Proceedings of the 4th Annual South-East European Doctoral Student Conference

Proceedings of the 4th Annual South-East European Doctoral Student Conference

Edited by
G. Eleftherakis
S. Hannam
E. Kalyva
A. Psychogios

ISBN 978-960-9416-00-9
978-960-9416-02-3

ISSN 1791-3578

Copyright: July 2009

© SOUTH-EAST EUROPEAN RESEARCH CENTRE

24, P. Koromila str., 54622, Thessaloniki, Greece
Tel: +30 2310 253 477
Fax: +30 2310 234 205
http://www.seerc.org

Preface

The 4th Annual SEERC Doctoral Student Conference (DSC2009) took place on July 6th and 7th in Thessaloniki, Greece, and was organised by the South-East European Research Centre.

The aim of the conference was to initiate an exchange of knowledge between young researchers and to help establish a network of scholars currently undertaking research in South-East Europe. Having identified academic isolation as a problem that many doctoral students face today, SEERC aims to bring researchers together for establishing collaborative links between disciplines, for testing the ground for innovative ideas and for engaging the wider academic community.

Building on the success of the past three conferences, this year's conference attracted a large number of submissions resulting in 80 presentations of both papers and posters. The audience of the conference expanded beyond the boundaries of South East Europe and we had presentations from UK, France, Germany, Holland, Slovenia, Hungary, Poland and as far as Ukraine, confirming the need for Doctoral Students to come together, discuss their experiences and gain external feedback to their work as well as listen to the progress and methodology of fellow PhD candidates.

The event commenced with welcome speeches from Dr. Iraklis Paraskakis (Director of SEERC Doctoral Programme and Chair of DSC2009) and Mr. Nikos Zaharis (SEERC Director) and was followed by the keynote speech, given by Professor Lila Leontidou of the Hellenic Open University. The title of the keynote speech was "*Shifting Boundaries of Europe: Ambivalence and Fluidity of Mappings and Cultural Identities of Modernity and Post-Nationalism*"

The scope of the conference was, again, multi-disciplinary spanning throughout the areas in which SEERC is doing active research and therefore it was divided into four parallel sessions:

- Enterprise and Regional Development
- Information and Communication Technologies
- Governance Politics and Society
- Risk Well Being and Cognition

There were just over 130 submissions and of these 67 were accepts as full papers and 13 for poster presentations. The full papers were divided as follows:

- 21 for the Enterprise and Regional Development track
- 17 for the Information and Communication Technologies track
- 19 for the Governance Politics and Society track
- 10 for the Risk Well Being and Cognition track

One of the main objectives of the conference has been to provide an opportunity for PhD students to receive advice from experts in their chosen field of research. This would not have been accomplished without the participation of the invited discussants. The list of the discussants according to the research track is as follows:

Enterprise and Regional Development

- Prof. Nora Karassavidou, (The School of Economic Sciences of the Faculty of Law, Aristotle University of Thessaloniki, Greece)
- Dr. Efi Nikolaidou, (Business Administration and Economics Department, CITY College, International Faculty of the University of Sheffield, Greece)
- Dr. Konstantinos Priporas, (Department of Marketing & Operations Management, University of Macedonia, Greece)
- Ms. Niki Glaveli, (The School of Economic Sciences of the Faculty of Law, Aristotle University of Thessaloniki, Greece)
- Dr. Ioannis Krassas, (Business Administration and Economics Department, CITY College, International Faculty of the University of Sheffield, Greece)

Information and Communication Technologies

- Dr Ilias Sakellariou (Department of Applied Informatics, University of Macedonia, Thessaloniki),
- Dr Kostas Dimopoulos (Dept of Computer Science, CITY College, International Faculty of the University of Sheffield, Greece)
- Ass. Prof. Panagiotis Bamidis, (Medical School, Aristotle University of Thessaloniki, Greece)

Governance Politics and Society

- Prof. Lila Leontidou, (School of Humanities, The Hellenic Open University, Greece)
- Prof. Ralph Negrine, (Department of Journalism Studies, The University of Sheffield, UK)
- Dr. Lia Papathanasiou, (English Language Support Unit, CITY College, International Faculty of the University of Sheffield, Greece)
- Prof. Nikos Hatzipantelis (School of Political Sciences, Aristotle University of Thessaloniki, Greece)
- Dr Sotiris Serbos (Democritus University of Thrace, Greece)
- Dr. Kalliopi Chainoglou (Department of International European Studies, University of Macedonia, Greece)

Risk Well Being and Cognition

- Dr. Angelos Rodafinos, (Department of Psychology, CITY College, International Faculty of the University of Sheffield, Greece)
- Dr. Vassilis Barkoukis, (Department of Physical Education and Sport Science, Aristotle University of Thessaloniki, Greece)
- Dr. Lambros Lazouras, (Department of Psychology, CITY College, International Faculty of the University of Sheffield, Greece)

- Dr. Suzzie Savvidou (Department of Psychology, CITY College, International Faculty of the University of Sheffield, Greece)
- Ms Aimilia Ypsilanti, (Department of Psychology, CITY College, International Faculty of the University of Sheffield, Greece)

SEERC would like to thank all the above named discussants for accepting our invitation and providing their valuable feedback to the PhD students that made their presentations.

Finally, I would like to also thank the members of the Organising Committee, the authors of the papers, all the presenters and participants and our colleagues at SEERC that contributed in making DSC2009 a successful event. We are looking forward to the announcement of the 5th conference.

Iraklis Paraskakis
Director of SEERC Doctoral Programme and Chair of DSC 2009 Conference

Organising committee:

Chair

Dr. Kostas Dimopolous

Committee:

Mr. Andrej Luneski
Dr. Fillipos Proedrou
Ms. Dialecthi Fotoplouo
Ms. Athanasia Halatzouka

Scientific Committees:

- Enterprise, Innovation and Development

Scientific Committee Chairs

Prof. Lenny Koh, *The University of Sheffield, UK*
Dr. Alexandros G. Psychogios, *City College – International Faculty of the University of Sheffield, Greece*

Scientific Committee

Dr. Kostas Alexandris, *Aristotle University of Thessaloniki*
Prof. Elena Antonacopoulou, *The University of Liverpool, UK*
Prof. Tommy Choy, *The Hong Kong Polytechnic University*
Prof. John Cullen, *The University of Sheffield, UK*
Dr. Dimitris Dadakas, *City College - International Faculty of the University of Sheffield, Greece*
Prof. Tsvetan Davidkov, *St Kliment Ohridski Sofia University, Bulgaria*
Dr. Denise Fletcher, *The University of Sheffield, UK*
Dr. Dimitris Georgarakos, *Goethe University, Frankfurt, Germany*
Prof. Antonis Makredemtres, *The University of Athens, Greece*
Dr. Efi Nikolaidou, *City College - International Faculty of the University of Sheffield, Greece*
Dr. Eytichis S. Sartzetakis, *University of Makedonia, Greece*
Dr. Leslie T. Szamosi, *City College - International Faculty of the University of Sheffield, Greece*
Prof. Demetris Vrontis, *University of Nicosia, Cyprus*
Dr. Alexandros Zangelidis, *University of Aberdeen, UK*

- Information & Communication Technologies

Scientific Committee Chairs

Dr. Konstantinos Dimopoulos, *City College - International Faculty of the University of Sheffield, Greece*
Dr. George Eleftherakis, *City College – International Faculty of the University of Sheffield, Greece*

Scientific Committee

Dr. Costin Badica, *University of Craiova, Romania*
Prof. Tudor Balanescu, *University of Pitesti, Romania*
Dr. Panagiotis Bamidis, *Aristotle University of Thessaloniki, Greece*
Prof. Nikos Basiliades, *Aristotle University of Thessaloniki, Greece*
Prof. Vera Behar, *Bulgarian Academy of Sciences, Bulgaria*
Prof. Betim Cico, *Polytechnic University of Tirana, Albania*
Dr. Antony Cowling, *University of Sheffield, UK*
Prof. Bekim Fetaji, *South East European University*, *FYROM*
Dr. Stylianos Hatzipanagos, *King's College London*, *UK*
Prof. Yannis Manolopoulos, *Aristotle University of Thessaloniki, Greece*
Prof. Roumen Nikolov, *University of Sofia, Bulgaria*
Dr. Miguel Nunes, *University of Sheffield, UK*
Dr. Ilias Sakellariou, *University of Macedonia , Greece*
Prof. Demetrios Sampson, *University of Piraeus, Greece*
Dr. Anthony Simons, *University of Sheffield, UK*
Prof. Demosthenes Stamatis, *Alexander TEI, GREECE*
Prof. Athena Vakali, *Aristotle University of Thessaloniki, Greece*
Dr. Ana Vasconcelos, *University of Sheffield, UK*
Dr. Anca Vasilescu, *University of Brasov, Romania*
Prof. Katerina Zdravkova, *University of Ss Cyril & Methodius, FYROM*

- Governance, Politics and Society

Scientific Committee Chairs

Professor Andrew Taylor*, University of Sheffield*
Sara Hannam, *City College – International Faculty of the University of Sheffield*

Scientific Committee

Professor Vasile Boari*, Babes-Bolyai University, Romania*
Agnes Brunet, *City College - International Faculty of the University of Sheffield, Greece*
Professor Simon Bulmer, *University of Sheffield*
Professor Theodore Chadjipandelis, *Aristotle University of Thessaloniki*
Professor Bob Deacon, *University of Sheffield*
Dr Christos Frangonikolopoulos*, Aristotle University of Thessaloniki*
Sara Hannam, *City College - International Faculty of the University of Sheffeild, Greece*
Professor Jan Horwath, *Univeristy of Sheffield*
Associate Professor Asteris Huliaras, *Harokopion University of Athens*
Professor Alison James, *University of Sheffield*
Professor Lila Leontidou, *Hellenic Open University*
Professor Ralph Negrine*, University of Sheffield*
Dr Lia Papathanasiou, *City College - International Faculty of the University of Sheffield, Greece*
Associate Professor Flore Pop*, Babes-Bolyai University, Romania*

Assistant Professor Radmila Popovic, *University of Belgrade*
Dr Nicola Rehling, *Aristotle University of Thessaloniki*
Assistant Professor Nicos Sifakis, *Hellenic Open University*
Professor Dimitrios Sotiropoulos, *University of Athens*
Dr Paul Stubbs, *Institute of Economics, Zagreb*
Professor Andrew Taylor, *University of Sheffield*
Dr Natalia Vlas, *Babes-Bolyai University, Romania*

- Risk, Cognition and Well-being

Scientific Committee Chairs

Professor Richard Eiser, *University of Sheffield*
Dr Efrosini Kalyva, *City College – International Faculty of the University of Sheffield*

Scientific Committee

Dr. Klio Semoglou, University of Western Macedonia
Dr. Demosthenes Panagiotakos, *Harokopeio University*
Dr. George Sideridis, *University of Crete*
Dr. Maria Georgiadi, *University of Crete*
Dr. Barkoukis Vassilis, *Aristotle University*
Dr. Tsorbatsoudis Haralambos, *Aristotle University*
Dr. Polyhronopoulos Evaggelos, *Harokopeio University*
Mr. William Myers, *University of Wolverhampton*
Dr. Angelos Rodafinos, *City College - International Faculty of the University of Sheffield, Greece*
Dr. Ana Vivas, *City College - International Faculty of the University of Sheffield, Greece*
Dr. Susie Savvidou, *City College - International Faculty of the University of Sheffield, Greece*
Dr. Martha Kelpi, *City College - International Faculty of the University of Sheffield, Greece*
Dr. Lambros Lazuras, *City College - International Faculty of the University of Sheffield, Greece*
Dr. Efrosini Kalyva, *City College - International Faculty of the University of Sheffield, Greece*

Table of contents

INFORMATION AND COMMUNICATION TECHNOLOGIES

ENTERPRISE, INNOVATION AND DEVELOPMENT

European Agriculture: A Aualistic Model Versus Endogenous Development

Antonio Sortino

Department of Biology and Agro-Industrial Economy, University of Udine, via delle Scienze 208, 33100 Udine, Italy, ***antonio.sortino@uniud.it***

Through this article, we aim to create a theoretical basis necessary to undertake the study of the dualistic model of agriculture (modernity/tradition), such as the European model of agriculture, and the hypothesis of its endogenous development.
We will start our paper studying the transformation of the agriculture from the rural society to the industrial society. The agriculture in the industrial society, unlike the agriculture at the time of rural society, is characterized by exogenous development paths dependent from the industrial sector. We will study the theoretical roots of the exogenous development and, consequently, two scholars of the modernisation: Schultz and Rossi-Doria. After that, our study will focus the roots of the endogenous approach to the rural development stressing the paradigm of the rural endogenous development. The paradigm refers to traditional agricultures and we will extend it to the modern agricultures. Finally, we will discuss of the neoclassical economics roots of the endogenous rural development.

Keywords. Rural endogenous development, Modernised agriculture, Traditional agriculture, European agriculture

1. Introduction

The European model of agriculture, defined for the first time on Agenda 2000, refers to a multifunctional model of virtuous agriculture that is progressively substituting the European productivist model of agriculture, that, in the recent past, has been supported by the Common Agricultural Policy. The European model of agriculture comprises heterogeneous realities and economic actors characterised by different local conducts. Schematizing this model, we can distinguish two agricultural typologies: modernised agriculture and traditional agriculture.

Modernised agriculture has integrally accepted technologies, inputs and more in general industrial values. It reflects faithfully the productivist model and at the same time it is a big producer of negative externalities (erosion of biodiversity, loss of agricultural landscapes, loss of traditional peasant culture, etc.).

By contrast, traditional agriculture typology encompasses those fragile and less favoured areas where the paradigm of modernisation could not be applied effectively, efficiently and profitably. For such reasons, traditional agriculture is firstly based on local elements, such as peasants' knowledge, biodiversity and traditional agricultural techniques. Therefore, traditional agriculture is virtuous: in fact, it unconsciously protects biological and cultural local elements, at the base of those agricultural activities, which are socially and environmentally sustainable.

The exogenous development of the agricultural sector, according to Slee (1993) and Lowe (2006), has a strong theoretical base that derives from the liberal, keynesian and marxist approaches. Moreover, the agricultural economics is historically connected with the study of the agricultural productivist model. Instead, the rural-sociology literature, in the last 20 years, is more interested to the less favoured areas. We refer, in particular, to the so-called paradigm of the "rural endogenous development" (Ploeg van der, 1993).

In both literatures, sociological and economic, the study of the dualistic agriculture is missing.

Through this article, we aim to create the theoretical basis necessary to undertake the study of the dualistic model (modernity/tradition) of agriculture such as the so-called European model of agriculture.

For this reason we will study the transformation of the agriculture because the economic development from the rural society to the industrial society (§1 and §1.2). The agriculture in the industrial society is characterized by exogenous development paths dependent from industrial elements and values. We will study the theoretical roots of the exogenous development (§2) and two of the main scholars of the modernisation in agriculture: Theodore W. Schultz (§3) and Manlio Rossi-Doria (§4). After that, our study will focus the roots of the endogenous approach to the rural development stressing the paradigm of the rural endogenous development (§ 5 and 6). This paradigm refers to less favoured areas and it will be extend to the modern agricultures. In this way we have created a new paradigm that comprises the two typologies of agriculture, modern and traditional. Finally, we will discuss of the neoclassical economics roots of the endogenous rural development (§ 7).

2. Agriculture in the economic development

As is commonly known, the economic growth has determined a sequence of different societies: first the rural one, later the industrial one, and today the post-modern one. The economist Colin Clark realized such phenomenon noticing a direct correlation between the growth of the pro capita income and the relative diminution of the weight of the primary sector inside the economic system (Clark, 1940). Logically, the Three-sector hypothesis of Clark foresees a second phase of development characterized by the prevalence of the tertiary sector.

Such quantitative analysis has some limits. For instance, it does not emphasize the qualitative transformations within the society. In fact, the rural society, the industrial

one and the post-modern society are qualitatively different because they were transformed to image of the dominant economic sector (in the temporal order: agricultural, industrial and tertiary sector). The social transformations, produced by the passage from one dominant sector to another, do not concern only production and exchange relations but the whole society personal relationships, languages, shared values, aesthetics, etc. The centrality of an economic sector is therefore evident in its ability to transform and to make itself similar to its surroundings, including the agriculture sector (De Masi, 2003).

Hence, a traditional agriculture has existed in the rural society period, a modernised one has existed in the industrial society period and we expect to have a new and different agriculture in the post-modern period. In spite of the radical changes, imposed from outside and consequently called exogenous ones, agriculture has got its own peculiarities that it keeps almost by force of habit. An example can be seen in the existence of a familiar productive structure, not always dominant but persistent, with a scale and an intensity that appear minute compared to other sectors. Unlike industrial activities, agricultural activity does not contribute to environmental deterioration and, in some cases, it even prevents the environment from being deteriorated. The environmental-economics literature indicates not deteriorating the environment as a positive externality. Some valid examples of positive externalities can be: production of salubrious and high-quality food, production of landscape, protection of environment and preservation of genetic local resources.

2.2 The agriculture from the rural society to the industrial society

The Three-sector hypothesis fails to explain the qualitative transformations of the society and its agricultural sector because the economic development. Such qualitative transformations, from the rural society to the industrial society, will be discussed in this paragraph.

Rural society mostly reflected agricultural sector in its practices, shared values, recognized models and relations: patriarchal family, good neighbourly practices, home place coinciding with workplace, conviviality and production of private consumption. They all encompassed and reproduced agricultural sector values, that were consequently extended to daily practices. In rural society, agriculture was not only the dominant economic activity, but represented the fundamental element of the entire economic system (Francovich and Mazzi, 1974).

The beginning of industrial processes, financed by the so-called primitive accumulation of capital, has progressively imposed new values created by the French Enlightenment and the Anglo-Saxon Pragmatism, such as rationalism, mechanism, secularism, etc. (De Masi, 2003). Industrialization practices, that have been entirely moulded by such new values, have initially ruled within industries; for instance with the standardization of products and processes, the specific qualification of the operators, the economy of scale, the myth of efficiency and the maximization of productivity, etc. (De Masi, 2003). When industrial practices have over-flown from manufactures, extending towards the core of social organizations and of civil living

(hospitals, prisons, schools, universities, army and maybe churches), imposing values, relations and "modern" rhythms, society has been eventually industrialized.

Therefore, modernisation is a combination of all those socio-economic and cultural changes originated from industrialized processes. These processes obviously have not left out the agricultural sector. Indeed, "farm" is progressively transformed into "industry" (Hardt and Negri, 2002): the integration of industrial inputs (industrial chemistry, seed production, mechanics, etc.) has increased together with the overall weight of the agro-industrial system. Besides, the final output is not so different from any other industrial product, within the seasonal and perishable limits of agricultural products. As is commonly known, the so-called commodities are defined by a high production, by relatively low production costs and by the almost total separation from territorial characteristics (altimetry, climate, local genetic variability, etc.).

In the final analysis, the modernised model of agriculture is based on three characteristics peculiar to the industrial sector: concentration, intensification and specialisation (Arnalte and Ortiz, 2006).

Evaluating positive aspects of the industrialization of agriculture is quite controversial. However, it cannot be denied that the industrialization of agriculture has sorted the millenary problem of food sustenance, at least in European Community, United States and Japan, even creating considerable agricultural surplus. But in the last decade, the limits of this agriculture have clearly exploded. Indeed, modernised agriculture is the cause of negative externalities and at the same time it cannot guarantee food security as it is testified by a decade of food scandals (e.g. bovine spongiform encephalopathy). Last but not least, another element of crisis in modernised agriculture is the so-called agricultural squeeze (squeeze of agricultural profits) essentially linked to the structural increase of variable costs and in particular of the energetic ones (Arnalte and Ortiz, 2006; Ploeg van der, 2006).

Besides, modernisation is not spread equally on all agricultural areas. In fact, some agricultural areas were excluded from modernisation because they were intrinsically incompatible with exogenous technologies. Modernised processes in such areas, like mountainous ones, have not found an environment suitable to their introduction. In these difficult areas, we can often assist to the development of an obstinate and involuntary tendency to keep together biological and cultural elements, based on local productive activity. The agriculture developed in this context is proved to be multifunctional and also a producer of goods and services for public use (mostly landscape and biodiversity).

Therefore, modernisation has created, both directly and indirectly, a dualistic agriculture between modernity and tradition such as the European model of agriculture. Summarising: modernised agriculture presents very high peaks of physical productivity, but lacks of socio-environmental inputs, due to the negative externalities produced. On the contrary, traditional agriculture is an unconscious producer of positive externalities. However, its decline has been caused by economic inefficiency which represents its weak point (at least as long as it is possible the monetization of the positive externalities to overcome the so-called "market failure".)

3. The theoretical roots of the exogenous development

According to Slee (1993) and Lowe (2006), there is a strong theoretical base for the rural exogenous development. The economic science, independently from the different typology of approach: 1) liberal, 2) marxist or 3) keynesian, has always considered rural development as a dependent development from external agents to the agricultural sector (Slee, 1993; Lowe, 2006).

1) In the liberal approach, agricultural sector is important for the economic development of the whole economic system. For example, in the models analyzed by Slee from the standard economy, the agricultural sector is seen as: a) a provider of food at non-inflationary prices; b) a purchaser of industrial inputs, c) a source of investment capital, d) a source of foreign exchange earnings to support the development process (Slee, 1993). Within the liberal approach the attention is given to the means by which modernization can be speeded up by means of credit tools, new technologies and infrastructures (Slee, 1993). Such changes are all likely to increase the extent of external control and incorporate rural areas more fully into national and global markets (Slee, 1993).

2) Not differently from the liberal approach, marxist approach considers the rural areas incapable of an endogenous and self-centered development. Those areas are <<viewed as enslaved beneath traditional rules, living an undignified, stagnant and vegetative life>> (Slee, 1993). The exogenous development of the rural areas should lead to the replacement of self-sufficiency practices with export-oriented crop production (Slee, 1993).

3) In a recent intervention, Lowe (2006) analyzes the keynesian approach to the development of rural areas. After the Second World War the principal economic problems of Europe, and particularly of its less favoured rural areas, are linked to the unemployment, poverty and the low quantities of circulating capital (Lowe, 2006). The keynesian solution to these problems is finalized to increase the rate of saving through investments and, particularly, through public interventions finalized to fill the so-called saving gap between reach and poor areas. This has happened, in particular way, through the transfer of economic resources to the poorest regions to create occupation, to increase the incomes and to stimulate the demand of industrial good (Lowe, 2006).

The model of application of the keynesian theory has entered in crisis in the years 70 of last century for different reasons: the continued industrialization of agriculture came up against the saturation of domestic markets, against ecological limits and against a greatly diminished capacity in the urban sector to absorb the surplus rural population (Lowe, 2006).

The liberal, marxist and keynesian approaches have different basis of departure: respectively the mechanisms of market (liberal approach) and the government planning (marxist and keynesian approaches). The political vision of the rural areas is however identical: the rural areas are considered marginal areas primarily destined to the production of food stuff by to furnish to the urban centers. Any hypotheses of economic growth, also for less favoured areas, is linked to such functions. Thus, the

development of rural areas can happen only thanks to exogenous processes of specialization, concentration and integration.

4. Theodore W. Schultz and its paradox

Among the main scholars of the modernization of the agricultural sector we intend to underline the thought of T. W. Schultz. According to Schultz, even though traditional agriculture has a low production level, it cannot produce more, consequently it cannot contribute valuably to the economic development. On the contrary, since modernised agriculture produces in plenty, it can contribute to the economic development more than it has thus far (Schultz, 1964).

Schultz indicates the solution of such a paradox: it is necessary to modernize traditional agriculture in developing countries through exogenous inputs (such as technology) mainly through investments on education and farmers' training. Hence, according to Schultz and paraphrasing Schumpeter, it is necessary to "destroy" traditional agriculture in poor countries to "create" economic growth.

The analysis of Schultz, for some aspects, is ingenuous (Calcaterra, 1995). Following Schultz, the traditional agriculture of the developing countries has the same potential of production of the modern agriculture of the developed countries (Rao, 1986; Calcaterra, 1995). The only difference is that the modern agriculture has had the chance to develop (Calcaterra, 1995).

Therefore, the solution proposed by Schultz is to transform the traditional agriculture in the modern one through injections of capital and technologies. This model has, at least for economists, two "seductive" aspects (Calcaterra, 1995). For first: the model results congenial for its conformity to the general principles of the economic growth and, particularly, for the use of capital and technology in a context governed by the prices (Calcaterra, 1995). For second: a model that ties the production growth to the increase of the factor productivity gives the perspective of an indefinite expansion of crops. This could minimize the problem of the natural resources and particularly of the land (Calcaterra, 1995).

5. Manlio Rossi-Doria and the actualization of his thought

Dualism in the European model of agriculture is a bequest of modernization processes in agriculture. Such dualism has already been identified and analysed by Rossi-Doria in 1958, who described <<la polpa e l'osso>> (beef and bone) in the agriculture of Southern Italy (Rossi-Doria, 1958). According to Rossi-Doria, beef is represented by modern agricultures insisting on fertile and productive lands, while bone is made up of mountainous areas, latifundium areas and, in general, areas characterised by unfavourable social and/or pedo-climatic conditions.

Rossi-Doria's lesson about the dualistic agriculture of beef and bone can be summarised in the following way: different realities must correspond to different

public policies (Rossi-Doria, 1958). This principle has not been applied yet at the European governance (De Benedictis, 2002).

For better intending the meaning of the distinction of Rossi-Doria it is necessary to know the context and the dominant scenery of agrarian policy in which he operates. The dominant paradigm in the agriculture after the second Postwar is the paradigm of "productivist modernization" (De Benedictis, 2002). With this concept, De Benedictis means the whole processes useful to increase the productivity of the production factors and, in last analysis, to increase the economic growth of the agricultural sector (De Benedictis, 2002).

We are in presence, according to De Benedictis, of a paradigm shift from the "productivist modernization" to the new concept of "qualitative modernization" of our times. With this last concept, De Benedictis refers to a new interpretative paradigm of the multifunctional agriculture in Europe that comprises several characteristics such as the environmental protection, salubriousness and high-quality food goods, quality production processes. The "qualitative modernization" and the multifunctionality of the European model of agriculture shows in its main functions (Arzeni, Esposti, and Sotte, 2001):

a)Food functions: the function of producing quality food in accordance with food safety both from a quantity and quality perspective.

b)Environmental functions: the function of producing positive externalities such as protection of landscape and biodiversity, reduction of all negative externalities produced by agriculture.

c) Rural functions: this function refers to an agriculture that can support the sustainable development of rural communities, in compliance with history and local culture.

De Benedictis in his article tries to actualize the thought of Rossi-Doria. Regarding the beef, De Benedictis denotes its weak points in terms of sustainability. The author cites a recent study (INEA 2001) that highlights how the expansion of the irrigated land has determined: i) drop in ground-water level; ii) the increase of agricultural land salinity; iii) quality water degradation. Not reversible phenomena connected with the increase of irrigation costs.

Regarding the bone and the less favoured agricultures, some relevant fact indicates, in the south of Italy, the qualitative sensibility of this typology of agriculture: a) the coexistence of "sustainability" of the productive activities and "success" in terms of profits and remuneration for work in some less favoured areas; b) the number of organic farm and the size the organic agricultural land; c) the success of institutional innovations such as the Leader programs particularly in the less favoured areas.

6. Original sources of endogenous development

According to Slee and Lowe, previous rural development policies have been influenced by different liberal, marxist or keynesian theoretical approaches (Slee, 1993; Lowe, 2006). These approaches utilize different analytical instruments, but share

the common goals: industrialization, specialization and integration to promote rural and dependent exogenous development paths for rural areas (Lowe, 2006)

The endogenous approach to the rural development, and alternative to liberal, marxist and keynesian policies, has arisen in Europe, according to Lowe, in the 80s of last century from a series of practical experiences of alternative development.

With the concept of "endogenous approach to rural development", Lowe indicates a sustainable development "led from inside" and based on local elements (natural and cultural). Lowe identifies four original sources (that are summarised just below) of endogenous development, which have represented an opposition "in daily practices" to liberal and keynesian paradigms (Lowe, 2006).

First source: the progress of some rural and semi-urban areas in the 70s and 80s with internal dynamics unknown until then. Some researches have tried to identify the key-success of these regions (flexible work markets, dynamic nets of small companies, community market, informative and technological atmosphere, local institutions regulations, rules and customs) and if the experience could be repeated in other regions (Iacoponi, 1993; Ploeg van der and Long, 1994). The most sensational example in Europe has concerned Italy after the '60 in the last century. Several areas located in the north-east and in the centre of Italy (the so-called NEC regions), characterised by a rural economy and by a society based on sharecropping, far from big industrialised centres, have experimented a rapid economic dynamism rooted in the rural and based on the nets of industrial and agro-industrial districts (Sotte, 2006). In this case, the development model, which has been less-sustainable in environmental terms, is alike to the endogenous development model, because it is based on local elements such as human capital, locally shared values, technological and informative atmosphere.

Second source: movements and regional agencies that, trying to overcome the failures of liberal and keynesian policies, have attempted to promote in local communities forms of development less dependent from exogenous resources. In particular, local action comprised rural diversification and support to local productive activities. Such movements and agencies can be found in the marginal areas of different European countries: Ireland, France, Scottish Highlands, Iceland and rural Wales, Austrian and Italian mountain communities and groups for the development of the villages in North Sweden (Lowe, 2006)

Third source: the debate on rural sustainability, including Cork declaration, that has attempted to match the economic development of rural communities with the environment protection and with the quality of life in rural areas, without which it is difficult to keep sustainable economic activities and employment; to avoid the depopulation of rural areas and to protect the environment (Lowe, 2006).

Fourth source: a boost to "self-reliance" and to local self-organization's abilities promoted by environmentalists' movements and development radical activists, that work at close contact with groups of particularly marginalized areas. These ones have transferred Schumacher's thought of "small is beautiful" to the community economics field, trying to reassert the local control on economic activities, to protect the community from the homologous effect of globalization (Lowe, 2006). Lowe lists some examples of such kind of development based on the concept of "small is

beautiful", present in the outskirts of South Spain, West Ireland, North Sweden (Lowe, 2006).

The exogenous sources: In order to partially integrate Lowe's ideas, we refer to two exogenous sources, that has advocated endogenous development in European rural areas. We will now discuss the combined action of the two phenomena: a) changes of the European society and the food style of its citizens, that have stimulated the endogenous development in rural areas; in particular we refer to the consumers' demand, that has moved from standardized products to a diversified range of customised and niche products (Sotte, 2006; Pieroni and Brunori, 2007); b) the contribution of new technologies, also IT, that have allowed small-middle companies to reach competitive levels (through external network economies), once reached only by large-sized companies (Sotte, 2006).

7. A sociological paradigm for the dualistic model of agriculture

The paradigm of the rural endogenous development has been elaborated in the last decades of 1900, within sociological and agricultural economics literature (Iacoponi, 1993; Ploeg van der, 1993; Slee, 1993). It attempts to interpret the birth of self-supported development processes and in particular rural areas excluded by modernised processes. This paradigm has developed further and its most recent reinterpretation has been renamed "New paradigm of rural development" (Ploeg van der, 2006), but its basic assumptions, that will be described now, are the same.

The development model taken from this paradigm is "self-centred", because mainly based on (but not exclusively) endogenous resources, traditional techniques and local knowledge, such as agricultural biodiversity, locally spread techniques and local informative atmosphere. At the same time, the development model is "conservative"; for instance, local elements, being at the base of the productive process, are preserved, protected and exploited to that end (Iacoponi, 1993; Sortino, 2007). The paradigm is based on the three following starting assumptions: 1) the local choice of the development options; 2) the check or local monitoring of the development process; 3) local appropriation of development assets (Slee, 1993).

Exogenous elements, as the so-called modern inputs, essential to the attainment of an efficient economics, are not uncritically refused but inserted, once they have been deconstructed and reconstructed, in accordance with the "local style of farming" (Ploeg van der, 1993). That means that a technology can be redesigned and modified instead of being applied integrally: its main components can be rearranged in order to distinguish it from the original exogenous model to adjust it to the local needs (Ploeg van der, 2006). Then, the last aspect is one of the paradigm's crucial points, able to distinguish the endogenous development from a simple endogenous preservation and from an exogenous development. The endogenous conservation, namely the attempt to preserve the countryside, managing it as it was a museum, indeed does not entail any insertion of external components into the local history or tradition, which in some contexts often transforms into folklore. The exogenous development, that has affected the most profitable agriculture, vice versa imposes a strong and fast change of the

natural environment through a copious integration of chemistry and technology, to increase the productivity of land, work and capital factors.

The paradigm of endogenous rural development, as we have seen, almost rises in the context of rural areas excluded by modernised processes. At the light of the European model of agriculture that we have defined dualistic, this paradigm has been refined and extended to modernised agriculture. It is reckoned that the latter agriculture typology, in the attempt of restoring the lost virtuosity, accepts the even partial return of virtuous elements from tradition (for example good agriculture practices, use of autochthonous seeds, crop rotations etc.), trying obviously not to come out from sustainable economic field.

Therefore, two patterns of endogenous rural development are set up. They will be summarized more clearly here below.

The first pattern, already analysed mainly in sociological literature, concerns the development of fragile agricultures (Ploeg van der, 1993). It is focused on local resources and it is supported, in order to achieve economic sustainability, by the introduction of exogenous elements duly accustomed in accordance with environment and "local feeling".

The second pattern of endogenous development concerns the qualitative evolution of modernised agriculture. With the perspective of converting environmental sustainability of productive activities, this agriculture typology prepares for the acceptance of past virtuous elements' return, duly readjusted to the new productive structure. In a preceding paper by the authors (Sortino and Chang, 2007) some indicators of the return of techniques have been underlined. For example: a reduction of synthesis products which are used in agriculture, expansion of organic agriculture; increasing the demand of traditional, typical and organic products.

The endogenous development pattern that concerns modernised areas is related to a sort of quantitative decrease in production. It does not always match with worse economic performances, since products from tradition mostly fetch higher prices in advanced economies. Therefore, in such development context of quantitative decrease, the term "post-productivism" is a perfect synonym for "post-industrial".

8. There are any neoclassical economics roots for the endogenous development?

The theoretical inspiration of this paragraph originates from the analysis of some researches made by Italian agricultural economists who, a few years ago, have studied the possible theoretical connections between the endogenous rural development paradigm and neoclassic economy (Iacoponi 1993; Romano, 1996).

The neoclassical economic theory comprises at its core the "endogenous growth theory", called also "new growth theory", born to fill some gaps from the tout court neoclassic growth theory elaborated mainly by Solow in the second half of '900. The main gap in Solow's theory was evident in its representation of country's growth

governed by exogenous elements: population growth rate and technical progress rate (Solow, 2000). According to this theory, every single country has reduced possibilities to be able to influence its own economic growth. In addition, the long term convergence of countries is conjectured at world level. A part from the initial conditions, world countries' economies should reach a similar stable state characterized, in particular, by the same gross domestic product (GDP) per capita (Solow, 2000).

The "new growth theory", without questioning any basic assumption of the neoclassical theory, considers that the economic growth of a country could also depend on endogenous lever like public and private investments on knowledge, education and training. In this way, the missing convergence is justified by the fact that, rich countries can invest big resources on human capital, consequently making their work-force more productive, whereas poor countries are obliged to renounce it. They do not have even the resources to guarantee primary education, that is of minor importance compared to the demand of satisfying other needs, even more essential (starvation and sub-nutrition in the countries of the Third World). Education, as we all know, in many countries, is considered to be a luxury.

The "new growth theory" aims to provide a descriptive scheme of reality (Pasinetti, 2000) and consequently of post-modern society and immaterial production, where knowledge, training, language, but also affection make more and more economic wealth (De Masi, 2003). Even the paradigm of endogenous rural development is particularly at its ease in the post-modern context. Such assertion is confirmed by the demand growth of those users enjoying landscape-environmental services, produced by endogenous agriculture as well as by the added value created by services included in food and agriculture assets (for example traditions, healthiness, environment protection) of the same agriculture. This appears as a good starting point to search for possible contact points between the two theories.

Our opinion is that both theories assign a fundamental role to "human capital" namely a growth and development force, as it is even corroborated by agricultural economics literature. Moreover, as Pasinetti says, the introduction of the concept "physical quantity" of human capital has occurred within the "new growth theory", <<without any apparent need to consider the logical foundations or the conditions under which it could quantitatively be represented>> (Pasinetti, 2000). All the more reason for validating this criticism in a context of endogenous rural development where traditional agricultural skills and "informal" knowledge, namely that one handed down from father to son (e.g. traditional cheese-making techniques), cannot be compared to "study hours" or "school years".

The differences between the "new growth theory" and the paradigm of endogenous rural development are instead several; many of these have been already identified and discussed in previous researches made by agricultural economists. We highlight the five main differences existing and we list them below.

A) The first difference is pretty obvious: the different meaning assigned to the terms "growth" and "development", in the literature of the last decades; indeed, the first term

mainly conveys the idea of quantity, while the second one refers to the idea of quality (life quality, food security, biodiversity etc.) (Romano, 1996).

B) In the new growth theory, there is no hypothesis of any development sustainability, something which is instead implicit in the theory of endogenous rural development (Romano, 1996).

C) Endogenous growth models have been formulated from simple assertions and deep simplifications (Solow, 2000). By contrast, the starting assertions of endogenous rural development cannot be easily summarized in simple mechanics to be inserted in models of economic growth.

D) The new growth theory is macroeconomic. However, in those rare cases when the endogenous rural development paradigm was applied, the analysis of the agricultural system and the rural system could not be expanded further (Romano, 1996).

E) The new growth theory does not entail, and does not aim to do it, the processes of adjustment of external elements into a non-modernized context or the processes of re-adjustment of traditional elements into industrialized agriculture that characterize, as we have seen, the two patterns of endogenous development identified on this article.

9. Conclusion

Our contribution is essentially theoretical and it has concerned the dualistic agricultural model such as the European agriculture and the hypothesis of its endogenous development. The distinction of the agriculture in traditional and modern is not new in literature: Manlio Rossi-Doria and Theodore W. Schultz, following different point of views, have already described agriculture activity in a similar way.

The theories that have been critically analysed in this paper take form from social sciences and, in particular, we have referred to the sociological paradigm of endogenous rural development, extending it to modernized and less virtuous agricultures. Moreover, we have created an economic background searching for possible contact points between the endogenous rural development theory and the so-called new theory of neoclassical growth, known in literature also with the name of theory of endogenous growth. The conclusion, in this last case, is that the tools offered by the neoclassic theory, which for its own nature promotes an exogenous and global development, are not the most suitable for the study of endogenous development phenomena that can entail traditional agricultural and, through the return of traditional techniques, modernized agriculture.

References

1. Arnalte, E. and Ortiz, D. (2006): ‘Tendenze dell’agricoltura spagnola. Difficoltà di attuazione di un modello di sviluppo rurale basato sulla multifunzionalità in agricoltura’, In: Cavazzani A. et al., Politiche,

Governance e Innovazione per le aree rurali, INEA, Collana Studi & Ricerche, Edizioni Scientifiche Italiane, Napoli.
2. Arzeni, A., Esposti, R. and Sotte F. (2001): 'Agriculture in Transition Countries and the European Model of Agriculture: Entrepreneurship and Multifunctionality', In: The World Bank, Šibenik-Knin and Zadar Counties: Framework for a Regional Development Vision, University of Ancona.
3. Calcaterra, E. (1995) 'Agricoltura e degradazione ambientale nel terzo mondo: un nodo cruciale', Agribusiness, Management & Ambiente, Vol. I, n.1.
4. Clark, C. (1940) The conditions of economic progress, London: Macmillan.
5. De Benedictis, M. (2002) 'L'agricoltura del mezzogiorno: <<la polpa e l'osso>> cinquant'anni dopo', La Questione Agraria n. 2.
6. De Masi, D. (2003) Il futuro del lavoro, fatica e ozio nella società postindustriale, Rizzoli, Milano.
7. Francovich, R. and Mazzi, M. S. (1974) Le campagne europee dopo il mille, Roma: La nuova Italia.
8. Hardt, M. and Negri, A. (2002) Impero: il nuovo ordine della globalizzazione, Milano: Rizzoli.
9. Iacoponi, L. (1993) 'Il distretto agro-industriale come modello di sviluppo endogeno', In: Panattoni A. La sfida della moderna ruralità, Pisa: Cnr-Raisa.
10. INEA (2001) 'Studio sull'uso della risorsa idrica, sulle produzioni agricole irrigate e sulla loro produttività', Working Paper, Roma.
11. Lowe, P. (2006) 'Concetti e metodi nelle politiche europee di sviluppo rurale, In Cavazzani A. et al., Politiche, Governance e Innovazione per le aree rurali, Napoli:INEA, Collana Studi & Ricerche, Edizioni Scientifiche Italiane.
12. Pasinetti, L. (2000) 'Critica della teoria neoclassica della crescita e della distribuzione', Moneta e Credito, n. 210
13. Pieroni, P. and Brunori, G. (2007) 'La (ri-)costruzione sociale del paesaggio nella campagna contemporanea: processi, problematiche, politiche per uno sviluppo rurale sostenibile, In Brunori G. et al.: La gestione del paesaggio rurale tra governo e governance territoriale, Milano: Franco Angeli.
14. Ploeg J.D. van der (1993) 'Dynamics and mechanisms of endogenous development: an introduction to the theme of the seminar, Camar conference, Chania.
15. Ploeg van der, J.D. and Long, A. (1994) Beyond modernisation: the impact of endogenous rural development, van Gorgun, Assen
16. Ploeg van der, J.D. (2006) Oltre la modernizzazione, processi di sviluppo rurale in Europa, Catanzaro: Rubbettino.
17. Rao, M. F. (1986) 'Agriculture in recent development theory', Journal of Development Economics, vol. 22, n. 1.

18. Romano, D. (1996): 'Endogenous rural development and sustainability: a European (non orthodox) perspective', 5th conference on Agriculture Food and Environment, University of Minnesota
19. Rossi-Doria, M. (1958) Dieci anni di politica agraria, Bari:Laterza.
20. Schultz, T. W. (1964) Transforming traditional agriculture, Yale University Press, New Alen.
21. Slee, B. (1993) 'Endogenous development: a concept in search of a theory', Camar Conference, Chania.
22. Solow, R. M. (2000): 'La teoria neoclassica della crescita e della distribuzione, Moneta e Credito, n. 210.
23. Sortino, A. (2007) 'L'inquadramento teorico del paradigma dello sviluppo rurale endogeno', Agribusiness Paesaggio & Ambiente, Vol. 10, n. 2.
24. Sortino, A. and Chang, M. (2007) 'The new endogenous agriculture. The return of traditional techniques and knowledge in the modernized agricultural economy', 18th ESEE conference, University of Life Sciences di Praga.
25. Sotte, F. (2006) 'Sviluppo rurale e implicazioni di politica settoriale e territoriale. Un approccio evoluzionistico', In Cavazzani A. et al., Politiche, Governance e Innovazione per le aree rurali", Napoli: INEA, Collana Studi & Ricerche, Edizioni Scientifiche Italiane.

Understanding the Role of Management Education in the Greek Context: Environmental Pressures and Reforms.

Dialechti Fotopoulou[1], Geoffrey T. Wood[3], Alexandros G. Psychogios[2]

[1] *SEERC, 17 Mitropoleos Str, Thessaloniki, Greece, di.fotopoulou@seerc.org*

[2] *City College, Business Administration and Economics Department, 13 Tsimiski Str, 54624 Thessaloniki, Greece, a.psychogios@city.academic.gr*

[3] *The University of Sheffield, Management School, 9, Mappin Str, Sheffield, S14DT, UK, g.t.wood@sheffield.ac.uk*

Business schools in Europe develop management education programmes by borrowing elements from the most influencing model of business education: the Anglo-Saxon type MBA degree. Nevertheless, it is argued that differentiation rather than isomorphism towards a perceived best practice in business school curricula can prove to be more beneficial and provide knowledge more relevant to management practice in different national contexts. This study seeks to investigate the role of Anglo-Saxon business education in promoting the dissemination of Anglo-Saxon/Liberal Market style business practices into South European economies, using the case of Greece. Framed by the literature on comparative capitalism, this study is based on a document study and in-depth interviewers of stakeholder groupings. It aims to provide further insights into the process of change in South European capitalism, the impact of management education on practice, and the extent to which the expansion of Anglo-Saxon style management education into the Mediterranean world provides the basis for emerging sets of complementarities underlying a protracted process of experimentation and change.

1. Introduction

A burgeoning body of literature suggests that business schools in Europe present similarities with the generalist Anglo-Saxon educational model. By considering that education is one of the channels of knowledge transfer to organizations and consequently one of the institutional drivers that shape organizational behavior, it is interesting to examine the role of Anglo-Saxon business education as a driver of Anglo-Saxon isomorphism in non Anglo-Saxon economies. Research on this area is quite limited and is mainly focused on Continental European countries while other countries like the Mediterranean ones are quite neglected. Therefore, by taking the

example of Greece, the aim of this paper is to examine whether management education in Mediterranean countries is indeed in a process of Anglo-Saxonization and to discuss the possible implications of such a process on management practice.

After an extensive literature review the underlying conclusion is that there are Anglo-Saxon influences in the field of management education due to various homogenizing pressures, especially in countries that are not highly developed economically, including the Mediterranean ones. Nevertheless, an unmodified isomorphism with the Anglo-Saxon model does not exist. National institutional characteristics hinder the homogenization of educational systems in different capitalisms and lead to local variations. Therefore, there is a need to further examine how this blending trend of local and Anglo-Saxon elements in management education affects the application of managerial practices in Mediterranean capitalisms.

The main consideration is that since various capitalisms have different institutional features that affect each other, they will probably have different educational needs. Hence, the usefulness of the generalist Anglo-Saxon education in a non-Anglo-Saxon business system is questioned especially in capitalisms with very distinct characteristics.

By focusing on Greece as an example of Mediterranean capitalism we conclude that it is a system that shares commonalities with Mediterranean countries but also has some distinctive features. It has been presented as a system that is highly externally dependent, state-led and characterized by dualism (few large and many small family-owned enterprises). These characteristics along with the high levels self-employment, the large unofficial economy and the unstable political situation are some of the institutional factors that affect significantly managerial practice in Greek organizations.

On the other hand, recent pressures mainly stemming from globalization, have emphasized the need of Greek organizations to move away from the traditional paternalistic management practices and this need has initiated a change towards a more strategic focus of managerial practices. Education is an important driver of transmitting such knowledge to firms. What is known by examining the general education system in Greece is that apart from certain improvements mainly initiated by the requirements of the Bologna Accord, it has responded slowly to environmental pressures and reforms. These slow responses have implications in the field of management education, which has actually started to develop in the 1990s.

What has not been extensively examined is how these new to the Greek reality management techniques are transplanted in Greek organizations and whether management education in Greece is indeed acting as a driver of Anglo-Saxon management isomorphism. For that reason an exploration of the structure and the content of management education in Greece as well as an examination of the managerial practices applied will provide an insight on what managers actually learn.

To conclude, this paper aims to raise the issue on whether management education in Greece is indeed following the influential Anglo-Saxon model of education as Mediterranean countries seem to do so, or it retains its own national peculiarities. Further research will provide evidence on that and will ultimately focus on whether managerial education acts as a driver of institutional isomorphism with the Anglo-Saxon prototype.

2. Understanding Education as an Institutional Driver

Education, and particularly the structure of the modern university, is regarded as a significant institutional factor in shaping what goes on in the modern complex environment (Hollingsworth, 2006). Management Education in particular is inevitably linked with the organization of work at four levels: at the individual level through personal participation in the training programmes, at the organizational level through company formal training programmes, at the inter-organizational level through formal training programmes sponsored by professional associations; and finally, at the societal level through the contents and procedures of the education systems, such as University curricula (Zeitz *et al,* 1999).

The importance of education as an institutional driver has been stressed out by Varieties of Capitalist (VOC) theorists. They examine internal relations within firms by incorporating most aspects of broader institutional analyses as well as structural and cultural factors (Brewster *et al,* 2008). The focus of VOC theories is to analyze institutional complementarities that lead to the formation of various capitalisms. It is argued that national contextual factors determine managerial practices as these factors shape a national "logic of action" (Rees and Edwards, 2006). They suggest that internationalization is indeed forcing organizations to change but this change is path dependent and reflects organizations' "institutional linkages" (Hall & Soskice, 2001; Whitley, 1999).

Cross-country comparisons in the VOC literature present various clusters of countries that have similar characteristics and therefore form various capitalisms. Education is one of these characteristics and when combining some of the most influencing models of capitalism in the VOC literature (Amable, 2003; Hall & Soskice, 2001; Whitley, 1999) we can come up with four main clusters in the European region regarding education and training (see table 2.1).

These four clusters follow the pattern of Amable's (2003) categorization. His study on education drifted towards country specific systems rather than heterogeneous (i.e., based on the standardization of the curricula, system flexibility, nature of vocational training – work or school based – and the availability of funding – public or private). The two extremes are traced between the "standardized" type of education in Germany and the Netherlands as opposed to the "differentiation" type of education in the USA and Canada. The other countries fall in between those two extremes.

An important outcome of Amable's (2003) study is that that both general education and vocational training impact on the interactions between labour markets and educational systems and that vocational training is a significant factor that differentiates countries. Differentiation is mostly attributed to the diverse institutional and regulatory environment of training in various countries. Firm size is also a significant actor, as this determines the financial capabilities of firms in investing on training. Finally, regarding education, the absence of competitive pressures in that field is, according to Amable (2003), the major reason for the lack of convergence.

LMEs/Compartmentalized/Market-Based	CMEs/Collaborative/Continental European
Education and Training Systems focus on General Skills Competitive Education System High Enrolment Rates in Tertiary Education Non-homogeneous secondary education system Low Strength of the Public Training System Weak National Vocational Training System Low Duration and Spend on Training	Education and Training focus on Industry or Firm Specific Skills Public Education System Emphasis on Secondary Education High Degree of Standardization High Strength of the Public Training System Intermediate to Strong National Vocational Training System Low to Intermediate Duration and Spend on Training
Social Democratic/Weak Vocational CME	**Coordinated Industrial District/Mediterannean**
Education and Training focus on Industry or Firm Specific Skills Public Education System High Public Expenditures on Tertiary Education High Aid to Students Limited Strength of the Public Training System Weak National Vocational Training System Lower than Average Duration and Spend on Training	Low expenditures for education, particularly tertiary education Weak education system Low enrolment rates Weakness in science and technology tertiary education High Strength of the Public Training System National Vocational Training System poor matching with employee demand Intermediate Duration and Spend on Training

Table 1. Models of Capitalism: Education & Training

In other words, while other institutional factors act as drivers of isomorphism between countries due to the existence of competitive pressures, this does not apply in the case of education, which is presumed as a field with limited competition. Nevertheless, evidence from the management education filed suggests that competitive pressures do exist. This is indicated by the fact that Universities nowadays both in Europe and the US have adopted a "global focus" in their operations (Beech, 2006). In addition, many scholars identify isomorphic tendencies with the Anglo-Saxon prototype in the field of management education (Mazza *et al*, 2005; Kipping *et al*, 2004; Hull, 2000).

Therefore, what needs to be further examined is whether there are significant Anglo-Saxon influences in the management education field and therefore to conclude on whether there are converging trends in business education. If management education is indeed in a process of Anglo-Saxonization, and given the heterodox tradition where different sets of countries have different characteristics and hence different educational needs, this raises questions regarding the relevance of an Anglo-Saxon management education paradigm in non-Anglo-Saxon business environments. In the discussion that

follows the level of convergence with the Anglo-Saxon model of management education and its possible implications are discussed.

3. Towards the Anglo-Saxonization of Management Education?

3.1 Business Schools in Europe and the process of Imitation

An issue that has received particular attention in academia is that business schools in Europe are becoming similar and close to the Anglo-Saxon type of education. Imitation seems to be stronger in countries that are not highly developed economically and especially in those in transition. In other words, in countries where institutions are more weakly embedded or more fluid. Those countries tend to replicate the practices of advanced countries and the US plays the most predominant role. Examples of these are the Mediterranean and the countries of Eastern Europe. (Usdiken and Yildirim, 2007; Kipping *et al,* 2004; Gemelli, 1998; Svetlicic and Cibron, 1996).

Despite their inherent structural differences in higher education, Mediterranean countries seem to follow similar Western models of education (Sultana, 2008). In particular, Mediterranean universities' trends towards privatization, entrepreneurship and massification reflect the situation in the northern and more economically developed economies. Universities in that region aim at establishing worldwide networks in various ways: they promote publications and attendance at conferences in order to strengthen the relationships with similar institutions world-wide (Sultana, 2008; Calleja, 2005). Overall, Mediterranean countries are found to be the main importers of transnational education in Europe (Van der Wende, 2001).

Managerial education has been affected by the general trend that exists in the higher educational sector of the Mediterranean towards modernization. Research has shown that although Universities in different countries of the Mediterranean are at different stages of development, they all seem to follow similar paths towards western structures and particularly Anglo-Saxon structures. Kipping *et al* (2004) have identified multiple "Americanizations of Management Education" in Mediterranean countries. They examined the interaction of American models of management education with France, Italy, Spain, and Turkey and found that -at least in terms of the content- American models influenced the educational systems of all four countries.

On the contrary, part of the literature suggests that business schools in Europe do differ. There are considerable differences both between US and European business schools as well as between business schools in various European countries. Differences between European and US business schools can be separated into three sets: *Institutional* (Culture, National Educational Drivers, Socio-economic organization, Language, Regulation, Standardization and Size), *Competitive* (Knowledge Transmission International Mindset, Governance and Values, Funding and Endowment, Corporate Links), and *Social Capital* Differences (Don Antunes and Howard, 2007; Ertl, 2006; Whitley, 2006; Javidan, 2005; Mazza *et al*, 2005; Huisman and Van Der Wende, 2004; Kipping *et al*, 2004; O'Hare, 2004; Hill and Houghton,

2001; Kumar and Usunier, 2001 Lam, 2000; Engwall and Zamagni, 1998; Engwall *et al*, 1996; Gemelli, 1996)

Despite the existence of heterogeneities and the fact that the interdependence of cultures, structures and practices hinder the comparison between distinct national models of Management education, there exists a point of convergence. As Ramirez (2004) concludes in his study on European Management Education, Training and Development in all countries examined is that despite the absence of a uniform set of practices, there seems to be a convergence toward Anglo-Saxon ones resulting from the pressures of international competition.

Overall, it is a fact that the development of US management programmes overseas is rather significant and that most business schools internationally offer –or at least is anticipated that they offer- MBA degrees (Pfeffer and Fong, 2004). In other words, although existing variations in the educational systems of various capitalisms prohibit the creation of duplicate management programmes, the expansion of Anglo-Saxon management programmes such as the MBA degree, serves as a channel through which management ideas are transferred (Mazza et al, 2005). Such a transfer might increase homogeneity between countries and since the main country of origin is the US, it probably has Anglo-Saxon influence

3.2 Questioning the Anglo-Saxon Management Education Model

According to Hall and Soskice (2001), the Anglo-Saxon educational model in tertiary education is quite competitive since it focuses on general skills. More specifically, the majority of educational programmes give much priority on the acquisition of general skills while industry-specific skills are quite neglected

This generality in educational programmes goes hand in hand with the management system applied in Anglo-Saxon countries as well as the pattern of work organization and employment relations. First of all, the strong academic orientation of the Anglo-Saxon education system hinders the generation of formal intermediate skills and qualifications among the workforce. This necessitates a hierarchical pattern of work organization. Moreover, managers at various levels of organizations are responsible for both leadership and administrative functions, giving more emphasis on leadership (Lam, 2000).

Manual and management functions are divided in the education system and therefore managers with an academic background in engineering cannot stay in the higher levels of hierarchy for long (Bykjeflot, 1998; Amdam, 1996). Rather, individuals with general credentials govern careers at the top of organizational hierarchies (Whitley, 1999). Moreover, the acquisition of a general management credential as the MBA allows for the high levels of manager mobility across firms as well as across sectors, as this is the post-war Anglo-Saxon pattern (Whitley, 1999). As Bykjeflot (1998) points out: "*The archetypal industrial leader in the USA today… is the general manager"* (p.73).

The emphasis on general abstract knowledge may be explained by the fact that Anglo-Saxon economies are considered "individualistic"; they are characterized by fluid,

short term relationships that impact on employment relationships (Haake, 2002). Corporate governance in Anglo-Saxon countries values the satisfaction of short-term shareholder interests' highly and treats managers as agents of the shareholders (Deakin *et al*, 2006). The implications of such corporate governance on employment relationship are low job security and high job turnover rates. Nevertheless, such characteristics provide no incentives for making education relevant to industry-specific skills (Georgen *et al*, 2007). In other words, short-termism makes non-specific organizational knowledge more valuable (Haake, 2002).

Nevertheless, such neo-liberal policies are not welcomed by all economies as some place greater emphasis on issues other than short-term profitability, such as social stability. Although evidence from employment relationships suggests that the expansion of the neo-liberal policy and its corresponding organizational practices have affected working life considerably (James and Wood, 2006) local variations do exist as evidenced by VOC theories. For instance, countries in other capitalisms have different patterns of employment relationships and hence educational needs.

Moreover, and despite their dominance, there is no actual evidence that the LME type economies perform better than the CME type economies (James and Wood, 2006). The observation of various performance indicators like GDP levels, science and technology in US, UK and the rest European countries does not reveal a "crushing" superiority of the liberal economies. The fact that the US presents a successful economic performance, does not guarantee that the liberal market economies are the most successful. And even if they are, this does not suggest that if their institutional peculiarities are incorporated in other types of capitalisms – like the CMEs or the MMEs- this will make these economies perform like the US (Amable, 2004).

Moreover, there is the view that the requirements of modern management involve the cooperation and teamwork between different specialist managers which contradicts the individualistic Anglo-Saxon managerial approach (Bjarnar and Gammelsoeter, 1996) and that the administrative practices developed in the US may not be appropriate for international management practices (Bigelow, 1994).

In addition, there are considerations on whether homogenization pressures with the Anglo-Saxon prototype will alter the distinct identity and image of various educational institutions located in different societies (Kumar and Usunier, 2001). Despite the useful sharing of various cultural and societal perceptions about management by foreign students in an internationalized educational institution (Yang and Rosa, 2001) there are concerns that the global focus of Universities might distance business education from nationalism (Sharma and Judy, 1996) and lead to an "outdated one-best-wayism" in management transferred mostly from Western Universities (Howe and Martin, 1998).

As Bjarnar and Gammelsoeter (1996) comment:

"...Management education at present is based on a paradigm combining individualism and generalism, as well as the belief that universal management techniques exist and can be conveyed to the individual through a business school education" (p.242).

As long as there is the belief that universal management techniques do exist and they can be transmitted through business schools, curricula in management education will continue to be broad, and knowledge will be generalist and individualistic (Bjarnar and

Gammelsoeter, 1996). Nevertheless, the existence of an Anglo-Saxon based "one-best-wayism" in managerial practices and management education can be seriously questioned.

It is argued that differentiation rather than isomorphism in business school curricula can prove to be more beneficial and provide knowledge more relevant to management practice in different national contexts (Pfeffer and Fong, 2004). There is the view that regional development should be a main concern of the Universities and therefore teaching and research should concentrate more on regional needs.

As Chatterton and Goddard (2000) comment: "…a strong and supportive regional economy will create a competitive University, and a strong University has more to offer a region" (p.481). The main argument is that Universities need to respond to the needs of a complex market place that includes several types of businesses, which in turn require several types of skills (Chatterton and Goddard, 2000). For instance, if University education in a country that consists of SMEs does not respond to the particular educational and training needs of SMEs, then regional development will be undermined.

Furthermore, it is shown that business performance is strongly affected by the way different countries are able to combine foreign expertise and their own national systems in management education. Countries that have managed an effective combination rather than adopting the American system of management education have been the most successful (Amdam, 1996). After all, as Sorge (2004) comments full convergence with any model cannot occur, no matter how influential that model might be.

In this respect, we can hardly argue that general business education will contribute greatly in the development of managerial skills, especially in business contexts that do not share the same institutional features with the Anglo-Saxon ones; nor can we argue that national traditions in management education should prevail. Rather, agreeing with Svetlicic and Cibron (1996) *"a productive combination of local and foreign expertise is probably the best"* (p.116).

Although the literature so far suggests a homogenization tendency with the Anglo-Saxon educational system, the research for non-Anglo-Saxon business systems (especially those in Southern Europe) is rather limited, and there is no strong evidence of convergence in such systems. For this reason it is important to investigate management education in non-Anglo-Saxon capitalisms, such as Mediterranean or Southern European capitalisms. Greece can be considered a prime example and the main characteristics of its business system are analyzed in the discussion that follows.

4. The Greek Case

4.1 The Greek National Business System

The literature on comparative capitalism argues that distinct varieties of capitalism or business systems, characterized by systemically specific institutional frameworks

persist despite the globalization pressures. These institutional factors – and associated sets of practices are complementary to each other and affect the way managerial practices are applied in various capitalisms. In this respect the aim of this section is twofold. Firstly, to identify the general characteristics of the Greek national business system (NBS) as these can provide the framework for understanding the organizational behavior of Greek firms and secondly, to examine education and training in Greece as part of its institutional complementarities.

Greece is not close either to the CME or the LME dichotomous model. Along with other Mediterranean countries, it has obvious differences with liberal economies (McMenamin, 2003) and has been characterized as one of the "latecomers" compared to other continental countries (Ferrera, 1996). Various surveys have indicated commonalities between countries such as Italy, Spain, Portugal, and Greece. Either called Mediterranean or South European Capitalisms (Georgen *et al*, 2007; Amable, 2003), or Mixed Market Economies (Molina and Rhodes, 2006), or State Capitalist Economies (Schmidt, 2002) these countries share commonalities that arise from their shared political history and are expressed in terms of economic structures, managerial practices, and other institutional features such as education and training.

Nevertheless, apart from Greece's commonalities with Mediterranean countries, it is not certain whether the Greek case is a characteristic example of the Mediterranean capitalism as its economic performance lags behind countries like Italy and Spain and numerous differences between Greece and other Mediterranean countries have been identified, especially in terms of economic structure.

The Greek business environment consists of various characteristics that are found in different NBSs (Psychogios and Wilkinson, 2007) and has been affected by various political and environmental factors (Holman, 2001). The Greek state scores low according to international measures of government effectiveness, and is not highly capable in delivering public goods. The costly government administration, the large size of the black economy, and the high levels of corruption and tax evasion are some of the reasons that hinder Greece's international competitiveness (Featherstone, 2009).

The Greek NBS is characterized by dualism. The dualism of the Greek NBS refers to the fact that it consists of few large businesses and many family owned SMEs (Mihail and Elefterie, 2006). This has the following implication: although subsidiaries of multinationals use managerial tools and techniques similar to their rivals abroad, the knowledge and usage of these techniques by family owned SMEs that are the majority in the Greek business system, are quite limited (Makridakis *et al*, 1997). Greece also faces an "entrepreneurial deficit". During the last fifteen years the number of new firms created annually remained stable while the number of firms that interrupted their operations increased dramatically (Papayannakis *et al*, 2008).

All the above characteristics shape Greece's managerial practices. Overall, managerial culture in Greece is traditionally based on paternalism with low levels of decentralization and high levels of formalization (Joiner, 2000). Formalization is more evident in SMEs compared to large firms (Spanos *et al*, 2001). Specialization is another important feature of the work culture in Greece since management techniques, tools and systems are highly dependent to the expert knowledge of the people involved (Psychogios and Wilkinson, 2007).

Nevertheless, the forces of competition have urged Greek SMEs to start shaping formal managerial layers staffed with professional managers and move towards more professional rather than paternalistic management styles which are more close to the Western type management style (Mihail and Elefterie, 2006; Spanos *et al*, 2001). SMEs realize that they need to move away from the authoritarian and paternalistic management style by including well-educated professionals in their management teams (Spanos *et al*, 2001).

Despite the recognition of management education's importance in the organization of work, research on this area is rather limited. More specifically the structure and the content of management education in Greece have not been extensively examined. Therefore it is important to examine how managerial theories and concepts are taught in Greek universities in order to further examine how these affect managerial practice.

4.2 Management Education

In Greece the demand for management education has increased considerably (Mihail and Elefterie, 2006). Managerial education in Greece started to gain popularity in the 1950's with the operation of Multinationals. At that time, the top managers selected for executive positions were either from abroad or had studied for an MBA abroad. Since these executives were paid high salaries compared to executives in traditional Greek companies, managers and students started to study abroad for an MBA or took executive courses to improve their managerial skills (Makridakis *et al*, 1997).

Overall it seems that Greek Universities responded quite slowly to the increased demand for business studies. Until the early 1990s only one accredited MBA programme was offered by a public University despite management education's growing popularity (Mihail and Elefterie, 2006). Similarly, the first Master degree in the field of HRM offered by a Greek University was launched in 2002 (Galanaki and Papalexandris, 2005). This was probably the reason why there has been a considerable development in the number of external management development seminars offered to managers in the 1980's (Papalexandris, 1988).

Nowadays, management education in Greece is offered by Universities, Technical Universities, and other private educational institutions (Colleges). More specifically, there are eighteen (18) Universities and seventeen (17) Technical Universities that offer degrees in the general fields of Economics and Management. Business Administration in particular is offered both at the undergraduate and the postgraduate level. Undergraduate degrees in Business Administration are offered by eight (8) Universities and eight (8) Technical Universities, while postgraduate degrees are offered by six (6) Universities and three (3) Technical Universities that cooperate with Anglo-Saxon Universities (the University of Staffordshire in the UK and the University of Kentucky in the U.S). Private colleges that offer business studies collaborating with foreign (mainly UK and US) Universities are numerous. Nevertheless, only nine of them are members of the Hellenic Colleges' Association (HCA, 2009).

The management educational system in Greece has been heavily influenced by western business school models (Savvas *et al*, 2001). Most managers' educational background

is on general sciences rather than on specific management practices such as HRM (Papalexandris, 1992). According to 2006 data, 25 % of managers in Greece had an engineering background (Hellenic Centre for Investment, 2006). This has an impact on the way managers apply certain practices in organizations. For instance, Psychogios and Szamozi (2007) have suggested that HRM in Greece lacks sophistication and that one of the main reasons for that is the existence of an educational gap and limited training on HR issues.

The Greek government has probably realized that Greece's entrepreneurial deficit and its overall poor performance is partly attributed to the fact that the Greek productive system is considered inefficient in exploiting scientific and research outcomes (Papayannakis *et al*, 2008) and that "*Greece has fallen behind in modernizing its educational institutions*" (Makridakis *et al*, 1997p.382), thus indicating a knowledge gap.

For that reason, the Ministry of Education is promoting entrepreneurship education in Universities through the Operational Programme for Educational and Vocational Training "Entrepreneurship Programs in Higher Education". Their actions mainly involve the adoption of best practices of other European Universities (Papayannakis *et al*, 2008).

Despite some indications of foreign influence such as this programme, there is no strong evidence that the Greek management education system is following the Anglo-Saxon prototype. What needs to be explored is whether management education in Greece is indeed acting as a driver of Anglo-Saxon management isomorphism. This can be evidenced by examining the structure and the content of management education programmes as well as the extend to which Anglo-Saxon educated managers apply Anglo-Saxon "best" practices.

The question is whether this adoption of best practices in management education is a trend in the Greek higher educational system and what are the implications of this. Considerations might be raised on whether management education should be structured as such due to the particularities of the Greek NBS and especially that of dualism. In other words, there is the view that apart from conventional wisdom in business studies, the particularities of management practice in Greece necessitate the development of knowledge material that will be more suitable for the Greek reality.

5. Further Research

5.1 Aims and Objectives

The literature suggests that contemporary management education in Europe has Anglo-Saxon origins and influences. Therefore the questions that emerge are whether management education acts as a driver of Anglosaxonization of management practices in non-Anglo-Saxon business systems, and whether the penetration of Anglo-Saxon ideologies into these systems has created increased demand for Anglo-Saxon style management education. Using Greece as an example of a non-Anglo-Saxon NBS the main research questions to be addressed are the following:

Do management education programmes in Greece have Anglo-Saxon influences in terms of their structure and content?

To what extend Anglo-Saxon educated managers in Greece adopt and implement Anglo-Saxon management concepts and practices?

5.2 Research Strategy

As mentioned earlier in the paper, research on whether management education in Greece has Anglo-Saxon influences is rather limited. In this respect, an exploratory research approach will be adopted since there are few earlier studies on the research problem (Hussey and Hussey, 1997). The research strategy of this study will be based on qualitative methods, since a qualitative approach is considered more appropriate for an exploratory study (Creswell, 2003). The data collection tools will involve a documentary analysis of curricula and semi-structured in-depth face-to-face interviews with academics and managers.

The documentary analysis will involve the examination of managerial degrees' curricula benchmarked against a sample of degree curricula from UK and US management schools. The focus of the study will be on MBA programmes. Certain criteria of comparison will be identified such as the aims of the programmes, the admission procedures and the student profiles, the structure of the programmes and accreditation issues. The comparison of the curricula will indicate whether there are significant similarities between Anglo-Saxon and non-Anglo-Saxon MBA programmes. The documentary analysis will be valuable as it will give an insight of the structure and the content of managerial education in Greece and will provide the basis for the interviews with academics that will follow.

Regarding interviews, they will refer both to academics and managers and they will be semi-structured. They will include open-ended questions as this type of questions require more extensive answers compared to closed ended questions and therefore are more suitable for the semi-structured interviews (Stroh, 2000).

As far as it concerns academics, module leaders of both Anglo-Saxon and non-Anglo-Saxon MBA courses will be interviewed. Module leaders will provide information on how their modules have been designed, structured, what are their learning outcomes and what are their basic teaching methods. The main aim is to understand whether non-Anglo-Saxon MBAs are influenced by the Anglo-Saxon prototype in these aspects. Overall, as with documentary analysis, interviews with academics aim at answering whether management education in Greece has Anglo-Saxon influences.

Interviews with Anglo-Saxon educated managers will also be conducted. The sample of managers will consist of MBA graduates. The main aim of these interviews is to examine whether managers apply the specific Anglo-Saxon management practices they have been taught. For that reason the main topics to be addressed will concern issues such as the involvement of lower and middle level managers in strategic planning, the leadership styles (i.e. autocratic or participative), the structuring of management hierarchies (i.e. decentralization), the use of value adding methods and techniques of HRM, and the use of control mechanisms (i.e. emphasis on people or on processes).

6. Conclusion

Educational systems are complementary to labor markets and shape organizational behavior. This institutional role of education has been identified by various institutional theories. Although the Rational Instrumentalist Approach provided an initial examination of education as an institutional factor, most of the research focused on mimetic explanations. A greater emphasis on education has been given in VOC and other heterodox approaches to institutions, where cross country comparisons have classified countries in particular clusters depending on various institutional drivers including education.

Although education- compared to other institutional factors - seems to be more country specific, still a clustering of countries with similar characteristics in their education and training systems exists. This clustering clearly distinguishes Anglo-Saxon countries from the Continental, Social Democratic, and Mediterranean ones and suggests that there is no strong evidence of convergence between the Anglo-Saxon and the rest systems of education. Nevertheless, evidence from the field of management education reveals that due to various homogenizing factors, and despite local peculiarities, there seems to be a level of convergence with the Anglo-Saxon prototype, especially in countries that are not so developed economically.

If that is the case, this raises questions on whether this homogenizing process in management education can prove to be beneficiary for non-Anglo-Saxon business environments that have different institutional characteristics and hence different educational needs. Hence, this study aims to examine whether management education in Greece has strong Anglo-Saxon influences and whether this affects the application of Anglo-Saxon managerial practices in its environment. Also, if evidence shows that Anglo-Saxon practices are indeed applied in the Greek context, the study will examine whether this has increased demand for Anglo-Saxon management education. In that case, the relationship between managerial practices and managerial education will be a two-way one.

To conclude, we need to remove the blinkers and put research forward by attempting to understand not only the extent to which management education in Greece is mainly affected by Anglo-Saxon prototypes, but also whether Anglo-Saxon management knowledge influences management practice.

References

1. Amable, B. (2003) *The Diversity of Modern Capitalism,* Oxford: Oxford University Press.
2. Amable, B. (2004) 'Reforming Europe: Is the Third Way the Only Way? Prisme No 3: Centre Saint-Gobain pour la Recherche en Economie. Obtained through the Internet : *http://www.jourdan.ens.fr/~amable/EngPrismeIIIFVC.pdf* [accessed 22/01/2008].
3. Amdam, Rolv Petter (1996) 'National Systems versus Foreign Models in Management Education: The Norwegian Case', *In*: Amdam, Rolv Peter (ed.),

Management Education and Competitiveness: Europe, Japan, and the US. London: Routledge.
4. Beech, N. (2006)' Intense, Vigorous, Soft and Fun: Identity Work and the International MBA', *Critical Perspectives on International Business,* Vol. 2, No.1, pp.41-57.
5. Bigelow, J. (1994) 'International Skills for Managers: Integrating International and Managerial Skill Learning', *Asia Pacific Journal of Human Resources,* Vol. 32, No. 1, pp.1-12.
6. Bjarnar, O. and Gammelsoeter, H. (1996) 'Managing Management Teams: a Challenge for Management Education', *In:* Amdam, Rolv Peter (ed.), *Management Education and Competitiveness: Europe, Japan, and the US*, London: Routledge.
7. Brewster, C., Wood, G., and Brookes, M. (2006) 'Varieties of Capitalism and Varieties of Firms', *In*: Wood, G. and James, P. (eds) *Institutions, Production and Working Life*, Oxford: Oxford University Press.
8. Byrkjeflot, H. (1998) 'Management as a System of Knowledge and Authority', *In:* Alvarez, J.L (ed.) The *Diffusion and Consumption of Business Knowledge*, London: Macmillan Press Ltd.
9. Calleja, J.J. (2005) 'Wounded by a Divide Syndrome: the Impact of Education and Employment on Euro-Med Cohesion', *Mediterranean Politics,* Vol. 10, No.3, pp. 407-426.
10. Chatterton, P. and Goddard, J. (2000) 'The Response of Higher Education Institutions to Regional Needs', *European Journal of Education, Vol.* 35, No. 4, pp.475-796.
11. Deakin, S., Hobbs, R., Konzelmann, S., Wilkinson, F. (2006) 'Anglo-American Corporate Governance and the Employment Relationship: a Case to Answer?', *Socio-Economic Review,* Vol.4 *No. 1*, pp.155-174.
12. Don Antunes and Howard, T. (2007) 'The Competitive (Dis) Advantges of European Business Schools', *Long Range Planning,* Vol 40, No. 3, pp.382-404.
13. Engwall, L.and Zamagni, V., eds., (1998) *Management Education in Historical Perspective*. Manchester: Manchester University Press.
14. Engwall, L., Gunnarsson, E. and Wallerstedt, E. (1996) 'Mercury's Messengers: Swedish Business Graduates in Practice', *In:* Amdam, Rolv Peter (ed.) *Management Education and Competitiveness: Europe, Japan, and the US*, London: Routledge.
15. Ertl, H. (2006) 'European Union Policies in Education and Training: the Lisbon agenda as a turning point?' *Comparative Education,* Vol. 42, No. 11, pp5-27.
16. Ferrera, M. (1996) 'The Southern Model of Welfare in Social Europe', *Journal of European Social Policy,* Vol. 6, pp. 17-37.
17. Galanaki, E. and Papalexandris, N. (2005) 'Outsourcing of Human Resource Management Services in Greece', *International Journal of Manpower,* Vol. 26, No.4, pp.382-396.
18. Gemelli, G. (1998) 'The "enclosure" effect: innovation without standardization in Italian postwar Management Education', *In*: Engwall, L. and Zamagni, V. (eds.) *Management Education in Historical Perspective*, Manchester: Manchester University Press.

19. Georgen, M., Brewsterm C. and Wood, G. (2007) Corporate Governance Regimes, Investments in Human Capital and Economic Growth. *ECGI Working Paper Series in Finance.* Finance Working Paper N. 188. Obtained through the internet: *http://papers.ssrn.com/sol3/papers.cfm?_abstract_id=1016363* [accessed 15/11/2007].
20. Haake, Sven (2002) 'National Business Systems and Industry Specific Competitiveness', *Organization Studies,* Vol.23, No.5, pp. 711-736.
21. Hall, P. and Soskice, D. (2001) Varieties of Capitalism: the Institutional Foundations of Comparative Advantage. New York: Oxford University Press.
22. Hellenic Centre for Investment (2006) Human Resources. Obtained through the Internet: *http://www.elke.gr/default.asp?V_DOC_ID=2438&V_LANG_ID=0* [accessed 9/1/09].
23. Hill, J. and Houghton, P. (2001) 'A Reflection on Competency-Based-Education: Comments from Europe', *Journal of Management Education, Vol.* 25, No.2, pp.146-166.
24. Hollingsworth, J.R. (2006) 'Advancing our understanding of Capitalism with Niels Bohr's thinking about Complementarity', *In:* Wood, G.T. and James, P. (eds) *Institutions, Production and Working Life*, New York: Oxford University Press
25. Holman, O. (2001) 'Semiperipheral Fordism in Southern Europe', *In*: Jessop, B. (ed) *Country Studies,* UK: Edward Elgar Publishing Limited.
26. Howe, S. and Martin G. (1998) 'Internationalization Strategies for Management Education', *Journal of Management Development, Vol.*17, No.6, pp.447-462.
27. Huisman, J. and Van Der Wende, M. (2004) 'The EU and Bologna: are Supra and International Initiatives Threatening Domestic Agendas?', *European Journal of Education,* Vol. 39, No. 3, pp.349-357.
28. Hull, G.S (2000) 'US MBA and Management Training Programs in Central and Eastern Europe', *Journal of Technology Transfer,* Vol. 25, pp. 319-327.
29. James, P. and Wood, G. (2006) 'Change and Continuity in Working Life'. *In:* Wood, G., and James, P. (eds*) Institutions, Production, and Working Life*, Oxford: Oxford University Press.
30. Javidan, M Stahl,G., Brodbeck,F. and Wilderom, C. (2005) 'Cross-Border Transfer of Knowledge: Cultural Lessons from Project GLOBE', *Academy of Management Executive*, Vol.19, No. 2, pp.59-76.
31. Joiner, T.A (2000) 'The Influence of National Culture and Organizational Culture Alignment on Job Stress and Performance: Evidence from Greece', *Journal of Managerial Psychology, Vol.* 16, No. 3, p. 229-242.
32. Kipping, M., Usdiken, B. and Puig, N. (2004) 'Imitation, Tension and Hybridization: Multiple Americanizations of Management Education in Mediterranean Europe', *Journal of Management Inquiry, Vol.* 13, No.2, pp 98-108.
33. Kumar, R. and Usunier, J.C. (2001) 'Management Education in a Globalizing World: Lessons from the French Experience', *Management Learning, Vol.* 32, No. 3, pp.363-391.
34. Lam, A. (2000) 'Tacit Knowledge, Organizational Learning and Societal Institutions', *Organization Studies,* Vol. 21, No.3, pp.487-513.

35. Makridakis, S., Caloghirou, Y., Papagiannakis, L., and Trivellas, P. (1997) 'The Dualism of Greek Firms and Management: Present State and Future Implications', *European Management Journal,* Vol. 15, No. 4, pp.381-402.
36. Mazza, C., Sahlin-Andersson K., and Strandgaard-Pedersen, J. (2005) 'European Constructions of an American Model: Developments of four MBA Programmes', *Management Learning, Vol.* 36, No. 5, pp. 471-491.
37. McMenamin, I. (2003) 'Is there an East-Central European Variety of Democratic Capitalism? Working Paper in International Studies Centre for International Studies. Obtained through the Internet: *http://webpages.dcu.ie/~ucis/2003_5.pdf* [accessed 22/01/2008].
38. Mihail, D. and Elefterie, K.A. (2006) 'Perceived Effects of an MBA Degree on Employability and Career Advancement: the Case of Greece', *Career Development International,* Vol. 11, No. 4, pp.352-361.
39. Molina, O. and Rhodes, M. (2006) 'Mixed Market Economies', *European Politics and Society*, Vol.5, No. 1, p.11-15.
40. O'Hare, D. (2004) 'Management Education and Development in the United Kingdom', International Journal of Applied Management Education and Development, Vol.1, No.1, Obtained through the Internet: *http://www.managementjournals.com/journals/education/article2.htm*,[accessed 15/10/2008]
41. Papalexandris N. (1992) 'Human Resource Management in Greece', *Employee Relations*, Vol. 14, No. 4, pp 38-52.
42. Papalexandris, N. (1988) 'Measuring the Effectiveness of External Management Training Seminar in Greece', *Management Learning, Vol.* 19, No.1, pp.22-29.
43. Papayannakis, L., Kastelli, I., Damigos, D., and Mavrotas, G. (2008) 'Fostering Entrepreneurship Education in Engineering Curricula in Greece. Experience and Challenges for a Technical University', *European Journal of Engineering Education,* Vol. 33, No. 2, p.199-210.
44. Pfeffer, J. and Fong, C. (2004) 'The Business School "Business": Some Lessons from the US Experience', *Journal of Management Studies,* Vol. 41, No.8, pp.1501-1520.
45. Psychogios, A. and Szamozi, L. (2007) 'Exploring the Greek National Business System: towards a Modernization Agenda', *EuroMed Journal of Business, Vol.* 2, No.1, pp 7-22.
46. Psychogios, A. and Wilkinson A. (2007) 'Exploring TQM awareness in the Greek National Business Context: between conservatism and reformism cultural determinants of TQM', *International Journal of HRM*, Vol.18, No. 6, pp.1042-1062.
47. Rees, C. and Edwards, T. (2006) 'National Systems and Management Action', *In:* Edwards, T. and Rees (eds), C., *International Human Resource Management*. London: Pearson Education Ltd.
48. Savvas, M. El-Kot, G. and Sadler-Smith, E. (2001) 'Comparative Study of Cognitive Styles in Egypt, Greece, Hong-Kong, and the UK', *International Journal of Training and Development, Vol.* 5, No. 1, pp.64-73.
49. Schmidt, V. (2002) *The Futures of European Capitalism*, Oxford: Oxford University Press.

50. Sharma, B. and Judy, A. R. (1996) 'Aspects of Internationalization of Management Education', *Journal of Management Development*, Vol. 15, No.1, pp. 5-13.
51. Sorge, A. (2004) 'Cross-National Differences in Human Resources and Organization', *In*: Harzing, A. and Ruysseveldt, J.V. (eds) *International Human Resource Management*, 2nd edition, London: Sage Publications.
52. Spanos, Y., Prastacos, G. and Papadakis, V. (2001) 'Greek Firms and EMU: Contrasting SMEs and Large-Sized Enterprises', *European Management Journal, Vol.* 19, No. 6, p.638-648.
53. Sultana, R. (2008) 'Higher Education in the Mediterranean: the University between Tradition and Modernity', Working Paper, Obtained through the Internet:*http://www.unifr.ch/ipg/ecodoc/conferences/DocuPDF_Conf_Inter/Sultana.pdf* [accessed 22/01/08].
54. Svetlicic, M. and Cibron, A. (1996) 'Management Education in a Country in Transition: the case of Slovenia', *In*: Amdam, Rolv Peter (ed) *Management Education and Competitiveness: Europe, Japan, and the US*, London: Routledge.
55. Usdiken, B. and Yildirim, O. (2007) 'National Business Systems and Educational Backgrounds of Salaried Top Managers: the Case of Turkish Family Business Groups. *Proceedings of the 23rd EGOS Colloquium*, Vienna, 5-7 July 2007.
56. Van der Wende, M. (2001) 'Internationalisation Policies: about New Trends and Contrasting Paradigms', *Higher Education Policy,* Vol. 14, No.1, p.249-259.
57. Whitley, R. (1999) Divergent Capitalisms: the Social Structuring and Change of Business Systems, Oxford: Oxford University Press.
58. Whitley, R. (2006) 'Understanding Differences: Searching for the Social Processes that Construct and Reproduce Variety in Science and Economic Organization', *Organization Studies*, Vol. 27, No.8, pp.1153-1177.
59. Yang, B. and Rosa, D. (2001) 'Predicting academic performance in management education: an empirical investigation of MBA success', *Journal of Education for Business,* September/October, pp.15-20.
60. Zeitz et al (1999) 'Distinguishing Adoption and Entrenchment of Management Practices: a Framework for Analysis', *Organizational Studies, Vol.*20, No.5, pp.741-776.

The Impact of Foreign Direct Investment on the Greek Economy. A Data Envelopment Analysis of the Greek Chemicals Sector and the Food Products And Beverages Sector

Pavlos Gkasis, Evangelia Desli and Persefoni Tsaliki

[1] *20, Skra Square, 55132, Thessaloniki, Greece*

[2] *Dept of Economics, Aristotle University of Thessaloniki, 54124,, Thessaloniki Greece, {pgkasis, desli, ptsaliki}@econ.auth.gr*

The current study considers technological diffusion via technological and productivity spillovers, as a main mechanism through which Foreign Direct Investment (FDI) affects an economy. More specifically, it investigates whether there are positive or negative spillovers from the operations of Multinational Enterprises (MNEs) both in the Chemicals and the Food Products and Beverages sectors of the Greek economy. The study uses Data Envelopment Analysis (DEA) to estimate technical efficiency of all the firms in these industries and in turn attempts to identify the determinants that influence the existed technical efficiency levels. In addition, in order to measure productivity spillovers a convergence test is applied to examine if technical efficiency levels of all firms operating in each sector converge over time. The data set used comprises of two panels of firms that operate in the Chemicals and the Food Products and Beverages sectors in the time period 2001-2007. The empirical analysis conveys to the conclusion that domestic firms' technical efficiency is positively affected by FDI presence in the chemicals sector and FDI generate positive spillovers contributing to the economic performance of the sector. On the other hand FDI presence in the food products and beverages sector is not appearing an evident effect on the performance of domestic firms, leading to the hypothesis that FDI presence in larger sectors does not promote the diffusion of technology.

Keywords. Technical Efficiency, Technological and Productivity Spillover, Multinational Enterprises, Technological Change, Data Envelopment Analysis.

1. Introduction

Foreign Direct Investment (FDI), along with financial capital movements and international trade are considered as the main mechanisms through which economic globalization has emerged during the last decades. In contrast to portfolio investment and international trade, FDI is considered to contribute to the production basis of the receiving country and the mechanism of achieving that is the transfer of technology from the investing country to the host economy (Nicholson, 2002; Barrell and Pain, 1997; Blomström and Kokko, 1997; Reddy and Zhao, 1990; Unger, 1988; Rosenberg, 1983; Mansfield et al., 1982). Hence, the diffusion of technology and the subsequent productivity spillovers generated by FDI inflows has been extensively discussed in economic literature.

The present study examines to what extent technological diffusion is the main mechanism through which FDI affects the Greek economy. More specifically, the analysis focuses on two different sectors of the Greek economy. The first sector to be analyzed is the chemicals sector which appears an extensive presence of MNEs. Additionally, the food products and beverages sector is analyzed in comparison to the chemicals sector. The first part of the data set used is a panel of 254 firms that operate in the chemicals sector for the time period 2001-2007. The second part of the data set is a panel of 971 firms of the food products and beverages sector for the same time period (2001-2007). Using firm-level data, we employ Data Envelopment Analysis (DEA) to estimate the technical efficiencies of all the firms in the two industries. In turn, we attempt to identify the determinants that influence the existed technical efficiency levels. Moreover, in order to measure productivity spillovers a convergence test is applied in order to examine to what extent technical efficiency levels of all firms in each sector converge over time. Finally, the implications of MNEs operating in two sectors with different economic characteristics are discussed. The current paper is organized as follows: Section 2 briefly reviews the studies referring to technological diffusion through FDI. Section 3 presents the methodology used in the empirical analysis and provides a description of the employed data sets. Section 4 reports the empirical results of the analysis and it contains a discussion of the main results and policy implications of the paper. Section 5 concludes.

2. Literature Review

Different aspects of FDI have been widely analyzed in economic literature; however the research on technological diffusion through FDI is still limited and the main reason for this can be found in the nature of technology itself. Moreover, as Krugman (1991) notes the analysis of technological spillovers is also a difficult task mainly due to the problems of measurement and indices' construction. Despite the pessimistic approach of data availability, many researchers use different approaches to unravel the process through which technology flows inside an economy. This has taken the form of an important (though compact) strand of literature which identifies technological diffusion as the main mechanism through which FDI affects economies worldwide. The bulk of these studies use either panel or cross section data sets; however, the employed

methodological approach affects the results to a great extent, as most cross-section studies present a positive relation between technological diffusion and FDI, while studies using panel data offer mixed results.

Cross-section studies that show a positive and statistically significant relation (Dimelis, 2005; Wei and Liu, 2003; Buckley et al., 2002; Driffield, 2001; Blomström and Sjöholm, 1999; Chuang and Lin, 1999; Sjöholm, 1997; Kokko, 1996; 1994; Kokko, et al., 1996; Blomström and Wolff, 1994; Blomström, 1986; Blomström and Persson, 1983; Globerman, 1979) outweigh those that have used cross-section analysis and show a negative relation (Barrios et al., 2002; Caves, 1974) . On the other hand, analysis of panel data has generated more vague results as some studies show positive results (Ruane and Ugur, 2005; Sinani and Meyer, 2004; Bouoiyour, 2004; Liu, 2002; Haskel et al., 2002; Girma and Görg, 2002: Liu et al., 2000) some show negative ones (Barry et al., 2005; Imbriani and Reganati, 2003; Zukowska, 2000; Flôres et al., 2000; Aitken and Harrison, 1999; Haddad and Harrison, 1993), while others have mixed results depending on the subsamples (all firms or only domestic) used in their analysis (Feinberg and Majumdar, 2001; Kinoshita, 2001; Djankov and Hoekman, 2000; Barrios, 2000).

Apart from the differences that stem from the employed econometric techniques another issue that arises is the data engaged in the analysis. The majority of studies on technological diffusion use either disaggregated data on the firm level or aggregated data on macroeconomic indicators. There is however an additional data generating process which is the application of input – output matrices (Javorcik, 2004; Damijan et al., 2003; Yudaeva et al., 2003; Schoors and van der Tol, 2002). While this kind of data is very helpful for the analysis of inter-sectoral relations as well as intra-sectoral ones as it focuses on linkages between sectors, it appears significant problems due to unavailability of data (Javorcik, 2004). An additional impediment on studying technological diffusion is the lack of reliable data (Sinani and Meyer, 2004; Damijan et al., 2003). According to Görg and Strobl (2001), the lack of credible firm level data is considered as the main reason leading to limitations when technological diffusion is being analyzed.

In order to address the difficulties presented above, DEA has emerged as an alternative methodological approach during the last decade for the evaluation of technological diffusion through FDI. The studies by Fu (2008), Ghali and Rezgui (2008), Mahadevan (2002), Hirschberg and Lloyd (2000), Sun, Hone and Doucouliagos (1999) have used the DEA framework to examine the existence of technological diffusion through FDI. While Fu (2008), Ghali and Rezgui (2008) and Mahadevan (2002) acknowledge the existence of technological diffusion through FDI as an important factor affecting the efficiency of local firms, Hirschberg and Lloyd (2000) and Sun, Hone and Doucouliagos (1999) find mixed results. All these studies though, come to address the role of absorptive capability on the extent that host economies will benefit from the inflow of FDI. Especially Hirschberg and Lloyd (2000) and Sun, Hone and Doucouliagos (1999), which focus on the manufacturing sector of the Chinese economy, find that the relation between technological diffusion and efficiency of local firms is positive for firms that are more technologically endowed and more open to the inflow of FDI. The role of the absorptive capability of the host economy is appearing as the most important determinant regarding the importance of technological diffusion regardless of the methodological framework used, either econometric or DEA.

Research on technological diffusion through FDI is very limited in the case of the Greek economy. The two studies of Barrios et al. (2002) and Dimelis (2005) use firm level data which is elaborated in a cross section econometric framework. Barrios et al. (2002) examine the Greek economy along with the Irish and Spanish economies. They use firm level data for the years 1992 and 1997 in a cross section framework and according to their results technological diffusion through FDI was found as statistically significant for Spain and Ireland but not for Greece. The study by Dimelis (2005) concludes that only firms that are technologically endowed will be able to follow the technologically advanced MNEs that operate in the economy and absorb positive spillovers from them. However, limitations exist in both studies for Greece since the issues at hand are observed for single years without allowing safe conclusions after taking into account the dimension of time on technological diffusion.

3. Methodological Issues

The econometric techniques used to examine the technological diffusion through FDI in an economy have not offered a clear picture of the diffusion process and also of the main variables to which technological diffusion can be attributed. The vast majority of studies use cross-section data sets which, however, omit time-varying effects and dynamic aspects of the impact of FDI on the economy. Consequently, cross section studies tend to offer biased results towards positivity for technological diffusion and FDI. Other studies use panel data sets, which is, in general, a more solid methodological approach as it captures time-varying effects and observes cross section data for a defined set of years, thus offering a clearer picture of the issue under consideration.

Methodology

Most studies on technological diffusion via FDI use stochastic models whose main properties can be summarised as follows: the production function is known and common for all firms, firms operate under constant returns to scale and terms of optimization, and there is no technical change (Fu, 2008). However, the presumption of a common production function implies that all firms structure their production process in the same way. Apart from that, it is not certain that all firms operate under constant returns to scale and that they are seeking for optimization. Finally, the data handling of large panel data sets under the hypothesis of lack of technical change for the whole period, is an assumption that is not accurate considering the way economies operate. For all these reasons, the stochastic models gradually have been substituted by others such as the deterministic parametric statistical and non-statistical frontier models, the stochastic parametric frontier models, and the non-parametric Data Envelopment Analysis (DEA) which focus on the measurement of technical efficiency.

The most prominent approach is the DEA which is a non-parametric statistical methodology for the estimation of technical efficiency of a Decision Making Unit (DMU). In order to examine the technical efficiency of a firm, a frontier is being calculated which represents the most technically efficient firms of the sample at hand. This frontier basically defines the technological level within which firms belonging to

the same sample operate. The main advantages of DEA compared to stochastic methodologies is that no production function is necessary to be estimated in the outset of the analysis, firms can operate under variable returns to scale, they can have multiple inputs and multiple outputs as well, they do not necessarily operate under any rules of optimization and technical change is endogenous in the methodology. On the other hand, the lack of standard errors is important for the skepticisms around the selection of DEA as an analytical tool (Ray, 2004).

The methodology applied in the current paper is a combination of DEA and regression analysis based on a panel dataset (Ray, 1988). In the first stage the technical efficiency scores (TE) for each firm are estimated from the following general model:

$$y_{it} = f_t(X_{it}) * TE_{it} \quad (1)$$

where yit is an output measure of the firm i at time t, Xit stands for the inputs used in the context of the current technological level f of the firm i at time t and TEit is technical efficiency scores of the firm i at time t. More specifically, our model is formulated:

$$y_{it} = f_t(L_{it}, K_{it}, M_{it}) * TE_{it} \quad (2)$$

In which the output is measured by total revenue, while the inputs are L for labor, K for capital and M for materials, sales and administrative expenses of each firm.

In the second stage of the analysis the efficiency scores are regressed against a set of variables

$$TE_{it} = a + b_1 OWN_{it} + b_2 SIZE_{it} + b_3 CR_{it} + b_4 PM_{it} + \varepsilon_{it} \quad (3)$$

where TE stands for technical efficiency scores; OWN is a dummy variable that differentiates domestic firms from foreign owned firms (domestic=0 and foreign=1); SIZE express the relative size of each firm measured by the employment of the firm over the total employment of the sector; CR is calculated as current assets divided by current liabilities for each firm and it captures the ability of the firm to meet its short term debts. Finally, PM stand for net profit margin of the firm and it equals net profit over total revenue.

Data

Aggregated macroeconomic data is insufficient to explain the process of diffusion as many databases appear incomplete time series data (especially at the firm level) and lack of efficient data on several indicators (Sinani and Meyer, 2004; Damijan et al., 2003; Görg and Strobl, 2001). Hence, the implementation of balance sheets data is more cohesive as data can be found for every year that a firm operates in the economy. Thus, microeconomic analysis at the firm level is considered in this study as the most appropriate methodological approach to capture the diffusion of technology through FDI at the intra-sectoral level.

Different data bases have contributed to the collection of data for the needs of the present analysis. The main volume of data was drawn from firms' balance sheets published by Hellastat S.A. and a list of MNEs operating in the chemicals and food products and beverages sectors, by the Bank of Greece. Additional data on labor cost and labor cost deflators was collected from Eurostat and the National Statistical Service of Greece (NSSG). The initial sample comprised of 469 firms of the chemicals sector of which 41 were listed as MNEs from the Bank of Greece. Due to lack of essential data, the final sample for the chemicals sector consists of 254 firms of which 34 are MNEs. Regarding the sector of food products and beverages the initial sample consisted of 2345 firms of which 47 are MNEs. This sector also appeared lack of essential data. Therefore, the sample has 971 firms of which 27 are MNEs. The analysis of the two sectors takes place on the 4-digit categorization of the NACE Rev. 1.1 index for the statistical categorization of economic activity. Finally, the analysis covers the period 2001-2007. The rationale behind analyzing two different sectors is to compare the effect that FDI has on a sector with extensive presence of MNEs compared to domestic firms, such as the chemicals sector, while for the food products and beverages sector there appears to be a smaller ratio of MNEs to domestic firms.

4. Empirical Results

We calculate the efficiency scores of the MNEs and domestic firms of operating in the chemicals sector. Table 1 reveals that MNEs' technical efficiency exceeds that of domestic firms by more than 10% every year. On 2004 the difference reaches as high as 14% which is decreasing until 2007.

	2001	2002	2003	2004	2005	2006	2007
MNE	0.88	0.85	0.86	0.85	0.85	0.84	0.84
Dom. Firms	0.76	0.75	0.75	0.71	0.74	0.73	0.72

Table 1**. Average Technical Efficiency Scores (Chemicals)**

The same conclusions can be reached for the food products and beverages sector. MNEs appear higher average Technical Efficiency scores than domestic firms for the whole period. The largest difference can be found for 2005 when the difference reaches 12%. It is evident that MNEs operate more efficiently regardless of the sector into which they operate.

	2001	2002	2003	2004	2005	2006	2007
MNE	0.86	0.86	0.84	0.85	0.84	0.84	0.84
Dom. Firms	0.78	0.79	0.78	0.76	0.72	0.73	0.75

Table 2: Average Technical Efficiency Scores (Food Products and Beverages)

In turn, we take into consideration three levels of technical efficiency in order to better understand the distribution of the TE scores between the different firms in the sector.

We set the 0.95 level of technical efficiency as our threshold for the almost perfect efficient condition. In turn, we characterize as moderately efficient firms those with TE scores above 0.80 and lower 0.95, whereas firms with TE scores below 0.80 are characterized as inefficient.

Focusing on the chemicals sector, above the 0.95 level of technical efficiency operate 8 out of 34 MNEs, that is 24% in our sample. With respect to the domestic firms, 19 out of 220 (8% of the sample) show an average technical efficiency above the threshold. Hence, MNEs tend to be more efficient than domestic firms. Moreover, 7 out of the 8 most efficient MNEs belong to the group of medium to large firms. With respect to the domestic firms only one out of the 19 most efficient firms belongs to that group. The rest are small and medium size enterprises. It appears, therefore, that large MNEs are more technically efficient than smaller MNEs in the chemicals sector which, apparently, make better use of their resources by taking advantage of economies of scale. In contrast, it seems that domestic firms do not rely so much on economies of scale but instead on other ways, i.e. innovation to achieve high levels of technical efficiency. This is evidenced, to some extent, by the large number of small and medium sized domestic firms.

In addition, domestic firms with TE score of over 0.95 show a stable trend throughout the whole period. Moderately efficient domestic firms with average TE score of over 0.80 and lower of 0.95 appear differences in their performance. Large and small domestic firms show a stable and mostly upwards trend in their scores whereas medium sized firms appear declining scores each year. Finally, the inefficient firms with TE lower than 0.80 do not demonstrate a common performance. Large and medium sized firms are declining in terms of efficiency, while small ones have stable and some increased scores of TE. With respect to MNEs we see that the efficient firms (TE > 0.95) are all stable for the whole period. The same happens with moderately efficient firms of large and medium size but not with smaller MNEs which show a decreasing tendency of TE. Inefficient MNEs show a decreasing trend for the period at hand. Table 3 presents a synopsis of the above for both the MNEs and domestic firms.

	Almost Efficient TE ≥ 0.95	Moderately Efficient $0.80 \leq$ TE < 0.95	Inefficient TE < 0.80
Large	Stable	Stable/Increasing	Decreasing
Medium	Stable	Decreasing	Decreasing
Small	Stable	Increasing	Increasing

MNEs

	Almost Efficient TE ≥ 0.95	Moderately Efficient $0.80 \leq$ TE < 0.95	Inefficient TE < 0.80
Large	Stable	Stable	Decreasing
Medium	Stable	Stable	Decreasing
Small	Stable	Decreasing	Decreasing

Table 3. Technical Efficiency Performance by Ownership and Size (Chemicals)

Findings are not so straightforward for the food products and beverages sector. Large and medium sized MNEs that operate efficiently, appear a stable level of technical

efficiency for the whole period of the study. Whereas, the other categories of the MNEs' group offer mostly mixed results with some firms appearing increasing levels of TE and others decreasing. Proceeding to the domestic firms' group, the almost efficient group of firms behaves in a similar fashion to the respective group of the chemicals sector. This is, all firms that have an average TE score of over 0.95 for the whole period are steadily efficient throughout the period, regardless of their size. Large domestic firms that are moderately efficient are generally stable for the period at hand and most of them appear increasing rates of TE. Medium and small sized domestic firms that are moderately efficient do not have a stable trend. The same is true for medium sized firms of the inefficient group of firms. These last findings verify the hypothesis that domestic firms operating in the food products and beverages sector do not appear evident changes in efficiency scores that could be attributed to the presence of MNEs in the sector. Apart, from the large sized moderately efficient firms that have a tendency towards stability and some increase their TE scores, all other groups operate vaguely showing that MNEs that operate in a larger sector do not affect domestic firms' efficiency to an important degree.

Domestic firms

	Almost Efficient TE ≥ 0.95	Moderately Efficient $0.80 \leq$ TE < 0.95	Inefficient TE < 0.80
Large	Stable	Stable/Increasing	Stable
Medium	Stable	Mixed Results/Stable	Mixed Results
Small	Stable	Mixed Results	Decreasing

MNEs

	Almost Efficient TE ≥ 0.95	Moderately Efficient $0.80 \leq$ TE < 0.95	Inefficient TE < 0.80
Large	Stable	NA	NA
Medium	Stable	Decreasing	Mixed Results
Small	NA	NA	Mixed Results

Table 4. Technical Efficiency Performance by Ownership and Size (Food Products and Beverages)

After calculating the technical efficiency scores (TE) we proceed with the second stage of our analysis which is the analysis of its determinants. As it is shown in equation 1, our model contains four basic explanatory variables, that is OWN (ownership structure), SIZE, CR (current ratio) and PM (profit margin). Additionally, there have been introduced two interaction variables of ownership along with the PM and CR variables in order to capture the importance of foreign ownership on these two variables and one year lagged TE since the current level of technical efficiency is partly a product of technical efficiency in the previous period. In all model specifications we estimate a panel data sample taking into account cross section weights. Before proceeding to the estimation of each model we checked the coefficient for the effect of each year on every variable and we haven't found that it is statistically significant. Following McDonald (2008) and Hoff (2007) we use Ordinary Least Squares (OLS) to analyze the sample as technical efficiency scores are fractional data

and not generated by a censoring process (McDonald, 2008) which would call for use of a Tobit analysis. In order to correct for heteroskedasticity we calculate White heteroskedastic-consistent standard errors.

	1st Model	2nd Model	3rd Model	4th Model
TE(-1)	0.837***	0.836***	0.842***	0.841***
	[0.025]	[0.024]	[0.026]	[0.026]
SIZE	1.387***	1.394***	1.363***	1.350***
	[0.190]	[0.192]	[0.182]	[0.175]
PM	0.024***	0.025***		
	[0.0051]	[0.005]		
OWN	0.014***	0.014***	0.007***	0.018***
	[0.002]	[0.001]	[0.002]	[0.003]
CR	-0.00078			
	[0.001]			
C	0.108***	0.107***	0.104***	0.104***
	[0.0198]	[0.019]	[0.021]	[0.021]
OWN*PM			0.117***	0.119***
			[0.026]	[0.024]
OWN*CR				-0.007***
				[0.001]
R-squared	0.994	0.994	0.995	0.996
Adj. R-squared	0.994	0.994	0.995	0.995
DW statistic	1.91	1.90	1.92	1.92
N	1367	1367	1367	1367

Table 5. Panel Generalised Least Squares Analysis with Cross Section Weights

Note: Figures in square brackets are Standard Errors. *,**,***, denote statistical significance at the 10%, 5% and1% confidence intervals, respectively. N=Sample size, Durbin-Watson= Durbin-Watson statistic

Table 5 reports the results of the four specifications for the chemicals sector. The results from the first specification which contains the four variables from equation 2

and TE lagged one year show that except for CR, all other variables are statistically significant at the 1% confidence interval. In the second specification, we drop CR and the model continues to perform with all variables being statistically significant at the 1% confidence interval. In the third and fourth specifications, we add consecutively the interaction terms of the OWN variable with the variables PM and CR. Both models performed well. In addition, the interaction variable OWN*CR is negative in both models indicating that MNEs do not need increased liquidity in order to obtain loans for their short-time responsibilities. On the other hand the positive and significant coefficient of the OWN*PM enhances the hypothesis that MNEs operate with increased efficiency as they have better control of the costs associated with their production process.

	1st Model	2nd Model	3rd Model	4th Model
TE(-1)	0.723***	0.724***	0.787***	0.787***
	[0.041]	[0.041]	[0.037]	[0.037]
SIZE	1.669***	1.677***	1.400***	1.397***
	[0.322]	[0.323]	[0.323]	[0.323]
PM	0.139***	0.138***		
	[0.016]	[0.016]		
OWN	0.006	0.006	0.005	0.008
	[0.005]	[0.005]	[0.005]	[0.005]
CR	0.000			
	[0.000]			
C	0.204***	0.202***	0.155***	0.155***
	[0.033]	[0.033]	[0.030]	[0.030]
OWN*PM			0.104***	0.111***
			[0.015]	[0.014]
OWN*CR				-0.001***
				[0.000]
R-squared	0.990	0.990	0.989	0.989
Adj. R-squared	0.990	0.990	0.989	0.989
DW statistic	2.110287	2.104	2.208	2.208
N	5071	5072	5072	5071

Table 6. Panel Generalised Least Squares Analysis with Cross Section Weights of Technical Efficiency (Food Products and Beverages)

Note: Figures in square brackets are Standard Errors.

*,**,***, denote statistical significance at the 10%, 5% and1% confidence intervals, respectively.
N=Sample size, Durbin-Watson= Durbin-Watson statistic

The different model specifications for the food products and beverages sector are presented at table 6. The most important finding of these specifications is that the ownership dummy used is not statistically significant for any of the four model specifications used. This implies that ownership structure does not play an important role to the TE score achieved by each firm. All other variables are statistically significant at the 1% confidence interval, as happened for the chemicals sector. These findings come to verify the analysis of the first stage that domestic firms are not evidently affected by the presence of MNEs in the sector.

5. Conclusion

The present study examines the effect of MNEs in the sector of chemicals and food products and beverages in the Greek economy during the period 2001-2007. At first, following the DEA method we calculated the TE scores for all firms in these sectors. In turn, with a use of an econometric model we attempted to investigate the determinants of the TE scores. The analysis for the chemicals sector showed that ownership is a statistically significant explanatory variable indicating that MNEs hold a strong position within the sector. In fact, the empirical analysis showed that MNEs outperform domestic firms in terms of technical efficiency (Table 1). Moreover, the efficient MNEs perform in a steady manner for the whole period of analysis whereas the almost efficient MNEs of small size are gradually more technically inefficient. The same is evident for the inefficient MNEs. The reason is probably the same for both of these instances. Small, almost efficient MNEs and inefficient MNEs independent of size are gradually outperformed by the more efficient ones. This can lead either to crowding out effects for these firms or even subsequent mergers and acquisitions.

Even more interesting, are the results for the group of the domestic firms. Firms belonging to the efficient group are stable with respect to their performance in TE scores. However, the moderately efficient group is probably the best example to present the effects of the technological diffusion through FDI. Within this category, large firms which have better levels of absorptive capability, better use of resources and most likely important market shares present stable and increasing TE scores through the seven year period. At the same time small moderately efficient firms have stable and increasing scores of TE as well, since they are forced to operate efficiently in order to survive. In contrast, medium sized firms display declining TE scores as they do not have the absorptive capacity to gain from the influx of FDI, while at the same time they are powerful enough not to be driven by the fear of closure in order to reach better levels of technical efficiency. As expected, inefficient domestic firms are all appearing declining TE scores. These are the firms that are crushed by the increased competition inside the sector and although driven by the instinct of survival do not have the adequate power to remain in the economy.

The food products and beverages sector appears different results. TE obtained at the first level of the analysis offer mixed results regarding the performance of domestic firms compared to that of MNEs, which is the focus of the current analysis. The moderately efficient group of large domestic firms is partly verifying the hypothesis of diffusion but results for other groups cannot lead to generalizations. Moreover, when proceeding to the second stage of the analysis it becomes evident that ownership structure is not an important explanatory variable for the TE scores obtained by the firms of the sample.

To conclude, MNEs are performing as expected for both sectors, which is more efficiently than domestic firms do. The analysis shows that the larger the sector is into which MNEs operate, the smaller are the effects of diffusion on domestic firms. One can hypothesize that MNEs' contribution to the local economy can be more fruitful when focused on smaller sectors of the economy.

Future research on technological diffusion through FDI with the use of DEA, will offer an alternative framework to the existing bibliography. It is believed that the implementation of DEA along with data of better quality and the improvement of the existing databases will offer better model specifications to examine the underlying relation between FDI and technological diffusion.

References

1. Aitken, B., and Harrison, A., (1999), "Do Domestic Firms Benefit from Direct Foreign Investment? Evidence from Venezuela." The American Economic Review, Vol. 89, No. 3, pp. 605-618.
2. Arcelus, F. J., and Arocena, P., "Convergence and Productive Efficiency in Fourteen OECD Countries: A Non-parametric Frontier Approach", International Journal of Production Economics, Vol. 66, No. 2, pp. 105-117.
3. Banker, R. D., Charnes, A., and Cooper, W. W., (1984), "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis", Management Science, Vol. 30, No. 9, pp. 1078-1092.
4. Barrell, R., and Pain, N., (1999), "Real Exchange Rates, Agglomerations, And Irreversibilities: Macroeconomic Policy and FDI in EMU", Oxford Review of Economic Policy, Vol. 14, No. 3, pp. 152-167.
5. Barrios, S., Dimelis, S., Louri, H., and Strobl, E., (2002), "Efficiency Spillovers from Foreign Direct Investment in the EU Periphery: A Comparative Study of Greece, Ireland and Spain", FEDEA Discussion Paper No. 2002-02, Spain.
6. Barrios, S., (2000), "Foreign Direct Investment and Productivity Spillovers: Evidence from the Spanish Experience", Working paper No. 2000-19. The Foundation for Applied Economic Research, Spain.
7. Barry, F., Görg, H., and Strobl, E., (2005), "Foreign Direct Investment and Wages in Domestic Firms in Ireland: Productivity Spillovers Versus Labour-market Crowding Out". International Journal of the Economics of Business, Vol. 12, No. 1, pp. 67-84.

8. Blomström, M., (1986), "Foreign Investment and Productive Efficiency: The Case of Mexico", The Journal of Industrial Economics, Vol. 35, No. 1, pp. 97-110.
9. Blomström, M., and Kokko, A., (1997), "How Foreign Investment Affects Host Countries", World Bank Policy Research Working Paper No. 1745, World Bank.
10. Blomström, M., and Persson, H., (1983), "Foreign Investment and Spillover Efficiency in an Underdeveloped Economy: Evidence from the Mexican Manufacturing Industry", World Development, Vol. 11, No. 6, pp. 493-501.
11. Blomström, M., and Sjöholm, F., (1999), "Foreign Direct Investment Technology Transfer and Spillovers: Does Local Participation with Multinationals Matter?", European Economic Review, Vol. 43, pp. 915-923.
12. Blomström, M., and Wolff, E. N., (1994), "Multinational Corporations and Productivity Convergence in Mexico' στο: Blomström, M., and Wolff, Edward N., Convergence of Productivity: Cross-national Studies and Historical Evidence, Oxford and New York: Oxford University Press, pp. 243-259.
13. Bouoiyour, J., (2004), "Foreign Direct Investment in Morocco", στο: Perrin, S. and Sachwald, F., eds. Foreign Direct Investment in Developing Countries: Leveraging the Role of Multinationals, Paris: AFD, pp. 149-167.
14. Buckley, P., Clegg, J., and Wang, C., (2002), "The Impact of Inward FDI on the Performance of Chinese Manufacturing Firms", Journal of International Business Studies, Vol. 33, No. 4, pp. 637-655.
15. Caves, R. E., (1974), "Multinational Firms, Competition, and Productivity in Host-Country Markets", Economica, Vol. 41, No. 162, pp. 176-193.
16. Charnes, A., Cooper, W. W., and Rhodes, E., (1981), "Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow Through", Management Science, Vol. 27, No. 6, pp. 668-697.
17. Charnes, A., Cooper, W. W., and Rhodes, E., (1978), "Measuring the Efficiency of Decision Making Units", European Journal of Operational Research, Vol. 3, pp. 429-444.
18. Chuang, Y.-C., and Lin, C-M., (1999), "Foreign Direct Investment, R&D and Spillover Efficiency: Evidence from Taiwan's Manufacturing Firms.", Journal of Development Studies, Vol. 35, No. 4, pp. 117-137.
19. Coelli T., Prasada Rao D.S. and Battese, G. E., (1998), An Introduction to Efficiency and Productivity Analysis, Boston: Kluwer Academic Publishers.
20. Damijan, J. P., Knell, M., Majcen, B., and Rojec, M., (2003), Technology Transfer through FDI in Top-10 Transition Countries: How Important are Direct Effects, Horizontal and Vertical Spillovers?, Working Paper No. 549, The William Davidson Institute, University of Michigan.
21. Dimelis, S. P., (2005), "Spillovers from Foreign Direct Investment and Firm Growth: Technological, Financial and Market Structure Effects", International Journal of the Economics of Business, Vol. 12, No. 1, pp. 85-104.
22. Djankov, S., and Hoekman, B., (2000), "Foreign Investment and Productivity Growth in Czech Enterprises", The World Bank Economic Review, Vol. 14, No.1, pp. 49-64.

23. Driffield, N., (2001), "The Impact on Domestic Productivity of Inward Investment in the UK", The Manchester School, Vol. 69, No. 1, pp. 103-119.
24. Farrell, M. J., (1957), "The Measurement of Productive Efficiency", Journal of the Royal Statistical Society, Series A, Vol. 120, pp. 253-281.
25. Feinberg, S. E., and Majumdar, S. K., (2001), "Technology Spillovers from Foreign Direct Investment in the Indian Pharmaceutical industry", Journal of International Business Studies, Vol. 32, No. 3, pp. 421-437.
26. Flôres, R. G., Fontoura, M. F., and Santos, R. G., (2000), "Foreign Direct Investment Spillovers: What Can We Learn from Portuguese Data?", Working Paper No. 4/2000, Graduate School of Economics, Universidade Tecnica de Lisboa.
27. Fu, X, (2008) "Foreign Direct Investment, Absorptive Capacity and Regional Innovation Capabilities: Evidence from China", Oxford Development Studies, Vol. 36, No. 1, pp. 89-110.
28. Ghali, S., and Rezgui, S., (2008), "FDI Contribution to Technical Efficiency in the Tunisian Manufacturing Sector", ERF Working Paper Series, No. 421.
29. Girma, S. and Görg, H., (2002), "Foreign Direct Investment, Spillovers and Absorptive Capacity: Evidence from Quantile Regressions", GEP Research Paper 02/14, University of Nottingham.
30. Globerman, S., (1979), "Foreign Direct Investment and 'Spillover' Efficiency Benefits in Canadian Manufacturing Industries", The Canadian Journal of Economics, Vol. 12, No. 1, pp. 42-56.
31. Görg, H., and Strobl, E., (2001), "Multinational Companies and Productivity Spillovers: A Meta-analysis With a Test for Publication Bias", The Economic Journal, Vol. 111, No. 475, pp. F723-F739.
32. Haddad, M., and Harrison, A., (1993), "Are There Positive Spillovers from Direct Foreign Investment? Evidence from Panel Data for Morocco", Journal of Development Economics, Vol. 42, No. 1, pp. 51-74.
33. Haskel, J. E., Pereira, S., Slaughter, and C., Matthew, J., (2002), "Does Inward FDI Boost the Productivity of Domestic Firms?" Working Paper, No. 8724. National Bureau of Economic Research, London.
34. Hirschberg, J. G., and Lloyd, P. J., (2000), "An Application of Post-DEA Bootstrap Regression Analysis to the Spill Over of the Technology of Foreign-Invested Enterprises In China", Department of Economics Working Paper #732, The University of Melbourne.
35. Imbriani, C., and Regganati, F., (2003), "The Effects of Foreign Direct Investments on Local Firms: Some Evidence from Panel Data for the Italian Manufacturing Sector", Paper presented at the 2003 Hawaii International Conference on Social Sciences, June 12-13.
36. Javorcik, B. S., (2004), "Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers through Backward Linkages", The American Economic Review, Vol. 94, No. 3, pp. 605-627.
37. Kinoshita, Y., (2000), "R&D and Technology Spillovers via FDI: Innovation and Absorptive Capacity", Working Paper No. 349. The William Davidson Institute, University of Michigan.
38. Kokko, A., (1996), "Productivity Spillovers from Competition between Local Firms and Foreign Affiliates.", Journal of International Development, Vol. 8, No. 4, pp. 517-530.

39. Kokko, A., (1994), "Technology, Market Characteristics, and Spillovers", Journal of Development Economics, Vol. 43, No. 2, pp. 279-293.
40. Kokko, A., Tansini, R., and Zejan, M., (1996), "Local Technological Capability and Productivity Spillovers from FDI in the Uruguayan Manufacturing Sector", Journal of Development Studies, Vol. 32, No. 4, pp. 602–611.
41. Krugman, P., (1991), Geography and Trade, MIT Press, Cambridge MA.
42. Liu, Z., (2002), "Foreign Direct Investment and Technology Spillover: Evidence from China", Journal of Comparative Economics, Vol. 30, No. 3, pp. 579-602.
43. Liu, X., Siler, P., Wang, C., and Wei, Y., (2000). "Productivity Spillovers from Foreign Direct Investment: Evidence from UK Industry Level Panel Data", Journal of International Business Studies, Vol. 31, No. 3, pp. 407–425.
44. Mahadevan, R., (2002), "A DEA Approach to Understanding the Productivity Growth of Malaysia's Manufacturing Industries", Asia Pacific Journal of Management, Vol. 19, pp. 587–600
45. Mansfield, E., Anthony, R., Mark, S., David, T., Samuel, W., and Peter, B., (1982), Technology Transfer, Productivity and Economic Policy, New York.W.W., Norton & Company Inc.
46. Nicholson, M.W., (2002), "Intellectual Property Rights and International Technology Diffusion", Paper prepared for "Responding to Globalisation" Conference at Boulder, CO.
47. Ray, S., (2004), Data Envelopment Analysis. Theory and Techniques for Economics and Operations Research, New York, Cambridge University Press.
48. Ray, S., (1988), "Data Envelopment Analysis, Non-Discretionary Inputs and Efficiency: An Alternative Interpretation", Socio-Economic Planning Sciences, Vol. 22, No. 4, pp. 167-176.
49. Reddy, N.M., and Zhao, L., (1990), "International Technology Transfer: A Review", Research Policy, Vol. 19, No. 4, pp. 285-307.
50. Rosenberg, N., (1983), Inside the Black Box: Technology and Economics, Cambridge, Cambridge University Press.
51. Ruane, F., and Ugur, A., (2004), "Foreign Direct Investment and Productivity Spillovers in Irish Manufacturing Industry: Evidence from Plant Level Panel Data", International Journal of the Economics of Business, Vol. 11, No. 3, pp. 53-66.
52. Schoors, K., and van der Tol, B., (2002), Foreign Direct Investment Spillovers Within and Between Sectors: Evidence from Hungarian Data. Working Paper No. 157, Ghent University, Belgium.
53. Sinani, E., and Meyer, K. E., (2004), "Spillovers from Technology Transfer: The Case of Estonia", Journal of Comparative Economics, Vol. 32, No. 3, pp. 445-466.
54. Sjöholm, F., (1997), "Technology Gap, Competition and Spillovers from Direct Foreign Investment: Evidence from Establishment Data", Working Paper No. 38, The European Institute of Japanese Studies, Stockholm, Sweden.

55. Sun, H., Hone, P., and Doucouliagos, H., (1999), “Economic Openness and technical Efficiency: A Case Study of Chinese Manufacturing Industries”, Economics of Transition, Vol. 7, No. 3, pp. 615-636.
56. Yudaeva, K., Kozlov, K., Melentieva, N., and Ponomareva, N., (2003), “Does Foreign Ownership Matter? Russian Experience”, Economics of Transition, Vol. 11, No. 3, pp. 383-410.
57. Unger, K., (1988), “Industrial Structure, Technical Change and Micro-economic Behavior in Less Developed Countries”, in Giannitsis, A., (1991), (eds), Economic Theory and Technology, Athens, Gutenberg Publications, pp 259-277.
58. Wei, Y., and Liu, X., (2003), “Productivity Spillovers among OECD, Diaspora and Indigenous Firms in Chinese Manufacturing”, Lancaster University Management School Working Paper 2003/008, The Department of Economics, Lancaster University Management School, Lancaster.
59. Zukowska-Gagelmann, K., (2000), “Productivity Spillovers from Foreign Direct Investment in Poland”, Economic Systems, Vol. 24, No. 3, pp. 223–256.

Personal Tax Wedge In Croatia: Does It Hurt Competitiveness?

Ana Grdović Gnip[1], *Iva Tomić*[2]

[1] *Department of Economics and Tourism Dr.Mijo Mirković, University Juraj Dobrila, Preradovićeva 1, 52100 Pula, Croatia, agrdovic@efpu.hr*

[2] *Department of Human Resource Economics, The Institute of Economics, Zagreb, J. F. Kennedy Square 7, 10000 Zagreb, Croatia, iva.tom@gmail.com*

Labour force competitiveness primarily depends upon the skills structure of labour force, the size of labour supply, and the institutional settings in the labour market. This paper notably focuses on the size of labour supply. The main objective is to find out how the size of the labour supply is affected by the size and the structure of total labour costs. Hierarchical and K-means cluster analysis based on 2006 data for 38 countries (EU and OECD members, plus Croatia) use tax wedge as one of the main competitiveness' factors. Results show that the countries can be classified into two clusters: (1) countries with low tax wedge and high employment rate, or (2) countries with high tax wedge and high unemployment rate. Croatia resulted as a country with high tax wedge and high unemployment rate. Further discussion in the paper is devoted to challenges of lower tax burden in Croatia.

Keywords. competitiveness, labour force, employment, unemployment, labour costs, tax wedge, hierarchical cluster analysis, K-means cluster analysis, discriminant analysis, Ward method, dendogram, Croatia.

1. Introduction

Competitiveness symbolizes the ability to perform in the market, i.e. to compete with other firms or countries in a globalized marketplace. Higher competitiveness of the entire economy should bring an increase of FDI, productivity, and political stability while reducing poverty and inequality. Many aspects can determine this, but labour force competitiveness is unquestionably one of the vital factors in achieving competitiveness of the whole economy. It primarily depends upon the skills structure of labour force, the size of labour supply, and the institutional settings in the labour market.

This paper primarily focuses on the size of the labour force; more precisely on the impact that labour costs have on the size of labour supply. This is especially important for the countries with more labour-intense industries, such as Croatia. Croatia is still behind the most advanced economies and competes for foreign investments with other

countries in Eastern Europe. Yet, low labour supply, accompanied with relatively high labour costs, constitutes its biggest constraint. In the paper, we (1) compare the Croatian labour force with other European countries regarding the size and structure; (2) assess the size of the tax wedge in the OECD and EU countries together with Croatia; and (3) analyse the impact of the tax wedge on unemployment and employment rates.

The paper is organised into six parts. In the second part, after introduction, the authors try to provide some insight into main determinants of labour force competitiveness. Third part of the paper gives more detailed look at the supply side of the Croatian labour market. In the fourth part, detailed analysis of the impact of labour costs on the size of labour supply is provided. Tax wedge, as the difference between the costs of labour that are paid by the employer and the net salary that goes to the employee, is defined here and used for further empirical analysis. Practical implications and recommendations are given in the fifth part of the paper. Sixth part is devoted to concluding remarks.

2. Labour force competitiveness

Competitiveness can be defined in many ways, but it implies sustainable productivity and growth potential in all cases. If applied to the whole country, higher competitiveness should bring an increase of FDI, reduction of poverty and inequality, increase of productivity, and political stability - what all finally leads to an increase of the standard of living for the entire population (Bejaković, 2004). It is no wonder then how, recently, all governments pay lots of attention to different competitiveness reports and have even special advisory bodies or government agencies that tackle competitiveness issues. This is especially true for the less developed countries that need to catch-up with more developed ones. However, competitiveness depends on many factors that influence each other, and for a country to be competitive, all of these individual factors need to be competitive on their own.

From the point of view of the entire economy, one of the most important factors is labour force competitiveness (Obadić, 2004). Nevertheless, measurement of the labour force competitiveness is not as easy as it may seem at first sight. There exist a number of definitions of the labour force or human resources competitiveness and they all take different indicators for determining how exactly competitive labour force in one country or region is. Most commonly used measure is definitely the one that compares unit labour costs in home country relative to competitors (Ark van, Stuivenwold, and Ypma, 2005; Gora et al., 2006; Tica and Jurčić, 2006). Yet, this definition shows only one part of the whole aspect of labour force competitiveness. The efficiency of use of labour as a resource depends upon the size of the labour supply, flexibility of the labour market, and the qualification structure of the labour force (Scharle, 2003). Naturally, there are numerous ways that these three can be improved in order to increase competitiveness of the entire labour supply.

2.1 Qualification structure of the labour force

Labour force skills are certainly among the chief determinants of the competitiveness in the labour market. That is not surprising since better-educated and trained human resources should be more productive what ultimately determines their competitiveness in the market. FDI inflow brings new technologies and ways of doing business, but this can only be adopted by the qualified labour. Even so, the exact level of labour force skills in some firm or country is hard to identify. Imperfect measure for the skills structure of the entire labour force is the so-called educational attainment. Table 1 shows shares of active population in the working-age population (between 15 to 64 years) by the highest level of education attained for new member states and some of the accession countries of the EU in the year 2007. It is easily observable how the core of the working-age population consists of the people with secondary education, while those with tertiary education constitute less than 30%, or even less than 20% of the total labour force. Moreover, the share of those with primary education in some countries is still surprisingly large (Cyprus, Malta, Turkey, Romania). This could partly be explained by the fact that these data include persons that could not get to a higher degree of education because of their age.

GEO	Level of education*		
	Primary	Secondary	Tertiary
EU27	25,11	49,43	25,24
EU25	25,40	48,62	25,75
EU15	28,14	44,93	26,65
NMS10	16,95	59,36	23,66
CY	23,43	40,59	35,98
EE	10,47	56,25	33,28
LT	8,08	59,98	31,93
BG	16,14	59,64	24,22
SI	14,49	62,05	23,46
LV	13,96	62,43	23,32
PL	9,63	68,84	21,52
HU	14,01	65,51	20,47
HR	16,65	64,73	18,62
MT	61,12	21,51	17,37
SK	7,22	77,70	15,08
CZ	7,03	78,75	14,21
RO	21,64	64,46	13,90
TR	64,42	22,25	13,32

Table 1. Structure of active population by the share of highest level of education attained (%), 2007

Source: Eurostat, authors' calculation.
*Primary: ISCED 0-2; Secondary: ISCED 3-4; Tertiary: ISCED 5-6.

However, the next table shows that the share of persons in the age group between 15 to 24 is around 10% what definitely cannot explain high share of primary educated people in total labour force. Table 2 illustrates the age structure of active labour force for the mentioned countries. Another interesting thing observable in this table is quite low share of the people aged 55 to 64 in the active labour force. This is probably due to early retirement schemes, different disability pensions, and similar benefits in most of these countries. This could pose a serious problem because almost all of these countries face demographic change with their population (as well as labour force) getting older. In order not to worsen labour force competitiveness, the inclusion of this older group into the labour market should become one of the main policies for the near future

GEO	Age group							
	15-24	25-29	30-34	35-39	40-44	45-49	50-54	55-64
EU27	11,52	11,89	12,80	13,77	14,03	12,93	11,29	11,76
EU25	11,62	11,83	12,73	13,62	14,18	12,92	11,30	11,80
EU15	11,86	11,42	12,41	13,75	14,47	12,87	11,07	12,16
NMS10	11,67	12,97	13,43	13,22	12,55	12,84	11,64	11,67
EE	11,89	12,12	12,30	12,19	12,32	13,33	12,06	13,79
CZ	8,34	12,63	14,87	12,70	12,89	11,82	13,23	13,52
LV	13,45	11,80	12,08	12,50	12,32	13,31	11,24	13,28
BG	8,54	10,02	12,69	14,65	14,53	15,68	11,01	12,88
PL	16,58	10,26	10,89	13,17	13,27	12,48	10,82	12,52
CY	10,17	14,77	13,80	12,60	13,10	12,63	10,46	12,47
LT	9,28	11,72	13,02	13,76	14,41	14,44	10,93	12,42
TR	10,56	10,81	10,60	12,15	14,10	15,27	14,17	12,31
SI	10,30	14,01	14,55	16,89	10,51	12,17	11,17	10,39
HU	7,58	14,19	15,17	14,71	11,86	12,27	14,22	10,00
MT	18,51	14,96	13,52	10,10	11,36	11,18	10,46	9,92
HR	11,12	14,45	14,14	12,44	13,27	13,35	12,83	8,42
SK	10,63	13,25	14,12	13,58	13,50	14,77	11,81	8,35
RO	11,18	14,96	14,67	12,62	12,35	13,65	12,35	8,22

Table 2. Structure of active population by age groups (%), 2007

Source: Eurostat, authors' calculation.

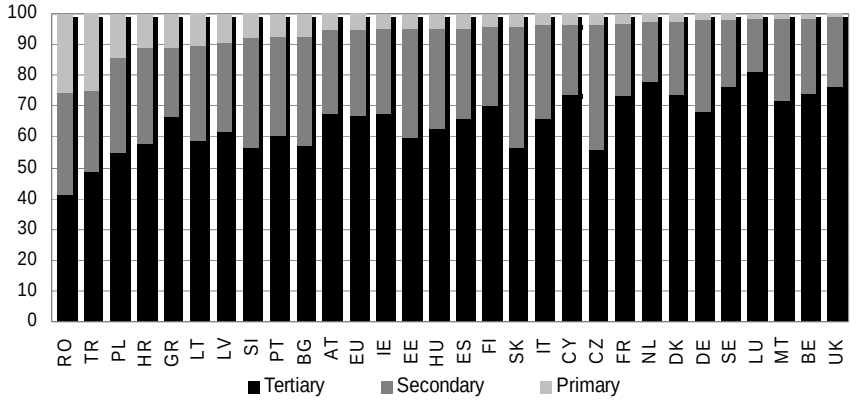


Figure 1. Share of employed persons by sector* (%), 2007

Source: Eurostat.
*Primary sector: NACE A-B; Secondary sector: NACE C-F; Tertiary sector: NACE G-Q.

Another indicator of the qualification structure of labour force could be sectoral distribution of employed population. Figure 1 shows these data for the EU countries plus Croatia and Turkey. Evidently, most advanced countries of the EU have the highest share of employed in tertiary sector. For some of the EU new member states (Czech Republic, Slovakia, Estonia, Bulgaria, and Slovenia) secondary sector (industry) still holds a large part of the employed, while primary sector has significant importance only in the least developed countries (Romania and Turkey). This distribution implies higher share of high-skilled workers in advanced countries and higher share of low-skilled workers in less developed countries. It will be very hard to catch-up if this structure of labour supply skills persists in the future.

2.2 The size of the labour supply

The size of the labour supply is probably the most important determinant of labour force competitiveness. Labour force could be educated and labour market institutions and regulations could be flexible, but if there are not enough people who would work, labour market will not be competitive. That is why it is very important to have adequate population structure so as to expect increasing competitiveness.

Figure 2 presents structure of the total population of the EU countries plus accession countries, Croatia and Turkey, by their economic activity. In almost all of the countries, at least 50% of the total population is inactive. This implies high share of dependent population, with the emphasis on the increasing share of dependent elderly population. In addition, among the active population different shares of employed and unemployed population are visible. Again, "old" member states show relatively better result comparing to the new ones and accession countries. Special cases are Turkey, with very low share of active population (around 35%), and Sweden, with very high share of active population (around 70%).

Nevertheless, even if we ignore the part of the inactive population bellow 15 years and above 64 years of age (working-age population), we still have very high inactivity rates. They range from 20% in Sweden to 50% in Turkey, with the average for the EU27 at around 30%.

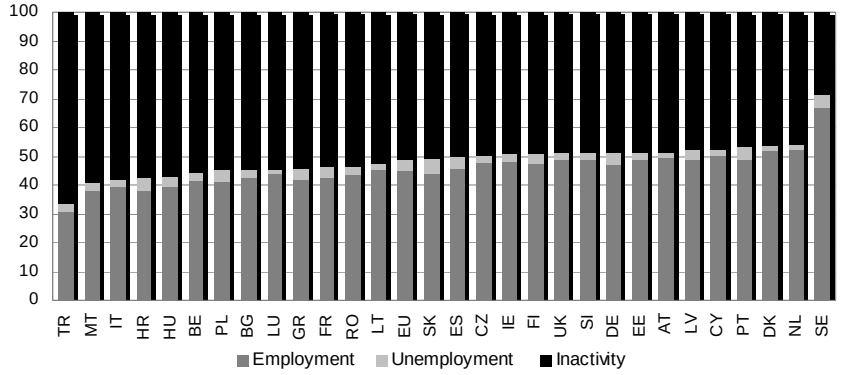


Figure 2. Total population structure by economic activity (%), 2007

Source: Eurostat.

Table 3 shows the main reasons for the inactivity among the working-age population (between 15 to 64 years.) For the most part, inactive population in this age group is still in some form of education or training. On the other hand, in countries like Czech Republic, Slovenia, and Croatia high share of inactive population in this group is already retired. This is the result of, already mentioned, early retirement schemes and disability pensions in these countries. Other reasons include illness, disability, or some family responsibilities. Interesting part of this group are those who "think no work is available" because they represent some discouraged people who would work but think that there is no vacant positions for them.

2.3 Flexibility of the labour market

Presented indicators of the labour markets in EU, and especially in CEE countries, show that the situation in most of these countries is not favourable. There is a very high rate of inactivity with inadequate skills structure of the labour supply. The question that arises here is – can this bad situation in the labour market be the result of its rigidity? Quite often, low level of competitiveness in the labour market is connected with its rigidity (Nestić, 1998; Scharle, 2003). The solution to the problem is liberalization of the labour market, i.e. its flexibility. But, what kind of flexibility, or more precisely – how far should the liberalization of the labour market go? For example, labour markets are never flexible enough for the employers, while employees like to have higher degree of protection in the labour market (Cazes and Nesporova, 2007).

GEO	Own illness or disability	Other family or personal responsibilities	Looking after children or incapacitated adults	In education or training	Retired	Think no work is available	Other reasons
EU	12,7	9,6	8,2	32,3	20,7	4,1	12,2
EU15	11,5	11,1	8,6	30,4	21,1	4,0	13,3
NMS10	14,0	10,5	9,1	38,8	20,1	2,9	6,2
BG	13,7	6,4	7,3	33,8	19,6	11,6	7,6
CZ	3,8	0,6	15,2	41,0	37,1	0,5	1,7
EE	17,9	0,0	13,6	47,8	13,0	3,8	0,0
CY	11,9	28,3	9,4	38,8	7,9	1,6	2,1
LV	13,8	7,8	8,1	41,3	15,0	4,7	8,9
LT	17,9	6,2	4,5	49,7	12,4	2,1	7,2
HU	16,0	1,9	10,6	33,5	26,3	4,2	7,4
MT	6,8	34,3	7,3	20,5	15,0	0,0	55
PL	19,9	7,8	7,6	36,4	18,8	5,2	4,4
RO	11,4	10,3	2,7	32,7	13,0	3,2	26,8
SI	16,2	5,9	1,9	37,6	32,4	2,0	4,0
SK	15,4	1,6	12,9	41,7	23,4	1,6	3,3
HR	8,4	12,3	2,8	35,6	32,8	5,6	2,5
TR	5,3	54,8	3,8	16,2	9,8	3,0	7,0

Table 3. Main reasons for not seeking employment among inactive population (%), 2007

Source: Eurostat.

European countries are often considered to be highly socially protective, particularly in comparison with the US or some other advanced OECD countries. However, for the former transition countries employment protection legislation, together with other policies and institutions in the labour market, was even more rigid. These countries had no exposure to international competitive environment and were socially protected up to a degree that market-based economies could never achieve. After the transition process started, one of the main issues was to deregulate labour market. According to Cazes and Nesporova (2007), already by 2004 employment protection legislation was on average more liberal in CEE countries than in the "old" EU countries, but still more rigid than in OECD countries. Yet, this did not bring expected improvements in other indicators of labour market competitiveness.

Same research (Cazes and Nesporova, 2007) proved that there is no statistically significant influence of employment protection legislation and collective barging on the trends in employment and unemployment. So, they propose a combination of flexibility and security – flexicurity - in the labour market in order to increase

competitiveness. Comparing this research to some other propositions (Scharle, 2003), the correlation between labour market flexibility and its competitiveness remains ambiguous.

Clearly, to achieve competitiveness in the labour market one has to be aware of many things that influence it. This chapter indicated the most important factors in determining labour force competitiveness – skills structure of the labour force, labour market institutions/regulations, and the size of the labour supply. However, the rest of the paper will be dedicated only to the third factor – the size of the labour force – with special accent on the size of the Croatian labour supply.

3. Croatian labour supply

Croatian labour market is considered to be pretty rigid, actually it is among the countries that have the most rigid employment protection legislation in Europe (Rutkowski, 2003). With inadequate qualification structure of its labour force (Bejaković, 2004) and with very low activity the situation seems even worse. In addition, demographic ageing is even deteriorating the current situation. As Obadić (2004) states, this implies low utilization of human resources what leads to lower levels of production and economic welfare, and finally, to lower level of competitiveness in the labour market.

Figure 3 shows some recent trends in Croatian labour market. It is noticeable that activity rate stayed quite constant during the observed period (2002-2007). However, the structure of the active population changed. Employment rate increased while the unemployment rate fell during this period. The positive trend in the labour market is a result of high economic growth during the same period . However, due to the world financial crisis, this trend should slow down in 2008 and 2009. One more interesting thing observable in Figure 3 is the share of long-term unemployment (>1 year) in total unemployment. This share crossed over 60% of total unemployment in 2007, which is definitely something to worry about.

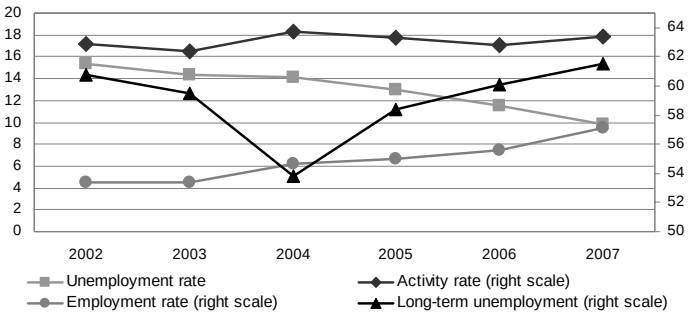


Figure 3. Trends in Croatian labour market (% 15-64), 2002-2007

Source: Eurostat.

Evidently, the labour supply in Croatia is low. Employment and activity rates show that Croatian labour force is well underused what surely undermines its competitiveness. This problem is even more evident if we look at a specific age and education groups. (Figures 4a and 4b).

Low employment rates for the two end age-groups (15-24 and 55-64) indicate low flexibility of the Croatian labour market with inexistence of the possibility for part-time employment or some other schemes for those in education or those with decreased abilities. In addition, the employment rate for female population is lower in each age group with the highest difference in these end-groups.

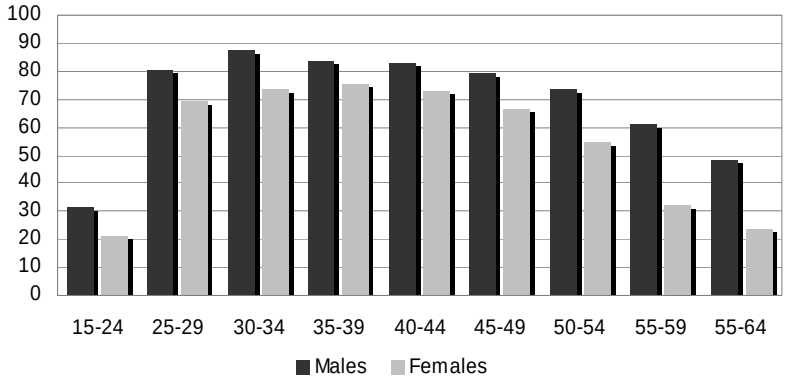


Figure 4a. Employment rate by age cohorts (%), 2007

Source: Eurostat.

Figure 4b shows that females are equal to males when it comes to employment rates only in the group of people with tertiary education. Here, the employment rate is well above average (around 80%), while for those with only primary education the employment rate is 40% for males, and 30% for females.

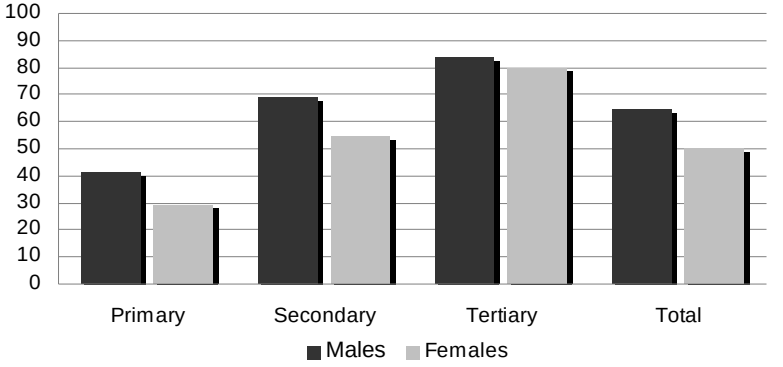


Figure 4b. Employment rate by highest level of education attained (% 15-64), 2007

Source: Eurostat.

All this shows the scarcity of labour supply in Croatia. There could be a number of reasons for this situation. Perhaps labour demand is too low, and the discouraged workers are becoming inactive. Possibly different benefits schemes motivate people to go out of the labour force or, maybe, there are just not enough people who could actually work in Croatia. Following chapters focus on the effect that labour costs have on the labour supply with the empirical results and policy recommendations.

4. The impact of labour costs on the size of labour supply

Main issue that we are concerned about here is the justification of the size of the labour costs, with emphasis on Croatia. It is well known how the price of labour is less important in modern, technologically advanced economies. However, in labour-intense industries this factor is very important. Being that Croatia is still behind the most advanced economies and competes for the foreign investments with other countries in Eastern Europe and its total labour costs are relatively high. Of course, the impact of labour costs onto the size of the labour force is highly dependent on the elasticity of supply and demand curves. For example, social security contributions paid by employers usually affect labour supply while contributions paid by employees influence demand in the labour market (Nestić, 1998). However, this is very hard to determine for the entire economy and it is most often calculated on a sectoral basis (Gora, Radziwill, Sowa, and Walewski, 2006). Still, the results show that the elasticity is higher for labour-intense industries with lower-skilled workers. As already shown (Figure 1), many of the NMS countries, together with Croatia and Turkey, have a relatively high share of employed persons in primary and secondary sector what implies greater influence of labour costs onto the size of the labour supply.

Comparing Croatia with EU countries for the year 2007 it is easily observable that Croatia has one of the highest total labour costs. This is even more evident if we compare costs and the productivity of labour (Figure 5). In this case, the differences are even greater. Again, FDI in Croatia comparing to some of the NMS is much lover . One of the reasons for that could be high labour costs. However, looking only at total labour costs, one can be easily misled due to the fact that total labour costs are composed quite differently in each country.

If we want to determine the impact of the total labour costs on the size of the labour force - we should differentiate between total labour costs and net earnings. Here, the difference between net earnings in Croatia and the selected countries (Table 4) is somewhat smaller. Moreover, structure of labour costs differs greatly among countries. For instance, while for both the EU15 and the NMS10 social security contributions paid by employer make the highest share in the difference between total labour costs and net earnings, for Croatia and Slovenia the biggest part of this difference constitute social security contributions paid by employees.

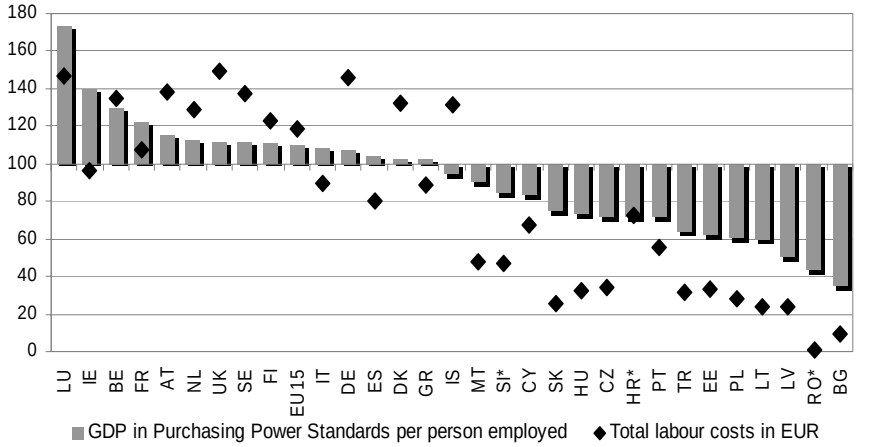


Figure 5. Labour productivity and total labour costs (EU27=100), 2007

Source: Eurostat.
*Total labour costs for Romania, Slovenia and Croatia are for 2006.

GEO	Total labour costs	Gross earnings	Taxes	Social Security-employees	Social Security-employers	Net earnings
EU25	106,08	105,43	105,75	105,25	109,61	105,42
EU15	118,98	118,51	121,19	116,67	121,56	118,51
NM10	29,88	28,18	14,58	37,85	39,09	28,13
BG	9,15	8,72	0,00	7,99	11,45	9,92
CZ	34,53	30,31	15,12	27,95	57,35	32,57
EE	32,89	29,23	41,99	5,61	52,64	31,79
CY	67,38	75,11	0,00	34,91	25,58	91,39
LV	23,56	22,50	38,65	14,94	29,30	21,81
LT	23,58	21,30	39,22	4,71	35,93	21,98
HU	32,69	28,54	16,68	35,80	55,09	28,71
MT	47,41	52,23	2,09	32,19	21,35	61,91
PL	28,01	27,56	10,51	50,14	30,44	25,66
RO	1,19	1,09	0,69	1,35	1,68	1,09
SI*	46,61	47,97	32,29	76,98	39,64	44,4
SK	25,89	24,31	4,02	24,03	34,43	26,84
TR	31,96	31,17	43,74	34,49	36,23	29,00
HR*	72,83	74,25	74,92	107,84	65,55	68,09

Table 4. Structure of total labour costs (EU27=100), 2007

Source: Eurostat, CBS.
*Data for Romania, Slovenia and Croatia are for 2006.

In this way, we can define the so-called tax wedge, which measures the difference between the costs of labour that are paid by the employer and the net salary that goes to the employee. The effect of labour costs on the size of the labour supply and, consequently, on the competitiveness of labour force, is best substituted by this measure. It is shown (Cazes and Nesporova, 2007) how high salary taxes contribute to, not only to higher unemployment, but also to lower participation and employment in the labour market. Thereby, one can determine how much the government can influence human resource competitiveness, via the size of its labour supply, by increasing/decreasing taxes and/or social security contributions.

Following figures (Figure 6a 6b) show simple regressions of the effect that tax wedge (as a percentage of total labour costs) has on the employment and the unemployment rate. We can conclude that those countries with higher tax wedge typically have lower employment and higher unemployment rates. Even though this analysis is too simple to give straightforward conclusion, the same result has already been confirmed by some other empirical analyses (Dolenc and Vodoprivec, 2005).

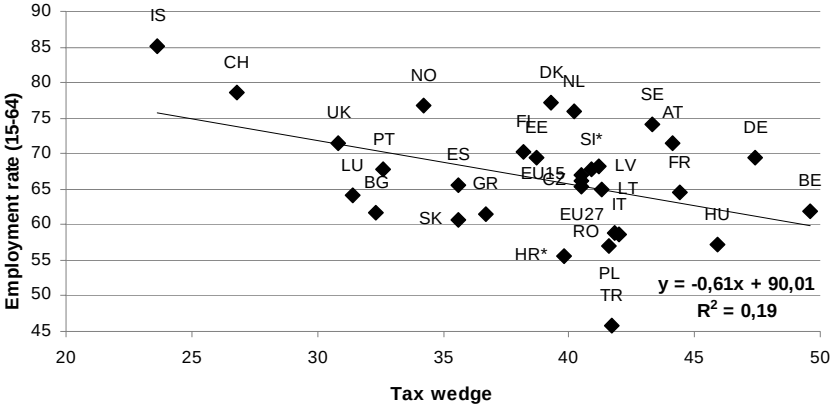


Figure 6a. Tax wedge effect on the rate of employment (15-64), 2007

Source: Eurostat, CBS.
Data for Slovenia and Croatia are for 2006.

The positive correlation between tax wedge and unemployment rates (Figure 6b) together with the negative correlation between tax wedge and employment rates among countries (Figure 6a) are considered in a deeper and accurate empirical analysis, i.e. cluster analysis. This analysis gathers cases in relatively homogenous groups based on some dependent characteristics. In our case, the dependent characteristics are tax wedge, employment and unemployment rates in 2006, while cases are represented by an efficient sample of 38 countries, i.e. members of the European Union – 27 countries; members of the Organisation for Economic Co-operation and Development (OECD) that are not part of the EU – 10 countries; and Croatia. Data are sourced from the Eurostat official data, OECD official data and the Croatian Bureau of Statistics official data plus additional authors' calculations (used data are available in Appendix 1).

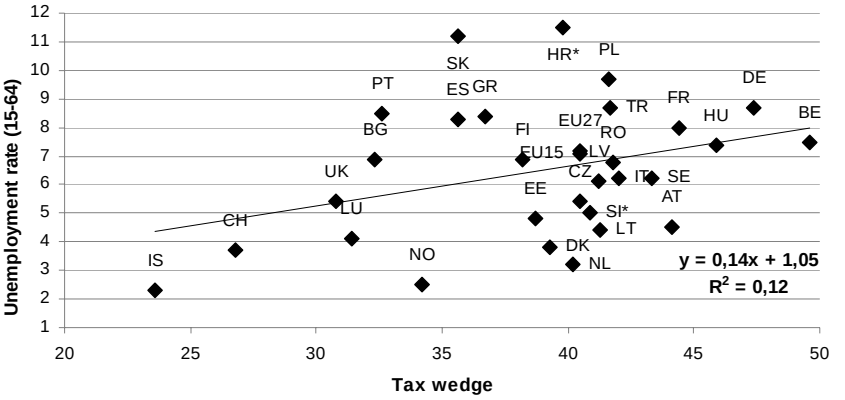


Figure 6b. Tax wedge effect on the rate of unemployment (15-64), 2007

Source: Eurostat, CBS.
Data for Slovenia and Croatia are for 2006.

Hierarchical clustering using Ward method and squared Euclidean distance as a measure function, grouped countries into two main clusters depending on the size of tax wedge, employment and unemployment rates. Table 5 shows some descriptive statistics for the two clusters.

The first cluster is a group of twelve countries that registered a lower tax wedge and a higher employment rate. Among these countries, three are members of the European Union (United Kingdom, Malta and Cyprus), one more is still from the European continent (Switzerland), while the majority are highly developed countries members of the OECD. The second cluster gathered 26 countries, i.e. 68% of the whole sample, with higher tax wedge and higher unemployment rate. Countries from this cluster all registered a high unemployment rate but with a great variance between tax wedges among countries. Figure 7 shows the hicherarchical clustering using the dendogram.

As dendogram shows, this cluster is made of two additional "sub-clusters", first one with 14 countries and an average tax wedge of 41,5% of total labour costs, and the second one with 12 countries and an average tax wedge of 45,4% of total labour costs. Evidently, Croatia is among the countries with higher tax wedge and higher unemployment rate.

Ward method		Mean	Std. deviation	Countries
Cluster 1	Employment rate	70.9875	7.49794	12 countries
	Unemployment rate	4.6250	1.13388	AUS US JP UK CAN CH IS IE KO NZ CY MT
	Tax Wedge	25.2083	6.65110	
Cluster 2	Employment rate	64.1423	7.16509	26 countries
	Unemployment rate	7.6654	2.64846	ES LT CZ LV SI EE LU PT AT SE FI DK NL NO DE FR BE HU GR RO IT SK HR PL BG TR
	Tax Wedge	42.8923	5.45395	
Total	Employment rate	66.3039	7.86128	38 countries
	Unemployment rate	6.7053	2.67824	
	Tax Wedge	37.2079	10.13139	

Table 5. Cluster statistics using Ward method in hierarchical clustering

Source: Authors' calculation

In order to show the appropriate values of tax wedge, unemployment and employment rate in the cluster centres and the distance between these centres, K-means cluster analysis is carried out. This analysis showed the same cluster participation for all countries except Norway that switched place from cluster two to cluster one, meaning from a cluster with high tax wedge to the one with lower tax wedge. This is due to the fact that despite the relatively high Norwegian tax wedge it registers also a high employment rate. That is why Norway may be called a "marginal case" and two different ways of clustering gathers it in a different groups. Table 6 shows values at initial and final cluster centroids.

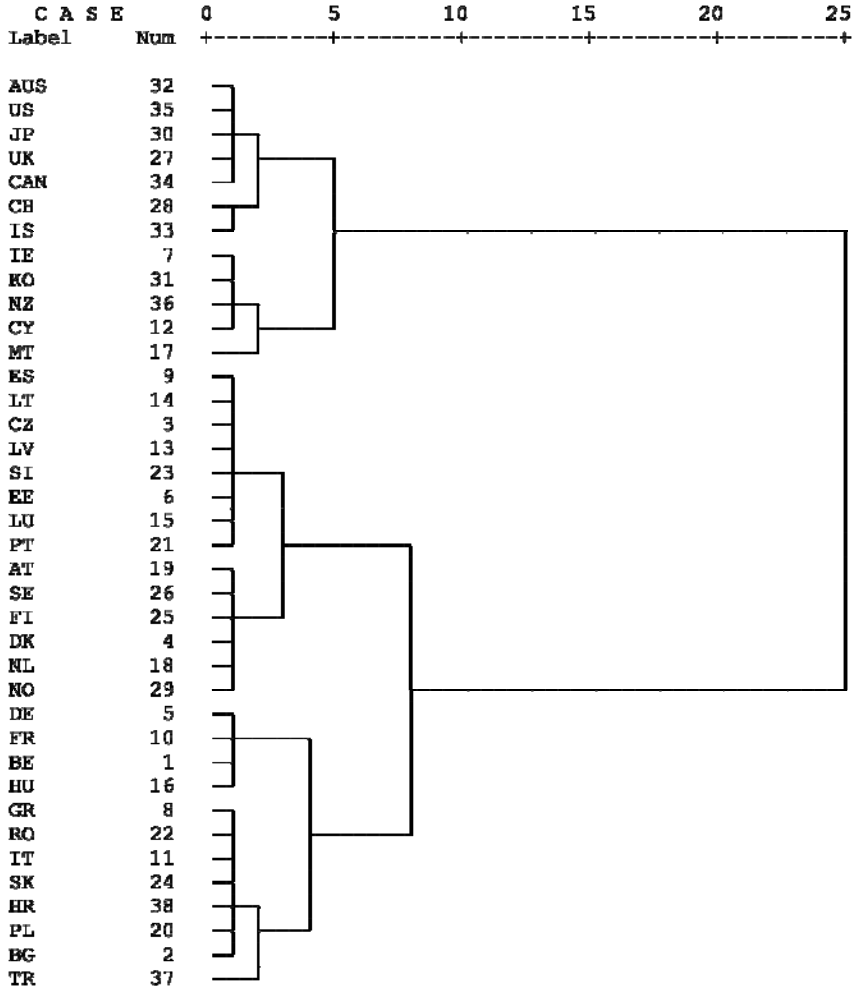


Figure 7. Dendogram using Ward method in hierarchical clustering

Source: Authors' calculation

K-means clustering		Initial Cluster Centres	Final Cluster Centres	Countries
Cluster 1	Employment rate	69.60	71.33	13 countries AUS US JP UK CAN CH IS IE KO NZ CY MT NO
	Unemployment rate	4.70	4.53	
	Tax Wedge	11.90	26.14	
Cluster 2	Employment rate	61.00	63.69	25 countries ES LT CZ LV SI EE LU PT AT SE FI DK NL DE FR BE HU GR RO IT SK HR PL BG TR
	Unemployment rate	8.30	7.84	
	Tax Wedge	55.40	43.12	

Iteration History				
	Iteration 1	Iteration 2	Iteration 3	
Cluster 1	13.802	1.433	0.000	Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is .000. The current iteration is 3. The minimum distance between initial centers is 44.488.
Cluster 2	12.532	0.745	0.000	

Table 6. K-means Clustering statistics

Source: Authors' calculation

The initial cluster centres are vectors with values based on the tax wedge, employment and unemployment rates, which refer to the countries with lower tax wedge (cluster 1) and countries with higher tax wedge (cluster 2). Final cluster centres are reached doing three iterations, when the process of redistribution of the units stops and there are no further changes to the cluster centres . The variability among countries is minimised when the centre of the tax wedge in cluster one assumes the value of 26,14% of total labour costs, while the centre of the tax wedge in cluster two assumes the value of 43,12% of total labour costs.

Both, hierarchical and K-means clustering gave the same result regarding clustering countries in different groups. One country membership can also be obtained using a discriminant function. The discriminant analysis based on all given countries except Croatia, i.e. 37 countries, resulted in the following function to predict a membership in the two mutually exclusive forecasted clusters:

$$DF = -2{,}806 - 0{,}053ER + 0{,}040UR + 0{,}162TW,$$

where: DF – discriminant function; ER – employment rate; UR – unemployment rate; TW – tax wedge;

Countries with a positive value of the discriminant function are those with high tax wedge and high unemployment. The higher the tax wedge, the higher the probability that the country will be classified in the cluster with higher unemployment rate, because it is already known that higher taxes reduce employment. The discriminant function for Croatia using available data in 2006 (see Appendix 1) is:

$$DFhr = – 2{,}806 – 0{,}053*55{,}6 + 0{,}040*11{,}5 + 0{,}162*39{,}8 = 1{,}1548$$

Croatia's membership to the cluster with lower employment rate due to higher tax wedge was shown once again.

5. Practical implications and recommendations

Croatia is clustered among countries with higher tax wedge and higher unemployment rate. Lowering the tax wedge by decreasing the income tax or social contributions, would result in a "better" position for Croatia.

This time of world economic crisis affects Croatia as every other country. It is theoretically known that a government should opt for expansive fiscal or monetary policy to retrieve the economy out of crisis in these times. Expansive fiscal policy also means lower tax burden. Croatian government could reduce social security contributions, as they have the greatest share in the total tax burden. But the expenditure side of the budget has also to be seen. Lower social contributions' revenue would surely, at least in the short run, result in a cut off on the budget's expenditure side.

Knowing that the Croatian budget is "socially oriented", because a great part of the expenditures are addressed to socially endangered groups, lower revenues would even worsen the situation. It is also important to mention that Croatia might be seen as a "special case" because, as table 7 shows, there is an important gap between the registered and the LFS (based on ILO methodology) unemployment rate in Croatia, comparing to other CEE countries.

The average difference between the registered unemployment rate and the LFS unemployment rate for Croatia in the selected period shown in table 7 is 5,5% which is much higher than the average of every other shown country. It means that a large number of workers in Croatia are registered as unemployed, and working for their employees without being registered. This might also occur when total labour costs are high as it is in the Croatian case (Nestić, 1998). Lowering the labour costs, i.e. income tax and/or social contributions, may, especially in this case, benefit in the long run. Labour taxes reduction should increase employment and result in higher competitiveness of the labour supply, and consequently in higher competitiveness of the entire economy. If competitiveness is measured regarding labour force and labour costs, it might be said that Croatia should reduce labour costs in order to enhance its competitiveness especially among South and East European countries as its main competitors.

2000		2001		2002		2003	
Registered	ILO	Registered	ILO	Registered	ILO	Registered	ILO
8.8	8.8	8.9	8.1	9.8	7.3	10.3	7.8
8.6	6.4	8.0	5.7	8.0	5.8	8.3	5.9
15.1	16.1	17.4	18.2	20.0	19.9	20.0	19.6
17.9	18.6	18.6	19.2	17.5	18.5	15.6	17.5
12.0	7.0	11.8	6.4	11.3	6.3	11.0	6.7
17.9	16.9	17.3	19.7	16.3	17.8	13.5	13.7
10.5	7.1	8.8	6.6	7.8	8.4	7.2	7.0
22.3	16.1	22.8	15.8	21.3	14.8	18.7	14.3
2004		2005		2006		2007	
Registered	ILO	Registered	ILO	Registered	ILO	Registered	ILO
9.5	8.3	8.9	7.9	7.7	7.1	6.0	5.3
9.2	6.1	9.1	7.2	9.2	7.5	10.1	7.4
19.0	19.0	17.6	18.0	14.8	13.9	11.2	9.6
13.1	18.1	11.4	16.2	9.4	13.3	8.0	11.0
10.4	6.3	10.2	6.6	8.6	6.0	7.3	4.9
12.2	12.0	10.7	10.1	9.1	9.0	6.9	6.9
6.3	8.0	5.8	7.0	5.2	7.2	4.1	6.4
18.5	13.8	17.8	12.7	16.7	11.2	14.4	9.6

Table 7. Registered and LFS unemployment rate (%), 2000 – 2007

Source: Croatian National Bank

6. Conclusions

The European labour market, on average, proved to be quite rigid, with high share of inactive population and inadequate qualification structure of its labour force. Croatia is one of the countries with highest inactivity rates, highest share of employed persons in primary sector and highest share of retired persons among the working-age population. Additionally, Croatia has one of the highest total labour costs, especially comparing with labour productivity.

This paper demonstrates the theoretical implication that the reason of high unemployment rate can be found in the high tax wedge. Cluster analysis showed that high unemployment rate is evidenced in countries with higher tax wedge, while higher employment rate is registered in countries with lower tax wedge.

Labour taxes reduction positively enhance employment and result in higher competitiveness of the labour supply, and consequently in higher competitiveness of the entire economy. When competitiveness is measured regarding labour force and labour costs, it might be said that Croatia is not in a favourable position among other South and East European countries as its main competitors, and should reduce labour costs, in order to enhance its competitiveness.

Croatia is a specific case concerning unemployment, because the difference between the registered and LFS unemployment rate for Croatia is very high, much higher than in other EU or OECD countries. This is another confirmation that high labour costs result in higher unemployment rate, especially in Croatia.

By reducing the tax wedge, Croatia would probably face more challenges than benefits in the short run, because of the decreased budget revenues. But, in the long run it should benefit from a lower tax burden. The labour market would adjust to the labour cost changes, employment should increase while unemployment would show a decreasing trend. This way, not only Croatia would result in a higher labour supply competitiveness, but it should also enhance the budget revenue level prior to lower tax wedge, when employment rates were low and unemployment rates high.

Opting for such an expansive fiscal policy measure, i.e. lower personal tax wedge, especially in the time of a new world economic crisis and for a labour intense industry such as Croatian, together with other expansive fiscal and/or monetary measures, government would surely boost its economy.

References

1. Ark, B. van, Stuivenwold, E., and Ypma, G. (2005) 'Unit labour costs, productivity and international competitiveness', GGDC Research Memorandum 200580, Groningen Growth and Development Centre, University of Groningen. Obtained through the Internet: http://ggdc.eldoc.ub.rug.nl/FILES/root/WorkPap/2005/200580/gd80.pdf, [accessed: 7/2/2009].
2. Behar, A. (2009) 'Tax wedges, unemployment benefits and labour market outcomes in the new EU members' AUCO Czech Economic Review 3, pp. 69-92. Obtained through the Internet: auco.cuni.cz /mag/article/download/id/59/type/attachment, □accessed: 2/2/2009□
3. Bejaković, P. (2004) 'Labour Force Competitiveness in Croatia: Status and Problems', In: Bejaković, P. and Lowther, J. (eds.), The Competitiveness of Croatia's Human Resources, (pp. 1-13), Zagreb: Institute of Public Finance.
4. Cazes, S. and Nesporova, A. (2007) Flexicurity: a relevant approach in Central and Eastern Europe, Geneve: International Labour Office.
5. Daveri, F. and Tabellini, G. (1997) 'Unemployment, Growth and Taxation in Industrial Countries', IGIER Working Paper, No. 122. Obtained through the Internet: http://ssrn.com/abstract=52787 or DOI: 10.2139/ssrn.52787, [accessed: 26/1/2009].
6. Dolenc, P. and Vodoprivec, M. (2005) 'Porezni klin u Sloveniji: međunarodna usporedba i preporuke za ekonomsku politiku', Financijska teorija i praksa, Vol. 29, No. 3, pp. 283-297.
7. Gora, M., Radziwill, A., Sowa, A., and Walewski, M. (2006) 'Tax Wedge and Skills: Case of Poland in International Perspective', CASE Reports, No. 64. Obtained through the Internet: http://www.case.com.pl /upload/publikacja_plik/10223856_rc64.pdf, [accessed: 26/1/2009].

8. Muysken, J., Van Veen, T., and De Regt, E. (1999) 'Does a shift in the tax burden create employment?', Applied Economics, Vol. 31, No. 10, pp. 1195-1205.
9. Nestić, D. (1998) 'Tržište rada i oporezivanje: smanjenjem doprinosa protiv nezaposlenosti', Privredna kretanja i ekonomska politika, Vol. 8 No. 63, pp. 27-62.
10. Nickell, S. (2003) 'Employment and Taxes', CESifo Working Paper Series, No. 1109. Obtained through the Internet: http://ssrn.com/abstract =489443, accessed 27/1/2009
11. Obadić, A. (2004) 'Comparison of Basic Macroeconomic Indicators on the Labour Markets of a Select Group of Countries', In: Bejaković, P. and Lowther, J. (eds.), The Competitiveness of Croatia's Human Resources, (pp. 47-59), Zagreb: Institute of Public Finance.
12. Rutkowski, J. (2003) 'Analiza i prijedlozi poboljšanja tržišta rada u Hrvatskoj, Financijska teorija i praksa, Vol. 27, No. 4, pp. 495-513.
13. Scharle, A. (2003) 'Competitiveness and the Labour Market', Ministry of Finance Working paper, No. 4. Obtained through the Internet: http://pdc.ceu.hu/archive/00002571/01/4_eng_040205.pdf, [accessed: 24/1/2009].
14. Tica, J. and Jurčić, Lj. (2006) 'A Relative Unit Labor Cost: Case of Accession Countries', FEB - Working paper series, No. 6. Obtained through the Internet: http://web.efzg.hr/RePEc/pdf/Clanak%2006-06.pdf, [accessed: 30/1/2009].
15. Vehovec, M. (2008) 'New Perspectives on Extending Working Life: Challenges for Croatia and Slovenia', In: Vehovec, M. (ed.) New Perspectives on a Longer Working Life in Croatia and Slovenia, (pp. 19-34), Zagreb: The Institute of Economics, Zagreb and Friedrich Ebert Stiftung.

Is the Choice Experiment Suitable Approach for Expert Support of Decision-Making?

Ondřej Vojáček[1], *Iva Pecáková*[2]

[1]*Department of Environmental Economics, University of Economics in Prague, W. Churchilla 4, 130 67 Praha 3,* ***ovojacek1@centrum.cz***

[2]*Department of Statistics, University of Economics, Prague, W. Churchilla 4, 130 67 Praha 3, pecakova@vse.cz*

In the paper discrete choice models are discussed and applied to empirical data. The main goal of the paper is to find out whether the various discrete choice models provide the analyst with robust and reliable estimates of values of natural goods or values of changed quality of such goods. Our results indicate that the yielded value estimates using the choice experiment method and discrete choice models are applicable in expert support of decision-making on allocation of public resources to such goods, because they are stable and robust. These issues are investigated for the marginal willingness of Czech visitors to the Macha lake beaches to pay for the water quality and beach characteristics.

Keywords: Logistic regression, choice experiment, welfare measure, random utility theory, public goods

JEL: C01, C13, C16, D61, H40

1. Introduction

Neo-classical environmental economics assumes that the market provides insufficient quantities or quality of public environmental goods. The solution offered by this economic school is the additional allocating of public funds to such goods (Hampl, 2001). The economic question is, however, what is the optimum rate of resources to be invested from public funds. The answer to that question largely derives from the value of the natural goods which is being (in neoclassical economics paradigm) derived using the methods developed for that purpose. One of these methods is called choice experiment (Bateman et. al., 2002). The choice experiment enables to derive the values of natural resources or the values of the changes of the quality of these resources. For the value estimates using the choice experiment method the discrete choice models are used. A lot of attention is being nowadays paid to develop and refine these models.

In this paper discrete choice models are discussed and applied to empirical data. These issues are investigated for the marginal willingness of Czech visitors to the Macha lake beaches to pay for the water quality and other beaches characteristics.

Given the large sums of money allocated world-wide to the conservation and renewal of natural resources, the authors feel free to pronounce the opinion that the debate contributing to the understanding of the discrete choice models and their application is very useful to economics.

The main objective of this paper is to apply various types of discrete choice models on a practical example of primary data analysis gained from environmental research and discuss obtained results. The main hypothesis of this paper is that regardless of the discrete choice model applied, the analyst gets very similar results in respect of the model parameters and welfare measures.

The structure of the paper is as follows. The economic foundation of the choice experiment and discrete choice models is explained in the first chapter. Section two summarizes the site characteristics, sample strategy, survey design and variables used in the choice experiment. In the third part selected discrete choice models are briefly discussed and applied to the data. The same chapter discusses the results of the model estimates in terms of their fitness and estimated values. The last (fourth) chapter concludes.

2. The Economic Foundation of the discrete choice models

Discrete choice modelling complies with Lancaster's new approach to consumer theory and with the random utility theory (McFadden, 1974). According to this theory consumers derive their utility not from the product as such, but from the attributes by which the product can be described. According to Lancaster, each product can be described in terms of its attributes (Lancaster, 1966).

The second foundation of the choice experiment method is the random utility theory (RUM). It is based on the hypothesis that individuals make choices based on the attributes of the alternatives (an objective component) along with some degree of randomness (a random component). This random component arises either because of randomness in the preference of the individual or because of the fact that choice experiment attributes do not include all preferences of the individual.

The obvious objective in the discrete choice modelling is to analyze the individual's choice in relation to the characteristics of the product. Application of the choice experiment method is connected to a sociologic survey. In the survey several different products described in terms of their attributes are offered to the respondents. One of the attributes is always the price of the product or similar measure of the value. The consumer then has to make a „trade-off" between the changes in the attribute levels and the price. One of the possibilities always offered to the consumer has to be the "opt out/status quo" alternative (no change at no costs). An "opt out" alternative is essential for the prediction of the demand. The individual utility maximization problem can be written as follows:

$$\max_{Z,m\in C} U_i^*(Z, X_m),$$
$$while...ZP \le Y_i - c_m$$

Where: U_i is the utility function,
Z is a vector of consumer goods,
P is the corresponding price vector,
Y_i is the income,
X_m is the attribute vector of the m-th alternative,
c_m is the cost of the m-*th* alternative,
C is the choice set.

If we express the conditional indirect utility function of the individual *i* as follows:

$$U_{im}(P, Y_i - c_m, X_m), \text{ then}$$

under the assumption of the linearity of the indirect utility function, the linear utility function can be expressed as follows:

$$U_{im} = P\lambda + X_m\beta + \gamma(Y_i - c_m) + \alpha_i + \varepsilon_{im}, \text{ where}$$

U_{im} is the utility of the individual *i* from the alternative *m,*
λ and β are coefficient vectors,
γ is the marginal utility of income,
α_i is an individual specific component,
ε_{im} is an unknown individual/alternative specific component.

The respondent *i* choose alternative *m* over alternative *j* if:

$$U_{im} > U_{ij}, \forall_j \in C; j \neq m.$$

As can be seen the decision of the alternative is not affected by P λ, γ Y_i, α_I, so the decision rule (for the individual respondent) can be expressed as:

$$X_m\beta - \gamma c_m + \varepsilon_{im} > X_j\beta - \gamma c_j + \varepsilon_{ij}, \text{where}$$
$$\forall j \in C; j \neq m$$

Now if we include the random utility theory (which is based on the hypothesis that individuals make choices based on the attributes of the alternatives along with some degree of randomness) and consider the random sampling of the respondents, then ε_{im} $(m = 1,\ldots,M)$ can be interpreted as random terms. The probability of the respondent choosing alternative *m* can be than written as:

$$P(m:C,\beta,\gamma) = P(X_m\beta - \gamma c_m + \varepsilon_{im} > X_j\beta - \gamma c_j + \varepsilon_{ij}; \forall j \in C; j \neq m)$$

The random component arises either because of randomness in the preferences or simply because the choice experiment attributes do not cover all preferences of the individual. If the individual chooses the alternative *m* over the alternative *j*, then U_m must be greater than U_j. Based on repeated observations of the choices, one can examine how the levels of various attributes affect the probability of choice. The obvious objective in the discrete choice modelling is to analyze the probability of individual's choice in relation to the changing characteristics of the (environmental) product.

Different assumptions about the distribution of the error term leads to the different econometric models. There is not much scope in this paper to discuss the econometric issues behind these models and thus with respect to the given hypothesis laid out in the paper, the attention is focused on brief model description and then results presentation and final discussion. But first of all the data, variables and sampling strategy is briefly described, then the models are briefly discussed and applied further.

2. The study area, sampling strategy and survey design

Macha Lake is situated in the Central Bohemia Region. It lies almost 100 km north of Prague. Its size is approximately 305 hectares. The average depth of the lake is 2.29 meters. There are four paid beaches and several public beaches by the lake. In good weather there are up to ten thousand visitors per day on these beaches. There are also several other access points to the lake, but only a few people choose them in comparison with the paid beaches. Close to the Lake are two small towns called Doksy and Stare Splavy.

The water eutrophication leading to Cyanobacterium problems is the main factor affecting the quality of the lake's water. This problem has been caused mainly by the lake utilization in the second half of the 20th century, when the Macha Lake was used for intensive fish farming and therefore was fertilized by super phosphates. The high nutrient content in the sediments at the Lake bottom is responsible for the Cyanobacterium problems.

In order to apply the choice experiment approach it is necessary to obtain relevant information from visitors of the recreation site. Individual data are usually obtained by administering a survey. Therefore a questionnaire was constructed. The study population was the population of the visitors to the paid beaches, which amounts approximately to 5/6 of all visitors to Macha Lake.

The survey on the Macha Lake beaches was realized from July to August 2007. A pre-test was carried out in June 2007. Several in-depth interviews were made in Prague focusing on waterside recreation. Then a pilot version of the questionnaire was prepared. Two pilot surveys (about 10 respondents per pilot) were carried out in June in order to improve and finalize the questionnaire and to test the sampling strategy in the terrain. The final survey was realized in July and August. The questionnaire was administered on-site to visitors on the four paid beaches located by the Lake. Respondents were intercepted on each of the four beaches randomly and interviewed

by trained interviewers face-to-face. Respondents were selected randomly throughout the day on each beach separately. The survey resulted in a total of 331 completed questionnaires.

The attributes and their levels used in the final sampling are plotted in Table 1. For a detailed discussion of the choice of the attributes and the choice experiment design see Vojacek and Melichar (2007).

Attribute	Attribute Levels
Beach overcrowded	Yes No
Water quality	Clear water Slightly polluted water Polluted water
Beach equipment	Yes No
Entrance Fee	CZK 40 (EUR 1.6) CZK 80 (EUR 3.2) CZK 150 (EUR 6)

Table 1. Choice experiment attributes and their levels

The explanatory variables used in the modelling exercise were as follows; see Table 2.

Variable	Description of variable	Variable values
Crowd	Overcrowded beach	1=yes; 0=no
Clear	Water pollution level – clear	1=yes; 0=no
Dirty	Water pollution level – dirty	1=yes; 0=no
Noteq	No equipment at the beach	1=yes; 0=no
Fee	Entrance fee	40; 80; 150
Opt_out	The dummy for the opt-out alternative	0 = no, 1 = opt-out

Table 2. MNL model; Explanatory variables

The parameter for Clear water ("Clear") can be interpreted as a indication of the relative change in the people's preferences in comparison with the attribute level "Slightly polluted water" (the basic attribute level). This coefficient can also be expressed as a relative change in the probability of selecting a certain locality when water quality changes from slightly polluted to clear water and all other attribute levels remain unchanged. Similarly the variable "crowd" can be interpreted as the indication of the relative change in the people's preferences in comparison with the attribute level "not overcrowded beach".

3. Discrete choice models: discussion and empirical application

3.1 Multinomial logit model

The idea of how to make a link between the economic theory and the discrete choice models was originally introduced by McFadden in 1974. He assumed the error terms are independently identically distributed (i.i.d.) with the so-called Gumbel (or type I extreme value) distribution with the cumulative distribution function. This assumption leads to the conditional (later called multinomial) logit model (MNL model), where the probability of choice of the alternative *j* by the individual *i* is:

$$\pi_{ij} = \frac{\exp(V_{ij})}{\sum_j \exp(V_{ij})} = \frac{\exp(\mathbf{x}'_{ij}\boldsymbol{\beta})}{\sum_j \exp(\mathbf{x}'_{ij}\boldsymbol{\beta})}$$

If these random variables are distributed identically and independently (IID) and follow the Gumbel (type I extreme value) distribution, then their difference follows the logistic distribution. The problem with the MNL model arises from the IID assumption. The odds of choosing an alternative *i* over an alternative *j* do not depend on the other alternatives in the choice set or on the values of the explanatory variables. The IID however is often violated and thus the simple and easy interpretable MNL models are rarely appropriate for real choice problems.

This occurs also in the Macha Lake data analysis. The MNL model performed well (see the table 3 below), however the IIA assumption was violated using the Hausman and McFadden test of the IIA (Hensher et. al., 2005).

If this occurs, although the model is statistically significant with the Log likelihood function of the value -2305.512 and all the parameters of the model are significant at the 0.1 % level and have expected sign (that is people prefer clear water, they have negative preferences towards beaches which are not equipped and prefer lower entrance fees) this model can't be consider as valid.

3.2 Nested logit model

If the IID (and thus also IIA) is violated the researcher has to continue analyzing data in some other way, using different econometric models and tests of their appropriateness. In this situation the Nested logit (NL) model can be an alternative for the analyst, while it partially relaxes the IID assumption. It enables to model the choices in the hierarchical structure. The "NL tree structures are determined so as to accommodate differences in variance or scale that exist in the unobserved components of the utility expressions (i.e. on econometric and not behavioural grounds)" (Hensher et. al., 2005).

The choice alternatives are structured into several (K) groups (so-called nests) B_1, B_2, …, B_K. IIA holds within each nest, but it does not hold for alternatives among different nests. The vector of unobserved utility $\boldsymbol{\varepsilon}_i' = [\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{iJ}]$ has a generalized extreme value distribution with the cumulative distribution function

$$F(\boldsymbol{\varepsilon}_i) = \exp\left[-\sum_{k=1}^{K} \left(\sum_{j \in B_k} \exp(-\varepsilon_{ij} / \lambda_k) \right)^{\lambda_k} \right].$$

The parameter λ_k is a measure of the degree of independence in unobserved utility among the alternatives in nest k; full independence among all the alternatives in all nests ($\lambda_k = 1$) reduces the nested logit model to a multinomial logit model and the probability of choice for the alternative $j \in B_k$ is now

$$\pi_{ij} = \frac{\exp(V_{ij} / \lambda_k) \left(\sum_{c \in B_k} \exp(V_{ic} / \lambda_k) \right)^{\lambda_k - 1}}{\sum_{k=1}^{K} \left(\sum_{c \in B_k} \exp(V_{ic} / \lambda_k) \right)^{\lambda_k}}.$$

The unobserved factors have the same correlation for alternatives within a nest and no correlation for alternatives in different nests. As can be substantiated from the formula above, the IIA holds within each of the K nests of alternatives but not across nests.

To apply the NL model to the Macha Lake data, the choice problem needs to be reformulated to the hierarchical nested structure which is not easy task with our data. It seems to be possible to be done in the only way - the respondents could be organizing the choice problem in two stages: first whether they are willing to visit any of the offered localities and second which one of the offered localities they would visit.

The NL model again performed well. All the parameters of the model were significant at the 0.1% level and they had the same signs and sizes as in the MNL (see again table 3). In specifying the NL model, the correlation between the “choice “ and “opt_out” variables needs to be restricted, so that the correlation problem between the nest and the explanatory variable is avoided. However according to the log-likelihood ratio test comparing MNL and NL models (the test criterion value 2.252 and critical chi value of 5.99) the hypothesis that the models are the same could not be rejected. In other words the NL model does not significantly improve the model estimation, what is clear also from the estimated coefficients. The test of the IV parameters of the NL model confirms again this result. The IV test indicates that the NL model collapses into a single branch which is equivalent to a MNL model. Although the NL model did not perform better that the MNL the estimated parameters can be considered to be valid.

3.3 Probit model

Except NL model, probit model (PM) provides analyst with another possibility in data analysis, when the MNL model fails to pass the IIA assumption (Hausman, 1978). It bears more flexibility compared to the MNL. It assumed the random component is normally distributed. Despite the model does not rely on the IIA assumption it is not used in discrete choice data analysis as frequently as the NL or MNL models. The reason for this may be that the model estimation is time-consuming and problems with the log likelihood maximization occur.

The basic assumption of the probit model is that the unobserved utility components are normally distributed with the density

$$f(\boldsymbol{\varepsilon}_i) = \frac{1}{(2\pi)^{J/2}\left|\Omega\right|^{1/2}} \exp\left[-0,5\boldsymbol{\varepsilon}_i'\boldsymbol{\Omega}\boldsymbol{\varepsilon}_i\right],$$

and with a mean vector of zero means and known covariance matrix $\boldsymbol{\Omega}$. The choice probability of the alternative *j* can be then expressed as

$$\pi_{ij} = F(\boldsymbol{\varepsilon}_i) = \int_{\boldsymbol{\varepsilon}_i} f(\boldsymbol{\varepsilon}_i) d\boldsymbol{\varepsilon}_i \ .$$

With a full covariance matrix, various patterns of correlation and heteroskedasticity can be accommodated according to need, so that the IID and IIA are relaxed. However, the probabilities of choice can be expressed only in the form of integrals and they must be evaluated numerically through simulation. Also, the model interpretation is not as straightforward and intuitive as in the logit models. The linear combination of observed factors – the representative utility – in this model is a probit, i.e., a percentile of the normal distribution. For the estimated probit model for the Macha Lake data see again table 3.

As is clear from the estimated model the parameters of the probit model do not differ much from those in the MNL and NL models. Also the log-likelihood ratio test indicates that the probit model does not significantly improve the model fit.

3.4 Model comparison

The models comparison is shown in table 3 below. This parameter estimate overview confirms the previous finding that the MNL, NL and Probit models bring in principle the same values of the estimated parameters despite the fact that they differ in terms of their assumptions.

All the models have a favourable fit as they reach approximately an R-square of 0.3 which equals approximately to 0.6 – 0.8 for the linear regression model. The interesting finding is, that despite the MNL model fails in passing the IIA assumption, the other two applied models did not improved the value estimates at all. It confirms the hypothesis pronounced by some other authors, that the MNL model is quite robust even in the case it doesn't pass the IID (IIA) assumption. With regard to the estimated

parameters which doesn't differ nearly at all it can be proved that also the welfare measures which would be counted based on the applied models would not significantly differ.

	MNL	NL	PM
CROWD	-0.419 (-6.159)	-0.419 (-6.159)	-0.419 (-6.357)
CLEAR	0.686 -8.639	0.686 -8.639	0.686 -8.921
DIRTY	-2.438 (-25.703)	-2.438 (-25.703)	-2.438 (-20.152)
NOTEQ	-1.991 (-22.115)	-1.991 (-22.115)	-1.991 (-17.659)
FEE	0.011 (-14.335)	0.011 (-14.335)	0.011 (-14.449)
OPT_OUT	-2.75 (-22.463)	-2.75 (-22.463)	-1.979 (-20.007)
Log likelihood	-2305.512	-2304.386	-2302.676
Adjusted R2	0.287	0.316	0.300
Observations	2997	2997	2997

Table 3: Model comparison (t-statistic in brackets)

Note: ***, **, * = Significance at 1%, 5%, 10% levels

These welfare measures can be counted as follows:

$$WTP_{\Delta X} = \beta / \gamma ,$$

where β is the variable coefficient; γ is the marginal utility of income (fee variable coefficient) (Hanemann, 1984).

Thus, the point estimates of WTP for a change in the attribute level can be derived by calculating the marginal rates of substitution between the change in a given attribute and the price attribute - that is, by dividing the coefficient of the attribute by the coefficient of the entrance fee attribute. This is done for the MNL model in the Figure 1 (for more details see Vojacek and Melichar, 2007).

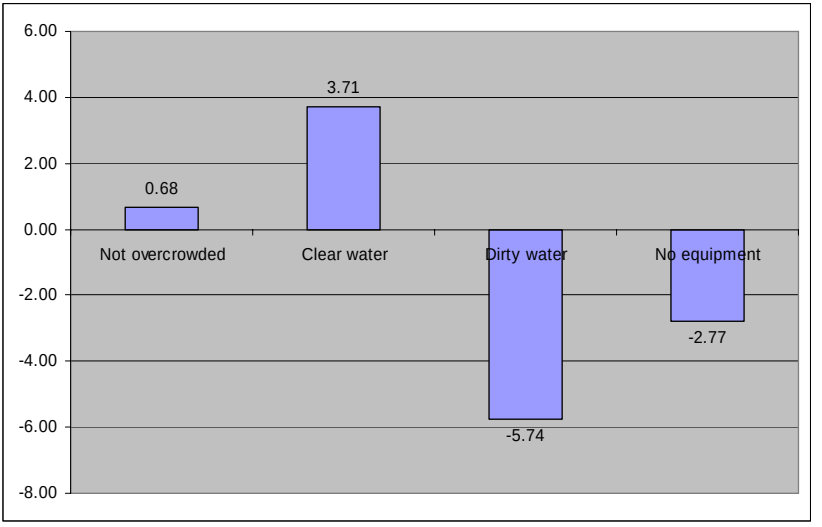


Figure 1. Marginal willingness to pay for a change in the attribute levels (EURO 2008)

4. Discussion and Conclusions

The paper has discussed, developed and compared a few discrete choice models with respect to the stated objectives of the paper. The multinomial logit model performed well. All the parameters of the model were significant at the 1% level. However, the multinomial logit model failed to pass the Hausman and McFadden test of the IIA, which is crucial for its validity.

One could then expect that the introduction of other, more advanced models would be significantly beneficial for the data analysis. However, this expectation was confirmed neither by the nested logit model nor by the probit model, which relax the discussed IID assumption within the distribution of the random component. These models did not bring any improvement to the data analysis, and also the estimated parameters of both the nested logit model and the probit model were almost identical. These findings comply with the results of the log-likelihood ratio tests and in the case of the NL model also with the test of the IV parameters. The main hypothesis of this paper was thus confirmed at least in terms of the discrete choice models applied.

Our analysis arrives at the conclusion that the choice experiment using discrete choice models is a well-developed method yielding robust estimates of values of environmental goods, and that if applied correctly, it can be used in support of decision-making regarding allocation of public funds.

The authors of the paper dare to say that also high data quality issue played significant role in the coefficient significance and robustness of the parameter estimates. Unfortunately the data quality as such is not much discussed in the other papers.

Future ambition of the authors is to further extend analysis also to the random parameters logit models (mixed logit models) and the latent class models and also carry out the cost-benefit analysis based on the particular models estimates and heterogeneity found in the data.

Acknowledgements

The research in this study was supported by the grant of the University of Economics in Prague: IGA (internal research project funding system) – Analysis of the environmental behaviour of consumers, and the research could not be done without the support of the Charles University Environment Canter in Prague. This support is gratefully acknowledged.

References

1. Bateman I. J., Carson R. T., Hanemann M., Hanley N., Hett T., Jones-Lee M., Loomes G., Mourato S., Özdemiroğlu E., Pearce D. W., Sugden R., Swanson J. (2002): *Economic Valuation with Stated Preference Techniques: A Manual*, Cheltenham: Edward Elgar, UK
2. Domencich T. and McFadden D. (1975): *Urban Travel Demand A Behavioural Analysis*, New York and Oxford: Elsevier
3. Hampl M. (2001): Trojí přístup k veřejným statkům, *Finance a úvěr*, 51(2), 111-125
4. Hanemann M. (1984): Welfare evaluations in contingent valuation experiments with discrete responses, American Journal Of Agriculture Economics, 66(3), 332 - 341
5. Hausman J.A., Wise D.A.(1978): A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences, *Econometrica 46*, 403-426
6. Lancaster K. (1966): A new approach to consumer theory, *Journal of Political Economy*, 74, 132-157
7. Louviere J.J., Hensher D.A., Swait J. (2000): Stated Choice Methods: Analysis and Applications in Marketing, Transportation and Environmental Valuation, Cambridge: Cambridge University Press
8. Manski Ch. F. (1977): The structure of random utility models, *Theory and Decision* 8 (3), 229-254
9. McFadden D. (1974): Conditional logit analysis of qualitative choice behaviour, in Zarembka, P. (ed.), *Frontiers in econometrics*, New York: Academic Press
10. Vojáček O., Melichar J. (2007): *Valuing Water Quality – Choice Experiment Approach*, In: ŠAUER, Petr (ed.). Environmental Economics, Policy and International Relations. Prague : Nakladatelství a vydavatelství litomyšlského semináře, Prague, 109–120

The Construction of a Behavioural Observation Scale for University Teachers' Evaluation

Ionescu Maria Claudia

Psychology Department, Faculty of Social-Humanistic Sciences, University of Pitesti, Str. Targu din Vale, nr. 1, Corp I, etj. IV, 110040, Pitesti, Arges, Romania, iclaudia_psiho@yahoo.com

The objective of this research paper is to create a valid and reliable instrument designed to be used for teachers' assessment in university. This instrument is a Behavioural Observation Scale (BOS), intended to be used by students as assessors of university teachers' professional performance, since students' ratings could be valid and reliable sources of teaching efficiency evaluation. The final version of BOS includes 51 items, divided into 8 dimensions or factors (Knowledge in Specialized Area, Cognitive Flexibility, Communication Ability, Pedagogical Ability, Ethical Conduct, Relationship with Students, Emotional Maturity, and Didactical Technology). Statistical analysis indicates high levels of internal consistency for all dimensions (Alpha Cronbach's coefficients range from 0.8009 to 0.9319). This instrument could be used in university teachers' assessment in order to give teachers feedback on their professional efficiency or in human resources decision-making.

Keywords: university teacher; performance evaluation; Behavioural Observation Scale; teacher's evaluation by students.

1. Introduction

Professional performance evaluation is an important process in Human Resources Management. Performance evaluation consists in systematic description of the activity of a person or group, identifying strong and weak points (Cascio, 1991). In a broad perspective, Manolescu (1998) defines performance evaluation as a cognitive action, process or activity through what an evaluator assesses or estimates an individual's performance against predetermined standards, and also against his mental representation, his values, and his view on that performance.

Various types of decision-making are based on results of performance appraisal. In higher education, performance appraisal is useful in offering teachers a feedback on the efficiency of their didactical activity. This is one way the quality of future professionals' instruction could be increased by. Performance appraisal includes also

an economical problem both for university and for students, who are the beneficiaries of educational system.

2. Literature Review

Teacher's professional efficiency can be operationalized by different factors such as: teacher's actions, knowledge, creativity. Research studying this topic pointed out that teacher's performance is a multidimensional construct that includes: effective performance of the professional role (behaviours in accordance to role expectations such as holding different academic titles or using adequate teaching methods) and effective teaching (outcomes of teacher's actions on students' learning and academic performance) (Blanton, Sindelar, and Correa, 2006).

De Jong and Westerhof (2001) consider that important dimensions of teacher's behaviour include: motivation and clarity (motivating behaviour, clarity of instruction), teaching ability (offering examples, correlating different information, addressing questions), individual counselling (assisting students during class, whenever needed), group control (group management, correcting off-task behaviours), metacognition (the way teacher approaches students' tasks and errors).

Another conceptualization of teacher' performance consists in four main areas of standards: planning and preparation of students' learning, creating a learning environment, teaching for learning, and professionalism (Holtzapple, 2003).

Teacher's behaviour is related to students' performance and represents an important condition for teaching quality. Most efficient teaching abilities, related to students' performance, are considered to be: effective use of error correcting procedures, use of specific feedback, efficient time management, clear knowledge presentation, communication of didactical objectives, investigation of students' understanding, efficient management of students' class activity (Scheeler, 2008).

Wilkerson et al. (2000) studied the relationship between teachers' performance, assessed by 360^0 feedback technique including evaluations from students, principal and self-evaluations, and students' results on four subject areas. Students' evaluations on teachers' performance were identified as the best predictor of students' outcome.

Teacher-student relationship is considered an important factor of teacher's performance. This relationship could be influenced by several aspects: attributes of the persons involved in relationship (sex, age), the way each person represents himself this relationship, communication processes, external factors such as: class or school climate, rules, discipline, class size (Lee, 2005).

Conflict and closeness are important characteristics of teacher-student relationship. Authenticity is a concept related to the quality of this relationship (Laursen, 2005). The authentic teacher considers his students as human beings, not as a simple resource for didactical process.

Research productivity is considered to be another factor involved in teacher's performance. There are conflicting points of view regarding this matter. There are studies that identified a positive correlation between teacher's research productivity

and teaching effectiveness (Hong, Xuezhu, and Ke, 2007) or argued that teaching and research promote each other. Research activity helps teachers to gain new knowledge and to stimulate their intellectual abilities. However, the correlation of these two variables is low.

Emotional functioning was another aspect investigated in connection to teachers' performance. Sutton (2004) studied the goals teachers regulate their emotions for and emotional regulation strategies they use and he pointed out that teachers considered that regulating their own emotions enhances teaching effectiveness and it is consistent with the ideal imagine of a teacher, from the emotional functioning perspective.

There are different ways to define teachers' professional efficiency; therefore, there are multiple methods to measure it: observation, students' performance, checklists, standards, assessment of didactical process. Case study could be a successfully used method in teacher's evaluation, as an alternative to paper-pencils test (Denner et al., 2002).

Teachers' evaluation by students is widely used with the finality to improve instructional process and to facilitate students' learning activity. Practical objectives of evaluation pursue different aspects: students' evaluations are a basis for faculty evaluation or for assessing faculty personnel retention, career advancement or reward. Students' evaluations are in many cases the only measure of teaching ability (Simpson and Siguaw, 2000).

Teachers can be assessed by external, independent evaluators, other than students. This is a costly activity and often a controversial one. Scores can vary as a result of differences between evaluators, although they are trained for evaluation. There are doubts about the validity of teachers' evaluations based only on hierarchical supervisor's assessments. A study that investigated a large sample of students pointed out that they can make a more subtle discrimination of teacher's performance (Peterson, Wahlquist, Bone, 2000). Other possible external raters could be colleagues. Research shows that if colleagues respect one another and trust each other, their ratings could be very helpful in mutual improvement of performance (Keig, 2000).

Teacher's evaluations by students have several advantages (De Jong and Westerhof, 2001): they are less costly than external evaluators and students' perception is based on daily interaction with teachers. This way, students can identify teacher's representative behaviours. Other points of view stress that using students as evaluators has several weaknesses: problems concerning validity and reliability of evaluations, students' ability to perceive in a differentiated manner teacher's behaviour, other factors influencing the ratings (students' number, gender), students' subjectivity (teachers with a sense of humour or those who are lenient could receive more favorable assessments). Teacher's evaluation could become an assessment of teacher's popularity (De Jong and Westerhof, 2001).

Reliability of students' evaluations could be assessed by item analysis (homogeneity) or by inter-rater reliability. Research shows that reliability of evaluation increases when students' ratings are aggregated. The more students assess teacher's performance, the more reliability coefficient increases.

Aleamoni (1999) investigated 16 myths related to students' evaluation of teacher's professional performance. These myths are discussed in reference to different research studies on this topic. Aleamoni shows, among other things, that: students' evaluations are reliable, without great variations (correlation coefficients range from 0.70 to 0.89); students are able to make fine discrimination in assessment, they can appraise differently the same teacher in areas concerning professional performance (professional preparation, organisation, stimulation of interest) in contrast to aspects related to teacher's personality (sense of humour, warmth).

The objective of evaluation does not appear to affect students' ratings. A study investigating this aspect (Fossey, 1999) showed that the validity of students' ratings was not influenced by the goal of the evaluation, which the students were informed about.

There are enough evidence to support the validity and reliability of students' ratings of teachers' professional performance. Professional efficiency raise not only the question of having well-trained evaluators, but also the problem of creating valid and reliable instruments and measurements, since the main sources of bias in performance assessment are the rater and the evaluation system.

3. Method

The objective of this research paper is to create a valid and reliable instrument designed to be used for teachers' assessment in university. This instrument is a Behaviour Observation Scale (BOS), intended to be used by students as evaluators of university teachers' professional performance. Studies investigating the characteristics of teacher assessment by students show that this appraisal is valid and reliable and it is used for performance improvement.

Participants

This research involved 100 participants, divided into 2 groups.

The first group was formed by 50 university teachers who were asked to participate in this study as experts. Almost all of them were female teachers (96%). The participant ranged in age from 28 to 45 years (M=38.56, SD=5.35). A single criterion was used in order to select this sample. This consisted in a minimum 5 years period of teaching experience. Participants' teaching experience ranged from 5 to 15 years (M=12.4, SD=3.04)

The second group consisted in 50 undergraduate Psychology students. The criterion used for this sample was students' participation in Human Resources Management course. The course syllabus included the specific area of instruments and techniques used in professional performance evaluation, so the students were familiar with the construction methodology of such an instrument.

Procedure

The construction methodology of a Behavioural Observation Scale (BOS) includes several steps (Pitariu, 2006):

1. A group of specialists or experts (5-10) meet and they are asked to elaborate individually a set of factors or dimensions that allow to qualify a person as an efficient professional in a specific work area. Experts could use job analysis (especially job description) in order to identify these factors.

The redundant dimensions are excluded and the remaining ones are included in a single list. Then the experts are asked to elaborate a definition for each dimension.

This procedure is repeated with one or two groups of experts in order to obtain a list with the most relevant factors for that specific professional area. The criterion used to select the factors is the frequency they appear with in all experts' lists. A number of 8-15 dimensions is sufficient.

2. The list with all the retained factors and their definitions is given to another group of 20-30 experts who are asked to offer different examples of professional behaviour for each factor. Behavioural examples are gathered randomly in a single list, after the redundant or the common ones are eliminated.

3. The list of dimensions and the list of behavioural examples are offered to another group of experts. Their task is to reassign each behavioural example to a single dimension. This reassigning procedure has to complete two conditions: a behavioural example needs to be correctly reallocated by 2/3 (67%) of experts, in order to be retained; a dimension needs to have correctly reassigned 60% of the initial behaviour examples, in order to be kept.

4. The remaining dimensions and their behavioural examples are collected in a single brochure. The first page contains the instruction needed for the instrument completion. Each factor and the behavioural examples that operationalized it are edited on a different page.

The instrument is completed by several evaluators in order to pre-test it. In this step, statistical procedures are computed. Descriptive statistics, inter-item correlations, consistency analysis, inter-factors correlations are computed in order to asses psychometric qualities of the instrument.

Factors and behavioural examples that not complete the statistical requirements are eliminated. In this way, the elaboration of the final form of the instrument is concluded.

Results

Teachers' professional performance is conceptualised in this research as a multiple criterion, meaning that professional performance is a multidimensional concept, each dimension or factor having an equal importance to the global performance.

The intention was to create an instrument that could be used in professional performance evaluation for all university teachers, independently on their specific professional area or their academic position.

In order to create a Behavioural Observation Scale for university teachers' performance evaluation, the methodological approach included several steps:

First step

A group of 15 experts, working individually, were asked to identify and define the most important dimensions or factors that express university teacher's professional performance, from students' perspective. The experts were teachers specialized in different academic areas: Psychology, Economics, Letters and Humanities, Engineering. They were introduced to a brief explanation regarding the methodological aspects involved in creating an evaluation instrument for teachers' professional performance.

20 undergraduate Psychology students were also involved in this task, in order to collect a wide range of relevant factors.

The two groups were asked to offer a brief definition for each factor.

At the end of this sequence, 11 dimensions, the most frequently mentioned, were identified and defined:

1. **Knowledge in Specialized Area**: various, recent information in specific professional area and in connected fields.
2. **Cognitive Flexibility**: the ability to organize and structure information; ability to easily pass from abstract to concrete levels and reverse; analytical and synthetic ability.
3. **Communication Ability**: the ability to express a clear, coherent message; listening ability; compatibility of different forms of message (verbal, nonverbal, paraverbal); feedback rendering.
4. **Pedagogical Ability**: the ability to guide students in developing professional competencies; the ability to enhance students' motivation for study; an interactive teaching style; error correction in order to improve students' activity.
5. **Ethical Conduct**: conformity to deontological principles and organizational rules.
6. **Relationship with Students**: the ability to develop trustful relationships with students; helping and supporting behaviours.
7. **Research Potential**: initiatives on theoretical and practical level; stimulation of students' research activity.
8. **Emotional Maturity**: emotional equilibrium; stress tolerance; self-awareness; self-control.
9. **Professional Interest**: permanent attention on recent theoretical and practical developments in professional area; participation in scientific events.
10. **Didactical Technology**: efficient and adequate use of didactical methods and means; didactical objectives communication and pursuit.

11. **Group Control**: efficient management of students' group; correction of off-task students' behaviour.

The main hurdle encountered in this step consisted in the fact that experts were confronted with the difficulty to identify factors or dimensions with a relatively high level of generalization. The most common tendency was to identify behaviours, meaning concrete aspects of professional performance.

Second step

Another group of 20 experts (university teachers specialized in different areas) and 30 undergraduate Psychology students were asked to elaborate behaviours, in the form of sentences, which operationalized each of the dimensions identified in the first sequence.

Behavioural examples repeatedly offered by participants were eliminated. Any grammar or expression errors were corrected, so that the meaning of the item should not have been altered.

At the end of this step, 131 behavioural examples were gathered.

For exemplification, 5 behavioural examples will be mentioned, for each factor:

I. Knowledge in Specialized Area

1. He/she has updated scientific information in his/her professional area.
2. He/she uses many concepts from scientific specialized area.
3. He/she has updated scientific information in his/her professional area and in connected areas.
4. Teacher uses specialty language in a correct and adequate manner.
5. Teacher clearly explains all the aspects that he/she presents.

II. Cognitive Flexibility

1. In his/her explanations, teacher easily combines abstract and concrete aspects.
2. Teacher presents information in a logically structured manner.
3. Teacher easily makes analysis and synthesis of scientific information.
4. He/she offers examples in order to support abstract knowledge.
5. He/she is able to reformulate abstract information in order for the students to understand it.

III. Communication Ability

1. Teacher is able to adapt his/her language depending on students' comprehension ability.
2. He/she is coherent and clear in expression.
3. It is easy for the students to understand what teacher intends to transmit.
4. Teacher identifies students' understanding difficulty from their nonverbal language.
5. Teacher asks for clarifications when he/she does not understand something.

IV. Pedagogical Ability:

1. Teacher stimulates students' implication in class activity, by adopting an interactive teaching style.

2. Teacher makes obvious the relationship between current tasks or subjects and the competencies required by the profession.
3. Through the working manner adopted in class, teacher encourages students' critical, independent thinking.
4. Teacher offers feedback about students' activity in order to increase their performances.
5. Teacher approaches students' errors as learning opportunities.

V. Ethical Conduct

1. Teacher grades his/her students depending on his/her liking or disliking, in a subjective manner.
2. He/she answers the phone during classes.
3. Teacher behaves equitably to all the students, without discrimination.
4. Teacher is late for classes.
5. Teacher has a dignifying behaviour.

VI. Relationship with Students

1. Teacher establishes with the students relationships of mutual respect and consideration.
2. He/she is an easily approachable person.
3. Teacher offers support and understanding when students face any sort of problem.
4. He/she is liked by the students.
5. Teacher has a sense of humour.

VII. Research Potential

1. Teacher offers many ideas for research.
2. He/she involves students in research projects.
3. Teacher appreciates students for their ideas in research area.
4. He/she is involved in different research projects.
5. Teacher offers useful suggestions regarding students' research activities.

VIII. Emotional Maturity

1. Teacher looses his/her calm whenever interrupted.
2. Teacher expresses in an aggressive manner the things he/she is disturbed by.
3. He/she is a calm and even-tempered person.
4. He/she does not like to be contradicted.
5. Teacher does not accept points of view different from his/her own.

IX. Professional Interest

1. Teacher asks students for feedback on his/her professional activity in order to improve it.
2. He/she is always interested in recent knowledge in his/her field.
3. Teacher keeps students informed about scientific events.
4. Teacher participates in various scientific events.
5. Teacher recommends updated bibliography.

X. Didactical Technology

1. Teacher presents current objectives at the beginning of each class.
2. He/she efficiently organizes time so that current class objectives could be achieved.
3. At the beginning of the term, teacher informs the students about evaluation criteria.
4. Evaluation tests cover all the subject area.

5. Didactical content is different from the named objectives.

XI. Group Control

1. Teacher efficiently organizes students' group in didactical activity context.
2. Teacher encourages students' positive task behaviours.
3. He/she easily maintains order in class.
4. Teacher establishes and communicates rules for class activity.
5. He/she mediates conflicts between students.

Third step

The next task for another group of 15 experts (university teachers) was to reassign each behaviour to the appropriate dimension. They received two lists: one comprising the 11 dimensions and the other comprising the 131 behavioural examples, in a random succession.

In order to keep a behavioural example, 10 (67%) experts should have been reassigned it correctly and in order to keep a dimension, 60% of the initial behavioural examples should have been reallocated to it.

Two dimensions were eliminated in this step: "Research Potential" and "Professional Interest". In these two cases, the reallocation procedure did not satisfy the second condition. Two possible reasons could explain this situation. The two dimensions were too similar; therefore it was difficult to discriminate their behavioural examples. The other explanation consists in the fact that the items generated for "Professional Interest" dimension were ambiguous behavioural examples. The supporting evidence was that many experts systematically reallocated these items to another factor ("Knowledge in Specialized Area").

From all 131 behaviours examples, 67 were maintained. Almost half of them (64) were eliminated because they failed the requirement to have been correctly reallocated by 2/3 of the experts.

Fourth step

A first version of BOS was created in order to pre-test it. The first page presented the instructions. Every dimension, along with its items was edited on a different page. A five-point response scale was attached to each item, evaluating the frequency of behaviours. Each point had a verbal label: 1 – Never, 2 – Sometimes, 3 – Regularly, 4 – Often, 5 – Always.

67 undergraduate Psychology and International Relations students were asked to complete the instrument in order to evaluate a certain teacher. They were also asked to report any difficulty they would encounter in using the appraisal scale. The purpose was to identify unclear behavioural examples. Several students reported understanding difficulty on 7 items: one from "Cognitive Flexibility" dimension (In his/her explanations, teacher easily combines abstract and concrete aspects.), one from "Relationship with Students" dimension (Teacher's emotional implication in teacher-student relationship is minimal.), two from "Emotional Maturity" (Teacher is able to

identify his/her own limits. Teacher rapidly changes his/her emotional state.) and three items from "Didactical Technology" dimension (Teacher follow the course syllabus. Teacher uses modern didactical means such a video projector. Didactical methods teacher uses are adequate to proposed objectives.).

Related to these behavioural examples, students reported difficulty in item comprehension or in item evaluation, arguing that some teacher's behaviours are not so explicit or easily observed by them. Another comment was made regarding the item relevance to teacher's professional performance.

These 7 items were eliminated in this step. Another item was reformulated. "Teacher does not stand to be contradicted" became "He/she is disturbed when contradicted". In this case, the problem was the verbal negation, which raised the difficulty to answer to this item.

The intermediate version of BOS, comprising 9 dimensions and 60 items, was created in this sequence.

Fifth step

In order to asses the internal consistency for each dimension of the scale, 30 teachers had been evaluated by 293 students. Teachers were specialized in different areas (Psychology, History, Journalism, Philosophy, Economics, Law, Letters and Humanities) and had different academic titles. 5 (16.7%) of them were males and 25 (83.3%) were females. Their age ranged from 24 to 43 years (M=29.77, SD=4.18). Their experience in academic teaching varied from 1 to 12 years (M=5.4, SD=2.74).

293 students evaluated the teachers. 57 (19.5%) of them were males and 236 (80.5%) were females. Their age ranged from 18 to 53 (M=22.14, SD=4.88). 40 (13.7%) of them were in the first academic year, 99 (33.8%) were in the second, and 154 (52.6%) in the third academic year.

26 teachers had 10 students' evaluations, 3 teachers had been evaluated by 9 students, and one teacher had 6 evaluations.

For each teacher, a final score was computed for each item. This score represented the mean of all 10, 9 or 6 responses given by students to an item. Based on this data, Alpha Cronbach's coefficients were computed in order to evaluate the internal consistency for each dimension, as an indicator of scale reliability. The following coefficients were obtained:

- **0.8311** for I. Knowledge in Specialized Area;
- **0.8188** for II. Cognitive Flexibility;
- **0.9256** for III. Communication Ability;
- **0.9147** for IV. Pedagogical Ability;
- **0.8096** for V. Ethical Conduct;
- **0.8993** for VI. Relationships with Students;
- **0.8677** for VII. Emotional Maturity;
- **0.7098** for VIII. Didactical Technology;
- **0.5258** for IX. Group Control**.**

Inter-item Correlations and Corrected Item-Total Correlations showed some negative correlations. Items that negatively correlate to total score of a dimension have to be eliminated, because they do not bring useful information regarding the construct that is measured (Popa, 2008).

3 items, from three different dimensions were eliminated, because of their negative correlation to the dimension score. This way, three Alpha Cronbach's coefficients had increased:

- **0.8203** for Ethical Conduct;
- **0.9319** for Emotional Maturity;
- **0.8009** for Didactical Technology.

"Group control" dimension has been eliminated in this step. One of its items was negatively correlated to total score. If this item had been deleted this would not have increased Alpha Cronbach's value over 0.70, the cut-off value used in order to consider that a dimension has an acceptable level of internal consistency. Even if the elimination of another item had raised Alpha Cronbach's coefficient to an acceptable value, this would not have been a valid solution, since the number of remaining items would have been 4. This would results in the end in the elimination of that factor, taking into account that a number of at least 5 items is recommended in order to operationalize a construct (Constantin, 2004).

Correlations between the 8 remaining dimensions are presented in table 1. The intensity of these highly significant correlations shows that BOS dimensions are distinct factors of teachers' performance.

For the majority of items, the desirable end of the five-point response scale is 5 – Always. There are few items reverse-scored. These items are:

- 1, 2, 3, 5, 7 for Ethical Conduct;
- 1, 3, 5, 7 for Emotional Maturity;
- 4 for Didactical Technology.

On each factor of BOS, score is computed as a mean of all items scores. A total score can also be computed as a mean of the 8 dimensions scores. Dimensions have equal weights in total score.

The final version of BOS is presented in the Appendix.

4. Conclusions and limitations

The main limitation of this study consists in the small-sized sample used in order to assess psychometric qualities of BOS. A future research should increase the sample size.

All the experts who took part in this research are teachers from the same university. This way, the evaluation scale is most suited for this university. In other university, teacher's professional performance could be defined differently, according to

organisational specificity. Therefore, before using this BOS, a preliminary analysis is necessary.

This instrument can be used as a part of a complex evaluation system used in university teachers' performance appraisal. An important part of a teacher's job is didactical activity. Therefore, performance evaluation is not complete without students' perspective on their teachers' professional effectiveness.

Correlations

		I	II	III	IV	V	VI	VII	VIII
I	Pearson Correlation	1.000	.826**	.784**	.856**	.578**	.752**	.529**	.615**
	Sig. (2-tailed)	.	.000	.000	.000	.001	.000	.003	.000
	N	30	30	30	30	30	30	30	30
II	Pearson Correlation	.826**	1.000	.791**	.794**	.478**	.825**	.494**	.582**
	Sig. (2-tailed)	.000	.	.000	.000	.008	.000	.005	.001
	N	30	30	30	30	30	30	30	30
III	Pearson Correlation	.784**	.791**	1.000	.835**	.611**	.857**	.562**	.652**
	Sig. (2-tailed)	.000	.000	.	.000	.000	.000	.001	.000
	N	30	30	30	30	30	30	30	30
IV	Pearson Correlation	.856**	.794**	.835**	1.000	.686**	.851**	.535**	.663**
	Sig. (2-tailed)	.000	.000	.000	.	.000	.000	.002	.000
	N	30	30	30	30	30	30	30	30
V	Pearson Correlation	.578**	.478**	.611**	.686**	1.000	.691**	.849**	.378*
	Sig. (2-tailed)	.001	.008	.000	.000	.	.000	.000	.039
	N	30	30	30	30	30	30	30	30
VI	Pearson Correlation	.752**	.825**	.857**	.851**	.691**	1.000	.721**	.579**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.	.000	.001
	N	30	30	30	30	30	30	30	30
VII	Pearson Correlation	.529**	.494**	.562**	.535**	.849**	.721**	1.000	.215
	Sig. (2-tailed)	.003	.005	.001	.002	.000	.000	.	.255
	N	30	30	30	30	30	30	30	30
VIII	Pearson Correlation	.615**	.582**	.652**	.663**	.378*	.579**	.215	1.000
	Sig. (2-tailed)	.000	.001	.000	.000	.039	.001	.255	.
	N	30	30	30	30	30	30	30	30

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 1. Correlations between BOS dimensions

This scale could be used not only in decisions concerning human resources in a university, but also as an opportunity for feedback on teacher's performance, useful in improving some aspects of didactical activity.

This study investigates only one aspect of university teachers' performance. It could be the starting point in creating a complex and complete evaluation system, including other performance criteria for teachers, such as research activity or efficiency in administrative tasks.

References

1. Aleamoni, L.M. (1999) 'Student Rating Myths Versus Research Facts from 1924 to 1998', *Journal of Personnel Evaluation in Education,* Vol. 13, No. 2, Boston: Kluwer Academic Publishers, pp. 153-166.

2. Blanton, L., Sindelar, P., Correa, V. (2006) 'Models and Measures of Beginning Teacher Quality', *Journal of Special Education*, Vol. 40, No. 2, pp. 115-140.
3. Cascio, W.F (1991) Applied Psychology in Personnel Management, New Jersey: Prentice Hall Inc.
4. Constantin, T. (2004) *Evaluarea psihologica a personalului*, Iasi: Polirom.
5. De Jong, R., Westerhof, K.J. (2001) 'The Quality of Student Ratings of Teacher Behaviour', *Learning Environments Research,* Vol. 4, pp. 51–85.
6. Denner, P., Miller, T. Newsome, J., Birdsong, J. (2002) 'Generalizability and Validity of the Use of a Case Analisys Assesment to Make Visible the Quality of Teacher Candidates', *Journal of Personnel Evaluation in Education*, Vol. 16, No. 3, pp. 153-174.
7. Fossey, R. (1999) 'Personnel Evaluation in Higher Education: A Mounting Call for Accountability', *Journal of Personnel Evaluation in Education*, Vol. 13, No. 2, pp. 115-118.
8. Holtzapple, E. (2003) 'Criterion-Related Validity Evidence for a Standards-Based Teacher Evaluation System', *Journal of Personnel Evaluation in Education*, Vol. 17, No. 3, pp. 207–219.
9. Hong, W., Xuezhu, C., Ke, Z. (2007) 'On the relationship between research productivity and teaching effectiveness at research universities', *Frontiers of Education in China,* Vol. 2, No. 2, pp. 298-306.
10. Keig, L. (2000) 'Formative Peer Review of Teaching: Attitudes of Faculty at Liberal Arts Colleges toward Colleague Assessment', *Journal of Personnel Evaluation in Education*, Vol. 14, No. 1, pp. 67-87.
11. Laursen, P. F. (2005) 'The Authentic Teacher', *In* Beijaard, D., Meijer, P.C., Morine - Dershimer, G., Tillema, H. (ed.), *Teacher Professional Development in Changing Conditions* (pp. 199-212), Dordrecht: Springer.
12. Lee, S. W. (ed.) (2005) *Encyclopedia of School Psychology,* Thousand Oaks: Sage Publications.
13. Manolescu, A. (1998) *Managementul resurselor umane.* Bucuresti: RAI.
14. Peterson, K.D., Wahlquist, C., Bone, K. (2000) 'Student Surveys for School Teacher Evaluation', *Journal of Personnel Evaluation in Education,* Vol. 14, No. 2, pp. 135-153.
15. Pitariu, H.D. (2006) Proiectarea fiselor de post, evaluarea posturilor de munca si a personalului: un ghid practic pentru manageri. 2nd edition, Bucuresti: Irecson.
16. Popa, M. (2008) Statistica pentru psihologie. Teorie si aplicatii SPSS. Iasi: Polirom.
17. Scheeler, M.C. (2008) 'Generalizing Effective Teaching Skills: The Missing Link in Teacher Preparation', *Journal of Behavioral Education*, Vol. 17, No. 2 / June, pp. 145-159.
18. Simpson, P., Siguaw, J. (2000) 'Student Evaluations of Teaching: An Exploratory Study of the Faculty Response', *Journal of Marketing Education*, Vol. 22, No. 3, pp. 199-213.
19. Sutton, R. E. (2004) 'Emotional Regulation Goals and Strategies of Teachers', *Social Psychology of Education,* Vol. 7, pp. 379–398.
20. Wilkerson, D.J., Manatt, R.P., Rogers, M.A., Maughan, R. (2000) 'Validation of Student, Principal, and Self-Ratings in 360^0 Feedback for Teacher

Evaluation', *Journal of Personnel Evaluation in Education*, Vol. 14, No. 2, pp. 179-192.

Appendix

Behavioural Observation Scale for University Teachers' Performance Evaluation

Instructions:

On the following pages you will find a set of appraisal scales based on which you are asked to evaluate professional performance of a certain teacher.

Each evaluation scale refers to a certain type of behaviour. There are 8 dimensions or factors, each comprising several behavioural examples.

You have to decide for each item, the frequency of that behaviour in your teacher's professional activity. You can choose one of the following answers:

1 – Never;
2 – Sometimes;
3 – Regularly;
4 – Often;
5 – Always.

When you complete this questionnaire, please take into account and try to avoid some of the systematic appraisal errors:

Don't be too lenient or too severe in your evaluation! Use all the five points of the answering scale!

The 8 dimensions represent distinct aspects. When you answer, consider each dimension independently! Do not mark the same answer to all the aspects! Each person has his own qualities and imperfections.

Be aware of your liking or disliking regarding the teacher you evaluate and try to leave it aside!

Be as sincere and as less subjective as you can!

Thank you for your sincerity and for your willingness to collaborate!

Knowledge in Specialized Area

1. He/she has updated scientific information in his/her professional area.
2. He/she uses many concepts from scientific specialized area.
3. He/she has updated scientific information in his/her professional area and in connected areas.
4. Teacher uses specialty language in a correct and adequate manner.
5. Teacher presents specialty knowledge and necessary explanations.
6. He/she supports his/her opinions with scientific arguments.

7. Teacher offers a lot of information, along with relevant examples.

Cognitive Flexibility

1. Teacher presents information in a logically structured manner.
2. Teacher easily makes analysis and synthesis of scientific information.
3. He/she offers examples in order to support abstract knowledge.
4. Teacher presents scientific theories and explanations starting from examples.
5. Teacher follows certain logic (like chronological order of theories) in subject presentation.
6. He/she is able to reformulate abstract information in order for the students to understand.

Communication Ability

1. Teacher is able to adapt his/her language depending on students' comprehension ability.
2. Teacher's voice is adequate to his/her verbal message.
3. He/she is coherent and clear in expression.
4. He/she nonverbally or by intonation emphasizes different subject aspects.
5. It is easy for the students to understand what teacher intends to transmit.
6. Teacher identifies students' understanding difficulty from their nonverbal language.
7. Teacher easily expresses his/her ideas.

Pedagogical Ability

1. Teacher stimulates students' implication in class activity by adopting an interactive teaching style.
2. Teacher makes clear the relationship between current tasks or subjects and the competencies required by the profession.
3. Through the working manner adopted in class, teacher encourages students' critical, independent thinking.
4. Students' knowledge and abilities are developed in his/her class.
5. Teacher stimulates his/her students to learn more.
6. Teacher approaches students' errors as learning opportunity.

Ethical Conduct

1. Teacher grades his/her students depending on his/her liking or disliking, in a subjective manner.
2. He/she answers the phone during classes.
3. Teacher behaves equitably to all the students, without discrimination.
4. He/she misses his/her classes giving students no motivation for the absence.
5. Teacher has a dignifying behaviour.
6. Teacher is late for classes.
7. He/she respects his/her promises.

Relationship with Students

1. Teacher establishes relationships of mutual respect and consideration with the students.
2. Teacher offers support and understanding when students face any sort of problem.
3. Teacher shows empathy to his/her students.
4. He/she is an easily approachable person.
5. He/she is liked by the students.
6. Teacher treats all the students as equals.
7. He/she creates a positive climate, which favors intellectual activity.

Emotional Maturity

1. Teacher looses his/her calm whenever interrupted.
2. Teacher expresses in an aggressive manner the things he/she is disturbed by.
3. He/she is a calm and even-tempered person.
4. He/she is disturbed when contradicted.
5. Teacher supports his/her opinions with calm and determination.
6. Teacher changes his/her/ own opinions and attitudes from one day to the next.

Didactical Technology

1. Teacher informs students about current objectives at the beginning of each class.
2. He/she efficiently organizes time so that current class objectives could be achieved.
3. At the beginning of the term, teacher informs the students about evaluation criteria.
4. Evaluation tests cover all the subject area.
5. Teacher proposes for debate topics concordant with course syllabus.

Micro-Insurance and Entrepreneurial Innovations

Pavol Minarik and Josef Mladek

[1]*Department of Institutional Economics, Vysoká škola, ekonomická v Praze (UEP), nám. W. Churchilla 4, 130 67 Praha 3, {pavol.minarik, josef.mladek}@vse.cz*

This paper aims at providing an institutional analysis of the various options of micro-insurance provision. Micro-insurance is one segment of microfinance, or the sector that provides financial services to the poor. Next to the better known micro-credit, micro-insurance plays a significant role in the development efforts. Geography-wise, cases in particular from the significant market of India are discussed. Different institutional arrangements are analyzed: (1) institutions driven by local communities, (2) Private seeking of value by commercial corporations, (3) Communities distorted by external interventions and (4) Hampered (commercial) markets communities The aim of this paper to shed some new light of the implications of particular institutional design on the outcomes.

Key words: micro-insurance, entrepreneurship, innovation, development, trust, state intervention, subsidies, India.

1. Introduction

Microfinance is usually defined as the sector that provides financial services to the poor or those who are excluded for some reason. The "family" of micro-financial services is wide and it keeps growing. The core segment is the micro-*credit*, but there are new emerging segments such as micro-insurance, or micro-payments. Covering all the issues of this dynamic sector is hardly possible in a single paper. Therefore, we will focus our attention only towards certain institutional aspects of microfinance. From the high number of microfinance segments most consideration will be given to microinsurance. Geography-wise, cases in particular from India will be discussed.

The methodology of this paper draws on the approach of analytical narrative as introduced by a diverse group of social scientists: Robert H. Bates, Avner Greif, Margaret Levi and Jean-Laurent – more in Bates et al. 1998. This also allows the research to tap some soft „sociological data", as advocated for example in (North, 2005). The analysis is driven by looking into action of an individual. Functioning of markets is perceived to be a process of innovation and entrepreneurial discovery rather than one-stop shifts of equilibriums.

The contribution of microfinance towards development is popular in the topical literature. Many poor people are not able to purchase micro-finance services. This is

either a result of high fees charged or no services provided at all. Following from this, it is then argued that a commencement of provision of microfinance services to the previously marginalized people can improve risk management, help to spur investment, and thus start-off economic growth. That is why, innovative entrepreneurial solutions in providing these services in a new (cheaper) way – *given the institutional constraints* - open up new frontiers and widen the choice and scope of mutually beneficial transactions in the society.

Even more importantly, micro-finance may serve as an important tool for coping with the current financial and economic crisis. The global poor are generally more vulnerable to negative shocks than the rest of the society. Thus provision of risk management through micro-insurance may serve as an important factor to mitigate the negative implications of the current crisis.

2. Institutional Approach to Development

Before we start with our investigation of micro-insurance institutional arrangements, it will be beneficial to briefly articulate institutional approach towards development.

As William Easterly would point out, billions of dollars have been spent on what is called international development aid. Yet, it is the countries that should be supposed to flourish the most, because they have received most of the aid, that are still among the poorest today. If development aid provided by governments (or supranational, government-sponsored, bodies) has not been successful, are there any alternatives?

Economic progress can be both indented and unintended and it can be carried-out by both forced (or imposed) and voluntary coordination. Consequently we need to deal with the issues of (A) International (official) aid, (B) Economic policy reaching the set goals, (C) Private charity and (D) Consequences of self-interested human action, as the simple 2x2 matrix shows bellow:

	Progress unintended	Progress indented
Forced coordination	**(A)** International aid	**(B)** Economic policy reaching the set goals
Voluntary coordination	**(D)** Consequences of self-interested action	**(C)** Private charity

(A) International aid or intergovernmental deals?

Politicians are, as are other people, profit maximizers given the constraints they face (Brennan and Buchanan, 2000). What makes them special though is their non-market privilege to influence lives of others, a privilege they may lose at any time, when they lose their office. Politicians thus have ample incentives to grab as much as they possibly (safely) can. As public choice economists would put it, government officials and their potential allies (e.g. interest groups) have incentives to exert their power to gain concentrated benefits and disperse the costs on others (Olson, 1982). It is the government and those who make the ultimate decisions about official development aid. A critical cynic could bureaucracy point out that the official "aid" cannot be perceived

as any sincere aid given by people to other people, rather it is a bunch of propaganda to mask deals between government officials and their allies across different countries.

(B) Economic policy reaching the set goals?

Some bureaucrats and politicians, however, may genuinely be willing to help others. They may have non-selfish incentives. Furthermore, let's suppose that they are successful in keeping their power and offices against any not-so-well-intending competitors. Although such a scenario must be considered rather unlikely, it is necessary to stress that even having the "good politicians" in power is not sufficient condition for successful government-directed economic progress. Well-wishing politicians or bureaucrats forcing their good will on others would *face the information problem* (Hayek, 1945). Aid workers and the development agencies do not often bear (and know) *real* costs of their action and their payout is not (contrary to the private profit-seeking sector) driven by the success of the (developmental) undertakings. Aid programs do not sell products to people. There are no prices and profits *to guide* optimization of the programs, which is otherwise the easiest way, for consumers to automatically express their tastes and desires. Forced development *lacks* the *feedback mechanism* of consumer demand and *drive for efficiency* of comparing the revenues earned with the incurred costs. Consequently, official development aid faces the problem of calculation in socialist economy (Mises, 1920).

Furthermore, official development aid by imposing new rules and creating subsidized competition for private businesses *weakens already existing institutional structures directed at generating peace and prosperity.* For example, an unforeseen consequence of food aid in some countries was a decrease of the price of agricultural products and a subsequent decrease in domestic farming production. By creating perverse incentives, governmental aid may destroy wealth production and instead increase dependence and encourage rent seeking.

(C) Private charity

The market for private aid is flourishing and individuals seem to be voluntarily willing to give their own money to help others. For example after the Asian tsunami in 2004, private charity proved to be a thriving alternative to the official funding. At the same time, it should be noticed, that private charity initiatives may in principle suffer from the same mistakes committed by governmental planning of development. Private charity organizations are also islands of planning (Coase, 1937). Yet, when competing against other private providers of charity, private aid entrepreneurs have incentives to learn from errors and improve their „products". There is a market driven weeding-out mechanism, which keeps the least efficient players out and makes it difficult for the private charities to grow beyond the size justified by consumer demand. This is a substantial difference from any programs financed by taxation revenues gained by coercion. Private charity organizers *do not have to be angles* by conviction. To secure their position in the market they are forced not to waste resources and provide services their compassionate customers ask for. However, charity is not necessarily needed to secure economic development. When properly coordinated, *even self-interested* action can lead to productive results.

(D) Progress as an unintended consequence of self-interested action

Human self-interested action coordinated by price signals and property rights has tremendously improved living standards of most people on the planet during the past two centuries. Removing obstacles to (cross-border) cooperation and competition has lifted millions of people out of poverty and improved living standards of others. Yet, there are still many parts of the world isolated from the benefits of trade and division of labor. What are the mechanisms that can direct self-interested human action towards more economic growth and progress and towards (de-facto) deregulations?

Even "bad" governments may be willing to conduct reforms in their self-interest

Fiscal revenues are often an important factor of wealth enjoyed by the ruling elite. Therefore even "bad" governments can have incentives to pursue pro-market reforms in order *to increase taxable economic activity* and thus their own well-being. In contrast, official aid provides additional resources to the recipient governments and therefore discourages them from pro-market reforms.

Pro-market reforms are generally a step in the right direction. But still, these are *government*-lead initiatives or *plans* of economic (pro-market) reform. The critiques of globalization fail to understand that the reform results are thus still far from "pure" capitalism. Poor peasants, whose property rights are not respected because of governmental plans of development, may in reality suffer, because of such governmental programs of planned „capitalism".

Entrepreneurs create value by overcoming obstacles caused by governmental interventions

Entrepreneurs obviously lack the power to *de iure* abolish regulations. Yet, through their profit seeking activities they may, without necessarily having an intention to do so, cause the regulations to have less significant impact on the economy and the society.

For instance, due to restrictive labor (and other) regulations, some people were artificially made relatively poor, which means that their transfer wages are low. Finding ways how to better employ the (relatively) poor (in a more profitable way) is thus one of the greatest and potentially very profitable opportunities for entrepreneurs. The growth of IT-enabled services in India is a prime example of such a private response to the labor market restrictions. Using innovative technologies and changing their business models, the entrepreneurs were able to access a huge pool of (previously untapped) Indian high-skilled professionals.

This is, as it will be mentioned later, also the case of the microfinance sector. New services can be provided even if improper institutions are still in place. Entrepreneurs managed to partially overcome the burden of high cost of regulation by coming-up with innovative, less costly, business models.

Final thoughts on the framework

When governmental coercion is involved, (A) is the worst-case scenario, but (B) is the most optimal attainable option. Bad results of good intentions of (A) and (B) can make people refuse giving their money voluntarily (C). As a result (C) may be currently less

significant, compared to what it would have been if forced coordination did not exist. Furthermore, in a free-market system, (D) is likely to out-compete or merge with (C).

3. Matrix of Institutional Options in Microfinance

In this paper we argue that self-interest can motivate both communities and commercial corporations to build-up institutions and mechanisms to efficiently provide microfinance services. Explicit monetary profits, perceived value of recognition by other members of the community, or value delivered to compassionate donors of a voluntarily funded (that is not funded by coercion, such as taxation) NGO may all coordinate human action in a productive direction.

However, these spontaneously grown-up institutions can be distorted by interventions directing self-interest in an unproductive direction. These can be a result of a direct impact of (often well indented) *governmental policies* or more indirectly a consequence of *governmental funding* provided to "private" NGOs channeling to them taxation revenues, obviously primarily un-earned through voluntary transactions.

It is also useful to draw a distinction between community driven and commercial (or profit driven) arrangements in microfinance. This is not to say that there are no other options. Surely, governmental and NGO operated non-profit MFIs (microfinance institutions) are also active in the sector. But for our analysis of self-interest-driven institutional arrangements, we consider these options (the communal and commercial) of spontaneous enterprising in the microfinance as most fertile. It should also be noted that the two options of institutional arrangements ((1) and (2)) are not necessarily in any kind of contraposition against each other. As we will show later, when guided by proper incentives, they often cooperate together in their self-interest.

Based on this line of reasoning, we can come-up with a following matrix:

	Community institutions	**Commercial** institutions
Genuine self-interest	1. Community driven institutions	2. Private value seeking
Distorted self-interest	3. Distorted communities	4. Hampered (com.) markets

3.1 Community driven institutions

Mechanisms of sharing risks were well in place before the modern commercial corporations started offering its insurance policies or bureaucrats began crafting governmental programs. For example: „[r]ural India has had a tradition of households that share income, which reduced the impact that the death of a family member would have on the household's economic condition. Religious and social institutions also have a history of providing some benefits to the very poor and marginalized" (Athreye and Roth, 2005, p. 16). Such community-based models allowed a "[g]roup of people get together and essentially develop their own insurance scheme in which they pool their own funds, develop their own rules, and run their own scheme." (Athreye and Roth, 2005, p. 5). Obviously the people in their own interest strived to come up with

such institutional solutions that would fit best their local societal (and other) conditions.[1]

Prime example of such micro-insurance enterprises can be found in so called burial societies as those in southern India. For a small weekly payment they provide coverage of costs associated with a death of a family member (Morduch, 2004). It is important to underline the ability of such societies to *economize on operating costs and costs of signaling trust*.[2] In fact these institutions arose in circumstances, under which the *marginal* costs of signaling trustworthiness and collecting funds were very low. As (Morduch, 2004) points out: "Burial societies can operate this way because they are based out of local institutions where people *already gather* weekly; the societies can thus collect small payments as *part of other activities*."[3]

With local communities sharing risks in-between themselves, the issue of insufficient risk diversification may well arise. If this was to become really a serious issue and members of the community had a choice over a provider of the service, than the competing providers of such schemes would have an interest to arrange for a solution to decrease the risk-volatility that would give them an advantage in the competition process (e.g. pooling of risks with providers servicing neighboring or more distant communities).

3.2 Private value seeking

Recently, profit driven commercial corporations have shown that they are capable of providing value not only to their shareholders, but also (indirectly) to their clients as they compete for revenues. For example the Ugandan operations of AIG provide insurance covering 1.6 million lives in the East African country (Mc Cord et al., 2005). Moreover AIG has managed to generate 20 % profits on the premiums collected. AIG is perhaps *the* most striking example, but if profits can be made, others are likely to follow and enter the market.

Later on in part 5, we deal with the issues of responsiveness to consumer demand, incentives to innovate and other issues, that are *common* to commercial corporations (2) and community-based voluntary schemes (1). The question of competition and

---

[1] High switching costs of transferring to another provider of micro-finance services may limit the policing power of competition. However, switching costs should be understood as *a result of a dynamic process of entrepreneurial discovery*. People that are at a particular time locked-into a specific arrangement and suffer from opportunistic behavior of their current service provider represent *a profit opportunity* for entrepreneurs that can un-lock these arrangements and provide more desirable solution.

[2] For obvious reasons the issue of moral hazard was not relevant in the case of burial societies. Yet, the situation is different in case of other insurance products. For how voluntary societies coped with this issue in Continental Europe and United States of America, see Chalupníček and Dvořák (2007).

[3] Italics added.

interaction between the community-driven (1) and commercial arrangements of insurance (2) is dealt with in part 6.

3.3. Distorted Communities

Unfortunately the institutions based on bottom-up ideas and on social capital (as described in (1)) are often distorted by both indented and unintended consequences of governmental interventions.

Weak institutions have crippled many developing economies and have also hurt (micro)-insurance. Incomplete property rights, slow contract enforcement and compulsory licensing tend to make it difficult for non-commercial projects *to grow beyond their informality* and to rely more on external institutions outside their "native" communities.

Gradual process of building-up trust versus stop-and-go policies

Subsidies provided by governments may distort traditional mechanisms of mutual trust and risk transfer. The process of building-up trust is arguably a long-term one. But as Churchill and Garand, 2006, p. 580) point out: „A change in government policy may result in an abrupt cancellation of the subsidy." Consequently, that renders any previous investment in trust as useless. Specific examples are provided in (Latortue, 2006, p. 484):

> The Government of Karnataka[4] partially subsidized the premiums of Yeshasvini Trust. However, when the subsidies stopped, the premiums had to be doubled from Rs. 60 (US$1.35) to Rs. 120 (US$2.70) per adult, and the number of policyholders plummeted from 2.2 million to 1.45 million.
>
> Initially, UNDP fully subsidized the premium for members [of Karuna Trust] below the poverty line and other disadvantaged groups. Many members were not aware they were insured. When the premium subsidy was removed two years later, about 70 per cent of the members wanted to drop out – they did not want to start paying for services that had previously been free and felt that the premium was too high.

Subsidies for NGOs adversely affect communities

Governmental subsidies are not a result of the NGOs succeeding in competition for funds from voluntary donors, but instead they are a consequence[5] of taxation revenues. Therefore the subsidies-financed NGOs may be missing a significant mechanism

[4] Karnataka is one of the federative states of the Union of India.

[5] Sometimes this consequence can be of a seemingly indirect nature – for example when the subsidy is provided by a supranational organization that itself is funded by several governments.

policing the quality of services they provide to the receiving communities.[6] Thus, some NGOs may continue adversely impacting the communities, even if it is not justified by demand of the voluntary donors. IMF (2005, p.12) gives voice to the critical opinions on subsidizing micro-insurance and argues that subsidies that are not successfully phased out may have dire consequences:

> „[R]elaxation of budgetary constraints on MFIs may hinder their performance and create unwarranted dependence on external funding, especially if MFIs lack adequate transparency and accountability. ….
> [A]s nonprofit organizations, MFIs may not have appropriate ownership structure, staff, operational systems, and incentives to improve the efficiency of their operations."

Fernando (2005, p. 226) investigates the provision of micro-*credit* directed at empowerment of women. In the conclusion he finds that the program did not meet its objectives:

> "[W]omen borrowers have become even more rigidly imprisoned within the power structures that oppress them."

2.4 Hampered commercial markets

The commercial sector has been hit by implications of bad external institutions even harder than (informal) community-based institutions, because it finds it more difficult to hide under the veil of informality. This can be illustrated again on the Indian experience. In 1950s then Prime Minister Nehru stated his policy by declaring: "we require life insurance to spread rapidly all over the country and to bring a measure of security to our people."[7] However according to Athreye and Roth (2005): "Despite Nehru's desires, in the decades following nationalization, insurance products were designed primarily for those with regular income streams, i.e., those in formal employment." In a more general statement, IMF (2005, p.22) also notes that inhospitable institutional environment can make it more difficult for entrepreneurs to provide financial services to the poor:

> "Frictions in financial markets may prevent efficient transactions from taking place which may affect the poor disproportionately. Such market failures are likely to be more

[6] This does *not* mean a *general* objection towards subsidies as a tool of provision of cheaper or better-quality services to the clients of subsidized services. We merely wish to point out to the fact that subsidies may alter the market environment. Drawing on the "Cantillon effect", subsidies are like newly created money. That means, it is crucial, who gets them first. Subsidies can be channeled to (micro-finance) institutions or to their clients, who then may pick their favorite provider of the service (or satisfy a completely different, more acute, desire - in case the subsidies are not tied). So, if subsidies are granted to the organizations, these may stay in business even if their survival is not justified by their ability to compete and satisfy demand for charity.

[7] Quoted in Roth and Athreye (2005).

> prevalent in developing countries due to various institutional weaknesses.“

The IMF paper (*ibid*., p. 12, 22) also notes the affect of governmental subsidies *crowding-out* commercial activities:

> „[S]ubsidization of MFIs could create a distortion, resulting in unfair competition with traditional financial institutions, and curtailing the development of a more sophisticated and commercially oriented financial sector. In other words, financial sector deepening may be actually hindered by improper subsidization of microfinance activities, that is, in areas that can be potentially covered by commercially-oriented institutions. ….
>
> Subsidization of MFIs may also relax their budgetary constraints and create unfair competition with traditional financial institutions, preventing them from entering the microfinance niche.“

The nationalization of the insurance sector in India made suffer the majority of consumers without a job in formal sector and resulted in what is still today a market with low saturation. To respond to the failure of its own policies, the government imposed a new set of policies, further distorting the market. As a result, insurers are required to sell a certain share of their policies in rural areas, i.e. to the poor segment of Indian population. If the companies fail to meet these requirements, they may loose their license. Unfortunately, the city poor may consequently suffer from the attention of insurers being *artificially* driven away from them into the rural areas.

But that is not the only issue resulting from this regulation. The regulation *transforms* the *opportunities* associated with the rural markets *into a burden*. As Athreye and Roth (2005) claim:

> „There have been unverified reports that some insurers are dumping poorly serviced products on clients solely to meet their targets. As soon as they have met their targets, some have immediately stopped selling microinsurance. This practice is difficult to regulate, as it is harder to police the quality of insurance sold and serviced to the poor than its quantity. It would certainly be socially unfortunate if the regulation resulted in a mass of poorly serviced products sold at a loss, to enable insurers to concentrate on their more profitable products. This situation would not result in meaningful sustainable financial deepening, but more akin to charity, forced on insurers to allow them to do business in India.”

2.5 The need for genuine self-interest

Are commercial projects the only institutional approach for success in microinsurance? As it has been shown in the first part, they are not. In this part, we argue that *whatever*

the organizational arrangements (e.g. community based *or* corporations), the microfinance segment needs to be centered on genuine self-interest in order to flourish. Self-interest restricted by well defined property rights encourages: (I) Competing commercial corporations seeking shareholder-value to provide desired services to the consumers; (II) Privately funded NGOs competing for the funds provided by voluntary donors to provide maximum "charitable" value for the donation received and thereby strive to service the communities in need (III) Competing communities constrained by reasonably low switching costs to provide services desired by their members.

In all these situations the providers tend to minimize costs and increase revenues. An unintended consequence of their selfish action is greater responsiveness to consumer demand and a genuine interest in employing innovative technologies and business models. In all of this, the providers are constrained by their (potential) competitors who may offer a more desirable service. So, in a situation of reasonably low switching costs, the insurance providers are unlikely to succeed in increasing their profit margins at the expense of their consumers.

Responsiveness to diverse and varying consumer demand

The effort to earn revenues drives competing insurers to provide value to the consumers. This is particularly beneficial, as demand for specific insurance products differs across particular sub-segments. As Cohen and Sebstad (2006, p.43) note: "To better accommodate these diverse and changing needs, competition and rivalry over satisfying consumer demand can be nothing but helpful." To give an example, in its self-interest, the commercial Tata-AIG group considered the specificity and particularities of demand for micro-insurance. As Athreye and Roth (2005, p. 23) put it: „Before designing the social products, the preferences of the target market were determined".

Avoiding the shift of the cost burden on the consumers

Shifting costs upon the consumer can make the product in a result far less advantageous. Churchill (2006, p. 18) points out that "The indirect costs of accessing and using that product, including transportation and the opportunity costs of lost wages**,** may be much higher than the actual cost." More generally, Buczkowski, Mc Cord and Saksena (2006, p. 214) suggest that "*A balance must be maintained* between the efficiency of the insurer, and the cash flow and transaction costs of the policyholder. Without an acceptable balance between the two, micro-insurance may not succeed."

However, there is no better way to police this issue, than the opportunity of consumers to choose their provider of insurance. If the consumers can choose between providers of micro-insurance, they are likely to pick the one who offers them a better deal including *all* the (at that time known) costs.

Self-interested cost-cutting as an incentive to implement innovative technologies, improve bussines model, limit operating costs…

Morduch (2004) reminds that contracts in micro-insurance are „generally for small amounts and damages have to be assessed by insurers on an individual basis; scale economies are thus limited". These circumstances further increase the pressure on

micro-insurance institutions to cut costs. Garand, Roth, Rutherford (2006) notice that "additional innovation is required to provide better long-term products for the lowincome market". "Experimenting with technologies (including ATMs with biometrics, smartcards, palm pilots and point-of-sale devices)" is one of the options according to Churchill, (2006, p. 17). „Besides branchless banking, electronic bank account deductions can be used if *one* member of the household has a formal sector job and bank account" Roth, Garand and Rutherford (2006, p.106). More orthodox methods may include saving on labor costs, premium collecting costs or receipt administration costs. "Micro-insurance is labor intensive, therefore simplification of operations can make it easier to sell and service products by less expensive workers" (Churchill, 2006, p. 18). "If convenient, premiums can be collected less frequently" (Morduch, 2004). "In order to cut administration costs, policyholders do not receive a statement or a receipt[8] from Tata-AIG in India, instead agents deal with this information" (Athreye and Roth, 2005, p. 24).

As it is evident, the options to cut costs are many. However, to persist in cost cutting on the various margins, self-interest to generate profits (or other form of value) and competition between microfinance providers is necessary. Self-interest results into cost cutting, which forces competing micro-finance providers to stick with generating value. Consequently, self-interest makes it difficult for microfinance providers to divert from their goal of generating value by incurring costs that are unjustifiably higher than those incurred by their competitors.

Yet, still it could be perhaps argued that, while cutting costs in order to increase its profits, self-interested providers of (micro)-insurance may also try to shift costs on their clients or lower the quality of the provided services. Obviously the insurance providers may try to explore these options, yet they would be constrained by their competitors who would take-away their costumers.

The value of a brand in a trust-related industry

In general, trust is important in insurance, as it is an industry with deferred and conditional contract performance. But trust is even more significant in the context of *micro*-insurance that serves poor in developing nations, where official contract enforcement is generally weak. Moreover, the marginalization of micro-insurance clients is another factor in favor of fostering trust outside the official legal and court system. As Churchill, Cohen (2006, p. 180) put it: „because the poor are not a powerful or influential customer base, so insurers do not have significant incentives to keep them happy."

Signaling trust

For the micro-insurance to serve its purpose, a solution for trust signaling needs to be discovered. Churchill and Cohen (2006, p. 180) suggests: „For microinsurance, perhaps the most effective way of conveying this message is through branding –

[8] Still, such a measure is policed by the willingness of Tata-AIG to avoid costs resulting of miss-informed policyholders or intruders among its network of distributors of its policies.

associating the insurer with something that is trusted by the poor." Churchill and Cohen (2006, p. 181) point out the reliance of AIG on the recognized Indian brand of Tata: „When AIG entered the Indian market, it was fortunate to start a joint venture with the Tata group of companies, one of the most respected and trusted Indian industrial conglomerates. When Tata-AIG entered the low-income market, it exploited the Tata brand: agents selling micro-insurance assured potential clients that such a large company would have little interest in stealing their miniscule (in relative terms) premiums."

The capital value of a brand as an incentive for long-term investment

Brand facilitates signaling of trust, but it also provides corporations with long-term incentives to invest in the markets (i.e. in the poor) and than harvest the yield over long time. Consequently, commercial companies can discover value not only in present everyday operations, but also in the future distant cash flows. This may explain, why commercial corporation such as TATA-AIG was the "only example of creating general awareness that emerged from the case studies…, which produced brochures explaining insurance *without*[9] actually mentioning the insurer or its product. The literature was disseminated by its NGO partners, and their credibility in the low-income market helped raised the standing of insurance as a viable intervention for the poor." Churchill and Cohen (2006, p. 182).

Balancing the economies of size and local knowledge

Different factors influence the cost function of particular insurers in a specific way. So, there is a rationale for the size of micro-insurance businesses to be spontaneously (re-)discovered, as the conditions in the market develop..

Volume of operations usually affects costs and revenues. Considering the economies of scale may suggest that the higher the volume the lower the unit costs and the higher the profit margin. However, smaller-scale operations and reliance on local knowledge of communities may help to improve risk management and subsequently decrease costs. So, there seems to be a trade-off between savings from the big scale and loosing the value of specific knowledge[10] associated with small, localized operations. Price competition and varying costs are likely to guide the competing organizations in estimating the approximate optimal size and structure of their operations. In the following paragraphs we discuss arguments for the prevailing economies of scale in the micro-insurance business. For a discussion on how commercial companies can combine big scale operations with reliance on local knowledge and enforcement facilitated by local communities, please see part 6.

Economies of scale of risk management

For a risk transfer to be efficient and for a risk estimate to be reasonably accurate, the insurance industry relies on the Law of Large Numbers. Bigger scale of insurance operations consequently leads to more accurate risk assesments and lower insurance premia paid by the poor. As Churchill (2006, p. 17) notices, „When projections can be

[9] Italics added.

[10] Such as risk associated with particular clients.

estimated with a high degree of confidence, then the product pricing does not have to include a large margin for error, making it more affordable for the poor."

Economies of scale of information processing

The accurate risk pricing is (I) very technical, (II) requires assistance from an actuary, (III) data must be carefully managed, (IV) modeling techniques can be used to price products more accurately (Garand and Wipf, 2006a, p. 253). This suggests that there are further significant economies of scale from employing appropriate information system and reliance on knowledge and expertise. Similarly, Botero et al. (2006, p. 591) also signify the role of information processing: "Insurance is an information-processing business. The raw materials are customer data, product information, transaction details, investment records and so on. Even before the birth of the computer, large insurance companies drove the development of sorting, tabulating and calculating machines to improve efficiency."

Subsidies may discourage sticking near the optimal size

However, subsidies may favor particular NGOs, thus diverging the size of the subsidized insurer from the size justified by its cost function and its ability to satisfy consumer demand. IMF (2005, p. 12) discusses the effects of subsidies and notes that:

> [P]rovision of financial services to the very poor is a costly activity relative to the volume of resources channeled. Thus, a disproportionate part of the resources could be dissipated in administrative and operative expenses.

In other words, subsidies may allow the insurer not to make use of the accessible economies of scale by altering the size of its operations.

2.6. Global corporations vs. communities - competition or cooperation?

Commercial enterprises seeking to improve their revenues and lower costs have ample interest to cooperate with organizations such as cooperatives or informal communal groupings. These incentives for commercial corporations result from opportunities in *decreasing distribution costs* or *better risk managemen*t stemming from *local knowledge* and *informal enforcement of rules*. The partnering institutions may then benefit from the ability of big commercial corporations to *diversify risks* and provide *expertise at relatively low unit costs*.

Entrepreneurial discovering in the institutional environment

Distribution and premium collecting costs are high relatively to premiums charged, if we consider the low amounts of micro-insurance payments. Churchill (2006, p. 19) argues that to lower the costs one of the strategies "[I]s to collaborate with another organization that already has financial transactions with low-income households so the insurer can leverage existing infrastructure to reach the poor". "Group insurance is more affordable than individual coverage, but how does one find groups of people in the informal economy?" asks Churchill (2006, p. 23). Churchill, Liber and Wipf (2006, p.153) notice that "It is relatively easy if the targeted population is a well-organized

group that can accommodate group insurance arrangements, but is quite challenging if it is not because of the higher delivery and claims costs." Churchill (2006, p. 23) points out the various levels of organization in society: "Even though the informal economy is sometimes known as the disorganized sector, there are groupings out there that could be used, such as women's associations, informal savings groups, cooperatives, small business associations and the like."

The lesson learned should be that micro-insurance entrepreneurs can discover value in informal institutions in the society in order to decrease their operational costs. They can *foster (informal) institutional fabric*, where it (partially) already exists or they can try to create it on the margins where such process is associated with relatively low costs. Now, let us have a look at specific examples of AIG operations of AIG-Uganda and in India (the TATA-AIG joint venture).

In Uganda, AIG cooperates with local microfinance institutions. Self-interest in pushing-up its own bottom line has led the AIG to search for appropriate distribution channels. Botero, Mc Cord and Mc Cord, (2005, p. vi.) point-out that: "Almost all of the major MFIs in Uganda offer the AIG product." Ibid. (p. 44): "AIG Uganda is one of the first large insurance companies to extend coverage to the lowincome market by partnering with numerous microfinance institutions."

Such voluntary partnerships tend to provide value for both parties by economizing on the costs of expertise and information systems. Armendáriz and Morduch (2007, chapt.6, p. 168) view the cooperation from the reverse perspective of the cooperating MFI (called FINCA): „A major part of the success for FINCA Uganda stems from the partnership with AIG. The partnership spares FINCA staff from having to deal with the technical side of insurance provision (calculating actuarial tables, calculating appropriate reserves), avoids extra regulation, and ensures that risks are diversified. As a large insurer, AIG has means to spread risks across its many policies and can reinsure with ease. …. Were FINCA to go it alone, it would also have to find a way to protect itself in the event of larger-than-expected obligations."

In India, AIG at first wanted to cooperate with local MFIs in a similar way such as in Uganda. However, high costs resulting from such ventures made it soon clear that this was not a profitable option in the highly regulated Indian environment. Why? Insurers are required to sell a certain percentage of their policies in the rural areas. Consequently the demand for cooperating MFIs was too strong and there were too few reliable MFIs to work with (according to Athreye and Roth 2005, p. 21).

In its response to this *signal of high costs*, TATA-AIG developed an *innovation* - its micro-agent model. Consequently, groups of several local insurance agents were being established. These foster the insurance processes[11] and the process of building up trust. Criteria for the selection of micro-agents at Tata-AIG include various measures to pre-select candidates that *in their self-interest will behave honestly* and generate most value for the company. These measures seek to minimize the *adverse selection* problem and include[12]: (I) residency in the community in which she will sell and service policies,

[11] That is selling the policies, collecting premia payments, dealing with insurance claims etc.

[12] For a complete list of criteria see Churchill and Cohen (2006, p. 175).

(II) being married (III) good track record of integrity (i.e. trust of others in the community). As it is obvious, a successful candidate needs to have strong ties with the local community; so any *deceitful action* would be *too costly* for her *to pay-off*. Moreover, insurance agents are discouraged from cheating because of their involvement in self-help-groups (SHG). These rural informal organizations foster mutual-support of their members. High costs of breaking up with SHG make consequently dishonest action less likely. As Athreye and Roth (2005, p. 26) also point out: „The agent is also usually a member of an SHG, which provides for some monitoring of agent integrity."

The *monitoring of trustworthiness* is multi-layered. The agents in the group tend to check each other. Moreover they are checked by trusted NGOs, which can use its local knowledge and recommend candidates for new agents for a fee. Churchill and Leftley, (2006, p. 275) notice that, "Tata-AIG also involves non-governmental organizations (NGOs) in the process of identifying micro-agent candidates since locally they are often in a good position to assess the extent to which individuals are respected in their communities." Cooperation with credible NGOs also generates trust into the commercial company itself. TATA-AIG has "collaborated with local NGOs that helped strengthen its local credibility." Churchill and Cohen (2006, p. 181).

The unintended consequence of cost cutting is provision of better-paid jobs to the rural poor: (Athreye and Roth 2005, p. 21) "The CRIG model creates an insurance distribution infrastructure in low-income neighborhoods; it creates a new profession, the micro-agent, with new livelihood opportunities for poor women."

3. Conclusions

We have discussed both the community-based and commercial options of micro-insurance provision. We have noticed the benefits of entrepreneurs entering the micro-finance markets or communities discovering novel solutions and arrangements. At the same time, the salvations for *every* microfinance project may not be fully deserved, if microfinance projects are not driven by bottom up solutions from communities or competing entrepreneurs striving to satisfy real demand. If microfinance is directed by governmental „grand plans" or NGOs financed from tax revenues (and therefore not necessarily always fully responsive to the competition for donors' funds), then such development projects may not result into an improvement.

Proper institutional framework is crucial. Weakened self-interest tends to result into distorted communities and hampered markets. We have also discussed the benefits of self-interest driven by proper rules for responsiveness to demand, cost cutting, balancing the size and using the value of brand for signalling trust and long-term commitment. Finally, we have shown that self-interested commercial corporations tend to cooperate with local MFIs, NGOs and self-help groups.

By exploring various institutions arrangements, we can explain what may work and what can not:

(1) Well-indented development policies may have some unintended consequences. For example the stop-and-go nature of some subsidies programs may well distort the gradual process of building up trust in communities.

(2) To satisfy diverse and varying consumer demand, foster innovation and limit shifting costs on the consumer, competing self-interested entrepreneurs (shareholders) limited by well-defined property rights are needed. Numerous issues are consequently discussed including the implications of a corporate brand in this trust-related industry which are significant.

(3) Finally, it is shown how global corporations have the self-interested & profit-maximizing incentives to cooperate with local communities in order to better manage their own risks and cut costs. Some corporations tend to rely not only on the official institutional framework and its enforcement, but they have been also looking into innovative ways how to tap the informal institutional framework - for example in order to better monitor trustworthiness.

There are two general limitations of the research. (1) Some soft data is used. (2) The sources of data/inputs for analysis are rather secondary, than instead being based on research and evidence gathering conducted on the spot. This however, we would argue, is *off-set* by logical deductive reasoning of the analytic narrative and structured framework used to approach the analysis. Also as far as the use of "soft data" is concerned, it has been argued (for example in North, 2008) in favor of its usage in order to provide the ability of more encompassing analysis of complex issues in the society.

At the very end of the paper, after a discussion of the issues and constraints of micro-insurance provision, the alternatives of insurance should be pointed-out. Certainly, micro-insurance is not the only way for the poor to deal with risks. Morduch (2004) argues that having "a cushion of savings to fall back on" can be a solution for risks that are difficult to insure. These include the frequent and serious risks of bad weather and agricultural risks. Profit driven, competing, micro-insurance can without doubts generate value for millions of people. Yet, the very poor might also gain from an institutional change that sidelines bad institutions in favour of productive institutions, which allow for more trade and deeper division of labour and therefore lead the very poor away from the risk of falling down into absolute misery and poverty.

Acknowledgments

This paper has been supported by an internal grant of University of Economics in Prague, IGA 33/08.

References:

1. Armendáriz, B, and Morduch, J. (2007): *The Economics of Microfinance*. MIT Press.

2. Athreye, V. and Roth J. (2005): ‘’TATA-AIG Life Insurance Company Ltd. India’, *CGAP Working Group on Microinsurance, Good and Bad Practices*, Case Study No. 14., September 2005.
3. Bates, R. et al (1998): *Analytic narratives*. Princeton University Press.
4. Botero, F. et. al. (2006): ‘The future of microinsurance’, *In*: Churchill, C. (ed.), *Protecting the poor*: *A microinsurance compendium*, (pp. 583–601), International Labour Organization.
5. Botero, F., Mc Cord, J. and, Mc Cord, M. (2005): ‘AIG Uganda – A Member of the American International Group of Companies’, *CGAP Working Group on Microinsurance, Good and Bad Practices*, Case Study No. 9. April 2005.
6. Brennan, G. and Buchanan, J. (2000): *The Power to Tax: Analytical Foundations of a Fiscal Constitution*. Indianapolis, IN: Liberty Fund, Inc., 2000. Obtained through the Internet: http://www.econlib.org/library/Buchanan/ buchCv9c7.html, [accessed 29 /7/2008].
7. Buczkowski, G., Mc Cord, M. and Saksena, P. (2006): ‘Premium collection: Minimizing transaction costs and maximizing customer service’, *In*: Churchill, C. (ed.), *Protecting the poor*: *A microinsurance compendium*, (pp. 197–215), International Labour Organization.
8. Chalupníček, P. and Dvořák, L. (2009): ‘Health Insurance before the Welfare State: The Destruction of Self-Help by State Intervention’: *The Independent Review*, v. 13, no. 3, Winter 2009, pp. 367–387.
9. Churchill C., Liber, D. and Wipf, J. (2006): ‘Product design and insurance risk management’, *In*: Churchill, C. (ed.), *Protecting the poor*: *A microinsurance compendium*, (pp. 146–173), International Labour Organization.
10. Churchill, C. (2006): ‘What is insurance for the poor?’, *In*: Churchill, C. (ed.), *Protecting the poor*: *A microinsurance compendium*, (pp. 12–24), International Labour Organization.
11. Churchill, C. and Cohen, M. (2006): ‘Marketing microinsurance’, *In*: Churchill, C. (ed.), *Protecting the poor*: *A microinsurance compendium*, (pp. 174–196), International Labour Organization.
12. Churchill, C. and Garand, D. (2006): ‘Strategies for sustainability’, *In*: Churchill, C. (ed.), *Protecting the poor*: *A microinsurance compendium*, (pp. 564–582), International Labour Organization.
13. Churchill, C. and Leftley, R. (2006): ‘Organization development in microinsurance’, *In*: Churchill, C. (ed.), *Protecting the poor*: *A microinsurance compendium*, (pp. 270–287), International Labour Organization.
14. Churchill, C. and Roth, J. (2006): ‘Microinsurance: Opportunities and pitfalls for microfinance institutions’, *In*: Churchill, C. (ed.), *Protecting the poor*: *A microinsurance compendium*, (pp. 452–468), International Labour Organization.
15. Cohen, M. and Sebstad, J. (2006): ‘The demand for microinsurance’, *In*: Churchill, C. (ed.), *Protecting the poor*: *A microinsurance compendium*, (pp. 12–24), International Labour Organization.

16. Enarsson, S., Wirén, K. and Almeyda, G. (2006): 'Savings- and credit-linked insurance', *In*: Churchill, C. (ed.), *Protecting the poor*: *A microinsurance compendium*, (pp. 112–129), International Labour Organization.
17. Fernando, J. (2005): 'Micro-Credit and Empowerment: Visibility Without Power', *In:* Fernando, J. (ed.), *Prospects and Perils of Micro-Credit: Cultural Politics of Empowerment*. United Kingdom: Routledge.
18. Fischer, K. and Qureshi, Z. (2006): 'Cooperatives and insurance: The mutual advantage', *In*: Churchill, C. (ed.), *Protecting the poor*: *A microinsurance compendium*, (pp. 336–356), International Labour Organization.
19. Garand, D., Roth, J. and Rutherford, S. (2006): 'Long-term savings and insurance', *In*: Churchill, C. (ed.), *Protecting the poor*: *A microinsurance compendium*, (pp. 254–269), International Labour Organization.
20. Garand, D. and Wipf, J. (2006a): 'Pricing microinsurance products', *In*: Churchill, C. (ed.), *Protecting the poor*: *A microinsurance compendium*, (pp. 238–253), International Labour Organization.
21. Garand, D. and Wipf, J. (2006b): 'Risk and financial management',*In*: Churchill, C. (ed.), *Protecting the poor*: *A microinsurance compendium*, (pp. 254–269), International Labour Organization.
22. Hayek, F. (1945): 'The Use of Knowledge in Society Edition': *American Economic Review,* XXXV, No. 4; September, 1945, pp. 519-30.
23. International Monetary Fund (2005): *Microfinance*: *A View from the Fund***,** January 25, 2005
24. Latortue, A. (2006): 'The role of donors', *In*: Churchill, C. (ed.), *Protecting the poor*: *A microinsurance compendium*, (pp. 470–487), International Labour Organization.
25. Mc Cord, M. et al. (2006): 'The partner-agent model': *In*: Churchill, C. (ed.), *Protecting the poor*: *A microinsurance compendium*, (pp. 357–377), International Labour Organization.
26. Mises, L. (1920): 'Economic Calculation in the Socialist Commonwealth'. *In:* Hayek, F. (ed.), *Collectivist Economic Planning*. London: Routledge & Kegan Paul, 1920.
27. Morduch, J. (2004): *Micro-insurance: the next revolution?*, New York University Paper. Obtained through the Internet: http://www.microfinancegateway.org/redirect.php?mode=link&id=21189, [accessed 20 /2/2008].
28. North, D. (2005): *Understanding the Process of Economic Change*. Princeton University Press.
29. Olson, M. (1982): The *Rise and Decline of Nations.* Yale University Press. September *1982.*
30. Qureshi, Z. (2006): 'Governance', *In*: Churchill, C. (ed.), *Protecting the poor*: *A microinsurance compendium*, (pp. 288–306), International Labour Organization.
31. Roth, J., Garand, D. and Rutherford, S. (2006): 'Long-term savings and insurance', *In*: Churchill, C. (ed.), *Protecting the poor*: *A microinsurance compendium*, (pp. 94–110), International Labour Organization.

The Participation of Immigrants in Greece's Labor Force

Karamanli Theodora

Department of Balkan, Slavonic and Oriental Studies, University of Macedonia, Thessaloniki, Greece. Address: Chilis 41, 55132, Thessaloniki, tkara@uom.gr

In this paper we investigate the factors which affect the possibility of immigrants' participation in the Greek labor force. We used original data collected from the research we carried out all over Greece for the period of 2005-2006. The detailed econometric analysis of these data, proves that age, marital status and children in the family affect significantly the possibility of women participating in the labor force. The variable which shows the legal status of immigrants' residence in Greece contributes positively to the possibility of their participating in the labor market, only when it refers to illegal immigrants.

Furthermore, we gauge the factors which have an impact on the possibility of the immigrants being employed. From the total number of immigrants who are part of the Greek labor force, illegal ones stand the best chance of being employed. To a considerable extent, the results of the present research reflect features of the Greek economy, such as the demand for cheap labor of low specialisation.

JEL classification: J21, E24, F22, O52

Keywords: Labor Force participation, Unemployment rate, Immigration, Greece

1. Introduction

Since the beginning of the 90's Greece has become the host country to a great immigration influx. Hundreds of thousands of economic immigrants have chosen Greece as a destination for settling down and seeking employment. (Angrist and Kugler, 2003; Baldwin-Edwards, 2004; Venturini, 1993). In Greece and Italy, two countries that developed quickly, especially after the 70's and in which an intense informal economic activity and black labor functioned, there were jobs not occupied by natives. Those posts attracted immigrants. (see: Baldacci et al., 1999; Polyzou and Altsitzoglou, 2007; Pugliese, 1990; Reyneri, 2001).

According to Baldwin et al (1999) and King (2000), migration after the 80's to Southern Europe countries, is characterized by unskilled labor demand, especially in

the illegal labor market. The particularities of the Greek economy, such as: i) the seasonal character of several activities (tourism, fishery), ii) the existence of numerous small family businesses, which encourages the non-institutional occupation iii) the large number of self-employed (Cavounidis, 2006) and iv) the young natives' unwillingness to take on menial jobs with no social status (Malheiros, 1999), constituted a strong attraction factor for a lot of immigrants. Immigrants, to a large extent, entered Greece illegally and in the beginning, at least, they were illegally occupied in the labor market. (see Baldwin, 2004; 2005; Lianos et al., 1996).[1]

From OECD research for Greece (2005), it is concluded that the increase in the labor supply from 5% to 10% within 10 years, was smoothly absorbed without increasing unemployment. More specifically, taking into consideration that the greatest part of the immigrants is occupied in the constructions, agriculture and services sector – mainly as paid house help – the results show that in most businesses, the wages were shaped on the basis of the only labor supply by the immigrants due to lack of interest from Greeks.

From the first investigations that concern Greece's labor market, after the voluptuous immigrant entrance, is the Lianos, Katseli and Sarris (1996) research. From their study, it appeared that illegal immigrants did not replace Greek workers. Furthermore, the immigrants' employment contributed by 1% to an increase in the GDP. In their investigation, Sarris and Zografakis (1999), use a general balance model and they demonstrate the redistributive immigration consequences on a small open economy, such as that of Greece, through a theoretical and empirical analysis. They reach the conclusion that immigrants have a positive effect on the salaries of the skilled Greek employees, on Greek GDP and on occupation, while, on the contrary, the immigrants' impact on the unskilled natives' wages and those occupied in agriculture is negative. The charting and legalization of illegal immigrants will reduce the existent redistribution effects.

The research that has up to now been conducted focuses on the impact that immigration has on the labor market and on Greek GDP. In this paper we focus on the factors which have an impact on the possibility of the immigrants being employed and in particular of those who are illegal. Moreover we compute the unemployment rate they face. The information that comes to light in the current research directly concerns the Greek labor force. The immigrants into Greece, from 1997 and onwards, comprise a significant part of the total labor force and affect the labor market. From 1997 until 2004, the part of the labor force that consisted of immigrants, increased from 3,7% to 6,9% In the year 2004, 14% of the insured employees in IKA (Social Insurance Institution) were immigrants. The insurance funds and part of the national tax profits depend on immigrants.

In order to analyze the data, we first present tables with statistical facts in section 2. The partial influence of every single explanatory variable on the immigrants' decision to participate in the Greek labor force is analyzed in section 3. Our findings are that the

[1] According to estimates by Lianos et al (1993), illegal immigrants approach the number of 470.000, most of which are Albanian. In 2004, illegal immigrants are estimated to 150.000 by Baldwin, whereas according to Konti they range between 230.000 and 300.000.

participation of immigrants in Greece's labor force is associated with gender, age, family status, family income, education and legal status. In section 4 we report and discuss the results of our estimation which refer to the factors that affect unemployment among immigrants. Our main findings are that the duration of the immigrants' residence in Greece, as well as their intention to return to their home countries, the height of their family income, their residence status, are connected with the possibility of their being employed in Greece. Section 5 concludes the paper with a discussion of the implication of our results.

2. Data description

Data used in this paper are unique and come from in situs research conducted all over Greece, within the framework of doing the thesis of this writer, during the period between June 2005 and September 2006. A total number of 1440 personal questionnaires were completed. The population of the present research comprises immigrants aged 16 to 65 who come from the Balkans, countries of Eastern and S.A. Europe, Asia, Africa and North America.[1]

So as for bias to be avoid, were selected immigrants of both genders, all ages, able to work, who are employed in every kind of job, who live in urban and rural areas, married or not, and immigrants who have or do not have their family with them. In our effort to have the most representative sample with regard to the immigrants' characteristics, field work covered all twelve regions of Greece. We relied on the national census of 2001 for the registered aliens that reside in the regions above who are not natural born Greek citizens.[2] The sample that emerged is random, unbiased and representative Table 1 provides a picture of the immigrant sample that emerged from the *in site* research. The variables presented in this Table are the same as those used in the econometric estimation of the models.

Labor force includes every active worker, anyone seeking for employment, albeit unsuccessfully, and worker that have been given leave of absence by their employers for personal reasons. Not included in the labor force are those who do not work because they do not seek employment, those occupied in household chores without pay, pensioners and those unable to work due to serious health problems. (Berkowitz and Jonson, 1974). The rate of participation in the labor force is the population percentage above 16 years of age, who is part of the labor force. In most studies for the investigation of the participation percentage in the labor force, participation takes value 1 if a person does participate in the labor force and value 0 otherwise. For the needs of

[1] In the sample above are not included E.E. citizens (15), nor citizens of the U.S.A., Canada and Australia. Also excluded are repatriated people from former Soviet Union countries. However, Greeks from Northern Epirus are included, as they still lack the granting of Greek nationality.

[2] According to the 2001 census, the population of Greece is 10.964.020 of which the aliens are 797.091 or, else, 7,27% of the total population. In the population of the 797.091 aliens who were registered by the 2001 census there is no differentiation among immigrants, refugees and seasonal immigrants. The aliens who leave in Greece come from 120 countries in total. The 77,2% of the total immigrants who leave in Greece, comes from 10 countries.

the present research, we investigated the factors which affect the possibility of the immigrants' participation in the labor force of Greece. The variable we created for this purpose – Dummy LFP – takes value 1 for every immigrant who works or wants to work but cannot find a job during the investigation period and value 0 for all the rest. The fact that all immigrants questioned constitute a potential labor force rids us from incidental problems that occur after the sample selection, as well as possible bias in sample selectivity (Goldenberg, 1981; Green, 1981).

Next, the factors that affect the immigrants' chance of being employed were investigated, in particular of the illegal ones, who participate in the labor force. For our purpose, the key question in this section of the survey was the following: "Do you work now in Greece?" The possible answers to this question could be either "Yes, I work" or "No, I don't". Table 2 presents a picture of the employed immigrants out of the whole immigrant sample. From the 1440 immigrants questioned within the research framework, 2 refused to respond and 182 stated that they do not work, either because they cannot find employment or because they do not want to work. The remaining 1256 immigrants, that is 85,5% of the total sample, were working during the conduction of the research. This percentage alters a lot, if we separate the sample in legal and illegal immigrants. For the legal ones the percentage is 84,8% while for the illegal is 91,6%.

If, from the total number of immigrants who do not work, we consider only those who want to work but cannot find a job, and if we deduct pensioners, students, women who choose to look after their houses and families and immigrants who due to chronic disease cannot work, we have a better view of the occupation of the immigrants participating in the labor force in Greece. Table 2 depicts the occupation of legal and illegal immigrants in the total of the labor force. From the 1160 legal immigrants who both want and are able to work, 1125 find employment, that is 96,9%, and from the 133 illegal immigrants, 131, that is 97,7% find employment. From the combination of the facts of Table 2, it emerges that the participation in labor force, in the total of our of immigration sample population percentage, is equal to 90%, namely 89,5% for legal and 93,7% for illegal immigrants.[1]

3. Immigrants' participation in the labor force- Econometric estimation

The descriptive analysis provides a satisfying image with regard to the immigrants' occupation in Greece. For further analysis a more systematic research of the relationship every single explanatory variable on the dependent one is deemed necessary, ceteris paribus. For the above estimation the Probit method is used. The analysis is completed with the estimation of marginal effects in the variable medians, which show us the way in which the possibility of the immigrants' participation in the Greek labor force alters at the change of an independent variable, ceteris paribus. The results from the estimation are presented in Table 3.

[1] 1294 immigrants in a total of 1438 participate in the labor force.

Next, we divided the immigrant population into male and female in order to specify the factors which affect the immigrants' participation in the labor force on the basis of gender. A likelihood ratio test rejects the zero hypothesis, which means regression coefficients differ significantly between male and female and this leads to the conclusion that the divided model is regarded as the most appropriate. The results of the regression after the division appear in Table 3 as well. In the econometric analysis that follows greater emphasis is placed on the results of divided model, for the reason that has already been supplied. For certain variables in the regression of males the coefficients have not been estimated because all immigrants of the respective categories participate in the labor force.

Our econometric analysis proves that gender is statistically significant at a 1% significance level. A male immigrant's likelihood of participation in the Greek labor force is by 7,6 percentage points greater that the female counterpart and is in accordance with bibliography (see Blau and Kahn, 1995; 2001; Stier and Madel, 2003). An increase in the immigrant's age by one year enhances the possibility of their participation in the labor force by 1,8 percentage points, but at a decreasing rate, according to the theory. This positive relationship is anticipated. A person's independence from their family, their professional ambitions and the constantly increasing on a par with age economic needs force immigrants to pursue participation in the labor force. Married immigrants, in contrast with unmarried ones, are less likely to participate in the labor force. The increase in the number of children in the family, decreases the likelihood of participation in the labor force. The negative signal in the above two variables, comes from the women's working behavior, as we see after the segregation of the sample in male and female.

The variables of age, family status, and the number of children in the family have statistical significance only as regards female immigrants. It is noticed that an increase in the age of the females by 1 year enhances the likelihood of their participation by 4,6 percentage points, whereas the fact that they are married reduces the likelihood of their participation by 15,4 percentage points with regard to single, widowed or divorced females. The need of married females for the ascertainment of survival means is by large covered by the husband's occupation, whereas the females themselves choose to be occupied with child rearing and household care. Conversely, single females will pursue to participate in the labor force in order to cover their financial needs. Furthermore, most single women are younger and pursue a place in the labor force which will enable them to meet their professional aspirations (Aaronson et al., 2006). Having one more child decreases the likelihood of a female's participation in the labor force by 4,4 percentage points. Child rearing is considered a female's duty and has priority over their participation in the labor force.

The existence of financially dependent individuals on the immigrant in the home country is positively connected with the likelihood of their participation in the labor force in Greece. As regards the home country, no one nationality is statistically significant in the total sample. The home country bears statistical significance only in the case of immigrant Bulgarian males, who are less likely to participate in the labor force than the Albanians. Muslim immigrants are by 1,8 percentage points more likely to participate in the labor force compared with Christian immigrants. The above finding could be attributed to the fact that the institution of the family is stronger among Muslims, as is the model 'man, head of the family that bears all financial

responsibility'. This forces the male to be included in the labor force. Most male Muslims, as heads of their family, undertake its financial support. Moreover, Muslim immigrants come from unprivileged countries and they are more disciplined. By comparing living and working conditions in their country of origin, they accept any job they might be given in Greece.

Immigrants who have completed secondary education, by comparison to immigrants that have studied at tertiary level of education, have a greater possibility of participating in the labor force by 2,2 percentage points. The negative relationship which is noticed between tertiary education and labor force participation may be due to the discouragement from their participation in employment which do not match their qualifications (Curtain, 2001; McKenzie, 2001). In the Greek labor market there is a surplus of graduates of tertiary education while the demand is for unskilled labor. After the sample division, it is noticed that the level of education is significant only in the case of females. Female immigrants that have completed secondary education are more likely to participate in the labor force compared with those that have studied at tertiary level. This is to say there is a negative relationship between the female immigrants' years of education and their decision to participate in the labor force.

The relatively great demand for unskilled labor within the Greek labor market, in combination with the fact that educated immigrants to Greece cannot find work that matches their qualifications, accounts to a certain extent for the negative relationship between years of education and decision to participate in the labor force. Abbot–Chapman et al., 1997, attribute the negative impact of education on the decision to participate in the labor force to the individual's ambitions in connection with the decision for such participation[1]. Although other studies exist, which led to a similar conclusion (see, for example, Berkowitz and Johnson, 1974; Stern, 1996), the exactly opposite result was expected from the theory of human capital. (see Bowen and Finegan, 1966).[2]

Immigrants who know Greek language writing have by 2,3 percentage points a smaller likelihood of participating in the labor force, compared with those that have no such knowledge. Greek language writing and reading knowledge stands for an investment on the part of the immigrant. If the outcome of this investment is not satisfactory, or is not the anticipated one, the immigrant might withdraw from the labor force. Greek employers appreciate the immigrants' ability to speak the Greek language rather than their ability to read or write.

The amount of the net family income positively affects the immigrant's likelihood of participating in the Greek labor force. This suggests that more financial sources constitute an incentive for the immigrant to participate in the labor force. In one of his researches, Konstantopoulos (2005) argues that income beyond wages is a motive for prolonged participation in the labor force. Also, immigrants who entered Greece with

[1] The results of Abbot–Chapman et al., 1997 refer to male population.

[2] Based on the Human Capital Theory we expect that the higher educational level, implies higher productivity of a person, benefits on the employers' side, when they offer work to people with the before mentioned skills and as a consequence higher likelihood of participation of these people in the labour force.

the intention of providing financial support to their families in the country of origin, compared with those who left their home country due to low wages, stand better chances of acquiring a working place within the host country's labor force. Besides, the most feasible way of providing financial support to their families is through the income they will earn from being employed in the host country. Those, however, who left their home country for reasons other than the previous, are less likely to find a place within the labor force.

Variables that show the immigrant's legal framework are significant in the case of the immigrant male only. Illegal immigrants are more likely to participate in the labor force in comparison with legal ones, as the former have fewer demands, smaller power to negotiate and are eager to accept any kind of work. In a research of theirs, Venturini, (1989) and Natale, (1990), prove that immigrant participation in the labor force is tightly connected with their legal status, since they are more likely to be occupied in the informal economy sector. Upon entering a foreign country, illegal immigrants confront greater difficulty in finding employment than their legal peers. For this reason, a tendency to make use of immigrant networks created by earlier immigrants is observed (Borjas, 1994; Ethier, 1986). In their turn, pre-existing immigrants prefer to employ immigrants from their own country of origin, as these come from the same social environment, the same culture and speak the same language, which intensifies network externalities (Carrington et al, 1996; Chiswick and Miller, 1996; Church and King, 1993; Marks, 1996).

4. Factors which affect immigrant unemployment – Econometric estimation

In the analysis that follows, the question "Do you work now in Greece?" constitutes the dependent variable and addresses immigrants participating in the labor force. Despite the fact that on the basis of the likelihood ratio test marginally more appropriate is considered the gender segregated model, greater emphasis is placed on the analysis that concerns the total of the immigrants, because in the regression of the total sample, gender is not statistically significant.

Regression results are summarized in Table 4. As it has already been said, the immigrant's gender does not significantly affect their likelihood of being employed in Greece. The greatest statistically significant possibility to be employed has those immigrants who are aged 21–35, in comparison with immigrants who are younger. Married immigrants, as well as an immigrant's years of education, are negatively connected with the likelihood of their being employed in Greece. The negative signal in married immigrants is due to the females, as it emerges from Table 4 regression, and it is probably caused by the females being refused entrance into certain professions. With regard to years of education, the negative signal expresses the demands of the Greek labor market; namely, it reveals that the Greek labor market requires immigrants with little education. According to Wooden (1999), in developed economies it is noticed that full-time employment posts are gradually replaced by part-time ones, many of which require a low level of education. The fact that people with tertiary education have increased qualifications acts as a deterrent to their being hired, as

employers will select under-qualified people in their effort to reduce production costs. Besides, an over-qualified post candidate is likely to refuse employment of this sort.

Immigrants who live and work illegally in Greece were more likely to be employed at the time this research was conducted, since the uncertainty stemming from their illegal residence status drives them to accept any kind of employment they are offered. According to Bailey (1985), unregistered immigrants are more willing to compromise with lower wages on account of their legal status than they would if they were legally employed. This probability is greater and more significant when it concerns male immigrants. Employers, on the other hand, in their effort to minimize production cost, hire illegal immigrants more easily. As a result, illegal immigrants with low education are preferred to legal ones with high education. By contrast, males who hold the Special Identity Card of Greek Ancestry, appear as less likely to be employed than legal immigrants.

Immigrants and namely the male ones that send money to their home country are more likely to be employed. Given the fact that immigrants in Greece are economic immigrants, the likelihood of remitting money to their country of origin is greater for those employed than those out of work. The immigrants' length of residence in Greece is linked positively with their employment likelihood. The more the years they reside in Greece the better they adapt to the labor market demands, the better they acquire professional skills, learn the Greek language and are more likely to participate actively in the Greek labor force. The variable which describes an immigrant's intention to return to his home country is connected negatively with the likelihood of their being employed, since the chance of remaining unemployed in the host country constitutes a powerful incentive for returning home. Family income is positively connected with the immigrants' employment chances. Also, it is linked directly with man's living conditions and quality of life, both of which are greatly enhanced by means of wages.

On interpreting certain points which appear in Table 4, we could argue that the area of residence is statistically significant in the case of male immigrants only. Those who reside in urban areas are less likely to be employed, as unemployment mainly strikes city-dwelling immigrants. Furthermore, there is a negative relationship between the males' years of education in the home country and their employment probability in Greece. This fact reflects Greek labor market demands for labor force of low education or for education that differs from what the immigrants' country of origin provides.

Finally, the shorter the employment length of male and female immigrants, besides present employment, the more likely it is for them to be employed at the time of the research conduction. In combination with the positive relationship which connects length of employment in Greece with the likelihood of employment, it is concluded that most immigrants are economic immigrants who, after changing several jobs in the beginning of their life as immigrants in Greece, are now steadily employed in their present job and have done so for a long time.

5. Conclusion

In the present paper we estimated, with the Probit econometric analysis and we analyzed the factors that affect the immigrants' decision to participate in the Greek

labor force, as well as the factors that modulate the unemployment percentage that they confront. The data were obtained from an in situs research all over Greece, which was conducted in 2005-2006.

Our econometric results prove that age, family income, existence of financially dependent people in the country of origin and secondary education, as opposed to tertiary, are positively related with the female's participation likelihood in the labor force, while family status and the number of children negatively. With regard to male immigrants, Muslims, in relation with Christians, and illegal immigrants, in relation with legal ones, stand a greater chance of participating in the labor force.

The years of formal education of the male immigrants and their working experience in Greece besides their present employment, are negatively connected with the likelihood of their being occupied, which reveals that the Greek labor market requires labor force of low education and skill. Married female immigrants are less likely by 3,2 percentage points to find occupation. Also, less likely to be employed are the male who reside in urban areas. Total family income is positively related with the occupational likelihood of both men and women.

Our investigation reveals that 15 years after the massive entrance of immigrants into Greece, illegal immigrants, as opposed to legal ones, are more likely to become part of the Greek labor force and have the same - if not more - occupational opportunities as the legal immigrants. As long as the labor market continues to be supplied with illegal labor, social security funds will continue to lose part of their revenue and the State will suffer the same in terms of income lost from taxes. Illegal immigrant employment further aggravates an already large informal economic sector in Greece, with all the negative effects that the latter can have on the economic growth of Greece.

According to an OECD report (2005), part of the economic activity in Greece continues to be rooted in informal economy, which is largely supplied with illegal immigrant labourers, most of which are unskilled. Even though businesses can have only short-term financial benefits from hiring low-cost unskilled or illegal immigrant labor, by doing so they are led to a decrease in their productivity and to a delay in dealing effectively with their structural problems. The fact that businesses thus delay modernization renews the demand for cheap labor and sustains informal economy. Of course, the so far realized legalisation of formerly illegal immigrants, within the framework of the immigration policy, constitutes a positive step towards the smoothening out of the above situation.

Without doubt, illegal immigration into Greece has benefited the country's economy. Low labor and production costs have motivated an increase in production and profits for some businesses. From now on, priority should be given to the competitiveness of the economy and to the improvement of its structural problems. The credibility of European countries, with regard to establishing a framework of legal economic immigration, depends on their ability to combat illegal labor and illegal immigration. So, if the aim is the healthy development of the country, it is suggested that tighter control should be made at all Greek borders and other points of entrance of illegal immigrants. Also, bureaucracy and the cost of legalizing residence and labor should be reduced, motives should be given for legal occupation, and businesses that employ labor illegally should be penalized. It cannot be denied that all legalization of formerly

illegal immigrants that has so far been realized constitutes a positive step forward. Nevertheless, it has not led to an improvement of the situation yet.

Variables	Variable description	Number of obs	%
Gender: male	1 if individual is male; 0 otherwise	823	57,2%
Female	1 if individual is female; 0 otherwise	617	42,8%
Age group: 16-20	1 if individual is 16-20 years old; 0 otherwise	138	9,6%
21-35	1 if individual is 21-35 years old; 0 otherwise	568	39,4%
36-50	1 if individual is 36-50 years old; 0 otherwise	623	43,3%
51-65	1 if individual is 51-65 years old; 0 otherwise	111	7,7%
Married: yes	1 if individual is married; 0 otherwise	903	62,7%
Number of children: 0	1 if individual has no children; 0 otherwise.	525	36,5%
1	1 if individual has 1 child; 0 otherwise.	265	18,4%
2	1 if individual has 2 children; 0 otherwise.	460	31,9%
3	1 if individual has 3 children; 0 otherwise.	132	9,2%
more	1 if individual has more than 3 children; 0 otherwise..	58	4,0%
Dependants in country of origin: yes		896	62,2%
Country of origin: Albania	1 if individual owns the citizenship of Albania; 0 otherwise.	914	63,5%
Bulgaria	1 if individual owns the citizenship of Bulgaria; 0 otherwise.	125	8,7%
Georgia	1 if individual owns the citizenship of Georgia; 0 otherwise.	52	3,6%
Romania	1 if individual owns the citizenship of Romania; 0 otherwise.	44	3,1%
Russia	1 if individual owns the citizenship of Russia; 0 otherwise.	50	3,5%
Poland	1 if individual owns the citizenship of Poland; 0 otherwise	44	3,1%
Other	1 if individual owns the citizenship of other country of questionary; 0 otherwise	261	14,5%
Place of residence: Urban	1 if individual is urban resident; 0 otherwise.	1114	77,4%
Religion: Christian	1 if individual is Christian; 0 otherwise.	999	69,4%
Muslims	1 if individual is Muslim; 0 otherwise.	360	25,0%
Others	1 if individual has an other religion; 0 otherwise.	81	5,6%
Education level: Elementary till 9 years)	1 if individual has gone up to elementary school; 0 otherwise	420	29,2%
Secondary (9-12 years)	1 if individual has gone up to secondary school; 0 otherwise.	629	43,7%
Higher	1 if individual has a high schooling degree; 0 otherwise	391	27,1%
Speak Greek	1 if individual speaks well Greek language; 0 otherwise.	1335	92,7%

Write Greek	1 if individual writes well Greek language; 0 otherwise	771	53,5%
Legal status: legal	1 if individual leaves legally in Greece; 0 otherwise	1069	74,2%
Illegal	1 if individual leaves illegally in Greece; 0 otherwise.	139	9,7%
Ε.Δ.Τ.Ο.[1]	1 if individual holds the Special National Identification Card but is not natural born Greek citizen, has political asylum, is seasonal worker; 0 otherwise.	232	16,1%
Send money at home: yes	1 if individual send money in home country; 0 otherwise	614	42,6%
Intend to leave: yes	1 if individual intend to return home; 0 otherwise	756	52,5%
Reason you immigrate:			
Unemployment at home	1 if individual left his country because of unemployment; 0 otherwise	557	38,7%
Political crisis	1 if individual left his country because of political crisis; 0 otherwise.	142	9,9%
Low wages	1 if individual left his country because of low wages; 0 otherwise	226	15,7%
Support family	1 if individual left his country to support financially his family; 0 otherwise	152	10,6%
Other	1 if individual left his country in order to study in a better university, to leave in abroad, to met his family/friends in Greece; 0 otherwise.	363	25,2%

Table 1. Sample characteristics through the variables used in regressions

Source: Data base from *in site* research by the author.

	of the total immigrant population			of the total immigrant labor force		
	Legal	Illegal	Total	Legal	Illegal	Total
I work	1125	131	1256	1125	131	1256
I do not work	171	11	182	35	3	38
Total	1296	142	1438	1160	133	1294
Non workers as % of the total	15,2	8,4	14,5	3,1%	2,3%	3,03%

Table 2. Occupation of legal and illegal immigrants of the total immigrant population and of the total immigrant labor force

Source: Data base from *in site* research by the author.

[1] The variable refers to holders of the Special National Identification Card who are not natural born Greek citizens.

Independent Variables	Coeffi -cients	Marginal Effects	Coeffi -cients	Marginal Effects	Coeffi -cients	Marginal Effects
Gender -Male	0,869***	,076***	Males		Females	
Age	0,242***	,018***	0,146	0,001	0,232***	,046***
Age^2	-0,003***	-0,001***	-0,002*	-1,87e-06	-0,003***	-0,001***
Married	-0,633***	-0,042***	0,829	0,001	-0,882***	-0,154***
Number of Children	-0,202***	-0,015***	-0,031	-0,001	-0,225**	-0,044**
Financially dependent individuals in home country	0,511***	0,034***	0,845**	0,001	0,391**	0,070**
Country of Origin						
Albania reference group						
Bulgaria	-0,210	-0,018	-1,03*	-0,006	-0,107	-0,022
Georgia	0,318	0,018	-0,465	-0,001	0,392	0,062
Romania	0,051	0,004			-0,100	-0,021
Russia	-0,278	-0,026			-0,403	-0,096
Poland	-0,188	-0,016			-0,249	-0,056
Other country	-0,186	-0,015	-0,314	-0,001	-0,248	-0,054
Religion						
Christian reference group						
Muslim	0,271	0,018*	0,540*	0,001	-0,022	-0,004
Other religion	0,119	0,008	-0,196	-0,001	0,187	0,033
Education						
High reference group						
Elementary	0,054	0,004	-1,112	-0,001	0,023	0,004
Secondary	0,305**	0,022**	0,347	0,001	0,303*	0,060*
Knowledge of Greek Language						
Speak	0,141	0,011	0,498	0,001	-0,097	-0,018
Write	-0,318**	-0,023**	-1,003***	-0,001	-0,185	-0,036
Legal status						
Legal reference group						
Illegal	0,253	0,015	0,677*	0,001	0,192	0,034

ΕΔΤΟ owner	0,016	0,001	-0,371	-0,001	0,090	0,017
Total, net, family Income	0,001***	0,001***	,003***	2,20e-06	0,001***	0 ,001***
Main reason you left your country						
Low wage reference group						
Unemployment for a long time	0,098	0,007	0,251	0,001	0,044	0,009
Political crisis	0,173	0,011	0,614*	0,001	0,076	0,014
To support financially my family	0,508	0,026**			0,334	0,055
Other reason	-0,457**	-0,041*	-0,031	-0,001	-0,560**	-0,122*
Constant	-4,036***		-2,569*		-3,376***	
Log-pseudolikelihood	-296,02619		-54,561222		-213,67133	
Pseudo R^2	0,3706		0,5887		0,2689	
Prob>chi2	0,0000		0,0000		0,0000	
Number of observations	1440		693		617	

Table 3. Estimation Results of the Probit Model – Factors that affect the likelihood of immigrants' participation in the Greek labor force. Dependent Variable: «Labor Force Participation» (Dummy LFP)

Source: Data base from *in site* research by the author.
Footnotes:1) *,**,*** mark significance levels 10%, 5% και 1% respectively.
2) Standard error corrections for heteroskedasticity have been made.

Independent Variables	Coeffi-cients	Marginal effects	Coeffi-cients	Marginal effects	Coeffi-cients	Marginal effects
Gender -Male	0,208	0,003	Males		Females	
Age 16-20 reference group						
Age 21-35	0,573*	0,006	0,818	0,0002	0,483	0,0116
Age 36-50	0,377	0,004	0,461	0,0001	0.669	0,0220
Age 51-65	0,451	0,003	0,352	0,0001		
Place of residence: Urban	-0,180	-0,002	-1,209***	-0,0002	0,206	0,0064
Family status: Married	-0,811***	-0,008*	-0,366	-0,0001	-1,366***	-0,0322***
Years of formal education	-0,066**	-0,001	-0,098***	-0,0001	-0,057	-0,0015
Knowledge of Greek	0,511*	0,011	0,729	0,0005	0,667*	0,0353

Language (speaking)						
Country of origin						
Albania reference group						
Bulgaria	0,238	0,002	-		-0,024	-0,0007
Georgia	-0,449	-0,009	-0,160	-0,0001	-0,252	-0,0089
Romania	-0,685	-0,019	-		-0,721	-0,0415
Russia	0,534	0,003	-		0,182	0,0041
Poland			-		-	-
Other country	0,099	0,001	0,567	0,0001	-0,133	-0,0040
Legal status						
Legal reference group						
Illegal	0,467*	0,003	1,168***	0,0001	0,334	0,0066
ΕΔΤΟ owner	-0,328	-0,005	-0,574**	-0,0002	0,130	0,0032
Working Experience						
Employment length in Greece besides present employment	-0,178***	-0,002**	-0,312***	-0,0001	-0,111***	-0,0030*
Employment length in home country	0,018	0,001	-0,014	-0,0001	0,033	0,0009
Years residence in Greece	0,056*	0,001	0,097**	0,0001	0,024	0,0006
Total, family, net income	0,002***	0,001***	0,003***	0,0001	0,001***	0,0001***
Send money in home country	0,697***	0,008*	1,043***	0,0003	0,161	0,0041
Leave Intention	-0,410**	-0,005	-0,696**	-0,0001	0,050	0,0014
Constant	0,892		0,939		0,844	
Log-pseudolikelihood	-109,11506		-40,8905		-48,6788	
Pseudo R^2	0,3259		0,565		0,2435	
Prob>chi2	0,0000		0,0000		0,0000	
Number of observations	1206		690		410	

Table 4. Estimation Results of the Probit Model. Factors affecting the employment likelihood of immigrants that constitute part of the labor force. Dependent Variable: «Do you work now?»

Source: Data base from *in site* research by the author.

Footnotes: 1) *,**,*** mark significance levels 10%, 5% και 1% respectively.

2) Standard error corrections for heteroskedasticity have been made.

References

1. Aaronson, S., Fallick, B., Figura, A., Pingle, J. and Wascher, W. (2006) 'The Recent in Labor Force Participation Rate and its Implications for Potential Labor Supply', Division of Research and Statistics Board of Governors of the Federal Reserve System, *Booking Papers on Economic Activity,* 1-65.

2. Abbott-Chapman, J., Easthope, G. and O' Connor, P. (1997) 'The Influence of Student Gender and Parental Socio-economic Status on Post–school Career Paths', *Australian Journal of Social Research,* Vol 4, No 1, pp.2-26.
3. Angrist, J. and Kugler, A. (2003) 'Protective or counter-productive? Labor market institutions and the effect of immigration on EU natives', *The Economic Journal,* Vol. 113.
4. Bailey, T. (1985) 'The Influence of Legal Status on the Labor Market Impact of Immigration, *International Migration Review,* Vol 19, pp 220-238.
5. Baldacci, E., Inglese, L.and Strozzi, S. (1999) 'Determinants of Foreign Workers' Wages in two Italian Regions with High Illegal Immigration', *Labour,* Vol 13, No. 3.
6. Baldwin-Edwards, M. and Arango, J., eds., (1999), *Immigrants and the informal Economy in Southern Europe*, London: Frank Cass.
7. Baldwin-Edwards, M. (2004) 'Albanian emigration and the Greek labour market', *South East Europe Review*, Vol.7, No. 1. Available from: http://www.migrants.gr/files/RECOURCE_123.pdf
8. Baldwin-Edwards, M. (2004) 'Statistical Data on Immigrants in Greece', a study prepared for IMEPO, Ministry of the Interior. Available from www.mmo.gr
9. Baldwin-Edwards, M. (2005) 'Migration into Southern Europe: Non-legality and labor markets in the region', *MMO Working Paper* No. 6, Dec.2005.
10. Berkowitz, M. and Johnson, W.G. (1974) 'Health and Labor Force Participation', *The Journal of Human Resources,*Vol 9, No.1, pp 117-128.
11. Blau, D.F. and Kahn, L.M. (1995) 'The Gender Earning Gap, Some International Evidence', *In*: Freeman R.B. and Katz L.F.(Eds), *Differences and Changes in Wage Streetcars,* The University of Chicago Press.
12. Blau, D.F. and Kahn, L.M. (2001) 'Understanding International Differences in the Gender Pay Gap', *Working Paper 8200*, National Bureau of Economic Research
13. Borjas, G.J. (1994) 'The Economics of Immigration', *Journal of Economic Literature,* Vol.32, No. 12.
14. Bowen, W.G. and Finegan, T.A. (1966) 'Educational Attainment and Labor Force Participation, *The American Economic Review,* Vol.56 No.1/2, pp 567-582.
15. Carrington W.J., Detragiache E. and Vishwanath T. (1996) 'Migration with endogenous moving costs', *American Economic Review*, Vol.86, pp.909-930.
16. Cavounidis, J. (2006) 'Labor Market Impact of Migration: Employment Structures and the Case of Greece', *International Migration Review*, Vol.40, pp. 635-660.
17. Chiswick, B. and Miller P. (1996) 'Language Skills and Earnings among Legalized Aliens', *paper presented at the 1997 American Association Meetings*, New Orleans, mimeo.
18. Church J.and King I. (1993) 'Bilingualism and network externalities', *Canadian Journal of Economics*, Vol.26, pp. 337-345.
19. Ethier E.J. (1986) 'International Trade Theory and International Migration', *American Economic Review*, Vol.75, No.4, pp. 691-707.
20. Goldberger, A.S. (1981) "Linear Regression After Selection", *Journal of Econometrics,* 15(3).

21. Curtain, R. (2001) How young people are faring 2001: Learning, work and in between, Curtain Consulting, Melbourne.
22. Greene, W. (1981) 'On the Asymptotic Bias of the Ordinary Least Squares Estimator, *Econometrica,* Vol.49, No.2.
23. King, R. (2000) 'Southern Europe in the changing global map of migration', *In*: R.King, G. Lazaridis and C.Tsardanidis (eds), *Eldorado or fortress?: migration in Southern Europe,* Macmillan New York: St. Martin's Press, Basingstoke
24. Konstantopoulos, S. (2005) The Correlates of Work in a Post-AFDC World: The Results from a Longitudinal State-Level Analysis, North-western University, USA.
25. Lianos, T., Sarris, A, and Katseli, L. (1996) 'Illegal Immigration and Local Labor Market: The case of Northern Greece', *International Migration,* Vol.34, No.3.
26. Malheiros, J.M. (1999) 'Immigration, Clandestine Work and Labour Market Strategies: The Construction Sector in the Metropolitan Region of Lisbon', *In*: Martin-Baldwin and Arango (eds), *Immigrants and the informal Economy in Southern Europe*, London: Frank Cass.
27. McKenzie, P. (2001) 'Pathways for youth in Australia', paper presented to the Conference on Vocational Training and Lifelong Good and Gone: the Great Black Migration, Bloomington, Indian University Press
28. Natale, M. (1990) 'La presenza straniera', *Learning* in Australia and Germany, Australia Centre, University of Potsdam, Germany, 29-31 May.
29. Marks C. (1989) *Farewell- We're In*: Natale M. (ed) *Economia e popolazione*, (pp.351-451), Milane: F. Angeli:.
30. OECD, (2005) *Economic Surveys: Greece*, *2005*, issue 12, pp. 117-140.
31. Pugliese, E.(1990) 'Gli immigrati nel mercato del lavoro', *Polis*Vol.1.
32. Reyneri, E. (2001) 'Migrants in Irregular Employment in the Mediterranean Countries of the European Union', Geneva: ILO, *Paper IMP* No.41. Available from: http://www.ilo.org/public/english/protection/midrant/downloaded/imp/imp41.pdf
33. Sarris, A. and Zografakis, S. (1999) 'A computable general equilibrium assessment of the impact of illegal immigration on the Greek economy', *Journal of Population Economics,* Vol.12, pp.155-182.
34. Stern, S. (1996) 'Semiparametrric estimates of the Supply and Demand effects of Disability on Labor Force Participation', *Journal of Econometrics*, Vol.71, pp.49-70.
35. Stier,H. and Madel H. (2003) 'Inequality in the Family: The Institutional Aspects of Wives' Earning Dependency', *Paper presented in the ISA Researce Committee on Social Stratification* (RC28) Tokyo, 2003.
36. Venturini, A. (1993) 'Changing Patterns of Labour Migration in Southern Europe', paper presented for the conference Economic Change in Southern Europe by the social Science Research Council.
37. Wooden, M. (1999) Review of research: Impediments to the Employment of Young people, NCVER, Leabrook, SA.
38. Πολύζου, Α. και Αλτσιτζόγλου, Ν. (2007) Πολιτικές ένταξης των μεταναστών: Η Ευρωπαϊκή Εμπειρία. Η περίπτωση της Ιταλίας, Εκδόσεις ΙΜΕΠΟ, Αθήνα.

The Contribution of Corporate Social Responsibility in the Adoption of Solar Power from Greek Companies

Lazaridou, Elissavet

University of Coventry, lazaride@coventry.ac.uk, ella@rc.auth.gr

A company's activity often creates negative economic, social and environmental repercussions. The development of Corporate Social Responsibility aimed to minimise any damage caused from enterprises to people outside and within the firm (Hopkins, 2007). Because of the increasing awareness about the disruption of the ecosystem the environmental aspect of CSR is in the front line of the engrossment. For the solution of many environmental problems, recent years started an extensive use of renewable energy forms, as solar power. The installation of photovoltaic systems for the production of solar power is embraced from many corporations and the scientific world as solar power does not have emissions, is a clear energy and reduces the firm's dependence on third parties. This research's intension is to investigate the commitment of Greek companies to the environmental aspect of CSR, through their readiness to adopt solar power. Furthermore, the author's attempt is to determine the indicators that have a positive or negative contribution in the adoption. The study is based upon a survey with questionnaire. The sample for the questionnaire technique is Greek firms with manufacture activities. On the whole, 63 questionnaires, distributed in two Greek counties, Thessaloniki and Kozani, were thoroughly answered.

Keywords: Corporate Social Responsibility, environmental CSR, solar power, Greek companies

1. Introduction

Environmental issues as the climate changes, the global warming and the energy problem are nowadays in the front line of the engrossment. In that frame, it is expected from the business sector to have a more ethical environmental performance (Haines et al, 2007). This aspect relied on the "externalitiy" theory. Every act of a business organization produces some "externalities" (Mankiw, 2001, p. 256). With the word externality we mean the impact of an individual's acts on the prosperity of his/her neighbours (Mankiw, 2001). If these externalities are negative it is moral to try to reduce them.

Logical outcome of the above approach is the development of a corporate social responsibility (CSR) theory. There is a lack of a broadly agreed definition for CSR. Hopkins (2007 p 15) offers the following definition: "CSR is concerned with treating the stakeholders of the firm ethically or in a responsible manner. Ethically or responsible means treating stakeholders in a manner deemed acceptable in civilized societies. Social includes economic and environmental responsibility". Stakeholders are groups of people both outside and within the organization. It could be mentioned that the further aim of corporate social responsibility is to increase continuously the standard of living and parallel to preserve the company's profitability, for all the people outside and within the firm (Hopkins, 2007).

In another definition stating by the World Business Council for Sustainable Development (1998), CSR is defined as "continuing commitment by business to behave ethically and contribute to economic development while improving the quality of life of the workforce and their families, as well as of the local community and society at large". According to the European Commission (2001, p 8) a social responsible should "not only fulfilling legal expectations, but also going beyond compliance and investing more into human capital, the environment and relations with stakeholders".

Therefore, a company should adopt fair and just labor practices friendly to their stakeholders. "Stakeholders are groups and individuals who can affect, or are affected by, the achievement of an organization's mission" (Freeman 1984, p 52). In addition, a firm should evaluate that social policy and environmental performance practices which will not disturb the ecosystem and will cause the minimum environmental pollution.

2. Theoretical framework

Many researches plead for the adoption of CSR. McWilliams and Siegel (2001) argue that companies which apply CSR policies are rewarded with their employees' loyalty – and loyalty is the gear for a better business performance. There is also strong evidence that as the wave of activists increasing the last decades, companies that are concerning about moral issues have positive effects in public relations, and an increasing reputation (Kovacs, 2006). Balabanis, et al (1998) in their study for the impact of CSR in the economic performance found positive relation between the implementation of CSR practices and financial ratios as the gross profit to sales ratio and also the overall profitability. As far as the behavior of consumers is concerned, preliminary studies suggest that consumers expect the companies to be social conscious and reward the social responsibility with their buying behavior (Becker-Olsen et al, 2006).

Concerning the environmental aspect of CSR, the compliance of organizations in the principles of environmental responsibility is increasing, especially among large companies (Berry and Rondinelli, 1998; Rondinelli and Berry, 2000). In their research Chapple, Morrison Paul and Harris (2005) about the benefits and costs of the waste reduction in United Kingdom's manufacturing, from 1991 – 1998, found that the waste minimization is caused either from environmental CSR or pecuniary reasons. Furthermore, environmental CSR activities are positively correlated with reduction in transaction costs, emission and risk according to Cruz (2008).

In regard to the environmental issues many techniques have been developed friendly to the ecosystem for the production of energy. One of these concerns the utilisation of solar power or solar energy. Solar power is the energy that derives from the sunlight. Due to Reinhardt and Vietor (1996), solar power consists a steady energy source equivalent to 1 Kilowatt-hour (KW) per square meter (m^2) at the Earth's acreage. Wallner (1999) in his study for the industrial development suggest the industrial ecology model as the only that ensures sustainable development and the solar power as the only serviceable power form in the service of this purpose as it is not related to the product or industrial process. Solar power is preferred as it is considered as a safe and almost absolute clean energy source (Tsoutsos et al, 2005). Solar power has many applications; can be used to produce lighting, generating electricity, heating and cooling.

The technologies that exploit the solar energy are separated into passive solar energy technologies, active solar energy technologies, solar thermal technologies, solar ponds and photovoltaic cells. Passive solar technologies are used for lighting, heating and cooling of buildings. Active technologies are used for cooling and heating of structures. From their part, the solar thermal technologies are used to provide heating by gathering the sun rays. Furthermore, solar ponds are devices that store and provide low-temperature thermal power with as a consequence the production of heating or even electricity. Finally, photovoltaic cells or photovoltaic arrays are devices that are solid-state, capable of converting the light from the sun directly into electricity (Reinhardt and Vietor, 1996). That is the great advantage of photovoltaic cells. Due to the little environmental effect of the other solar technologies or because of their small viability, the expansion of the photovoltaic cells is highly increasing the last years (Bahaj, 2002).

A country which has a high ratio of sunlight and therefore constitutes a good example for the diffusion of solar photovoltaic cells is Greece. The photovoltaic cells receive the sunlight and convert it into solar electricity. The performance of these systems depends on the sunlight that exists in the locality of installation (Pearce, 2006).

According to the European Commission (2007) Greece is a country among them with the highest ratio of sunlight in Europe but conversely the second from the end in the list of EU countries in accordance with the gross electricity generation from photovoltaic systems. The first position is possessed by Germany and follows Spain, Holland, Italy, France and Austria. It is indicative of Greece's capability for the production of solar power the fact that the average annual output per square meter in Germany is between 700-1200 kilowatt-hours (KWh) while in Greece with conservative estimations the produced kWh will be 30-40% higher than in countries of central Europe.

3. Methodology

In this research, the author's intention is to investigate the readiness of Greek enterprises to adopt a green energy, as it is solar power through the installation of solar photovoltaic systems. Moreover, to determine if this possible readiness for the adoption of solar power become as part of their environmental performance. The

motivation is that Greece constitutes a country that rarely has been investigated about the CSR practices of its companies and additionally.

The conversance of Greek companies to the adoption of solar power will be considered through the analysis of questionnaires that were distributed in two Greek counties, Thessaloniki and Kozani. These areas were chosen as the first includes the second largest city of Greece with the pollution an industrial city has. The second one is a rural area of North Greece rich in deposits of lignite. In this area there are four (4) power-houses that generate electricity from the lignite resources, with positive and negative consequences the economical growth and the environmental aggravation of the area, respectively (Public Power Company, 2007).

The questionnaire has been used to generate data for the environmental performance issues and the solar power. The sample has been selected with the two-step process of stratified sampling technique (Malhotra and Birks, 2007) without regard to the company's size or the type of corporate body. This survey was conducted of 63 corporations with manufacture activity which have website. From the county of Thessaloniki were 44 and 19 from the county of Kozani. The questionnaires have been distributed by hand to each respondent or answered by telephone between 1 and 21 of February, 2008. The sample was stratified by activity and website possession because of the larger environmental effects and power consumption of the manufacture companies and a positive intention in new technologies that indicates the website possession.

The population of interest in that research had been estimated as the members of the Industrial Chambers of these areas. Thessalonikis Chamber of Small Industries had 40000 members (SICT, 2007) and Thessalonikis Chamber of Commerce and Industry, 240 Industries (CCIT, 2007) and Kozanis Chamber had 1661 small industries as members (KCCSI, 2007). At this point it should be mentioned that Greece is a country with lack of extended heavy industry. In contrary, a large number of small industries are existed. In North Greece where the counties of Thessaloniki and Kozani are extended the corporations of Middle Tension according to Public Power Company are 3500 (PPC, 2007). Middle tensions corporations are the most depended on power and have the biggest power consumption.

The questionnaire included pre-coded, closed choice questions. These questions were used as they were easier to be answered and additionally, easier the data to be analysed (Lancaster, 2005). The questionnaire was short with twenty four questions in Greek language which have been selected to serve the research's scope. The questionnaire was consisted of three parts. The first contained questions concerning demographic characteristics of respondents, the second tried to demonstrate the attitude and practice of companies for environmental CSR and the third the readiness of the companies to adopt solar power and the key factors that provoke a possible slow readiness of the firms or enhance the adoption of solar power.

As far as the demographic characteristics of the respondents is concerned, they were displayed in four questions about the position the respondent held in the company, the industry in which belonged the organisation, region of the Greek area where the company operated and organisation's size. The first three questions can be met in several similar studies (Papasolomou-Doukakis et al, 2005). The fourth demographic

question was referred to the company's size and for the measurement was used the European Union's size ranking with the number of employees (European Environment Agency, 1998).

Five questions were multiple-choice questions; two open-ended referring to demographic characteristics and one followed the rank order scaling. In that case the respondent was asked to rank possible reasons that enhance the adoption of solar power according the criterion of importance (Malhotra and Birks, 2007). Moreover, in that particular question the respondent had the choice to mark other reason of the offered.

For the rest fifteen questions a five point Likert-style rating scale was used. The reader had the possibility to express his opinion in a level from strongly agree to strongly disagree. These questions discoursed a possible existence of CSR practice's in enterprises and the companies' stance about solar power. Two of them about the possible commitment of the organization at the environment with actions as recycling, conservative resource use, reduction of emissions and if it is a firm's obligation, actions for the environmental protection, were met in a similar form in the research of Papasolomou-Doukakis et al, (2005). Additionally, the next three questions were relied on Unite Nation's (1993); Rondinelli and Berry's (2000) ranking of environmental performance practices. Part three of the questionnaire had reference to companies' stance about solar power.

The data analysis of the questionnaires was become with the usage of statistic methods. For the practical applications of these methods SPSS Version 15 software was used.

The information for quantitative data is summarised in the following presented analysis. Starting with the demographic analysis of the answered questionnaires, 63 companies, industries and light manufactures, participated in the survey. The corporations belonged to 19 different industries. The industry of each of them is summarised in the following table (Table 1). The ranking of the industries has been made according to the London Stock Market classification system. The industry with the most respondents was the Food Products industry, with 16 corporations.

From the respondents 20 had the position of president in the organisation, 12 this of manager director, 18 belonged to the upper management, 12 to the other management and 1 to the rest personnel.

From the companies, 44 were operated in the county of Thessaloniki and 19 in the county of Kozani, whereas 41 corporations engaged 2 to 50 employees, 15 engaged 51 to 250 employees and 7 companies 251 and up. Furthermore, from the 63 enterprises, only two had already installed some form of renewable energy.

3.1 Factor Analysis

The items of the five point Likert - style scale were subjected to principal components analysis (PCA) using SPSS. Prior to performing PCA, the appropriateness of data for factor analysis was assessed. Examination of the correlation matrix revealed the presence of many coefficients of .3 and above. The Kaiser-Mayer-Oklin value was .687, exceeding the suggested value of .6 (Pallant, 2006) and Bartlett's Test of

Sphericity reached statistical significance, supporting the factorability of the correlation matrix. Using scree test it was decided to retain two components for further investigation.

Industry

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Food Products	16	25,4	25,4	25,4
	2 Brewers	2	3,2	3,2	28,6
	3 Farming & Fishing	1	1,6	1,6	30,2
	4Commodity Chemicals	1	1,6	1,6	31,7
	5 Specialty Chemicals	3	4,8	4,8	36,5
	6 Steel	2	3,2	3,2	39,7
	7 Industrial Machinery	6	9,5	9,5	49,2
	8 Building Materials & Fixtures	8	12,7	12,7	61,9
	9 Construction	2	3,2	3,2	65,1
	10 Durable Household Products	3	4,8	4,8	69,8
	11 Electrical Components & Equipment	4	6,3	6,3	76,2
	12 Furnishings	2	3,2	3,2	79,4
	13 Clothing & Accessories	2	3,2	3,2	82,5
	14 Electricity	1	1,6	1,6	84,1
	15 Integrated Oil & Gas	1	1,6	1,6	85,7
	16 Automobiles	1	1,6	1,6	87,3
	17 Paper	1	1,6	1,6	88,9
	18 Aluminium	3	4,8	4,8	93,7
	19 Containers & Packaging	4	6,3	6,3	100,0
	Total	63	100,0	100,0	

Table 1: Industrys' classification

The two component solution explained a total of 48.3%of the variance, with Component 1 contributing 33.3% and Component 2 contributing 15% (Table 2). The interpretation of the two components consisted of solar power readiness items, loading strongly on Component 1 and environmental CSR performance items, loading strongly on Component 2. Additionally, there was a week positive correlation between the two factors (r = +.22).

The reliability of the scale measured with Gronbach alpha coefficient which has to be reported above .7 (Pallant, 2006). In the current research, the Gronbach alpha coefficient of the environmental CSR performance was .79 and of the solar power readiness .80 (Please see Appendix III). Therefore, the internal consistency of the Likert –style scale was satisfactory. The result entailed the reliability of the scale.

The preliminary analyses of the continuous variables in Likert style ranking gave us some interesting results. First of all, the first five variables concerned the general environmental CSR practice of the company. The respondents had a trend to declare that their company had a commitment to the environment, with a mean of 3.43 and that is corporations' obligation the environmental protection, with a mean of 4.02.

Total Variance Explained

Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings (a)
Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
4,994	33,294	33,294	4,994	33,294	33,294	3,641
2,249	14,991	48,24	2,249	14,991	48,284	3,100
1,538	10,254	58,539	1,538	10,254	58,539	3,275
1,208	8,055	66,594	1,208	8,055	66,594	1,386
,965	6,437	73,030				
,782	5,215	78,245				
,710	4,736	82,981				
,636	4,238	87,219				
,465	3,103	90,323				
,395	2,632	92,954				
,303	2,019	94,974				
,260	1,736	96,710				
,221	1,475	98,185				
,166	1,107	99,22				
,106	,708	100,000				

Table 2. Extraction Method: Principal Component Analysis.

However, for the variables concerning specific CSR practices the mean is under the neutral point (mean=3). Particularly, the mean is 2.84 in the variable concerning the company's alliance with groups of interest, 2.86 the variable concerning environmental

philanthropic activities of the firm and a mean of 2.74 for the incentives that firm provides to the employees.

As concern the solar power variables the trend was to have answers that almost agree with the installation of photovoltaic cells if exist state subsidies and the full scale usage of the solar power with means 3.87 and 3.84 respectively. Relatively, exceeded the neutral point (mean = 3), the variables that referred in easier installation of photovoltaic systems in companies to residential areas (mean = 3.49), in benefits that companies would gain in long term from solar power adoption (mean = 3.40) and in positive contribution for the adoption of produced solar power's sell price in PPC (mean = 3.13)

At the opposite, the lowest means had the variables that referring to forming visit in a photovoltaic systems' company and the potential total substitution of conventional power with solar in corporations with means 2.03 and 2.46 respectively. Additionally, under the neutral point were the variables that referred in compulsory installation of photovoltaic systems for companies (mean = 2.78), in proactive development of photovoltaic systems in a firm with its own capital (mean = 2.81). The variable referred in respondents' past searching for information concerning solar power and photovoltaic systems, reached almost the neutral point (mean = 2.98).

Descriptive Statistics

	N	Min	Max	Mean	St. Deviation
Commitment to Environment	63	1	5	3,43	1,241
Obligation the Env. Protection	63	1	5	4,02	1,055
Alliance with Interest Groups	63	1	5	2,84	1,019
Environmental Philantr. Activities	63	1	5	2,86	1,105
Incentives to Employees	62	1	5	2,74	,974
Full-scale Usage	61	1	5	3,84	,860
Past Information	62	1	5	2,98	1,248
Past Visit in Phot. Company	62	1	5	2,03	1,130
Easier Installation for Companies	63	1	5	3,49	,948
Companies Take on Cost	63	1	5	2,81	1,162
Installation Compulsory for Comp.	63	1	5	2,78	1,170
Determination of Sale Price	63	1	5	3,13	1,143
Subsidy from State	63	1	5	3,87	1,198
Substitution of Conventional Power	63	1	5	2,46	,964
Beneficial in Long-term	63	1	5	3,40	1,056
Valid N (listwise)	59				

Table 3: Descriptive statistics

To examine if there is a statistically significant difference between the mean scores of the respondents in Thessaloniki and Kozani, independent-samples t-tests were conducted. The comparison of the self-esteem scores indicated that there was no significant difference in scores of Thessaloniki and Kozani for any continuous variable as the Sig. (2-tailed) value was above .05 in all cases. Conducting the independent – samples t-test with number of company's employees to be the categorical independent variable, there were some items statistically significant different (Sig.(2-tailed)< .05) (Pallant, 2006). In the comparison between small and medium size companies, according the number of employees, there was a significant difference in the variable concerning the substitution of conventional power (Sig. (2-tailed) = .032). For the same variable there was significant difference (Sig. (2-tailed) = .007) in the comparison between small and large companies, whereas there was no significant difference between the medium size and large companies.

3.2 Reasons that Enhance Solar Power Adoption

Social responsibility seems to be less associated with the reasons that motivate the adoption of solar power. For the analysis of the ranking of possible reasons that enhance solar power adoption, cross-tabulations have been created between the categorical variables of the reasons that help the solar power adoption from the one and the region or company's size from the other.

The respondents named as the most important reason the state subsidies with a total percentage of 57.1%. It should be mentioned that the state subsidy chose as most important reason the 61.4% of the corporations from the county of Thessaloniki and 47.4% of the corporations from the county of Kozani. Having on mind the company's size, the 58.5% of the small firms had the state's subsidy as their first choice of the reason that motivate solar power adoption, 60% of the medium size and 42.9% of the large companies.

As second important reason, the most preferences gathered the cost of oil with a total percentage of 34.5%. This was the choice of the 40% of the firms from the county of Thessaloniki whereas county of Kozani ranked as the second important reason the purchase of solar power from the Public Power Company (PPC). For the same categorical variable the cross-tabulation with the company's size indicated that the 31.6% of the small companies and the 45.5% of the medium size considered the cost of oil as second important reason, whereas the large ones chose the state subsidy with 50%.

In addition, third most important reason the respondents nominated the 'purchase of solar power from PPC' with a total percent of 38.9%. Analytically, the 43.6% of the firms from Thessaloniki had this choice whereas from Kozani as third important reason had the cost of oil with 46.7%. According to the company's size the percentage was 36.8% and 54.5% for the small and medium size companies respectively. Companies with more than 250 employees ranked as third important reason the environmental responsibility with 60%. The following figure (1) indicates the ranking of the reasons that enhance the adoption of solar energy according the data analysis.

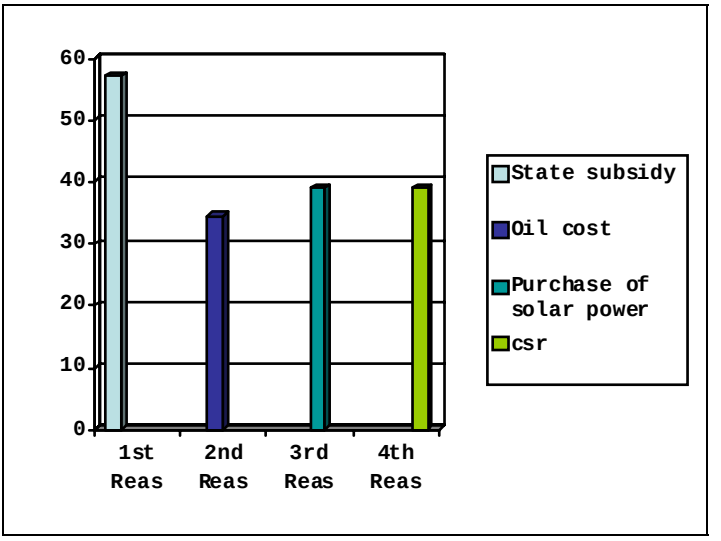


Figure 1. Reasons that motivate the adoption of solar power

Hence, the environmental responsibility was ranked fourth in the reasons that motivate the adoption of solar power with a total percent of 38.9%. The percentages according the region were 35.9% for Thessaloniki and 46.7% for Kozani. 42.1% of the small firms chose the environmental responsibility as fourth important reason. Relatively, the 36.4% of the medium size companies had the same choice, whereas the large firms ranked as fourth reason the purchase of solar power from PPC with a 60% percentage.

Finally, two respondents named other reasons as motivations for the adoption of solar power. Sunlight was the one and had been ranked as third in importance, whereas the cost of photovoltaic cells is the other and had been ranked fifth by the respondent that noted.

3.3 Negative Indicators in the Adoption of Solar Power

On the other hand an attempt became to determine the reasons of the slow adoption of solar power by Greek companies. It seems that the respondents considered a combination of variables as responsible for the limited utilization of solar power.

Analytically, the absence of information pleaded the 19% of the respondents, the sparse motivation the 17.5%, the cost of the photovoltaic cells the 11.1% and the low cost of the conventional power for the consumer the 9.5%. However, the majority with a percent of 38.1% chose the combination of 'all the above reasons' whereas a 4.8% declared that they don't knew (Figure 2). It should be mentioned that if someone examine the choices excluding the 'all the above option' the small size firms considered the sparse information as the main reason for the slow adoption with a

24.4%, the medium size companies then low cost of the conventional power with 20% and the large companies the sparse motivation with 28.6%.

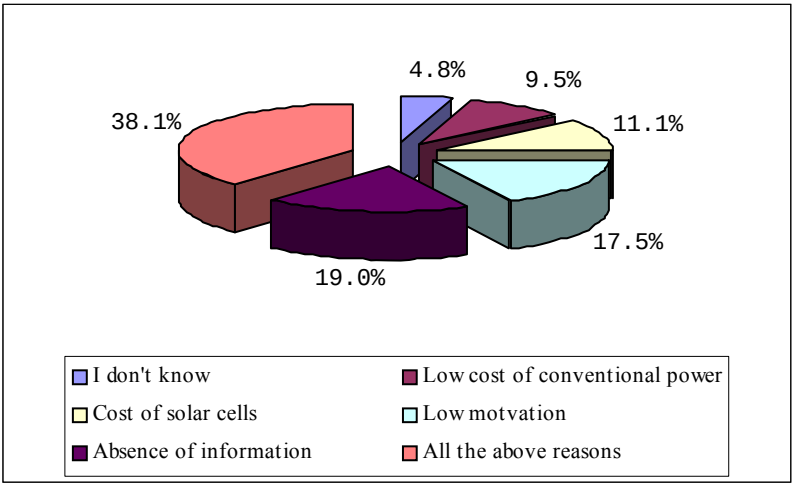


Figure 2. Indicators of the slow adoption of solar energy

4. Conclusion and Recommendations

Solar power produced by photovoltaic systems as an almost absolute clean energy source have become a comparable competitor of the conventional power produced from lignite, oil and gas, in many countries where the most corporations demonstrate a commitment to the environment. Applications of photovoltaic systems have become a part of the business practices that constitute corporate environmental responsibility (European Commission, 2007).

This paper examined the readiness of Greek companies to apply solar power as a part of their corporate social responsibility. Furthermore, it determined the compliance of Greek companies with the principles of environmental corporate social responsibility.

The findings from the questionnaires and interviews data analysis indicate that most enterprises associate the adoption of solar power with direct corporate benefits derived from the installation of photovoltaic systems. Hence, whereas they believe that is companies' obligation the protection of the environment, there will be long-term benefits from the adoption of solar photovoltaic systems, consider that the potential corporation's adoption of solar power could be achieved mainly through the state's subsidy. Furthermore, corporations appear to be reluctant to develop proactively photovoltaic systems only with their own capital. In order to apply solar power according to respondents' ranking, critical reason is the state subsidies, follows the cost of oil, the purchase of produced solar power from PPC and finally the environmental responsibility. Therefore, there is no significant correlation between the readiness of Greek enterprises to adopt solar power and their CSR performance.

On the other hand, negative indicator in the adoption of solar power is a combination of variants, including lack of information about solar energy form, lack of motivation, high cost of the photovoltaic systems and low purchase cost of conventional power. Especially, small size companies mention as main individual reason the lack of information, whereas the medium and large employees respectively, mention the lack of motivation as main reason.

The findings demonstrate that there is no statistically significant difference between the mean scores of the respondents in the two counties the questionnaires was distributed. Moreover, there is no statistically significant difference among the respondents according to the company's size. The only difference concerns the potential total substitution of the conventional power with solar between the small size companies, according the employee number, and the medium or large size.

The evaluation of the results should become, taking into account the absence of legislation for the renewable energy forms and the monopoly of PPC in power sector. The liberalisation of the power sector with the act 2773/1999 and especially, the new rule 13310/2007 that regulates the renewable energy market that was applied in 2008 is expected to renew the interest in renewable forms. However, it will be beneficial this law to provide incentives so as the companies to consume the energy that they produce and not only to sell it to the grid operator (PPC).

Another aspect of Greek market that should be considered is the large number of SMEs. For small companies is less attractive to spend time and money for the adoption of a CSR policy. According to World Business Council for Sustainable Development (1998) corporations are necessary to develop an integrated method to social responsibility and environmental management. Starting with the company's culture which must have a social responsible orientation, CSR should be diffused in the leadership, the employees – partners and should be proved continuously with positive actions. Therefore, many small companies hesitate to undertake environmental and general CSR management.

However, CSR approach reduces the negative externalities of business operations on the stakeholders. Especially, the environmental management contributes in the sustainable development significantly. The benefits are important for the companies too, as the literature review has indicated; hence responsibility and profits are not conflicting meanings. Greek companies should take advantage the sunlight of their country and substitute part of the internal consumption of conventional power with solar, contributed in the sustainable development of the country.

In the case of Greek SMEs, what is recommended is an extended public campaign to inform the managers about the benefits of the solar power adoption, even to a restricted extent. Moreover, it is essential to inform SMEs about the social and environmental practices that they have the possibility to demonstrate, as responsibility is expected not only from the large companies but also from them.

As far as the large firms are concerned, they dispose the means and the capital to implement an intensive use of solar power. The sooner the law 13310/2007 will be applied the better the promotion of solar and other renewable energy forms in the Greek market. Furthermore, the above could be the basis for an extensive and accomplished CSR performance.

Finally, it is suggested that this study could be repeated in three years. According the Rule 2001/77/EC, Greece is expected to reach the gross annual electricity consumption at 21% share of renewable energy in 2010 (European Union, 2004). Therefore, further research could identify more compliance with the principals of CSR from the corporations and different level of solar power adoption.

References

1. Bahaj, A. S (2002) 'Means of Enhancing and Promoting the Use of Solar Energy', *Renewable Energy*, 27:1, 97 - 105
2. Balabanis, G, Hugh, C. P, Jonathan, L (1998) 'Corporate social responsibility and economic performance in the top British companies: are they linked?', *European Business Review*, 98:1, 25 - 44
3. Becker-Olsen, L. K, Cudmore, B. A, Hill, R. P (2006) 'The Impact of Perceived Corporate Social Responsibility on Consumer Behavior', *Journal of Business Research*, 59, 46 – 53
4. Berry, M. A, Rondinelli, D. A (1998) 'Proactive corporate environment management: A new industrial revolution', *Academy of Management Executive,* 12:2, 38– 50
5. Caroll, A.B (2000) 'The four faces of corporate citizenship', in Richardson, J.E. (EDS) *Business Ethics 01/0*, Dushkin/McGraw-Hill. Guilford, CT, 187-91
6. Chapple, W, Morrison Paul, J.C, Harris, R (2005) 'Manufacturing and corporate environmental responsibility: cost implications of voluntary waste minimisation', *Structural Change and Economic Dynamics,*16, 347 – 373
7. Cruz, J. M (2008) 'Dynamics of Supply Chain Networks with Corporate Social Responsibility through Integrated Environmental Decision-making' *European Journal of Operational Research,* 184, 1005–1031
8. European Commission, Directorate- General for Energy and Transport (2004) *Electricity from Renewable Energy Sources: Encouraging Green Electricity in Europe*, ISBN 92-894-6666-9, Eur-Op, Luxembourg
9. European Commission, Directorate- General for Energy and Transport (2007) *Energy and Transport in Figures: 2007,* Eur-Op, Luxembourg
10. European Commission, Directorate- General for Energy and Transport (2007) *Renewables Make the Difference*, ISBN 978-92-79-06361-9, Eur-Op, Luxembourg
11. European Commission, Directorate – General for Energy and Transport (1998) *Photovoltaic solar energy : best practice stories,* Eur-Op, Brussels
12. European Commission, Directorate- General for Employment and Social Affairs (2001) *Green Paper: Promoting a European Framework for Corporate Social Responsibility*, ISBN 92-894-1478-2, Eur-Op, Luxembourg
13. European Environment Agency (1998) *Environmental Management Tools for SMEs*, ISBN 9291671061, Eur-Op, Luxembourg
14. Freeman, R.E. (1984) *Strategic Management: A Stakeholder Approach*, Pitman, Massachusetts
15. Haines, A, Smith, K. R, Anderson, D, Epstein, P. R, McMichael, A. J, Roberts, I, Wilkinson, P, Woodcock, J, Woods, J (2007) 'Policies for accelerating access to clean energy, improving health, advancing

development and mitigating climate change', *The Lancet*, 370:9594, 1264 – 1281
16. Hopkins, M. (2007) Corporate Social Responsibility and International Development: Is Business the Solution?, Earthscan, London
17. Kovacs, R (2006) 'Inderdisciplinary Bar for the Public Interest: What CSR and NGO Frameworks Contribute to the Public Relations of British and European activists', *Public Relations Review*, 32, 429 – 632
18. Kozanis Chamber of Commerce and Small Industries (2007) *List of Members: 2007*, KCCSI, Kozani
19. Lancaster, G (2005) Research Methods in Management: A Concise Introduction to Research in Management and Business Consultancy, Elsevier Butterworth – Heinemann, Oxford
20. Malhotra, N. K, Birks, D.F (2007) *Marketing research: an applied approach*, (3rd edn), Prentice Hall, Harlow
21. Mankiw, N.G (2001) *Economical principles/Volume A'*, Typothito – G.Dardanos, Athens
22. McWilliams, A, Siegel, D (2001) 'Corporate social responsibility: a theory of the firm perspective', *Academy of Management Review*, 26:1, 117-27.
23. Pallant, J (2007) *SPSS Survival Manual*, (3rd edn), McGraw – Hill, Maidenhead
24. Papasolomou-Doukakis, I, Krambia-Kapardis, M, Katsioloudes, M (2005) 'Corporate social responsibility: the way forward? Maybe not!: A preliminary study in Cyprus', *European Business Review*, 17:3, 263 - 279
25. Pearce, J.M (2006) 'Catalyzing mass production of solar photovoltaic cells using university driven green purchasing', *International Journal of Sustainability in Higher Education*, 7:4, 425 – 436
26. Public Power Company SA (2007) *Annual Report: 2007*, PPC, Athens
27. Reinhardt, F.L, Vietor, R. H. K (1996) *Business management and the natural environment*, South-Western College Publishing, Cincinnati, Ohio
28. Rondinelli, D. A, Berry, M. A (2000) 'Environmental citizenship in Multinational corporations: Social Responsibility and Sustainable Development', *European Management Journal*, 18:1, 70 – 84
29. Thessalonikis Chamber of Commerce and Industry (2007) *List of Members: 2007*, CICT, Thessaloniki
30. Thessalonikis Chamber of Small Industries (2007) *List of Members: 2007*, SICT, Thessaloniki
31. United Nations Conference on Trade and Development (1993) Environmental Management in Transnational Corporations: Report on the Benchmark Corporate Environmental Survey, ST/CTC/149. United Nations, New York
32. Wallner H. P (1999) 'Towards Sustainable Development of Industry: Networking, Complexity and Eco-clusters' *Journal of Cleaner Production*, 7, 49–58

Price Rigidities: Evidence from Slovakia

Katarina Lukacsy[1] *and Julius Horvath*[2]

Economics Department, Central European University, Nador u. 9, Budapest 1051 Hungary, [1]***lukacsy_katarina@phd.ceu.hu***, [2]*horvathj@ceu.hu*

This paper empirically estimates the price-setting patterns in the Slovak republic to provide missing evidence on price rigidities for a transition economy and to compare it with selected developed economies. We first examine the frequency and size of price changes to find that they vary across products remarkably: the frequency is largest for food products and smallest for services, price decreases are not uncommon and their size is comparable to that of price increases. We argue that unconditional means are insufficient, though, so we conduct survival time analysis where we condition the price change probabilities on various macroeconomic indicators to separate the true price rigidity from the default one. We find that once we clear the effects of the default price rigidities, the changes in probabilities of price changes become relatively comparable across product categories which is not the case with the popular unconditional means.

Keywords. price duration, frequency of price changes, competing risks, repeated multiple events, stratified gap-time Cox model

1. Motivation

Marketplace and its price system are *the* institutions of the free market economy and the ability of prices to adjust to changes in market conditions indicates efficiency of such economy. From the macroeconomic point of view, the ability of prices to adjust plays an important role in explaining the short-run monetary non-neutrality, hence it is important for studying and conducting the monetary policy. It is, therefore, crucial to understand how, why and when prices change, or alternatively, how, why and when prices do not change.

It is a known phenomenon that prices adjust with a delay to changes in economic conditions and several models have been set up to show that price rigidities determine the response of inflation and output to economic shocks which has implications for monetary policy. For instance, Rotemberg (1982) assumes positive costs of changing prices which introduce rigidities necessary for the existence of correlations between the output and nominal shocks thus making the rigidities central to the explanation of the business cycle. Hence, examining price rigidities promises shedding light on some macroeconomic issues and if examined and compared across countries, it could possibly play a role in the formulation and conduct of the euro-zone accession policy if significant heterogeneity proved among the candidate countries. However, as the focus

of this paper is mainly on the empirical part of the story, we do not go in deeper details of these models.

The empirical literature examining the extent of this phenomenon is not too extensive, even less so for the transition economies. The aim of this paper is to fill this gap and provide an assessment of the extent of price rigidities in a transition economy as well as its comparison with selected developed economies.

Originally, most of the papers dealing with price rigidity considered specific products and thus this research was limited. First studies in this area typically presented the reader with evidence that price stickiness is significant. For example Cecchetti (1986) presents evidence that magazine prices change once in 1.8-year period to more than ten-year period. Similarly, Kashyap (1995) reports that price changes of mail-order catalogue goods are unchanged for a period slightly more than a year. Bils and Klenow (2002) find that in about half of their data prices are unchanged for over four months.

What is more important about these studies, though, is that their results pointed to big differences in the frequency of price changes across different goods. Indeed, recently, several studies appeared which are better suited to judge the actual extent of price rigidity as they use large and complex sets of data, such as Bils and Klenow (2002), Dhyne et al. (2006) or Dhyne et al. (2005). In the context of transition economies this topic is dealt with in Gabriel and Reiff (2006). Similarly to these studies, the aim of this paper is primarily to characterize the average frequency and size of price changes at the store level in Slovakia.

However, we do not limit our study to this. We claim it is insufficient to analyze unconditional means as they do not differentiate between the default and the true rigidity in prices. The former captures stagnation of prices in case economic conditions also stagnate while the latter arises if prices fail to change in spite of economic conditions having changed. Indeed, firms are reluctant to adjust prices every time economic conditions change due to price adjustment costs such as input costs, implicit contracts, menu costs, etc. To capture merely this type of rigidity we use conditional methods.

To be more specific, we conduct survival time analysis attempting to explain the price change hazard rates by various macroeconomic variables such as wage, productivity, currency value, oil prices, market competition level, etc. The nature of our data gives rise to competing risks model with repeated events hence we employ a stratified gap-time Cox model and allow for variation clustered by products and by the type of a price change.

The structure of the paper is as follows: In Section 2 we describe our data set, in Section 3 we present our methodology and results on the unconditional frequency and size of price changes and compare them to those of the old euro-zone and U.S. As we claim that it is insufficient to examine such unconditional means, we then conduct conditional analysis. In Section 3 we present the reader with several macroeconomic indicators which various theoretic economic models identify as price level determinants and then use these as explanatory variables in survival time analysis. In the end we conclude.

2. Description of the Data

We use data set provided by the Slovak Statistical Office (SSO) which is used for computing the national consumer price index. It covers all the Slovak CPI products and spans from January 2002 till December 2007 with monthly frequency. From the total of 735 goods and services 113 are collected centrally so the main portion of prices is collected in a decentralized way. Each product is collected in several stores in each of the 38 major Slovak districts. This adds up to 17.937 stores altogether.

Each individual observation consists of information on the product code, the year and the month of observation, the store identifier and the price quote. We structure these data according to the Classification of Individual Consumption by Purpose (COICOP) as shown in Table 1. The reason for such classification is to avoid the assumption of all products being homogeneous in favour of assuming homogeneity only within a product category which is also justified by the results of Canetti et al. (1998) where the authors find systematic differences by industry but none among them.

	General Category (COICOOP group)	Weight in Total CPI	Number of Products	Number of Stores
1	Food and non-Alcohol Beverages	158.032	138	722
2	Alcohol Beverages and Tobacco	45.219	12	369
3	Clothing and Footwear	43.610	101	2,121
4	Housing, Water, Fuels	282.971	49	1,727
5	Furniture, House Equipment	54.189	93	2,665
6	Health Services	25.729	39	658
7	Transportation Services	94.818	81	1,493
8	Postal Service and Telecom.	37.152	22	145
9	Recreation and Culture	79.864	77	3,020
10	Education	15.134	7	473
11	Hotels and Restaurant	69.452	44	1,553
12	Other Goods and Services	93.830	72	2,991
	Total	1000.000	735	17,937

Table 1. Product categories in CPI

3. Static Unconditional Analysis of Price Changes

In this section we focus on simple descriptive statistics on the frequency and size of price changes as well as these decomposed to frequency and size of price increases and price decreases. The reason why we not only examine price changes in general but also decompose them is that prices may adjust asymmetrically, as has been shown for instance in Toolsema and Jacobs (2007) who find that the response of the mortgage rate is stronger if the cost increases than if it decreases.

We define that a price did not change in two consecutive periods if the store, the product code and the price quote are equal for two subsequent periods. A price spell is defined as the continuation of equal price quotes. In this data set SSO does not

distinguish a sale from regular price change so we consider any sale as a regular price change.

In what follows we present our results on five main indicators to characterize price setting in Slovakia: frequency of price changes, average duration of price spells and frequency and average size of price increases and decreases. We present these in Table 2 for each of the twelve categories separately and give a short commentary to each:

	COICOP category	Freq. of price changes	Dur. of price spells	Freq. of price increases	Freq. of price decreases	Share of price increases	Share of price decreases	Size of price increases	Size of price decreases
1	Food, non-Alc Beverages	38%	3.8	21%	17%	57%	43%	13%	11%
2	Alc Beverages, Tobacco	24%	6.1	14%	10%	52%	48%	8%	7%
3	Clothing, Footwear	18%	9.2	10%	8%	53%	47%	15%	12%
4	Housing, Water, Fuels	8%	17.5	6%	2%	71%	29%	20%	14%
5	Furniture, House Eqpment	16%	10.7	8%	8%	49%	51%	13%	10%
6	Health Services	14%	15.0	9%	5%	64%	36%	27%	15%
7	Transportation Services	36%	11.2	18%	17%	51%	49%	11%	16%
8	Postal Service, Telecom.	2%	34.4	1%	2%	8%	92%	61%	28%
9	Recreation, Culture	14%	17.3	7%	7%	57%	43%	17%	12%
10	Education	4%	30.0	3%	0%	76%	24%	52%	18%
11	Hotels, Restaurant	5%	23.7	4%	1%	69%	31%	18%	14%
12	Other Goods and Services	8%	21.8	5%	3%	60%	40%	35%	20%
	Total	17%	15.0	10%	7%	59%	41%	19%	14%

Table 2. Aggregate results

Frequency of price changes:

The highest frequency of price changes are observed for the category of *Food and non-alcoholic Beverages* where more than one third of prices change in a given month. One fourth of prices of *Alcoholic Beverages and Tobacco* change in a given month and *Clothing and Footwear* together with *Furniture and Household Equipment* experience less than one fifth of prices change in a given month. *Housing, Water, Fuels* , *Postal Service, Telecom* , together with *Education* , *Hotels, Restaurant* and *Other Goods and*

Services all exhibit one-digit price change frequency meaning prices are relatively rigid in these industries.

For comparison, in the euro-zone prices in the *Non-energy industrial* and *Services* sector change relatively rarely (the corresponding frequencies are 9% and 6% respectively) but values for the U.S. show smaller price rigidity there (22% and 15% respectively). Similarly, in the U.S. the food products exhibit higher probability of a price change in a given month (48% in *Unprocessed food* category and 27% in *Processed food* category) than is the case for the euro-zone (28% and 14% respectively) and Slovakia (around one third of all food prices change in a given month).

Average duration of price spells:

On average a price of *Food and non-alcoholic Beverages* lasts less than four month which means relatively flexible prices. At the same time, prices in *Alcoholic Beverages and Tobacco* last slightly longer than half-year and are followed by *Clothing and Footwear* where prices change every three quarters. In *Furniture and Household Equipment* and *Transportation Services* the prices remain unchanged for almost one year. In the rest of the categories prices exhibit strong rigidities as they change every 1.5 years or more. For comparison, the overall price duration in the euro-zone was 13 months and in the U.S. 6.7 months so prices in Slovakia, with the overall price spell duration of 15 months are relatively rigid.

Interesting to mention here are the results from Canetti et al. (1998). There, the authors also find that there are systematic differences by industry: services adjust most slowly, trade sector most rapidly. Additionally, after asking: How often do the prices of your most important products change in a typical year?, they find that the median is 1.4 times per year, with strong mode at 1 (annual price changes are by far the most often), 78% of the GDP is re-priced quarterly or less, 10% less than once per year and same amount more than weekly, more than 1% is re-priced more than once daily.

These results show that reporting pure means may not be sufficiently and it might be more interesting to examine rather the distributions. We address this point in this paper later.

Frequency of price increases and decreases:

Frequency of price increases and decreases: Generally, price increases are more frequent than decreases which goes in line with results of many past studies using data from various selected markets (gasoline, agriculture products, mortgage prices (Toolsema and Jacob, 2007; Peltzman, 1998). We observe relatively high frequency of price increases for the category of *Food and non-alcoholic Beverages*, *Alcoholic Beverages and Tobacco*, *Clothing and Footwear* and *Transportation Services*. Remarkably, the frequency of price decreases for these categories also reaches the highest values and is comparable in size with the frequency of price increases. Similar pattern can be observed for the low values of price increases. If we plot the two frequencies against one another, we indeed see that where price increases are frequent, price decreases are frequent, too and vice versa (Figure 1).

Average size of price increases and decreases:

The price increases are sizable compared to the prevailing inflation rate, which was 4.9% during the period in question. We observe the largest price increases for the category of *Postal services and Telecommunication* and *Education* where these average up to more than 50% of the price. Besides the category of *Alcoholic Beverages and Tobacco* all the categories exhibit double-digit values of sizes of price increases and the same is true for the sizes of price decreases. We can also notice, that where price increases are large, price decreases are large, too and vice versa. Similar pattern could be observed in case of frequencies, though in case of sizes, the correlation only seems to be strong for the lower values and for the higher values we see rather dispersed correlations (Figure 1).

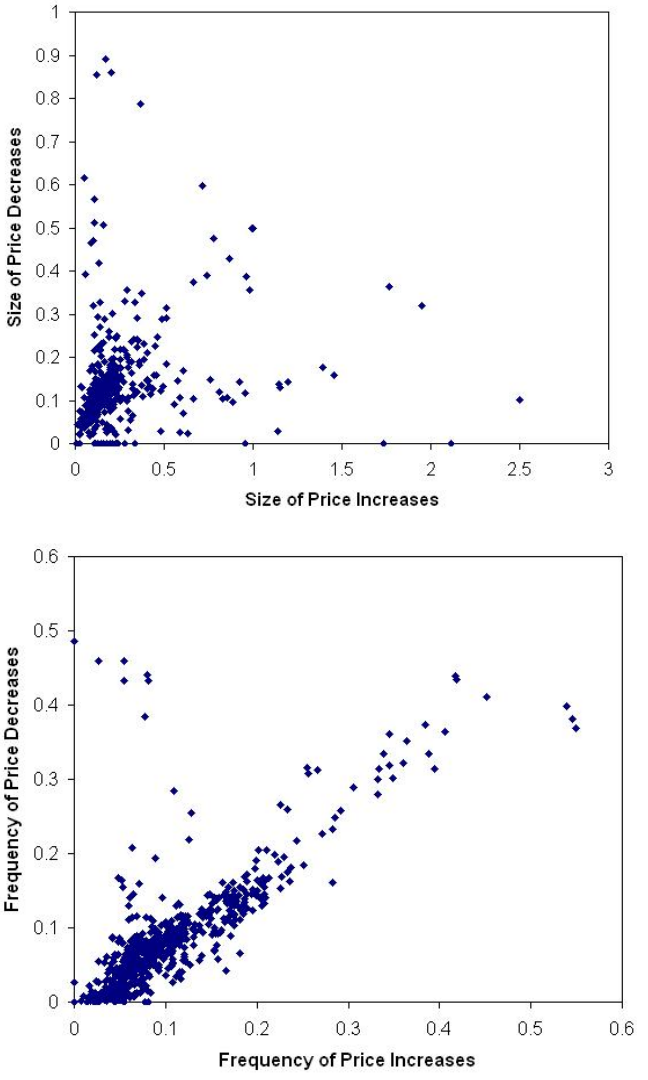


Figure 1. Correlation of frequencies and sizes of price increases vs. price decreases

In this section we examined different descriptive statistics of store-level prices in Slovakia. We found, that trade-sector products' prices go through more frequent adjustment than services do. We also saw that price decreases are not uncommon and even relatively sizable. What is more, we saw that the price changes' patterns are similar to those of some of the developed countries, though in absolute terms differences occur.

4. Dynamic Conditional Analysis of Price Changes

We believe it is insufficient to analyze unconditional means as presented in the previous section since they do not differentiate between the default and the true rigidity in prices.

The default rigidity arises when price remains unchanged after economic conditions have not changed while the true rigidity arises when prices fail to change in spite of economic conditions having changed. The default price rigidity is not truly interesting from the economics point of view while the true price rigidity is the one we like to focus on. Hence, we attempt to separate the two by conditioning the probabilities of price changes on various macroeconomic indicators which determine when, how often and by how much prices change. We try to identify these indicators in what follows and then use them as explanatory variables in survival time analysis regressions.

4.1. Price Level Determinants

Income, Wages, Energy/Oil prices, Search Costs, Productivity, Menu Costs and Currency Value are among the most trivial price level determinants. Among some additional are:

Managerial Thinking Costs: directly affect prices and were defined by e.g. Zbaracki et al. (2004)

Hierarchical Delays: are a kind of costs as defined by Hicks (2007).

Inflation: Although it is defined as the change in the price level, it still can have effects on the price changes through its lags or leads (expectations). Indeed, for example Ball and Mankiw (1994) built a model to give an explanation for prices to rise more speedily than fall with the driving mechanism being inflation. Similarly, Buckle and Carlson (1996) find evidence of price asymmetry systematically related to inflation. We conduct a short analysis to examine whether periods of different inflation rate exhibit differences in price changes' patterns. We find that inflation does seem to have influence on the probability of price changes as well as on their size. For details, refer to Appendix.

Price Lag: Rather than in its differences, price may also be a factor in its lag. Indeed, Eckstein (1971) used lagged prices as important explanatory variables of current prices even when all those variables which explain the desired prices were included.

Market Competition: Under perfect competition, prices are same for all the firms. With competitive markets prices are forced towards the average so that firms do not lose their shares of the market. As a result, an individual firm's price-setting strategy is an increasing function of the average price in the market. Indeed, Caglayan et al. (2006) shows that price dispersion is lower in more competitive environment. Baye et al. (2004) as well as Warin and Leiter (2007) find that price dispersion is greater when smaller number of firms list their prices on internet price-comparison site. Similarly, Gerardi and Shapiro (2007) find a negative effect of competition on price dispersion and Axarloglou (2007) finds that the likelihood of price adjustments is closely related to the intensity of market competition among price setters. We construct an index to

capture this effect as a ratio of the size of the small businesses in a region to the area of the region by which we construct a density-type of an index.

Deviation from the Average Price in the Region: Related to the above discussion, we hypothesize that at the store level the closest competitors' prices play a role in price-setting. This, naturally, can only be the case if the store is a price-taker. As we do not have any store-specific information, we assume all price setters to be price takers. We separate the deviation to positive and negative deviation as we believe the sign of the deviation matters since we analyze positive and negative price changes separately. We pose an assumption of quadratic functional form on this variable as we hypothesize that small deviations have smaller effect on the hazards than large deviations. This variable is only used in its first lagged value as we believe the price setters react to deviations with one period delay.

Size of the Last Price Change: We include this variable because Carlton (1986) finds a negative correlation between the frequency of price changes and the average absolute price change.

Time Elapsed since the Last Change: Price changes

GDP Gap: The choice of this variable comes due to the basic tenet of the sticky-price model that the higher the deviation of the output from the natural rate, the higher will be the deviation of the actual price from the expected.

In what follows we condition on these variables in the survival time analysis in order to separate the true price rigidities from the default ones. Not all the variables can be included, though, for reasons such as multicollinearity, unavailability or no reasonable proxy.

4.2. Survival Time Analysis – estimation of hazard rates

We now turn to dynamic conditional analysis of price changes in order to separate the true price rigidity from the default one. We conduct survival time analysis of the price spells with the aim to estimate the probabilities of price changes conditional on various macroeconomic indicators.

We proceed in our analysis separately for each product category, as they were defined above. In each product category we have $i = 1,2...n_k$ multiple-tuple observations (k = 1 , 2 .. . 12 denotes the product category index) consisting of:

• duration time of a spell t_i

• indicator c_i which takes on value 1 for uncensored observations and 0 for the censored observations.

• event type J (realized as j_i = 1,2 if price increase or price decrease respectively) which is the event corresponding to the i_{th} observation and j = 0 if the i_{th} observation is censored. (Left-censored observations are excluded from the analysis. As noted in [15] "If the censoring mechanism is random (i.e. the parameters of the censoring mechanism do not contain any information on the parameters of the duration model) discarding left-censored spells implies only a loss of efficiency in the estimation of the

parameters." We assume the censoring mechanism random and exclude the left-censored observations.)

- vector of covariates x_i which can depend on time.

Our case with two possible ways of termination of a spell (price increase or price decrease) gives rise to competing risks model. What is more, the repeated-event nature of our data gives rise to so called gap-time model in the group of Variance-Correction Models.

With T being continuous random variable representing survival time, the discrete-time parallel of the hazard-rate model with the cause-specific hazard rates can be written in the following form:

$$h_j(t_i, x_i) = Pr(T = t_i, J = j \mid T \geq t_i, x_i)$$

The two events, price increase and price decrease, are mutually exclusive so we have the probability of event j not occurring at duration t_i

$$Pr(T > t_i) = \prod_{s=1}^{t_i} \left(1 - \sum_{j=1}^{2} h_j(s)\right)$$

and the probability of event j occurring at duration t_i

$$Pr(T = t_i, J = j) = Pr(T = t_i, J = j \mid T \geq t_i, x_i) Pr(T = t_i) = \\ = h_j(t_i) \prod_{s=1}^{t_i - 1} \left(1 - \sum_{j=1}^{2} h_j(s)\right)$$

Using these and taking into account the two possible outcomes, the overall likelihood function then looks as follows:

$$L = \prod_{i=1}^{n_k} \prod_{j=1}^{2} [P_r(T = t_i, J = j)]^{I[j_i = j]} [P_r(T > t_i)]^{(1 - I[j_i = j])}$$

where j_i is the event type corresponding to the i_{th} observation and $I[.]$ is indicator function. Apparently, observations terminated in an event contribute to the overall likelihood function by probability equal to the first term and observations being (or considered) censored contribute by probability equal to the second term. The likelihood in terms of hazard rates can be then written as follows (with weights excluded for simplicity):

$$L = \prod_{i=1}^{n_k} \prod_{j=1}^{2} \left[h_j(t_i) \prod_{s=1}^{t_i-1} \left(1 - \sum_{j=1}^{z} h_j(s) \right) \right]^{I[j_i=j]} \left[\prod_{s=1}^{t_i} \left(1 - \sum_{j=1}^{z} h_j(s) \right) \right]^{(1-I[j_i=j])}$$

We employ the basic Cox model and we do so for one specific reason – the interpretation of the estimated coefficients is very intuitive and easy. Unlike most of the other duration models, here one can express the estimated coefficients as hazard rates readily to receive the percentage change in probability of an event corresponding to a change in an explanatory variable. We extend this hazard model by time-varying covariates in their concurrent values and their first lags:

$$\log h_{ij}(t_i, x_i) = \alpha_j(t) + \beta_{1j_i} x_i(t) + \beta_{2j_i} x_i(t-1)$$

where $x_i(t)$ are the covariates we condition on with $x_i(t-1)$ being their lags and $x_i(t) = x_i$ being time-invariant regressors. In this specification the regression constant $a_{ij}(t)$ represents the baseline hazard, i.e. the hazard corresponding to all covariates' values equal zero, $h_{0j}(t) = e^{aj(t)}$ and is different for each event type and product category. We stratify the variables *j* by event type and by products, i.e. observations corresponding to the same product and experiencing the same event type are assumed to be in the same stratum. Stratified estimates then yield equal coefficients across strata but with baseline hazard unique to each stratum.

4.3 Results

In Table 3 (Appendix) we present the results from this model for all the product categories. The numbers correspond to the percentage change in a probability of a price increase or a price decrease after a one-percent increase in a variable value (if the variable is an index, such as e.g. *market competition index* , the numbers correspond to the percentage change after a one-unit increase in its value).

We see that the estimated effects are generally strongly significant with insignificance occurring typically where regulated products are present, e.g. in *Education* , *Recreation and Culture* or *Postal Service and Telecommunication* . We can also notice that the sign of the effects is not typically opposite for the price increases and the price decreases which justifies the separation of the two – were the effects opposite, we could just as well analyze price changes altogether.

Generally, the magnitude of the effects is below 10% and not rarely too small to be reported (though still significant). Although this may seem too small, these values actually represent quite strong effect of the underlying explanators: Take for example *wages* to see that one-percent increase rises the probability of a price increase of food products by 0.1% and of a price decrease by 0.1%. This also means that if wages

double (a 100% increase), this probability goes up by 10% for price increases and down by 10% for price decreases so the estimated effects are in fact rather large.

This kind of interpretation cannot be applied to index-type variables such as *market competition index*. There the estimated effect corresponds to a one-unit increase so we have probability of a price change generally dropping by up to 5% as a result of market competition index growing by 10 units. For demonstration, this index ranges from 3.9 to 8.7 for seven out of eight regions and is equal 84.9 for the last one – the capital Bratislava region. This means that as for market competition prices are relatively variable in seven regions and are much more stable in the capital district.

One can also notice, that although we hypothesized quadratic effects of the deviation from the local average price, this typically did not seem to hold; either the quadratic term was too small or it was not significant. At the same time, though, we still estimate its linear effect as significant in most of the cases.

What is important to note about the size of the estimated effects, though, is that they are similar in their magnitude over the product categories. This is a result in odds with that coming from the unconditional analysis where we could see the probabilities of price changes varying much across the categories. Once we cleared the effects of the default price rigidities (those arising after no conditions have changed), we see that the changes in probabilities of price changes across product categories are relatively comparable. Hence, we have shown, that it is indeed not sufficient to derive conclusions from unconditional means only, as has been practice in most of the empirical literature.

5. Conclusions

In this paper we empirically estimated the price-setting patterns of final goods and services at the store level in the Slovak republic. We applied two approaches, unconditional means and conditional analysis. The reason to perform the first one was to be able to compare the price-setting patterns of a transition economy with those of developed economies as most of the existing literature on this topic employs this method. We applied the second approach - conditional analysis - in this paper to demonstrate that the former method was not sufficient.

We first obtained the following results: Frequency of price changes varies across products; one third of food prices change in a given month, while less than one fifth of prices change in a given month in the clothes, footwear and household industry. As for services, less than ten percent of prices change per month. This pattern is roughly the same in the old euro-zone and U.S. though prices are slightly more flexible there. Frequency of price increases is comparable to the frequency of price decreases and although price decreases are typically smaller than price increases, they exhibit the same order of magnitude.

All in all, we saw that the price changes' patterns in a transition economy are similar to those of some selected developed countries, though in absolute terms differences occur. This finding could provide valuable evidence on efficiency on certain macroeconomic policy issues of the euro-zone if we continue to watch the performance of the transition economy examined here since it just recently introduced Euro as its

common currency. Then we might be able to derive conclusions on price rigidities affecting monetary and macroeconomic performance of new entrants and possibly adjust the accession criteria accordingly.

After applying the second approach - conditional analysis – we indeed found that after we cleared the true price rigidity of the default one by conditioning the price change probability on various macroeconomic variables, we did not observe as big differences across the product categories as we did with the unconditional method. As a result we conclude that the popular unconditional methods provide biased results and should be substituted by conditional approaches.

References

1. Alvarez, Luis J.; Dhyne, Emmanuel; Hoeberichts, Marco M.; Kwapil, Claudia; Le Bihan, Herve and Patrick (*2005)* 'Sticky prices in the euro area - a summary of new micro *evidence',* Working Paper Series No. 563, European Central Bank.
2. Axarloglou, Kostas (2007) 'Thick markets, market competition and pricing dynamics: evidence from retailers' Journal of Managerial and Decision Economics, Vol. 28, No. 4, pp.669–677.
3. Ball, Laurence and Mankiw, N. Gregory (1994) 'Asymmetric price adjustment and economic fluctuations*'*, The Economic Journal, Vol. 104, No. 423, pp. 247-261.
4. Baye, Michael R.; Morgan, John and Scholten, Patrick (2004) 'Price dispersion in the small and in the large: Evidence from an internet price comparison site', The Journal of Industrial Economics, Vol. 52, No. 4, pp. 463-496.
5. Bils, Mark and Klenow, Peter J. (2002) 'Some evidence on the importance of sticky prices', NBER Working papers No. 9069, National Bureau of Economic Research.
6. Buckle, R. A. and Carlson, J. A. (1996) 'Inflation an asymmetric price adjustment', Series of Center for International Business Education and Research (CIBER), Purdue University, Krannert School of Management, No. 96-013.
7. Caglayan, Mustafa; Filiztekin, Alpay and Rauh, Michael T. (2006) 'Inflation, price dispersion, and market structure', Working Paper Series No. 2006-03, Indiana University, Kelley School of Business, Department of Business Economics and Public Policy.
8. Carlton, Dennis W. (1986) 'The rigidity of prices' Journal of American Economic Review , Vol. 76, No. 4., pp. 637–658.
9. Cecchetti, Stephen G. (1986) 'The frequency of price adjustment: A study of the newsstand prices of magazines.', Journal of Econometrics, Vol. 31, No. 3, pp. 255–274.
10. Dhyne, Emmanuel; Alvarez, Luis J.; Bihan, Herve Le; Veronese, Giovanni; Dias, Giovanni and Hoffm, Johannes (2006) 'Price changes in the euro area and the united states: Some facts from individual consumer price data' Journal of Economic Perspectives, Vol. 20, No. 2, pp. 171–192.

11. Eden, Ben and Campbell, Je (2004) 'Rigid prices: Evidence from us scanner data', Working Paper Series No. 461, Society for Economic Dynamics.
12. Gabriel, Peter and Reiff, Adam (2006) 'The effect of the change in vat rates on the consumer price index', Journal of MNB Bulletin, Vol. 1, No. 2, pp. 14–20.
13. Gerardi, Kris and Shapiro, Adam Hale (2007) 'The effects of competition on price dispersion in the airline industry: a panel analysis', Working Paper Series No. 07-7, Federal Reserve Bank of Boston.
14. Giannelli, Gianna Claudia (1996) 'Women's transitions in the labour market: A competing risks analysis on german panel data', Journal of Population Economics, Vol. 9, No. 3, pp. 287–300.
15. Michael J. Hicks (2007) 'Hierarchical delays as a source of nominal price rigidities: evidence from the microcomputer industry.' Journal of Managerial and Decision Economics, Vol. 28, No. 7, pp. 803–815.
16. Kalbfleisch, J. D. and Prentice, R. L. (2002) 'The statistical analysis of failure time data' New York: Wiley, 2nd Edition.
17. Kashyap, Anil K. (1995) 'Sticky prices: New evidence from retail catalogs', The Quarterly Journal of Economics, Vol. 110, No. 1, pp. 245–74.
18. Eckstein, Otto (1971) 'Industry price equations' Discussion paper number 158, Harvard Institute of Economic Research.
19. Peltzman, Sam (2000) 'Prices rise faster than they fall.' Journal of Political Economics, Vol. 108, No. 3, pp. 466-502.
20. Rotemberg, Julio J. (1982) 'Monopolistic price adjustment and aggregate output', Journal of Review of Economic Studies, Vol. 49, No. 4, pp. 517–531.
21. Toolsema, Linda A. and Jacobs, Jan P. A. M. (2007) 'Why do prices rise faster than they fall? With an application to mortgage rates', Journal of Managerial and Decision Economics, Vol. 28, No. 7, pp. 701– 712.
22. Warin, Thierry and Leiter, Daniel B. (2007) 'An empirical study of price dispersion in homogenous goods markets', Working Paper Series No. 0710, Middlebury College, Department of Economics.
23. Zbaracki, Mark; Ritson, Mark; Levy, Daniel; Dutta, Shantanu and Bergen, Mark (2004) 'Managerial and customer costs of price adjustment: Direct evidence from industrial markets', Journal of The Review of Economics and Statistics, Vol. 86, No. 2, pp. 514-533.

Appendix

A. Discussions on Inflation

In this section we examine differences between price changes' patterns for high-inflation period as opposed to low-inflation period. Indeed, as can be seen in Figure 2 where we plot inflation for the U.S., euro-zone and Slovakia for the period of our analysis, inflation was much higher in Slovakia in the first half of the period and decreased to the level of the developed economies in the second half of the period.

We divide our sample into two parts: i) first half exhibiting higher values of inflation ii) second half exhibiting stable low values of inflation. We then recalculate the

frequencies and the sizes of price changes for each part separately to find that price changes are typically more frequent during the period when inflation is high. Refer to Table 4.

We find results consistent with our expectations: inflation seems to have influence on the probability of price increases and decreases just as well as on the size of the price changes. Therefore we find it justified to include inflation among the price level determinants. However, since inflation in a small open economy such as Slovakia is determined by such factors as oil prices, currency value or productivity, we include those instead.

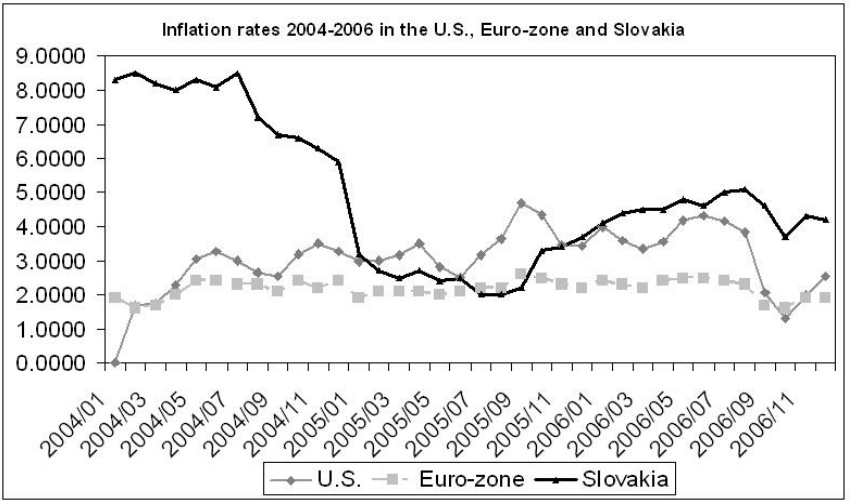


Figure 2. Inflation Rates 2004-2006 in the U.S., Euro-zone and Slovakia

COICOP category	1. Food, non Alcoholic Beverages		2. Alcoholic Beverages, Tobacco		3. Clothing, Footwear		4. Housing, Water, Fuels	
explanatory variables:	PrInc	PrDec	PrInc	PrDec	PrInc	PrDec	PrInc	PrDec
Wage	0.1***	-0.1***	-0.1**	-0.2**	-0.2***	-0.7***	0.9***	0.5***
Wage(-1)	-0.3***	-0.2***	-0.2***	0.2***	0.8***	1.0***	-0.8***	0.7***
Oil Prices	-≪0*	-≪0***	-≪0	-≪0	0.1***	≪0	0.3***	0.1
Oil Prices(-1)	+≪0***	-≪0**	0.1**	-≪0	-0.3***	-0.3***	-0.1***	-0.2***
Currency value	1.7***	0.9***	0.1	1.6***	4.0***	4.3***	4.1***	5.7***
Currency value(-1)	-1.4***	-0.5***	0.3	0.3	-0.6	-1.1***	-0.4	-0.8
Productivity	-3.3***	-2.6***	-2.1***	-3.6***	0.2	2.5***	4.7***	-5.5***
Productivity(-1)	2.8***	2.8***	1.9***	5.2***	1.8***	0.7	-6.1***	6.6***
GDP gap	2.6***	2.0***	2.1***	4.1***	0.3	-2.0***	-6.1***	6.5***
GDP gap(-1)	0.4***	-0.4***	0.1	-0.5**	-≪0	-0.6***	-1.1***	0.4
Last Price Adjst	-0.2***	+≪0	-1.4***	-≪0	≪0	-0.1***	-0.1***	≪0
Price(-1)	-0.1***	-0.1***	-0.4***	-0.9***	–≪0***	-≪0***	-≪0**	-≪0
Market Compet	-0.1***	-0.2***	-0.3***	-0.2***	-0.4***	-0.3***	-0.2***	-0.1**
NegDev frm Avg(-1)	0.3***	0.3***	0.8***	0.5***	-≪0	≪0	-0.2***	-0.6***
NegDev frm Avg(-1)Sqr	-≪0***	-≪0***	-≪0***	-≪0***	≪0**	≪0	≪0***	≪0***
PozDev frm Avg(-1)	-0.3***	-0.3***	1.1***	1.3***	0.1	-0.3**	0.7***	-0.7***
PozDev frm Avg(-1)Sqr	+≪0***	+≪0**	-≪0**	-≪0	≪0***	≪0***	-≪0	≪0**
EU entrance	7.0***	-0.6	-16.7***	-9.2***	-10.9***	-8.7**	1.0	-5.8
Tax-Law Passed	2.1**	-0.3	13.4***	14.0***	-18.2***	-26.2***	-13.9***	13.7*
Tax-Law Vetod	1.6	-1.8	8.8**	17.0***	-3.5	-4.4	-14.0**	-0.9

COICOP category	5. Furniture, House Equipment		6. Health Services		7. Transport. Services		8. Postal Service, Telecom.	
explanatory variables	PrInc	PrDec	PrInc	PrDec	PrInc	PrDec	PrInc	PrDec
Wage	0.4***	0.4***	1.3***	0.7***	0.3***	0.1***	0.4	-0.1
Wage(-1)	-0.4***	-0.1	-1.9***	-2.6***	-0.4***	0.3***	-1.0	-0.5
Oil Prices	0.2***	0.1***	0.5***	-≪0	0.2***	+≪0	-1.2***	-0.2
Oil Prices(-1)	-0.2***	-0.2***	-≪0	0.1	-0.1***	+≪0	1.8***	0.4**
Currency value	1.2***	0.7*	-3.2***	-2.5**	2.5***	0.6***	4.3	-1.4
Currency value(-1)	1.1***	1.6***	1.3	0.1	-1.6***	0.2	-2.8	3.2
Productivity	0.5	-1.5**	-5.6***	3.7***	0.1	-8.0***	1.8	-1.8
Productivity(-1)	0.3	2.8***	1.6*	-4.4***	0.7**	8.8***	-3.8	2.6
GDP gap	0.3	2.3***	4.9***	-4.4***	-0.2	8.0***	-5.5	0.7
GDP gap(-1)	0.1	-0.6***	0.2	0.9*	-0.8***	0.6***	1.8	-0.4
Last Price Adjst	-≪0	0.1***	-0.3***	-0.1***	-0.4***	-0.6***	0.7*	0.3*
Price(-1)	-≪0***	-≪0***	≪0***	+≪0**	+≪0***	+≪0***	-≪0***	+≪0
Market Compet	-0.4***	-0.3***	-0.3***	-0.3***	+≪0***	-≪0	0.1	-0.2
NegDev frm Avg(-1)	-≪0	+≪0	-0.4***	0.5*	-1.2***	-1.3***	0.8	+≪0
NegDev frm Avg(-1)Sqr	+≪0	+≪0	≪0*	-≪0*	+≪0***	+≪0***	-≪0	+≪0
PozDev frm Avg(-1)	-0.2	-0.8***	1.5***	0.1	-3.0***	-3.2***	0.8	-1.8
PozDev frm Avg(-1)Sqr	-≪0	-≪0	-≪0***	-≪0	+≪0***	+≪0***	-0.1	+≪0
EU entrance	-5.8	5.0	-20.4***	0.5	1.8	-5.7	86.1	6.1
Tax-Law Passed	-0.6	-13.2***	-15.0**	-55.2***	-13.3***	13.6***	57.0	31.2
Tax-Law Vetod	-4.4	-4.6	-41.8***	-59.9***	-10.3**	12.4***	144.8	15.2

COICOP category	9. Recreation, Culture		10. Education		11. Hotels and Restaurant		12. Other Goods, Services	
explanatory variables	PrInc	PrDec	PrInc	PrDec	PrInc	PrDec	PrInc	PrDec
Wage	0.3***	0.2**	-2.6**	-7.1*	-0.1	-1.4***	0.4***	0.3***
Wage(-1)	-0.4***	-0.4***	0.8	0.8	0.3	-0.3	0.1	$+\ll 0$
Oil Prices	0.2***	0.2***	0.5	0.8	-0.1**	0.4***	0.1**	$+\ll 0$
Oil Prices(-1)	-0.2***	-0.2***	-0.4	-0.7	-0.2***	-0.2**	-0.2***	-0.2***
Currency value	1.1**	0.4	20.1**	-6.2	4.5***	-2.6*	1.7***	2.3***
Currency value(-1)	1.8***	1.4***	-7.6	9.9	1.3**	8.6***	0.7*	0.9*
Productivity	0.4	1.1*	7.4	-15.0	-10.9***	-13.7***	-2.0***	-1.1
Productivity(-1)	$-\ll 0$	-0.3	-4.2	20.1	14.9***	15.5***	2.3***	2.7***
GDP gap	-0.1	-0.4	-14.2	11.0	10.5***	18.2***	2.5***	1.3*
GDP gap(-1)	-0.5*	-0.1	-16.9***	-3.3	-1.4***	2.9***	-0.6**	-0.9***
Last Price Adjst	$-\ll 0$	$+\ll 0$	-0.2**	-0.7**	-0.1***	-0.3***	-0.6***	-0.1***
Price(-1)	$-\ll 0$***	$-\ll 0$***	$+\ll 0$*	$-\ll 0$	0.1***	$+\ll 0$**	$+\ll 0$***	$-\ll 0$
Market Compet	-0.2***	-0.3***	-0.4**	-1.0	-0.6***	-0.5***	-0.3***	-0.2***
NegDev frm Avg(-1)	$+\ll 0$	0.1	-0.2	-2.7	-0.2***	0.5**	$-\ll 0$	0.1
NegDev frm Avg(-1)Sqr	$-\ll 0$	$-\ll 0$	$-\ll 0$	$+\ll 0$	$+\ll 0$	$-\ll 0$	0	$-\ll 0$
PozDev frm Avg(-1)	0.5***	-0.2	$-\ll 0$	-1.1	-0.7***	-0.9**	-0.2	-0.4***
PozDev frm Avg(-1)Sqr	$-\ll 0$	$-\ll 0$**	$+\ll 0$	$+\ll 0$	$+\ll 0$**	$+\ll 0$**	$+\ll 0$	$-\ll 0$
EU entrance	-5.0	-7.0*	40.89***	–	5.9	-32.3**	0.1	13.6***
Tax-Law Passed	6.2*	-4.1	-38.8	54.8	33.1***	28.3**	-2.7	-6.7*
Tax-Law Vetod	4.7	0.3	–	–	26.2**	58.8***	10.3**	-9.9*

***denotes significance at 99%, **denotes significance at 95%, *denotes significance at 90%

Table 3. Gap-time Cox Model – Price Increases and Decreases

	COICOP category	Frequency of price changes		Duration of price spells		Frequency of price increases		Frequency of price decreases	
		1	2	1	2	1	2	1	2
1	Food, non-Alc Beverages	40%	36%	3.8	4.7	22%	19%	18%	16%
2	Alc Beverages, Tobacco	30%	20%	4.9	5.7	18%	12%	12%	8%
3	Clothing, Footwear	19%	18%	8.6	9.2	11%	9%	8%	8%
4	Housing, Water, Fuels	4%	8%	10.8	17.8	3%	6%	1%	2%
5	Furniture, House Eqpment	17%	16%	10.4	11.0	9%	8%	8%	8%
6	Health Services	11%	15%	16.5	12.8	8%	8%	2%	7%
7	Transportation Services	9%	26%	9.6	12.4	14%	21%	10%	19%
8	Postal Service, Telecom.	1%	2%	1	33	0%	1%	0%	2%
9	Recreation, Culture	13%	14%	9.7	16.5	7%	7%	6%	7%
10	Education	5%	3%	21.2	26.7	4%	3%	1%	0%
11	Hotels, Restaurant	6%	4%	20.0	22.6	5%	3%	1%	1%
12	Other Goods and Services	11%	8%	14.6	21.5	7%	5%	4%	3%
	Total	15%	15%	9.8	15.0	10%	10%	7%	7%

	COICOP category	Share of price increases		Share of price decreases		Size of price increases		Size of price decreases	
		1	2	1	2	1	2	1	2
1	Food, non-Alc Beverages	58%	56%	42%	44%	12%	14%	10%	12%
2	Alc Beverages, Tobacco	62%	38%	38%	62%	8%	8%	7%	8%
3	Clothing, Footwear	55%	47%	45%	53%	13%	17%	11%	12%
4	Housing, Water, Fuels	43%	68%	57%	32%	23%	16%	11%	14%
5	Furniture, House Eqpment	50%	45%	50%	55%	10%	16%	10%	11%
6	Health Services	75%	53%	25%	47%	18%	32%	10%	16%
7	Transportation Services	48%	45%	52%	55%	8%	13%	6%	18%
8	Postal Service, Telecom.	0%	8%	100%	92%	20%	62%	11%	30%
9	Recreation, Culture	45%	53%	55%	47%	14%	20%	11%	13%
10	Education	74%	58%	26%	42%	64%	36%	14%	17%
11	Hotels, Restaurant	80%	49%	20%	51%	17%	19%	14%	13%
12	Other Goods and Services	66%	52%	34%	48%	28%	30%	9%	21%
	Total	52%	54%	48%	46%	16%	18%	10%	14%

Table 4. Aggregate Rresults after Accounting for Inflation

Development of a Green Supply Chain in the South Eastern Europe

Ioannis Mallidis

Department, Industrial Management Division, Aristotle University of Thessaloniki, Mechanical Engineering, 54124 Thessaloniki, Greece, mallidi@auth.gr

Southeastern Europe is an emerging market where several multinational retailers wish to enter. The reason is the increasing number of manufacturing facilities (e.g. Bulgaria, Romania and FYROM exhibit growth rates of 6%, 7% and 8% respectively) that has resulted in an increase of the GDP per capita and thus an increase of consuming capabilities. However, the induction of Bulgaria and Romania in the EU has changed the traditionally lenient regulatory environment of the area and thus, forced the supply chain regulators to take into serious consideration apart from the cost and time efficiencies, the green restrictions and performance indicators set by the European Union. The purpose of this paper is to study the current situation regarding alternative multi-modal routes for supplying the markets of Bulgaria and Romania and the associated cost, time and emissions for cargo transportation. Specifically, the paper proposes a decision making system for the selection of the optimum transportation network based on a scoring model that incorporates multiple optimization criteria (time, cost, CO_2 emissions).

Keywords. Supply chain network design, CO_2 emissions, green logistics

1. Introduction

The traditional supply chain strategies aim to organize cost and time effective supply chain configurations without taking in consideration environmental factors. Green practices are imposed by governments of the developed countries aiming to force private companies to adopt environmentally friendly operational strategies. Specifically in the European Union, these practices are organized under a common legislation framework and if not applied high cost penalties will be forced. Such penalties affect significantly the traditional perspective of time and cost effective supply chains creating the need for environmentally friendly practices. Supply chain operational policies that take in consideration environmental factors are extremely necessary today since the transport sector and specifically that of freight transportation is responsible for 20%-30% of total CO_2 emissions (Nakata, 2004). CO_2 emissions constitute the most important factor behind climate change which is almost invariably considered an issue of global interest. The influential Stern report estimates that

climate change may cause a permanent decrease in annual global GDP of between 5%-20% and thus actions must be taken (Koetse and Rietveld, 2008). Within the European Union, emission standards for new vans and trucks are getting stricter from Euro I in 1992 to Euro III, IV and V specifications, decreasing significally the permitted CO_2 and NO_x emissions in the atmosphere. These standards are only applied by the developed European Union countries. But what about the countries that have recently entered the European Union? A company that has invested or aims to do so in these countries must take into serious consideration green practices in order to meet the European Unions' standards. Inefficient intermodal transportation networks combined with poor rail and road infrastructures make the efforts for greening supply chains a real challenge. How can companies overcome these inefficiencies? How can they achieve to reduce their costs while at the same time improve their image to their customers through environmentally friendly policies? The answer lies behind the concept of optimization. Optimized operational policies applied in supply chain configurations can result in a significant decrease of CO_2 emissions and thus negative effects to the environment, by reducing the distances traveled and the frequency of freight consignments.

2. Literature Review

There is a vast number of papers that aim to propose environmentally friendly operational policies in production, transportation, distribution and reverse flows of products at the end of their lifecycle. In the majority, the environmental perspective of these papers is expressed through the effort of the minimization of distances travelled and thus the reduction of energy demands and CO_2 emitted in the atmosphere. In this sector you can find publications that aim to optimize transportation networks (Yu et al., 2008), quantify CO_2 emissions in the transport sector (Azar, 1996; McKinnon, 2006) and optimize reverse logistic networks and locations of Distribution Centers (Daskin et al., 2002).

3. System description

We assume that a major retailer company aims to expand its activities to the Balkans. The company imports white goods from Shanghai and supplies its two retail stores in Sofia (BG) and Bucharest (RO). It therefore tries to identify a supply chain configuration that is cost and time efficient while at the same time minimizes CO_2 emissions in order to meet EU standards and thus avoid penalties. The port of Shanghai constitutes the port of origin. The company has to choose between three ports of destination (Thessaloniki, Varna and Constanta) as a starting point of its supply chain network. Each port constitutes a start point of a different supply chain configuration. In other words the company has to choose between three supply chain configurations that correspond to each port of destination. The container is transported by ship from Shanghai to the port of destination and then to the central distribution center where it is deconsolidated. The retail stores set the orders and the products are distributed accordingly from the central distribution center. Each container order lot size is 300cntrs from Shanghai to the central DC. The orders set by the retail stores are

100cntrs from Sofia's retail store and 200cntrs from Bucharest retail stores. *The first step* is the calculation of the total time required to transport a container from the port of origin to the company's retail store. *The next step* will be the calculation of the total costs of the scenarios identified taking in consideration discounts based on quantities of containers transported by ship and rail. *The third step* will be the calculation of tones of CO_2 emitted from transportation means for each scenario. *The final step* will be the application of the scoring model to select the overall best *scenario. (In all scenarios distances are calculated with the use of Sonny map route planner and Google map).*

The first scenario examined, includes Thessalonikis container terminal as a start point of the supply chain configuration. The container is transported from Shanghai to Thessaloniki container terminal, loaded on wagon and then transported to the company's central distribution center in Lom. The container is then deconsolidated in the central DC where the picking procedure starts and an exit container is consolidated based on the orders placed by the retail stores. In order to supply Sofia's retail store the container is loaded on truck and transported from Lom central DC to Sofia's retail store. In order to supply Bucharest's retail store, the container is loaded on truck which used to transport it from Lom central DC to Lom inland terminal while gantry cranes are used to load the container on the barge ship. The container is then transported from Lom inland terminal to Giurgiu free zone port, loaded on truck by the terminals gantry cranes and transported to Bucharest retail store.

The second scenario examined, includes Varna container terminal as a start point of the supply chain configuration. The container is transported from Shanghai to Varna container terminal by ship and then transshipped on barge in order to be transported to Ruse inland terminal. The container is then loaded on truck by the terminals gantry cranes and transported to Ruse central DC. In order to supply Sofia's retail store the container is loaded on truck at the central DC and transported at Ruse inland terminal, while the terminals gantry cranes are used to load the container from the truck on the barge ship. The container is transported from Ruse inland terminal to Lom inland terminal, unloaded on truck by the terminals gantry cranes and transported to Sofia's retail store. In order to supply Bucharest's retail store, the container is loaded on truck from Ruse central DC and transported to the retail store in Bucharest.

The third scenario examined, includes Constanta container terminal as a start point of the supply chain configuration. The container is transported from Shanghai to Constanta container terminal by ship and then transshipped on barge in order to be transported to Giurgiu inland terminal. The container is then loaded on trucks by the terminals gantry cranes and transported to Giurgiu central DC. In order to supply Sofia's retail store the container is loaded on truck at the central DC and transported at Giurgiu inland terminal while the terminals gantry cranes are used to load the container from the truck on the barge ship. The container is transported from Giurgiu inland terminal to Lom inland terminal, unloaded on truck by the terminals gantry cranes and transported to Sofia retail store. In order to supply Bucharest's retail store, the container is loaded on truck from Giurgiu central DC and transported to the retail store in Bucharest.

3.1 Transportation Time

The time required to transport a container from Shangai to the retail store in each scenario is calculated based on the voyage time of each transportation mode, and the time spent at the inland and sea-ports. The voyage time is calculated based on the average speed of the transportation mode utilized in the case studies and the total kilometers traveled, while the residence time at the port is calculated based on the sea and inland container terminals loading and unloading capabilities without taking in consideration the ships residence time waiting to enter the terminals. The travel time by ship from Shanghai to Thessalloniki, Varna, Constanta, is 25 days (or 600h), 27 days (or 648 hours) and 29 days (or 696 hours) respectively. In Scenarios 2&3 we assume that the containers are discharged on the quay and loaded at the barge ship after 24 hours, with the corresponding loading rate of each terminal. The total time required to transport the container from Shanghai to each countries retail stores is indicated by the tables below:

	Travel times			
Route	Transportation mode	Average speed (km/h)	Distance (Km)	Travel time (h)
Shanghai-Thessaloniki	-	-	-	600.0
Th.Pa–Lom central DC	Rail	45	425.7	9.5
Lom central DC - Sofia retail store	Truck	60	147.0	2.5
Lom inland terminal - Giurgiu inland terminal	Barge	24	266.0	11.1
Giurgiu inland terminal- Bucharest retail store	Truck	60	64.7	1.1
Total travel time				624.2
	Residence times			
Terminal	Loading/Unloading rates (TEU/h)	no of containers		Residence time (h)
Th.Pa	75	300		4.0
Lom inland terminal (loading)	20	200		10.0
Giurgiu inland terminal (discharge)	20	200		10.0
Total travel time + Residence time				648.2

Table 1. Travel and Residence times in scenario 1

	Travel times			
Route	Transportation mode	Average speed (km/h)	Distance (Km)	Travel time (h)
Shanghai-Varna	-	-	-	648
Varna - Ruse centralDC	Barge	24	191.3	6.7
Ruse central DC - Lom inland terminal	Barge	24	288.5	12.0
Lom inland terminal - Sofia retail store	Truck	60	147.0	2.5
Ruse central DC - Bucharest retail store	Truck	60	74.3	1.2
Total travel time				670.4
	Residence times			
Terminal	Loading/Unloading rates (TEU/h)	no of containers		Residence time (h)
Varna container terminal (transhipment on barge)	50	300		6
Ruse inland terminal (discharge)	20	300		15
Ruse inland terminal (loading)	20	100		5
Lom inland terminal (discharge)	20	100		5
Waiting time at the terminal for transhipment				24.0
Total travel time + Residence time				725.4

Table 2. Travel and Residence times in scenario 2

	Travel times			
Route	Transportation mode	Average speed (km/h)	Distance (Km)	Travel time (h)
Shanghai-Constanta	-	-	-	696.0
Costanta - Giurgiu central DC	Barge	24	255.8	10.7
Giurgiu central DC - Lom inland terminal	Barge	24	274.1	11.4
Lom inland terminal – Sofia retail store	Truck	60	147.0	2.5
Giurgiu central DC - Bucharest retail store	Truck	60	64.7	1.1
Total travel time				721.7
	Residence times			
Terminal	Loading/Unloading rates (TEU/h)	no of containers		Residence time (h)
Costanta container terminal (transhipment on barge)	100	300		3
Giurgiu Inland terminal (Discharging)	20	300		15
Giurgiu Inland terminal (Loading)	20		100	5
Lom inland terminal (Discharging)	20		100	5
Waiting time at the terminal for transhipment				24
Total travel time + Residence time				773.7

Table 3. Travel and Residence times in scenario 3

Results

The total time required to supply both stores in scenario 1 is *648.2 hours,* in scenario 2, *725.4* hours while in scenario 3 the total time required is 773.7 hours.

Interpretation of results

It can be clearly stated that the high time differences between scenario 1 and scenarios 2 and 3 are due to, the respectively shorter shipping distances from Shanghai, combined with the use of rail transportation from Thessaloniki's container terminal to

the company's central DC. Of course the distances involved in the transportation networks are relatively short and thus do not affect the total travel time's significally. It is possible though, that in very long distances the time differences would increase significally due to the high speed differences between barge, trucks, rail (24, 45, 60 km/h respectively). Another very important factor that may affect negatively the time performance of the supply chains examined is closely related to inland and sea port delays due to strikes or malfunctions of cranes and terminal vehicles, bad road conditions and old trucks, bad rail conditions and wagons. Since data cannot be retrieved in this point we will not consider the above parameters while testing the time performance of the supply chain configurations examined.

3.2 Transportation Cost

The supply chain networks consist of a number of distribution centers, container terminals and retail stores. The container terminals and the central distribution centers as also the central distribution centers and the retail stores are connected via transport links. The transport links are either shipping links (barge transportation) or trucking links (land transport using trucks) or rail links. The costs of these scenarios involve the shipping costs from Shanghai to the port of destination, the free out costs at the port of destination, the lift on costs on the selected transportation modes at the port of destination, the transportation cost from the port of destination to the central distribution center and the transportation cost from the central distribution center to the retail stores. The free out costs per container at Thessaloniki's container terminal are approximately 160€ and 100€ in Varna and Constanta container terminal (Th.PA IT Department). The lift on wagon costs in Thessaloniki's container terminal are assumed to be 30€ while the transshipment costs in the terminals of Varna and Constanta are assumed to be 50€ extra per 20 foot container. Truck costs are calculated based on a logarithmic equation that derives from truck offers given from various transportation agencies.

$F(x) = 293,53\ln(x) - 1059,9$ Where x represents the kilometers traveled.

The rail costs are calculated again based on the below regression equations that derives from offers per km given by the Greek railway organization. For every 60TEU or 30 wagons we have a block train which results in a 5% discount. We use data from the Greek railway organization in this case study since rail transportation is used only in scenario 1 where Thessaloniki's container terminal constitutes the starting point of the supply chain network.

$$F(x,c) = \begin{cases} 0.626x + 144.78, & c < 60 \\ 0.595x + 137.55, & c \geq 60 \end{cases}$$

Where x represents the distances traveled.

The freight rate per container from Shanghai to the corresponding ports of destination in the scenarios examined is calculated based on the regression equations presented below. The regression equations derive from different offers provided based on

different container order quantities set. It is important to indicate that freight rates decrease as the volume of the containers transported increase.

- The regression equation for the route Shanghai-Thessaloniki is: *F(c) = 1499.5 – 0.6646c*
- The regression equation for the route Shanghai-Varna is *F (c) = 1710-0.662c*
- The regression equation for the route Shanghai-Constanta is *F (c) = 1595-0.6714c*

c represents the volume of the containers transported. So setting the number of containers ordered at c we calculate the discounted freight rate. Regarding the barge freight rates, they are calculated to be approximately 0.17$/Teu/km in 2000 (United Nations Economic and Social Commission for Asia and the Pacific ESCAP). Taking in consideration the inflation rate of 3% annually we assume that the price from January 2009 is equal to $0.17*(1.03)^9 = 0.2218$ $ or 0.18€. The total cost breakdown of the scenarios compared is indicated by the tables presented below:

Route	Transportation Mode	Distance	Cntrs	Cost (€)
Shanghai – Thessaloniki	Ship		300	391,000.0
Free out costs Th.PA			300	48,000.0
Lift on wagon costs			300	9,000.0
Th.Pa - Lom central DC	Rail	425.7	300	117,252.45
Lom – Sofia retail store	Truck	147.0	100	40,494.0
Lom – Giurgiu inland terminal	Barge	266.0	200	9,576.0
Giurgiu inland terminal-Bucharest retail store	Truck	64.7	200	32,810.0
			Total Cost	648,132.0

Table 4. Total costs of scenario 1

Route	Transportation mode	Distance	Cntrs	Cost (€)
Shanghai- Varna	Ship		300	453,420.0
Free out cost Varna container terminal			300	30,000.0
Extra Transshipment costs			300	15,000.0
Varna – Ruse central DC	Barge	191.3	300	10,330.2
City limit expenses Ruse inland terminal-Ruse central DC	Truck		300	24,000.0
Ruse central DC - Lom inland terminal	Barge	288.5	100	5,193.2
City limit expenses Ruse central DC-Ruse inland terminal	Truck		100	8,000.0
Lom inland terminal – Sofia retail store	Truck	147	100	40,494.0
Ruse central DC - Bucharest retail store	Truck	74.3	200	40,932.0
			Total Cost	627,369.4

Table 5. Total costs of scenario 2

Route	Transportation mode	Distance	Cntrs	Cost (€)
Shanghai – Constanta	Ship		300	480,000.0
Free out costs Constanta container terminal			300	30,000.0
Extra transshipment costs				15,000.0
Constanta - Giurgiu central DC	Barge	255.8	300	13,813.2
City limit expenses Giurgiu inland terminal-Giurgiu central DC	Truck		300	24,000.0
Giurgiu central DC - Lom inland terminal	Barge	274.1	100	4,933.8
City limit expenses Giurgiu central DC - Giurgiu inland terminal	Truck		100	8,000.0
Lom inland terminal - Sofia retail store	Truck	147	100	40,494.1
Giurgiu central DC – Bucharest retail store	Truck	64,7	200	32,810.0
			Total Cost	649,051.1

Table 6. Total costs of scenario 3

Results

In scenario 1 the total costs are 648,117.0€

In scenario 2 the total costs are 627,369.4€

In scenario 3 the total costs are 649,051.1€

Interpretation of results

The total costs in scenario 1 are approximately 20,000€ higher than the total costs of scenario 2 and approximately 1,000€ lower than the total costs of scenario 3, even though the shipping costs from Shanghai - Thessaloniki are significantly lower than the shipping costs from Shaghai- Varna and Shanghai- Constanta. The reason behind this cost difference is due to the signicantly high rail costs on one hand and the significantly low barge costs on the other, indicating that barge transportation constitutes a very cost effective transportation mode.

3.3 CO_2 emissions

Before moving into the calculation of the total CO_2 emissions of the scenarios examined, this paper will present the environmental performance (in CO_2 emissions) of each transportation mode involved in this case study separately. There are many papers that attempt to calculate CO_2 emissions for different transportation modes. The data used in this case study regarding CO_2 emissions derive from a presentation conducted

by Maersk line on the Stanford Supply Chain Conference (2008). Barge CO_2 emissions per TEU/km are not presented in the above presentation and are estimated based on a barge/rail ratio which is approximately common in all CO_2 emission papers. The barge/rail ratio is approximately 1.8, therefore the CO_2 emissions for each transportation mode are: The total CO2 of the scenarios compared are indicated in the tables below (the total grams of CO_2 emitted in each route of the scenarios tested are converted in tons by dividing them with 1,000,000):

Transportation modes	Grams of CO_2 emissions/km/TEU
Rail	147
Barge	266
Truck	472
Ship	84

Table 7. Grams of CO_2/km/TEU per each transportation mode

Route	Transportation Mode	Distance (km)	Cntrs	CO_2emissions (tons)
Shanghai – Thessaloniki container terminal	Ship	17,300.0	300	436.0
Th.Pa - Lom central DC	Rail	425.7	300	18.8
Lom – Sofia retail store	Truck	147.0	100	6.9
Lom – Giurgiu inland terminal	Barge	266.0	200	14.2
Giurgiu inland terminal- Bucharest retail store	Truck	64.7	200	6.2
			Total CO_2	482.1

Table 8: CO_2 emissions in scenario 1

Route	Transportation Mode	Distance (Km)	Cntrs	CO_2 emissions (tn)
Shanghai- Varna	Ship	18,215.0	300	459.0
Varna – Ruse central DC	Barge	191.3	300	15.3
Ruse central DC - Lom inland terminal	Barge	288.5	100	7.7
Lom inland terminal – Sofia retail store	Truck	147.0	100	7.0
Ruse central DC - Bucharest retail store	Truck	74.3	200	7.0
			Total CO_2	496.0

Table 9. CO_2 emissions in scenario 3

Route	Transportation mode	Distance (km)	Cntrs	CO_2 emissions (tn)
Shanghai – Costanta container terminal	Ship	18,400.0	300	464
Costanta - Giurgiu central DC	Barge	255.8	300	20.4
Giurgiu central DC - Lom inland terminal	Barge	274.1	100	7.3
Lom inland terminal - Sofia retail store	Truck	147.0	100	6.9
Giurgiu central DC - Bucharest retail store	Truck	64.7	200	6.1
			Total CO2	504.7

Table 10. CO_2 emissions in scenario 3

Interpretation of results

The results clearly indicate that if the company selects scenario 1 as a supply chain network configuration, it will achieve a higher level of environmental performance compared to options 2&3. The reason behind this performance is due to:

1. Shorter shipping distance from Shanghai

2. Use of rail transportation from Thessaloniki's container terminal to the central DC. Based on table 9 rail transportation constitutes the second most environmentally friendly transportation mode after deep sea transportation.

Scenario / Criteria	Scenario 1	Scenario 2	Scenario 3
Cost (€)	648,132.0	627,370	649,051.1
Time (h)	648.2	725.4	773.7
CO_2 emissions(tn)	482.1	496	504.7

Table 11. Total cost, time, CO_2 emissions of the scenarios tested

4. Development of the scoring model

The scoring model (Xanthopoulos, 2008), is a simple procedure through which total grades are set to alternative scenarios of a problem. The total grade is based on the sum of grades that are set to each one of the multiple criteria that are taken in consideration. The steps that are followed in order to apply this methodology are:

Step 1: Identification of the evaluation criteria *i* of the examined problem

Step 2: Set the weighting factor w_i at each criterion based on the subjective importance of the criterion for the manager. The waiting factor receives a value between 0-1

Step 3: Identification of the alternative scenarios *j* of the examined problem

Step 4: Grading (S_{ij}) of each alternative scenario $_j$ for each criterion *i.* The grading scale is not fixed and is free for format. eg. It can take values from 0-1, 0-5, 0-10, etc. In this case study it receives values between 0-10.

Step 5: For each alternative scenario *j* we calculate the corresponding weighted score (SCORE*j)*:

$$SCORE_j = \sum_i W_i S_{ij}$$

Step 6: Final selection of the alternative *j* with the highest weighted score.

4.1 Application of the model in our case study

		Cost [i=1]		*Time [i=2]*		*CO_2 Emissions [i=3]*		
	Scenarios	*[S_{1j}]*	*[W_1]*	*[S_{2j}]*	*[W_2]*	*[S_{3j}]*	*[W_3]*	*Score$_j$*
j=1	*Scenario 1*	*9.977*	*0.5*	*10*	*0.1*	*10*	*0.4*	*9.99*
j=2	*Scenario 2*	*10*	*0.5*	*9.88*	*0.1*	*9.97*	*0.4*	*9.97*
j=3	*Scenario 3*	*9.976*	*0.5*	*9.80*	*0.1*	*9.95*	*0.4*	*9.95*

Table 12: Total score in each scenario

Interpretation of results

The results indicate that the scenario with Thessaloniki's container terminal as the port of destination has the highest grade. If a company selects Thesssaloniki's container terminal as a start point of its supply chain configuration to the Balkans, it would achieve lower delivery times and higher environmental performance compared to supply chains that have Varna and Constanta container terminals as a start point. But the cost performance in scenario 1 is worse than that in scenarios 2 and 3 due to relatively high rail cost charges of the Greek-railway organization. If the Greek railways managed to provide competitive rail freight charges in cooperation with Bulgaria railways, Thessaloniki can fully utilize its strategic positioning in the Balkans and thus pull container flows from the other terminals in the Balkans.

5. Conclusions

The negative effects of climate change affect singifically the quality of our life and even threaten our existence. The transport sector is responsible for the release of a

significant level of CO_2 emissions in the atmosphere, constituting supply chains as an important source of environmental pollution. All decisions from now on, tactical or operational must seriously take in consideration the environment. A good start is the reduction of CO_2 emissions. This paper aimed to provide a simple approach in the selection of both efficient and environmentally friendly supply chains on a tactical level, taking only in consideration the CO_2 emissions factor.

References

1. Azar, C. (1996) 'Discounting and distributional considerations in the context of global warming', *Ecological Economics*, Vol. 19, pp. 169-84.
2. Daskin, M.S., Coullard, C. and Shen, Z.-J. M. (2002) 'An inventory-location model: Formulation, solution algorithm and computational results'. *Annals of Operations Research*, 110(1), 83–106.
3. Koetse M. J. and Rietveld, P., (2008) 'The impact of climate change and weather on transport: An overview of empirical findings', *Transportation Research Part D: Transport and Enviroment*, 14, (3), pp. 205-221.
4. McKinnon, A. (2006). CO_2 Emissions for Freight Transport: An analysis of UK Data. Report prepared for the Climate Change Working Group of the Commission for Integrated Transport. Obtained through the internet address: *http://www.cfit.gov.uk* [accessed 18/5/209].
5. Nakata, T. (2004) 'Energy Economic Models and the Environment', *Progress in Energy and Combustion Science*, Vol. 30 (4), pp. 417-415.
6. Rothengatter, W. (2000) 'External Effects of transport', *In*: Polak, J.B. and A Heertje (Eds). *Analytical Transport Economics: An International Perspective*. Edward Elgar: Cheltenham, U.K. and Northampton, Mass, pp. 79-116.
7. United Nations Economic and Social Commission for Asia and the Pacific (UNESCAP), Obtained through the Internet: www.unescap.org [accessed 18/5/2009].
8. Xanthopoulos, A. (2008) *Optimization of strategic design, production planning and operation planning of reverse supply chains*. Phd thesis. Department of Mechanical Engineering, Aristotle University of Thessaloniki.
9. Yu, B., Yang, Z. and Yao, B. (2008) 'An improved ant colony optimization for vehicle routing problem', *European Journal of Operational Research*, Vol. 196(1), pp. 171-176.

The Role of Banks in the Greek Monetary Policy Transmission

Aikaterini Markidou[1] and Eftychia Nikolaidou[2]

[1]South East European Research Center (SEERC) – A Research Center of the University of Sheffield and City Liberal Studies, 17 Mitropoleos Str, 54624 Thessaloniki, Greece, AMARKIDOU@SEERC.ORG

[2]City College, Affiliated College of the University of Sheffield, Business Administration and Economics Department, 2 Kalapothaki Str, 54624 Thessaloniki, Greece, ENIKOLAIDOU@CITY.ACADEMIC.GR

This paper investigates the relevance of the credit channel of monetary policy transmission in Greece by employing a SVAR model on both aggregated and disaggregated data and estimating the response of bank loans to different macroeconomic shocks. By distinguishing between households and firms instead of focusing on the response of total bank credit to a tightening monetary policy shock and by employing a SVAR methodology following the work by Safaei and Cameron (2003), this paper identifies structural models to study bank credit in Greece as a source of macroeconomic variation for the period 1980-2005. The findings suggest that the credit channel in Greece for the period 1980-2005 is inoperative when monetary base is considered as the main monetary policy variable. On the other hand, when interest rate is used to capture the role of monetary policy variable, there is promising evidence that the credit channel might be present and bank credit to individuals seems to be more vulnerable compared to bank credit to firms.

Keywords: Monetary policy transmission; bank lending; SVAR model; Greece

JEL classification: E44; E52; G21

1. Introduction

Monetary policy transmission mechanisms have been a subject of theoretical and empirical research over the last two decades in an effort to better comprehend how monetary policy affects the real economy[1]. A central point of this research has been the

[1] See Bernanke and Gertler (1995) and Kashyap and Stein (1993,1997) for excellent surveys.

role played by banks in the transmission of monetary policy, inclusive of the effect of monetary policy on bank lending. Despite the general agreement on the active role of banks in the transmission of monetary policy, there is an extensive dispute over the exact role that banks play. The basic idea behind the concept of the credit channel is that central bank impulses affect output, as an upshot caused by shifts in the supply of loans. This comes in contrast with the traditional money view where there is no reference to loan supply shocks. According to the money channel theory the focus is placed on the effects of monetary policy on the attraction of investment and saving that predominate when financial markets are complete. The major shortcoming of money view is the considerable difficulty in the identification of a quantitatively meaningful effect on aggregate spending and investment that the theory indicates it should influence. The credit channel has emerged to fill the gap. Even though the credit channel and the money channel share a common starting point in highlighting the relevance of financial considerations, they are reckoned to be complementary, which implies the simultaneous coexistence of the two transmission channels. Two subchannels of the credit channel have been presented in the literature, the balance sheet channel (BSC) and the bank lending channel (BLC). The BSC states that shifts in monetary policy can severely affect the creditworthiness of borrowers due to information asymmetries that consequently affect the bank loan supply and thus borrowers' investment and spending decisions stemming from the inability of borrowers to raise funds from any provider. The BLC, in turn, rests on the assumption that banks have a dual nature as holders of reserved-backed deposits and as originators of loans and this can affect the availability of bank loans. It is generally accepted that commercial banks are the major source for investment expenditure financing and consumption financing of small firms and households, particularly valid for most of the euro area countries that are prevailed by rather small-sized banks and strong customer relationships. The structure of a banking system is of great importance since it can create the required conditions for a BLC to be operative.

The credit channel theory has considerable appeal in Greece due to the dominance of small size bank-dependent firms with no alternatives to nonbank finance and the specific features of the Greek financial system. The Greek capital market was rather restricted compared to other EU countries and the Stock Exchange was significantly immature until mid 1990s. Moreover, banks were to a great extent government regulated and characterised by financial interventionism, restricting particular economic sectors from accessing bank credit, imposing restrictions on bank capital flows and reserve requirements as well. Mutual funds experienced considerable growth only towards the end of the 1990s after the systematic liberalisation of the financial markets[2]. As a result, it is easily anticipated that the Greek banking sector is of primary importance for the transmission of monetary policy and that the bank lending channel (BLC) exerts a dominant role particularly during the period prior to the liberalisation and transformation of the banking system in 1996.

Early literature has investigated the possible repercussions of monetary policy by studying the relationship between money and output as well as bank loans and output

[2] For more details regarding the deregulation of the Greek banking sector see Kashyap and Stein (1997) and Brissimis et al. (2001).

mainly by using correlations or through Granger-causality tests. It has approximately deduced in favor of the money channel. During the 1990s numerous studies have attempted to assess the existence of the BLC. Although the magnitude of financial market imperfections –the credit view in general- has been recognized by a substantial number of surveys, the empirical evidence for the existence of a BLC is much less definite. A very influential study that made use of aggregated US data and supports the BLC is that of Bernanke and Blinder (1992). Other SVAR studies supported the presence of the credit channel are that of Kashyap et al. (1993), Gibson (1997) for the U.S. where the composition of bank balance sheets has an impact on the strength of the transmission mechanism, whereas Carlino and DeFina (1999) following the theory proposed by Peek and Rosengren (1995), offered contrasting outcomes revealing that the credit channel is not present at the state level and greater concentration of small banks minimise responsiveness to monetary policy shocks at the level state. This is in contrast to the theory advocated by Kashyap and Stein (1994)[3]. Ramaswamy and Sloek (1998) and Clements et al. (2001) reported that a credit channel is present in most EU countries. Output responses of core countries react slower but deeper than in periphery countries. It is interesting to notice for instance that Gerlach and Smets (1995), Ehrmann and Worms (2001) revealed similar results against the efficacy of the credit channel in Germany while Hortemoller (2003) offers a contradictory outcome. Agung (1998) observed that the BLC is operative in Indonesia and consumer loans of all bank categories drop following monetary contraction. On the other hand, findings of the empirical part of Bacchetta and Ballabriga (2000) and Dedola and Lippi (2000) lie between the two opposite findings providing fairly supportive evidence for the BSC and inconclusive or partially consistent for the BLC.

This study attempts to analyse a macro-dynamic system utilising the SVAR approach that examines different bank credit measures (heterogeneity of the loan types) in Greece. Even if there are some drawbacks in the use of macro aggregated data, empirical research based on such data might still yield valuable insights[4]. So far, tests of the credit channel based on the importance of bank and other macroeconomic characteristics in determining the response of loans to a shift in monetary policy have been limited to U.S. and other developed EU countries' data. To our knowledge, studies involving macroeconomic time-series for the case of Greece have not attracted much attention by researchers due to data deficiencies. The only Greek study on the BLC is that of Brissimis et al. (2001) however it makes use of bank level panel data and analysed the effects of bank characteristics in terms of size and liquidity on the supply of bank loans. The remainder of the paper is organised as follows. Section 2 describes the data and presents the SVAR model. Subsequently, the estimated structural relationships are set out in section 3 along with the impulse response

[3] The hypothesis of Kashyap and Stein (1994) states that small banks can easier confront higher costs in attracting CDs and other non-deposit sources to counteract a loss in reserves resulting from monetary policy tightening. Consequently, small banks are forced to contract bank loans by relatively more than large banking groups.

[4] The limitations inherent in studies based on aggregated data are associated with the identification problem, whether or not the cutting off of credit following a monetary contraction is a result of deterioration in bank supply or decline in borrowers' demand for bank credit.

functions to the structural shocks and the variance decompositions. Section 4 provides concluding remarks.

2. The model

This section summons up the specification of the benchmark models that have been employed by Safaei and Cameron (2003). The small-scale models include 5 variables to capture the economy and the domestic sector of Greece. Although a larger SVAR that makes use of more variables may allow for more affluent interactions, a more parsimonious model is likely to be easier to estimate and more constant and more reliable to grasp key macroeconomic interactions. Impulse responses of macroeconomic variables to monetary shocks are estimated so as to investigate the credit view of monetary policy for the case of Greece over the period 1980-2005,

$$A_0\varepsilon_t = u_t \qquad [1]$$

where **A** is the matrix of short-run coefficients, $\boldsymbol{\varepsilon}_t$ is the reduced form residual resulting from the first step estimation of the VAR model and $\boldsymbol{u}_t$ is the independent structural disturbance term of the structural model. The **A** matrix contains n^2 parameters. Hence, in a contemporaneous version of a five variable model implying five equations, $n^2 - (n+1)\ n/2 = n\ (n-1)/2$ unique elements or to put it differently $((5^2 - 5)/2 = 10)$ restrictions are required to identify the structural parameters (orthogonal shocks) on the covariance matrix. Therefore, just-identification of the parameters in matrix A calls for exactly 10 restrictions (9 zero restrictions and an equality restriction) on the off diagonal elements of the matrix. In order to set the short-run restrictions that are necessary for the identification of the structural model, economic theory is applied for this purpose. In this way, the "aggregate demand" relationship is included in the contemporaneous restrictions. It is generally accepted that IS shocks (u_y) have a straightforward negative impact to real output which is greater comparing to that of price level. Taking into consideration shocks in output growth resulting from shocks in the growth rate of real credit ($\varepsilon_{ci} - \varepsilon_p$) is central to the theoretical underpinning of credit constraints that arise from the ability and willingness of banks in the bank lending procedure along with the cost of borrowing money which severely affects both the level of investment, real economic activity and the level of output[5]. Furthermore, the "aggregate supply" function is embraced, describing the shifts in price level and the corresponding impact in inflation rate stemming from shocks in aggregate demand. The thought behind shocks in aggregate demand is compatible with the short-run Philips curve, thus it is logical to recognize that inflation shocks (u_p) entail economic changes that drastically affect the cost of inputs derived from adverse shocks associated with factors of production such as oil prices, energy costs, unemployment

[5] Vermeulen (2002), Hubbard (1998), Boyd and Smith (1997), Oliner and Rudenbusch (1996) and Greenwald and Stiglitz (1993) are indicative examples in the literature providing empirical evidence of asymmetric effects resulting from monetary shocks that actually affect the supply of loans which in turn determine investment spending and the level of economic activity (balance sheet channel).

rate, tax regime or technological advances (i.e. information processing and communications technology)[6].

In addition, a policy reaction function caused by the monetary policy decisions set by the Central Bank should be considered as well. In particular, two policy variables are taken into account, namely monetary base and interest rate that make distinction between two classes of models (monetary base - B models and interest rate – R models). As far as Class B models are concerned, nominal reserve money responds in shifts of real output, price level as well as M1 balances (narrow money) reflecting a monetary base reaction function. It is reasonable to regard as it is also influenced by shocks in monetary base emerging from foreign exchange rate distractions[7]. Referring to Class R models, the policy variable selected is the interest rate. More precisely, interest rate is considered and is taken for granted to react to price and money balances but not to output at least in short-term. Responsiveness to output is added in the reaction function due to the just-identification of the model. Nonetheless, it seems rational to presume that in the bounds of the quarter, no data for real GDP is on hand. On contrary, data referring to inflation and money supply are offered in monthly basis.

Also, an illustration of the money stock function and specification of the demand for money balances are embodied. Many factors that determine the demand for money balances are found in the literature. Nevertheless, the level of real GDP, the level of prices, the level of interest rates accompanied by the pace of financial innovation turn out to exhibit greater influence to the money stock function. For this reason, structural disturbances u_m represent such financial innovation factors like financial system/institutional innovations, process and product innovations that are not unambiguously incorporated in the models.

An effort is made to incorporate a "credit supply" function attempting to capture the credit view of monetary transmission mechanism originated by the influential study of Bernanke (1983)[8] and empirically tested by King (1986). To be more specific, in the credit supply function, responses in price shocks, monetary base shocks and money demand shocks are taken into account. Monetary base tightening weakens the accessibility of funds to the financial system and to banks in particular. This results to credit constraints of borrowers as a consequence of draining bank reserves depressing the amount of loanable funds. In this way, an attempt is made so as to identify the credit view of monetary transmission mechanism. In addition, the inclusion of money demand can be rationalised based on the fact that unanticipated money shocks have an effect on monetary aggregates and bank deposits that represent the principal source of

[6] See Ball and Mankiw (1995).

[7] A considerable amount of papers take into account the exchange rate in investigating the transmission mechanism of monetary policy, i.e. Taylor (2001) and Leitemo and Soderstrom (2005).

[8] Bernanke (1983) examined the correlations between bank lending and economic activity and the consequences of output on bank lending. In this framework, the author investigated the impact of the Great Depression in the United States for the period 1930-1933. However, in his SVAR model, credit grants an unresponsive prospect, which cannot disentangle the supply-versus-demand puzzle.

funds for bank lending. This can lead to the deterioration of investment funds imposed by the negative influence on loanable funds which force bank-dependent borrowers to contract investments. Structural credit shocks, u_{ci}, correspond to exogenous financial innovations or monetary policy regulatory innovations.

Given the short-run restrictions analysed above based on economic theory, the matrices for the two different classes of models B and R are as follows:

$$\begin{vmatrix} 1 & a1 & 0 & 0 & -a1 \\ -a2 & 1 & 0 & 0 & 0 \\ -a3 & -a4 & 1 & -a5 & 0 \\ -a6 & -a7 & 0 & 1 & 0 \\ 0 & -a8 & -a9 & -a10 & 1 \end{vmatrix} \begin{vmatrix} \varepsilon_y \\ \varepsilon_p \\ \varepsilon_b \\ \varepsilon_m \\ \varepsilon_{ci} \end{vmatrix} = \begin{vmatrix} u_y \\ u_p \\ u_b \\ u_m \\ u_{ci} \end{vmatrix}$$

By multiplying the matrices we obtain 5 equations that are clearly displayed in the structural equations [2] to [6] for the Class B models. On purpose, the time subscripts are neglected to have a more comprehensible presentation.

$$\varepsilon_y = a1\,(\varepsilon_{ci} - \varepsilon_p) + u_y \qquad [2]$$
$$(+)$$

$$\varepsilon_p = a2\varepsilon_y + u_p \qquad [3]$$
$$(+)$$

$$\varepsilon_b = a3\varepsilon_y + a4\varepsilon_p + a5\varepsilon_m + u_b \qquad [4]$$
$$(+) \quad (-) \quad (?)$$

$$\varepsilon_m = a6\varepsilon_y + a7\varepsilon_p + u_m \qquad [5]$$
$$(+) \quad (+)$$

$$\varepsilon_{ci} = a8\varepsilon_p + a9\varepsilon_b + a10\varepsilon_m + u_{ci} \quad (i = 1,2,3) \qquad [6]$$
$$(+) \quad (-) \quad (?)$$

In equations [2] to [6] ε_y, ε_p, ε_b, ε_m and ε_{ci} correspond to the reduced form errors, "innovations" in the related variables. Similarly, u_y, u_p, u_b, u_m and u_{ci}, specify the orthogonal structural shocks. Respectively, the coefficients $a1$, $a2$,……,$a10$ are the structural parameters that measure instantaneous reactions and identify the short-run relationships among the shocks in variables. Lastly, c_i (i = 1,2,3) symbolize the three alternative measures of bank credit. The corresponding relationship of the structural parameter is specified in the parentheses underneath the coefficients and a question mark (?) under a parameter implies theoretical uncertainty on the appropriate sign related to the economic relationship of that parameter.

To this point, class B models considering the monetary base have been clarified. Regarding class R models reckon with interest rate, approximately similar theoretical relationships are applicable. Following the same procedure, the contemporaneous

restrictions referring to Class R models lead to the matrices depicted below

$$\begin{vmatrix} 1 & a1 & 0 & 0 & -a1 \\ -a2 & 1 & 0 & 0 & 0 \\ 0 & -a3 & 1 & -a4 & 0 \\ -a5 & -a6 & -a7 & 1 & 0 \\ 0 & -a8 & -a9 & -a10 & 1 \end{vmatrix} \begin{vmatrix} \varepsilon_y \\ \varepsilon_p \\ \varepsilon_r \\ \varepsilon_m \\ \varepsilon_{ci} \end{vmatrix} = \begin{vmatrix} u_y \\ u_p \\ u_r \\ u_m \\ u_{ci} \end{vmatrix}$$

Similarly, by multiplying the matrices we obtain another 5 equations that can be viewed distinctly in the structural equations (7) to (11) for the Class R models. Again, the time subscripts are omitted so as to facilitate the understanding of the structural relationship between the variables.

$$\varepsilon_y = a1\,(\varepsilon_{ci} - \varepsilon_p) + u_y \qquad [7]$$
$$(+)$$

$$\varepsilon_p = a2\varepsilon_y + u_p \qquad [8]$$
$$(+)$$

$$\varepsilon r = a3\varepsilon_p + a4\varepsilon_m + u_r \qquad [9]$$
$$(+) \quad (?)$$

$$\varepsilon_m = a5\varepsilon_y + a6\varepsilon_p + a7\varepsilon_r + u_m \qquad [10]$$
$$(+) \quad (+) \quad (-)$$

$$\varepsilon_{ci} = a8\varepsilon_p + a9\varepsilon_r + a10\varepsilon_m + u_{ci} \quad (i = 1,2,3) \qquad [11]$$
$$(+) \quad (-) \quad (?)$$

Equations [7] and [8] are precisely the same with Equations [2] and [3]. In Equation [9] where the policy response function is represented, instead of the monetary base, the interest rate is considered. Turning to Equation [10], the money demand function is enriched by permitting impacts arising from interest rate shocks, leaving the other elements unchanged. Consequently, in Equation [11] interest rate shocks instead of monetary base are set to influence credit supply function, and to put it differently interest rates shocks are behind the credit slowdown. An alternative method employed by Sims (1986) named Maximum Likelihood method (ML) and further developed by Amisano and Giannini (1997) is employed to estimate the parameters in Equations [2] to [11] for both classes and to calculate standard errors. The log likelihood is maximized by the method of scoring (with a Marquardt-type diagonal correction).

2.1 Data and variable selection

Five endogenous variables are encompassed in the SVAR models to detect the macroeconomic responses of the Greek economy to monetary innovations in the spirit of other studies of the monetary policy transmission process: output (*y*), price level (*p*), money stock (*m*), credit (*c*) and a monetary policy variable. There are two classes of models under consideration namely Class B and Class R. For Class B model the

monetary policy variable that has been selected is monetary base (*b*), whereas in Class R model is bank rate (*r*). In order to disentangle the impact of monetary policy, three different measures of bank credit (*c*) are considered that is to say credit to households (*c*1), credit to firms (*c*2) and total bank credit (*c*3). Hence, there are three models in class B and class R respectively. Seasonally adjusted quarterly time series for Greece over the period 1980:1 to 2005:4 are used. The time series in the SVAR analysis are: real gross domestic product (*y*), the consumer price index (*p*), the monetary base (*b*), overnight interbank rate (*r*), narrow money stock M3 (*m*), bank credit to households (*c1*), bank credit to firms (*c2*) and total bank credit (*c3*) used as three alternative measures of credit. The overnight interbank rate is the bank rate. For the analysis of macro data, economic aggregates such as money stock and monetary base are obtained from monthly statistical bulletins published by the Bank of Greece (BoG). Variables regarding bank credit are based on monthly data provided from the statistics department of Bank of Greece (BoG) as well as monthly data on the overnight interbank rate. Real GDP is taken from the National Statistical Service of Greece (NSSG).

3. Estimation results

Having ensured stationarity of all variables, the estimation results based on the Maximum Likelihood method (ML) demonstrate the estimated structural relationships for both classes of models in Table 1 and 2. The significance of the estimated coefficients have been determined taken into account the p-values along with the *z*-statistic that allow us to perform the test at the 5% significance level, a p-value lower than 0.05 is taken as evidence to reject the null hypothesis of a zero coefficient and a z-statistic greater than 2 at the same significance level respectively. To begin with the monetary-base models, it is apparent that the estimation results of B2 and B3 models are quite similar, however they are not significantly different than that of B1 model. More importantly, on average 7 out of 8 expected signs indicated in the previous section confirm the economic theory behind the equivalent relationship of the structural parameters for all B models.

The positive sign and the significant coefficient of credit in the aggregate demand equation of all models of B class are in accordance with the economic theory. In particular, they verify the fact that credit constraints do have an effect on either the level of investment or real economic activity and the level of output. The positive and significant coefficient of output in the aggregate supply equation for all B models, also confirms the economic theory, implying that the short-run Philips curve holds to be true in Greece. Aggregate demand innovations are important determinants of the course of inflation, at least in the short run. Referring to monetary base reaction function, the response of monetary base to output in all models is positive and significant indicating that price level targeting has been undertaken with the objective to achieve price stability. The positive but insignificant coefficient of price level to monetary base in B1 model can be interpreted as a lack of inflation targeting initiated by the monetary authority. In B2 and B3 models, the negative insignificant coefficient of price level implies that monetary policy can actually affect monetary variables such as price level. Furthermore, the positive but insignificant response of monetary base to money demand expresses the altering of interest rates by the monetary authority as

being able to affect the money supply and keep it under it's control by mandating specific types of interest rates that may lead to liquidity shocks. The negative coefficient of output in the money demand function is unexpected suggesting that the level of real output does not exhibit any influence to the money demand. On the other hand, the significant but positive coefficient of level of prices in B2 and B3 models proves to be a factor that determines the demand for money balances.

All coefficients referring to credit supply function are insignificant but the majority of them assert the economic theory. More specifically, in all B models the credit supply is negatively affected by monetary base. It is notable that the negative sign in monetary base cannot affirm the existence of the credit view of monetary transmission mechanism for any form of credit. In other words, the respective negative coefficients in all B models point to the fact that the credit view seems not to be relevant in the individual, corporate or total credit implying a non-operative role of the credit view for any credit category. Also, is notable in household credit the decrease of credit supply resulting from inflation growth. Additionally, the theoretical uncertainty prevailing the effect of money demand in the credit supply function is evident due to the fact that leads to contradictory sings in the coefficients of B1 model on the one hand and that of B2 and B3 models on the other hand. In particular, model B1 points to a positive impact of money demand shocks on the credit supply function while models B2 and B3 underline a negative impact correspondingly. Having examined the B models it is reasonably to assume that the credit view cannot be supported when monetary base is taken into account as a monetary policy variable even if being insignificant.

Turning to class R models, the results of the estimated coefficients for the 3 different models are fairly comparable as they are presented in Table 2. In the aggregate demand function there is a significant and positive impact of credit in R1 and R3 models supporting the economic theory in a way similar to monetary-base models. It is remarkable that all remaining coefficients in the aggregate supply function, interest rate function and money demand function are statistically insignificant but like in monetary base models the majority of the expected signs confirm the economic foundation imposed in the contemporaneous restrictions. More precisely, the aggregate supply function is characterized by the insignificant short-run impact of aggregate demand on price level. In the interest rate function which reflects the monetary policy variable for this class of models, it is evident that interest rate responds positively nonetheless insignificantly to price level in R1, R2 and R3 models, which indicates once again absence of inflation targeting by the monetary authority following the corresponding interest rate shock. In the same function, shocks in money demand lead to negative and insignificant effects on interest rate only in R2 model which is contrary to the responses of monetary-base models discussed above. In this sense, it is logically to assume that when interest rate is taken into account as a monetary-base variable, money demand shocks affect interest rates in a similar way than they affect monetary base.

Looking at the money demand function, it doesn't verify the expected positive relationship between money demand and output in any R model, however there is validation for the expected positive relationship between money demand and price level in R1 and R3 models since nominal money demand is proportional to the price level in the economy (notable exception is the independent relationship between real money demand and price level). Additionally, there is validation for the expected

reverse relationship between money demand and level of interest rates in R1 and R3 models. Lastly, the estimated coefficients of the credit supply function offer valuable insights for the credit view in the Greek economy.

Model B1

$\varepsilon_y = 1.313\ (\varepsilon_{c1} - \varepsilon_p) + u_y$ $\quad \sigma(u_y) = 0.058$
(0.629) (0.036)

$\varepsilon_p = 0.073\ \varepsilon_y + u_p$ $\quad \sigma(u_p) = 0.011$
(0.039) (0.059)

$\varepsilon_b = 0.342\ \varepsilon_y + 0.601\ \varepsilon_p + 0.268\ \varepsilon_m + u_b$ $\quad \sigma(u_b) = 0.076$
(0.148) (0.713) (0.229)
(0.020) (0.399) (0.241)

$\varepsilon_m = -0.158\ \varepsilon_y + 0.414\ \varepsilon_p + u_m$ $\quad \sigma(u_m) = 0.033$
(0.063) (0.312)
(0.013) (0.184)

$\varepsilon_{c1} = -0.218\ \varepsilon_p - 0.019\ \varepsilon_b + 0.037\ \varepsilon_m + u_{c1}$ $\quad \sigma(u_{c1}) = 0.014$
(0.180) (0.020) (0.046)
(0.225) (0.337) (0.413)

Model B2

$\varepsilon_y = 0.717\ (\varepsilon_{c2} - \varepsilon_p) + u_y$ $\quad \sigma(u_y) = 0.053$
(0.376) (0.056)

$\varepsilon_p = 0.053\ \varepsilon_y + u_p$ $\quad \sigma(u_p) = 0.010$
(0.025) (0.035)

$\varepsilon_b = 0.453\ \varepsilon_y - 0.230\ \varepsilon_p + 0.340\ \varepsilon_m + u_b$ $\quad \sigma(u_b) = 0.070$
(0.141) (0.683) (0.230)
(0.001) (0.736) (0.140)

$\varepsilon_m = -0.107\ \varepsilon_y + 0.590\ \varepsilon_p + u_m$ $\quad \sigma(u_m) = 0.031$
(0.061) (0.296)
(0.080) (0.046)

$\varepsilon_{c2} = 0.164\ \varepsilon_p - 0.004\ \varepsilon_b - 0.016\ \varepsilon_m + u_{c2}$ $\quad \sigma(u_{c2}) = 0.017$
(0.185) (0.025) (0.058)
(0.375) (0.872) (0.777)

Model B3

$\varepsilon_y = 1.189\ (\varepsilon_{c3} - \varepsilon_p) + u_y$ $\quad \sigma(u_y) = 0.054$
(0.531) (0.025)

$\varepsilon_p = 0.068\ \varepsilon_y + u_p$ $\quad \sigma(u_p) = 0.010$
(0.031) (0.028)

$\varepsilon_b = 0.446\ \varepsilon_y - 0.183\ \varepsilon_p + 0.326\ \varepsilon_m + u_b$ $\quad \sigma(u_b) = 0.071$
(0.145) (0.698) (0.228)
(0.002) (0.792) (0.153)

$\varepsilon_m = -0.108\ \varepsilon_y + 0.626\ \varepsilon_p + u_m$ $\quad \sigma(u_m) = 0.031$
(0.063) (0.305)
(0.087) (0.040)

$\varepsilon_{c3} = 0.070\ \varepsilon_p - 0.010\ \varepsilon_b - 0.029\ \varepsilon_m + u_{c3}$ $\quad \sigma(u_{c3}) = 0.01$
(0.168) (0.020) (0.046)
(0.676) (0.624) (0.519)

Table 1: Estimated structural relationships for monetary-base models

Note: The numbers in parentheses are asymptotic standard errors and P-values respectively. σ_u's are the estimated standard errors of the structural disturbances.

R1 model responses in a way similar to B1 model where there is a negative relationship, yet not significant, between credit and price level innovations. Nevertheless, the given economic theory is justified in R2 and R3 models. Once again there is no definite impact of money demand shocks to credit supply, as it is evident in the interest rate models. In R1 model, is apparent the positive effect of money demand shocks to credit supply, while the opposite prevails R2 and R3 models, rationalising thus the theoretical ambiguity on the issue. The unexpected positive response, even though significant, of credit supply to interest rate innovations in all models implies that banks in Greece induce bank loan supply in case of interest rate increase brought about by tighter monetary policy. This is in contrast with the economic underpinnings and an alternative explanation can be found in the parallel shifts of overnight interbank interest rates, deposit and lending interest rates and to put it differently similar interest spread across different types of interest rates. Admittedly, the impact of interest rate shocks is very determined and it is noticeable the comparable impact to all types of credit, however credit to individuals give the impression of being less affected by interest rate innovations compared to firm credit and total bank credit.

To conclude, in spite of a considerable amount of insignificant estimated coefficients in R class models mainly, the general impression for the credit view in Greece is quite prompting. A large portion of the estimated coefficients conforms to the imposed restrictions of the economic structure and the outcomes are pretty stable for the 3 alternative credit measures. Having examined the estimated structural relationships, the next section deals with the structural impulse response functions to study the consequences of the different shocks and to provide evidence in support of the theoretical model.

3.1. Structural impulse response functions

In the previous section, the structural parameters for the two classes of models have been estimated. The next step involves the study of the behaviour of the macro-dynamic system incorporated in the models. In other words, this section observes the econometric outcomes of the structural specification. For this reason, impulse response functions depict the dynamic impact following a one-time shock to one of the innovations on current and future values of the system's endogenous variables. The impulse responses facilitate to bring to light the interaction of credit with the rest of the economy. The impulse responses are estimated setting the impulse to one standard deviation of the residuals, thus ignoring the correlation in the VAR residuals. Impulse responses' standard errors are estimated by using Runkle's (1987) Monte Carlo simulation method and are plotted for a time period of 12 quarters (12 quarters/4=3 years) for all variables. The vertical axes depict changes in the variables from their baseline. The dashed lines are 90% confidence intervals obtained by employing Kilian's (1998) technique. It is reasonable to assume that for a model with five variables, 25 impulse response functions are obtainable. Since the impulse responses

for B2 are somewhat akin to B3 model, only the impulse responses for B1 and B2 are reviewed[9].

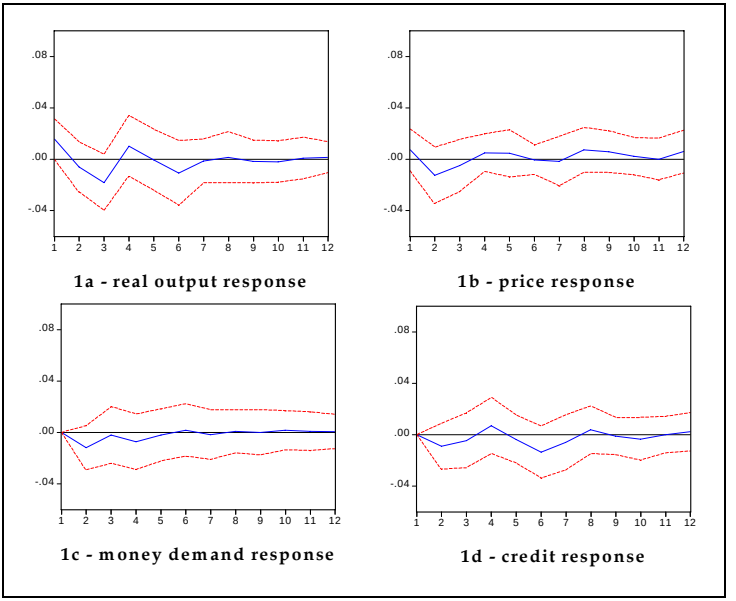


Figure 1. Responses to a monetary policy shock – B1 model

Note: The dashed lines enclose the one-standard deviation of the residuals that have been calculated from the asymptotic distributions of the responses.

The adverse monetary shock of model B1 presented in Figure 1 has a rather weak negative effect initially on real output, price level, money demand and credit supply as indicated in the economic theory. Approximately 3 quarters after the monetary shock, real output, price level, credit and money demand reveal a minor almost insignificant rise, which then tend to approach the zero baseline. Especially, the indifferent long-run response of real output is well suited in the economic theory. The negative impact of credit to monetary base resulting from contractionary monetary policy is in favour of the credit view however this outcome is in contrast with estimated structural relationships at least in the short run.

Responses to a credit shock are depicted in Figure 2 where it is noticeable the positive effect on real output that converges to zero in the following periods, the moderate short-term negative impact on money demand, the initial scarcely positive effect on monetary base is accompanied by a relatively unaffected period of no effect and the permanent adverse effect on price level for all the forecast horizon. Hence, the impulse responses present an inconsistency with the estimated structural relationships in the

[9] The impulse responses for the interest-rate models also do not significantly deviate from that of monetary-base models.

credit supply function and the economic theory as well. Generally speaking, monetary base positively responds to credit shock according to the impulse responses, however, the estimated structural coefficient points to an insignificant but negative impact which fits the economic theory. In other words, a paradox is observed in this case. Monetary base should have the same effect with price level and real output, nevertheless it seems to have an independent response. Moreover, the corresponding response of money demand as implied by policy reaction function should exhibit a positive effect, though the impulse response of money demand is associated with a rather negative effect through the 12 quarters following the credit shock. It is reasonably to assume that the dynamics of the monetary policy shock and money demand may not be captured in the corresponding functions of the structural model.

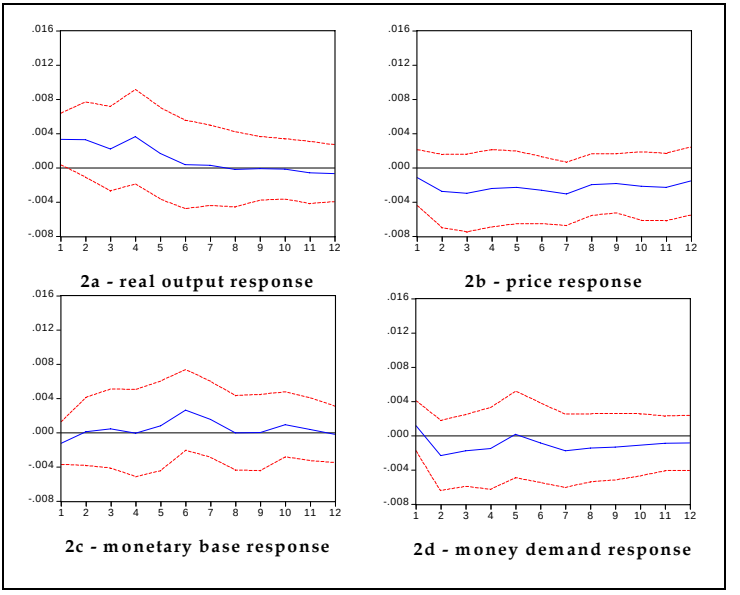


Figure 2. Responses to a credit shock – model B1

Note: The dashed lines enclose the one-standard deviation of the residuals that have been calculated from the asymptotic distributions of the responses

In Figure 3 there is roughly no difference in the impulse responses to a monetary policy shock in B1 and B2 models. Needless to repeat, the relevant impulse responses in B2 model exert similar characteristics like in B1 model. The only contrast can be found in credit response where right after the monetary shock credit seems to remain unaffected for about 2 quarters, then scarcely rises for another 2 quarters to decline and persists in a negative direction for the remaining period. The price response is the only one that is in line with the structural coefficients whereas real output and money demand responses pursue a reverse direction that then estimated ones.

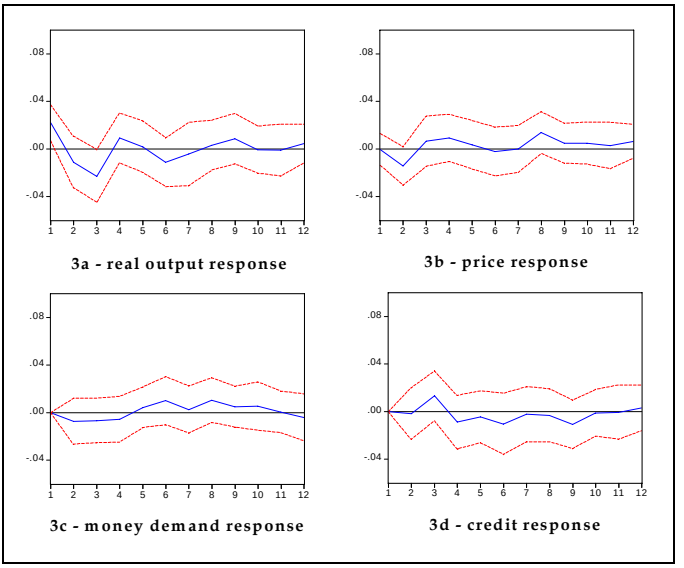


Figure 3. Responses to a monetary policy shock – B2 model

Note: The dashed lines enclose the one-standard deviation of the residuals that have been calculated from the asymptotic distributions of the responses

Figure 4. Responses to a credit shock – model B2

Note: The dashed lines enclose the one-standard deviation of the residuals that have been calculated from the asymptotic distributions of the responses

Surprisingly, there is a differentiation in the responses of B1 and B2 models and they can be found in the case of a credit shock. To be more specific, in Figure 4, it is clear that money demand in B2 model responds in an opposite way than that in B1 model and in particular it remains positive for almost 8 quarters after the credit shock, and then it becomes fairly neutral. The price response in B2 model is not that negative compared to the response in B1 model, which has a permanent negative effect. Nonetheless, the price response cannot be considered to have the expected response as indicated in the credit supply function. Likewise, the monetary base response is more dynamic in B2 model. Unexpectedly, the impulse responses referring to monetary base are in contrary with both the economic theory and the estimated coefficient in the credit supply function. More importantly, a credit shock has a positive effect in the real output with a tendency to reach the baseline following the fourth quarter. The two models present a difference in terms of the response of real output to a credit shock. The real output response in household credit is somewhat more definite than that of corporate credit. The respective effect on household credit model is steadily positive for the first 3 quarters and then there is another positive impact that lasts for another 3 quarters before becoming neutral. This can be taken as an indication of greater credit constrains for households in Greece. A possible explanation for this incident is that household credit in form of consumer and housing loans is a relatively new activity of the banking operations in Greece hence the openness to bank loans attracts the interest of the Greek banks basically following the transformation of the banking system after 1996.

There are some implications for the credit view after presenting the impulse responses functions for the 2 classes of models that may contribute to the investigation on the bank lending channel of transmission mechanism and shed some light on the issue for the case of Greece. The feeling extracted from the responses is that there is no explicit evidence for the existence of the credit view just encouraging indication no matter if monetary base or interest rate is used as a monetary policy variable given the presence of certain anomalies. To sum up, the findings of the impulse responses support the view that households instead of firms give the impression of being more credit constrained. This argument is based on the findings referring to the impact of bank credit shock in real output. However, credit constrains to households are somewhat predicted by the structural economic theory, nonetheless it can be verified by the organization and structure of the Greek banking system.

3.2. Variance decompositions

An alternative method to interpret system variables and the properties of the models is the variance decomposition. To assess the importance of the different shocks for the macroeconomic variables, the forecast error variances of the variables are decomposed with reference to the shocks. Given that factorisation is based on structural orthogonalisation, the forecast standard errors are identical to those from the Cholesky factorisation since the SVAR is just identified. It is representative for a variable to elucidate all its forecast variance in short run horizons and smaller percentages in long run horizons. The results for different time horizons are presented in Tables 3 and 4 only for B1 and B2 models, since there is little variability between corporate and total bank credit for both classes of models.

Variable	Quarter (s) ahead	IS shock	Aggregate supply shock	Monetary policy shock	Money demand shock	Credit supply shock
Real output	1	100.0	0.0	0.0	0.0	0.000
	2	93.7	1.3	1.1	3.6	0.3
	4	92.0	1.3	2.5	3.9	0.3
	8	83.0	2.3	6.6	5.3	2.8
	10	82.4	2.7	6.8	5.3	2.8
	12	81.9	3.0	7.0	5.3	2.8
Price	1	0.6	99.4	0.0	0.0	0.0
	2	0.7	93.6	0.1	4.3	1.3
	4	1.5	89.9	1.1	5.1	2.4
	8	2.4	84.3	2.2	6.7	4.4
	10	2.1	82.3	3.5	7.5	4.6
	12	2.5	81.6	3.7	7.2	5.0
Monetary base	1	3.9	0.9	95.2	0.0	0.0
	2	4.1	3.1	89.5	2.1	1.2
	4	9.5	3.5	82.2	2.7	2.1
	8	9.9	4.0	78.8	2.6	4.7
	10	9.9	4.4	78.3	2.5	4.9
	12	9.8	4.8	78.0	2.5	4.9
Money demand	1	5.1	1.9	1.3	91.7	0.0
	2	4.5	1.7	3.6	89.2	1.0
	4	5.0	8.3	4.5	81.3	0.9
	8	4.8	12.2	7.7	71.3	4.0
	10	4.7	12.5	7.9	70.5	4.4
	12	4.7	14.3	7.8	68.9	4.3
Creditto persons	1	5.6	0.7	0.7	0.7	92.3
	2	7.1	2.9	0.5	2.2	87.3
	4	10.4	6.2	0.5	3.1	79.8
	8	9.7	11.1	2.6	4.0	72.6
	10	9.4	12.5	2.8	4.6	70.7
	12	9.4	13.8	2.7	4.8	69.3

Table 3. Variance decompositions for the B1 model (Proportion of forecast error variable for variable)

It is another paradox that in the short and long run, all shocks that correspond to each variable's own shocks are important determinants for their forecast errors. For example in Table 3 it is apparent that aggregate supply, monetary policy, money demand and credit supply shocks contribute only a small part to the forecast error variance of real output during the whole forecast period. At the end of time horizon, IS shock accounts for 81,9 per cent of total variation in real output, AS shock for 3 per cent, monetary policy shock for 7 per cent respectively. In the study of Safaei and Cameron (2003) IS

shock contributes approximately 45 per cent to the total real output variation after 40 quarters following the shock while aggregate supply shock gradually increases its influence reaching a peak up to 22 per cent.

Variable	Quarter(s) ahead	IS shock	Aggregate supply shock	Monetary policy shock	Money demand shock	Credit supply shock
Real output	1	100.0	0.0	0.0	0.0	0.0
	2	91.4	1.2	0.4	6.1	0.9
	4	84.0	2.7	1.5	6.6	5.2
	8	71.7	4.5	5.5	8.7	9.6
	10	71.2	4.7	5.7	8.7	9.7
	12	70.2	5.4	5.7	9.2	9.5
Price	1	2.2	97.8	0.0	0.0	0.0
	2	2.2	89.7	0.1	7.9	0.1
	4	3.3	83.1	0.2	8.5	4.9
	8	4.8	76.2	2.0	12.0	5.0
	10	5.2	73.0	2.6	13.5	5.7
	12	5.4	73.1	2.6	13.0	5.9
Monetary base	1	8.7	0.0	91.3	0.0	0.0
	2	9.8	3.4	85.8	0.9	0.1
	4	16.8	4.6	73.3	1.8	3.5
	8	15.8	6.3	69.1	4.2	4.6
	10	16.1	6.5	67.0	4.6	5.8
	12	15.9	6.9	66.8	4.7	5.7
Money demand	1	2.5	3.8	2.1	91.6	0.0
	2	2.4	3.3	8.5	82.7	3.1
	4	4.3	8.5	8.5	74.8	3.9
	8	3.5	14.3	11.3	66.7	4.2
	10	3.4	14.9	10.9	65.6	5.2
	12	3.4	16.1	11.2	64.1	5.2
Credit to firms	1	5.5	1.5	0.0	0.1	92.9
	2	9.6	1.4	0.3	0.1	88.6
	4	11.6	2.6	1.0	1.8	83.0
	8	11.8	3.3	4.0	2.7	78.2
	10	11.5	3.3	7.1	2.7	75.4
	12	11.9	3.6	7.1	2.7	74.7

Table 4. Variance decompositions for the B2 model (Proportion of forecast error variable for variable)

Overall, the variance decompositions for household bank credit are comparable to some of the impulse responses presented in the previous section. However, some of the variance decompositions are in contrast to the estimated results and the impulse response functions. The money supply shock has a very modest impact on real output and even smaller to the variance of price level and the contribution of aggregate supply shock is noticeable only to money demand and household credit. It is remarkable that credit shock is not an important factor in bringing about essential variations in other variables.

The variance decompositions for B2 models are shown in Table 4. There are several differences but not very determined between household and corporate credit. For corporate credit, credit supply and money demand shocks are greater determinants of real output compared to that of household credit and this influence is apparent in the second quarter after the initial shock. Money demand shock plays a more active role in the variation of price level of corporate credit. There is also a considerable increase in the influence of IS shocks over the monetary base. In the same way, the contribution of monetary policy shock on money demand is greater for B2 model compared to B1 model. Finally, the role of aggregate supply shock is deteriorated for B2 model and on the other hand, the significance of monetary policy shock is triple than that of B1 model.

The outcomes from Table 3 and 4 support the belief that IS shocks along with money demand shocks play an important role in the variation of major macroeconomic variables and to a lesser extent monetary policy shocks. On the other hand, credit is assigned by a limited contribution in general macroeconomic fluctuations.

4. Concluding remarks

This paper examined the importance of bank credit as a critical financial source of variation on output and other major macroeconomic variables in Greece for the period 1980-2005. Following a SVAR approach that imposes contemporaneous constraints in the structural models and distinguishing between household, corporate and total bank lending based on two different classes of models for two different monetary policy variables, the outcomes offer sensible rationale for the specifications embodied in the SVAR models. A reasonable number of precise estimates of the structural parameters is observed, which implies consistency with the theoretical grounds. The feeling derived from the impulse responses fairly fitted the choice of credit measure.

However, lack of similar studies for the case of Greece does not permit comparison on the drawing conclusions although the results of this study mostly match those of Brissimis et al. (2001), the only Greek study that addresses the credit channel by using panel bank level data. The outcomes of this study clearly pointed to the significant impact of monetary policy on the supply of bank loans and on aggregate economic activity in general. However, in this study when monetary base is considered as a monetary policy variable, the structural estimates are not in favor of the existence of the bank lending channel in Greece. So inconclusive evidence calls for further investigation on the issue. A likely extension of this study would be to use a sample

taking into consideration the years following the financial liberalisation in 1996 capturing this way the advances that took place in the banking sector. This may prove to be useful to uncover the credit channel in Greece.

References

1. Agung, J. (1998) "Financial deregulation and the bank lending channel in developing countries: The case of Indonesia". *Asian Economic Journal*, 12 (3), p. 273-294.
2. Amisano, G., and Giannini, C. (1997) "*Topics in structural VAR econometrics*". 2nd ed. Springer-Verlag: New York.
3. Bacchetta, P., and Ballabriga, F. (2000) "The impact of monetary policy and banks' alance sheets: Some international evidence". *Applied Financial Economics*, 10 (1), p. 15-26.
4. Ball, L., and Mankiw, N.G. (1995) "Relative-price changes as aggregate supply shocks". *The Quarterly Journal of Economics*, 110 (1), p. 161-193.
5. Bernanke, B.S., and Blinder, A.S. (1992) "The federal funds rate and the channels of monetary transmission". *American Economic Review*, 82 (4), p. 901-921.
6. Bernanke, B.S., and Gertler, M. (1995) "Inside the black box: The credit channel of monetary policy transmission". *Journal of Economic Perspectives*, 9 (4), p. 27-48.
7. Bernanke, B.S. (1983) "Non monetary effects of the financial crisis in the propagation of the Great Depression". *American Economic Review*, 73 (3), p. 257-276.
8. Boyd, J.H., and Smith, B.D. (1997) "Capital market imperfections, international credit markets, and noncovergence". *Journal of Economic Theory*, 73 (2), p. 335-364.
9. Brissimis, S.N., Kamberoglou, N.C., and Simigiannis, G.T. (2001) "Is there a bank lending channel of monetary policy in Greece? Evidence from bank level data". *ECB Working Paper No. 104*.
10. Carlino, G., and DeFina, R. (1999) "The differential regional effects of monetary policy: Evidence from the U.S. States". *Journal of Regional Science*, 39 (2), p. 339-358.
11. Cecchetti, S.G. (1995) "Distinguishing theories of the monetary transmission mechanism". *Federal Reserve Bank of St.Louis Review*, 77 (3), p. 83-97.
12. Clements, B., Kontolemis, Z.G., and Levy, J. (2001) "Monetary policy under EMU: Differences in the transmission mechanism?". *International Monetary Fund Working Paper* 01-102.
13. Dedola, L., and Lippi, F. (2000) "The monetary transmission mechanism: Evidence from the industries of five OECD countries". *Banca d'Italia Temi di Discussione* 389.
14. Ehrmann, M., and Worms, A. (2001) "Interbank lending and monetary policy transmission: Evidence for Germany". *ECB Working Paper Series* 73.
15. Gerlach, S., and Smets, F. (1995) "The monetary transmission mechanism: Evidence from the G-7 countries". *CEPR Discussion Paper* 1219, July 1995.
16. Gibson, M.S. (1997) "The bank lending channel of monetary policy transmission: Evidence from a model of bank behavior that incorporates

long-term customer relationships". *Board of Governors of the Federal Reserve System, International Finance Discussion Paper* 584.
17. Greenwald, B.C., and Stiglitz, J.E. (1993) "Financial market imperfections and business cycles". *The Quarterly Journal of Economics*, 108 (1), p. 77-114.
18. Hortemoller, O. (2003) "Further VAR evidence for the effectiveness of a credit channel in Germany". *Applied Economics Quarterly*, 49 (4), p. 359-381.
19. Hubbard, R.G. (1998) "Capital-market imperfections and investment". *Journal of Economic Literature*, 36 (1), p. 193-225.
20. Hubbard, R.G. (1995) "Is there a credit channel for monetary policy?". *Federal Reserve Bank of St.Louis Review*, 77(3), p. 63-77.
21. Kashyap, A.K., and Stein, J.C. (1994) "The impact of monetary policy on bank balance sheets". *NBER Working Paper* 4821.
22. Kashyap, A.K., and Stein, J.C. (1997) "The role of banks in monetary policy: A survey with implications for the European monetary union". *Federal Reserve Bank of Chicago Economic Perspectives*, 22 (5), 2-18.
23. Kashyap, A.K., and Stein, J.C., and Wilcox, D.W. (1993) "Monetary policy and credit conditions: Evidence from the composition of external finance". *The American Economic Review*, 83 (1), p. 78-98.
24. Kilian, L. (1998) "Small-sample confidence intervals for impulse response functions". *Review of Economics and Statistics*, 80 (2), p. 218-230.
25. King, S. (1986) "Monetary transmission: Through bank loans or bank liabilities?". *Journal of Money, Credit & Banking*, 18 (3), p. 290-303.
26. Leitemo, K., and Soderstrom, U. (2005) "Simple monetary policy rules and exchange rate uncertainty". *Journal of International Money and Finance*, 24 (3), p. 481-507.
27. Oliner, S.D., and Rudebusch, G.D. (1996) "Is there a broad credit channel for monetary policy?" *Federal Reserve Bank of San Francisco Economic Review*, 1, p. 3-13.
28. Peek, J., and Rosengren, E.S. (1995) "Is bank lending important for the transmission of monetary policy?". *Federal Reserve Bank of Boston Conference Series*, 39, pp. 47-68.
29. Ramaswamy, R., and Slok, T. (1998) "The real effects of monetary policy in the European Union: What are the differences?". *IMF Staff Papers*, 45(2), p. 374-396.
30. Runkle, D.E. (1987) "Vector autoregressions and reality". *Journal of Business and Economic Statistics* , 5 (4), p. 437-432.
31. Safaei, J., and Cameron, N.E. (2003) "Credit channel and credit shocks in Canadian macrodynamics – a structural VAR approach". *Applied Financial Economics*, 13 (4), p. 267-277.
32. Sims, C. (1986) "Are forecasting models usable for policy analysis?". *Federal Reserve Bank of Minneapolis Quarterly Review*, 10 (1), p. 2-16.
33. Taylor, J.B. (2001) "The role of the exchange rate in monetary-policy rules". *American Economic Review*, 91 (2), p. 263-267.
34. Vermeulen, P. (2002) "Business fixed investment: Evidence of a financial accelerator in Europe". *Oxford Bulletin of Economics and Statistics*, 64 (3), p. 213-231.

Voluntary Employee Turnover Phenomenon in American MNC Subsidiary in Ghana: Is Pay the Force Behind Employee Exit Decisions?

Nana Yaw Oppong

Faculty of Business and Law, Leeds Metropolitan University, UK, PO Box AX 517, Takoradi, Ghana, nana.oppong@yahoo.com

Concerns have been raised about employee exits in Newmont Ghana Gold Limited probably due to how close the individual exits are – within a period of 14 month, leading to loss of some talented professional employees whose productivity is difficult to match. This study looks into "why the exits". The paper focuses on the evaluation of the company's pay and investigates if it contributes in any way to the turnover phenomenon, using predominantly primary data. The results show that dissatisfaction with pay contributes to the turnover phenomenon but how strong the factor is could not be concluded since no other factors were taken into consideration therefore opening the grounds for further research. Results also revealed the collective nature of the phenomenon and senior management's response, influence by parent company's home country HRM practices. The paper concludes with a four-point recommendation which could be implemented to check the turnover trend.

Keywords. Employee commitment, Newmont Ghana Gold Limited (NGGL), point-factor pay, employee turnover, alternative pay plans

1. Introduction

Turnover is an important organizational phenomenon that has invited close attention for decades (Bartunek, Huang, and Walsh, 2008). Research have revealed that voluntary employee turnover is expensive, and particularly so in the case of talented employees whose productivity is difficult to match (Boxall, Micky and Rasmussen, 2003) and there is therefore the evidence that high levels of voluntary turnover adversely affect business performance. In their study on the relationship between job satisfaction and employee turnover, Tang, Tim, and Tang (2000) said that job satisfaction is an affective response to specific aspects of the job and has a direct effect on turnover as well as an indirect effect through organizational commitment. They continue that job satisfaction and organizational commitment are the most immediate determinants of turnover intentions.

In this paper however, focus will be on pay (extrinsic feature of job satisfaction) and whether it has relationship with the voluntary turnover pattern in Newmont Ghana Gold Limited (NGGL), American multinational subsidiary in Ghana. Commitment in this paper will however not be considered as a determinant but as a result of employee behaviour resulting from their (dis)satisfaction leading to whether to stay (committed to the organization) or quit (voluntarily turn over). I will therefore consider them as two extremes but not one as determinant of the other. Rather, the absence of one means the other.

The three important items to consider in this paper are employee commitment, pay and turnover. These will be considered one after the other, and make a comparative assessment of the NGGL pay plan with other alternative pay systems. The conclusions will therefore be applied to assess whether or not the current pay system is a determining factor of the turnover trend – can the NGGL pay system induce commitment or influence turnover? In discussing turnover, I will try to assess whether the current trend has been individual or collective in nature as each has different organsational impact, most especially on the remaining employees' exit decisions.

2. NGGL Pay System

The NGGL pay system could be described as the Hay pay system or traditional American pay plan. Before the American business was seriously challenged by global competitors in the early 1980s, established base pay was as a result of a point-factor job evaluation, which according to Boyett and Boyett (1998) about 95% of workers in major US companies were paid based on this. NGGL pay system is also based on the point-factor. This point-factor, structured so that pay is attached to the jobs that employees perform (Guthrie, 2000) results from the following considerations.

- Preparation of job description for all positions and each job description rated according to common factors such as working conditions, knowledge required to perform the job, problem solving requirements, and accountability for performance.
- Total points accumulated, based on the applicable requirements, are used to establish salary level for each job i.e. jobs with high point scores are paid more than others with lower point scores;
- The resultant point-factor pay is then combined with annual merit increase that results from individual performance appraisal. According to companies usually budgeted for annual increases that at least equal to what other companies (probably in the same industry) were paying.

In his book *Strategic Pay,* Lawler (1990) revealed that American businesses were pleased with the point-factor pay system for the following reasons:

- It ensured high level of fairness and internal equity as every job in the company was paid based on one system;

- The point-factor approach ensured centralized control as the system ensured easy comparability across a single organization and also to compare jobs across different companies;
- A well-established system that consulting firms, most notably the Hay Group, created large and complex databases that silenced the messy issues of pay, and made the system fit for the description 'blame-shifting' tool. Managers could use the excuse that it wasn't their fault that employees were paid what they have, but the system made them do it;
- Linkage of individual's pay to their performance appraisal meant that outstanding performers were being rewarded and mediocre performers punished.

Boyett and Boyett (1998) assessing the point-factor pay system said that it was suitable for organizations which top executives made all decisions, middle and lower managers were responsible for implementing decisions and workers were made to obey and follow. This system however became less suitable when American companies began to compete globally and needed to become flexible, responsive (to local environments) and innovative (from the traditional pay practices) because the point-factor has some shortfalls as summarized below.

- It encourages people to do just their job descriptions as the system is based on rigid job description.
- It strictly concentrates on what should be done (the job) thereby depersonalizing people who are equated to set of duties and therefore ignoring who they are and what they can do. The job being so important than the person, people could even lie when creating job description to get points required to receive good pay.
- During the time when change becomes crucial, the point-factor system could make restructuring difficult as all job descriptions would have to be re-written and this would most likely result in in-fighting as people would try to maintain their responsibilities, accountabilities and other point-fetching factors.
- The system, linking base bay to individual's performance appraisal is demotivating as appraisal could be highly subjective judgment of supervisors. Boyett and Boyett added that this problem is even of more concern as companies insist that ratings across the organization has to be 'normally' distributed, implying that the only way for some employees to get higher ratings is for others to get lower ratings.

It is beyond the scope of this paper to make assessment of whether or not this pay system is good. The aim of this paper is to consider NGGL pay (that happens to be the point-factor) and to consider its relationship with the company's current turnover trend. The literature could however be used for further research into this type of pay system, generally or as pertains to NGGL.

3. Alternative Pay Plans

Surprisingly, the Hay Group, a 'big name' behind the point-factor system and described as American's premier point-factor consulting firm, realized by the middle of the 1990s that the system was dead in most American companies (Boyett and Boyett, 1998). Flannery, Hofrichter, and Platten (1996) for instance admitted that the vast gap between organization change and compensation required that while some organizations needed to modify their current Hay-type system, others required to create entirely new nonpoint-factor systems. Flannery, Hofrichter, and Platten were vice presidents of the Hay Group and this view coming from Hay Group executives needs be taken a bit more seriously.

Around the time of the Hay vice presidents' assessment of the pay system, many compensation gurus had also realized and advocated that the point-factor system should not only be abandoned by most companies but should also be replaced with skill-based and group-based compensation (Boyett and Boyett, 1998; Guthrie, 2000). The knowledge or skill-based pay considers base pay that is tied to the skills an individual possesses rather than the job he holds. Comparing this pay plan with the point-factor system, Guthrie (2000) said that although there may be pay adjustments (e.g. merit increases) associated with performance, large pay adjustment with the point-factor are only accomplished when employees change jobs. In contrast, the knowledge or skill-based pay system, pay is largely function of the breadth and/or depth of knowledge and skills usually specific to the organisation's operations. The system is therefore designed to promote investment in firm-specific human capital. This affirms the view of Cunningham (2007) that talent is highly influenced by the type of industry and the nature of its work, and advised that talent could best be developed if it is done organization-specific.

Group-base pay plan, on the other hand, is system where payouts are contingent on goals or targets of a group or unit. Citing Hansen (1997) and Weiss (1987) review of relationship between group-based pay plan and retention, Guthrie (2000) wrote that aggregate productivity effects observed were largely due to their effect on lower ability employees while the performance of the most able employees tended to decrease. Guthrie (2000) also concluded in his study to examine the relationships between the use of alternative pay practices and employee turnover rates that, turnover will be higher when group size is larger. This is a function of two factors that:

1. In smaller groups, individuals may perceive a higher likelihood that their individual efforts will affect the group outcome and therefore recognized;

2. Free riding – where group members who do not perform tend to take advantage of the efforts of their performing peers – will occur with greater frequency as group size increases. In group performance pay plan with increasing existence of free riding leads to employee turnover.

Employee Commitment

Some authors also refer to this as organizational commitment (by employee). These two terms are used interchangeably.

Wright and Kehoe (2008) define organization commitment as the strength of an individual's identification with and involvement in a particular organization. Mowday, Porter, and Steers (1979) and Wright and Kehoe (2008) consider organization commitment as of three components: a strong belief in and acceptance of organizational goals and values (identification); a willingness to exert considerable effort on behalf of the organization (involvement); and a strong desire to maintain membership in the organization (loyalty). Allen and Meyer (1990) on their part define organizational commitment as the feeling of obligation to stay with the organization: feelings resulting from the internalization of normative pressures exerted on an individual *prior to entry* or *following entry*. Allen and Meyer (1990) also explained commitment with the identification of three distinct components as follows:

1. Affective commitment – extent to which individual identifies himself with the organization;

2. Continuance commitment – the individual's need to continue working for the organization based on the perceived costs associated with leaving the organization either by personal sacrifice or limited opportunities elsewhere;

3. Normative commitment – commitment that is influenced by society's norms about the extent to which people ought to be committed to the organization.

Allen and Meyer (1990) therefore grouped people wanting to stay with their organizations into those that do so because they want to (affective commitment), because they need to (continuance commitment) or because they feel they ought to (normative commitment).

This is of interest to human resource (HR) practitioners, the assumption being that employee commitment enhances performance and organizational effectiveness. High profile exits can undermine the everyday promise that "people are our most precious asset". Gbadamosi (2003) revealed that when individuals work in organizations they display signs of any of the following two categories of effect or attitude:

1. Those who love their job but hate the organization for which they work; and

2. Those who hate their job but love their employing organization, thereby displaying a strong loyalty towards the organization.

Gbadamosi (2003) tried to identify factors of organizational commitment of employees in Africa, especially managers, and concluded that the practical significance of the factors depend on the extent to which an organization needs to enhance the commitment of its employees through a well-packaged incentive system that would induce commitment. Gbadamosi's assessment is on the premise that Africans, probably due to the high levels of poverty, are generally motivated by economic rewards than intrinsic satisfaction. Mullins (2005) explains economic rewards as extrinsic factors those that can be seen and/or easily quantified such as pay, bonuses, and material goods. These are what people usually look for when they are being employed. Unlike economic/extrinsic reward, intrinsic satisfaction is derived from the work itself. This is difficult to realize in the worker as this is purely personal-orientation to work. This includes such factors as advancement, interest in the job, and job enrichment. These,

Dr. Frederick Hertzberg (Two Factor Theory) considers as motivators – factors that must be put in place to influence one's behavior towards improved job performance (Sachau, 2007). Abraham Maslow (Need Hierarchy Theory) however, sees economic factors as what should be provided before the person will be influenced by the intrinsic factors to give off his best (O'Connor, College, and Yballe, 2007).

Both Maslow and Hertzberg in their respective theories contend that intrinsic factors are the actual motivators but Maslow points out that extrinsic or economic factors are to be satisfied before intrinsic factors can work as motivators. These motivational theories are pointer to the fact that job satisfaction enhances employee commitment. This is confirmed by Mullins (2005) and Torrington, Hall, and Taylor (2005) who make strong link between job satisfaction and motivation but Mullins is quick to reiterate that satisfaction is not the same as motivation and describes job satisfaction as more of an attitude, an internal state that could be associated with a personal feeling of attachment, either qualitative or quantitative. Mumford (1991) examines job satisfaction in the following two ways.

1. In terms of the fit between what the organization requires and what the employee is seeking; and

2. In terms of the fit between what the employee is seeking and what he is actually receiving.

Managing commitment is therefore an important HR task since commitment improves work performance and reduces turnover of employees in organizations. Explaining why, Gbadamosi (2003) said that a satisfied and committed work force translates into higher productivity due to fewer disruptions caused by absenteeism or the loss of good employees.

4. Employee Turnover

Employee turnover is a situation where employees leave the services of the company (not committed). Turnover is an important organizational phenomenon that has invited close attention for decades. Review of literature for this paper has revealed extensive research on turnover by Boxall, Micky and Rasmussen (2003); Boyett and Boyett (1998); Guthrie (2000); Mitchell et al. (2001); Mowday, Porter, and Steers (1982); Tang, Tim, and Tang (2000); and Wright and Kehoe (2008). Tang, Tim, and Tang (2000) in their study on the relationship between intrinsic job satisfaction and voluntary turnover also disclosed that other authors had done extensive work on employee turnover.

Generally, employee turnover is seen as extremely costly and most employers are better served with lower employee turnover than a higher one. However, are all employee turnovers costly to employers? Guthrie (2000) placed employee turnover into two categories – voluntary and involuntary. Voluntary turnover is controllable i.e. employers would like to prevent if they want to, and actually can do so. Involuntary on the other hand is one that can not be controlled such as death (natural cause) and retirement (legal requirement). Any of these two categories could also be either functional i.e. beneficial to the organization; or dysfunctional i.e. harmful to the

organization. This is a pointer to the fact that not all employee turnovers that employers view as harmful or costly but much will have to be done when turnovers appear to be harmful or costly.

For the purpose of this paper focus will be on voluntary turnover (employees who leave for other jobs, especially) and at the same time dysfunctional (that seriously affects the organization) so that those that are rather beneficial to NGGL will not be necessary to attempt to control. I would therefore like to term the type of employee turnover to be considered in this paper as *dysfunctional voluntary turnover.*

The various studies considered generally reveal that dissatisfaction may 'push' the employees to look for alternative employment (Mitchell et al., 2001), whereas the perception of attractive alternative job opportunities may 'pull' them to consider alternative employment (Tang, Tim, and Tang, 2000).

The general premise of research on turnover is that it is an individual phenomenon because individuals independently decide to leave their organizations, regardless of whether antecedents to their decision are individual or organizational. However in recent years, as identified by Bartunek, Huang, and Walsh (2008), recognition that focuses solely on the experiences of individuals ignores other crucial levels of analysis, especially the group level – collective turnover, which has more severe consequences for organizations than does individual turnover. Collective turnover is the turnover of two or more organizational members in close temporal proximity based on shared social processes and decisions to leave an organization. The phases or interaction processes leading to collective turnover have been identified as presented in Figure 1 and three-point explanation below.

1. Two or more members of organizational group experience unresolved dissatisfaction with some aspect of their organization despite complaints to supervisors/managers (exercise of voice);

2. These individuals' negative experiences are transformed into shared group perceptions through group sensemaking and emotional contagion that occurs in cohesive groups that view themselves as superior to others';

3. When such group members feel no hope of resolvability but have other opportunities available and few constraints on leaving, they collectively decide to leave and thus start looking for placements and eventually leave (turn over). At this stage the escalating interaction process moves from exercise of voice to exit tactics. The possibility of turnover here is thus characterized by the availability of external alternatives and the lack of constraints on leaving, representing the boundary conditions of the process (as shown in Figure 1).

5. Methodology

The aim of the study is to find out what influence employees exit decisions in an American MNC subsidiary in a foreign land. In this light, the American HRM practices are explored to more deeply understand how HRM systems influence commitment (Wright and Kehoe, 2008). Researchers (Anakwe, 2002; Bjorkman and Budhwar, 2007; Gardiner, 1998; and Jackson, 2002) conclude that while MNCs policies and

practices influence how the African employee works, the employee's cultural values also impact on how he does his work. Relating to specific country HRM practices, Royle (2004) found out that American MNCs have strong influence in determining the 'one best way' of conducting business, which confirms Tayeb's (2005) conclusion that Americans are domineering in their HRM practices, and indeed in many other aspects of their management in relation to their foreign subsidiaries.

The American national culture characteristic is individualism, placing value on freedom and belief that individuals can shape and control their own destinies. People pursue mainly those of the immediate family, and their in-group does not include commitment to the workplace (Chew, 2004). Management on their part, according to Collings (2003) emphasize individual's progression by moving from one company to another in pursuit of success.

Data for analysis is largely from primary sources – from Ellipse (specialized computer software used by NGGL for employee data management); and face-to-face interviews with 12 serving senior and managerial staff. Reference was also made to exit interview notes from personal files of resigned employees within the period of study. Focus however was on employees' responses to the questions on 'pay' and their relationship with their exit decisions. Categories of employees were limited to senior and management staff. Although exit interview notes analysed related as well to resigned senior and management staff from other departments, turnover trends in the finance and mining departments alone were considered.

6. Data Analysis

To ascertain the turnover situation, resignations from the finance and mining departments were considered, although other departments also had turnover rates that call for attention. Focus is on these two departments because employees in each of these departments exhibit similar relevant skills and professional backgrounds that may induce similar behavioural patterns towards compensation and exit decisions. Source of this data is Ellipse generated resignations which were cross-checked with archived resignation files.

Table 1 shows national senior and management staff resignations from the Finance department from November 2006 to February 2008 (period of the study). Two striking developments could be identified here. There were four resignations between 19 December 2006 and 3 January 2007, a period of about three weeks. Also, between 3 December 2007 and 4 February 2008 three senior positions as Assistant Site Controller, Senior Financial Analyst, and Senior Accountant all resigned.

Table 2 paints a rather shocking picture of turnover in the mining department. Eleven (11) senior and management staff resigned between 8 January 2007 and 8 September 2007, a period of nine months. A breakdown of the trend further depicts how close to each other the exits were. Three (3) Mining Engineers resigned from 20 July 2007 to 3 August 2007, a period of three weeks, while between 25 May 2007 and 4 July 2007 five employees (including four Production Geologists) tendered in their resignations.

To find out if salary/pay is a major contributory factor to the turnover trend, interview notes of twelve (12) resigned senior and management staff were analysed. I collated responses to the question on salary - *Salary: fair for work done? Compare with salaries elsewhere? Increase keyed to performance?* – and assessed if responses related in anyway to the departing employees' pay (Table 3). The responses, as written down by the employees themselves, are itemized in the first column and assessment of the individual responses whether or not they are pay-related (Yes/No) also shown in the second column. Out of the 12 respondents, 11 could be said to have some levels of dissatisfaction about NGGL pay. In item 3, the respondent reveals the approximate salary to receive from his/her prospective employers, if he finally decides to join them from NGGL. With regard to item 7 respondent admits that NGGL salary is favourable but would like more to be paid to compensate for the remoteness of the mine, showing that employees could also leave for opportunities in the cities but not necessarily for improved pay. Deducing from respondent number 11, the 'unfairness' of the salary was raised even before accepting the offer of employment. This could mean that this person started looking for his next job from his first day in NGGL. It is therefore not a surprise that this employee resigned in his sixth month.

Questions in Tables 4 and 5 were considered for analysis not because the responses were expected to reveal relationship with pay but because of their perceived likelihood to reveal some turnover and commitment management hints. Unlike responses in Table 3, Tables 4 and 5 responses are not up to 12 because not all employees answered all the questions but they all answered the question in Table 3.

Table 4 question reveals general reasons for turnover, some of which match the findings from my literature review. Table 5 question is also important for the turnover/commitment management because they may reveal what employees expected from NGGL that they couldn't receive but could be present in their new or prospective employments.

I had interviews with a total of 12 senior and management staff. The interviews were however not restricted to only the mining and finance departments though more than a half came from the two departments. The interviews were not structured and interviewees were not prompted of the purpose of the interviews. Interviewees were engaged informally, usually in the form of chats. Responses/comments were memorized and salient points jotted down upon return to the office. As purpose of interviews was not disclosed to interviewees employees expressed opinions without fear of being quoted or intimidated. Table 6 shows summary of these responses/concerns.

7. Discussions and conclusions

- Considering the proximity of exits in the sample departments used in this research, NGGL turnover is mostly collective in nature. This, as described by Bartunek, Huang, and Walsh (2008), has very serious repercussions on the manpower position of any organization that finds itself in similar situation as talented professionals are seen through the exit doors, probably at a faster rate

than they could be replaced. NGGL turnover described to suit collective turnover possibly should be of more serious concern to all stakeholders.

- The point-factor pay system, based on rigid job description and encourages people to do just their job descriptions, has its own retention mechanism. It has the power to retain those who were hired by virtue of their ability to perform the prescribed job without necessarily having skills that could allow them move to other companies that pride skills and knowledge. This however does not work to retain professional/technical employees who would not only be paid for what is done, but what is possessed and how to use it. This suggests the point-factor is not necessarily inappropriate but may not be a suitable option to retain senior and management staff, target employees of this study.

- The point-factor pay does not go beyond antecedent behaviour to enhance consequence behaviour of employees. This consideration is vital because all management ultimately comes down to managing behaviour, since it is through behaviour of people that all things get done (Wright and Kehoe, 2008). There are two ways of influencing behaviour: *antecedent* and *consequence*. Antecedent is managing performance by telling other people what to do or not to do. Relating to pay, this could coincide with the point-factor system that prescribes just what to do without much employee initiative, as use of acquired skills and knowledge is not emphasized. With antecedent behaviour therefore, employees (usually professionals) change their behaviour temporarily – behaviour is not sustainable. A better way of influencing behaviour however is consequence which, in addition to providing an antecedent to get behaviour started, employers can also provide something after the behavior occurs to increase or decrease the probability that it will be repeated – to sustain the behaviour. This could be in the form of performance related rewards and progression based on skills and knowledge. The study has shown lack of this as the NGGL's reward system is based on rigid job description and encourage people to do just their jobs without regard for skills and knowledge as evidenced in item 3 in Table 6.

- By design, knowledge or skill-based pay system encourages and rewards employees to increase their depth of knowledge and skills related to their employers' particular function of operation. This enhances the value of uniqueness of employees to their employer and always feel valued, situation likely to enhance consequence behaviour (Boyett and Boyett, 1998). Since the skill-based pay is designed to facilitate company-specific sets of skills, it is most unlikely that other companies will find the particular array of skills acquired by employees valuable. The acquisition of company-specific skills represents investment because employees are worth less to other companies. This makes the company's skilled employees less attractive to other companies or even to those in the same industry if skills development is well planned and implemented. The NGGL point-factor pay system defeats this and job attitudes of the target employees combine with job alternatives to predict indents to leave, which are direct antecedent of turnover (Mitchell et al., 2001).

- Pay, resulting from the point-factor system could be concluded as a contributory factor to the NGGL turnover phenomenon. As resulted from research by Guthrie (2000), large adjustment with the point-factor are only accomplished when employees change jobs. This assessment of Guthrie is probably based on the fact that pay (attached to the job) is already determined and limits of merit increases also defined by the system; and these make substantial increases clearly difficult. Once an employee realizes that he cannot get what his professional skills and knowledge could get him because of the dictates of the system, employees leave in search of higher pay – if external opportunities are available and they are not constrained from leaving (Mitchell et al., 2001). This is confirmed by the exit interview notes analyzed, specifically the question relating to pay (Table 3) where more than 90% of respondents reveal various levels of dissatisfaction (Mitchell et al., 2001). Some responses to the other two questions also reveal dissatisfaction with pay. For instance, response number 1 in Table 5, *overhaul of salary structure could help retain some calibre of professionals* is a strong caution.

8. Recommendations

Since this paper is basically on employee turnover, recommendations concentrate on how to improve retention but not to attack or praise the NGGL pay plan, though the total compensation plan could be made to respond to the local (Ghanaian) environment (another investigative paper could take care of this).

8. 1 Exit interviews

Efforts to prevent the exit of talented employees may be too late, but a well thought-out exit interview process may help prevent similar departures. With the company's month-long notice period, exit interview could be conducted during the second or third week of the notice period, when the emotion of resigning has died down but before the final week when people are either very busy handing over of actively disengaging with the company. As talent shortages bite, the right approach to exit interviews can enable companies to make the right organizational development interventions to retain their most valued employees and pre-empt damaging departures. This can be part of the company's talent management challenge and to ensure that exit data informs organizational development and retention strategy, it must feed into other organizational processes such as employee engagement surveys.

CIPD (2007) reveals that new hires are most committed during their first few months on the job but satisfaction level starts declining after the sixth month. The research further reveals that by the end of their second year, instead of focusing on the job, employees are more likely to be searching for a new one. Talent retention plans are therefore necessary from when the individual is engaged, including trying to identify what could lead to diminishing satisfaction especially during periods of induction and probation.

The talent retention team could have periodic meetings to review the exit data from across the company and the result used to shape retention and induction policies. Feedback, especially to line managers, will be necessary to allow them take pre-emptive action for retention.

8.2 Employee Engagement Survey

Not only exit interviews that could help control further employee turnover but also employee engagement survey, another organizational process. It will be beneficial for NGGL to embark on a survey of other companies, mostly in the mining industry (probably through consultants) to understand what are used to bait professional/technical employees from NGGL. Many concepts measured in an engagement survey could be turned into questions in exit interviews. After all, what attract people to other companies are the same factors that (when absent) repel them from their current employers. For instance, unofficial reports from some employees had it that three of the finance department employees (could be more) who left in 2007 joined the Chirano Gold Mine. What makes Chirano attractive?

Using the rivalry principle to explain design of HR planning process, Boxhall and Purcell (2008) reveal that labour markets are competitive and intelligent rivals will attempt to recruit the best workers and build the best management process. This suggests that the company's management team should attempt to understand the HR strengths and weaknesses of key competitors. To achieve this, NGGL needs to have good data on the employment strengths and weaknesses of rivals in the industry and also to improve conditions, the better options of which attract the company's professionals to other companies. Boxhall and Purcell (2008) advise something similar as they are of the view that whereas knowing competitors' strengths and weaknesses is a competitive question, it should not only be seen as such but should be noted that although rival firms may be competing in the labour market they also have common interest in improving labour facilities in many industries.

8.3 HR Business Partner Role

Investigations show that dissatisfaction on the part of two or more individuals who are part of an organizational group might trigger collective turnover if not addressed appropriately, and sometimes promptly. Managers are therefore cautioned to accord particular attention to most of the issues regarding their relationships with employees. Managers have the responsibility to try to understand the reasons behind individual employees' dissatisfaction and pay attention to them. When attention is paid to employees' dissatisfaction, managers should also know that individuals may impact other organizational members decision-making about turnover because employees are always in some kind of social relationship in organizations and their perceptions and behaviours similarly influence other members of their cohesive groups. This means that how managers act or react to employees' dissatisfaction may have implications for group as well as individual responses. For instance, research by Krackhardt and Porter (1985) as cited by Bartunek, Huang, and Walsh (2008) explored how the departure of some employees influenced the attitudes of the employees remaining in an

organization based on their relational embeddedness. Employees thus perceive similarity between themselves and departing workers as another antecedent factor influencing individual turnover decisions.

For line managers to efficiently perform this role, NGGL HR will have to play business partner role (Francis and Keegan, 2006) by partnering with line managers to reach their goals through effective turnover control strategy formulation and strategy execution. This will not only mean HR professional support for line managers but also the assurance that HR has not relocated its role to line managers who neither have the time nor the training to give this HR function the priority it requires (Francis and Keegan, 2006).

8.4 Strategic Approach

The situation may require regional senior management involvement or even global corporate strategy. To this point there seems to be little concern from senior management, as what is done so far about the phenomenon is more of employee champion role rather than strategic one. Or is it because it is less significant considering the proportion in the light of the whole Newmont Mining Corporation (the parent company)? At a 28 January 2008 meeting of Ahafo Mine management staff to introduce the new Regional Vice President of African Operations, Jeff Huspeni, a management staff wanted to know the incoming Vice President's plans about the turnover trend. The outgoing Vice President, Gordon Nixon replied that the phenomenon was not unique to NGGL but a worldwide development and added that workers all over the world are playing leapfrog game. Jeff, on his part, skillfully declared that it's a challenge for the whole organization, without a hint on the way forward. This may suggest the senior management's blithe attitude towards the situation, portraying an American HRM practice that emphasize individual's freedom to pursue one's career progression by moving from one company to another in pursuit of success.

Once identified as a global issue becomes a strategic issue if the company is to redesign this to its competitive advantage. There is cause to worry because the issue casts slur on HR department's efforts to recruit the right people (that are committed) and to retain them. It also stains the image of the company that it does not uphold the global business promise that "people are our most precious asset", if seeing them leaving is the least thing that could bother the company. This is more of (strategic) talent management issue.

References

1. Allen, N. and Meyer, J. (1990) 'The measurement and antecedents of affirmative, continuance and normative commitment to the organization', *Journal of Occupational Psychology*, Vol. 63, pp.1-18.
2. Anakwe, U.P. (2002) 'Human resource management practices in Nigeria: challenges and insights', *International Journal of Human Resource Management,* Vol. 13 No. 7, pp.1042-1059.

3. Bartunek, J. M., Huang, Z. and Walsh, I. J. (2008) 'The development of a process model of collective turnover', *Human Relations*, Vol. 61 No. 1, pp. 5-38.
4. Bjorkman, I. and Budhwar, R. (2007) 'When in Rome? Human resource management and the performance of foreign firms operating in India', *Employee Relations*, Vol. 29 No.6, pp.595-610.
5. Boxall, P., Macky,, K. and Rasmussen, E. (2003) 'Labour turnover and retention in New Zealand: the causes and consequences of leaving and staying with employers', *Asia Pacific Journal of Human Resources*, Vol. 41 No. 2, pp.195-214
6. Boyett, J. H. and Boyett, J. T. (1998) *The Guru Guide: The Best Ideas of the Top Management Thinkers*, Denver, John Wiley & Sons, Inc.
7. Collings, D. G. (2003) 'HRD and labour market practices in a US multinational subsidiary: the impact of global and local influences', *Journal of European Industrial Training*, Vol. 27, No. 2, pp.188-200.
8. Cunningham, I. (2007) 'Talent management: making it real', *Development and Learning in Organisations*, Vol. 21, No. 2, pp.4-6.
9. Flannery, T. P., Hofrichter, D. A. and Platten, P. E. (1996) *People, Performance and Pay*, New York, Free Press.
10. Francis, H. and Keegan, A. (2006) 'The changing face of HRM: in search of balance', *Human Resource Management Journal*, Vol. 16, No. 3, pp. 231-249.
11. Gardiner, K. M. (1998) 'Managing in different cultures: the case of Ghana', *In* Towers, B. (ed.) *The Handbook of Human Resource Management, (pp.27-49),* 2nd Ed., Oxford, Blackwell Publishing Ltd.
12. Gbadamosi, G. (2003) 'HRM and the commitment rhetoric: challenges for Africa', *Management Decision*, Vol. 41, No. 3, pp.274-280.
13. Guthrie, J. P. (2000) 'Alternative pay practices and employee turnover: an organization economics perspective', *Group & Organization Management*, Vol. 25, No. 4, pp.419-439.
14. Jackson, T. (2002) 'Reframing human resource management in Africa: a cross-cultural perspective', *International Journal of Human Resource Management*, Vol. 13, No. 7, pp.998-1018
15. Lawler, E. E. (1990) *Strategic Pay*, San Francisco, Jossey-Bass
16. Mitchell, T.R., Holdom, B.C., Lee, T.W., Sablynski, C.J. and Erez, M. (2001) 'Why people stay: using job embeddedness to predict voluntary turnover', *Academy of Management Journal,* Vol. 44, No. 6, pp.1102-1121.
17. Mowday, R. T., Porter, L. W. and Steers, R. M. (1982) Employee organizational linkages: the psychology of commitment, absenteeism, and turnover, New York, Academic Press.
18. Mullins (2005) *Management and Organisational Behaviour*, 7th Ed., Harlow, Pearson Education Limited.
19. Mumford, E. (1991) 'Job satisfaction: a method of analysis', Personnel Review, Vol. 20, No. 3, pp.11-19.
20. O'Connor, D., College, M. and Yballe, L. (2007) 'Maslow revisited: constructing a road map of human nature', Journal of Management Education, Vol. 31, No. 6, pp.738-756.

21. Royle, T. (2004) 'Employment practices of multinationals in the Spanish and German quick-food sectors: low-road convergence?', *European Journal of Industrial Relations*, Vol. 10, No. 1, pp.51-71
22. Sachau, D. A. C. (2007) 'Resurrecting the motivation–hygiene theory: Hertzberg and the positive psychology movement', Human Resource Development Review, Vol. 6, No. 4, pp.377-393.
23. Tang, T. L., Kim, T. J. and Tang, D. S. (2000) 'Does attitude toward money moderate the relationship between intrinsic job satisfaction and voluntary turnover?', *Human Relations*, Vol. 53, No. 3, pp.213-245.
1.
24. Tayeb, H.M. (2005) International human resource management: a MNC perspective, Oxford, Oxford University Press,
25. Torrington, D., Hall, L. and Taylor, S. (2005) *Human Resource Management*, 6th Ed., Harlow, Pearson Education Limited.
26. Wright P.M. and Kehoe, R.R. (2008) 'Human resource practices and organizational commitment: a deeper examination', *Asia Pacific Journal of Human Resources*, Vol. 46, No. 1, pp.6-19.

Appendix

Figures and Tables

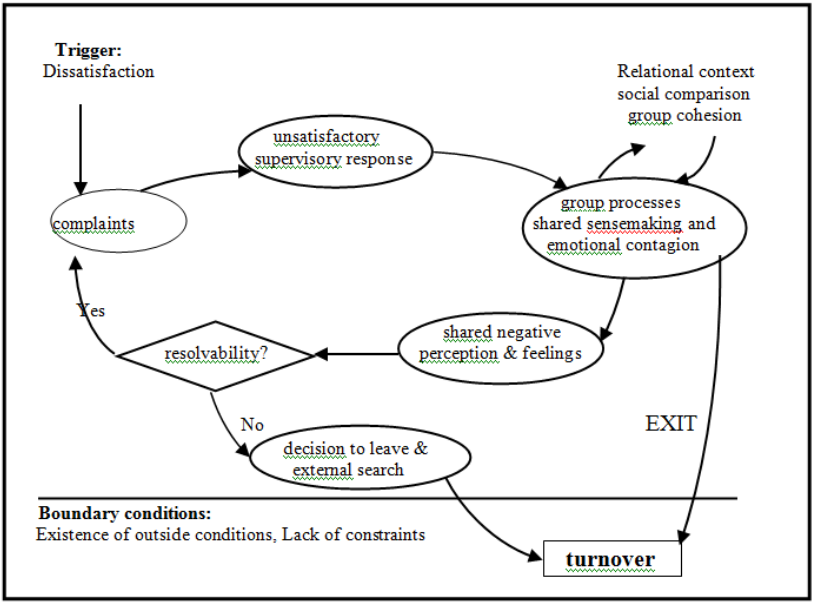


Figure 1. Process model of collective turnover

Source: Adapted from Bartunek, Huang, and Walsh, 2008

Position	Effective Date
Senior Analyst – Finance Systems	1 Nov 2006
Assistant Accountant Payables	19 Dec 2006
Senior Financial Analyst	22 Dec 2006
Account Payable Accountant	23 Dec 2006
Accountant - Payroll	03 Jan 2007
Financial Reporting Supervisor	28 Apr 2007
Manager – Management Accounting	5 Jul 2007
Manager – Capital and Business Planning	19 Jul 2007
Senior Financial Analyst	30 Dec 2007
Assistant Site Controller	01 Feb 2008
Senior Accountant	04 Feb 2008

Table 1: Resignations – Finance Department

Position	Effective Date
General Foreman – Load and Haul	08 Jan 2007
Survey Supervisor	15 Feb 2007
Production Geologist	25 May 2007
Project Geologist	20 Jun 2007
Mining Engineer III	20 Jun 2007
Project Geologist	22 Jun 2007
Project Geologist	04 Jul 2007
Mining Engineer	20 Jul 2007
Mining Engineer	24 Jul 2007
Senior Mining Engineer – Long Term Planning	03 Aug 2007
Shift Supervisor IV	08 Sep 2007

Table 2: Resignations – Mining Department

Question: Salary: fair for work done? Compare with salaries elsewhere? Increased keyed to performance?

Response	Pay Related?
Not fair for work done compared with salaries elsewhere. Much higher salaries elsewhere	Yes
Not really, as many other places have higher salaries. No, and this is the reason for job change – to higher paying company	Yes
Salary in NGGL needs to be adjusted to be competitive. I will receive about twice of NGGL salary if I accept to work with SRL, Sierra Leone	Yes
No, no. Salary structure should be reviewed	
Salary not commensurate with job done and position held	Yes
No. about three times less	Yes
Salaries are favourable but more needs to be done considering the location of the mine and other locations in Ghana for mining. Mining settlements are far from Kenyasi	Yes No
Compared to salary elsewhere, I was not well paid	Yes
No – believe should have been higher. Other areas pay more but I am not leaving because of that	Yes
Under paid, but there is room for increase in remuneration if I had stayed for long	Yes
Not fair compared with this position in other mining companies. It is far below expectation as indicated before acceptance of offer	Yes
Not fair for work done compared with salaries elsewhere	Yes

Table 3: Responses to exit interview question relating to 'pay'

Question: Summarize the major reasons for the separation

Was not satisfied with the role I am playing
Better opportunities out there
There is no uniformity in grades, JDs, salaries etc. Low salaries as compared to some similar mining companies. Official duty vehicle to facilitate work
I want to work outside Ghana and for international exposure. Want to pursue an MBA program
I am leaving due to family reasons. NGGL is a great place to work due to the training opportunities given to its employees. However, the salary structure should be reviewed
I think it's time I look elsewhere for the future of my family. There isn't any problem with the company but personal issues
To seek better offer elsewhere
Personal circumstance
External offer
No career progression

Table 4: Responses to exit interview question relating to major turnover

Question: What action by the company might have prevented the separation?

Overhaul of salary structure could help retain some calibre of professionals
Salary of about 12 million to 15 million Ghanaian cedis a month. Developmental and succession plans for advancement; official duty vehicle to facilitate work
Development/succession plan in place
My separation is purely on family ground and I feel nothing should prevent me from staying back in NGGL
I cannot think of any
Salary increase
Job training, recognition, satisfaction for others
If about a year ago when I indicated I had problems an option had been given to relocate to Accra, perhaps the separation could have been avoided
N/A
Develop potential in people to take up higher responsibilities when they come up

Table 5: Responses to exit interview question relating to how to control turnover

Senior management seem unconcerned about he phenomenon
If conditions of service are not improved most of us will leave , leaving the work to the few expatriate staff
We will continue to be dissatisfied and look elsewhere if qualifications and experience are not recognized and rewarded accordingly;
Many of us have regretted leaving our employers in the industry to join NGGL as promises made are not fulfilled
With reputation of the company and perceived job security it is difficult to understand why people leaving to companies some of whom have bleak future.
Some expatriate managers need to be replaced if key professionals in their departments or those yet to join are to remain.
Trying to import alien American HR and management practices make work rules so rigid, difficult and unfamiliar.
That makes the younger and less experienced professionals to get promoted and grow, though they risk getting experienced ones to learn from.
Why continued training while people continue to leave?

Table 6: Summary of responses/comments from senior/management staff interviewed

Revisiting Social Entrepreneurship in Croatia

Tina Lee Odinsky-Zec

Zagreb School of Economics and Management, Zagreb, Croatia

Purpose – This paper appraises the concept of Social Entrepreneurship (SE) in a Croatian context. The study examines the promotion and support levels of the nomenclature in a transistional economy.
Design/methodology/approach – Uses a combination of qualitative and quantitative methods including literature review, pre-existing case studies, follow up interviews and surveys.
Findings – SE does continue to exist amongst the participants in both active and nascent forms but requires more promotion to develop further.
Research limitations/implications – Limited only to the philanthropic community. This may not be representative of other SE activity occuring in other sectors.
Practical Implications – Proponents of social entrepreneurship must caution that they do not overlook smaller market SE activity as it is at the grassroots level that SE builds its momentum to affect greater change.
Originality / value – While case studies are useful tools providing a snapshot of activity, they also beg to be periodically revisited to provide an elongated view.

Keywords – social entrepreneurship, non-profit, development, Croatia

1. Introduction

Social Entrepreneurship (SE) as an academic discipline and scientific area of study has notably experienced explosive growth. SE activities can be plotted on a continuum ranging from profit to purpose. In particular, research suggests that SE occurs in all types of organizational structures in public, private and social economy enterprises. Equally impressive has been the rallying call from a wide variety of sectors to propel the catch phrase. The status of the social entrepreneur has been celebrated in such a way that it has turned the previously unsung hero into the heralded, "new hero[ine]" in many cultures (Oregon Public Broadcasting, 2005).

However, the imprint of SE is more of a patchwork across the globe rather evenly developed. This paper takes a decidely contextual approach in revisiting how social entrepreneurship has taken root in Croatia. In terms of revisiting, we will revisit people, places and ideas. First we look at Croatia and its place within greater Europe. We examine the literature of social entrepreneurship searching for fundamentals, frameworks and forces of expansion and contraction. Lastly, we will have a 2009

check-up report on enterprising nonprofit organizations recognized for their SE activities in 2006.

Moss, Lumpkin and Short (2008) suggest the correlation between the frequency of SE research in a particular nation and its acceptance. If we buy into their assumption then Croatia will certainly not be seen as wholly embracing social entrepreneurship yet. This particular paper will not rank Croatia's standing but will seek insight into the status of social entrepreneurship in this emerging market country.

2. Croatia: A Social and Entrepreneurial Synopsis

Croatia's popularized image as a former war torn country based on its fight and victory for independence from 1991 to 1995 has taken over a decade and a half to shake off. This external view brought with it an influx of support and aid in rebuilding itself. Now the nation has repositioned from a Balkan state to one that stradles both the Mediterrean and Central Europe and has regained some of its luster as a top tourist destination. This gross oversimplification is ironically important as it marks, on the one hand the development of civil society organizations in the country, as well as the possibility to build on some of the country's inherent qualities to produce authentic and sustainable initiatives for long term stability in the region.

Gojko Bezovan, one the leaders of the civil society movement of Croatia and a law professor in Zagreb stated in (2001,p.1) "There is limited tradition of civil society in Croatia — its development has been hindered by half a century of communism and totalitarian ideology coupled with a lack of experience with the concept of freedom of association. Citizens' civil engagement, for solving both individual and community problems has not been a common practice among the vast majority of citizens in Croatia. Most citizens consider the government/state responsible for solving their problems. Thus, passiveness and apathy exists in Croatia." While Dr. Bezovan and others have worked to change the future, this illustrates the challenges Croatian society faces internally. The Center for Development of Nonprofit Organizations also known as CERANEO is a research and action group which began addressing these issues in the late 1990's (interview with Mladen Ivanović in March 2009). Today, the registry of associations in Croatia has over 25,000 organizations (up from 20,000 in 2006) representing a wide range of social issues including but not limited to addictions, animals, children, culture, diseases, environment, handicapped, human rights, natural disasters, sports and others. There are several subdirectories which are promoting knowledge and resource exchanges among nonprofits as well as to solicit donations and more entrepreneurial forms of self-financing.

Since 1999 the Global Entrepreneurship Monitor (GEM) has been reporting on activities and as of the latest study 43 countries were being evaluated on common variables. From 2007 to 2008 the study reclassified country groups from income level to economic drivers. Croatia has been a part of the research since 2002. While it is seen as a country in transition from an efficiency-driven economy to an innovation-driven economy, it shares that discintion with only one other country, Hungary (Bosma et al, 2008). In the latest GEM report, while Croatian's are more likely to consider

entrepreneurship as a desireable career choice (70%) only one in ten are expected to start a business in the next three years.

The country is divided into three statistical areas which encompass over 20 entrepreneurial zones. Currently, The government has put a priority on stimulating entrepreneurship with the establishment of a one-stop shop for venture creation modelled after Ireland called HITRO.hr opened in 2005. Despite this and other attempts to reduce the bureacracy in starting and operating enterprises, it still takes 40 days to open a new business and that lags behind the world average of 38 (The Heritage Foundation, 2009). The Index of Economic Freedom for Croatia concludes for the second year in a row that Small and Medium Enterprises (SME) activity is curtailed by the opacity of administration and regulations at both the national and especially at the local level (The Heritage Foundation, 2008 as cited in Odinsky-Zec and Stubbs, 2009). Entrepreneurs face bureaucratic hurdles, a deficit of appropriate financial services, a shortfall of (non-financial) business support services, and want of a well-developed entrepreneurial culture (Salzano, 2002 as cited in Odinsky-Zec and Stubbs).

According to Keser (2008) the Republic of Croatia's Ministry of Economy strategies to develop overall entrepreneurship are of importance to national competitiveness. Towards those ends the government adopted a strategy from 2008 to 2012 in which they aim to rectify some of the pressing issues by encouraging SME development through incentivising an entrepreneurial environment, education, infrastructure, grants, favourable financing and guarantees. According to Morrison (2000, p.62-63) "the entrepreneur is motivated to create a venture, which reflects their vision and ambitions, and is prepared to review and reorganise their social environment to make it materialise." As such, the path for each entrepreneur can be either stimulated or deterred by the socio-cultural context in which they choose to manifest their ventures. The Croatian landscape presents many challenges for both social and entrepreneurial development.

3. Fundamentals: SE Definitions, the in-betweens

As is the case with entrepreneurship, social entrepreneurship defies a conclusive definition. As it means different things to different people, we selected four working definitions that had potential for possible acceptance in a Croatian context. We selected our definitions from Romania, Spain, the United States and the European Union. The logic behind each choice precedes each definition.

Romania which entered the European Union at the start of 2007 offers a model of a new member state, a status which Croatia hopes to achieve in the near future. In the context of Romania, Bibu and Orhei (2008) acknowledge the hybrid nature of social entrepreneurship combining a social mission with self-financing schemes. However, they also suggest that a redefining of social enterprises is needed to foster growth in Romanian society. The discourse in Romania currently focuses on nonprofit entrepreneurship rather than the more encompassing, social entrepreneurship (Bibu and Orhei, 2008).

The definition representing Spain is important as one of the authors of the definition has been designated as an early pioneer of the academic discipline of social entrepreneurship in Europe. In 2007, Johanna Mair became the recipient of the Ashoka Award for Social Entrepreneurship Education as a groundbreaking teacher and researcher who actively focuses on social and environmental solutions (The Aspen Institute, 2009). Mair and Marti (2006, p. 37) view social entrepreurship as "a process involving the innovative use and combination of resources to pursue opportunities to catalyze social change and/or address social needs."

The United States Small Business Administration (2007, p. 166) conducted comprehensive research tracing down the evolution of the definition of social entrepreneurship over a decade starting in the early to mid-1990's and settled on their own version "the practice of responding to market failures with transformative, financially sustainable innovations aimed at social problems.

The European Union has almost as many definitions of social enterprise and social entrepreneurship as it has member states. However research is now moving forward and there are several studies and organizations working to harmonise a definition (Heckl et al., 2007). In the interim, the European Commission, DG Enterprise & Industry (2009) address the issue with their temporal version: "social entrepreneurship is positioned between the traditional private and public sectors and combine societal purpose with the entrepreneurial spirit of the private sector."

Regardless of the place, the common denominator rests in the in-between as SE exhibits qualities of private, nonprofit and public sectors (Roper and Cheney, 2005). At once, leading to both confusion and creation of new persepectives and unexpected interpretations. "But with the melting of sector divisions, nonprofits, governments and businesses are blending sources and models of funding to create sustainable, and sometimes even profitable, social innovations (Phills Jr., Deiglmeier and Miller, 2008, p.41)."

A preliminary survey was used to test out these definitions among members active in the non-profit and entrepreneurship sectors in Croatia. Out of a small judgement sample of 55 respondents, almost 71% (39) respondents agreed with the EU definition that social entrepreneurship is positioned between the traditional private and public sectors and combine societal purpose with the entrepreneurial spirit of the private sector.

4. Framework: Towards A Conducive Environment

There are several countries specifically mentioned as models for social entrepreneurship development in Europe. The United Kingdom, Italy and Finland have all set up clear regulations that recognize and support social enterprises in their quest to compensate for shortfalls in society through innovation (Galera, 2006). The Croatian government did in fact develop a strategy for the development of social economy and nonprofit entrepreneurship as part of its 2007-2011 National Operational Strategy Plan for the Creation of a Seminal Framework on the Potential Development of Civil Society (Government of the Republic of Croatia, 2006). The contents of

which outline goals, measures and timelines that are certainly benchmarked as no less than 10 different European countries are mentioned throughout.

As discussed in Ivanovic's (2006) presentation, there remained barriers in legal frameworks, goverment lack of recognition of benefit potential, lack of incentives and support. As a follow up Mr. Ivanović personally updated his report and met with the author to discuss the progress to date. There have been some modest rule changes but most conditions in fact have not changed and limit the scope of activities to a very narrow range. The current sets of policies are rigid in wording and formal in procedure and thus there is a mix of progress. He did mention numerous organizations and individuals actively building overall capacities of Civil Society Organizations (CSO) and Non governmental organizations (NGO).

Civil society organisations utilise the social entrepreneurship model to help them strategically address finanacial independence and a sustainable future (Todres et al., 2006). In light of the economic crisis, it becomes even more relevant to have flexible models for social entrepreneurial activity. However, Croatian NGOs and CSOs have restrictions on revenue which is a barrier rather than a stimulant to development of the sector. The justification being that lawmakers in Croatia and other countries must avoid "grey areas" of business, societies and institutions. One such example of a best practices model is the UK's Community Interest Company (CIC) which is a new legal form allowing voluntary organisations to register as a company with limited liability, generating profit for the benefit of the community (Todres et al., 2006 and Bull, 2008).

A key issue as addressed by Grayson (2006) is deciding on channels that will be most successful at promoting repsonsible entrepreneurship. This would be done through tried and true advice and support centers for entrepreneurs and being ready to sacrifice the phraseology for more widespread adoption of the behavior. Korosec and Berman (2006) advocate for municipal support of social entrepreneurship. Social entrepreneurship activities use innovative and efficient methods to address areas of social need that are often unmet by the the municipalities themselves.

Curtis (2007) noted that many social enterprise organisations try to find validation by mimicking other organisationsal forms in the hopes of attracting the neccessary support from like-minded institutions. However, social entrepreneur can be instrumental in harnessing cross-sectoral cooperation. Trexler (2008) in his critical assesment about the continuing longevity of "social entrepreneurship" as a movement makes an important point that "beyond developing more effective ways to work with nonprofit organisations, the greatest contribution of social enterprise would be to end the counterproductive divide between public benefit and corporate identity."

As Croatia searches for its own identity in relations to developing social policies, their current strategy is to model many of the countries it has referenced. There have been numerous conferences inside the country presented under the title of nonprofit entrepreneurship. There are currently a plethora of fashionable buzz words required to garner support and funding. The "it" phrases are often being incorporated into documents without true understanding which leads to an apparent dysfunction at the point of implementation.

There is an urgent need for the nation to align strategies with practice not only in the domains which would support the promotion of social entrepreneurship but in regards

to the wider Croatian public sector. The Support for Improvement of Governance and Management otherwise known as SIGMA (2008, p. 2) gave an assessment of Croatia that included the following passage "Moreover, the Strategy seems on the one hand to be too ambitious (in terms of deadlines, areas of reform, and objectives) and, on the other, too generic (objectives are rather general, priorities are not prioritised, indicators are just barely qualitative, and budget commitment is missing or too vague). As a result, implementation would not reach the level of real needs. Furthermore, the strategy could remain just another interesting document creating a virtual reality" Reknowned Harvard Professor Michael Porter (2009) recently held a no-holds barred forum for the greater business community in Croatia. At the event, he made a strong case for replacing Croatia's current top down, outdated approach for economic development in favor of one that is a collaborative results oriented process, inclusive of civil society.

5. Forces: Push vs. Pull SE development

The influences on the development of social entrepreneurship can come from inside the organisations or outside. These internal forces pushing and the external forces pulling help shape and mold its progression. In one instance, the country has had candidate status in its path towards EU accession since 2004. Thus, it must put all national strategies in sync with EU member states, social economy initiatives included. Croatia's compliance to EU standards has pulled both social entrepreneurship and corporate social responsibility to the forefront (Cingula and Klačmer-Čalopa, 2007).

In her examination of entrepreneurial stories about creating organisational forms, Carrona (2007) contrasts the diffusion of information being protected in the for-profit sector due to its competitive nature, whereas the non-profit sector encourages use of their models for the common good and actually facilitates entrepreneurialism of others. The social capital has significantly more weight than financial capital among nonprofis. As Mcelroy, Jorna and van Engelen (2006, p. 134) state that "the only form of social capital more valuable than collectively held knowledge is the collective capacity to produce it."

Social entrepreneurship can exist without social enterprises. Thompson and Doherty (2006) in their original investigation into 11 cases of social enterprises, state that "social enterprise can be succintly defined as organisations seeking business solutions to social problems but differ from socially-oriented organisations, which while bringing value to their communities may remain dependent on donations and not fully embrace a market orientation." It can be speculated that the latter are simply nascent social entrepreneurs with all the potential and promise of recognized social enterpreneurs.

Austin, Stevenson and Wei-Skillern (2006) highlight the need for social entrepreneurs to manage growth and "resist the powerful demand-pull" suggesting that SE opportunities must be evaluated based on their potential fit in line with the inherent social mission and the organisation's resource capacity. Social enterprises must continually balance social impact with resource management. Manouerving the nuances across sectors and across terrains make social entrepreneurs the natural

equivalent of plants growing from rocks but are of vital importance because they prevent erosion. They seem to defy obvious forces and as such are a wonder to behold.

6. Methodology

This study is part of ongoing research in the area of social entrepreneurship in a Croatian context. In the previous section of this paper, an exploratory literature review was conducted to provide complimentary and contrasting views of social entrepreneurship development in the region from a variety of perspectives with the intention to be thought-provoking and stimulate further initatives. The literature provides the leap off point for which the story continues to unfold. It is further illustrated through case analysis supplemented with primary interviews and email correspondence. We follow the lead of Mair and Schoen's (2007) investigation into how social entreprises become self-sustainable by loosely comparing three business model aspects of: value network, resource strategy and customer interface. While their employment was applied to more widely documented social entrepreneurial organisations, we experiment with its usefulness in also examining lesser known organisations originated from a single country.

7. Revisiting Case Studies

Global networks embodying the social entrepreneurship movement include mammoths like the Skoll Foundation (*Uncommon heroes. Common good.*) and Ashoka (*Innovators for the Public Good*). While neither of these organisations are active in Croatia, the like minded Nonprofit Enterprise and Self-sustainability Team (NESsT) has been present. They describe their own mission as:

> NESsT works to solve critical social problems in emerging market countries by developing and supporting *social enterprises* that strengthen civil society organizations' financial sustainability and maximize their social impact.

The four pillars of their efforts are focused on philanthropic investment (NESsT Venture Fund), education (NESsT University), professional services (NESsT Consulting) and shopping portal (NESsT Marketplace). Through their educational initiatives they publish case studies and their representative office in Budapest covers Croatia.

Case studies originally identifying social entrepreneurship activity in Croatian based non-profits was conducted by NESsT. Those original case studies were all published in 2006 and as a follow up they were re-examined and a semi-structured interivew was created and administered to key personnel in each of the representative instituitions. The interviews were conducted by phone from Zagreb, four were conducted in English and one was conducted in Croatian. They averaged approximately 20 minutes in length with the shortest being 15 minutes and the longest being 45. All interviews took place between April 23 and April 30th, 2009.

The Croatian social enterprises covered include:

A. Eco Center Caput Insulae Beli, the Island of Cres.
B. Slap Association for Creative Development in Osijek.
C. Udruga MI in Split.
D. Udruga Slijepih Association of the Blind located in Sisak.
E. Zelena Akcija (Green Action) in Zagreb with other regional locations.

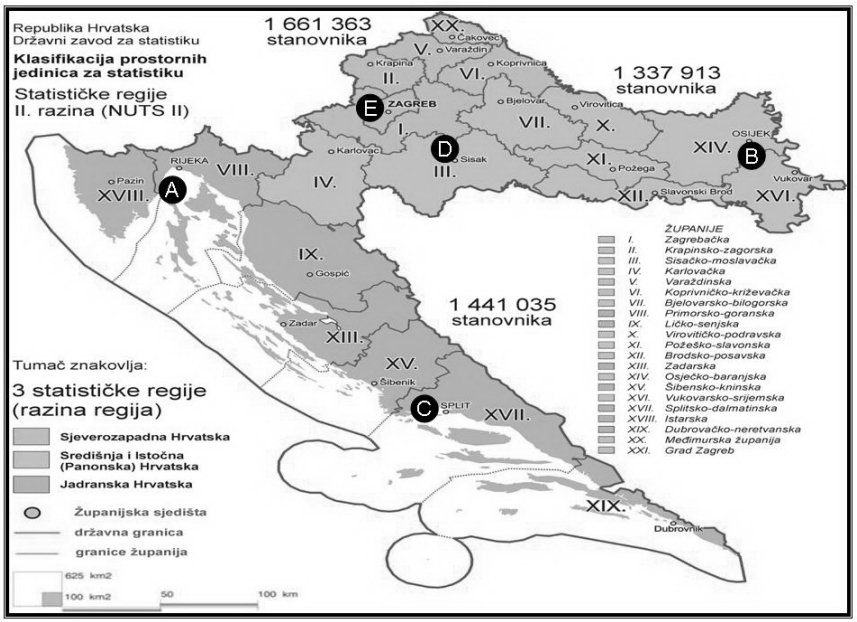


Figure 1. Map of Featured Nonprofits in Exisiting Croatian Entrepreneurial Landscape Adapted from (Keser, 2008)

The cases are geographically representative since they are located in different regions of Croatia from the islands to the cities as illustrated in Figure 6. While not all staff are still in place since the original case studies were conducted, the focus will be on comparing their organisational goals set in 2006 and the realisation of the goals as of 2009.

For consistency purposes we will review each organisation with a short summary of the original case followed by case development. As a quick reference tool and to lay out a side by side comparison of the cases we have labelled them A through E as they correspond to the original source material (NESsTa, NESsTb, NESsTc, NESsTd, and NESsTe, 2006, respectively). The case development updates are based on the personal communications between the author and key contact persons within each establishment, see Figure 7.

	A	B	C	D	E
	Eco Center Caput Insulae Beli	Slap Association for Creative Development	Udruga MI	Udruga Slijepih Association of the Blind	Zelena Akcija
Website	http://www.supovi.hr/english/index.php	http://www.pomakonline.com/	http://www.udruga-mi.hr/	http://www.udruga-slijepih.hr	http://www.zelena-akcija.hr/
Location	Island Cres, Croatia	Osijek, Croatia	Split, Croatia	Sisak, Croatia	Zagreb, Croatia
Geographic Scope	Local	Local/Regional/National	Regional	Local	International
Founder	Goran Sašić	Sonja Vuković +(10)	Nives Ivelja +		Toni Vidan +
Key Contact	Goran Sašić	Sonja Vuković	Nives Ivelja	Mirjana Vojnović	Daniela Jovanova Ivanković
Year of Foundation	1993	2000	1996	1955	1990
Start of Self-Financing	1995	2000	2002	2001	1994
Mission	Protect and preserve natural diversity as well as original values and cultural-historic heritage	To develop human resources for civil society, provide technical assistance to civil initiatives, develop models of self-sufficient non-profit enterprises and promote inter-sector cooperation.	To support the personal and social development of refugees through professional psychosocial work.	Supporting the education, employment and inclusion of blind persons in the Sisak area and Central Croatia.	Green Action is active on local, national and global levels and promotes the protection of nature and the environment, while advocating for sustainable development in Croatian and other communities. It also focuses on activities that foster public participation in decision-making processes with the aim of improving the quality of life in Croatia.
Core business	Protection of Eurasian Griffons	Intellectual Services	Social services and development	Quality of life for the visually impaired	Environmental advocacy and action
Organizational Structure	nonprofit	development oriented nonprofit	nonprofit	nonprofit	nonprofit
Number of Employees	7	6	10	2	12
Number of Volunteers	350	200+	54	as needed	30
Milestones	International recognition, local voice, managed growth	YES	Scalability of model	From 10 to 17% self-financed	Brand recognition
Target group	Human-Nature Relations	SME, NGO, CSO, women and youth	Senior Citizens, Volunteers, NGO's	152 members	250 members, 211 donors
Value Network	Shepards, Fisherman, Tourist Community	Partners BIOPA, PRONI, CREDE	World Bank, Croatian Government	Flower shops, Libraries, Fairs	Friends of the Earth
Strategic Resources	Expertise, Passion, Authenticity	Knowledge network, Human resources	Knowledge network, Human resources	Employment, Training of Members	Reputation, Grants
Customer Interface	Ecotourism and Education	Trainings, workshops, conferences, magazine portal	Meals on Wheels type program	Greeting cards and handmade toys	Activists, Education, Awareness

Figure 2. Comparison Chart of Nonprofit Organizations Profiled
Adapted from (Mair and Schoen, 2007)

Figure 2. Comparison Chart of Nonprofit Organisations Profiled Adapted from (Mair and Schoen, 2007)

A. Eco Center Caput Insulae Beli, the Island of Cres

In 1993, Eco Center Caput Insulae Beli (ECCIB) began their operations to protect the Eurasian Griffon Vulture on the northern Adriatic island of Cres. Its habitat and larger population on a nearby island was wiped out and drastic measures were needed in order to prevent its extinction. Initially, they began their self-financing with modest but mission-centered activities like educating school groups and accomodating volunteers interested in sustainable hands-on tourism. ECCIB has made significant progress since 2006 from a budget approximately 100,000 euros with 21% being self-financed to today where the budget hovers closer to 300,000 euros and has reached an impressive 70% self-financed.

The level of volunteers has remained at 350 per year but they have changed the structure so that the volunteers can have a greater impact by getting trained and staying for longer periods of time in the range of three to five weeks instead of the minimum one week.

While the founder Peter Susic has been preaching the ecotourism mantra to deaf ears for over a decade, the local community began to take serious notice only after legions of fans from the media began to offer accolades. International recognition has poured in from prestigous media outlets like CNN Traveller, Fodor's, BBC Wildlife, British Indepedent, Lonely Planet and British Times Travel section to name the recent articles some going as far as naming it a top ten eco-destination in the world. For an Adriatic destination to gain international press coverage while not unusual is especially noteworthy since it is centered on a particular destination not on the path of a typical tourist. The local and national tourism community is now listening and Mr. Susic and his team are positively effecting change to help others develop sustainable tourism development strategies and avoid the Mass tourism mistakes which would threaten the natural heritage.

Eco Center Caput Insulae Beli has taken the natural, cultural and historic treasures and highlighted the positive impact responsible entrepreneurship can have in impacting the wider community. By uncovering the potential of responsible tourism in the forms of sustainable eco-tourism and voluntourism ECCIB has taken an activity not widely recognised as a social entrepreneurial activity and made it both relevant and scalable. The opportunity to adapt and spread the concept is great as Croatia is a nation rich with geographic diversity including a coastline of more than a thousand islands.

Some key milestones have been the creation of unique trails that feature sculptures, words of wisdom and labryinths that have century's old historic and cultural foundations in the area. ECCIB has also worked hard to educate the value chain that contributes to preservation of biodiversity including the Griffon. In particular local shepards and fisherman play key roles in maintaining the food chain (Griffons feed on sheep carcasses) and the natural habitat (Griffons perch on cliffs, when the young fall into the sea they need to be rescued) and rehabilitation of injured birds takes place at the Eco Center before they are tagged and released once they are nursed back to health.

B. Slap, Association for Creative Development, Osijek

The vision of this organisation has remained steady from in inception in 2000 until today. It has leadership in place since its foundation and now neatly ties its core activities together under the umbrella of Intellectual Services. Slap offers civil society organisations consulting, and training services for a fee and is dedicated to promoting socially responsible entrepreneurship across sectors.

One of the first things asserted by one of the founders of Slap, Ms Sonja Vukovic, was that empowering others supercedes their own financial goals. Currently they have more projects and less time for their own income generating activities. They have a flexible policy which lets them create self-financing projects on an *as needed basis*. They have developed an online magazine portal which supports their promotion of nonprofit entrepreneurship and intersector cooperation.

A priority is manifesting socio-cultural relevant business models and they currently have a roster of over 20 different NGO's developing models of sustainability. One project known as Ethnobrand is a brand of handicrafts which combines the resources of female NGO's and a small business network to market four ethnocraft collections and offer sales through a webshop. This helps to employ women in previously war-affected areas and provide them with employment opportunities not readily available. Slap is adept at aligning strategic resources and creating synergy. They have worked in association with cluster initiatives like one called BIOPA and cooperatives to build up organic farming capacities in the fertile eastern region of the country. Slap provides creative human resource solutions in one instance loaning out staff and sharing salaries with partners to transfer intellectual services for project management and business plan development. They have taken on international interns but due to the students' lack of local market knowledge they no longer offer this option. They do however train local volunteers. The Ethnobrand has 40 volunteers while their youth employment support project that goes by the acronym YES has 120, with approximately 60 others utilises for assorted projects.

As well as expanding their program and reach, their operating budget has increased 6 fold to close to 2 million kuna or roughly 271,000 Euros which shows their growth and momentum.

C. Udruga MI, Split

Udruga MI is located in the Dalmatian coastal city of Split. The association was founded in 1996 primarily to help with refugees and displaced persons. As of the 2006 report, Udruga MI had three main activity areas 1. a center for refugees, 2. a volunteer center and 3. a senior services program. Initially, their programs with a staff of 13, but are now leaner with 10. One of which is dedicated full time to social entrepreneurship initiatives. They began with aid from the International Rescue Committee and continue to be reliant on foreign donations which now includes the Dutch government and the World Bank but also has ongoing support from Croatia's National Foundation

for the Development of Civil Society and the Croatian Ministry of Health and Social Care.

The organisation has continued to build its network and capacities. Today, they have two directions, the primary, which centers on social services, and the secondary which is focused on development. Their social services aid a growing elderly population aimed at improving the quality of mature living. Currently they serve 150 homebound seniors with home cooked meals and offer a weekly social calendar of activities for 400 to 500 mature adults. Their activities have influenced the Croatian government in launching a similar model for 70 other communities around the country with Udruga MI mentoring the projects initial phases. However, the pending crisis has put a strain on the organisation. While they have many projects in the pipeline that could help improve their self-sufficiency, some of their core activities are in jeopardy because government subsidies are in question.

The Udruga MI development program works on volunteer training and database maintenance, and provides a forum for civil society organisations of Split in cooperation with the City. In their advisory capacity they have a database of 600 organisations. One of the founders of the organisations, Nives Ivelja with close ties to active NGOs ran off a list of socially important, well run organisations that closed operations due to budget cuts. The reality that family violence centers, breast cancer awareness programs and eldery services were going to be casualties of the crisis had the leadership of Udruga MI on full alert.

D. Udruga Slijepih Association of the Blind, Sisak

The oldest institution in our group of case studies, the Association of the Blind was established in 1955 and started its first self-financing activities 46 years later in 2001. While its mission is to support education, employment and inclusion of blind persons in Sisak and Central Croatia, they were challenged to improve their effectiveness in light of diminishing governmental support. They were able to partner with a local artisan family to produce and market handmade wooden toys. While a percentage of the sale went to the association, previously unemployed members were trained and recieve a commission for their participation in production of the items. Udruga Slijepih had 143 members in 2006 and the number has increased to 153 in 2009. There are four members who work on the toy project. The association has gained increased recognition in the marketplace through these activities. However, members can not be fully employed under the current laws as they could jeopardise their own welfare benefits.

Since the original case study they have increased the percentage of their self-financing activities from 10 to 17. They have expanded their cooperation with other NGOs and artists to collectively sell and share in revenues and responsibilites. They have also expanded the channels of distribution from flower shops and holiday fairs to include libraries, street festivals and markets. The two full time staff gainfully applies their previous non-profit marketing expertise. It is encouraging to see they continue dialog with the government which is shown by the fact that hey have just completed a new business plan for 2009 accepted by the Croatian Ministry of Health and Social Care.

E. Zelena Akcija (Green Action), Zagreb

Zelena Akcija began in 1990 and combines advocacy and activism to raise awareness of issues facing the local, national, and international environment. It has enjoyed high brand recognition in Croatia. While there is general scepticism in the Croatian public about many nonprofits, especially ones with high profiles or those seeming to be doing too well, ZA has been able to build trust and develop a positive reputation due to its results. They were even able to capitalise on this by developing a logo and generating merchandise sales without tarnishing their image. Their product sales help them to connect with their customer base and raise awareness for their mission. Leadership from a founding member standpoint is still present from the original group, two are still active. Toni Vidan and Vladimir Lay have now gravitated towards specific areas of their interest. Mr. Vidan coordinates the energy program and Mr. Lay is an Advocacy Council Member. They also represent the international organisation, Friends of the Earth, in Croatia to add to their value network.

Since 2006, they have doubled the amount of full time staff from 6 to 12, remaining lean. Their number of volunteers is stable but now have a core of 20 to 30 students who assist in monthly activities. They offer programs such an environmental issue call center (Green Phone), island clean-up(Island of Šolta), nature protection zones: preservation and new designations, energy efficiency campaigns around Earth day and throughout the year as well as a permanent education and training center. Over the past three years they have created new combinations of resources in innovative ways which have garnered them additional support from large and individual donors. They took the property they rejuveniated on the island of Šolta and turned it into a Solar Academy for training, workshops and research in renewable energy. They also now have developed programs to deal with freshwater and river ecosystems; prevention, reuse and recyling of waste management; environmental law initiatives; urban intervention; and assists in monitoring international financial institutions in association with bankwatch.org.

They have expanded and diversified their methods of income generation but their two entrepreneurial ventures: an Eco-cafe and a Fair-trade eco-shop have hit roadblocks. Feasibility studies suggest both endeavors would be sustainable but not necessarily profitable, and thereby would be more for customer interface opportunities and mission transfer rather than spinoff ventures. ZA actively offers fee based consulting services, event management services, equipment rental, promotional merchandise sale in addition to membership fees, and online and offline donor drives. They have a roster of 250 members and a list of 211 donors which continues to rise.

8. Conclusions

Social entrepreneurship now requires its own dedicated glossary to define, categorise and pass on its potential meaning in different contexts. We looked at several definitions that could be relevant in a Croatian context and the working definition proposed by the European Union emerged as a clear favorite. Both influenced by the words were used to shape it (“social entrepreneurship is positioned between the

traditional private and public sectors and combine societal purpose with the entrepreneurial spirit of the private sector") and quite possibly the hope of EU inclusion it represents.

While a strategy on social entrepreneurship for Croatia has been presented the measurements of its effectiveness have not fully been addressed. The accountability for promotion and implementation requires the support of both formal and informal networks. The government, non-profits, education, business and industry all have a role to play in innovating social solutions for long term sustainability. The government has to put in place a truly enabling legal framework. The nonprofit sector needs to offer consolidated solutions to create more synergy across those serving similar missions. The education sector can research and profile nascent and active social entrepreneurs in and outside of the classrooms. The business community can recognise and reach out through corporate citizenship initiatives and provide resources, human and other to internal and external socially responsible activities. Development agencies at all levels; cities, islands, regions and inter-governmental, can continue to promote intersectoral cooperation.

This paper sought to give an overview of social entrepreneurship development in Croatia. The civil society sector was activated in the post-war aftermath but as the country transitions, the needs being addressed have not lessened but shifted. Thankfully its no longer a war zone making international headlines. Nevertheless social entrepreneurs are making progress towards fulfilling their social purpose and making inroads into their communities. As a result of this paper, we hope to encourage more rigorous and comparitive studies which can aid in bringing the country's SE aspirations into realisation. Since the social entrepreneurs are embedded not only in the the enterprises they start but also in the countries where they operate it seems logical that social entrepreneurs can play crucial roles in nations realising their potential.

References

1. Austin, J., Stevenson, H. and Wei-Skillern, J. (2006) 'Social and commercial entrepreneurship: Same, different, or both? ', Entrepreneurship Theory & Practice, Vol. 30, No. 1, pp.1-22.
2. Basic, M. (2008) 'Republic of Croatia, Ministry of Economy, Labour and Entrepreneurship, Map of entrepreneurial zones', Obtained through the Internet: *http://zone.mingorp.hr/English/index.htm/,* [accessed 10/05/2009].
3. Bayley, D., Bray, T.S.C. and Goračić, Lazetić, H. (2006) *Unplugged: Faces of social exclusion in Croatia*. Zagreb: United Nations Development Programme (UNDP) in Croatia.
4. Bezovan, G. (2001) Croatian Civil Society: On the Path to Becoming
5. a Legitimate Public Actor. Obtained through the Internet. [accessed 03/05/2009].
6. Bibu, N. A. and Orhei, L. (2008) 'Social entrepreneurship in the context of Romania's European integrationMunich Personal RePEc Archive. MPRA paper no. 9481. Obtained through the Internet: http://mpra.ub.uni-muenchen.de/9481/. [accessed 23/03/2009].

7. Bosma, A., Acs, Z.J., Autio, E., Coduras, A., Levie, J. and the *Global Entrepreneurship Research Consortium (GERA).* (2009) 'Global Entrepreneurship Monitor: Executive Report 2008', Obtained through the Internet: *http://www.gemconsortium.org/,* [accessed 12/03/2009].
8. Bull, M. (2008) 'Challenging tensions: critical, theoretical and empirical perspectives on social enterprise', International Journal of Entrepreneurial Behaviour & Research, Vol.14, No. 5, pp. 268-275.
9. Caronna, C. A. (2007) 'Entrepreneurial stories in the non-profit sector: A partial test and extension of cultural entrepreneurship theory', Paper presented at Annual Meeting of the American Sociological Association. Section on Organizations, Occupations and Work Regular Section. August 12, 2007. New York, NY.
10. Curtis, T. (2007) 'Finding that grit makes the pearl: A critical re-reading of research into social enterprise', International Journal of Entrepreneurial Behaviour & Research, Vol. 14, No. 5, pp.276-290.
11. Cingula, M.; Klačmer-Čalopa, M. (2007) 'Social entrepreneurship and social responsibility in Croatian companies', Presented at *The 6th International Symposium, Economy & Business. September 11 -14, 2007. Bourgas, Bulgaria.*
12. European Commission, DG Enterprise & Industry (2009) Social economy: Social enterprises. Obtained through the Internet: http://ec.europa.eu/enterprise/entrepreneurship/craft/social_economy/soc-eco_enterprises_en.htm, [accessed 04/03/2009].
13. Keser, J. (2008) 'Croatia: Regional differences and SME projects for competitiveness', Presented at the International Council for Small Business - World Conferenc. June 22-25, 2008. Halifax, Nova Scotia, Canada.
14. Korosec, R. L. And Berman, E. M. (2006) 'Municipal support for social entrepreneurship', Public Administration Review, May/June, pp. 448-462.
15. Galera, G. (2006) 'Emerging trends in Central, East and South-East European countries: General overview, Presented at the seminar *Emerging Models of Social Entrepreneurship: possible paths for social enterprise development in Central East and South East Europe. September 28-29, 2006. Zagreb, Croatia.*
16. Grayson, D. (2006). 'A conversation with DG Enterprise's Maive Rute about Responsible Entrepreneurship', EntreNews: The newsletter of EFMD's Entrepreneurship, Innovation and Small Business Network, Special Issue, January, p.3. Obtained through the Internet: http://www.efmd.org/attachments/tmpl_1_art_050222usmw_att_060627bxsd.pdf [accessed 21/1/2009].
17. Government of the Republic of Croatia (2006) Operational plan for national strategy creation of seminal framework for the development of civil society, 2007-2011. Obtained through the Internet: http://www.civilnodrustvo-istra.hr/fileadmin/datoteke/Korisni_dokumenti_2/Operativni_plan.pdf, accessed [05/4/2008].
18. Heckl, E., Pecher, I, Aaltonen, S. And Stenholm, P. (2007) 'Study on practices and policies in the social enterprise sector in Europe', The

Austrian Institute for SME Research and Turku School of Economics for the European Commission, DG Enterprise & Industry. Obtained through the Internet: http://ec.europa.eu/enterprise/entrepreneurship/craft/social_economy/doc/report_study_kmu_social_entreprises_fin.pdf, [accessed 21/1/2009].

19. Ivanović, M. (2006) 'Possible paths for social enterprise development in Central East and South East Europe: Croatia case study, Presented at the seminar *Emerging Models of Social Entrepreneurship: possible paths for social enterprise development in Central East and South East Europe. September 28-29, 2006. Zagreb, Croatia.*
20. Mair, J. and Marti, I. (2006) 'Social entrepreneurship research: A source of explanation, prediction and delight', Journal of World Business 41, pp. 36-44.
21. Mair, J. and Schoen, O. (2007) 'Successful social entrepreneurial business models in the context of developing economies: An explorative study', Int. Journal of Emerging Markets, Vol.2, No.1, pp.54-68.
22. McElroy, M.W., Jorna, R.J. and van Engelen, J. (2006) 'Rethinking social capital theory: A knowledge management perspective', Journal of Knowledge Management, Vol. 10, No. 5, pp. 124-136.
23. Morrison, A. (2000) 'Entrepreneurship: What triggers it? ', International Journal of Entrepreneurial Behaviour & Research, Vol.6, No.2 pp. 59-71
24. Moss, T.W., Lumpkin, G.T. and Short, J. (2008) 'The Dependent Variables Of Social Entrepreneurship Research', Paper presented at Babson College Entrepreneurship Research Conference (BCERC). June 4-7, 2008. Chapel Hill, North Carolina, USA.
25. NESsT (2006a) 'Eco Center Caput Insulae Beli: Becoming financially autonomous while capitalizing on hard assets and know-how', Nonprofit Enterprise and Self-sustainability Team Case Study Series. Obtained through the Internet: www.nesst.org, [accessed 03/02/2009].
26. NESsT (2006b) 'Slap Association for Creative Development: Balancing human resource capacity while earning revenues', Nonprofit Enterprise and Self-sustainability Team Case Study Series. Obtained through the Internet: www.nesst.org, [accessed 03/02/2009].
27. NESsT (2006c) 'Udruga MI: Fee for services – daily program activities meeting financial needs', Nonprofit Enterprise and Self-sustainability Team Case Study Series. Obtained through the Internet: www.nesst.org, [accessed 03/02/2009].
28. NESsT (2006d) 'Udruga Slijepih Association for the Blind: Providing employment for the organization's target group while obtaining organizational sustainability', Nonprofit Enterprise and Self-sustainability Team Case Study Series. Obtained through the Internet: www.nesst.org, [accessed 03/02/2009].
29. NESsT (2006e) 'Zelena Akcija(Green Action): Supplementing donor funds through several small-scale self-financing activities', Nonprofit Enterprise and Self-sustainability Team Case Study Series. Obtained through the Internet: www.nesst.org, [accessed 03/02/2009].
30. Odinsky-Zec, T. L. and Stubbs, P. Forthcoming (June 2009) 'Evolution of an organic champion: Social entrepreneurship, complexity theory and

leadership in transition- A Croatian case study', In: Goldstein, J.A., Hazy, J.K. and Silberstang, J. (eds.), *Complexity Science and Social Entrepreneurship: Adding Social Value through Systems Thinking*, Litchfield Park, AZ: ISCE Publishing.
31. Oregon Public Broadcasting (2005) The New Heroes. Obtained through the Internet: http://www.pbs.org/opb/thenewheroes/, [accessed 03/02/2008].
32. Phills, J.A. Jr., Deiglmeier, K. and Miller, D.T. (2008) 'Rediscovering social innovation', Stanford Social Innovation Review, Fall, pp. 34-43.
33. Porter, M.E. (2009) 'Competitive strategies for the Third Millennium', Presented at the Pliva Business Forum. *May 15, 2009. Zagreb, Croatia.*
34. Roper, J. and Cheney, G. (2005) 'Leadership, learning and human resource management: The meanings of social entrepreneurship today', Corporate Governance, Vol. 5, No. 3, pp. 95-104.
35. Salzano, C. (2002). Local economic development in the war-torn areas of Croatia: supporting the transition towards a participatory and decentralized market economy. Obtained through the Internet: http://www.ilo.org/dyn/empent/docs/F426156573/Case%20Study-Croatia.PDF, [accessed 10/02/2008].
36. SIGMA (2008). Croatia Public Service and the Administration Framework Assessment. Obtained through the Internet: http://www .oecd.org/dataoecd/48/0/41637118.pdf, [accessed 03/05/2009].
37. Stryjan, Y. (2005) 'The practice of social entrepreneurship: Theory and the Swedish experience', Wismar Discussion Papers, Hochschule Wismar, University of Technology, Business and Design, Wismar, Germany. Obtained through the Internet: www.hws-startupfuture.de, [accessed 03/04/2009].
38. The Aspen Institute (2009) Past Faculty Pioneer Award Winners. Obtained through the Internet: http://www.aspencbe.org/awards/ pioneers/ past_winners.html, [accessed 11/05/2009].
39. The Heritage Foundation (2009) Croatia information on economic freedom, 2009 World Economic Freedom Index. Obtained through the Internet: http://www.heritage.org/index/Country/Croatia, [accessed 09/05/2009].
40. Thompson, J. and Doherty, B. (2006) 'The diverse world of social enterprise: A collection of social enteprise stories', International Journal of Social Economics, Vol. 33, No.5/6, pp. 361-375.
41. Todres, M., Cornelius, N., Janjuha-Jivraj, S. and Woods, A. (2006) 'Developing emerging social enterprises through capacity-building', Social Enterprise Journal, Vol. 2, No. 1, pp. 61-72.
42. Trexler, J. (2008) 'If you just follow the latest fad, what happens when it fades?', Chronicle of Philanthropy, Vol. 21, No. 4, p. 33.
43. U.S. Small Business Administration (2007) Social Entrepreneurship and Government: A new breed of entrepreneurs developing solutions to social problems. Obtained through the Internet: www.sba.gov, [accessed 03/04/2009].
44. Waddock, S.A. and Post, J.E. (1991) 'Social entrepreneurs and catalytic change.' Public Administration Review, Vol. 51, pp.393-401.

Trends and Changes on the International Financial Market in Terms of Globalisation

Oana Mionel

Academy of Economic Studies, Faculty of International Business and Economics, 41, Dacia Blvd., District 1, Postal Code 010404, Bucharest, R omania,
panaoana2005@yahoo.com

From the outside, financial markets appear dry, technical and economic in nature – (all about) percentage declines, volume, margin call and paper losses. However, their inner mechanism is psychological. The purpose of this paper is to identify the changes and tendencies of the world market. The paper does not rely on an exhaustive approach, but it selects the relevant theoretical and practical elements regarding the most important present change and tendency of the world economy – the shift from the rational investment decision to the psychology of the financial market. The focus of this analysis is the investor's behaviour in making decisions at financial market level.

Keywords: globalisation, financial market, behaviour, currency market, psychology of the financial market, investor's behaviour.

1. The globalisation of the financial markets

The globalisation of the financial markets, phenomenon typical of the end of the 20th century, has led to a separation of the financial economy from the actual economy. This phenomenon has not occurred suddenly. There had been several successive stages that engendered what one calls today financial innovations and behaviouristic finances: in the first stage, the financial operations accompanied the flows of goods, and in the second stage, the financial activities either preceded or followed the flows of goods, fact that eventually led to the occurrence of the Euromarkets.

In time, the globalisation of the financial environment caused important changes at the world level,''the rule of the 3 D-s'' (deregulation, non-mediation, openness); the stress was laid on the competition between the world's big financial centres, innovations and implementation abroad; the development of the IT sector offers the actors on the world market information in real-time which is the essential factor in the decision-making processes; ''the professionalisation'' of the markets that amplifies the dynamics or tendencies of the market (there occurs an increase of the importance of the professional investors who have the same information and similar behaviours; the effects of their actions may provoke sometimes brutal shocks, because all reactions often coincide);

the growth and diversification of portfolios. The common topic of all these evolutions is the role of the market – market growth and development, the competition and the enhanced market freedom, the increase of market efficiency, the increased trust in market forces and mechanisms, as well as the creation of the best market techniques and tools. The interdependence of all these forces creates a global environment of creativity and enthusiasm. The financial markets have progressively increased becoming more and more sophisticated, integrated and efficient. Globalisation enhances the role of information in the financial operators' psycho- investment decisions.

2. Changes in the investment behaviour of the „players" in the financial market

According to the classical theory of financial market shares, operators who anticipate are able to „defeat the market", only if they have either exclusive accurate information, or priority over public information. Observing the old exchange rates and the others' behaviour influences the behaviour of each of the operators. Thus, surveys done after the October 1987 crash suggested that successive variations of the exchange rates were links whose association along a chain was facilitated by the fears of both the institutionalised and individual investors, motivated by the behaviour of individuals from the exterior (Schiller, 1993, p. 34). This kind of contagious process, generating a collective dynamics by means of expectations, significantly foregrounds the efficiency hypothesis that defines information as an objective reality while its various meanings are not considered important.

In the field of foreign currency market, there are significant differences between operators: they do not have the same time horizon, the same risk taking manner or the same targets. This variety of operators can be explained by the theory of knowledge. Within the efficiency hypothesis, information is perceived as an objective reality, while its various meanings are disregarded. Despite this opinion, it is a known fact that information is a stimulus for action only if its meaning is of a certain interest to the subject. One can constantly notice that two operators that simultaneously receive the same piece of information can formulate different conclusions with respect to its probable impact on prices. The heterogeneous reactions to information influences the predictions related to the future market evolutions. The neoclassical economic models of financial markets assume that participants make decisions on the basis of available information and that they employ this information fully in a way they consider appropriate (Oberlechner, 2004, p. 411). Starting from the premise that participants on the market are rational, the hypotheses of the efficient markets claim that prices always completely reflect existent information (Fama, 1970, p. 389). Rational decision and market behaviour on the financial markets are presumed to be natural and consequently they do not need to be explained (Cosmides and Tooby, 1994, p. 330). The premises of the concept of efficient market are the following: investors are rational (investors are risk reluctant and they want high profit assets for a certain risk level), markets are efficient (current exchange rates reflect all the available or public information), profits are independent (the changes of the exchange rates can be generated only by new information; the t day profit is not correlated with the t+1 day profit), markets have a

random walk (the probability of profit distribution is approximately the same as the normal distribution- the Gauss' bell/curve).

Over the last few years, the efficient market hypotheses have changed both theoretically and from the perspective of the empirical observations that contradict it. Actually, the premises of The Efficient Market Theory are not real. Investors are not always reluctant to risk and they do not react immediately to information. In some cases, they act later, following the trend. People do not relate to new information in a linear way (Shefrin and Statman, 2003). People can become even more interested in taking additional risks when investments are likely to fail (Kahneman and Tversky, 1979, p. 265) Thus, the abundance of market 'anomalies' has been compared to the hypothesis of rationality, which assumes that participants in the market have magical abilities to sense the solutions to the economic problems each economist is confronted with. On the foreign exchange market, many economists have noticed that, unlike the hypothesis of the full utilisation of unchecked information, and especially with respect to short term movements of the exchange rates, models based only on economical principles were not successful (MacDonald and Taylor, 1992, p. 13). For example, analyses of the exchange rates of increase and the evolution of commerce can at most predict the short term movements of the exchange rates (Frankel and Froot, 1990, p. 22.) and also "rational" economic concepts such as purchasing power parity are not considered useful to the operators on the foreign exchange market (Cheung and Chinn, 2000, p. 456). In the financial literature, many attempts have been made in order to „correct" the theory of the processing of information on the financial markets. For example, in 1992, the role of the waterfalls of information during financial crises was examined. It was thus found that in such moments, the investors should simply follow the others' behaviour and consequently the rules of the improvement decision are characterised by a mimetic behaviour of market participants who do not employ their own information (Bikhchandani, Hirshleifer and Welch, 1998, p. 160).

3. How does information influence the investor's behaviour?

The information available on the market is a set of public data, rendered in an objective form. This information becomes material information (it can influence the prices of the financial assets) when there occurs a combination of the investors' knowledge, experience and evaluations (Fama and Miller, 1974). Two important economists have worked together on an ambitiously large project: testing the information efficiency of the financial market. Within the same basic premise of the rational behaviour of the financial investors, the two researchers are now interested in the degree of integration of the previous, present and anticipated economic information in the present price of the financial assets (in their inherent value). This time, they are interested in the connection between information and value, which is the degree of determining value in relation to the information available in the moment of evaluation.

Even from 1990, the testing of the hypothesis regarding the financial market efficiency has focused on the independence of the variations of the stock exchange rates and it has built the pattern of the financial assets price movement in random distribution processes (the random promenade model) (Bachelier, 1900, p. 63). The scientific research initiated in 1990 was completed in 1965 in the important article The

Behaviour of Stock Market Prices in which the author structures the financial market into three categories of informational efficiency: weak- the market that enables the immediate integration of any available information regarding the previous evolution of the respective stock market prices into the prices of financial titles; semistrong – the market that enables both the integration of the previous information and the inclusion of all public information (transmitted via mass-media) into the present price; strong-the market that enables the inclusion of any possible information including privileged one (available only to managers) in the calculation of present prices (Fama, 1965, p. 57).

The variations of the exchange rates are therefore independent. It is useless for one to use the information contained by the history of stock market prices to forecast their evolution, because according to the hypothesis of the weak efficiency, all this (previous) information is already incorporated into the present price of the asset. A single new piece of information can determine a variation of stock market price. The one who owns this piece of information is the single person that can make an accurate prediction regarding the evolution the exchange rate and can obtain profit from balancing the two values. The increased competition among financial investors generates the increase of the market informational efficiency and the decrease of the opportunities to arbitrate the exchange rates. The variations of exchange rates become random and their predictability becomes more and more reduced. Information is a strategic variable of major importance. It is not free and it is not evenly distributed at the level of actors on the foreign exchange market (Hubert, 2001, p. 63). Therefore, obtaining it involves a cost that, in its turn, influences the result of the decisions. In addition, it is not available for everybody – some of the actors of the financial life own privileged information or have a much faster access to information.

4. How does the market „player" perceive risk?

The asymmetry of information generates differentiated positions of the economic actors in the process of price-value arbitration. Consequently, the asymmetry of information generates two categories of risk: the moral hazard risk, and the risk of adverse selection (the one who identified these types of risks is the American economist, winner of the Nobel Prize in 1972, Kenneth Arrow). Risk, defined by The Oxford English Dictionary as "hazard, danger; exposure to bad luck" is usually perceived as a variation of the profit of their investments.

Dealing with risks entails important consequences for the world economy. Certain investments are profitable only when the risks associated with them can be transferred to other economic actors. Possibilities of this kind are somehow limited. The reason is that by using derivate products (risk transfer instruments) one can only deal with known risks. On the other hand, the risks that generate big financial crises are considered rare and difficult to predict. Consequently, one cannot talk about an absolute protection against risks and the subsequent losses entailed by them. Thus, one can state that it is almost impossible to accurately predict the level of exchange rates between strong currencies in an interval of one year. This involves uncertain variables for which one cannot discover a governing law of probabilities, not even empirically. This is a consequence of the fact that they are caused by events such as an earthquake in Japan, a stock exchange crisis in the United States, the solving or the continuation of

institutional reforms in Europe. The general principle that generates the enhancement of derivate products is risk decomposition (Sharpe, 1964, p. 424). A complex risk is considered an articulation of elementary risks that can be dissociated. Hence, the players on the market have decided to create instruments (derivate financial contracts) by which they can decide on the type and degree of risk they are willing to take (Miller, 1997). Thus, one may infer that financial derivate products do not diminish risks, since they have no impact upon the risk factors. Their role is to transfer risks from the investors who are unwilling to take them to other investors who accept them. Thus, they enable an investor to accept both a higher risk level and an increased profit of the investment. The moral hazard risk is caused by the uncertainty resulted from the fact that certain economic agents could (discreetly) transmit false signals in order to obtain profit from the difference between the real value of the title and the market price influenced by the respective signals. The moral hazard risk is the focus of a new general theory: the theory of signals, developed by American economists (Grossman and Stiglitz, 1976). The risk of adverse selection occurs as a consequence of speculative behaviour of the market players.

In order to formulate predictions, individuals rely on information, procedures of analysis, and their processing. Thus, in this process of information analysis, investors make systematic mistakes of logic, repeated and predictable, that are the results of behaviourist and psychological predispositions. With respect to decision making, the consequence of these faulty predispositions is that the investor can process information in a wrong manner, by making decisions that are either too strong or too weak. An exaggerated emotional reaction caused by new information is generated by the phenomena of representation and generalisation: individuals tend to extrapolate the existent information at a certain time and consider it representative for the entire population (predictive stereotyping).

The exaggerate reaction can also be explained by the Prospect Theory (Daniel Kahneman – winner of the Nobel Prize in 2002 for integrating psychological research into economic science and especially with respect to the evaluation and the making of decisions in case of uncertainty and Amos Tversky, 1979). The scientific result of the two authors is considered fundamental in the field of behaviouristic theory of decision making in the context of risk and uncertainty. This theory suggests that when the investment heads to losses, the investors can become interested in taking additional risks.

5. The anticipations of the financial “players”

The behaviour of the participants in the market obviously depends on their anticipations. Psychologists and economists have studied the manner in which people form their expectations, in relation to the anticipations of the variations of the exchange rates (speculative actions) (Robert E. Lucas – The Nobel Prize for the theory of rational expectations on the behaviour of different participants in the economic changes). Speculation is based on selling or buying a certain currency with the intention to resell or re-buy it on a subsequent date with the hope to obtain a profit from the respective favourable difference of price. Most of the times, an action of this

kind entails risks, because the obtaining of profit is the result of anticipations. If the anticipations are faulty, losses may occur.

In the field of foreign exchange, speculative actions situate the participant in this kind of operations in an open position, i.e. he/she expects fluctuations of the exchange rates in both senses, but he/she considers that the evolution of the exchange rate is according to his/her anticipations. Hence, the "a la hausse" speculator anticipates that the foreign currency he/she buys at a certain time, at a transparent exchange rate, will be appreciated in a future moment, and hence its sale will create a profit. Similarly, the "a la baisse" speculator, relying on anticipations regarding the depreciation of the foreign exchange he/she owns, will sell the respective currency in favour of an appreciated currency. The key element for the success of the "a la hausse and "a la baisse" operations is the quality of the predictions (anticipations), i.e. both the direction of the foreign exchange rate and the moment when the appreciation or the depreciation have reached their extreme levels (the maximum or the minimum level).

Unlike the foreign exchange arbitration, within the speculative transactions the initiator of speculation does not start from an initial liquid. He/she relies on his/her prediction regarding the subsequent evolution of the exchange rate and he/she does not take into account the possibility he/she has according to the international protocol. This protocol stipulates that the speculator can postpone the payment with 24 hours, interval in which he/she will make a transparent transaction of selling currency. If he/she does not obtain this agreement, the initiator of the speculative transaction can obtain a very short term credit (for one day) in order to make the payment. The interest paid for this credit will diminish the profit resulted from the speculative transaction.

Speculative actions tend to diminish the difference between the transparent exchange rate and the forward exchange rate and this is why they may have stabilising effects under certain circumstances. The role of the speculative actions can sometimes be disturbing for the foreign exchange market. It all depends on the exchange rate protocol adopted by the respective state, as well as on the general economic situation. Another cause is represented by the wrong anticipations of a significant number of speculators. Speculative actions are frequent, but of medium dimension and they contribute to the stabilisation of the exchange rate. In case of an economy characterised by special imbalance conditions, where a monetary policy is applied in order to balance the economy, at least on medium term, the speculative actions are very sensitive to various measures of political economy, acting in accordance with the anticipations of the effects of these policies.

Hence, those individuals specialised in speculative actions regard a massive increase of the monetary mass as a signal that the inflation and depreciation of the national currency will continue. The same interpretation is assigned to the prolongation of the deficit of the balance of payments for more trimesters or years, followed by a government initiated agreement with the international financial organisations.

Depreciation becomes almost certain for speculators when the government resorts to measures resulting in legislative restrictive regulations (currency owners must comply with certain decisions of the Central Bank), nationalisation, the sudden increase of currency taxes, restrictions on foreign exchange. The governments' efforts to stop the instability of the internal foreign exchange market are closely monitored by those

individuals involved in speculative actions. Hence, when the government's actions are reassuring, based on sufficient resources and also done in accordance with declarations, the "speculative attack" is considered to enhance stability. It generally acts in the same direction as the central bank in that it sells the expensive currency and it buys the cheap one. Hence, by emphasising the supply, respectively the demand, it leads the exchange rate to a balanced value. If, on the contrary, the government actions contradict their own declarations or express lack of determination or lucidity, the speculative actions may disturb the situation on the foreign exchange market. It follows that in a context of a government supervised flotation of the exchange rate, the speculator of foreign exchange pays attention to both the evolution of the national currency rate in relation to other currencies, and to the Central Bank's actions and declarations regarding the stability of the exchange rate. The critical moments when this attention increases are represented by the economic or public changes: the coming of maximum import periods (during holidays), the beginning of the touristic season, international agreements with international financial organisations, changes of government, major social conflicts.

Speculation, in the negative sense of the term, begins only when the currencies' rates are disconnected from "their fundamental value" so that it can be appreciated via international comparisons of the purchasing power or productivity (Teulon, 1996). One can thus talk about the occurrence of "speculative bubbles", the concept of „ speculative bubble" was introduced in the economic literature by Heckscher (1931) and then by Keynes (1936) and the occurrence of a "casino economy". The bubble is the result of the behaviour of the agents who intervene on the financial markets. Their decisions to invest depend on the anticipated future profits. The optimistic or pessimistic nature of the predictions is linked with the subjective probability they envisage for the occurrence of a future event. When operators anticipate an increase of the inflation rate and therefore a depreciation of the exchange rate, they massively sell the unstable currency. The ensemble of these (selling) behaviours engenders a decrease of the currency value that passes the long-term balance level, obeying PPC (according to Rudgider Dornbush, an overshooting occurs).

The speculative action characterises the individual willing to engage in the game of fluctuations registered in the rates of various currencies, aware that on the basis of information he/she has access to or simply by means of intuition helped by prognosis methods, has great chances to succeed. As a conclusion regarding speculative actions, one can state that their success significantly depends on the ability to anticipate the evolution of the economic situation in general and that of the exchange rates in special.

6. Reducing risks by means of diversification

Portfolio selection- the theory of portfolio and the idea of risk are the basis of the financial theory in the field of portfolio analysis (Markowitz and Sharpe, 1990) The scientific work of the two researchers makes up the modern theory of portfolio, the tough nucleus of the financial theory. Along a decade, the two researchers defined two important elements that make up the modern theory of portfolio: the idea that the balance between profit and risk depends on diversification and the definition of risk (Markowitz, 1990).

The economic literature regarding the theory of portfolio options presents two more models: the first model, initisated by Tobin - A General Equilibrium Approach to Monetary Theory în 1969 and continuated of Branson, Halttunen şi Manson – Exchange rates in the short run: the dollar/ deutschmark rate, 1977, does not include foreign currency in the structure of the market participants' portfolio and enables the explanation of portfolio reassignments generated by the deficits or surpluses of the balance of the current transactions; the second model (the model with currency substitution) starts from the fact that participants in the foreign exchange market own a portfolio, both national and foreign currency – perfectly substitutable assets.

The model with substitution between currencies has two important differences in comparison with the first model: the investors create a portfolio made up of national and foreign currencies; the national and the foreign titles are perfectly substitutable. The hypothesis of static anticipations is no longer valid, in other words participants in the market do not anticipate a stability of the exchange rate (in this situation, the anticipations of the exchange rates are considered adaptable and rational).

The world economy can be equated with an ensemble of markets (especially currency and financial markets). In these markets connected by instantaneous communications media, the investors (banks, retirement funds, etc.) try to maximize the return of the asset portfolios, minimizing risks. The search for the best profit creates short term movements of capital. The latter engender additional offers related to certain foreign currencies and new demands related to other foreign currencies. Thus, the balance price (the exchange rate) changes within the exchange market. Since 1982, the disinflation process and the maintenance of high interests, led to transformations in the investment and saving behavior. Hence, private financing capacities were created in order to obtain high financial profits.

Studies done over the years also prove that the tendency to diversify registers higher and higher increases, due to the significant evolution of the international financial market: in Japan, the diversification process developed gradually and continued to develop once the markets were opened at the end of the 70's. At present in London, an international financial center acknowledged from the beginning of the 80's, there is a very strong tendency to diversify portfolios. Starting from 1982, the economic situation in the United States, - the low inflation, high remuneration of investments, the economic boom, encouraged the circulation of the international capitals available in this region. In this context, the United States of America gained a strong position that allowed it to maneuver the circulation of capital from other creditor – countries for its own profit, especially since they are expressed in dollars. In the case of the European Union one notices that the countries from the Euro Zone have achieved a massive diversification since 1993 as a result of the unification of the European capital market. The author of the theory of the risk reluctant portfolio describes the idea of portfolio diversification:" "Don't put all your eggs in one basket" (Krugman and Obstfeld, 2007).

7. Conclusions

The foreign exchange market has been subject to several forces and this fact has had important consequences: the market reacts to the contradictions between the economic and currency policies applied by the developed countries, sanctioning the disagreements caused by sheer national attitudes that do not take into account the impact on other countries; the evolution of the exchange rates is more and more based on a short term logic. Any slight variation of the anticipated profits may cause a significant and lasting reaction of the exchange rate.

References

1. Aglietta, M. (2001), *Macroéconomie financière*, Paris: La Découverte.
2. Andersen, T. (2006), Global Drivatives – A strategic risk management perspective, Boston: Prentice Hall.
3. Copeland, L. (1989), Exchanges rates and international finance, London: Pearson.
4. Cuthbertson, K. and Nitzsche, D. (2001), *Investments: spot and derivatives markets*, New York: John Wiley&Sons.
5. Eun, C. And Resnick B. (2007), *International financial management*, Boston: McGraw-Hill/Irwin.
6. Gaftoniuc, S. (2000), *Finanţe internaţionale*, Bucureşti: Editura Economică.
7. Herbst, A. (1992), Analysing and forecasting futures prices: a guide for hedgers, speculator and traders, New York: John Wiley&Sons.
8. Homaifar, G. (2004), Managing global financial and foreign exchange rate risk, New Jersey: John Wiley&Sons.
9. Hull, J. (2006), *Option, futures and other derivatives*, Upper Saddle River: Prentice Hall.
10. McDonald, R. (2006), *Derivatives markets*, Boston: Addison-Wesley.
11. Milles, R. (2004), Warren Buffet Wealth – Methods Used by the World`s Greatest Investor, New Jersey: John Wiley&Sons.
12. Oberlechner, T. (2004), *The psychology of foreign exchange market*, New York: John Wiley&Sons.
13. Solnik, B. (1991), *International investments*, Menlo Park: Addison-Wesley.
14. Solnik, B. And Jacquillant, B. (2002), Marches financiers: gestion des portefeuille et des risques, Paris: Dunod.
15. Journal articles
16. Barberis, N. and Sheleifer, A, (2003), 'Style Investing', *Journal of Financial Economics* 68, pp. 161-199.
17. Bikhchandani, S. and Sharma, S. (2001), 'Herd Behavior in Financial Markets', *IMF Staff Papers*, Vol. 47, No. 3, pp. 279-310.
18. Cosmides, L. and Tooby, J. (1994), 'Better than rational: Evolutionary Psychology and the Invisible Hand', *The American Economic Review*, Vol. 84, No. 2, pp. 327-332.
19. Fama, E. (1965), 'Random Walks in Stock Market Prices', *Financial Analysts Journal*, Vol. 21, No. 5, pp. 55-59.

20. Fama, E. (1970), 'Multiperiod Consumption – Investment Decisions', *The American Economic Review*, Vol. 60, No. 1, pp. 163-174.
21. Fromlet, H. (2001), 'Behavioral Finance – Theory and Practical Application', *Business Economics*, Vol. 36, No. 3, pp. 63-69.
22. Oberlechner, T. (2004), 'Information sources, news and rumors in financial markets: Insights into the foreign exchange market', *Journal of Economic Psychology*, No. 25, pp. 407–424.
23. Shefrin, H. and Statman, M. (2000), 'Behavioral Portfolio Theory', *Journal of Financial and Quantitavive Analysis*, Vol. 35, No. 2, pp. 127-151.

Breaking the "Glass Ceiling": The Interplay of Institutional Pillars as Antecedents of Organizational Innovation

Myrofora K. Rafailidou

Athens University of Economics and Business, Department of Marketing & Communication

The present study uses the lens of organizational innovation to focus on breaking the "glass ceiling" within a neo-institutional framework. First, the concept of the "glass ceiling" is defined and the extent of its existence is assessed. Second, the four pillars of the neo-institutional theory are presented. Third, a model for explaining the differential breaking of the "glass ceiling" in organizations is formulated using the four pillars of institutions as the hermeneutic framework. In particular, this model aims at responding to the question of what is the motivational basis for breaking the "glass ceiling" and what are the factors, mechanisms and processes involved with it. The present study argues that organizations use a mixture of expedience, social obligation, shared understanding, and emotional attachments as a motivational basis. Furthermore, it focuses on the role of internal and external factors related to this particular organizational change.

Keywords. glass ceiling, neo-institutional theory, four pillars of institutions, organizational innovation

1. Introduction

The study of organizational innovation, as compared to technological innovation, has been comparatively neglected. Neo-institutionalism has attempted to rectify this knowledge gap. The present study uses the lens of organizational innovation to focus on ways of breaking the "glass ceiling" within a neo-institutional framework.

During the last three decades, women's participation in the workforce has significantly increased, effectively changing the labour dynamics. It is estimated that nowadays women constitute more that 40% of the global workforce (International Labour Organization, 2008). However, despite the undeniable progress, women continue to comprise fewer than 5% of senior executives in most developed countries (International Labour Organization, 2004). A series of testing hurdles such as negative stereotypes, discrimination at the workplace**,** and social prejudice form a subtle yet also

quite strong barrier on the way towards the higher echelons for women. This invisible barrier that prevents women rising upward in management is called the "glass ceiling" (Hymowitz and Schellhardt, 1986; Morrison and Von Glinow, 1990; Tharenou, 1999).

The "glass ceiling" has serious implications for organizations. It is one of the main causes of the diminished utilization of a company's workforce, since it hampers the development and use of the talents of women. As a result, companies experience lost profits, lack of diversity in their workforce, decreased productivity, and inflated hiring and training costs (Schwartz, 1989; Adler, Brody, and Osland, 2000). Human resources are, after all, one of the few elements that can be used to create a strong competitive advantage that is unique and difficult to imitate elsewhere. However, the only way to achieve that is to fully employ the power of women in the workforce (Harel, Tzafrir, and Baruch, 2003). Companies that recruit, maintain, develop and promote the women of their workforce have been found to develop a strong competitive advantage (Caputo and Dolinski, 1998; Weiler and Bernasek, 2001; Collins, 2004). Given the positive economic effects of breaking the "glass ceiling", we would expect that companies, based on numerical calculations alone, would rush to embrace this practice. To the extent that they do not, we expect that, in addition to these calculations, other intra-organizational and environmental institutional mechanisms may shape this particular innovation.

Institutions, the basic subject of neo-institutional theory, provide society with specific guidelines that regulate its functions, and thus reassuring stability and order. These frameworks are transmitted and maintained through various means and they operate on a variety of levels, ranging from global systems to interpersonal interactions. Although they tend to resist to external pressures, they are not invulnerable to change. (DiMaggio, 1988; Oliver, 1991). Institutions are multifaceted, durable social structures, made up of symbolic elements, social activities**,** and material resources. Institutional systems are more or less comprised of the same elements but in different analogies; they consist of cultural-cognitive, normative, regulative and emotional elements that - along with associated activities and resources - create a unique combination.

The present study proposes a conceptual model for explaining the differential breaking the "glass ceiling" in organizations using the four pillars of institutions as the hermeneutic framework. The process of breaking the "glass ceiling" by an organization is motivated by a combination of regulative, normative, cultural-cognitive**,** and emotional mechanisms. Within this framework, the relationships between intra-organizational and broader environmental institutional mechanisms shape the reasons why various organizations embrace (or reject) the innovative practice of breaking the "glass ceiling" and why different organizations engage in that practice at different points in time.

This framework asserts that institutions are made up of diverse elements that differ in a number of ways, since they offer various levels of order and compliance, varying mechanisms, a range of empirical indicators and alternative approaches for establishing claims of legitimacy**.** These elements may not be aligned or may undermine the effects of each another; therefore, the dominant element varies among the different institutional systems as well as over time. Furthermore, the organizational boundaries are often invisible or easily penetrated, allowing alternative approaches to emerge and

support divergent models of behaviour. Competing rules or schemas open up possibilities for breaking the "glass ceiling" and lead to the adoption of new institutions. The proposed model will be tested with a combination of mainly qualitative and secondarily quantitative data.

2. Theoretical and Empirical Background

2.1 Glass ceiling

Nowadays, the ever-growing bibliography concerning the causes of the "glass ceiling" can be organized into two distinctive categories. The first category refers to differences that can be found on an individual level among members of the two sexes. These differences have to do with the nature, skills, and idiosyncrasies of men and women, that supposedly make males more successful managers than their female counterparts. The development of these differences is ascribed to two different origins, nature and society. The first group of explanations about the differences between men and women is based on evolutionary psychology and socio-biology (Browne, 1998; Fisher, 1999) and it has been the object of strong criticism from "traditional" social sciences. The second group of explanations regarding gender differences is based on the process of socialization that takes place from an early age during human development. According to this group, the differences of men and women are a by-product of the socialization procedure, since the two sexes are encouraged to do different things while also discouraged outright to do others. This procedure is thoroughly explained by the Standard Social Sciences Model. According to Toody and Cosmides (1992), this model is based on the assumption that the transmission of culture from generation to generation is conserved through learning, a widespread procedure that ensures that the child will grow into an adult of its culture. It is a collective procedure that is enforced upon the members of the new generation, in order to create "copies" of the older generations. Under this scope, the differences between men and women that create the "glass ceiling" are nothing more than an extension of the socialization procedure without further biological provenance.

Perhaps the most ambiguous subject on the field of differences between men and women is the managerial style of each gender. One of the most frequently-quoted articles on the subject (Dempster and Paton, 2002; Harris, 2002; Vilkinas, 2000; Zellner and Segal, 1992; Epstein et al., 1991) is Judy Rosener's "Ways Women Lead". The differences in the managerial style of men and women, according to Rosener, are a result of the different leadership styles that the two genders use. The women's style is leadership oriented to people, while the men's is leadership oriented to results. Rosener uses the term "interactive leadership" to name the women's leadership style. Over the years, there have been several studies and meta-studies that, in general, come to the same conclusion; that men and women do not have significantly different management styles (Alimo-Metcalfe, 1993; Vilkinas, 2000). Because of the large volume of relevant studies, it more efficient to examine meta-studies that collectively examines large numbers of individual studies to produce an aggregate result. These studies also confirm that the differences between men and women managers are not significant, and that their performance is not different (Powell, 1990; Tharenou, 2001).

Hermeneutical approaches based on gender differences (of biological or social cause) have been used not only by academics but by managers as well, in order to explain the absence of women from the higher managerial echelons. The use of norms that have been created by men and for men to measure the deviation of women from the male standards had several flaws. First of all, the research that was based on such frameworks did not find significant differences between men and women as was expected, while on the other hand certain traits that were allegedly associated with the "ideal manager" such as teamwork, cooperativeness or empowerment were attributed to women and not men. Therefore, if these assumptions were correct, women should be better managers than men and ergo compose the majority of the managerial ranks, something that was under no circumstances confirmed by the statistics (Tharenou, 1999). Another major flaw was the fact that this kind of research did not take into consideration any social or organizational factors such as stereotypes or social prejudice (Tharenou, 1999).

The second distinctive category of hermeneutical approaches refers to organizational and social causes. This category escapes the narrow frame of gender differences, expanding from the individual to a collective level. According to the explanations that this category includes, the major causes that create the "glass ceiling" have to do with the existence of strong, negative stereotypes for women in the workplace. These stereotypes affect women in a multitude of ways, including discrimination during the recruitment and selection procedure, difficulties in forming mentoring relations, or exclusion from important networks. The major characteristic of this explanation category is that it conveys a large part of the responsibility of the creation and maintenance of the "glass ceiling" from women themselves to society as a whole. Furthermore, it takes into consideration a large number of factors in order to explain the existence of the "glass ceiling", since it does not assume that organizations are neutral, but that certain assumptions about men and women are embedded into the organizational culture (Kakabadse A. and Kakabadse N., 1997). An example of a meta-research based on these principles by Jackson-Cooper (2001), found that the three major causal factors of the "glass ceiling" were specific organizational cultures, work-life imbalances, and tokenism-negative stereotypes.

Organizational culture is not gender-neutral. On the contrary, the composition of an organization's workforce affects the organizational culture, since the group that holds the majority – men or women – tends to push the organization into adopting several elements that favour the predominant sex (Jackson-Cooper, 2001). The authority in organizations that have a strongly "masculine" culture originates from the position that each individual has in the hierarchical ladder, while the culture itself has strong elements of individualism, absolutism, and top-bottom structures (Bajdo and Dickson, 2001). The strong pressure originating from the organizational gender composition that affects the organizational culture can render the latter hostile to the minority group –in most cases, women, and it can affect their career advancement. In a study between 400 women managers, 70% of them expressed the opinion that the male-crafted organizational culture hampered their advancement to the top, verifying the aforementioned assumptions about organizational culture (Zellner and Segal, 1992). To the extent that organizational culture tends to remain stable over time, it is difficult for women to overcome this disadvantage (Jackson-Cooper, 2001).

Strongly correlated with a corporation's organizational culture, are the unique recruitment and selection processes that the corporation employs. Many times, the recruitment procedures are poorly designed, leading to the exclusion of certain population groups, even though members of these groups may possess the required qualifications – women are often the victims of such methods (Adler, 1994). Furthermore, poorly designed selection procedures that exclude too many women or other minority members, create an adverse impact within the organization (Cook, 2004). Women often have to face the side-effects of adverse impact, especially in managerial-related occupations (Alimo-Metcalfe, 1990), as well as the negative stereotypes or prejudice of the selection managers. Davison and Burke (2000) found that the less training the selection managers had and the less information was given about the available positions, the stronger the negative stereotypes that were displayed against female candidates.

The problems that women have to face during the hiring selection process are quite similar to those of the promotion selection one. Promotions are important not only for the ordered functioning of a corporation, but also for its employees. They constitute a significant milestone in the vocational life of an individual, and they reward an individual's exceptional performance. Women are often the victims of irrational promotion selections (Cai and Kleiner, 1999). Unfair promotion practices that lead to the exclusion of women can lower their working morale as well as reduce their performance and their work satisfaction (Shrader and Blackburn, 1997).

Education, skills or special training can only offer entrance to the lower managerial ranks to women. The higher echelons of organizational hierarchies require more subjective criteria, such as participation in unofficial networks and access to powerful mentors (Adler and Izraeli, 1994). These criteria are vague, non-transparent, and they tend to favour men (Large and Saunders, 1995). For example, news about future vacancies for higher management jobs are usually circulated through unofficial networks that women are often excluded from and many of the selection criteria are designed in way that creates adverse and exclusionary effects on them (Adler and Izraeli, 1994).

Another possible factor for the creation of the "glass ceiling" is the problem of achieving work-life balance. All individuals are subjected to the side-effects of the internal conflict that contribute to and cause work life and personal life dilemmas or tensions. Greeunhas and Beutell (1985) describe this phenomenon as the internal conflict caused by the pressures of job and family roles that are in a certain degree, incompatible. Individuals with strong internal conflict may experience reduced work productivity, unhappiness and unrest in their personal life (Rau and Hyland, 2002). Women are especially vulnerable to work-life conflict pressures. For example, the long working hours that are required today in order to advance into the managerial ranks may cause women feelings of guilt because they cannot spend enough time with their children and families (Simpson, 1998).

The negative effect that a social role can have on another role is called the spill-over effect (Rothbard, 2001). The spill-over effect is directed from one's family role to one's work role that can create reduced performance, while an opposite-directed spill-over effect can create family turbulence and unsatisfactory parental performance (Frone, Yardley, and Markel, 1997). Since women experience motherhood and

childbirth, they are frequently vulnerable to falling victim to a work-life imbalance. Due to motherhood, women have to follow completely different career paths than men; also, these paths do not lead to the same destination, the top (Miller and Tsiantar, 1991). But even if women manage to juggle all their conflicting social roles, their success is not always recognized. The problem affects women that do not have families as well, since they are considered possible "victims" of work-life imbalance. For example, single women that had the same qualifications with men candidates for a promotion were more negatively evaluated compared to men, because men were considered to be a minimum risk choice (Mavin, 2001; White, 1995; Batt and Valcour, 2003).

Even though this problem has serious consequences not only for women but for the organizations as well, there are only very few studies around career paths that integrate the distinctive aspects of women's lives (Mavin, 2001; Still and Timms, 1998). Most of the existing career path models are based on the typical and linear course of a man's life: education-employment-retirement. These career paths that are so important for the evaluation and the advancement of an individual, are indiscreetly used both for men and women employees (Mavin, 2001). Women that have to deviate from that path due to motherhood, find themselves at a disadvantage with regard to their male colleagues when it comes the time for a promotion (O'Leary, 1997).

So far, several factors that lead to the creation of the "glass ceiling" have been mentioned. The connective link between them all is the negative stereotypes against women that prevail throughout social and organizational life. According to the basic cognitive model, individuals have a priori some fixed opinions on several matters. These opinions are called schemes or prototypes, and their main function is to act as filters in the information process in order to accelerate cognitive procedures. Stereotypes are a part of these prime schemas, and concern almost every expression of social life. In the business world, they are frequently used by managers or the corporate administration to justify or explain the inequalities between the men and women of their workforce. Traditionally, stereotypes that are associated with women are far more negative than those of men in terms of work skills and vocational behaviours (Wilson, 2003). Sometimes these stereotypes are so strong that women tend to accept them instead of trying to disavow them. A notorious stereotype is that women are worse managers than men and cannot perform as well. So far, several studies have debunked this stereotype (Tharenou, 1999). In fact, women may perform better than men, but because they are evaluated with stricter criteria than their male colleagues, their performance is constantly downgraded (Tharenou, 1999).

A set of stereotypes may form a unique stereotypical role (Kanter, 1977). Women can find themselves trapped into these roles not because they pursued them, but because they were imposed on them by their environments. Most of the time female managers are members of a skewed group, since they are a small minority compared to their men colleagues. In such cases, women automatically acquire a special token status that transforms them into symbols of their gender. They are not treated as unique individuals, but they are seen through a warped lens of stereotypes. This procedure is called tokenism and it has significant negative side-effects for women such as a strong pressure to perform well, stress, and unwanted attention (Kanter, 1977).

Apart from negative stereotypes, women have to face network seclusion. An individual's position in the social web, and respectively in an organizational web is called social capital and it may be profitable in more than one way (Brass, 1995). Social capital can be developed through participation to social networks, important hubs of information circulation – sometimes even more efficient than official corporation channels. Apart from fast access to important information, participants of a network can enjoy other privileges, such as support, increased attention, job-related advice, and encouragement. Networks are constructed under the influence of the organizational culture. Since men constitute the majority of the workforce, resulting in a strongly masculine organizational culture, this trend is passed on to the networks (Adler and Izraeli, 1994). As a result, women are often excluded or under-represented in these networks, missing the chance to increase their social capital or benefit from the unique advantages that networking has to offer (Maclaran, Stevens, and Catteral, 1997, Weiler and Bernasek, 2001).

2.2 Institutional theory

According to Scott (2001), institutions are multifaceted, durable social structures, made up of symbolic elements, social activities and material resources. Institutions are not invulnerable to change, however they tend to resist to external eroding pressures while at the same time they are transmitted across generations to be maintained and reproduced (Zucker, 1977). In a broader sense, the element of reproduction and transmission across generations can be construed in the wide framework of socialization that was mentioned earlier.

Institutional systems are variously comprised of regulative, normative, cultural-cognitive, and emotional elements that together with associated activities and resources create a unique combination (Scott, 2001). These elements have been over time identified by one or another scholar as a vital ingredient of institutions; Scott categorized them into what was later known as the system of three pillars of institutions. The emotional element is peripherally mentioned in that system, but for the purpose of the proposed research framework **it** is considered essential and, therefore, it is included as a unique forth pillar.

The pillars framework asserts that institutions are made up of diverse elements that differ in a number of important ways, since they offer different levels of order and compliance, varying mechanisms, diverse empirical indicators**,** and alternative approaches for establishing legitimacy claims. And although all institutions are composed of various combinations of these elements, the dominant element varies between the different institutional systems as well as over time (Scott, 2004). Different theorists also tend to privilege one or another of these elements. For example, most economists highlight regulative elements (North, 1990), early sociologists favoured normative elements (Selznick, 1949)**,** and more recent organizational sociologists emphasize the cultural-cognitive elements (Zucker, 1977; DiMaggio and Powell, 1991; Scott, 2001).

Every pillar has a unique content that defines it and differentiates it from the others. The regulative pillar contains explicit regulatory processes such as rule-setting, monitoring or sanctioning activities. It is concerned with the process by which rules are

established and how the society uses sanctions to control the behaviour of its members in an attempt to influence its own course. The sanctioning processes may operate through various informal and formal mechanisms. The regulative pillar reflects more the economists' view for the institutions, since according to its content; an organization will adopt an institution if this institution is expected to yield profit. Furthermore, because of the competitive interests that overrule almost every social function, e.g. the economy, the society needs to adopt institutions that will ensure the unhindered continuation of these functions, by posing specific rules on the social actors in order to maintain order (Scott, 2001).

The second pillar, the normative, functions in a mutually reinforcing way with the first pillar. It includes values and norms that introduce specific boundaries into social life. Values are socially acceptable conceptions of what is preferred or desirable, offering standards to which social structures or social actors' behaviour can be compared and assessed (Kluckhohn, 1951). Norms on the other hand specify how things should be done and they define socially legitimate means by which individuals or organizations can pursue these objectives. In a way, the normative pillar is concerned with the definition of socially desirable goals and the designation of the appropriate ways to pursue these goals. It is essential to identify the social groups that each value or norm is applicable to, since different sets of values and norms are accepted or imposed to each group and sometimes these sets can be so divergent that they can lead to social turbulence. Values and norms function on an organizational level as well as on individual one. Every organization has to choose which operational processes it will adopt according to its unique internal value system, but, at the same time it has to ensure that these processes are agreeable to the external institutions (Meyer and Rowan, 1977). In general, normative systems impose constraints on social behaviour but at the same time they establish social action by consolidating rights, responsibilities, privileges, and duties.

The cultural-cognitive pillar is concerned with the conceptions that constitute the nature of social reality and the procedures through which meaning is appointed. It emphasizes the cognitive dimension of social life, and how the external world stimuli and the responses of the individuals form internalized symbolic representations of the world (Scott, 2001). For example, Max Weber regarded action as social only to the extent that the social actor attached a specific meaning to the behaviour. To understand or explain any action, one must take into consideration not only the objective conditions but also the social actor's subjective interpretation of them. Compliance to the current institutions is not always a conscious procedure, since most of the institutional symbols or functions are deeply embedded within individuals through socialization that starts from an early age. In a similar way, organizations adopt institutions that are widely established into the social-cognitive environment are expected to operate simply because "this is the way things should be done" (Scott, 2001).

The final pillar, the emotional one, refers to the sentimental "bonding" that individuals as well as organizations on a collective level develop with specific practices, behaviours or relationships. This sentimental attachment is applicable to institutions as well, and it is especially important in times of change because it is one of the most common causes of resistance to change (Oliver, 1991; Scott, 2001). Each of the four aforementioned elements blends together with the others to create the unique identity

of a society or an organization. Furthermore, the broader cultural-cognitive, normative, regulatory, and emotional aspects of institutions shape the nature of competition and of markets, as well as the meaning of effective performance and efficient operation (Powell, 1991; Whitley, 1992). This equilibrium tends to be stable over time, but that does not mean it is not subject to change.

The feature of change is dominant throughout social life, and institutions do not comprise an exception. Institutional arrangements change all the time, since they are subject to eroding or degenerating forces, and require a continuous input of energy and resources to prevent decay and obsolesce. However, it is important to observe both the beginning and the end of institutions, since in the construction and deconstruction phases, conflict and agency mechanisms are likely to be more visible. There are several causes of change of the institutional arrangements. For one, the elements of institutions themselves may not be aligned and one may undermine the effects of the other. Furthermore, the boundaries of organizational fields are often invisible or easily penetrated, allowing alternative approaches to emerge and support divergent models of behaviour. Finally, since each institutional arrangement serves better the interest of one group compared to the other, underprivileged groups may mobilize and successfully promote new models of structures and actions. Related to the "glass ceiling" for example, if the sub-group of female employees of a corporation feels wronged due to negative stereotypes or excluded from the upper managerial echelons, it can demonstrate its opposition to the current situation, by choosing to either leave the company or reduce its productivity. In conclusion, it is obvious that competing rules or schemas open up possibilities for bargaining among social actors and groups that lead to the adoption of new institutions and the obsolesce of old ones.

3. The Institutional Pillars as a hermeneutic framework for breaking the "glass ceiling"

Organizations embrace new practices and set new policies and strategies. The question is why do they do it and with what mechanisms? In particular the present model aims at responding to the question of what is the motivational basis for breaking the "glass ceiling" and what are the factors, mechanisms and processes involved with it? The present study argues that organizations use a mixture of expedience or self-interest, social obligation, shared understanding, and emotional attachments as a motivational basis. Furthermore, the present model focuses on the role of internal and external factors related to this particular organizational change. The internal factors are the actors within the organization while the external factors are the societal institutions. Internal factors are based on coercive, normative, mimetic, and sympathetic mechanisms and use mainly the process of invention and negotiation. The external factors are based on a mixture of coercive, normative, and cultural mechanisms and use the process of imposition and diffusion. Organizations respond to change factors using a mixture of expedience or self-interest, social obligation, shared understanding, and emotional attachments as a motivational basis (Figure 1).

Based on the regulative pillar, organizations adopt the endeavour of breaking the "glass ceiling" because they judge that it will have an overall positive impact. The benefits for adopting this practice will overweight the cost of not adopting it. The motivational

basis is expedience and self interest. The benefits for breaking the "glass ceiling" vary. The "glass ceiling" creates significant costs for the corporations; it is one of the main causes of diminished utilization of a company's workforce since it leads to failure to fully develop and use the talents of women, lost profits, lack of diversity in the workforce, decreased productivity, and increased hiring and training costs (Schwartz, 1989; Adler et al, 2000). On the other hand, human resources are one of the few elements that can be used to create a strong to imitate competitive advantage since they are unique and difficult to copy. The only way to achieve that is to fully employ the power of women in the workforce (Harel, Tzafrir, and Baruch, 2003). Companies that recruit, maintain, develop and promote the women of their workforce have been found to consequently develop a strong competitive advantage (Caputo and Dolinski, 1998; Weiler and Bernasek, 2001; Collins, 2004). Given the positive economic effects of breaking the "glass ceiling" for the companies, we would expect that companies, based on calculation alone, would rush to embrace this practice. Another benefit of breaking the "glass ceiling" may be that the organization avoids costs stemming from internal or external factors which use coercive mechanisms to impose pressure for change; such coercive mechanisms are, for example, hiring laws outlining equal-opportunity employment.

Based on the normative pillar, organizations break the "glass ceiling" because this is the right thing to do; in this case, the motivational basis for breaking the "glass ceiling" is social obligation. Mechanisms of societal factors have diffused their norms and values about gender equality. Internal actors through inventive practices and negotiations have exerted normative, organizational changes. Female employees and the organization have binding expectations of which the realization makes the organization a credible employer.

Based on the cultural-cognitive pillar, organizations break the "glass ceiling" because they have internalized specific symbolic representations related to gender; internalization occurs through mimetic processes and is shaped by external cultural frameworks. Also internal actors who have internalized specific, symbolic representations related to gender help an organization's internalization process through invention and negotiation. Once the organization is based on the cultural-cognitive pillar for breaking the "glass ceiling", it is taken for granted that that this is the way the organization is doing things and the any other way is inconceivable.

Finally, organizations that break the "glass ceiling" may be based on the emotional pillar; in this case the motivational reasons are emotional attachments to this type of behaviour. Through the mechanism of sympathy, significant internal actors using invention and negotiation help organizations to break the "glass ceiling". The change is adopted because it has emotional meaning to individuals such as the CEO or the owner of the organization; it is change guided by "habits of the heart" (Bellah et al., 1985). Scholars such as Durkheim and Parsons have acknowledged the importance of emotional attachments as a motivational basis for action (Scott, 2001).

In conclusion, the goal of breaking the "glass ceiling" as being adopted by an organization may be motivated by a combination of expedience, social obligation, internalized meanings, and emotional attachments stemming from external, societal factors and/or internal actors using different mechanisms and different processes. The proposed model is contingent upon the combination of organization's motives and its

response to pressures from various external and internal forces, mechanisms and processes.

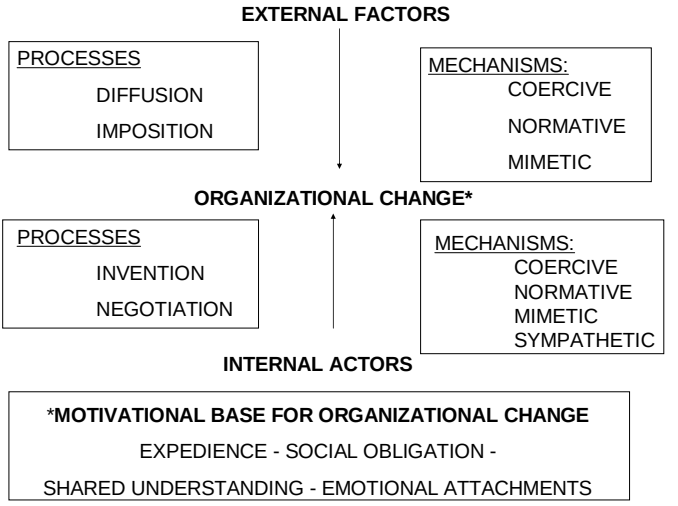


Figure 1. A neo-institutional model for breaking the glass ceiling

References

1. Adler, J. N. (1994) 'Competitive Frontiers: Women managing across Borders', Journal of Management Development, Vol. 13, No. 2, pp. 24-41.
2. Adler, J. N. and Izraeli, N. D. (1994) Competitive Frontiers: Women Managers in a Global Economy, Cambridge: Blackwell Publishing.
3. Adler, J. N., Brody W. L. and Osland S. J. (2000) 'The women's global leadership forum: enhancing one company's global leadership capability', Human Resource Management, Vol. 39, No. 2, pp. 209-225.
4. Alimo-Metcalfe, B. (1990) 'An investigation of female and male constructs of leadership and empowerment', Women in Management Review, Vol. 10, No. 2, pp. 3-8.
5. Alimo-Metcalfe, B. (1993) 'Women in Management: organizational socialization and assessment practices that prevent career development', International Journal of Selection and Assessment, Vol. 1, No. 34, pp. 68-83.
6. Bajdo, M. L. and Dickson, W. M. (2001) 'Perceptions of Organizational Culture and Women's Advancement in Organizations: A Cross-Cultural examination', Sex Roles, Vol. 45, No. 5, pp. 399-414.
7. Batt, R. and Valcour, P. M. (2003) 'Human Resource Practices as Predictors of Work-Family outcomes and Employee Turnover', Industrial Relations, Vol. 42, No. 2, pp. 189-220.

8. Bellah, N. R., Madsen, R., Sullivan, M. W., Swidler, A. and Tipton, M. S. (1985) Habits of the heart: individualism and commitment in American life, Berkeley: University of California Press.
9. Brass, D. J. (1995) 'A social network perspective on human resource management', Research in Personnel and Human Resource Management, Vol. 13, No. 1, pp. 39-79.
10. Browne, R. K. (1998) 'An Evolutionary Account of Women's Workplace Status', Managerial and Decision Economics, Vol. 19, No. 3, pp. 427-440.
11. Cai, Y. and Kleiner, H. B. (1999) 'Sex Discrimination in hiring: The «glass ceiling»', Equal Opportunities International, Vol. 18, No. 2, pp. 51-55.
12. Caputo, K. R. and Dolinsky A. (1998) 'Women's choice to pursue self-employment: the role of financial and human capital of household members', Journal of small business management, July 1998.
13. Collins, K. J. (2004) 'Women at the Top', Business & Economic Review, Vol. 5, No. 4, pp. 20-21.
14. Cook, M. (2004) Personnel selection: adding value through selection, Sussex: John Wiley & Sons Ltd.
15. Davison, K. H. and Burke, M. J. (2000) 'Sex discrimination in simulated employment contexts: a meta-analysis', Journal of Vocational Behavior, Vol. 56, No. 2, pp. 225-248.
16. Dempster, L. and Paton R. (2002) 'Managing Change from a Gender Perspective', European Management Journal, Vol. 20, No. 5, pp 539-548.
17. DiMaggio, J. P. (1988) Interest and agency in institutional theory – Institutional patterns and organizations: culture and environment, Cambridge: Ballinger.
18. DiMaggio, J. P. and Powell, W. W. (1991) Constructing an organizational field as a professional project – The new institutionalism in Organizational Analysis, Chicago: University of Chicago Press.
19. Epstein, F. C., Olivares F., Bass, B., Graham, P. and Schwartz, N. F. (1991) 'Debate – it is time to stop talking about gender differences? : Ways Men and Women Lead', Harvard Business Review, Vol. 69, No. 1, pp. 150-160.
20. Fisher, H. (1999) The First Sex: the natural talents of women and how they are changing the world, New York: Random House.
21. Frone, R. M., Yardley, K. J and Markel, S. K. (1997) 'Developing and testing an integrative model of the work-family interface', Journal of Vocational Behavior, Vol. 50, No. 2, pp. 145-167.
22. Greeunhaus, H. J. and Beutell, J. N. (1985) 'Source of conflict between work and family roles', Academy of Management Review, Vol. 10, No. 1, pp. 76-88.
23. Harel, H. G., Tzafrir, S. S. and Baruch, Y. (2003) 'Achieving organizational effectiveness through promotion of women into managerial positions : HRM practice focus', International Journal of Human Resource Management, Vol. 14, No. 2, pp. 247-263.
24. Harris, H. (2002) 'Think International Manager, Think Male: Why are women not selected for international management assignments?', Thunderbird International Business Review, Vol. 44, No 2, pp. 175-203.

25. Hymowitz, C. and Schellhardt, D. T. (1986) 'The «glass ceiling»: why women can't seem to break the invisible barrier that blocks them from the top jobs', The Wall Street Journal, March 24.
26. International Labour Organization (2008) Global employment trends for women, Geneva: ILO.
27. International Labour Organization, (2004) Breaking through the «glass ceiling»: women in management, Geneva: ILO.
28. Jackson-Cooper, J. (2001) 'Women's middle manager's perception of the «glass ceiling»', Women in Management Review, Vol. 16, No. 1, pp. 30-41.
29. Kakabadse, A. K. and Kakabadse, N. (1997) 'Best practice in the Australian public service (APS): an examination of discretionary leader- Best practice in the ship', Journal of Managerial Psychology, Vol. 12, No. 7, pp. 431-491.
30. Kanter, R. M. (1977) 'Some Effects of Proportions on Group Life: Skewed Sex Ratios and Responses to Token Women', American Journal of Sociology, Vol. 82, No. 5, pp. 965-990.
31. Kluckhohn, C. (1951) 'Values and value-orientations in the theory of action: An exploration in definition and classification' In: Toward a General Theory of Action, Cambridge, MA: Harvard University Press.
32. Large, M. and Saunders, N. M. (1995) 'A decision-making model for analysing how the «glass ceiling» is maintained: unblocking equal opportunities', The International Journal of Career Management, Vol. 7, No. 2, pp. 21-28.
33. Maclaran, P., Stevens, L. and Catteral, M. (1997) 'The 'glasshouse effect': women in marketing management', Marketing Intelligence & Planning, Vol. 15, No. 7, pp. 309-317.
34. Mavin, S. (2001) 'Women's career in theory and practice: time for change?', Women in Management Review, Vol. 16, No. 4, pp. 183-192.
35. Meyer, W. J. and Rowan, B. (1977) 'Institutionalized organizations: formal structure as myth and ceremony', American Journal of Sociology, Vol. 83, No. 4, pp. 340-363.
36. Miller, A. and Tsiantar D. (1991) 'Mommy tracks', Newsweek, November 25, pp. 48-49.
37. Morrison, A. M. and Von Glinow, M. (1990) "Women and Minorities in Management', American Psychologist, Vol. 45, No. 2, pp. 200-208.
38. North, C. D. (1990), Institutions, Institutional Change and Economic Performance, Cambridge: Cambridge University Press.
39. O'Leary, J. (1997) 'Developing a new mindset: the "career ambitious" individual', Women in Management Review, Vol. 12, No. 3, pp. 91-99.
40. Oliver, C. (1991), 'Strategic Responses to Institutional Processes', The Academy of Management Review, Vol. 16, No. 1, pp. 145-179.
41. Powell, N. G. (1990) 'One more time: do female and male managers differ?' Academy of Management Executive, Vol. 4, No. 3, pp. 68-75.
42. Powell, W. W. (1991) Expanding the scope of institutional analysis – The new institutionalism in Organizational Analysis, Chicago: University of Chicago Press.
43. Rau, L. B. and Hyland, M. M. (2002) 'Role conflict and flexible work arrangements: the effects on applicant attraction', Personnel Psychology, Vol. 55, No. 2, pp. 111-134.

44. Rothbard, P. N. (2001) 'Enriching or Depleting? The dynamics of engagement in work and family roles', Administrative Science Quarterly, Vol. 46, No. 1, pp. 655-684.
45. Rosener, B. J. (1990) 'Ways Women Lead', Harvard Business Review, Vol. 68, No. 6, pp. 119-125.
46. Schwartz, N. F. (1989), 'Management Women and the New Facts of life', Harvard Business Review, Vol. 67, No. 1, pp. 65-76.
47. Scott, W. R. (2001) Institutions and Organizations, London: Thousand Oaks.
48. Scott, W. R. (2004), Institutional theory: contributing to a theoretical research program, Palo Alto: Stanford University Press.
49. Selznick, P. (1949) TVA and the Grass Roots, Berkeley: University of California Press.
50. Schrader, B. C. and Blackburn, B. V. (1997) 'Women in management and firm financial performance: an exploratory study', Journal of Managerial Issues, Vol. 9, No. 3, pp. 355-372.
51. Simpson, R. (1998) 'Presenteeism, Power and Organizational Change: long hours as a career barrier and the impact on the working lives of women managers', British Journal of Management, Vol. 9, No. 1, pp. 37-50.
52. Still, L. and Timms, W. (1998) 'Career barriers and the older woman manager', Women in Management Review, Vol. 13, No. 4, pp. 143-155.
53. Tharenou, P. (1999) 'Gender differences in advancing to the top', International Journal of Management Reviews, Vol. 1, No. 2, pp. 111-132.
54. Tharenou, P. (2001) 'Going up ? Do traits and informal social processes predict advancing in management?' Academy of Management Journal, Vol. 44, No. 51, pp.1005-1015.
55. Tooby, J. and Cosmides, L. (1992) The psychological foundations of culture - The adapted mind: Evolutionary psychology and the generation of culture, New York: Oxford University Press.
56. Vilkinas, T. (2000) 'The gender factor in management: how significant others perceive effectiveness', Women in Management Review, Vol. 15, No. 5, pp. 261-271.
57. Weiler, S. and Bernasek, A. (2001) 'Dodging the «glass ceiling»? Networks and the new wave of women entrepreneurs', The Social Science Journal, Vol. 38, No.1, pp. 85-103.
58. White, B. (1995) 'The career development of successful women', Women in Management Review, Vol. 10, No. 3, pp. 4-15.
59. Whitley, R. (1992) The social construction of organizations and markets – Rethinking Organizations: new directions in organization theory and analysis, Los Angeles: Sage Publications.
60. Wilson, M. F. (2003) Organizational Behaviour and Gender, London: Ashgate Publising.
61. Zellner, W., Segal, A. T. (1992) 'Corporate women: Progress? Sure. But the playing field is far from level', Business Week, No. 3269, pp. 74-79.
62. Zucker, G. L. (1977) 'The role of institutionalization in cultural persistence', American Journal of Sociology, Vol. 42, No. 5, pp. 726-743.

Direct Foreign Investment in The Czech Republic

Milan Damborský and Gabriela Říhová

University of Economics, Prague, Faculty of Economics and Public Administration, Regional Studies Dpartment, mdamborsky@centrum.cz, rihg01@vse.cz

The target of this article is to describe briefly the development of direct foreign investment inflow into the Czech Republic and its relation with the regional development. The authors worked on the assumption that in case of economically powerful regions, the high inflow of direct foreign investment impacts positively on the other acceleration of economic growth and increase of interregional differences in the Czech Republic as well. And in case of economically poorer regions, the low inflow of direct foreign investment impacts as a factor of poor economic growth. The authors work with the regional data about the direct foreign investment which were released by the Czech National Bank in March 2008. These data, that are available in public up to now, describe the status of 2006. The limited data basis of Czech National Bank and Czech Statistical Office (like publishing the data with high lapse of time) limits significantly the possibility of economically oriented analysis currency evaluating the interregional differences.

Keywords. Direct foreign investment, localization factors, interregional differences, economic growth

1. Introduction

The significant administrative barriers, which had to allow the control of the capital flow in and out of a country itself, were made for a long term. In connection with prevailing of more liberating approach in the area of economic policy, there was a significant increase of capital flow between the single states in 80's.

In connection with the change of government attitude to the direct foreign investment, the public administration tries to form the capitally attractive environment for the foreign investors, i.e. to mobilize the localization factors in administrated area. The localization factors are the specific features of the given places that have the impact on the location of socio-economic activities (e.g. by way of the influence on rentability of economic activities). The localization is under the control of not just one, but more factors. The factors are not fixed, the new factors constantly arise and the other die out. The localization factors can be divided into several classifications. B. Grabow, D. Henckel and D. Hollbach-Grömig is the hard ones (measurable) and the soft ones (not measurable, the psychological and sociological substances). This classifications follow

up the older ones on the economic (objective) and uneconomic (subjective). According to H. A. Stafford, the investors, when selecting the region (they go after the maximization of sale), at first decide according to the place with the lowest costs (Damborský, 2007).

2. Methodology

There were used the method of data analysis (promulgated by CNB and CSO) for total impact appreciation of the foreign direct investments on the regional development, the study of the available documents about the former foreign direct investments and the proper research based on the investigation of the foreign direct investment impacts in the chosen examples.

The concept of the direct foreign investment in this article agrees with the concept of the direct foreign investment of the economic organisation OECD which is in accord with the concept of the EUROSTAT and MMF direct foreign investment: "The direct foreign investment reflects the resident intention of one economics (the direct investor) to gain the constant participation in the subject, which is a different economics resident than the economics resident (the direct investment). The constant participation implies the existence of a long-term relation between the direct investor and the direct investment and the important influence on the enterprise operation. The direct investment includes both the original transaction between both subjects and all following capital transaction between them and afilated enterprises registered in business register."

3. Results

In the initial period, most of the direct foreign investment was comprised from investment into the purchase of the state participation, was oriented on the standing production and on the expanding of assembling phases exportably oriented sections, mainly of the manufacturing industry. Afterwards and in the last period, the quality shift is obviously shown in the structure of the direct foreign investment, when the foreign companies invest into the areas with the high added value (the software development, the IT services, the call-centres, the shared centres for accounting and financial operations), which nowadays allow an application of the highly qualified labour work. After the initial significant deposits of the investors into the registered capital of the Czech enterprises, there are the gradual changes of the direct foreign investment structure, where the reinvest profit holds the considerable function that the investors insert into the further extent of the companies (MPO, 2008:57).

From the beginning of the 90's till the end of 1997, the direct foreign investment in the Czech Republic reached up to $9,200 000,000 which was less than in Poland ($14,600,000,000) and Hungary ($15,900,000,000). In connection with the privatization sales of the big Czech bank's state participation into the possession of the strong foreign strategic partners, with the sale of big companies' state participation, with the arrival of the commercial chain and accepting of the investment incentive system (including the income tax abatement, the grants on the creation of the new job

opportunities, the grants on the training and retraining of the employees and the grants for the municipalities on the technical equipment) there was a more massive inflow of the investment in 1998. In the recent years, the direct foreign investment has been moving around $6,000,000,000 per year. The years 2002 and 2005 were the exceptions, when the direct foreign investment reached more than $8,000,000,000 and the year 2003, when it reached only $2,100,000,000 which was given by the repurchase of the Eurotel stocks from the foreign investor (MPO, 2008:57, Toušek a kol. 2005:36).

From 1993, the main source of the direct foreign investment the reinvested profit. Its importance is still growing. This trend is connected with the positive expectation of the investors and profitability of the foreign investment (MPO, 2008:57).

The direct foreign investment had an important role at the creation of the standard market economy and, to a considerable extent, stand behind the good condition of the Czech economy with the changing industry structure and services development and constantly increasing competitiveness of its production on the foreign and home market. While the direct foreign investment is, considering to the small home market, oriented to the manufacturing industry especially expertly, the target of the direct foreign investment to the services was to prevail on the home market (MPO, 2008:57).

The direct foreign investment is nowadays the significant source of the job production and not only at the previous investors, but also at the home companies which gain the supplier contracts. The economists agree with the opinion, that the foreign investment has the positive impact on the growth of the productivity of labour in Czech economics. Thanks to the direct foreign investment, there is the growth of the employment rate, growth of the productivity of labour, recovery and improvement of the labour market, the positive spillover effects (MPO, 2007A). The high direct foreign investment is becoming the growth pole (e.g. Skoda Car, JSC).

The direct foreign investment inflow into the Czech Republic reached $185,300,000,000 in 2007 (i.e. $9,300,000,000) and it was interannually by 36.4% higher. From this, Kč 57,500,000,000 (i.e. $2,800,000,000) averaged the increase of the subsidiary companies' registered capital, Kč 130,600,000,000 (i.e. $6,400,000,000) amounted the reinvested profit, investment in form of the rest of capital decreased (by 2,800,000,000). The investments from the EU countries predominated significantly the rest. They came from the Netherlands (Kč 53,400,000,000), Luxembourg (Kč 23,800,000,000), Austria (Kč 21,700,000,000), France (Kč 19,900,000,000) and Germany (Kč 15,400,000,000). The highest capital inflow from the out-European countries was from Korea (Kč 7,900,000,000) – in connection with the Hyundai car factory construction in north-Moravian town of Nošovice, USA (Kč 7,000,000,000) – in form of reinvested profit and the rest of capital – and Japan (Kč 5,000,000,000). From the branch point of view, the increase of registered capital by the foreign owners took place especially in the property purchase (big development projects) and providing enterprises services branch. It was headed towards the financial intermediation and the chemical products and two-track motor vehicle production (MPO, 2008:57).

The investors, who moved their products into the Czech Republic, are motivated to move their activities further to the west into the countries like Romania, Bulgaria,

Ukraine and into the countries in the south-eastern Asia. In connection with this movement, the question of reconversion is being solved. The bigger problem with reconversion of the areas arisen under the terms of the direct foreign investment inflow in spite of that we can record the part outflow of the foreign investors from the Czech Republic hasn't set in yet. The progress of the investor departure can be supported with evidence on the investment development under the terms of fi. Flextronics, Brno that remained behind the expectation.

The movement process (delocalization) of the labour-demanding production and assembling activities into the developing countries with the low costs is characteristic for Western Europe and Northern America already from the 60's. This can be supported with evidence e.g. on the quick textile or electronic industry development in the eastern-Asian countries, e.g. Southern Korea, Taiwan, Malaysia that absorbed the significant capacity of the direct foreign investment, especially from the USA (Dicken, 2003, Ženka, 2008:1).

Nowadays, the traffic accessibility, an offer of the appropriate areas for business, the corresponding offer of labour are emphasised like the basic localization factors present in the Czech Republic for the direct foreign investment. Especially the "traffic services" localization factor is very important for the high direct foreign investment in the Czech Republic. This increases the importance of the region traffic infrastructure, especially the connection on the international highway and railway system (Toušek, 2005:36). In the past years, privatization was very important motivation and with it connected unique investment opportunity.

Region	accumulated state of the direct foreign investment per one inhabitant in 2006
Praha	Kč 748,028
Middle-Bohemian region	Kč 157,672
Liberecký region	Kč 109,755
Region Vysočina	Kč 99,895
Moravian-Silesian region	Kč 98,263
South-Bohemian region	Kč 96,420
Plzeňský region	Kč 96,063
Pardubický region	Kč 75,080
Ústecký region	Kč 73,851
Jihomoravský region	Kč 61,843
Karlovarský region	Kč 53,592
Zlínský region	Kč 50,250
Olomoucký region	Kč 41,512
Královehradecký region	Kč 39,252

Table 1. Accumulated state of the direct foreign investment per one inhabitant in 2006, the source: CNB, recalculation the team of the authors 2008

Not all regions are attractive for this direct foreign investment. The most attractive region for the direct foreign investment is Prague. It has also the best values in most of the state macroeconomic index as well. It shows the lowest unemployment rate (2007,

2.5%), the highest GDP per one inhabitant (2006, Kč 666,815) and net disposable income per one inhabitant (2006, Kč 204,845). Prague is also the Czech Research Centre.

The second most attractive region for the direct foreign investment is the Middle-Bohemian region. This region can be marked as the Czech most attractive region of migration. In the Middle-Bohemian region, the amount of the inhabitants increased by 5% from 2000 till 2007, which is the highest value between the regions, note the data CSO sources (recalculation Milan Damborský).

The third highest value of the accumulated state of the direct foreign investment in 2006 per one inhabitant reached the Libereský region, Kč 109,755 (CNB, 2008; recalculation the team of the authors, 2008). The Liberecký region belongs to the ordinary regions according to the view of the socioeconomic values.

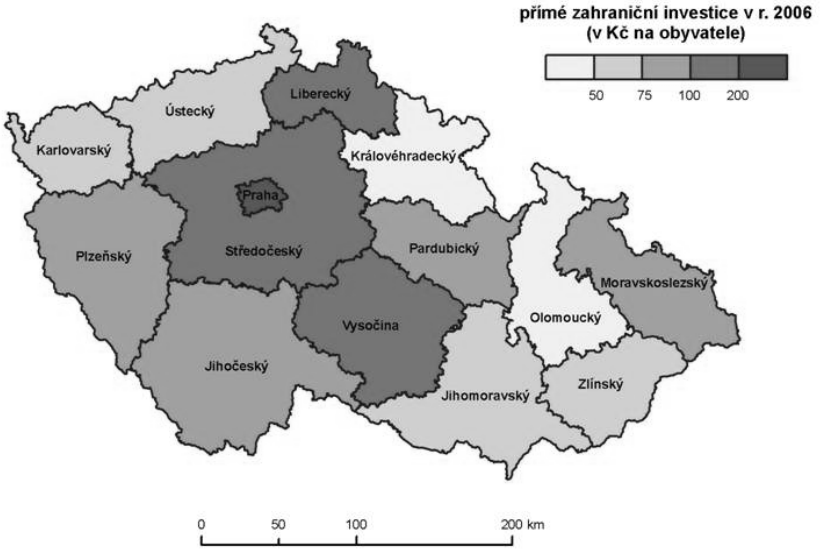


Figure 1. The direct foreign investment in 2006, data source: CNB 2008, an author: Tomáš Krejčí

4. Conslusion

From the above-mentioned presumptions implies that the direct foreign investment has the positive impact on the regional economics. The enterprises owned by foreign capital reach the higher productivity on one worker than at the home enterprises. Foreign capital brings the technological and managerial know-how that allows the more effective functioning of an enterprise. It indirectly impacts on both the foreign investment and the increasing productivity of the whole concrete branch because there is an overflow of know-how from abroad on the home enterprises. The direct foreign

investment contributes to the knowledge of economics development. The capitally powerful foreign enterprises offer the possibility of cooperation in area of innovations, research and development to the academic sphere, especially to the Universities. A contract with a foreign investor means a chance for the home supply houses to participate in the global production chains and to gain the access to the further orders.

According to the above-mentioned can be stated that in case of the capital Prague, Middle-Bohemian, Královehradecký, Zlínský, Karlovarský, Olomoucký and Ústecký regions the direct foreign investment inflow contributes to the regional increase of disparities, i.e. in case of the economically efficient regions the high direct foreign investment inflow impacts positively on the further acceleration of the economical growth and in this way also on the increase of the interregional differences in the Czech Republic and in case of the poorer regions the low direct foreign investment inflow impacts as one of the factors of the poor economic growth. In case of the South-Bohemian, Plzeňský, Moravian-Silesian regions the direct foreign investment inflow contributes to the decrease of the regional disparities and in case of the Liberecký, Pardubický, South-Moravian and Vysočina regions it is impossible on the basis of the comparison of the position according to the direct foreign investment inflow and the position according to the socio-economic level to set whether the direct foreign investment inflow is the reason of the increase or decrease of the Czech regional differences.

References

1. Blomström, M. and Kokko, A. (1998) ´Multinational Corporations and Spillovers´ *Journal of Economic Surveys*, No. 12., pp. 247-277.
2. Česká národní banka (2008). Přímé zahraniční investice 2006. Internet: http//www.cnb.cz/, [accessed 15. June 2008]
3. Český statistický úřad (2008). Internet: http//www.czso.cz /, [accessed 15. June 2008].
4. Damborský, M. (2007) Location and Spatial Development. In: 2nd Central European Conference in Regional Science, European Regional Science Association, pp. 160 – 179.
5. Damborský, M., Jetmar, M., Kouřilová, J., Říhová, G. (kol. autorů, 2008) ´Přímé zahraniční investice v ČR´. Prezentace na semináři Přímé zahraniční investice v ČR a SR. Vysoká škola ekonomická v Praze. 24. June 2008.
6. Dicken, P. (2003), Global Shift: Reshaping the Global Economic Map in the 21st Century, London: Sage, pp. 632.
7. Hlavenka, J. (2002), *Flextronics to balí: když je přepobídkováno*, Internet: http://www.zive.cz/default.aspx?article=107640, [accessed 12. July 2002].
8. Jetmar, M., Kouřilová, J. (2006) ´Přímé zahraniční investice jako specifický nástroj regionální politiky´, *In*: Jetmar, M., Kouřilová, J. (ed), *Management regionální politiky a reforma veřejné správy*, (pp. 234 – 240), Linde Praha.
9. Ministerstvo pro místní rozvoj (2006), Strategie regionálního rozvoje ČR (2007 – 2013). Internet: http://www.mmr.cz/, [accessed 1. April 2006]
10. Ministerstvo průmyslu a obchodu (2007), Vliv přímých zahraničních investic (tisková zpráva), Internet: http://www.mpo.cz/, [accessed 14. May 2007].

11. Ministerstvo průmyslu a obchodu (2007), Vyhodnocení investičních pobídek ve zpracovatelském průmyslu, Internet: http://www.mpo.cz/, [accessed 13. April 2007]
12. Ministerstvo průmyslu a obchodu (2008), Analýza vývoje ekonomiky ČR a odvětví v působnosti MPO za rok 2007, Internet: http://www.mpo.cz/ [accessed 25. March 2008].
13. Toušek, V., Smolová, I., Fňukal, M., Jurek, M., Klapka, P. (2005) *Česká republika – portréty krajů*, Ministerstvo pro místní rozvoj ČR, pp. 136.
14. Vančura, M. (2007) ´Direkte Auslandsinvestitionen, Investitionsanreize und Zielregionen in der Tschechischen Republik´, *In*: Jurczek, P. (ed.), *Beiträge zur Kommunal - und Regionlaentwicklung, Sachsen und Tschechien - zwischen Konkurrenz und Zusammenarbeit aus geographischer Sicht*, (pp. 131 – 146), Heft 45, Chemnitz, Technische Univerzität Chemnitz.
15. Wokoun, R. a kol. (2008) *MasterCard Czech Centers of Commerce*, Internet: http://www.centrarozvoje.cz/, [accessed 25. March 2008].
16. Ženka, J. (2008) ´Riziko delokalizace zpracovatelského průmyslu Česka: Regionální aspekty´, *In*: Ženka, J. (ed.), *Geografie – sborník České geografické společnosti*, (Vol. 113, No. 1., pp. 1 – 19).

Assessing HRM Policies and Practices in Different National Contexts within the International Hospitality Industry: A Survey of HR Managers of a Global Luxury Hotel Chain

Giovanni Oscar Serafini

South East European Research Centre, 17 Mitropoleos Street, 54624, Thessaloniki, Greece, giserafini@seerc.org

This article aims at laying the theoretical grounds to an ensuing research work in the field of HRM applied to the global luxury hotel industry. The review of literature leads to the conclusion that business and context specificities determine a gap between HRM in theory and HRM in practice. This divergence can be reduced provided that HRM assumes a strategic role with the coherent support of both top leadership and line managers, which is essential in order to maximize the contribution of the human capital to ongoing competitive advantage and organizational success.

Originality: This paper aims to enhance understanding of the HRM model used by a global luxury hospitality concern as well as of the implementation of HRM policies and procedures across the differing national contexts where the company operates.

Value: To the best of the researcher's knowledge, there has been no empirical study addressing HRM policies and procedures in a global hospitality chain while investigating HR professionals exclusively.

Keywords: HRM policies and procedures; HRM implementation; HR and organizational performance; employee commitment; strategic HRM; global organization; luxury hotel industry; service quality; organizational culture; leadership; business strategy.

1. Introduction

In recent years, the number of organizations providing service has been increasing impressively in the Western world. In particular, statistics reveal that, while three out of four jobs in the US relate to the services sector, about half of the family expenses concern services and nine out of ten new jobs created belong to the service industry (Goodman and Steadman, 2002). Also, this emphasis towards services provision is further highlighted by the fact that even the manufacturing sector is steering heavily towards it (Dwyer, 2008, Gebauer, 2007).

This research paper focuses on hotels, a thriving area of the services industry which by definition is labor-intensive. In fact, the success of hospitality organizations critically relies on its workforce and its capabilities in terms of dedication to work, commitment, resilience, creativity and attitude (Hoque, 1999). Consequently, successful workforce management represents an essential pre-requisite to ensure the sustained attainment of organizational goals.

Notwithstanding some cautious voices urging for more conservative estimations (Leiper, 1999), still the hospitality industry is one of the most important economic sectors of the global economy, as in 2008 it employed over 225 million people around the world and generated 9.6% of global GDP (WTTC, 2009). Thus, it represents a most fertile ground for enquiry in International Human Resources Management (IHRM) and, especially, for comparing how firms manage their people in different national contexts.

To date, there is limited work on the international hospitality industry, and the opportunity to examine closely a major global luxury hotel chain would prove a valuable case study. This conviction is rooted with the fact that commitment to service excellence does compel hospitality organizations to approach workforce management from a strategic standpoint thus allowing human resource management policies and practices to emerge (Maxwell et al., 2004).

This paper seeks to assess the Human Resource Management (HRM) policies and practices in different national context within a global luxury hotel chain in order to gain information on differences and similarities across countries, between unit-level and corporate-level HR managers, as well as precious insight pertaining to IHRM. Thus, the research instrument to be used will be a survey of its HR managers at both headquarters and field levels. The company under consideration is a global de-luxe hotel chain that, for the purpose of safeguarding anonymity, will be referred to as the 'participant company' or the 'hotel company'.

The structure of this full paper for the 4th SEERC Doctoral Conference is made up of four main sections. Initially, there is a review of HRM literature in general which then shifts focus towards HRM applied to the hospitality industry and the challenges faced by organizations spread throughout the world. Second, it illustrates the intended research purposes, design and methodology before examining the forecast empirical evidence to be gathered through a survey of HR managers. The last part of the study is centred around possible findings and research limitations.

In the final analysis, by obtaining access into a selected hospitality concern, study outcomes would prove to be of great potential value and interest both internally to the organization for its perusal as well as externally, by adding further knowledge to be used by academics and practitioners alike.

2. Literature Review

2.1 Theoretical Perspectives of HRM

Nowadays HRM is considered to be a central management activity which has to be totally integrated with business planning, while its main focus being employees considered as individuals (Storey, 1993; Guest, 1987; Hendry and Pettigrew, 1986). One of the strategic goals of HRM is to create an organizational environment whereby the workforce feels bound and committed to business vision, mission and goals through comprehensive employee consultation, communication and involvement. Particularly, according to Guest (1987), since strategic integration, employee commitment, flexibility and quality represent fundamental HRM principles, employers applying HRM practices are better set to respond to the fast-changing economic and social environment. To illustrate, selection, training, empowerment teamwork and reward systems are certainly instrumental for the attainment of service quality and customer care.

Historically, the concept of HRM managed to establish itself in the early nineties by drawing mainly from two converging schools of thought. On the one hand, Barney's (1991) resource-based theory of the firm and, on the other, the core competencies approach conceived by Prahalad and Hamel (1990). In fact, they both maintain that sustained competitive advantage is intimately connected to the internal resources of a firm, provided that four conditions are met. While these resources must be value-adding to the organization, they have also to be unique or uncommon. Then, competitors should be unable to imitate these resources which, last, cannot be substituted. Together with Storey (2007), the authors agree that human resources thoroughly satisfy these requirements and enable the organization to achieve superior performance as compared to the other players in the market. Thus, successful management of human resources results from aligning the workforce with organizational objectives while nurturing their competencies and skills so as to perform both effectively and efficiently. The deriving results are far reaching: while, on the one hand, organizations enhance their sustained competitive advantage, on the other, consistent and coherent HRM contributes to the well-being of both individual employees, their families and society at large (Guest, 2002).

Nonetheless, there remains to identify if there is one best way to strategically manage human resources, which is at the center debate between the 'universalistic' and the 'contingency' schools of thought (Delery and Doty, 1996). On one extreme lie the 'universalistic' perspective which contends that specific sets of HR practices always being better performance if applied correctly "across the population of organization" (p. 805). On the other extreme, the 'contingency' viewpoint maintains that organizational strategy is the main contingency factor that affects HRM

implementation. Thus, the latter perspective features more elaborate arguments than the former by virtue of more complex interactions involved (Venkatraman and Prescott, 1990; Venkatraman, 1989). Somewhere in the middle of the continuum the 'best practice', bundles' or 'best fit' policies and practices are located, which aim at developing employee performance in order to ensure a stable service quality and competitive advantage (Boxall and Purcell, 2000). These in-between perspectives, by drawing from experience and sound evidence, demonstrate that, provided comparable structures and contexts, successful organizations have the tendency to apply the same policies and methods. This statement is corroborated by Huselid's findings (1995) from a sample of more than one thousand firms in various US industrial sectors, whereby it was shown that specific high performance HRM work practices affected positively the firm's financial performance as highlighted by measurable business outcomes.

2.2 Debates of Hard Vs. Soft HRM

Since early academic debates, HRM has frequently been illustrated as a concept with two main and distinct variants: hard and soft (Truss et al., 1997). This dichotomy sees the 'hard' standpoint emphasizing on the strategic, quantitative and calculative management of human resources, thus considering them as an economic factor in production. On the other hand, 'soft' HRM draws from the Human Relations school of thought, thus focusing on motivation, communication and leadership. In particular, according to Stace and Dunphy, 'soft' HRM "is directed primarily to human fulfillment rather than productivity and profit" (1991).

Whatever the HRM models, all approaches are concerned with strategic issues. While 'hard' models usually tend to have a more intense focus on linking the HRM strategy to the overall corporate strategy, still a varying degree of emphasis on strategic fit can be seen in both variants, thus demonstrating that they are not necessarily mutually exclusive (Wright and Snell, 1998).

2.3 Management of the HR flow

Recruitment and Selection are two of the most critical activities of HRM in any organization during any economic circumstances, even in a challenging period such as the current financial crisis (Fernández-Aráoz et al., 2009). In fact, if relevant policies, procedures and practices are well thought-out, they can considerably contribute to successful organizational performance, positive employee relations and a favourable public image thus strengthening competitive advantage (O'reilly III and Pfeffer, 2000).

Recruitment and selection are two sides of a major HRM task often referred to as 'staffing' (Carless et al., 2009), which is related with ensuring the organization attracts and employs the right people for the right job and at the right time. More in detail, recruitment is a process aiming at attracting a pool of candidates for a particular position with appropriate qualifications from which it is attainable and practical to select and hire potential employees (Breaugh and Starke, 2000). The employee in-flow process is to be considered thorough if it is compounded with a sound selection

system made up of appropriate techniques and methods aiming at selecting, appointing and initiating a new and competent workforce member or members. Central aspect of selection is the ability to precisely match a candidate to a specific position (Behling, 1998) while having previously defined the criteria for job success.

Despite the criticality of recruitment and selection, authors are still worryingly reporting about insufficient policies and procedures in place. Particularly, more than a decade ago, Graves and Karren (1996) highlighted the shortcomings of the most widespread selection technique, the employment interview, a as it is vulnerable to recruiters' subjectivity. Currently, there are academics (Carless et al., 2009) that still are highlighting poor and uninformed recruitment and selection practices by HR professionals, even though research has demonstrated the development of highly effective techniques to minimize any inefficiencies. Apparently, the authors say, responsibility falls onto the HR professionals who often lack the academic knowledge to refine their skill, thus determining what is referred to as research–practice gap (p. 105).

Lately, much has been written about talent management, a subset of policies and procedures aimed at ensuring the sustained ability of an organization to attract key employees who are of vital importance for competitive advantage in the knowledge economy. Indeed, McKinsey consulting firm since 1998 (Chambers et al., 1998) referred to the 'war for talent', meaning for human resources featuring key skills and knowledge, holding a noteworthy value on the market and who expect continuous development opportunities. Nonetheless, till now there is still confusion about the definition of talent management, especially as some approaches tend to propose again something that anyway used to be done in the past through sound personnel management, only that now it is adapted to the imperatives of the current business environment (Lewis and Heckman, 2006).

Last, as any human relationship is doomed to end, employment relationship is certainly no exception. Nonetheless, there is a wide variety of circumstances that may lead to such development: retirement, dismissals or, even, involuntarily exits as in the cases of redundancies due to downsizing, business decline or even technological advances and globalization. Whatever the case may be, it rests upon HR professionals to ensure compliance to legal requirements as well as to organizational policies and practices without foregoing that the whole unpleasant process must be handled sensitively, even by recurring to outplacement services (Freeman, 2009).

2.4 Development of Human Resources and Organizational Performance

The application of policies and practices related to the professional growth of the workforce is another set of responsibilities of an HRM function aligned with the learning organization model (Senge, 2006) and focused to promote continuous improvement. These learning and development initiatives include both formal and informal in-house activities as well as external education opportunities (Garavan and Mccarthy, 2008).

Interestingly, the employee development framework illustrates most evidently how much the success of HRM initiatives are dependant upon total organizational efforts in

the same direction (Dixon, 1998). In fact, human resources development (HRD) activities must be supported by the organization in its entirety while demonstrating a passionate and enthusiastic engagement to constant improvement through ongoing learning. Furthermore, the leadership must consistently cascade the vision to all organizational levels and ensure it is understood and shared while assuming a facilitating, coaching and supportive role towards personal and team developments (Wilson, 2008). Leaders should be always keen in developing themselves in the first place, something that compels organization to feature management development programs for middle and senior managers as well as directors. Nonetheless, the key element of support towards HRD is to be found in the CEO, meaning the person who sets the organization to work towards the vision, mission and strategy and, by being unceasingly aware of the external environment, accommodates the organizational structure accordingly (Lafley, 2009). To this end, free and open communication throughout the total organization is instrumental in order to maintain workforce alignment towards the attainment of common goals (Williams, 2002) and nurture a trustworthy work climate (Ruppel and Harrington, 2000).

Once the elements referred above are solidly in place, then an organization can address itself coherently and decisively towards learning and development. In fact, the commitment to ongoing development can be translated into a lifelong learning philosophy whereby employees are motivated to acquire innovative ideas and knowledge, try what they learn as well as to take charge of their personal growth (Barsh et al., 2008). The results in an organizational culture that nurtures human resources development become also evident in the forward-looking way decisions are taken within an environment that welcomes innovation, experimentation and the influence of external stimuli while mistakes are seen as an inevitable outcome on the road to ongoing improvement (Jassawalla and Sashittal, 2002). The benefit of this organizational culture shows especially in the collaborative way the workforce would embrace the adoption of new technology (Martins and Terblanche, 2003; Mirvis et al., 1991), as well as during challenging times when business sustainability is tightly linked to its employees' engagement with the organization (Chander et al., 2000).

As Kotter wrote in his last book (2008), organizational success is attained if an organization is pervaded by a widespread 'sense of urgency'. These times of intense competition require a workforce that faces challenges and relentless change needs without fear and that is willing to question itself in terms of its embedded core knowledge base as well as its familiar and conventional policies and procedures. Necessary pre-requisites for successful change management to occur is again the all-important organization culture which fosters the transition of employees from the old to the new coupled with the existence of enabling policies and procedures and the provision of appropriate tools for making change a reality (Schneider et al., 1996).

This leads to the critical relationship between ongoing organizational development and the intellectual capital as well as knowledge management capabilities which is valid for any organization. First of all, knowledge becomes meaningful if its holders understand not only their responsibility to update and develop it further but also are eager to share it, something that relies on their psychological motivators (Cabrera et al., 2006) According to Smedlund (2008), the knowledge system of a firm is a highly articulated web encompassing explicit, tacit and even potential knowledge which does represent a vital competitive advantage once it is continuously enriched through the

elaboration of new information. This system, then, must facilitate the diffusion of knowledge throughout the organization thus enabling the organization to change incessantly (Nonaka et al., 2006) while being aware of market needs and trends. Consequently, benchmarking procedures assist further in understanding organizational positioning versus the competition as well as facilitates the identification and possible adoption of best practices (Zairi and Al-Mashari, 2005; Jarrar and Zairi, 2001).

Nonetheless, human resources development systems cannot be considered thorough if they are not supported by policies and procedures that ensure sound measurement and assessment of changes in employee attitude, performance and commitment to ongoing development (Lähteenmäki et al., 2001). Furthermore, these policies and procedures are instrumental in enabling comparisons among individual as well as across team performances, thus triggering a healthy sense of competition, a willingness to challenge familiar and conventional ways together with an attitude eager to take advantage of ongoing development at all levels (Rousseau, 1998).

Finally, human resources development can be further encouraged by implementing appropriate pay scheme and incentives to improve performance, corroborate motivation towards ongoing learning and professional growth which ultimately strengthen the employees' satisfaction with their jobs. Appropriately rewarding employees for their developmental efforts which translate into improved performance represent a concrete action by the organization towards recognizing employees for their talent as well as their valuable contribution to overall organizational success. Since 1988, though, Baker et al. (1988) warned that such incentive systems must be closely scrutinized as they are all too often uneconomical to an organization, something that Beer and Cannon (2004) agreed upon by demonstrating that indeed, if not applied appropriately, the costs may be more than the benefits as occurred to Hewlett-Packard in the mid-1990s.

2.5 Human Resource Management Roles

In his remarkable work, Ulrich (1998; 1997) has offered a thorough vision of how the HRM function should develop so as to successfully contribute towards the attainment of an organization's strategic excellence goals. Thus, he points to the role of the HR director as business partner, a concept initially brought forward a decade before by Tyson in 1985, as suggested by Armstrong (2000).

Ulrich identifies four distinct and coexistent roles for the HR professional with associated deliverables. Initially, the HR professional has to be an 'Administrative Expert' meaning that they should be highly competent with the way HRM work is organized and accomplished, while delivering personnel administration services efficiently and effectively (e.g., recruitment, selection, training, compensation, benefits, work force planning, and performance management) by taking advantage of new technologies, improved methods and innovative work solutions such as shared services and outsourcing wherever appropriate. Then, drawing from their institutional role vis-à-vis the organization's people resources, HR managers are requested to be 'Employee Champions' towards senior management by virtue of their in-between position in the organizational ladder. While they would strongly represent workforce concerns to the upper echelons of management, by the same token they should

maximize employee contribution by focusing on commitment and job performance. To this end, career planning, training and development, and mentoring are among the many programs that align employee competencies with business objectives, and therefore add value. Third, these times of intense competition at business, regulatory and global levels compel the HR professional to assume a 'Change Agent' role, whereby they would represent the catalyst towards ongoing organizational transformation, processes improvement and the shaping of a culture embracing change. This would determine ongoing competitive advantage on condition that HR professionals assist the management team develop and cascade throughout the organization explicit visions of the future while developing procedures that encourage and reward behaviors which are coherent with goal achievement, and overcome any resistances to change. In this role, it is evident that HRM adds value by assisting organization move from the old status quo into new and more competitive dimensions. Last, HR managers, by being 'Strategic Partners' should bind with both senior and line management in a trustworthy partnership aimed at the accomplishment of business strategies, thus contributing to the sustained demonstration of organizational mission. This, in turn, translates into value creation as appreciated by the organizational stakeholders, namely investors, line managers and employees (Ulrich and Brockbank, 2005). In order to be effective in this role, HR managers must be conversant with organizational and financial matters combined with business savvy. As a result, when HRM policies and practices become of value by aligning themselves to the goals of line management, only then can HR managers expect a senior executive role at the leadership table along with the other organizational functions.

2.6 Cross- Cultural Dimensions

While the preceding conceptualization by Ulrich tends to have a universal applicability, a landmark research by Hofstede (1980) about twenty years earlier defined that HRM policies and practices may not be valid across different countries because of divergence from local laws, customs and cultures as consistently demonstrated by Zhang (2003). Equally, cultural differences could affect how the HR function is deployed across borders by global organizations. By drawing from an impressive amount of 116,000 questionnaire replies gathered following a survey involving IBM executives in 1980 he was then able to study how cultures differentiate from country to country. Thus, while he initially identified four cross-cultural dimensions, he later (Hofstede, 1993) included a fifth dimension. The five cross-cultural dimensions identified by Hofstede are Uncertainty Avoidance, Masculinity-Femininity, Individualism-Collectivism, Power Distance, and Short-Long Term Orientation.

First, Uncertainty Avoidance relates to the extent people are at ease with ambiguity. Personalities with high levels of uncertainty avoidance confer great value to stability and certainty, and tend to appreciate methodologies that heighten a sense of security in unclear situations (e.g., detailed work plans). Individuals who score low in this dimension are more comfortable with risk and are more prone to move forward without having the full picture of what is lying ahead.

Then, Masculinity-Femininity illustrates the degree that personalities show masculine (e.g., independence, dominance) or feminine values and behaviors (e.g., interdependence, empathy, openness). High masculine cultures feature clearly differentiated gender roles, independent performance, achievement, and ambition, while feminine cultures are characterized by equal gender roles, quality of life, and helping others.

Thirdly, the author considers Individualism-Collectivism as being the dimension that refers to the degree according to which individuals emphasize their individual needs and wants versus the needs of the group. Thus, responsibility and achievement by individuals belong to cultures that tend towards the individualism side of this cultural continuum, while cultures that are closer to collectivism focus on group efforts, teamwork, and group membership.

Following is the fourth dimension that deals with Power Distance, which reflects the degree that people accept differences among individuals as sensible and expected. Therefore, cultures featuring high power distance determine workplace environments whereby subordinates and their superiors constitute two distinct groups separated by a considerable status gap. To the other extreme, people belonging to low power distance cultures experience fewer status barriers in their reporting relationships because superiors are more open.

Last, the fifth dimension is represented by the Long-Short Term Orientation, whereby short-term orientation emphasizes the present or the past, honors customs and tradition, and utilizes resources to meet current needs. On the other hand, environments featuring a long-term orientation are geared towards the future by saving resources to meet future emergencies.

Based on the vast amount of data collected, Hofstede (1993, 1980) has been able to plot many countries along the five cross-cultural dimensions according to the scores recorded thus managing to produce extremely valuable information material featuring national profiles that are of both theoretical and practical benefit to IHRM. Thus, any organization planning its operational expansion on a global scale may attain valuable understanding of successful management approaches and practices in overseas locations by referring to the country profiles defined by Hofstede.

In the final analysis, HR managers are presented with two set of powerful tools aiming at enhancing their value to organizations. On the one hand, Ulrich's roles concept strengthens HR strategic capability while, on the other, Hofstede's cross-cultural dimensions compel the HR professional to reflect on how they can be more effective across different cultural settings in a global reality.

2.7 HRM Implementation

Since the early Eighties, academics have been elaborating (Miles and Snow, 1984) on the relationship between strategy and HRM thus concluding that the former can indeed emphasize to a considerable degree the implementation of specific aspects of HRM and their consequent impact on business performance. In this same line, (Subramony, 2006) highlights that the overall business strategy and the economic approach of

business dictate which human resource management practices are accepted and which are rejected. In fact, Marginson et al. (1994) contend that integrated and consistent HRM implementation may be frustrated by limiting financial controls and expectations. Moreover, in times of recession as the current financial crisis, HRM is subject to even higher pressures imposed by the need to downsize or restructure business following the decision by management to cut on labor costs, which, due to their higher predictability than future revenues, allow for a more convenient way to improve profitability and, ultimately, competitiveness (Cascio et al., 1997). Despite the expected benefit, Cascio (1993) warns management about the need of careful downsizing execution as, in the author's own words,

> "study after study shows that following a downsizing, surviving employees become narrow minded, self-absorbed and risk averse. Morale sinks, productivity drops, and survivors distrust management. In fact, this constellation of symptoms is so common that it has taken on a name of its own: survivors' syndrome" (p. 100).

To this end, Tricker (2009) highlights the explicit responsibility of board of directors in the way they exercise corporate governance and their ultimate power to steer organizational strategies. Their decisions, it is contended, must feature professionalism, integrity and trust.

Indeed, HRM can critically contribute to organizational effectiveness and 'agility' (Shafer et al., 2001) provided that the structure it relates to and the organizational systems support these goals. To this end HRM can capitalize on the benefits of information technology which enables it to develop administrative efficiencies while reducing costs. Furthermore, international organizations recognize e-HRM as instrumental towards the standardization, harmonization and alignment of HR policies and processes throughout (Strohmeier, 2007).

Nonetheless, no structure, system, tools, policies or procedures are enough for sound HRM implementation unless there are solid organizational values that focus on people management and leadership that engage line managers towards the implementation of HR-aligned practices (Purcell and Hutchinson, 2007). While the involvement of line managers in the delivery of HR initiatives may appear to somewhat downplay the role of HR specialists (Watson and Maxwell, 2007), still Larsen and Brewster (2003) consider line manager's co-operation an essential factor towards the establishment of the HR business partnership, a viewpoint further strengthened by Lucas et al. (2004) who stress that line management play a key role in the introduction of HRM innovations. More to this, Drucker himself characteristically warned executives that focus on human resources shall be the top priority as quoted by Flaherty (1999):

> "The future will not just happen if one wishes hard enough. It requires decision—now. It imposes risk—now. It requires action—now. It demands allocation of resources, and above all, of human resources—now. It requires work—now" (p. 122)

2.8 HRM in the Hospitality Industry

2.8.1 Overview

The preceding literature review assists in highlighting key activity areas of concern to HRM. The focus now shifts towards the hospitality industry in an attempt to illustrate the characteristics of this particular service sector and the way HRM has adjusted itself in order to represent a most valuable business partner at all levels.

As demonstrated above, it is the ultimate business strategy that affects the mission of HRM. Indeed, Ford and Heaton (2001) contend that, due to the considerable intangible aspect of the service experience (Lashley, 1998), hospitality industry has managed to develop unique competencies in service provision that businesses in any other industrial sectors may consider worth implementing. More specifically, the hotel industry is particularly capital- and labor-intensive, as opposed to other sectors of the services industry, such as marketing and consulting. Further, the logistics and supply chain processes can be as sophisticated as in manufacturing businesses (Dimou et al., 2003).

Exactly because the hospitality industry is focused in providing service excellence in an extremely competitive environment (Dubéa and Renaghanb, 1999), organizations within this industrial sector aim to build a solid culture of service so as to guide every decision by having the customer in mind and ensure that any customer experience is managed successfully (Teare, 1995). In turn, this reflects on managerial skills of hospitality professionals which differ substantially from those required in the manufacturing industry. Particularly, hospitality service delivery systems demand managers to develop unique skills in areas such as organizing, staffing and commanding (Bowen and Ford, 2004).

Consequently, being customer satisfaction and service delivery consistency the main strategic challenges faced by any hospitality concern, staffing is a top HRM priority with the goal to hire employees featuring a strong service attitude (Ladhari, 2009). Taking the local UK hotel market as an example, however, a survey by Price (1994) highlighted that, despite these recruitment efforts, still personnel practices were poor in the mid-nineties as they were in the mid-eighties as highlighted by Kelliher and Johnson (1987), a worrying reality further confirmed by McGunnigle and Jameson (2000). Nonetheless, in the mid-nineties that same local market indicated that large and foreign-owned hotels were featuring a more advanced approach to personnel management, even if not thorough in proper HRM terms (Kelliher and Johnson, 1997). Indeed, Lucas (2002) demonstrated that HRM in the hospitality industry as compared to all industries and services in Great Britain, featured a considerable divergence, especially in terms of 'retaining control/cost control' approach to management leading to a very 'hard' HRM in practice which was a cause to employee dissatisfaction. More recently, however, Wilton (2006) showed that the application of formal or strategic approaches to employee relations was directly proportional to the hotel size.

On the whole, despite any shortcomings, the hospitality industry recognizes that the commitment of their workforce to service quality delivery is key to competitive advantage and sustained organizational success (Yang and Cherry, 2008; Worsfold, 1999). As a result, HR managers together with line managers must recognize the importance of maintaining a favourable psychological climate (Amenumey and Lockwood, 2008), to avoid the creation of toxic workplace environments leading to decrease employee performance and, even, burnout (Ledgerwood et al., 1998). Thus, the contribution of sound HRM to organizations is greatest whenever its policies and practices are coherently applied (Kelliher and Riley, 2002), especially by providing the opportunity for employees to unleash all their potential through empowerment (Lashley, 1995). Indeed, time and again research has underlined the closely-knit relationship between sound people management and quality service provision (Maroudas et al., 2008 ; Pallet et al., 2003 ; Haynes and Fryer, 2000).

2.8.2 International Luxury Hotel Chains and HRM

Same as occurs in any other industrial sector, hospitality firms do have the strategic option to expand their business operations internationally, or even globally. Contractor and Kundu (1998) found that the 'entry mode' into new markets is determined by country-, environment- and firm-specific variables. According to the researchers, 65.5% of hotel international operations are contracted through non-equity arrangements, such as franchise agreements, management contracts and strategic alliances. Consequently, these arrangements enable hotel organizations with successful and unique service brands to expand globally. Further, a most recent study by Graf (2009) clearly indicates that stock markets react abnormally whenever the entry mode of the hotel chain fits with the specificities of the host country. In particular, the author found that investors react very favorably to new management contracts in developing countries and new franchise agreements in developed countries, also because of the anticipated benefits resulting from the hotel firms' strategic orientation, contractual control and formality of business operations (Yan et al., 2007).

For luxury hotel chains, emphasis on quality throughout is of utmost importance and, therefore, HRM plays a strategic role in strengthening this vision within the organization (Maxwell et al., 2004). In 1999 Hilton, for instance, introduced a worldwide customer service quality initiative named 'Equilibrium' which prompted HRM to develop a strategy branded 'Esprit' that encompassed, beyond a successful staffing function, policies and practices targeting specifically the areas of assessment, compensation, benefits, recognition, career tracking and discipline (Maxwell and Lyle, 2002). Other hotel chains, have put HRM at center stage with regards to their business strategy along with marketing and operations as in the case of Accor (Aung, 2000), or, even, have achieved remarkable talent management through the dynamic involvement and support of the hotel General Manager to a critical HR area of responsibility (Yeung, 2006).

2.8.3 The nature of HRM in Global Luxury Hotel Chains

The section above illustrated the easiness with which nowadays a successful business concept can expand globally, even to risky business contexts, by means of favourable and relatively safe contractual agreements. Given the speed with which a hotel

company can become truly global, it might be inferred that HRM would become global as a consequence. Indeed, Brewster et al (2005) contend that International HRM applied by MNCs could be enabled to assume a global strategic value. In particular, the authors identify the factors determining the shift of HRM from international to global as being: talent management and employer branding, global leadership through international assignments, international workforce management and assessment of HR contribution. Further, Hughes (2002) maintains that the next stage of strategic HRM is universal HRM by capitalizing on 'best practices' perceived to be valid universally, such as establishing a service-oriented culture, nurturing a solid human capital base, motivating employees, and offering employees the opportunity to make a difference to service. Another framework used to identify the HRM model of a company operating internationally is offered by transnational HRM (THRM) (Dickmann et al., 2009). According to this process-based viewpoint, THRM relates to knowledge networking and standardization of HRM policies and procedures in MNCs., thus allowing for deeper understanding of business contexts and the identification of international management practices to be developed.

Nonetheless, despite the global expansion of luxury hotel chains and the subsequent attempt to globalize the HRM function, a closer examination of local contexts demonstrates the existence of unique challenges pertaining to the political, national and cultural settings, which requires the local hotel HR professional to adapt accordingly (Naama et al., 2008; Costa, 2004; Lu and Chiang, 2003).

Furthermore, labor markets critically affect the quality of service professionals: to this end, governments do contribute to the establishment of MNCs through, among others, the education of local nationals in hospitality vocational professions (Baum and Szivas, 2008; Kusluvan and Karamustafa, 2001). Even in developed countries such as Korea, research has shown that Training Managers working in international chain hotels modify training programs in order to make it fit to the local culture (Lim and Wentling, 1998). Equally important, labor shortage is a ubiquitous challenge to be found even in populous China, however with its distinct peculiarities related to lack of qualified labour force and the unwillingness for university graduates to join the industry (Zhang and Wu, 2004). On the other hand, surveys by Magnini and Honeycutt (2003) and Shay and Baack (2004) demonstrate the high failure rate of expatriated hotel managers as well as discrepancies between the expatriated managers' self-perception of performance and subordinate-rated managerial effectiveness.

In conclusion, the above mosaic of sample challenges compels HRM to adjust to the 'think global, act local' perspective whereby central offices originate wide-scope HRM policies and procedures and then it falls upon the local hotel HR manager to decide and act the best way they see fit to the local context which, incidentally, they know better than anybody else in the HRM function. An example of such an approach is offered by Zuehl and Sherwyn (2001) who contend that, after analyzing employment termination practices in a sample of countries, MNCs may identify common employment termination practices as general guidelines to be considered.

3. Research Outline

Ten years ago, Olsen and Roper (1998) stated that research in the hotel industry was at an 'embryonic stage'. Nowadays, however, substantial knowledge has been gathered since then, highlighting a highly diversified HRM reality across this specific industrial sector. A key issue brought forward by researchers is the gap between HRM theory and the reality of HRM applied in the field. For instance, Groeschl and Doherty (2002) highlighted inconsistencies in the application of the all-important appraisal process, undermining the benefits it is supposed to offer. More recently, (Davidson et al., 2006) an investigation of 4-5 star luxury domestic and multinational corporation (MNCs) hotels operating in the otherwise developed Australian hotel market indicated that, despite the availability of high quality labor supply, the hotels' HRM policies and practices were weakly applied. This leads to the survey results by Kelliher and Riley (2002) who indicated that the impact of HRM is highest when organizations apply policies and procedure coherently.

The case study organization is a global luxury hotel group: for the purpose of confidentiality, the name of the company is disguised. It will hereafter be referred to as the 'participant company' or 'the hotel company'. It belongs to a diversified international hospitality group, acting as the owner-manager, franchiser and management company of international deluxe hotels, resorts, and mixed-use real estate developments.

Therefore, the aim of this research paper is the assessment of HRM policies and practices in different national contexts within the international hospitality industry by surveying HR managers of a global luxury hotel chain. In particular, the ambition is to gain a thorough insight on HRM by gathering responses from HR managers operating both at the hotel-unit level and in the administration offices. Research goals are multiple, as the wealth of information will enable to draw useful insight in terms of the 'strength' of the HRM system (Bowen and Ostroff, 2004) and organizational commitment (Wright and Kehoe, 2008), for instance. Ultimately, however, the researcher will attempt to identify the model of HRM applied among the already existing such as International HRM, Global HRM and Transnational HRM, or, especially in the light of the current turbulent times, if it implements an HRM model of its own kind.

In the final analysis, using a case study organization that operates globally within the fast growing service sector it is hoped it will enable the exploration of the relationship between corporate guidelines and application of HRM policies and practices to different national contexts, thus contributing to the debate on the nature of HRM in the hotel industry.

4. Methodology and Research Design

A quantitative survey will be used to examine the approaches and implementation of HRM policies and practices in a global luxury hotel chain. As Saunders et. al., (2003), have pointed out the quantitative survey "is a popular and common strategy in business

and management research", and surveys enable "easy comparisons and appear to be authoritative". (p. 92)

In this case, the aim is to make direct comparisons between human resource management policies and practices implementation across different national contexts of a same global luxury hotel chain. The survey is a useful tool for analyzing and assessing HR practice and the HR function in a systematic way and will allow the researcher to gather similar data and draw useful conclusions. The researcher will seek to look at the extent to which HRM policies and practices are being adopted across the hotel chain at both field and administration levels. In fact, the survey is going to include hotel properties present across the world in 45 countries and the self-administered questionnaires will be sent by electronic mail to all human resource professionals of the hotel company. Given the widespread presence of hotel properties and administration offices around the world, the author is confident that the minimum useful number of 30 respondents for statistical analysis purpose as advised by 'The Economist' (Saunders et al., 2003) will be comfortably attained thus yielding a sample population much larger than this threshold.

Indicatively, the questionnaire will ask respondents to give demographic information on themselves and the workforce they influence and details about their hotel property. Then, following sections will deal with the extent of the adoption of HRM policies and practices by asking, to illustrate, about conditions of employment, job design, recruitment and selection, training, appraisal, quality management, communication and consultation, compensation and pay systems as well as satisfaction with various HR elements within the organization.

Last, information will be sought in order to assess the HRM strategy in terms of the extent to which the formal HRM strategy is being adopted, the degree of strategic integration of HR with business strategy and to reveal the approach of the hotel company in relation to its either ethnocentric or polycentric or geocentric manning strategy.

5. Possible Findings

Apart from revealing information on the profile of HR professionals operating in the hotel group, important findings could occur in relation with the degree of ownership and accountability of HRM policies and procedures in areas such as selection, training and development, employee motivation, recognition, performance management, organizational commitment and organizational culture.

Also, as identified in the literature, useful insight will be obtained with regards to the all-important role of line managers in the sustained implementation of HRM policies and procedures, the extent to which they perceive HR as being important to the organization, as well as the nature of barriers to HRM implementation in general.

Next, HR managers' response will illustrate how they perceive and act upon Company decisions and directives being cascaded by Headquarters; thus the survey may shed light on any possible communication gaps related to local implementation. Being the hotel group a global organization, useful insight will be obtained on how people

management is occurring in different countries and contexts because the survey will allow the interpretation of results according to location and regions, thus enabling comparison against what is known from other countries based on practical and academic literature. Last, given that the hotel group features 3 hotel luxury brands, useful information will emerge as to how people management is occurring in different brands.

6. Research Limitations

Although the research is planned to focus on all the properties of the hotel company around the globe, still the sample will be limited to the the 45 countries in which it operates. Then, responses are going to be obtained only from the HR professionals exclusively, leaving out of the research process both top management and all other levels of staff. Exactly because this research asks company HR professionals to assess the HRM function, there is the real possibility that results might be biased.

Indeed, due to the involvement of HR management exclusively, the researcher will be able to reach only the intended human resource (HR) policies while foregoing the implemented HR practices (Khilji and Wang, 2006).

Despite this single industry analysis has its merits, nonetheless it cannot be easily generalized as opposed to multi-industry studies that provide the opportunity to facilitate the understanding of issues more broadly. Thus, being this study hotel specific, caution is advised in drawing general conclusions.

References

1. Amenumey, E. K. & Lockwood, A. (2008) Psychological climate and psychological empowerment: An exploration in a luxury UK hotel group. *Tourism & Hospitality Research,* Vol. 8, No. 4, pp. 265-281.
2. Armstrong, M. (2000) The name has changed but has the game remained the same? *Employee Relations,* Vol.22, No. 6, pp. 576 - 593.
3. Aung, M. (2000) The Accor multinational hotel chain in an emerging market: through the lens of the core competency concept. *The Service Industries Journal,* Vol. 20, No. 3, p. 43-60.
4. Baker, G. P., Jensen, M. C. & Murphy, K. J. (1988) Compensation and incentives: practice vs. theory. *Journal of Finance,* Vol. 43, No. 3, pp. 593-616.
5. Barney, J. (1991) Firm resources and sustained competitive advantage. *Journal of Management,* Vol. 17, No. 1, p. 99-120.
6. Barsh, J., Capozzi, M. M. & Davidson, J. (2008) Leadership and innovation. *McKinsey Quarterly,* No. 1, pp. 36-47.
7. Baum, T., Szivas E. (2008) HRD in tourism: a role for government? *Tourism Management,* Vol. 29, No. 4, pp. 783-794.
8. Beer, M. & Cannon, M. D. (2004) Promise and peril in implementing pay-for-performance. *Human Resource Management,* Vol. 43, No. 1, pp. 3-20.

9. Behling, O. (1998) Employee selection: will intelligence and conscientiousness do the job? *Academy of Management Executive,* Vol. 12, No. 1, pp. 77-86.
10. Bowen, D. E. & Ostroff, C. (2004) Understanding HRM–firm performance linkages: the role of the "strength" of the HRM system. *Academy of Management Review,* Vol. 29, No. 2, pp. 203-221.
11. Bowen, J. & Ford, R. C. (2004) What experts say about managing hospitality service delivery systems. *International Journal of Contemporary Hospitality Management,* Vol. 16, No. 7, pp. 394-401.
12. Boxall, P. & Purcell, J. (2000) Strategic human resource management: where have we come from and where should we be going? *International Journal of Management Reviews,* Vol. 2, No. 2, p. 183.
13. Breaugh, J. A. & Starke, M. (2000) Research on employee recruitment: so many studies, so many remaining questions. *Journal of Management,* Vol. 26, No. 3, pp. 405-434.
14. Brewster, C., Sparrow, P. & Harris, H. (2005) Towards a new model of globalizing HRM. *International Journal of Human Resource Management,* Vol. 16, No. 6, pp. 949-970.
15. Cabrera, Á., Collins, W. C. & Salgado, J. F. (2006) Determinants of individual engagement in knowledge sharing. *International Journal of Human Resource Management,* Vol. 17, No. 2, pp. 245-264.
16. Carless, S. A., Rasiah, J. & Irmer, B. E. (2009) Discrepancy between human resource research and practice: Comparison of industrial/organisational psychologists and human resource practitioners' beliefs. *Australian Psychologist,* Vol. 44, No. 2, pp. 105-111.
17. Cascio, W. F. (1993) Downsizing: What do we know? What have we learned? *Academy of Management Executive,* Vol. 7, No. 1, pp. 95-104.
18. Cascio, W. F., Young, C. E. & Morris, J. R. (1997) Financial consequences of employment-change decisions in major US corporations. *Academy of Management Journal,* Vol. 40, No. 5, pp. 1175-1189.
19. Chambers, E. G., Foulton, M., Handfield-Jones, H., Hankin, S. M. & Michaels Ill, E. G. (1998) The war for talent. *McKinsey Quarterly,* No. 3, pp. 44-57.
20. Chander, G. N., Keller, C. & Lyon, D. W. (2000) Unraveling the determinants and consequences of an innovation-supportive organizational culture. *Entrepreneurship: Theory & Practice,* Vol. 25, No.1, pp. 59-77.
21. Contractor, F. J. & Kundu, S. K. (1998) Modal choice in a world of alliances: analyzing organizational forms in the international hotel sector. *Journal of International Business Studies,* Vol. 29, No. 2, pp. 325-357.
22. Costa, J. (2004) The Portuguese tourist sector: key challenges for human resources management. *International Journal of Contemporary Hospitality Management,* Vol. 16, No. 7, pp. 402-407.
23. Davidson, M., Guilding, C. & Timo, N. (2006) Employment, flexibility and labour market practices of domestic and MNC chain luxury hotels in Australia: Where has accountability gone? *International Journal of Hospitality Management,* Vol. 25, No. 2, pp. 193-210.
24. Delery, J. E. & Doty, D. H. (1996) Modes of theorizing in strategic human resource management: tests of universalistic, contingency, and

configurational performance predictions. *The Academy of Management Journal,* Vol. 39, No. 4, pp. 802-835.
25. Dickmann, M., Muller-Camen, M. & Kelliher, C. (2009) Exploring standardisation and knowledge networking processes in transnational human resource management. *Personnel Review,* Vol. 38, No. 1, pp. 5-25.
26. Dimou, I., Archer, S. & Chen, J. (2003) Corporate development of international hotel firms: a transaction cost economics and agency theory approach. *International Journal of Tourism Research,* Vol. 5, No. 2, pp. 147-149.
27. Dixon, N. M. (1998) The responsibilities of members in an organization that is learning. *The Learning Organization,* Vol. 5, No. 4, pp. 161-167.
28. Dubéa, L. & Renaghanb, L. M. (1999) A look at the industry's overall best-practice champions. *The Cornell Hotel and Restaurant Administration Quarterly,* Vol. 40, No. 6, pp. 16-26.
29. Dwyer, J. (2008) Are you being served? [manufacturer-turned-service provider]. *Engineering & Technology,* Vol. 3, No. 9, pp.69-71.
30. Fernández-Aráoz, C., Groysberg, B. & Nohria, N. (2009) The definitive guide to recruiting in good times and bad. *Harvard Business Review,* Vol. 87, No.5, pp. 74-84.
31. Flaherty, J. E. (1999) *Peter Drucker: shaping the managerial mind.* San Francisco, John Wiley & Sons.
32. Ford, R. C. & Heaton, C. P. (2001) Lessons from hospitality that can serve anyone. *Organizational Dynamics,* Vol. 30, No. 1, pp. 30-47.
33. Freeman, K. W. (2009) The right way to close an operation. *Harvard Business Review,* Vol. 87, No. 5, pp. 45-50.
34. Garavan, T. N. & Mccarthy, A. (2008) Collective learning processes and human resource development. *Advances in Developing Human Resources,* Vol. 10, No. 4, pp. 451-471.
35. Gebauer, H. (2007) An investigation of antecedents for the development of customer support services in manufacturing companies. *Journal of Business-to-Business Marketing,* Vol. 14, No. 3, pp. 59-96.
36. Goodman, B. & Steadman, R. (2002) Services: business demand rivals consumer demand in driving job growth. *Monthly Labor Review,* Vol, 125, No. 4, p. 3.
37. Graf, N. S. (2009) Stock market reactions to entry mode choices of multinational hotel firms. *International Journal of Hospitality Management,* Vol. 28, No. 2, pp. 236-244.
38. Graves, L. M. & Karren, R. J. (1996) The employee selection interview: a fresh look at an old problem. *Human Resource Management,* Vol. 35, No. 2, pp. 163-180.
39. Groeschl, S. & Doherty, L. (2002) The appraisal process: beneath the surface. *Journal of Human Resources in Hospitality & Tourism,* Vol. 1, No. 3, pp. 57-75.
40. Guest, D. (2002) Human resource management, corporate performance and employee wellbeing: building the worker into HRM. *Journal of Industrial Relations,* Vol. 44, No. 3, pp. 335-358.
41. Guest, D. (1987) Human resource management and industrial relations. *Journal of Management Studies,* Vol. 24, No. 5, pp. 503-521.

42. Haynes, P., Fryer, G. (2000) Human resources, service quality and performance: a case study. *International Journal of Contemporary Hospitality Management,* Vol. 12, No. 4, pp. 240-248.
43. Hendry, C. & Pettigrew, A. (1986) The practice of strategic human resource management. *Personnel Review,* Vol. 15, No. 5, pp.3-8.
44. Hofstede, G. (1980) Motivation, leadership, and organization: do American theories apply abroad? *Organizational Dynamics,* Vol. 9, No. 1, pp. 42-63.
45. Hofstede, G. (1993) Cultural constraints in management theories. *Academy of Management Executive,* Vol. 7, No. 1, pp. 81-94.
46. Hoque, K. (1999) Human resource management and performance in the UK hotel industry. *British Journal of Industrial Relations,* Vol. 37, No. 3, pp. 419-443.
47. Hughes, J. M. C. (2002) HRM and universalism: is there one best way? *International Journal of Contemporary Hospitality Management,* Vol. 14, No. 5, pp. 221-228.
48. Huselid, M. A. (1995) The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of Management Journal,* Vol. 38, No. 3, pp. 635-672.
49. Jarrar, Y.F. & Zairi, M. (2001) Future trends in benchmarking for competitive advantage: a global survey. *Total Quality Management & Business Excellence,* Vol. 12, No. 7, pp. 906-912.
50. Jassawalla, A. R. & Sashittal, H. C. (2002) Cultures that support product-innovation processes. *Academy of Management Executive,* Vol. 16, No. 3, pp. 42-54.
51. Kelliher, C. & Johnson, K. (1987) Personnel management in hotels - some empirical observations. *International Journal of Hospitality Management,* Vol. 6, No. 2, pp. 103-108.
52. Kelliher, C. & Johnson, K. (1997) Personnel management in hotels - an update: a move to human resource management? *Progress in Tourism and Hospitality Research,* Vol. 3, No. 4, pp. 321-331.
53. Kelliher, C. & Riley, M. (2002) Making functional flexibility stick: an assessment of the outcomes for stakeholders. *International Journal of Contemporary Hospitality Management* Vol. 14, No. 5, pp. 237-242.
54. Khilji, S. E. & Wang, X. (2006) 'Intended' and 'implemented' HRM: the missing linchpin in strategic human resource management research *International Journal of Human Resource Management,* Vol. 17, No. 7, pp. 1171-1189.
55. Kotter, J. P. (2008) *A sense of urgency.* Boston, Harvard Business Press.
56. Kusluvan, S., Karamustafa, K. (2001) Multinational hotel development in developing countries: an exploratory analysis of critical policy issues. *International Journal of Tourism Research,* Vol. 3, No. 3, pp. 179-197.
57. Ladhari, R. (2009) Service quality, emotional satisfaction, and behavioural intentions: a study in the hotel industry. *Managing Service Quality,* Vol. 19, No. 3, pp. 308 - 331.
58. Lafley, A. G. (2009) What only the CEO can do. *Harvard Business Review,* Vol. 87, No. 5, pp. 54-62.

59. Lähteenmäki, S., Toivonen, J. & Mattila, M. (2001) Critical aspects of organizational learning research and proposals for its measurement. *British Journal of Management,* Vol. 12, No. 2, p. 113.
60. Larsen, H. H. & Brewster, C. (2003) Line management responsibility for HRM: what is happening in Europe? *Employee Relations,* Vol. 25, No. 3, pp. 228-244.
61. Lashley, C. (1995) Towards an understanding of employee empowerment in hospitality services. *International Journal of Contemporary Hospitality Management,* Vol. 7, No. 1, pp. 27 - 32.
62. Lashley, C. (1998) Matching the management of human resources to service operations. *International Journal of Contemporary Hospitality Management,* Vol. 10, No. 24, pp. 24-33.
63. Ledgerwood, C. E., Crotts, J. C. & Everett, A. M. (1998) Antecedents of employee burnout in the hotel industry. *Progress in Tourism and Hospitality Research,* Vol. 4, No. 1, pp. 31-44.
64. Leiper, N. (1999) A conceptual analysis of tourism-supported employment which reduces the incidence of exaggerated, misleading statistics about jobs. *Tourism Management,* Vol. 20, No. 5, pp. 605-613.
65. Lewis, R. E. & Heckman, R. J. (2006) Talent management: a critical review. *Human Resource Management Review,* Vol. 16, No. 2, pp. 139-154.
66. Lim, D. H. & Wentling, R. M. (1998) Modification of transferred training programmes for multinational chain hotels in Korea. *International Journal of Training and Development*, Vol. 2, No. 1, pp. 17-27.
67. Lu, Z., Chiang , D. (2003) Strategic issues faced by Ontario hotels. *International Journal of Contemporary Hospitality Management*, Vol. 15, No. 6, pp. 343-345.
68. Lucas, R. (2002) Fragments of HRM in hospitality? Evidence from the 1998 workplace employee relations survey. *International Journal of Contemporary Hospitality Management,* Vol. 14, No. 5, pp. 207-212.
69. Lucas, R., Marinova, M., Kucerova, J. & Vetrokova, M. (2004) HRM practice in emerging economies: a long way to go in the Slovak hotel industry? *International Journal of Human Resource Management,* Vol. 15, No. 7, pp. 1262-1279.
70. Magnini, V. P. & Honeycutt, E. D. (2003) Learning orientation and the hotel expatriate manager experience. *International Journal of Hospitality Management,* Vol. 22, No. 3, p. 267.
71. Marginson, P., Edwards, P. K., Armstrong, P. & Purcell, J. (1994) Strategy, structure and control In the changing corporation: a survey-based investigation. *Human Resource Management Journal,* Vol. 5, No. 2, pp. 3-27.
72. Maroudas, L., Kyriakidou, O. & Vacharis, A. (2008) Employees' motivation in the luxury hotel industry: the perceived effectiveness of human-resource practices. *Managing Leisure, Vol.* 13, pp. 258-271.
73. Martins, E.C. & Terblanche F. (2003) Building organizational culture that stimulates creativity and innovation. *European Journal of Innovation Management,* Vol. 6, No. 1, pp. 64-74.

74. Maxwell, G. & Lyle, G. (2002) Strategic HRM and business performance in the Hilton Group. *International Journal of Contemporary Hospitality Management,* Vol. 14, No. 5, pp. 251-252.
75. Maxwell, G., Watson, S. & Quail, S. (2004) Quality service in the international hotel sector: A catalyst for strategic human resource development? *Journal of European Industrial Training,* Vol. 28, No. 2/3/4, pp. 159-182.
76. McGunnigle, P. J. & Jameson, S. M. (2000) HRM in UK hotels: a focus on commitment. *Employee Relations,* Vol. 22, No. 4, p. 403-422.
77. Miles, R. E. & Snow, C. C. (1984) Designing strategic human resources systems. *Organizational Dynamics,* Vol. 13, No. 1, pp. 36-52.
78. Mirvis, P.H., Sales, A.L. & Hackett, E.J. (1991) The implementation and adoption of new technology in organizations: the impact on work, people and culture. *Human resource Management,* Vol. 30, No. 1, pp. 113-139.
79. Naama, A., Haven-Tang, C. & Jones, E. (2008) Human resource development issues for the hotel sector in Lybia: a government perspective. *International Journal of Tourism Research,* Vol. 10, No. 5, pp. 481-492.
80. Nonaka, I., Von Krogh, G. & Voelpel, S. (2006) Organizational knowledge creation theory: evolutionary paths and future advances. *Organization Studies,* Vol. 27, No. 8, pp. 1179-1208.
81. O'Reilly III, C.A. & Pfeffer, J. (2000) Unlocking the hidden value in organization. *Employment Relations Today,* Vol. 27, No. 2, pp. 63-80.
82. Olsen, M. D. & Roper, A. (1998) Research in strategic management in the hospitality industry. *International Journal of Hospitality Management,* Vol. 17, No. 2, pp. 111-124.
83. Pallet, W.J., Taylor, W.W. & Jayawardena C. (2003) People and quality: the case of Delta hotels. *International Journal of Contemporary Hospitality Management,* Vol. 15, No. 6, pp. 349-351.
84. Prahalad, C. K. & Hamel, G. (1990) The core competence of the corporation. *Harvard business review,* Vol. 68, No. 3, pp. 79-93.
85. Price, L. (1994) Poor personnel practice In the hotel and catering industry: does it matter? *Human Resource Management Journal,* Vol. 4, No. 4, pp. 44-62.
86. Purcell, J. & Hutchinson, S. (2007) Front-line managers as agents in the HRM-performance causal chain: theory, analysis and evidence. *Human Resource Management Journal,* Vol. 17, No. 1, pp. 3-20.
87. Rousseau, D. M. (1998) Why workers still identify with organizations. *Journal of Organizational Behavior,* Vol. 19, No. 3, pp. 217.
88. Ruppel, C. P. & Harrington, S. J. (2000) The relationship of communication, ethical work climate, and trust to commitment and innovation. *Journal of Business Ethics,* Vol. 25, No. 4, pp. 313-328.
89. Saunders, M., Lewis, P. & Thornhill, A. (2003) *Research methods for business students.* Harlow, Pearson Education.
90. Schneider, B., Brief, A. P. & Guzzo, R. A. (1996) Creating a climate and culture for sustainable organizational change. *Organizational Dynamics,* Vol. 24, No. 4, pp. 7-19.
91. Senge, P. M. (2006) The fifth discipline: the art and practice of the learning organization. New York, Random House.

92. Shafer, R. A., Dyer, L., Kilty, J., Amos, J. & Ericksen, J. (2001) Crafting a human resource strategy to foster organizational agility: a case study. *Human Resource Management,* Vol. 40, No. 3, pp. 197-211.
93. Shay, J. P. & Baack, S. A. (2004) Expatriate assignment, adjustment and effectiveness: an empirical examination of the big picture. *Journal of International Business Studies,* Vol. 35, pp. 216-232.
94. Smedlund, A. (2008) The knowledge system of a firm: social capital for explicit, tacit and potential knowledge. *Journal of Knowledge Management,* Vol. 12, No. 1, pp. 63-77.
95. Stace, D.A. & Dunphy D. C. (1991) Beyond traditional paternalistic and developmental approaches to organizational change and human resource strategies. *International Journal of Human Resource Management,* Vol. 2, No. 3, pp. 263-283.
96. Storey, J. (Ed.) (2007) *Human resource management: a critical text.* Cengage Learning EMEA.
97. Storey, J. (1993) The take-up of human resource management by mainstream companies: key lessons from research. *International Journal of Human Resource Management,* Vol. 4, No. 3, pp. 529-553.
98. Strohmeier, S. (2007) Research in e-HRM: review and implications. *Human Resource Management Review,* Vol. 17, No. 1, pp. 19-37.
99. Subramony, M. (2006) Why organizations adopt some human resource management practices and reject others: An exploration of rationales. *Human Resource Management,* Vol. 45, No. 2, pp. 195-210.
100. Teare, R. (1995) The international hospitality business: a thematic perspective. *International Journal of Contemporary Hospitality Management,* Vol. 7, No. 7, pp. 55 - 73.
101. Tricker, B. (2009) *Directors: an A-Z guide.* London, Profile Books.
102. Truss, C., Gratton, L., Hope-Hailey, V., Mcgovern, P. & Stiles, P. (1997) Soft and hard models of human resource management: a reappraisal. *Journal of Management Studies,* Vol. 34, No. 1, pp. 53-73.
103. Ulrich, D. (1998) Delivering results: a new mandate for human resource professionals. Boston, Harvard Business Press.
104. Ulrich, D. (1997) Human resource champions: the next agenda for adding value and delivering results. Boston, Harvard Business Press.
105. Ulrich, D. & Brockbank, W. (2005) *The HR value proposition.* Boston, Harvard Business Press.
106. Venkatraman, N. (1989) The concept of fit in strategy research: toward verbal and statistical correspondence. *The Academy of Management Review,* Vol. 14, No. 3, pp.423-444.
107. Venkatraman, N. & Prescott, J.E. (1990) Environment, strategy, co alignment: an empirical test of its performance implications. *Strategic Management Journal,* Vol. 11, No. 1, pp. 1-23.
108. Watson, S. & Maxwell, G. A. (2007) HRD from a functionalist perspective: the views of line managers. *Advances in Developing Human Resources,* Vol. 9, No. 1, pp. 31-41.
109. Williams, S. L. (2002) Strategic planning and organizational values: links to alignment. *Human Resource Development International,* Vol. 5, No. 2, pp. 217-233.

110. Wilson, C. (2008) Performance coaching and training in the workplace. *Manager: British Journal of Administrative Management,* No. 61, pp. 26-27.
111. Wilton, N. (2006) Strategic choice and organisational context in HRM in the UK hotel sector. *Service Industries Journal,* Vol. 26, No. 8, pp. 903-919.
112. Worsfold, P. (1999) HRM, performance, commitment and service quality in the hotel industry. *International Journal of Contemporary Hospitality Management,* Vol. 11, No. 7, pp. 340-348.
113. Wright, P. M. & Kehoe, R. R. (2008) Human resource practices and organizational commitment: a deeper examination. *Asia Pacific Journal of Human Resources,* Vol. 46, No. 1, pp. 6-20.
114. Wright, P. M. & Snell, S. A. (1998) Toward a unifying framework for exploring fit and flexibility in strategic human resource management. *The Academy of Management Review,* Vol. 23, No. 4, pp. 756-772.
115. WTTC (2009) http://www.wttc.org/bin/pdf/original_pdf_file/exec_summary_2009.pdf.
116. Yan, Y., Child, J. & Chong, C. Y. (2007) Vertical integration of corporate management in international firms: implementation of HRM and the asset specificities of firms in China. *International Journal of Human Resource Management,* Vol. 18, No. 5, pp. 788-807.
117. Yang, H. & Cherry, N. (2008) Human resource management challenges in the hotel industry in Taiwan. *Asia Pacific Journal of Tourism Research,* Vol. 13, No. 4, pp. 399-410.
118. Yeung, A. (2006) Setting people up for success: how the Portman Ritz-Carlton hotel gets the best from its people. *Human Resource Management,* Vol. 45, No. 2, pp. 267-275.
119. Zairi, M. & Al-Mashari, M. (2005) The role of benchmarking in best practice management and knowledge sharing. *Journal of Computer Information Systems,* Vol. 45, No. 4, pp. 14-31.
120. Zhang, H. Q. & Wu, E. (2004) Human resources issues facing the hotel and travel industry in China. *International Journal of Contemporary Hospitality Management,* Vol. 16, No. 7, pp. 424 - 428.
121. Zhang, M. (2003) Transferring human resource management across national boundaries. *Employee Relations,* Vol. 25, No. 6, pp. 613-626.
122. Zuehl, J. J. & Sherwyn, D. S. (2001) Identifying uniform employment termination practices for multinational employers. *Cornell Hotel and Restaurant Administration Quarterly*, Vol. 42, No. 5, pp. 72-85.

Kinship Network and Technology Transfer in Albania: Theoretical Review and Policy Recommendations

Mamica Skenderi[1] *and Esmeralda Gassie*[2]

[1] *South-East European Research Center (SEERC), 17 Mitropoleos str., Thessaloniki, 54624, Greece, mskenderi@seerc.org*

[2] *Graduate Centre of Business F2-134, Kemmy Business School, Centre for European Studies, University of Limerick, Co. Limerick, Republic of Ireland, Esmeralda.Gassie@ul.ie*

This paper attempts to gain some insights on the important role that kinship network might have on the diffusion of technology through foreign direct investment, an issue that has not attracted the attention of researchers so far. Policy makers are often focused on assessing the flexibility and efficiency of the legal and institutional investment environment. International trade theory conveys precisely the idea of free movement of capital, without linguistic, social or cultural distinctions. Focusing on the case of Albania, this article shows the extent to which the economic activities are embedded in cultural and social contexts, irrespective of a particular geographical area. In considering the social dissemination channels of technology, the kinship network emerges as an important explanatory factor of the absorption capacity, and its modalities, in socially dense countries. The foundations of the technology transfer process are questioned through the analysis of the international joint venture process. This paper shows that the kinship network can play an important role and influence economic processes. More specifically, it emphasises the importance of involving the local socio-cultural networks for business development, notably for international joint ventures. The former practices of fully owned foreign subsidiaries find their limitations in economies where numerous pecuniary and non pecuniary resources are channelled through local social networks. Nonetheless, the kinship network can also have other limiting effects for business development. A fine balance has to be found between the positive and the negative effects of the kinship network, on the grounds of interrelating socio-cultural structures to economic activities.

Keywords. foreign direct investment, kinship network, technology transfer

1. Introduction

Since the fall of communism, accession to the European Union is a strong motive for policy makers in Albania. It has become the foundation of economic development policies. In fact, economic development models converge towards the approach of the international financial institutions such as the World Bank or the International Monetary Fund. But when applied to post communist countries, many voices were raised on the costs of transition reforms based on those principles (Brada et al. 2004, 2006) and their inadaptability to South Eastern European countries (Lavigne 1999, Petrakos 2003, Turkes and Gukguz 2006). In order to prove the validity of the critics made to the application of standard socio-economic development frameworks to Balkan countries, the overall objective of this paper is to evaluate the transition mechanisms and channels of convergence towards the market economy, through an assessment of technology transfer processes in Albania.

The assumptions made in the rest of the paper are explained below. On one hand, the achievement of the market economy is approximated by FDI. The first part of this thesis will proceed to the analysis of the technology transfer theory and empirical studies in Albania. The importance of technology transfer as an important factor for stimulating growth, is widely embraced by economists and policy makers. On the other hand, according to the neoclassical economics, the existence of an influential social structure, approximated by the kinship network, is considered as a non-market phenomenon. The recent phenomenon of resurgence of the kinship network would be explained as accompanying a feeble state intervention, chaotic and uncertain business environments, shaky private property legislation, corruption etc. The latter problems still exist because of the transition period, which would be a 'pretext' for such instability. Consequently, the recurrent institutional factors (infrastructure/ legislation/ fiscal types of obstacles), take over the growing debate on the importance of specific characteristics of the Albanian society and culture.

Although many authors have emphasized the role of the kinship network in business development, its influence on technology transfer has not yet been studied. One of the reasons for the lack of studies in this area is the difficulty in measuring and approximating the phenomenon. On one hand there is economic theory which explains the technology transfer through the development of foreign direct investment in a given country. On the other hand, qualitative sociological and anthropological surveys have shown the importance of social structure.

The main research questions of this thesis are qualitative: (how) is the transition from socially based (kinship network, family business) to market organised activities (foreign investment) occurring in post communist, future EU member Albania? What is the viability of the technology transfer model and can its performance for the Albanian case be assessed?

2. Technology Transfer and FDI: The Economic Theory Perspective

Technological progress plays a crucial role in the economic growth and can also stimulate economic development and industrialization (UNECE 2001). Many countries lack the research and development resources and skills required to develop their own indigenous product ad process technology, which is particularly true for the less developed economies. Therefore, FDI represents one important way to access advanced technology. Albania, being one of the poorest countries of South-East Europe and particularly part of the Western Balkan countries, is far away from the international technology frontier and needs technology to stimulate growth. The country could benefit at a large extent from FDI technology transfer. This section brings some insights on the methods of technological transfer, accompanied with a dose of reality on FDI transfer in Albania.

2.1 Theoretical Framework on Technological Transfer to Local Economies

The technology that is transferred through FDI can take two forms: hard technology and soft technology (Portelli and Narula, 2003). Hard technology consists of physical investment: plants, equipments, and machineries. Hence, hard technology is supposed to include aspects of embodied knowledge in the machinery and equipment. On the other hand, soft technology includes: knowledge, management/organization system, and production processes. Soft technology is supposed to include aspects of disembodied knowledge as a result of the transfer of operation skills.

FDI can affect the enterprises of the host country, by transferring technology, in two ways: directly and indirectly. Foreign enterprises contribute in a direct injection of foreign capital, technology and foreign management skills to their affiliates, which in turn leads to higher productivity. This is referred to as the direct effect of FDI. The positive direct effect of FDI is examined and confirmed empirically in a vast number of studies. However, foreign enterprises do not affect only their affiliates, but also other firms in the same sector or even in other sectors. The indirect effects are referred as spillovers or externalities. According to Blomstrom and Kokko (1998), spillovers may take the form of positive and negative externalities arising from inward foreign investment. Spillovers can occur: internally between firms that may be in direct competition with the foreign firms (intra-industry spillovers); and externally (inter-industry spillovers) to other firms in the host economy, which are vertically integrated with the foreign enterprises, such as suppliers (backward spillovers) and customers (forward spillovers).

There have been identified at least four ways in which technology and know-how might be diffused from foreign enterprise to other firms in the economy: demonstration – imitation effect, competition effect, foreign linkage – cooperation effect, and training effect (Teece, 1997). First, the proximity of local firms to foreign enterprises can sometimes lead to demonstration effect. When foreign firms introduce new products, processes, as well as organizational forms, they provide to other local enterprise a

demonstration of increased efficiency and productivity. Second, the transfer of capital and technology stimulates competition in the local market. Domestic enterprises face a greater competitive pressure, which induces them to introduce new products in order to protect their market share and adopt new management methods so as to increase productivity.

Third, cooperation between foreign enterprises and upstream suppliers and downstream customers increases spillovers. So as to improve the quality standards of their suppliers, backward linkage channel (vertical spillovers) operates through: direct knowledge transfer from foreign enterprises to their local suppliers, as well as higher requirements regarding product quality and on-time delivery, which in turn provide incentives to domestic suppliers to improve production and technology. Fourth, knowledge can be transferred indirectly through the movement of labor. When MNEs subsidiaries hire domestic workers, the human capital may be enhanced further through organization of training facilities and on-the-job learning. Consequently, indirect effects arise when local personnel trained in the foreign subsidiary decide to leave the firm and move to other domestic firms or help establish new business. Moreover, benefits may arise too if superior management skills of foreign MNEs stimulate local suppliers, distributors and competitors to improve their own management skills. Therefore, human capital can spillover from foreign enterprises to other enterprises as skilled labor moves to domestic firms or decide to open their own enterprises.

However, spillovers "depend crucially on the conditions for local firms" (Blomstrom and Kokko, 2002, p. 177). Positive spillovers are realized only if local firms have adequate social capabilities and absorptive capacities to absorb foreign technologies and skills. Thus, for relatively backward countries, it is often difficult to build the necessary social capabilities and absorptive capacities that allow domestic firms to take advantage of the spillovers potentially available in the economy (UNECE, 2001). Literature has found mixed empirical evidence on the existence of spillovers. Though, evidence is strongest in case of vertical linkage, particularly "backward" linkage with local suppliers (foreign firms provide them with technical and financial assistance, training and other support).

Joint venture is viewed as the "commercial agreement" between two or more entities in order to facilitate work and cooperation in order to accomplish a common aim, through the management of common resources (Norwood and Mansfield, 1999). An IJV implies the creation of an entirely new company from different parent companies, and located in the host country. Theory supports that the higher the dissimilarities between local and foreign markets, the higher the preference of foreign investors towards joint ventures (Duarte and Canal, 2004). Thus, the argument on technology transfer through IJV is stronger when the joint venture is between parent company from a developed economy and parent company from a developing or transitional economy. In this case, there is larger technology gap between foreign and local parent, and the joint venture relies mostly on technological knowledge provided by foreign parent (Meyer, 2004). The transfer of know-how from foreign parent to the IJV would improve the performance and competitiveness of the later.

2.2 Technology Transfer in Albania: bridging theory and reality

Technology transfer in Albania from FDI remains limited due to many reasons, such as limited contact of foreign companies with the local economy, low levels of FDI, low cost FDI, limited levels of regional FDI. However, there is some evidence of spillovers in the manufacturing sector in case that there is contact with the local companies, and particularly local suppliers.

Albania currently provides much of the low cost 'labour force' of Italy and Greece, its major trading partners, but the terms of trade are not favourable to Albania. Instead, its relative trade dependency and the lack of sufficient technology transfer keeps the country trapped in a cycle of low intensive production, without further development of the agricultural or industrial sector. Currently more than half of its GDP derives from trade (mainly textile exports) and the services sector (particularly construction). The imbalances in Albania's trade and investment regime are reflected in the continuing high current account deficit which, in 2006, reached 7.6 per cent of GDP. Financing of the deficit continues to depend in large part on remittances from Albanians abroad, estimated at around €500 million or 13 per cent of GDP in 2006 or approximately 55 per cent of the trade deficit. It is in this key aspect that the Albanian economy differs from other peripheral states in Europe.

	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Albania	53	70	90	48	45	41	143	174	207	178	338	277	325
Bosnia and Herzegovina	n.a.	n.a.	n.a.	n.a.	67	177	146	119	265	381	668	521	423
FYROM	19	9	11	16	118	32	176	442	77	95	157	100	351
Serbia and Montenegro	n.a.	n.a.	n.a.	740	113	112	50	165	475	1,260	1,029	2,090	5,128
SEE	72	79	101	804	343	362	515	900	1,024	1,914	2,192	2,988	6,227

Figure1. FDI inflows in South Eastern Europe (in USD million) (Source: Eurostat)

One of the main reasons for lack of technology transfer is the low levels of serious foreign investments, as well as the ability to maintain these initial serious investments in the country. However, this is difficult to achieve unless corruption is addressed in a proper way. Besides other factors, corruption is one of the major impediments for attracting and keeping serious FDI in Albania. Corruption can be fought in many ways, yet another way of looking at it is involving those networks that are accused of corruption, i.e. kinship network to participate.

3. Kinship network[1] investment: thinking the unthinkable?

The current perspective of the European Union on enlargement converges towards the regional integration of the Western Balkan countries, as a preliminary step before accession to the European Union. Therefore the Western Balkan countries are working towards the achievement of regional economic integration. The recent conflicts would not facilitate this process. Still, according to regional integration models, an increase in regional investment would be key to economic growth. In particular, firms suffer from a lack of capital. Regionally based investment would enhance firm development as it was the case for the European Union member countries. But Albania is the least regionally integrated country amongst the Western Balkan countries. Its main trading partners are Italy and Greece. Consequently, compliance with the European Union policies is an imperative. Initiatives such as 'one stop shop' or improvement of the business environment have been implemented in order to attract investors. The success of these policies has been quite limited as foreign direct investment has not sufficiently increased. Albanian firms might possess inherent characteristics which would restrict foreign investment.

3.1 The involvement of the kinship network in business matters

Previously, it was shown that for the technology transfer process, the ideal business type would be a joint venture with a local firm. Based on the assumption that the local firm has the knowledge and the resources on which the foreign partner can count, it would be a more efficient solution. Nevertheless, as it will be shown below, the effect of the kinship network is a two faced coin, and the negative or positive impacts, are not a foregone conclusion.

The Business Environment and Enterprise performance survey conducted by the EBRD and the World Bank in 2002 shows that 8.2 per cent of the 170 surveyed CEOs claim that the largest shareholders are family members and 64.7 per cent are individuals (EBRD and WB 2002). A more recent study on financing problems of Albanian SMEs, emphasized that one third of the interviewed managers were family members and a preference was given to local partners compared to foreigners (Bitzenis and Nito 2005). In addition, of all 226 respondents, 41 percent borrowed from friends and acquaintances. These figures show that the kinship network is involved in funding, management and possibly ownership and employment.

[1]The kinship network is not to be confused with 'clans'. Groups which claim common descent from a common ancestor, even if they cannot demonstrate exactly how this descent came about, are known as clans. They differ from lineages, therefore from the kinship network, as members may not be able to state their exact links to each other' FOX, R. (1983) *Kinship and marriage : an anthropological perspective*, Cambridge University Press. In the case of Albania, Ireland or Scotland, clans have been the basis for the constitution of independent units, politically organized and economically sustained. So the denomination 'clan' suggests a politically inspired concept which is historically outdated, compared to the socio-biological kinship network, and is not the basic unit of the contemporary Albanian society, as it has often been claimed.

Not only this, but 56 percent of the surveyed entrepreneurs stated that they were not planning any merger activity. On the other hand, regional investors, in particular from Italy, express their scepticism in relation to increasing investment in Albania (Esposti et al. 2001). They mainly use the labour force through migration agreements and tend to bring Albanian workers in Italy rather than transferring the production processes in their neighbouring countries. This reluctant behaviour has repercussions on the level of FDI that is not dependent on the infrastructure/ business environment in Albania. Very often it is claimed that FDI in Albania remains limited mainly due to a relatively insecure investment environment, poor infrastructure (Kaltani 2007), heavy administrative procedures (Hashi 2001, Hashi and Mladek 2000, Hashi and Xhillari 1999), corruption in the public administration (Feilcke-Tiemann 2006, Gërxhani and Schram 2004) and the judiciary and relatively high taxes (Bitzenis and Nito 2005, Muent et al. 2001). Nevertheless, there is potential for FDI growth in Albania due to the country's advantages such as proximity to key EU markets and low cost labour force. Away from these frequently cited arguments, the kinship network could also foster economic development and be cited as an important factor. If used in a constructive way, it can generate spillover effects and save on transaction costs at a larger extent than other measures currently undertaken.

Rauch (2001) has drawn parallels between domestic and international networks on technology transfer. The link between the two networks could be diaspora, or ethnic and kin networks. Recently, Bandelj (2008) extends social capital to social structure, power and culture in a study of the determinants of foreign direct investment for receiving/ host post communist countries in transition. She recognises that this economic transaction between the investor and the host partner, can partly be explained as a socially constituted process. Personal networks, shared ethnic origins and cultural 'matching', contrary to cultural similarities, are more important determinants for FDI than formal institutional arrangements between countries. Based on an empirical approach, Bandelj uses dummy variables of the registered emigration from host to investor countries, the presence of the investor country national ethnic minority in the host country and finally language.

However, kindred involvement in business is believed to engender negative consequences for the firm's growth (Bartlett and Bukvic 2002). Cappuyns (2006) elucidated several consequences of the kinship involvement in business development, without quantifying the extent of the impact. The factors that are mentioned, draw the limits of the positive aspects of the kinship network involvement in business. An over reliance on kindred would cause a limitation of the number of joint ventures and partnerships with firms from other countries or regions, hindering the process of technology transfer and innovation. There would only be a low involvement in professional associations and gatherings. In addition, there is a tendency to keep business information for the involved kinsmen and to recourse to the kinship network funding and remittances, which would limit the firm's growth. Other resources, such as the labour force, are also accessed through the kinship network. The economic decisions made do not always depend on efficiency criteria. Generally speaking businesses involving kindred present a risk aversion that would limit their resources and development (Rus 2002). Iosifides (2007) observes that the new Albanian immigrants in Greece have a path dependent activity, which follows the one of their

kindred already settled in the host country. These studies show that the kinship network can become a barrier for the firm's growth or its creation.

3.2 Recommendations for policy makers

In Albania, which has made FDI the priority in its economic policy, there is a need for foreign investors to be based on local networks for understanding the complex social norms and values involved in business relations. Given the recognised importance of the social network, international joint ventures would be more successful if foreign partners were 'playing' with local actors rather than being individually confronting a different business and social environment from their own. In that perspective, the creation of agencies outside the host country would facilitate the contact with foreign investors. Chambers of commerce or investment agencies are very often too dependent on government regulations and funding. They might not be able to react quickly to the appearance of investment opportunities. Other organisations base their activities on the unfounded conclusion that the existence of common economic interests would automatically stimulate partnership and cooperation amongst different cultural groups / nations (Efthymiadis 2006).

Iosifides (Iosifides et al. 2007) found that Albanian immigrants in Greece use their ethnic, kin based-, network to develop their business. The importance of the diaspora in the perspective of international joint ventures is central. The creation of a special centre, in the foreign partner's home country, which would link ethnic communities in the Diaspora with potential investors on a voluntary and informal basis, might foster the emergence of familiarity and closeness, an element so important for trust building. Peer pressure in both partner countries would guarantee against possible issues of free riding, more so than government control, which has proven to be more of an obstacle in the transition to the market economy. At a more general level, social engineering is necessary when envisaging socio-economic projects with a multicultural dimension.

For the next couple of years, the Albanian government intends to focus on SME development, export promotion, FDI attraction and minerals' sector development (Ministry of Economy 2007). Of all these objectives, SME development will take 50 to 60 percent of the foreseen budget and FDI promotion 10 percent. Nevertheless, the methods and aims set, show no preoccupation on how respondent will Albanian economic operators be. Instead, the strategies follow word by word the progress reports of the European Commission together with business environment indicators of the OECD, or the Doing business indicators of the WB. The mismatch between conceptual frameworks and empirical facts might hinder the efficiency of EU and national policies at two levels. First there is incompatibility between the issue targeted by policy makers and the problem in reality. This engenders very often frustration from policy makers as they see that the implementation of their programs takes a long time. But it does as well affect the legitimacy of the programs and the institutions from the point of view of the population targeted (Doja 2001). The inclusion of cultural and social peculiarities in designing theoretical models and economic policies has recently been embraced by economists, but this actually deserves more attention and developments.

4. Conclusion

According to regional integration models, an increase in regional investment would be central to economic growth. In particular, firms suffer from a lack of capital. Regionally based investment would enhance firm development as it was the case for the European Union member countries. But Albania is the least regionally integrated country amongst the Western Balkan countries. Its main trading partners are Italy and Greece. Initiatives such as 'one stop shop' or improvement of the business environment have been implemented in order to attract investors. The success of these policies has been quite limited as foreign direct investment has not sufficiently increased. Albanian firms might possess inherent characteristics which would restrict foreign investment.

Amongst the four channels of technology transfer through international joint ventures, demonstration, competition, cooperation and training effects, not one would be functional if the local social network is not willing to adopt or cooperate with foreign partners, for cultural reasons. Those mechanisms might come up for various reasons but it would seem that one way of avoiding such a phenomenon to occur is to get introduced by a local agent and then try to reproduce the occasions of contacting the local partner.

This paper tried to show that not only there is a structural rationale behind the involvement of the kinship network at all levels of business partnerships but that there are empirical foundations of similarly organised social networks that have similar effects on economic activities. The general statement made on the relationship between kinship structure and economic growth in Albania (Bogdani and Loughlin 2007) is on a necessary disconnection, if not in the short run then in the long run. This statement is based on the experience of socio-demographic transformations in the Western countries. In socially dense countries, the inclusion of such networks acts as a reverse conditionality on the adoption of market rules.

The effects of the kinship network are multifaceted. On the one hand, relying on the kinship network would make the partnership with foreign businesses more limited as the most trustable partner is the one that is geographically and culturally closer. On the other hand, if a foreign investor can get trustworthy in the eyes of local businessmen, there are large networks that become available and it can be assumed that business performance is increased. Certainly the biological and affine relations do matter in setting partnerships, not only when there is an institutional void but whenever policies are designed to be effective by involving local traditional social groups. This paper aimed to demonstrate that it recognising the kinship network's influence in business activities does make a difference. Nevertheless, further empirical and modelling-based research is needed to measure the exact implications.

References

1. BANDELJ, N. (2008) From Communists to foreign capitalists : the social foundations of foreign direct investment in postsocialist Europe, Princeton, N.J. ; Oxford, Princeton University Press.

2. BARTLETT, W. & BUKVIC, V. (2002) What are the Main Barriers to SME Growth and Development in South-East Europe. IN BARTLETT, W., BATEMAN, M. & VEHOVEC, M. (Eds.) Small enterprise development in South-East Europe : policies for sustainable growth. Boston, Mass. ; London, Kluwer Academic Publishers.
3. BITZENIS, A. & NITO, E. (2005) Financing problems and the lack of entrepreneurship in Albania: findings from small and medium enterprises. Global Business and Economics Review, 7**,** 409-38.
4. BLOMSTROM, M. and A. KOKKO (1998) Multinational Corporations and Spillovers. Journal of Economic Surveys, 12, p. 247-277.
5. BLOMSTROM, M. and A. KOKKO (2002) "FDI and Human Capital: A Research Agenda". OECD Technical Papers, No. 195.
6. BOGDANI, M. & LOUGHLIN, J. (2007) Albania and the European Union: the Tumultuous journey towards integration and accession, London ; New York, I. B. Tauris.
7. BRADA, J. C., KUTAN, A. M. & YIGIT, T. M. (2004) The Effects of Transition and Political Instability On Foreign Direct Investment Inflows: Central Europe and the Balkans. Working Papers (William Davidson Institute) -- University of Michigan Business School**,** 1-40.
8. BRADA, J. C., KUTAN, A. M. & YIGIT, T. M. (2006) The effects of transition and political instability on foreign direct investment inflows. Economics of Transition, 14**,** 649-680.
9. CAPPUYNS, K. (2006) Internationalisation of family businesses through strategic alliances: an exploratory study. IN POUTZIOURIS, P., SMYRNIOS, K. & KLEIN, S. (Eds.) Handbook of research on family business. Cheltenham, Edward Elgar.
10. DOJA, A. (2001) Democracy and stability in Southeast Europe: Human, cultural and social factors. Temps Modernes, 56**,** 147-166.
11. DUARTE, C. and E. CANAL (2004) "The Choice Between Joint Ventures and Acquisitions in Foreign Direct Investments: The Role of Partial Acquisitions and Accrued Experience". Thunderbird International Business Review, 46 (1), p. 39-58.
12. EBRD & WB (2002) Business Environment and Enterprise Performance Survey. IN EBRD & WB (Eds.) The BEEPS II Interactive Dataset: Enterprise Survey in Transition.
13. EFTHYMIADIS, N. (2006) The Experience of the Private Sector in Doing Business and Investing in Eastern Europe and the Black Sea. Journal of Southeast European & Black Sea Studies, 6**,** 225-242.
14. ESPOSTI, R., SOTTE, F. & TAFFI, M. (2001) Development and competition between territories, Marche and the Balkans: the case of Albania. Ancona, Italy.
15. FEILCKE-TIEMANN, A. (2006) Albania: Gradual Consolidation limited by
16. Internal Political Struggles. Journal of Southeast European & Black Sea Studies, 6**,** 25-41.
17. FOX, R. (1983) Kinship and marriage : an anthropological perspective, Cambridge University Press.
18. GËRXHANI, K. & SCHRAM, A. (2004) Albanian political-economics: consequences of a clan culture. EconWPA; Development and Comp Systems

19. HASHI, I. (2001) Financial and institutional barriers to SME growth in Albania: results of an enterprise survey. MOCT-MOST: Economic Policy in Transitional Economies, 11**,** 221-38.
20. HASHI, I. & MLADEK, J. (2000) Fiscal and regulatory impediments to the entry of new firms in five transition economies. Journal of East-West Business, 6**,** 59-94.
21. HASHI, I. & XHILLARI, L. (1999) Privatisation and transition in Albania. Post-Communist Economies, 11**,** 99-125.
22. IOSIFIDES, T., LAVRENTIADOU, M., PETRACOU, E. & KONTIS, A. (2007) Forms of Social Capital and the Incorporation of Albanian Immigrants in Greece. Journal of Ethnic & Migration Studies, 33**,** 1343-1361.
23. KALTANI, L. (2007) Complementary Reforms and the Link between Trade Openness and Growth in Albania. Post-Communist Economies, 19**,** 225-253.
24. LAVIGNE, M. (1999) The economics of transition : from socialist economy to market economy, Basingstoke ; New York, Macmillan ; St. Martin's Press.
25. MEYER, K. E. (2004) "Perspectives on Multinational Enterprises in Emerging Economies". Journal of International Business Studies, 35, p. 259-276.
26. MINISTRY OF ECONOMY (2007) Strategjia e zhvillimit te biznesit dhe investimeve (2007-20013). Tirane.
27. MUENT, H., PISSARIDES, F. & SANFEY, P. (2001) Taxes, competition and finance for Albanian enterprises: evidence from a field study. MOCT-MOST: Economic Policy in Transitional Economies, 11**,** 239-51.
28. NORWOOD, S. and N. MANSFIELD (1999) "Joint Venture Issues Concerning European and Asian Construction Markets of the 1990s". International Journal of Project Management, 17 (2), p. 89-93.
29. PETRAKOS, G. (2003) Peripheral European Transitions: Performance, Structure and Trade Relations in the Balkan Region. South-Eastern Europe Journal of Economics, 1**,** 41-64.
30. PORTELLI, B. and R. NARULA (2003) "Foreign Direct Investment Through Acquisitions and Implications for Technological Upgrading. Case Evidence from Tanzania". Globalization as Transformative Process Working Paper Series, No. 05/03, University of Oslo, Norway.
31. RAUCH, J. E. (2001) Business and Social Networks in International Trade. Journal of Economic Literature, 39**,** 1177-1203.
32. RUS, A. (2002) Social Capital and SME Development. IN BARTLETT, W., BATEMAN, M. & VEHOVEC, M. (Eds.) Small enterprise development in South-East Europe : policies for sustainable growth. Boston, Mass. ; London, Kluwer Academic Publishers.
33. UNECE (2001) "Economic Growth and Foreign Direct Investment in the Transition Economies". UNECE Economic Survey of Europe, No.1, Chapter 5.
34. TEECE, D. J. (1997) "Technology Transfer by Multinational Firms: The Resource Cost of Transferring Technological Know-How". Economic Journal, 87, p. 242-261.
35. TURKES, M. & GUKGUZ, G. (2006) The European Union's Strategy towards the Western Balkans: Exclusion or Integration? East European Politics & Societies, 20**,** 659-690.

Why the Innovation Policies are Important in Small and Medium-Sized Enterprises? A Comparison Between European Union and Turkey

Başak Söylemez

Turkish Industrialists' and Businessmen's Association (TUSIAD), Refik Belendir Sok. 111/11, Çankaya, Ankara, 06540, Türkiye, basaksoylemez@gmail.com

In order to achieve sustainable development, it is necessary to produce high value added products while activating economic structure and social dynamics. Recent developments in world economy have shown that small and medium-sized enterprises (SMEs) can create the major dynamics for the economic progress. Recognizing the importance of SMEs, most of the countries, especially developed ones, have started to implement the regulations that will aid the promotion, development and protection of SMEs. Turkey, as candidate country to European Union, can use the advantages of SMEs to gain competitiveness in the European Union market and also in the global markets. Today, the most effective method to reach this aim is to promote innovation, while investing on research and development. In this paper, the importance of innovation in SMEs will be evaluated. Afterwards, the comparison will be made between Turkish and European Union SMEs regarding the innovation policies.

Keywords: small and medium-sized enterprises, research and development, framework programme, innovation, technology, Turkey, European Union.

1. Introduction

Currently, technological growth and emergence of information technology have changed the dimension of the competition from inexpensive labour and having access to the natural resources, to production technologies and original products that can compete in the international markets. Therefore, to have a competitive industry that can compete in the international markets it is important to have customer oriented production, contemporary management understandings, research and development activities, technological developments and innovative studies.

Innovation is today more crucial than ever if a business is to survive and prosper. Markets are changing faster and level of competition is increased by emerging economies such as China and India. Hence, in order to have technology and innovation oriented economy; small and medium-sized enterprises (SMEs) are the key elements.

Therefore, in developed countries the economy is mainly composed of small and medium-sized enterprises. Where the sustainable developments of SMEs are especially important, it is necessary to produce high value added products while activating economic structure and social dynamics. At this point technology and innovation are the key factors that determine the sustainable development and competitive advantage (World Bank, 2007). Technology and innovation are also important for obtaining total factor productivity and for reaching sustainable long-term economic growth targets (World Bank, 2006). In countries like Turkey, where economic structure is mainly composed of SMEs, it is important to have appropriate circumstances for technological development and innovation, encourage cooperation between organisations and direct customers to buy new products from domestic SMEs. This is the reason why it is important to develop technology and innovation oriented firms that focus on new products, processes and services while taking advantage of knowledge and risking their potential for research and development.

Since Turkey established a customs union with European Union (EU) in 1996, she is a part of European single market for industrial goods. Therefore, Turkish SMEs have to compete not only with their domestic rivals but also with the rivals in the EU member states (Yücel, 2006). This necessity forced SMEs to determine their existing technology and innovation problems and to think possible solutions for them. By solving these problems new research and development projects can be planned and Turkey can increase its long-term productivity and competitiveness.

In this paper, Turkish and European small and medium-sized enterprises are examined and mainly innovation polices regarding to them are evaluated. The main objective of this research is to participate in Turkey's European Union harmonization process while emphasizing on the necessity of increasing the competitive powers of small and medium-sized enterprises.

2. Small and medium-sized enterprises in European Union

2. 1 Definition of small and medium-sized enterprises

In order to improve SMEs consistency and effectiveness and to limit distortions of competition, EU has generated common definition for SMEs. Even though the usage of this definition is voluntary, Commission is inviting member states, European Investment Bank (EIB) and European Investment Fund (EIF) to apply this common definition as widely as possible. The aim of this common definition is to lead all Community policies, EIB and EIF activities regarding to SMEs (Sarı, 2005). Also this definition is used in member states as a grand incentive concerning SMEs.

The common SME definition was first recommended in April 1996. That definition has been widely applied throughout the EU. However, taking into account of economic developments since 1996, the Commission adopted a new recommendation on 6 May 2003 (Commission Recommendation, 2003). The common definition that is summarized in Table 1 is a new definition, which was put into effect on 1 January 2005. The new definition takes account of various types of relationships between

enterprises. It also helps to promote innovation and foster partnerships, while ensuring that only those enterprises, which genuinely require support, are targeted by public schemes. Clear instructions are given on how to treat particular relationships between an SME and other enterprises or investors when calculating the enterprise's financial and staff figures. Essentially, the new definition takes into account an SME's ability to call on outside finance. However, the staff headcount thresholds remain at their previous levels, since to raise them would have diluted the measures tailored for SMEs.

Enterprise Category	Headcount	Annual Turnover	Annual Balance Sheet Total
Medium-sized	< 250	≤ € 50 million	≤ € 43 million
Small	< 50	≤ € 10 million	≤ € 10 million
Micro	< 10	≤ € 2 million	≤ € 2 million

Table 1. The new thresholds

By this definition, the scope of the SMEs definition is clearly market out. Thus, the self employed, family firms, partnerships and associations regularly engaged in economic activity may be considered as enterprises. So, it is the economic activity that is the determining factor, not the legal form. With this definition it is ensured that, the support measures are granted only to those enterprises, which genuinely need them. For this reason, it introduces methods to calculate the staff and financial thresholds to gain more realistic picture of the economic situation of an enterprise. To this end, a distinction has been introduced between different types of enterprises: autonomous, partner and linked. Each corresponds to a type of relationship, which an enterprise might have with another.

2. 2 Innovation policies regarding to small and medium-sized enterprises

Innovation holds the key to face most pressing socio-economic and environmental challenges, such as global warming, globalisation and job creation. Therefore, Commission Vice-President Günter Verheugen, who also heads the Enterprise and Industry DG noted that; "Europe needs to become a truly knowledge-based and innovation friendly society, where innovation is not feared but welcomed, is not hindered but encouraged; where it is part of our society's core values and seen to work for the benefit of all citizens" (White, 2006, p.3). As known this innovation friendly society and economy idea is the core element of the Lisbon Strategy.

Research and Development (R&D) Framework Programme (FP) is one of the Union's most powerful instruments to reach the aims of Lisbon Strategy. The new Framework Programme, FP7, which is planned to run from 2007-2013, refocuses efforts to address the Lisbon Strategy. In fact, the instruments to be used and the research themes it will cover are very much in line with FP6. Whilst the R&D Framework Programmes are well known across Europe after more than 20 years of existence, EU support initiatives in other fields of business that are less well known. As mentioned in Lisbon Strategy Europe's development depend on improving competitiveness and innovation therefore,

Commission has put forward proposals for a new framework programme focused on these two areas.

Approved by the Council and European Parliament, the "Competitiveness and Innovation Framework Programme (CIP)" will support innovation and growth over the period 2007-2013. Its overall budget will be a different order, with the proposal calling for € 4.2 billion over the seven years. The € 3.62 billion CIP budget, the largest part, worth € 2.17 billion, is earmarked for the Entrepreneurship an Innovation Programme (EIP) that targets SMEs in particular (ESN, 2006). This programme aims to achieve its objectives through the following actions;

- Enterprise Europe Network: This network combines and builds on the former Innovation Relay Centers (IRC) and Euro Info Centers. It offers concrete and effective solutions to entrepreneurs and companies in more than 40 countries. The Enterprise Europe Network is unique both in terms of its wide geographic reach and of the wide range of integrated services it provides to SMEs and other business actors (European Commission, 2008a). They mainly provide enterprises with a range of quality services to help make them more competitive.
- Access to finance for SMEs: Mainly this financial support is gained from EU financial instruments, which are managed by the European Investment Fund (EIF) in cooperation with financial institutions. These EU instruments target companies in different phases of their lifecycle. EIF uses market-based instruments to provide venture funding and guarantees to promote the creation and growth of European small businesses.
- Eco-innovation: By this action innovative products, processes and services aiming at reducing environmental impact, preventing pollution or achieving a more efficient and responsible use of natural resources will be supported. The pull of eco-innovation can be enhanced by environmental policy, notably through a well-designed regulation and the development of market-oriented instruments. Eco-innovation can also be promoted by fostering co-operation between research and enterprises in promising areas, such as construction, water-management, bio-industries, carbon capture and storage or recycling (European Commission, 2008b).

There are other initiatives, like Gate2Growth. This program was set up with the prime objective to support innovative entrepreneurs in Europe. It aims to help entrepreneurs by fostering networking and exchange of experience and good practice at the European level. Therefore, it provides tools, infrastructure and support services directed to innovative entrepreneurs as well as to their supporters.

Also INNO-Policy TrendChart tracks innovation policy developments in all 27 EU member states, and Iceland, Norway, Switzerland, Croatia, Turkey, Israel, Brazil, Canada, China, Japan, United States of America (USA) and India compromising the Network of National Correspondents (European Commission, 2008c). Its main objective is to improve understanding at European level of how EU member states design and deliver innovation policy in response to specific challenges inherent in their national innovation systems.

There is also Innovating Regions in Europe (IRE) network, which offers regions a joint platform for collaboration and exchange of ideas on the development of regional innovation policies. So, it helps to exchange of experience and good practice among European regions that are enhancing their capacity to support innovation and competitiveness among regional firms through the development and implementation of regional innovation strategies and schemes. The IRE network aims to create an inter-regional learning process for the implementation of regional innovation support actions and it seeks to promote trans-regional innovation projects.

Another initiative is the ten-year lending plan started by the European Investment Bank (EIB), with the European Investment Fund (EIF), as a trial in 2000 in response to the Lisbon agenda, is Innovation 2010 initiative (i2i). Through the EIF support to innovative SMEs and entrepreneurship is ensured, notably through venture capital activities. EIB's overall objective for i2i is to mobilise up to € 50 billion over the current decade.

Also, there is European Regional Development Fund (ERDF) which is the largest Community financial instrument benefiting SMEs. It aims to reduce disparities in the development of regions and to support social and economic cohesion within the European Union. In order to strengthen the creation and competitiveness of SMEs, the ERDF co-finances activities in a broad range of areas, including innovation.

In addition to mentioned initiatives there are number of incentive applications and regulations that support and encourage the innovation. For example, EIB and European Commission have joined forces to set up the Risk Sharing Finance Facility (RSFF). RSFF supports investment in innovation, high-risk research, technological development and demonstration projects through loans and guarantees. By reinforcing the financing capacity of EIB in the area of research, the Community will mobilize efforts of a large number of European banks and financial institutions into research and innovation, thus increasing private investment and funding in this area.

Although there are number of incentives focused on improving innovation, for EU-27 the proportion of enterprises with innovation activity was 41.5 % in industry and 37.0 % in services (Parvan, 2007). For each size-class the proportion of enterprises engaged in innovation activity was generally higher in industry than in services.

Regardless of the sector, for EU-27, enterprises with some form of innovation activity were 33.2 % among small enterprises, 39.6 % among medium-sized enterprises (Parvan, 2007). The larger an enterprise, the more likely it is to capitalize on innovation. In order to boost the economic development of member countries and as a whole of EU, definitely the proportion of enterprises with innovation activity should be increased (European Communities, 2004).

Approximately half of the medium-sized enterprises and a third of small enterprises carry on innovation activities. As shown in Figure 1, in Switzerland, Luxemburg, Iceland, Belgium and Germany local innovation activities are dominant. However, that proportion is less than 20 % in Slovenia, Slovakia, Poland, Denmark and Greek.

In general, financial resources, knowledge, human capital and management competences are the barriers to innovation and investment (European Commission, 2006a). These areas are equally important. Specific types of SMEs have specific types

of needs and what may be available to meet these needs may depend more on the financing system within individual countries than on the characteristics of the firm such as size, sector, age and profitability. Therefore while developing SME oriented policies, divergence precautions should be adopted. At this point the most effective tool that will beacon is eco-innovation.

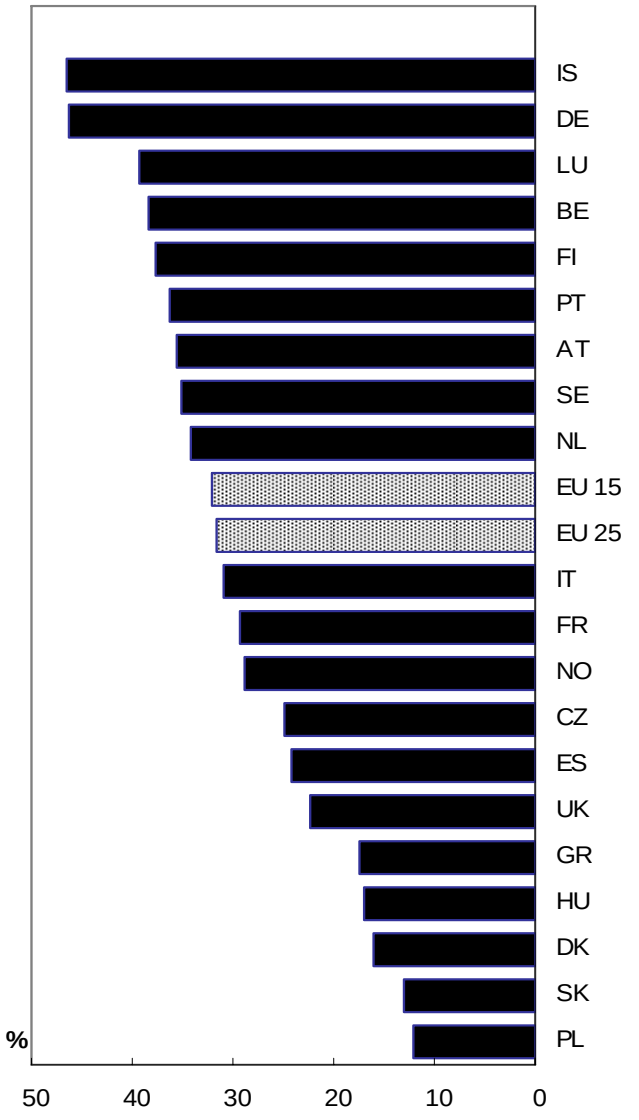


Figure 1. 1998–2000 Innovation activities of SMEs

Source : OECD, 2006.

3. Small and medium-sized enterprises in Turkey

3. 1 Definition of small and medium-sized enterprises

In order to establish a standard in developing policies for SMEs, on 19 September 2005 Council of Ministers has released a regulation, about definition, classification and qualification of SMEs. As summarized in Table 2, this common definition was issued in compliance with the SME definition of the EU.

Enterprise Category	Headcount	Annual Turnover	Annual Balance Sheet Total
Medium-sized	50-249	≤ € 15.15 million	≤ € 15.15 million
Small	10-49	≤ € 3 million	≤ € 3 million
Micro	0-9	≤ € 606,000	≤ € 606,000

Table 2. New SME regulation in Turkey

This common definition does not only define financial thresholds but also states out the autonomy of the SMEs. Financial thresholds and autonomy criteria were determined in accordance with European Union's SME definition. But set significantly lower limits on annual turnover and total assets than EU.

Additionally, the definition takes the increasing number of micro enterprises into account by setting financial thresholds for them. This development aims to encourage the adoption measures addressing the specific problems micro enterprises face, especially during the start-up phase. This definition has again been modified effective 18 May 2006 to eliminate differences in SME definitions used by various institutions in Turkey.

3. 2 Innovation policies regarding to small and medium-sized enterprises

Supreme Council for Science and Technology (SCST), which is the highest science and technology policy making body in Turkey, held its eleventh meeting on March 10th 2005. In this meeting Supreme Council decided to translate Oslo Manual, which was originally prepared by OECD to define innovation, innovation activities, innovating firm and all the related innovation concepts in a systematic manner, to Turkish language and use it as a reference to all the assessments.

Oslo Manual describes, technological product and process innovations as implemented technologically new products and processes and significant technological improvements in product and processes (OECD and Eurostat, 2006). A technological product and process innovation has been implemented if it has been introduced on the market or used within a production process. Technological product and process

innovations involve a series of scientific, technological, organizational, financial and commercial activities. This wide definition of innovation comprises an extended set of innovation. However, until today there was not a common definition for innovation or innovation activities, therefore Oslo Manual is so important for Turkey in the path of innovation.

Parallel to the importance given by The Scientific and Technological Research Council of Turkey (TÜBİTAK) to the innovation, Supreme Council for Science and Technology has argued the topic from a different perspective in the meeting on September 12th 2006. In this meeting, it was concluded that Turkey needs a strategy and an action plan that is oriented to improve innovation activities in order to improve its innovation performance. Therefore it was decided to prepare a "National Innovation Strategy and Action Plan" in coordination with TÜBİTAK. National innovation system will cover the following areas;

- Enterprise, productivity and innovation
- Transfer of knowledge and technology to the firms
- Competitiveness
- Infrastructure and workplace
- International collaboration
- Governance and coordination

Although Turkey does not have a uniform innovation policy, the subject gains importance day by day and started to be discussed in all the platforms. For the period of 2002-2004 the innovation action of the enterprises was 29.7 % for small, 37.2 % for medium-sized and 43.5% for the large enterprises (Turkstat, 2006). The sectoral composition is shown in Figure 2. Even though Turkey does not have a uniform innovation policy, it is interesting that, the enterprises with more than 10 employees in manufacturing and service sector respectively did 34.58 % and 25.90 % of technological innovation. Doing technological innovation is directly proportional to enterprise size therefore large enterprises do more technological innovation than small enterprises.

In this scope, in the Ninth Development Report (2007-2013) the importance of innovation was emphasized and its essential role in efficiency and competitive power was mentioned (State Planning Organisation, 2006). Also it is pointed out that all the necessary legal and institutional regulations should be done in order to boost the National Innovation System.

The SME Strategy issued in 2007 is the only strategy which puts emphasis on the development of entrepreneurship, technology and innovation based start-ups (European Commission, 2008d). One of the action item that is identified in the document is, Development of Technology and Innovation Capacity. This gives responsibility of developing schemes to support the establishment of new technology-based firms to TÜBİTAK.

As well known, innovation is the one of the most important component of the competitive economy, that is why it should be supported and its international transfer should be ensured. Therefore, in the harmonization period, with scope of Sixth Research and Development Framework Programme Turkey has established Innovation

Relay Centers. One of the IRC is located in Aegean and the other one is located in Anatolia.

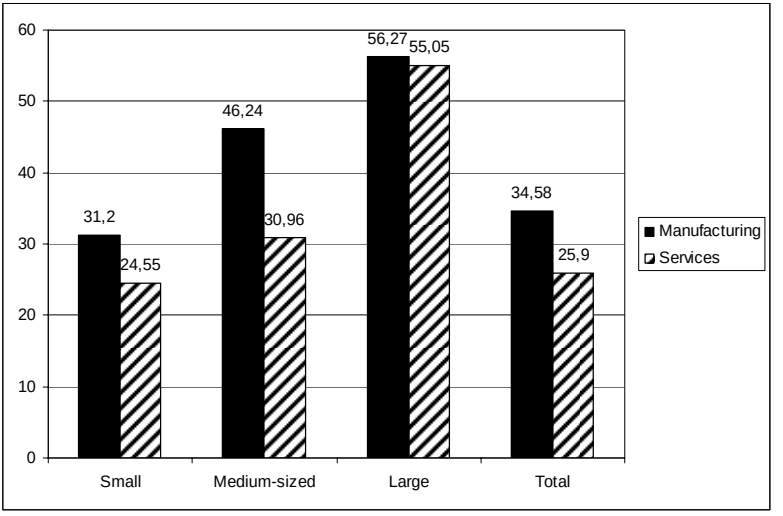


Figure 2 Technological innovation in manufacturing and services sector

Source: Turkstat, 2006.

Other than this, in the scope of FP6, in Mersin there are some efforts to create "Regional Innovation Strategy (RIS)". This project is directed by Mersin Governorship and coordinated by Middle East Technical University (METU) Techno Park with consortium of Mersin University, Mersin Chamber of Commerce and Industry, Mersin Tarsus Organized Industrial Zone and Business Innovation Center of Epirus. It is the first attempt in Turkey for supporting infrastructure and human capital to increase the innovation activities on regional base.

Neither Turkey has national or regional innovation policy nor a special program that supports innovation activities. For example, in SME Funding Program, which is the one of the supporting programs of TÜBİTAK- Technology and Innovation Support Programme (TEYDEB), innovation activities are supported indirectly while supporting R&D activities of SMEs.

4. Comparison of innovation policies regarding to small and medium-sized enterprises between European Union and Turkey

Today, goods and services should be different from their competitors in order to succeed in the market. Innovation is the key to this differentiation. Therefore, innovation is gaining importance for commercialization and transfer of the knowledge.

Innovation activities may differ from each other. Like some of the enterprises enforce well defined innovation activities like product development but others may enforce less

common innovation activities like improving marketing and/or production process. Both of these firms can be considered as innovating firms because innovating firm is the one that has implemented technologically new or significantly technologically improved products or processes during the period under review. It should be mentioned that, technology and knowledge to advance innovation activity needs collaboration between firms. During the process of diffusion, one firm's new or improved product may become another firm's intermediate good while producing new or improved product. Therefore, new technology, process management and product development needs close coordination between firms and foundations that form the national innovation system.

In 1995 Innovation Relay Centers (IRC) were established in EU. These centers used to help to incitement innovation by stimulating technology transfer. On the other hand in Turkey the establishment of these centers was a part of FP6 and was established in year 2002. In Turkish system the effective coordination and flow of knowledge and human resources between universities, firms and research centers is done by IRC, which are insufficient considering the country's economic and demographic structure.

Moreover in EU there are some mechanisms that support cooperation and communication between research centers and business. In Turkey, there is limited communication between research institutions and firms due to the limited availability of intermediaries facilitating exchanges between the industry and research communities. Also there is a lack of incentives to stimulate collaboration between universities and enterprises. So, the cooperation between the business and research communities is weak (European Commission, 2006b).

In Turkey, regulation flows from governmental agencies to innovation organizations and financial flows from the government to private sector are insufficient. The lack of regional and sectoral approaches in innovation policy making and implementation, the insufficient number and diversity of financial institutions for innovation and the small number of financial and innovation intermediaries are the most important shortcomings of the system. Also there is a cultural difference between EU and Turkey regarding to innovation. In EU the importance of innovation has been understood earlier and a necessary infrastructure has been constructed. On the other hand for Turkey, innovation is a new concept. Therefore, financial structure to support solely innovation activities is not developed. That is why innovation oriented funds are added to the existing support programs. But these funds are insufficient to support innovation activities. Hence, current situation is difficult for SMEs to be involved in innovation activities in Turkey.

The increase of competition conditions, economic development and EU negotiation process let the innovation activities to gain importance in Turkey. In this respect, systems, programs and regulations that support and encourage national innovation systems have been proposed however national innovation system has not been yet developed.

The most important aspect that supports the efforts on innovation activities is the enterprises that adopt innovation in all dimensions. As shown in the Figure 3, the percentage of firms that adopt innovation activities in EU 15 is 44 % while this

percentage is 31 % in Turkey. This figure points out that Turkey is giving importance to innovation activities but in general not all the firms adopt innovation.

As a result, innovation performance of Turkey, as a developing country, is significantly lower than EU average. However, the innovation development rate of Turkey (4.1 %) is above the EU average (2.1 %). Considering these findings and the sector-specific composition of Turkey, it can be concluded that; there is a lack of private sector investment in innovation and there is a lack of financial support for innovation investments. But taking the innovation development rate into account, it can be pointed out that the concept of innovation is gaining importance from day by day.

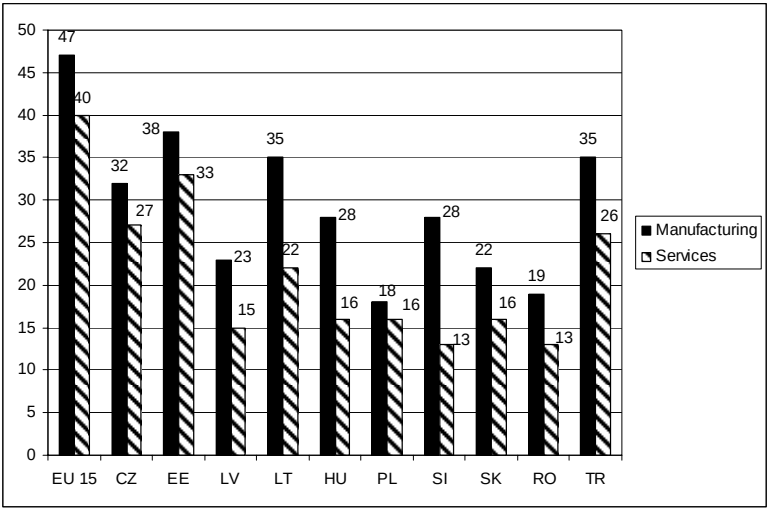


Figure 3 In 2004, firms with innovation activities

Source: Crowly, 2004.

5. Conclusion

In order to support competitive force of small and medium-sized enterprises, it is important to have SME oriented regulations. Those regulations will support the country's economic conditions while easing to do business in the country.

Since the establishment, European Union has given the necessary importance to the SMEs. It has generated number of regulations, action plans and support programs. At every level it has given importance to research and development activities regardless to the sector and it has established technology and innovation oriented SME policies.

However, in Turkey, the role and the importance of SMEs in economic and social life is only recently being realized. Up to November 2005, the lack of a uniform definition of SMEs consistent with EU definitions had hampered data collection and comparative policy/programme development on SMEs. Before the common definition of SMEs, firms in the service sector cannot be counted as "SME". So they could not benefit from

the financial and technical support schemes, which are designed for SMEs. Through this definition firms in the service sector gained the SME statue.

In the harmonization process, Turkish SMEs are forced to compete with their rivals in the unsuitable conditions. Therefore, in order to maintain its economic development and to gain advantage over its rivals, Turkey should focus to improve its innovation activities and should put emphasis on the conclusion of "National Innovation Strategy and Action Plan" activities.

References

1. Commission Recommendation (2003) 'Concerning the definition of micro, small and medium-sized enterprises', *EU Official Journal*, OJ L 124, 2003/361/EC, pp. 36-41.
2. Crowly, P. (2006) 'European TrendChart on Innovation Annual Innovation Policy Trends and Appraisal Report', *Community Innovation Statistics*, November 2004, No. 12, pp. 1-2.
3. ESN (2006) 'Entrepreneurship and Innovation Support Firmed Up', *European Innovation*, September 2006, ISSN. 1830-4338, p. 26.
4. European Commission (2006a) Reporting Intellectual Capital to Augment Research, Development and Innovation in SMEs, Belgium.
5. European Commission (2006b) European TrendChart on Innovation Annual Innovation Policy Trends and Appraisal Report. Obtained through the Internet: *http://trendchart.cordis.lu*, [accessed 10/11/2006].
6. European Commission (2008a). About the Network. Obtained through the Internet: *http://www.enterprise-europe-network.ec.europa.eu*, [accessed 5/2/2009].
7. European Commission (2008b). Putting Small Businesses First. Obtained through the Internet: *http://ec.europa.eu/enterprises/enterpreneurship*, [accessed 25/3/2009].
8. European Commission (2008c). European Innovation Progress Report 2008. Obtained through the Internet: *http://www.proinno-europe.eu*, [accessed 20/3/2009].
9. European Commission (2008d). INNO-Policy TrendChart-Policy Trends and Appraisal Report. Obtained through the Internet: *http://www.proinno-europe.eu*, [accessed 1/4/2009].
10. European Communities (2004) Innovation in Europe (THEME 9 Science and Technology), Luxemburg.
11. OECD (2006). OECD Science, Technolgy and Industry Scoreboard 2005-Towards a Knowledge-Based Economy. Obtained through the Internet: *http://puck.sourceoecd.org*, [accessed 25/10/2006].
12. OECD and Eurostat (2006) Oslo Kılavuzu Yenilik Verilerinin Toplanması ve Yorumlanması için İlkeler, Ankara.
13. Parvan, S. V. (2007) 'Is Europe growing more innovative?', *Community Innovation Statistics*, 2007, No. 61, pp. 1-6.
14. Sarı, A. O. (2005) *Avrupa Birliği'nde KOBİ Destek Mekanizmaları ve Türkiye*, İstanbul: Economic Development Foundation.

15. State Planing Organisation (2006) *Dokuzuncu Kalkınma Planı (207-2013)*. Ankara.
16. Turkstat (2006). 2002-2004 Yılları Sanayi ve Hizmet Sektörleinde Teknolojik Yenilik. Obtained through the Internet: http://www.tuik.gov.tr, [accessed 17/3/2009].
17. Yücel, İ. H. (2006) Türkiye'de Bilim ve Teknoloji Politikaları ve İktisadi Gelişmenin Yönü, Ankara: State Planing Organisation.
18. White, D. (2006) 'A Holistic Blend for Europe's Innovation Pie', *European Innovation*, November 2006, ISSN. 1830-4338, pp. 3-4.
19. World Bank (2006) 'Turkey', *Country Economic Memorandum,* Vol. 2, No. 33549-TR, pp. 175-203.
20. World Bank (2007). Science, Technology & Innovation Capacity Building in Value-added Export Enterprises Through Private-Public Instruments. Obtained through the Internet: *http://sitesources.worldbank.org/ EDUCATION/Resources*, [accessed 20/5/2008].

Measuring the Value of a Knowledge Intensive Project

Gabriel Takács

Selye J. Research Center, Selye J. University, Komárno, 945 01 Slovak Republic,
takacsg@selyeuni.sk

Setting a price for a project, which we do not know what outcome it will bring is like paying for the product what we have no information of. This paper is a thesis research plan summary what has the goal to gather practical data from field and create a measurement tool which project managers can use to assess the value of the entire project team work prior the work is started. Typically in white collar workforce management areas the end result of a team work is a vaguely defined outcome which can not be marketed. We can find several areas where project teams gather to accomplish a task which has no competition and no real price setting mechanisms in work. My purpose in this paper is to show the complexity of measuring projects in the meaning of setting the right price for the project and therefore the outcome as well, which itself includes new knowledge being created. My goal is to create a measurement tool or set of tools - based on empirical evidence – that would gather all aspects of the workflow and knowledge being added and it's connectivity to the customer's interest in the outcome.

Keywords. knowledge management, white-collar work force, productivity measurement, project management, knowledge creation

1. Introduction

The *project* term is used widely in the business field. It is referred to a finite process which brings a change to the original status quo. Generally speaking the project is a process completed within a given time period having a goal (or goals) while creating value. It has a beginning and an end usually set a priori and it's result can get either service or product characteristics. Another criteria is that this result must be different from the previously existing one. It is either new or returning to an older state, but different in at least one sense of dimension. Otherwise the project would be a nil project having nothing to do during its realization phase. A project differs from the continuous operations of management and running a business unit or other process even though these processes can have service or product results. These latter ones are continuous and produce the same result compared to a previous state, which means nothing really has been done, no step forward, no new knowledge, no new procedure has been used. Therefore new management techniques should be used while realizing a

project than running continuous business operations (Ireland, 2006). These fall under the discipline of *project management*.

A project must not be necessarily a business term. Usually it is used in the business field, where technical parameters can be defined to characterize the project realization and end product. Such project can be easier to price based on market values of resources used in the project. Broadly we can call any finite process a project which has a goal and set constrains in time and resources.

My focus is on business projects where is a need to set it's value for all parties initiated in it's accomplishment. To step forward I will examine those projects that are more knowledge intensive, where the end result is hardly marketable and where we can not find competitor or other price setting mechanism. Several such cases can include but are not limited to: tailor made software development projects, business process reengineering, team consulting, creative projects. These mind intensive projects call for white collar workforce and knowledge creation. In the upcoming sections I will illustrate how a project is constructed, where the key elements are to determine it's value for all parties included in the project accomplishment.

2. Problem description

How should we value the project a priori when we do not know the final result, neither we can determine the end product's or service characteristics, nor we have information about the value of knowledge being created during the realization phase neither we have information of future resistance and adaptability, we do not know the impact of the project on the user and market strength.

The less information we have the harder it is to measure the value of a project. Therefore in order to crate a measurement tool which could be used universally in various types of knowledge intensive projects we should be aware of the scale of information we need to obtain before starting the evaluation.

The basic concept valuing the project is the known cost-schedule-scope metrics shown in figure 1.

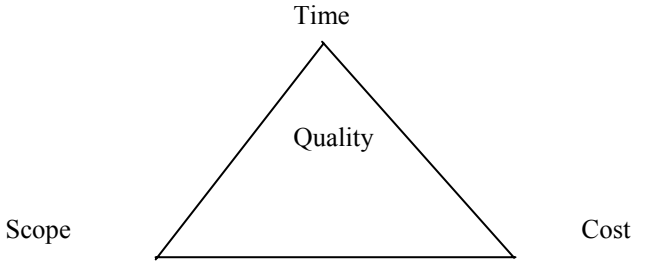


Figure 1: Project management triangle for measuring projects

Source: Lock, 2007

This metrics has been used in evaluation of project's success. Based on this three dichotomy paradigm the project must be delivered on time within budget and must be within scope, while providing quality to the customer/ developer. Several authors call for expanding this metrics with other variables. Roger (1999) and Levinson (2008) call for softer variables where customer expectations play a greater role in evaluation. Levinson (2008) even suggests that the use of the rigid cost-schedule-scope metrics causes problems in the quality and is counter productive in terms of creativity and knowledge creation.

Especially large scale projects need more attention because we can see a lack of efficiency in terms of fitting into allocated budgets, meeting scheduled deadlines, releasing lower quality products. According to Keil (1995) nearly one-third of software projects exceed their costs or schedule. Some of these projects has the tendency to "take on a life of it's own". Eventually they are redirected and represent waste of organizational resources.

Managers often see the success when the project fits the schedule, budget and requirements. Anything else is considered to be a failure. Another problem with these metrics are that they don't take into account the mercurial nature of the project delivery. Project requirements may change for a variety of reasons and during the lifetime of the realization more information is released. Interdependencies and complexity pose another threat on cost and time constrains. Such factors are not clear, neither measurer precisely, so project sponsors often see them as wasted and factors causing projects to fail. From the rigid metrics of Figure 1 we need to move toward a more sophisticated and knowledge oriented framework.

Projects are a great source of new knowledge creation. People put together into teams are the ultimate source of creativity and innovation. Why limit their human mind and built in evolution just to fit into budgets and time frames. Certainly these rigid factors are crucial in business world, but on the long run they might be overwhelmed by brilliant ideas and brand new knowledge created. My emphasis is to select those factors prom project management framework that rather positively influence the overall outcome of a project in terms of knowledge and long run success.

Instead of measuring individual productivity I will focus on team productivity where not so easily captured measures should be implemented. Instead of labor intensive improvements I will search for motivators in white collar workforce management. While understanding the importance of technical tools being used under project management, I will try to eliminate their effect on value creation. Merely because these technical tools are available for everyone, so noone gains advantage of using them. I will rather focus on structural and managerial techniques that could facilitate work and could result in a higher quality project either in terms of the finished product or in terms of project process and the road toward accomplishment.

3. Project framework

In this paragraph I will take detailed view on various aspects of a project. First I begin with analyzing project from the point of view of all parties involved in the completion of the project. Then I analyze the structures and processes related to a project. This

view is rather classical but necessary to point out critical factors that can have influence on success and determine the value of the project. Knowledge is described in connection to projects and it's importance is highlighted. Creativity is introduced as the real value and novelty created during project implementation. All of these sections are closely linked and provide the base for creating the measurement tool in the upcoming section.

3.1 Parties and win-win concept

For my purposes I will define three major groups in project management. Each of them is a stakeholder in the project's success (Boehm and Bose, 1994) identifying their win conditions (or value propositions) about the project. There can exist more than three major critical stakeholders of a project, but for my purposes I will define three classes:

First is the customer or sponsor, who pays for the project and is usually the main risk taker in the project. It influences the project's end result characteristics before the planning and during the realization phase. This stake holder has the right to chose many aspects of the project and has major influence as a customer with it's rights. Taking high responsibility in the planning phase.

Second there is the project development team, who actually conducts the project, together with the sponsor plans the project and later on executes the project. The team can be set up by the project manager, who is responsible for the success and win-win negotiation. The project team develops knowledge, can be characterized by size, structure, motivation, level of authority. This stakeholder takes the risk of quality and meeting constrains. Itself can not be evaluated only to the limits given by customers and later on after implementation by the used. During the project process phase the developer team can change in characteristics as influenced by outside or inside factors.

Third there is the user, who as a stakeholder in the project has the least influence but also has the fewest to risk. Mostly the customer of the project sells the outcome of the process to user and thus the user has little contact with the developer team. This though might be a disadvantage if we are building a long term relationship between the user and the customer via the project. The user is the ultimate decision maker and is critical in success of the project. In some cases though, not even the user is capable of determining whether the project is a success or not. Such cases might be when the developer team is influenced heavily by the user and favors their preferences over the customer's ones. Then really the project serves the user but could drive the customer into bankruptcy or can not keep up the project because of loose constrains.

There is an interaction between these three major stakeholders and if these interactions lead to the highest possible optimum, then we can talk about ultimate success of the project. This is a win-win situation for all of them. The customer gets a quality product with new knowledge and creative ideas, keeping all constrains under limits, while the project development team receives the desired compensation and the user's needs are satisfied on a higher level. All these happen in an environment which is hostile, changing and unpredictable. If all these are done, the project can be highly valued and the outcome is mutually beneficial for all of them.

3.2 Project management process and structure

According to Uppal (2008) at most companies or other organizations projects are executed through the use of some sort of project management process. This is a systematic method for evaluating and executing projects. It is a structured process with clearly defined five phases. Each five phases has its own objective and deliverables. These phases are:

1. **Conceptual phase** – occurs before the proposal is classified as project.
2. **Analysis phase** – the purpose of the analysis phase is to select the preferred project option
3. **Definition phase** – the purpose of the definition phase is to finalize the project scope, cost and schedule and get final project funding
4. **Execution phase** – the purpose of the execution phase is to produce an operating asset consistent with scope, cost and schedule
5. **Evaluation phase** – the purpose of the evaluation phase is to evaluate asset operation, in order to validate the economic justification for the project, and ensure performance to specifications and maximum return to the stakeholders.

Once all three parties agree on the outlines of the upper phases (users are not directly involved in decision making, only the customer represent indirectly and partially that point of view in the negotiation process), the major guidelines are written in a contract. Such contract includes the followings:

Purpose, general description, scope description by project breakdown structure, assumptions and allowances, estimate exclusions and exceptions, execution strategy, pricing, contingency, benchmarking and reference documents.

The phases are connected to parties as follows:

The project customer has the responsibility and the authority for approving the project mission and business goals, providing early funding and resources, and managing the approval process for the project. The project development team has the responsibility for defining the project boundaries and business objectives and for ensuring that items affecting business strategy, economics and risks are managed. (Uppal, 2008)

Lock (2000) defines six stages of a project which differs from the above only by adding a phase in between 3 and 4, the definition phase and execution phase. By Lock there we should handle the “purchase of goods and services” as a unique stage in project life cycle. Keeping in mind the importance of procurement I can not take into account this separation, since in that phase every project manager faces the same market conditions and can obtain the same goods and services. Therefore it does not play a significant role in project evaluation and project pricing. Neither it will have effect on overall level of new knowledge and creativity.

Analysis phase and definition phases are mostly the project manager’s responsibility. Generally speaking these are the planning phases. In this stage the overall structure is defined and the way the project will be executed is planned. How rigid or flexible the planning phase is have a great impact on the final outcome, thus on the success of the project. The planning should rather be standing on solid data and forecasts. By doing

so we can judge all constrains as precisely as we can. A major part in analyzing the background is the environment mapping. This broader perspective allows the project manager to place the project to a broad situation and try to estimate the impact it has on the environment and vice versa. The quality of the environment study determines the overall outcome of the project.

3.3 Knowledge

Knowledge at the individual level can be understood as a cognitive capability of mental or physical action (Nonaka, 1994). For project management purposes the distribution and share of such individual knowledge plays a role. If we take a person with a given set of knowledge and place her into a team of other people with their own set of different type of knowledge the crucial issue will be how they deal with each other's knowledge (Argote and McEvily and Reagens, 2003). Research has shown that productivity of knowledge workers is associated with individual characteristics, such as breadth and depth of knowledge, work efficiency, capability of learning, and motivation (Reagans and Zuckerman, 2001). The productivity of such team members also depends on how they use accumulated and second source knowledge. Evidently when we create such process where people can share and distribute their own knowledge and make it public (within the project domain), where the cost of acquiring is zero (no exchange is necessary) the project team will be facilitated and internal boundaries will fall. This will build trust and openness. Because of tasks in knowledge intensive projects are complex, requiring coverage of various field, collaboration becomes another crucial issue. Collaboration is when people meet and engage in mutual problem solving approach. An extreme collaboration can be one sided where one is a mentor only providing information and sharing her knowledge, while the other party only perceives and accepts new knowledge and applies in work. If motivation is set, it can be beneficial for both of them.

Advanced information technology has greatly influenced ever aspects of knowledge intensive work environment. Project management uses these IT techniques to facilitate work to meet the time constrain. It is obvious that faster and robust software and processors can gain faster performance whatever type of projects we are talking about, but as said previously I will not evaluate projects based on these IT technologies, since they can be obtained by everyone, so no real competitive advantage can be gained by using them. I will not reject the other help provided by information technology and that is how it helps in knowledge intensive environment in knowledge sharing and distribution. An example (Dyer and Nobeoka, 2000) could be a project team widespread globally working on the same project utilizing all technology available to spread their piece of work and building a knowledge database. Virtual teams can use personal and cultural background knowledge to enrich the overall project if needed. Broader (global) perspective facilitated by IT means new ideas shared rapidly. IT functions also support several phases, first in planning, forecasting and later at controlling. Sophisticated monitoring techniques can give managers insights at every stage online. Last but not least knowledge seeking is eased by IT and web technologies. Searching for databases, publicly available information of secondary sources can be easier using IT. All above in terms of richer knowledge rather than being fast and quick gathering information.

3.4 Creativity

Creativity is described (El-Murad and West, 2004) as “creative thinking” or “ability”, “problem solving”, “ imagination”, “innovation”. They are obviously important aspects of valuable projects. Creativity is the ability to produce work that is novel, original, new or unexpected. Novelty itself is not sufficient, the work must be also valuable and useful to be considered creative.

Creativity in contrast to Knowledge is considered riskier. Creative workers in projects are needed when a large proportion of the success is based on how novel the product will be or how novel the implementation process will be. In their work content a high proportion of new ideas will override not only structures and systems, but disciplines as well. Therefore creativity can be destructive and risky in terms of the rigid cost-schedule-scope metrics. But not in terms of long run success and quality work. Creativity’s goal is not only to be new or interesting, if we stick to the value creation concept, creativity becomes the real driving force of all projects and will be considered as a must. Creativity in contrast to knowledge can not be shared and distributed. It refers to individual characteristics, such as intelligence, broad interest, intuition, self-confidence, attraction to complexity etc. (Amabile, 1983). Several authors (Oldham and Cummings, 1996) in creativity imply that work context such as task complexity, deadlines, goal orientation, perceived evaluation, supervisory styles improve creativity. By increasing job complexity and enhancing supportive management style can improve creativity.

Either personal or work characteristics determine creativity, in project management both can be measured and valued. As long as we consider the white collar workforce to be the one who determines the projects success internally creativity plays a major role in this process.

4. Toward the universal metrics and research methodology

Based on opinions of previously mentioned authors (Levinson, 2008; Roger, 1999) it seems that the classic metrics what is imposed on project evaluation, the cost-schedule-scope does not provide objective value on success. Moreover it seems to be counter productive and demotivates project developers to produce a quality project.

Therefore I will suggest a new metrics system based on investigation and empirical research. This research intends to cover all aspects of a project and has a long run success focus with a win-win optimum for all parties involved.

My goal is to come up with a metrics system that could be useful for all parties involved, especially useful for the project sponsor to evaluate any knowledge intensive project a priori to the start and later on during evaluation and completion. The main feature of this universal valuation metrics would be that it motivates all parties to reach a better result in all phases.

Model

First it should be measured how knowledge intensive the project is. This should be done before starting the project in the conceptual phase. This metrics is independent

from other more complex investigative evaluation metrics. Simply we need to know the knowledge intensity of the project to be able to state that we will step forward from the cost-schedule-scope metrics system. On one extreme a non knowledge intensive project would have a 0 knowledge level, where the project itself is technical, rather mechanical with 0 need to implement thinking. This kind of project is merely a computation, construction or other type of work, which could be done by tools, computers or other equipment without human interference. Such projects can be exactly and precisely evaluated by the existing cost-schedule-scope metrics. That is because we can measure and plan the outcomes and processes with 100% certainty and the project sponsor does not need from the project developer to put new knowledge into the project.

Examples of projects with few or even 0 level of knowledge could be: small scale building projects, operative projects, administrative projects, things to do "list" style of projects and so on.

On the other extreme we can think of a project which is high on knowledge or extremely knowledge intensive, where the project sponsor demand novelty and innovative thinking, where even the plan phase and conceptual phase is novel, by stating that we would even start to think about the project in a new way, plan it in a new and novel way, next we demand a think thank style of work from the project team where knowledge sharing and distribution is the key to successful completion, where creativity is at all level and at every member. Then we want the user of the project to be involved in a new way, where we require the user to step forward to a higher level and give them the end result of the project to be used in a different and better way than he ever expected. Such knowledge intensive project can not be measured in terms of cost-schedule-scope, rather new metrics are needed. Examples could be: tailor made software solutions, new large scale construction projects in difficult environments, problem solving in corporate consultancy, complex technical boundaries to be broken, product innovations, etc.

Obviously my proposed universal metrics could be used in knowledge intensive projects, where we keep the cost-schedule-scope metrics to shade, not to harm the creative thinking while going through all phases of the project.

Evaluation of the project then should start by answering the following question:

Q 1:What is the project's knowledge intensity and novelty level?

Answer: Empirical testing and field research should give the answer to where to set the limit for cost-schedule-scope metrics and where to start to implement new metrics:

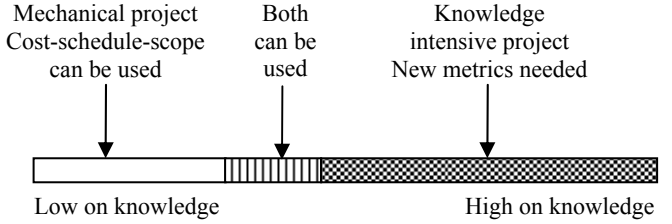


Figure 2: The level of knowledge intensity and novelty of a project

Earlier mentioned three major stakeholders are involved in a project, the project sponsor, the developer, and the user. Each has different set of values and prepositions. The following metrics involves all of them because we estimate that the win-win optimum is the desired, so all opinions and point of view should be involved.

I take the following phase break down of a project to be the base for the evaluation: conceptual, analysis, definition, execution, evaluation. It is important that we do evaluate at each phase and not only once as it is suggested by the traditional cost-schedule-scope metrics. This one measures the project's success only at the end, therefore does not play a motivation role, rather a negative motivator, thus not leading to the best result.

My concept is a process evaluation and not a result evaluation, although the result is evaluated as well, only in terms of other aspects. Rather a long run point of view is used when evaluating the result to get the product life time cost as a base and not an absolute number at the time of ending the project. This long run and overall phase evaluation concept has a major advantage: it motivates all parties to give their best work at all time, before, during and after the project. This is a positive motivator scheme in-built to the metrics.

One might think that it is useless to evaluate a project at it's conceptual or analysis phase, where the concept is only forming and various impacts act on the project. This is nothing else than the environment of the project, which as we saw from Oldham and Cummings, (1996), Shalley (2001) the context plays a major role in creativity, which enhance new knowledge being added to existing ones and facilitate knowledge sharing and distribution. In these early phases of the project several outside effects act positively: intense competition, threats and unsecure conditions of the market, forecasted business opportunities, and other volatile details of the environments. These enhance the project and could be constrains and motivators. How strong these outside forces are, how they change the work habits of the parties play a role in overall project success, thus should be counted in the evaluation metrics.

The definition phase is the most important part and should be taken care of the evaluation metrics as well. A good plan sometimes is more valued than the execution itself. A very well done planning and structure and organization procedures could yield to better results than if we have poor planning but good execution. Rather focus on planning and project definition than paying for expensive execution. This does not mean that execution is not important and we should neglect it, it just point out that planning is a step further in importance.

Execution phase is the phase which is purely under the project manager's competence. It is the developer's area to bring the most out of the project. This needs good motivation factors to enhance knowledge and novelty creation. Once again in knowledge intensive projects cost, schedules and scope is shaded and plays minor role in evaluation of execution. Going out of scope could yield interesting market opportunities, being late could yield to more mature results and being on advantage on long run, being out of budget if necessary could rather be a motivator than a bad signal. All this should be measured in a context and for all parties involved. User's preferences on the long run is should count the most.

The evaluation phase is rather a control phase, where the project parties launch the project's results and when the project itself end with the desired result. Definitely this must pay a major role in evaluation and not the project previous parts. This evaluation phase shows the financial return and return on investment on the long run. Taking this phase into evaluation makes all parties, especially the developer cautious and focus on long run success. Any changes or reopening the project and amendments comes automatic in this phase, but if previous parts score high this would less likely to happen.

The following table itself is the universal metrics table to be used in knowledge intensive projects. It should be a constant evaluation starting from phase one till five, counting with value prepositions of sponsor, developer, and user. Each question in the matrix should be carefully examined and given a score, which would add up to the final score of the project. This final score comes up in time as the project continue, not valuing anything too early, neither late, juts on time, when it happen. This could function as a motivator. All parties should be aware of the evaluation, knowing it in advance motivates to a better result. It follows the principle that self evaluation is contrasted to supervisor evaluation to gain a break even point where all parties reach their win-win strategy (Nakata and Fukuda, 1999). Two parties are confronted with the same question and the common answer generates the score. The sponsor sees thinks from the other side as the developer, the developer sees it from other side as the user, so confronting self evaluation with supervisor evaluation could generate a more objective score. The feedback principle is a key concept and the evaluation's continuity is the other one. The final score is reached only after the final phase, therefore financial settlement is due only after the final score is reached.

Phases:	Project sponsor	Developer	Developer	User
Conceptual Weight: 0-100%	Competitiveness of market Technical background Scenario preciseness Score: 0-10		Existing problems Purchasing power Score: 0-10	
Analysis Weight: 0-100%	Contract characteristics Developer involvement Score: 0-10		User availability Existing knowledge base Score: 0-10	
Definition Weight: 0-100%	Flexibility of structures Staffing Motivation scheme Score: 0-10		Developer-user benefits of share User diversity Score: 0-10	
Execution Weight: 0-100%	Trust and adhesives inside Knowledge share Technology enhancement Score: 0-10		Knowledge share User involvement Score: 0-10	
Evaluation Weight: 0-100%	Marketability and return Maintenance Resistance to copy Score: 0-10		Safety Add on possibility Simplicity Score: 0-10	

Table 1. Universal metrics – areas to evaluate by phases

Source: Own design

To get the scores we need to prepare additional questionnaires towards all parties in all areas. Sometimes several questions will be needed to get one score. This additional questionnaire can be project are specific, therefore in each industry could be varied. Empirical research is needed to validate the following table's reliability, and if needed modifications would be necessary. An estimated minimum 100 knowledge intensive project evaluation with this metrics could yield to a basic result, which would be used to amend the necessary question areas in each phase.

5. Methodology

In each phase a short questionnaire is prepared for those involved in the project from all parties. First the sponsor and developer are confronted the same questionnaire in each phase to self evaluate first their point of view and then the counterparty. The same is done with the developer and user on different topics where each of them presents their own perspective. Only quantitative questions should be asked and a final score between 0 and 10 should be generated in each phase, 0 indicating a priceless, valueless result, while 10 indicating a mutually prospering and expected high quality outcome. There are 10 areas, 2 in each phase, one for sponsor-developer, one for developer-user. Evaluation is done throughout the project phases and the final 0-100 score is reached by adding the mid scores and weighting them with the previously set weights *10 after adding the numbers. These weights represent how each phase is valued by the parties. All weights should be set in a negotiation process. The final number would represent the overall universal value of the project. This final evaluation number could be used as a base for financial settlement and revenue distribution throughout the project. Certain cost-schedule-scope limits could be set outside of this metrics as base starting point, but that is not necessarily needed. I expect from the upper metrics to be useful even without the traditional cost-schedule-scope metric, but that would require several experimental research. That kind of research is out of my possibilities and certainly may contain risk elements at this phase. First ex post or side by side empirical research is necessary to prove the upper table's consistency. Field research in knowledge intensive industries should be used as base.

6. Conclusion

The evaluation tool presented in this paper is intended to fulfill two purposes. First it is to help in pricing the project and find it's real value under win-win conditions based on long run perspective. Second it has to fulfill a strong motivation toward a highly knowledge intensive and creative thinking. Such goals show beyond the existing cost-schedule-scope metrics and are more flexible and require more involvement. It might be even measured by an outsider or consultancy. Therefore getting a broader perspective. The research will focus to organizations in Central and Eastern Europe and Western style corporations operating in these geographical areas, who are active in knowledge intensive industries. The research will be short sighted on cultural background because of the persons involved in filling the questionnaires and their common style of thinking. In a future research a more global perspective should be tested. According to my hypothesis, cultural differences and diversity may play a

major role in a quality project work (Kamath és Likert, 1994). In practical world there are many needs in customers, project sponsors that are latent, simply because the customer does not really have the solution and knowledge background, but only the basic need. The solution thus can range out to various possible outcomes fully under control of the project team and interaction between the three parties. Then the project team's responsibility is to find the best solution to the need, bearing the full responsibility to the future add-ons and lifespan of the outcome. What price should we set to this kind of project is given by the universal metrics.

References

1. Amabile, T. M.,(1983) 'The social psychology of creativity – a componental conceptualization', *Journal of Personality and Social Psychology*, Vol. 45, No. 2, pp.357- 376.
2. Argote, L. and McEvily, B. and Reagens, R.,(2003) 'Managing knowledge in oragnization: An interactive framework and review of emerging themes', *Management Science*, Vol. 49, No.4, April, pp. 571-582.
3. Boehm, B. and Bose, P., (1994) 'A collaborative spiral software process model based on theory W', *Proceedings, ICSP 3, IEEE*, October, Reston, Virgina, U.S.A..
4. Dyer, J. H. and Nobeoka, K., (2000) 'Creating and managing a high-performance knowledge-sharing network: The Toyota case', *Strategic Management Journal*, Vol. 21, pp. 345-367.
5. El-Murat, J. and West, D.C., (2004) 'Definition and measurement of creativity', *Journal of Advertising Research*, Vol. 44, Issue 2, p188-201.
6. Ireland, L.R. and David Cleland (2006) *Project Management: Strategic Design and Implementation*, McGraw-Hill Professional, ISBN: 007147160X., pp. 100-110.
7. Levinson, M., (2008) 'Common project management metrics doom IT departments', *Network Asia*, Vol. 4, Issue 9, pp.8-8, 2/3.
8. Lock, D., (2007) *Project Management*, Gower Publishing, ISBN: 0566087723.
9. Kamath, R.R. and Likert, J.K., (1994) 'A second look at japanese product development', *Harward Business Review,* Jan-Feb., pp. 154-170.
10. Keil, M., (1995) 'Pulling the plug: software project management and the problem of project escalation', *MIS Quarterly*, Vol. 19, No. 4., pp. 421-447.
11. Nakata, T. and Fukoda, H. and Yang, Q., (1999) 'Construction of an indicator for measuring white-collar productivity', *Journal of Japan Industrial Management Association*, Vol. 50, No. 1, pp. 20-26.
12. Nonaka, I., (1994) 'A dynamic theory of organizational knowledge creation', *Organizational Science*, Vol. 5, No. 1, pp. 14-37.
13. Oldham, G. R. and Cummings, A., (1996) 'Employee creativity: Personal and contextual factors at work', *Academy of Management Journal,* Vol. 39, No. 3, pp. 607-634.
14. Reagans, R. and McEnvily, B., (2003) 'Network structure and knowledge transfer: The effect of cohesion and range', *Administrative Sciences Quarterly*, Vol. 48, pp. 240-267.

15. Reagans, R. and Zuckerman, E.W., (2001) 'Networks, diversity, and productivity: the social capital of corporate R&D teams', *Organizational Science*, Vol. 12, pp. 502-517.
16. Uppal, K. B., (2008) 'Project management, cost engineering, project definition, action plans of what?', *Paper presented at AACE International Transactions – 52nd Annual meeting, June 29-July 2, 2008,* Toronto, Canada

Insolvency of European Airlines. A Typical Case of Cross-Border Insolvency

Tsanidis Vassileios

Democritus University of Thrace, Law Faculty, Department of Labor and Business Law, Miaouli 15, GR-55236, Thessaloniki, vassilis.tsanidis@gmail.com

Air transport was until recently a very fragmented sector. State protectionism ensured that national flag carriers would enjoy privileges, which were against any kind of competition. Globalization and the increase of international commerce urged to the liberalization of the aviation sector market. In Europe this liberalization occurred gradually, so that the national airlines would have the time to adjust to the new environment. Airlines are now being treated as regular companies; therefore, due to the world economic crisis they are open to the risk of financial failure. The insolvency of an air carrier in Europe is most of the times a complicated issue, due to the fact that it involves assets and commercial activity in other member-states. For that reason, it is essential to know the procedure which has to be followed in order to avoid conflict of laws in this typical case of cross-border insolvency.

Keywords. air transport, bankruptcy, European Community competition law, market liberalisation, European Union integration, cross-border insolvency

1. Introduction – special characteristics of air transport.

Air transport has particular features, different from those of transport in general. Therefore, it is essential to take into account these particularities before examining the legal position of the aviation sector.

Despite its international character, air transport has or better used to have a strong national character as well. Historically, this goes back to the First World War, when air space was for the first time recognized as part of a country's territory. Flying over a country depended only on the government's approval. This of course led to the dependence of airlines on the governments, which controlled them financially. Due to this fact every State, even the small ones, had to establish its national airline that would operate as flag – carrier throughout the world and serve the national feelings as a symbol of progress for the whole country (Dagtoglou, 1994).

This sector of transportation was considered very sensitive and therefore was left out from every integration process at the European Union (E.U) level. The legal basis for this exemption was article 80 par. 2 of the European Community (E.C) Treaty, which limits the full application of the Title V (ex IV) regarding the European policy for

transportation only to rail, road and inland waterway transportation but allowing the Council to decide unanimously whether, to what extent and by what procedure appropriate provisions may be laid down for sea and air transport (Nettesheim, 2002).

2. Liberalization process of the air transportation sector

Until recently air transport was one of the most regulated sectors within the E.U. Protectionism of national air-carriers by governments was very common and in some Member-States the above mentioned airlines enjoyed even a status of monopoly. Competition and its rules did not apply having as a result the fact that the market - wherever that existed – operated in a fragmented manner (Kostopoulos, 2005).

Signs of change began to appear in the mid '80s. As political support grew, two decisions of the European Court of Justice (ECJ) urged the Council to adopt the required provisions for the creation of the single market in aviation sector. The first decision was the "Nouvelles Frontiéres" decision (European Court of Justice, 1986) , through which the ECJ declared that according to art.80 par.2 of the E.C Treaty, air transport is excluded only from the provisions of the Title V of the Treaty and not from the whole Treaty. The significance of this decision lies upon the fact that it declared that E.C competition rules should be also applied to air transportation. With the second decision "Ahmed Saeed" (European Court of Justice, 1989), the ECJ took a step further. It stressed out that liberalization should be accompanied by a consistent protection of free competition, so that European aviation could survive competition at international level (Dagtoglou, 2004).

The gradual liberalization of the intra-community commercial aviation was implemented through the introduction, on a phased basis, of three packages of measures aiming at the creation of the single European sky (Camesasca, Goeteyn and Soames, 2004) . The first package came into force in 1988, the second in 1990 and the most significant third one in 1993 with the exemption of the cabotage protection which was lifted in 1997(Hohenstatt and Jestaedt, 1992). The creation of the single aviation market in Europe can be seen as a continuation trend that had taken place in the U.S. during the decade of 1970(Morell, 1998). The latter process, due to its radical character, was called deregulation, instead of liberalization, a term which was used by the European Commission in order to stress out the integration process that was implemented gradually (Button and Drexler, 2005 ; Kostopoulos, 2005).

The liberalization of the aviation sector in both sides of the Atlantic (and in most countries worldwide) meant that traditional airlines which relied on state aid, tax exemptions, restrictions on market access etc. had to adjust and learn to operate in a new more competitive environment(Meunier, 2005). In Europe, though, even after the creation of the single market in air transport, the European Commission and the ECJ in many cases preserved parts of the previous fragmentation of the airline industry by recognizing to some European national air-carriers originating from the big member-states a leading role in the European air transportation (Kostopoulos, 2005). But the privileged treatment of the national flag carriers by the above mentioned institutions, couldn't change the new economic landscape, where new and most of the times low - cost airlines entered the market and increased the competition.

3. European Air Company

The definition of the term "European air company" is very crucial for the European insolvency of airlines. Only European air carriers could enjoy the benefits from the liberalization of the market and the creation of the single European sky (Guinchard, 1998 ; Mühlke, 1995; Nettesheim, 2002).

In July 1992, European Council adopted the Regulation 2407/92 on licensing of air carriers where the definition of the term European airline was to be found. Pursuant to article 2 par. c of the Regulation 2407/92, operating license means an authorization granted by the Member - State responsible to an undertaking, permitting it to carry out carriage by air of passengers, mail and/or cargo, as stated in the operating license, for remuneration and/or hire. Pursuant to par. b of the same article, air company is an air transport undertaking with a valid operating license.

The European air company is an air carrier that has already received an operating license by the competent authorities of a Member - State. More specifically, article 4 paragraph 1 states that no undertaking shall be granted an operating license by a Member - State unless:

(a) its principal place of business and, if any, its registered office are located in that Member - State; and

(b) its main occupation is air transport in isolation or combined with any other commercial operation of aircraft or repair and maintenance of aircraft.

4. Airline Bankruptcies in the European Union and in Switzerland

The new circumstances in the aviation market in conjunction with the downturn of the world economy in the latest years lead the traditional full service airlines, which lost substantial market shares, to significant economic difficulties. For that reason, bankruptcy or government inspired rescue operations were inevitable not only for small airlines but also for big air companies which in the past used to be national flag carriers and enjoyed monopolies in their own countries.

The European Commission noticed that in October 1996, 25 new air companies were founded and 16 of them disappeared very soon (Chatzinikolaou, 2005). The first insolvency case, concerning a previous national flag carrier, was that of Swiss-Air. In October 2001 Swiss-Air filed for bankruptcy. Of course the events of 11th September in the U.S together with the increasing competition and the establishment of many low – cost airlines had a fatal impact on the economic viability of the above mentioned company (Balfour, 2003). The Airline was kept alive until March 31, 2002 by the Swiss federal government. On April 1, 2002 the successor Airline, Swiss International Air Lines was founded, taking over most of the routes, planes and staff of former Swissair.

Little later its alliance partner the Belgian air company Sabena announced that it was on the verge of insolvency too. The Belgian government after the permission of the Commission granted a rescue loan, which was insufficient to save Sabena, but enabled the creation of the company named Brussels Airlines, based namely on the surviving

Sabena subsidiary (Camesasca, Goeteyn and Soames, 2004). Alitalia and Olympic Airways faced also many financial difficulties. Especially Olympic Airways one of the biggest air carriers in South-Eastern Europe faced excessive losses and therefore, in 2003 the government decided to restructure the Olympic Airways Group of Companies. The subsidiary, Olympic Airlines S.A**.** took over the flight operations of Olympic Airways and the remaining group companies merged and formed a new company, called Olympic Airways - Services S.A(Featherstone and Papadimitriou, 2007). The new airline also continued to face financial problems and for that reason it decided its privatization, a process that was concluded very recently with the agreement to sell the state air carrier Olympic Airlines to Greece's private Marfin Investment Group.

5. Cross-Border Insolvency in the European Union

When an air company has assets only in one Member-State and its commercial activity is restricted only in this State then the national law for bankruptcy of the same State applies and indicates the procedure that has to be followed in order to satisfy the creditors of the insolvent company. Airlines, though, are not companies whose trade activities are restricted in the territory of only one State. Most of the times they have assets in States different of that of their corporation, and in order to promote their business they open branches in several countries abroad. In case an air company – as it is very common - has developed commercial activity all over Europe, how could we find the competent national authority to deal with this issue and how countries, where the above mentioned company has assets or commercial activity could cooperate in order for the liquidator to satisfy as many creditors' claims as possible? (Warren, 1993 ; Wilmowsky, 2000) For that reason it is very interesting to examine the nature of the existing laws in E.U field and their interactions in case of cross-border insolvencies.

The legal basis of the approach of international insolvencies in the European area is article 220 paragraph 4 (now article 293) EC. This article gives Member States the potential to conclude international law agreements regarding simplification of procedures for the mutual recognition and execution of judiciary and arbitrary decisions. The special nature of international insolvencies and their importance for the security of transactions imposed its distinction from the rest matters of commercial and civil law (Fletcher, 1999 ; 2002).

At multilateral level, an effort was made by a special committee for the preparation of a Draft Treaty based on the provisions of the above mentioned article. At the beginning of 1980s the Committee introduced this Draft to the Council of Ministers for approval, but it was abandoned soon after the non-achievement of a general agreement (Virgos,1998).

A radical change regarding the legal basis of this matter was introduced by the Amsterdam Treaty, which gave the EU the potential to regulate cases related to the judicial cooperation on civil matters. Thus the Council of the EU, having as a legal basis articles 61c and 65 of the EC, adopted the Regulation 1346/2000 on insolvency proceedings. The adopted text was similar to the text of the Draft Treaty. Most

provisions remained in fact as they were, and simply their legal form was amended, in order to avoid, for their adoption, the consensus of all Member States (Fletcher, 2002).

The system regarding the European insolvencies that the Regulation 1346/2000 introduces is very close to the principal of universality. According to that principal, the debtor's insolvency in one Member-State has a direct impact on all his assets, located in other Member States as well. The insolvency proceedings, concerning the totality of the debtor's assets, is being provided by the law of the State where this procedure was opened (lex fori concursus), usually the State of the debtor's residency(Huber, 2002).

Moreover, all creditors, irrespective of their nationality, are being satisfied according to the way determined by the lex fori concursus(Lehr,2000). From the aforementioned definition it is evident that cross border insolvency proceedings based on the universality model, presupposes the mutual recognition of its results by all involved Member States, if it is to operate correctly. Thus, a State, which adopts the universality principle, has to give to its national law on insolvency proceedings, an international application (Leible and Staudinger, 2000). The latter is afforded from the moment the Member-State considers itself competent to open insolvency procedure in cases for example where the debtor's main establishment is located in its territory. On the other hand, this State has to accept the results of an insolvency opened in another State, due to the existence of the debtor's assets situated in its territory.

Taking into consideration the nature of the European Union and the diversity of national laws of the Member – States the principal of universality, which is based on the model one debtor, one insolvency procedure couldn't be able to be effective. It would surely affect the functioning of local economies and would probably create problems, due to the fact that every national legislation is not inspired by the same ideas regarding the issue of insolvency proceedings and thus a situation non-compatible with their legal tradition may arise. For that reasons the Regulation 1346/2000 adopts the system of controlled universality, because it restricts the universal effect of the lex fori concursus by enabling the opening of secondary insolvency proceedings in the Member – States where the insolvent air company has an establishment (Mastromanolis, 2001). These secondary insolvency proceedings act in parallel with the main insolvency procedure they have only local effect and there is no possibility to surpass their local impact and assert a universal application (Wimmer, 1998).

As it was noted before, air carriers in Europe are not companies which are restricted to the territory of one Member State. Due to the nature of their commercial activity very often they create establishments in other Member States and they participate in the development of the local economies. Therefore, the provisions of the Regulation 1346/2000 assert application in the insolvencies of air companies, regulating the interactions between national legislations that might appear in such cases. Taking into account the implications of the liberation of the European market of the aviation sector it wouldn't be strange if an insolvent air company, which used to be a national flag carrier, through the provisions of Regulation 1346/2000 would have to accept the effects of a law different of the law of its corporation or its centre of main interests for the establishments that might have in the other Member – States.

It is true that the article 3 par.1 of the Regulation 1346/2000 refers to the centre of the debtor's main interests as presupposition for the lex fori concursus. Taking distances from the incorporation doctrine, the above mentioned Regulation adopts the real seat theory, since it notes that the place of the registered office shall be presumed to be the centre of its main interests only in the absence of proof to the contrary. This goes in line with the provisions of the Regulation 2407/92, because as it was noticed earlier the licensing of air carriers is closely linked to their principal place of business. Both Regulation 1346/2000 and Regulation 2407/92 refer to the law applicable to the company's actual domicile (*real seat*). Therefore, in cases of insolvent air companies, the main insolvency procedure should be opened in the Member-State of their registered office only if there is no proof that the real seat of the air company is not located in any other Member – State (Mastromanolis, 2001).

6. Conclusion

Insolvency of airlines is a typical case of cross-border insolvency. It affects all the economies where the insolvent air carrier had developed economic activity. Liberalization of the European market of the aviation sector and the creation of the single European sky encouraged cross-border transactions and mergers of companies which used to be national flag carriers (Air France – KLM case). Taking into account the world economic crisis and the absence of state interventions, it is highly likely in the near future to deal with more insolvency cases of airlines in the European area. Therefore the existence of a procedural framework for the cross-border insolvencies in the EU is very crucial in order to avoid conflict of laws and to protect the faith of the investors in the aviation sector.

References

1. Balfour, J. (2003) 'Sate aid and the aviation industry - unfair competition'. Paper presented at the 2003 fall meeting of the American Bar Association, section of international law and practice. Transatlantic Dialogues in Law and Practice. October 15-18, 2000. Brussels, Belgium.
2. Button, K. and Drexler, J. (2005) The Implications on Economic Performance in Europe of Further Liberalization of the Transatlantic Market, Fairfax U.S.A : Center for Transportation Policy , Operations and Logistics, School of Public Policy, George Mason University.
3. Camesasca, P., Goeteyn, G. and Soames, T. (2004) 'European Aviation Law: New Wings Unfolding', Air and Space Law, vol.29, no.2, pp. 114-136.
4. Chatzinikolaou – Aggelidou, R. (2005) Air transport of passengers according to international conventions, the national and the community law of air transports, Thessaloniki : Sakkoulas (in Greek).
5. Dagtoglou, P. (1994) Air transport and the European Union, Essays and Comments, Athens : Sakkoulas.
6. European Court of Justice (1986) Criminal proceedings against Lucas Asjes and others, Andrew Gray and others, Andrew Gray and others, Jacques Maillot and others and Léo Ludwig and others. References for a preliminary

ruling: Tribunal de police de Paris - France. Joined cases 209 to 213/84, Judgment of the Court of 30 April 1986, European Court reports 1986.
7. European Court of Justice (1989) Ahmed Saeed Flugreisen and Silver Line Reisebüro GmbH v. Zentrale zur Bekämpfung unlauteren Wettbewerbs e. V. Case 66/86, Judgment of the Court of 11 April 1989, European Court reports 1989.
8. Featherstone, K. and Papadimitriou, D. (2007) 'Manipulating Rules, Contesting Solutions: Europeanization and the Politics of Restructuring Olympic Airways', Government and Opposition, vol.42, no1, pp.46-72.
9. Fletcher, I. (1999) Insolvency in Private International Law, National and International approaches, New York: Oxford University Press.
10. Fletcher, I. (2002) The Law of Insolvency, London: Sweet & Maxwell.
11. Guinchard, M. (1998) 'The European regulation of air transportation', Revue française de droit aérien et spatial, vol.205, no.1, pp.5-34(in French).
12. Hohenstatt, K. S and Jestaedt, T. (1992) 'The third package of measures for the liberalization of the air transportation', Europäische Zeitschrift für Wirtschaftsrecht, vol.4 , pp.116-119(in German).
13. Huber, P. (2002) 'The european regulation for insolvency', Europäische Zeitschrift für Wirtschaftsrecht, vol.16, pp. 490-495(in German).
14. Kostopoulos, K. (2005) The obligation of public services to airlines and airports in the framework of community competition law, Athens: Sakkoulas (in French).
15. Lehr, W. (2000) 'The new european regulation for insolvencies', Konkurs. Treuhand. Sanierung. Zeitschrift für Insolvenzrecht, vol. 4, pp.578-585 (in German).
16. Leible, S. and Staudinger, A. (2000) 'The european regulation for insolvency', Konkurs. Treuhand. Sanierung. Zeitschrift für Insolvenzrecht, vol.4, pp.533-577(in German).
17. Mastromanolis, E. (2001) 'The new EU Regulation 1346/2000: From the abstracts of national legislations towards a common European Insolvency Law', Business and Company Law (Δίκαιο Επιχειρήσεων και Εταιρειών), vol.12, pp.1220-1233(in Greek).
18. Meunier, S. (2005) Trading voices. The European Union in International Commercial Negotiations, Princeton, NJ : Princeton University Press.
19. Morell, P. (1998) 'Air transport Liberalization in Europe: The progress so far', Journal of Air Transportation World wide, vol.3, no1, pp.42-60.
20. Mühlke, H. H. (1995) 'Licensing of German air companies according to the new European regulation 2407/92', Zeitschrift für Luft- und Weltraumrecht, vol.44, no 2, pp.147-167(in German).
21. Nettesheim, M. (2002) 'Air Carriers and the Freedom to Provide Services in Europe', Zeitschrift für Luft- und Weltraumrecht, vol.51, no. 2, pp. 240-270(in German).
22. Virgos, M. (1998) The 1995 European Convention on Insolvency Proceedings: an Insiders View, The Hague: Kluwer Law International.
23. Warren, E. (1993) 'Bankruptcy Policymaking in an imperfect world', Michigan Law Review, vol. 92, pp.325-340.
24. Wilmowsky, P. (2000) 'Choice of law in international insolvencies – a proposal for reform', In: Basedow, J. and Kono, T. (ed.) , Legal Aspects of

Globalisation, Conflict of Laws , Internet, Capital Markets and Insolvency in a Global Economy, (pp. 197-213), The Hague , London , Boston : Kluwer Law International.

25. Wimmer, K. (1998) 'Particularities of secondary insolvency proceedings', Zeitschrift für Wirtschaftsrecht, vol.23, pp.982-989. (in German).

Key Factors that Influence the Quality in E-Commerce

Yannis E. Tzavlopoulos

Business Excellence Laboratory, University of Macedonia, Greece, 156 Egnatia st., 54006, Thessaloniki, Greece, tzavlop@gmail.com

E-commerce more than a decade represents a new way of conducting business transactions such as buying, selling, or exchanging products through internet. Although e-commerce began with a great prosperity of reshaping the marketing domain, it hasn't the rapid development as expected to have. One of the most important and emerging issues influencing the penetration of e-commerce is quality. This paper focuses on the critical factors in B2C transactions that influencing the quality in e-commerce. The improvements of those factors may result in more effective performance, greater customer satisfaction, and more rapid exchange.

Keywords – Electronic Commerce, Quality, Internet, B2C

1. Introduction

E-commerce is getting more and more popular to customers and the ever-growing number of companies providing B2C e-commerce services contribute to this trend.

It is obvious that potential e-shoppers will buy from a certain e-shop only if they feel that they will be able to fulfill their specific needs. This should be done in a secure, quality environment. Therefore doing online business alone does not necessarily guarantee competitive advantages. E-commerce firms depend on people visiting their sites, purchasing their products and, more importantly, becoming repeat customers (Smith and Merchant, 2001).

The key element to business achievement is quality (Dale, 1999). Without a quality management approach that guarantees quality from its systems, staff and suppliers, a business will not be able to deliver the appropriate level of service quality to satisfy its customers (Cox and Dale, 2001*)*.

Numerous studies show that higher perceived web site quality and customer service lead to higher profitability levels (Hoffman *et al*., 1995, Lohse and Spiller, 1999*,* Vanitha *et al.,* 1999; Tilson *et al.,* 1998; Xia *et al*., 2003). According to past studies, it seems reasonable to assume that in the new electronic environment, perceived web site quality is one of the most important factors influencing the consumer before and after buying.

However, as e-commerce is a relatively new concept, no particular body of literature specifically addresses the concept of e-commerce web site quality (Cao *et al., 2005)*. The question is "What constitutes web site quality or what makes a web site effective?" Without answering this question we are likely to continue floundering, wasting time and money creating some sites that are winners and many others that are losers. Business decision makers need to know some guidelines for developing their e-commerce presences (Cao *et al., 2005)*.

Although existing literature on the subject includes several recent works that address online purchase quality, there is no clear distinction made between what constitutes the dimensions of process and what constitutes the dimensions of result (Alzola and Robaina, 2005). What composes web site quality and the factors that affect customers' perceptions on the acceptance of the site are unclear (Lin and Lu, 2000). Various models and methods have been developed the latest years, such as SITEQUAL measuring the perceived quality of an online shop (Yoo and Donthu, 2001), WebQual based on QFD, which helps gain understanding about which qualities are considered by the user to be the most important in any given situation (Barnes and Vidgen, 2003) and e-SQ scale, which measures the degree to which a web-site facilitates effective and efficient purchasing (Zeithaml *et al., 2001)*.

This paper conducts an exhaustive literature review on this issue in order to help identify the main aspects influencing e-commerce quality and provide a sound methodological framework for measuring and improving it. The results of this study will then be used as a basis in order to empirically investigate this issue in the Greek market.

2. Literature review

Although e-commerce is not a relatively new way of doing business, no particular literature specifically focuses the concept of e-commerce quality.

However, through last year some different models come up with a main goal to evaluate the factors that influencing e-commerce.

This paper attempts to analyze all the published models concerning e-quality. For this reason we are making a reference able following a short review to all the relevant theories and models.

Under e-commerce context, the reliability dimension of SERVQUAL can also be explained/replaced by the critical concept **of trust**. Trust is crucial in the transactional,buyer-seller relationships of customers and e-vendor because of **the risk** and uncertainty of the online environment (Reichheld and Schefter, 2000; Mayer et al.,1995). Trust is an expectation that the e-vendor will not behave opportunistically by taking advantage of the situation (Gefen et al., 2003). It is customers' beliefs that the e-vendors will behave in a reliable/dependable, ethical and socially appropriate manner (Hosmer, 1995; Zucker, 1986).

However, in electronic commerce (e-commerce) environments, the concept of CPQ is

affected by differences in the interaction between the online retailer and the consumer.

Year	Authors	Model	Dimension Measured	Factors Measured
1984	Grönroos	CPQ	Consumer perceived quality	Measures consumer's expectation of service compared with the customer's perception of the service actually received.
1985	Parasuraman	SERVQUAL	Service quality	Reliability, responsiveness, assurance, empathy and tangibles.
1989	Davis	TAM	Web-site quality	TAM predicts that user acceptance of any system is determined by two factors: (1) perceived usefulness and (2) perceived ease of use
1992	Delone and McLean	IS success	Information Systems Quality	IS success model proposed six main dimensions: system quality, information quality, use, user satisfaction, individual impact and organizational impact.
2001	Yoo and Donthu	SITEQUAL	Online Purchasing experience	Aesthetic competitive value, ease of use, design, ease of ordering, corporate and brand equity, processing speed, security, product uniqueness and product quality assurance.
2001,2002	Barnes and Vidgen	Web site quality scale	Consumer perception of online experience	The scale included: usability of web site, information quality, and interaction between consumer and web site
2002	Loicano	WebQual (TM)	Web site quality	A scale incorporating 12 items of web-site quality: information fit-to-task, interactivity, trust, visual appeal, innovativeness, flow/ emotional appeal, design appeal, intuitiveness, response time, integrated communications, business process and viable substitute.
2003	Wolfinbarger and Gilly	eTailQ	Consumer perception of e-retailing quality	
2003	Park and Kim		Factors influence consumer purchase behavior	Information quality, user interface quality and security perception play important role in online consumer purchase behavior.
2005	Zeithaml	E-S-QUAL	E-service quality	E-S-QUAL scale consists of two parts: One for routine service encounters An auxiliary scale for service errors (known E-ResS-QUAL)
2005	Lee and Lin	Revised SERVQUAL	E-service quality	They established a scale of dimensions such as web site design, reliability, responsiveness, trust and personalization.
2007	Su, Li, Song and Chen	e-commerce CPQ		The study has identified six dimensions: Outcome quality Consumer service Process controllability Ease of use Information quality Web site design

Dimensions of e-quality

	Dimension	Key features	Description
1	Performance	Use	Ease of use of the web site Ability to get an overview of the structure Ease of navigation
		Content	Accuracy of information Concise nature of the information Timeliness of the information
2	Features	Search capability FAQ Interesting Links	
3	Structure	How information is presented	Appropriate keywords or subheadings How hyperlinks are used
4	Aesthetics	Appearance of the web site	Color combinations Type and size of fonts The animation The sound effects The clarity and readability of texts
5	Reliability	Consistent performance	Up to date information Accurate information
		Availability of web site	Accessibility Speed Quickly download
6	Storage capability	Capability to store information and make it easily available to customers	
7	Serviceability	Conflicts and complaints from customers are resolved	
8	Security and system integrity	Web site ability to safeguard and protect information that is provided to it	
9	Trust	The degree of confidence customers have in online exchanges	
10	Responsiveness	Responding to customer needs (e.g. through e-mail) Flexibility in policies (e.g. cancelations, refunds etc)	
11	Product/ Service differentiation and customization	Convenience Timely delivery Not attainable features	
12	Web store policies	Customer oriented policies	Warranties Refunding
13	Reputation	Past experiences Perceived site performance	
14	Assurance	Provide impeccable service	
15	Empathy	Individualized response to customer requests	

Apart from the absence of the ambience of a physical shop (such as temperature, lighting, and business equipment), the online retail environment also lacks person-to-person contact. This means that interaction does not include the attitude and expertise of a salesperson, which can have an important influence on CPQ (Crosby et al., 1990; Shamdasan and Balakrishan, 2000).

Some authors focus on aspects of web design, whereas others measure quality issues along the entire online shopping process – including information searching, online ordering, delivery and payment, and after-sales service. In general, studies in this realm can be categorized into two main streams:

(1) those that emphasize e-transactions; and

(2) those that focus on e-service.

From the exhaustive literature review we came up with the **five most commonly used e-quality** dimensions:

(1) ease of use;

(2) web site design;

(3) customization;

(4) responsiveness; and

(5) assurance.

These dimensions will be briefly discussed. **Ease of use** is an essential element of consumer usage of computer technologies (Davis, 1989; Morris and Turner, 2001; Venkatesh, 2000; Venkatesh and Davis, 2000), and is of particular importance for new users (Gefen and Straub, 2000). Ease of use is a determinant of service quality (Dabholkar, 1996) and is decisive for customer satisfaction, since it enhances the

efficiency of using the service (Xue and Harker, 2002).

In an e-tailing context, ease of use includes aspects such as functionality, accessibility of information, ease of ordering and navigation (Reibstein, 2002). Besides being easy to use, the company's site should be pleasing to the eye.

Thus, another quality dimension directly related to the user interface is **web site design** (Wolfinbarger and Gilly, 2003; Zeithaml et al., 2000), or **e-scape** (Van Riel et al., 2004). An often-cited benefit of online technologies is that the web site can be **personalized** to the user's needs, although this may be a challenging task, because of the lack of a human touch (Rust and Kannan, 2002). E-tailers should strive to customize their services to users' individual needs (Srinivasan et al., 2002), e.g. based on past purchases and other information provided by customers. Loyal customers can be a valuable source for service improvements, but companies often ignore such information (Finkelstein, 2003). As in a traditional service context, customers expect quick feedback on requests and when they suggest improvements.

Though r**esponsiveness** in general has a positive influence on e-satisfaction, it should

be noted that it may impact quality perceptions negatively if customers feel that they arebombarded with company e-mails (Zeithaml et al.,2000).

The fifth quality dimension is **assurance**, i.e. the customer's perceived security and privacy when using the e-tailer's services. Security and privacy are of serious concern to e-service customers (Rust and Kannan, 2002). Security concerns the risk of third parties obtaining critical information about the customer (e.g. access to credit card or bank account details), whereas privacy relates to the concern about the potential

misuse of personal information by marketers (Milne and Rohm, 2000). Privacy exists when customers can restrict the use of personal information. However, many customers are not aware of what information e-tailers collect, or where to look for opt-in or opt-out options. Milne and Rohm (2000), for example, found that less

than half of surveyed direct mail responders knew how to remove their name from the mailing list.

3. Limitations

The paper is based on the literature review and there was no research contacted in order to justify some of the data presented. If a company manages to improve its e-commerce quality, this will have a favorable effect on transaction results. In fact, the results may enhance the way that companies doing e-business.

4. Conclusions

Most relevant studies about quality in the internet have focused on web design aspects. Moreover, the existing literature regarding internet consumer behavior has not fully analyzed profits generated by higher perceived quality in terms of user satisfaction and loyalty.

By examining thoroughly all the possible factors influencing the quality in e-commerce we set up a framework and some practical guidelines for identifying and improving quality in e-commerce by changing some special factors.

Furthermore, the results of this study can be used as a basis in order to empirically investigate this issue in the Greek market.

References

1. Alzola L. and Robaina V. (2007) *"Measuring the results in B2C e-commerce"*, International Journal of Quality & Reliability Management, Vol. 24 No. 3, pp. 279-293.
2. Barnes, S.J. and Vidgen, R.T. (2002), *"An integrative approach to the assessment of e-commerce quality",* Journal of Electronic Commerce Research, Vol. 3, No. 3, pp. 114-27.
3. Dale, B.G. (Ed.) (1999), *"Managing Quality"*, 3rd ed.,Basil Blackwell, Oxford.

4. Davis, F.D. (1989), "Perceived usefulness, perceived ease of use, and user acceptance of information technology", MIS Quarterly, Vol. 13 No. 3, pp. 319-40.
5. Dabholkar, P.A. (1996), "Consumer evaluations of new technology-based self-service options: an investigation of alternative models", International Journal of Research in Marketing, Vol. 13 No. 1, pp. 29-51.
6. Hoffman, D.L., Novak, T. and Chatterjee, D. (1995), *"Commercial scenarios for the web:opportunities and challenges"*, Journal of Computer Mediated Communication, Vol. 1 No. 3,pp. 23-45.
7. Cao M., Zhang Q. and Seydel J. (2005), *"B2C e-commerce web site quality: an empirical examination"*, Industrial Management & Data Systems, Vol. 105, No. 5, 2005, pp. 645-661.
8. Cox J. and Dale B.G. (2001), *"Service quality and e-commerce: an exploratory analysis"*, Managing Service Quality, Vol. 11, No. 2, pp. 121-131.
9. Finkelstein, S. (2003), "Why Smart Executives Fail and What You Can Learn From Their Mistakes", Portfolio, New York, NY.
10. Lin, J.C.C. and Lu, H. (2000), *"Towards an understanding of the behavioral intention to use a web site"*, International Journal of Information Management, Vol. 20 No. 3, pp. 197-208.
11. Lohse, G.L. and Spiller, P. (1999), *"Internet retail store design: how the user interface influences traffic and sales"*, Journal of Computer Mediated Communication, Vol. 5 No. 2.
12. Milne, G.R. and Rohm, A.J. (2000), "Consumer privacy and name removal across direct marketing channels: exploring opt-in and opt-out alternatives", Journal of Public Policy and Marketing, Vol. 19 No. 2, pp. 238-49.
13. Morris, M.G. and Turner, J.M. (2001), "Assessing users'subjective quality of experience with the World Wide Web: an exploratory examination of temporal changes in technology acceptance", International Journal of Human-Computer Studies, Vol. 54 No. 6, pp. 877-901.
14. Gefen, D. and Straub, D.W. (2000), *"The relative importance of perceived ease of use in IS adoption: a study of e-commerce adoption"*, Journal of the Association for Information Systems, Vol. 1 No. 8, pp. 1-28.
15. Rust, R.T. and Kannan, P.K. (2002), *"The era of e-service"*, in Kannan, P.K. (Ed.), e-Service. New Directions in Theory and Practice, M.E. Sharpe, New York, NY.
16. Reibstein, D.J. (2002), *"What attracts customers to online stores, and what keeps them coming back?"*, Journal of the Academy of Marketing Science, Vol. 30 No. 4, pp. 465-73.
17. Srinivasan, S.S., Anderson, R.E. and Ponnavolu, K. (2002), *"Customer loyalty in e-commerce: an exploration of its antecedents and consequences"*, Journal of Retailing, Vol. 78 No. 1, pp. 41-51.
18. Smith, B.A and Merchant, E.J. (2001), *"Designing an attractive web site: variables of importance"*, Proceedings of the 32nd Annual Conference of the Decision Sciences Institute, San Francisco, CA.
19. Tilson, R., Dong, J., Martin, S. and Kieke, E. (1998), *"Factors and principles affecting the usability of four e-commerce sites"*, Proceedings of 4th Conference on Human Factors and the Web (CHFW), AT&T Labs, Florham

Park, NJ, April, available at: www.research.att.com/conf/hfweb/ proceedings/tilson/
20. Van Riel, A.C.R., Lemmink, J., Streukens, S. and Liljander, V. (2004), *"Boost customer loyalty with online support: the case of mobile telecoms providers"*, International Journal of Internet Marketing and Advertising, Vol. 1 No. 1, pp. 4-23.
21. Vanitha, S., Lepkowska, E. and Rao, B.P. (1999), *"Browsers or buyers in cyberspace? An investigation of factors influencing electronic exchange"*, Journal of Computer Mediated Communication, Vol. 5 No. 2.
22. Xia, Y., Ahmed, Z.U., Ghingold, M., Boon, G.S., Mei, T.S. and Hwa, L.L. (2003), *"Consumer preferences for commercial web site design: an Asia-Pacific perspective"*, Journal of Consumer Marketing, Vol. 20 No. 1, pp. 10-27.
23. Xue, M. and Harker, P.T. (2002), *"Customer efficiency: concept and its impact on e-business management"*, Journal ofService Research, Vol. 4 No. 4, pp. 253-67.
24. Yoo, B. and Donthu, N. (2001), *"Developing a scale to measure the perceived service quality of internet shopping sites (sitequal)"*, Quarterly Journal of Electronic Commerce, Vol. 2 No. 1, pp. 31-47.
25. Zeithaml, V.A., Parasuraman, A. and Malhotra, A. (2001), *"A conceptual framework for understanding e-service quality: implications for future research and managerial practice"*, working paper, report no. 00-115, Marketing Science Institute, Cambridge, MA.
26. Venkatesh, V. (2000), "Determinants of perceived ease of use: integrating control, intrinsic motivation, and emotion into the technology acceptance model", Information Systems, Vol. 11 No. 4, pp. 342-65.
27. Venkatesh, V. and Davis, F.D. (2000), "A theoretical extension of the technology acceptance model: four longitudinal field studies", Management Science, Vol. 46 No. 2, pp. 186-204.

The History Of Financial Crises Of Turkish Banking System: Solution Suggestions For 2008 Global Financial Crises

Okyay UÇAN[1], Özlem YARDIM[2] and F. Özlem ÖZKAN[3]

Faculty of Business And Administrative Sciences, University Of Çukurova / Turkey,
[1]oucan@cu.edu.tr , [2] ozlem.yardim@nevsehir.edu.tr, [3]oozkan@cu.edu.tr

There is a widespread conviction that financial crisis are rare events. But after the financial liberalization movements that started in 1980s, especially 1990s witnessed financial crisis with large impacts. The fact that the banking crisis and problems occupy the primary ranks in the economic agenda has two main reasons. First one is the serious consequences of crisis on local economies. Second one is the dispersion of crisis to other countries as the financial markets become more and more integrated. The aim of this study is to analyze the two financial crises that Turkey encountered in the months of November of 2000 and February of 2001 and in the light of the measures taken after these crises, to make some suggestions about the measures that must be taken in the Turkish banking sector. With this aim, after a short definition of financial crisis, the development of Turkish banking system will be analyzed briefly. How Turkey was dragged into crisis environment and how it came through these crises will be explored. Lastly, in the light of the previous experiences with crises, some information is going to be given about the measures that must be taken for global crisis of 2008.

Keywords. Financial Crisis, Turkish Banking System, Banking Crises in Turkey, Effect of Financial Crisis

1. Introduction

In the 1980s financial deregulation movements have caused to raise financial crisis transitivity. Consequently during the 1990s there has been variety of crisis which influences other countries at different regions in the world.

The crisis term that is derived from Latin is used to be synonym with depression and decline. It means current solutions against any problem become useless (Delice, 2003, p.59).

Economic crisis may be defined as a fluctuation out of a band in quantities and prices in the foreign exchange market good and services and factor markets (Kibritçioğlu, 2000). Economic crisis may originate in both cyclical reasons as supply surplus or demand shrinking in the real and financial sectors, and internal factors of the economic structure itself. However, reason of the economic crisis may not be an "economic issue". For instance, act of god in a country or an epidemic may cause the economic crisis. Moreover, rapid changes in the political, technological and ecologic fields are the reasons for economic crisis. In addition to those; government interference and abrupt cyclical movements in the price level, production and employment may cause crisis.

Because of the expectations about country's risky condition in the financial field, investors attempt to make speculative attacks so financial crisis starts and freshen due to the intensity of this attacks. As expectations worsen, panic starts and size of the crisis deepen. Financial crisis that affects banking system not only is the cause of inefficient and unproductive working of banks but also corrupts corporate structure. Consequently, following the problem encountered in the banking system, real economic indicators become worse. Thus, by spillover effect both financial crisis and real economic problems result in deeper crisis in the international arena. By means of this bidirectional effect in the literature, the studies about financial crisis become widespread.

About the causes of the financial crisis there are two fundamental approaches. They are first generation and second generation crisis models. Especially, in the 1980's first generation models were developed to explain the crisis in the Latin America countries. These theories emphasize the inconsistency between national economic policy and exchange rates (Krugman, 1979). However, in the beginning of 1990's, traditional theory remained incapable to explain the crisis of European Money System. So, second generation models were used to explain this problem. The theoreticians asserted that negative expectations of economic actors about sustainability of fixed exchange rate in the long run caused the crisis (Obstfeld, 1996). Unlike the other models, theories that explain the Southeast Asia crisis, emphasized the governmental guarantee for the firms' and banks' deficit (Krugman, 1997).

Lots of studies have been done about the causes of the financial crisis. Eichengreen and Bordo (1999) used data from 21 developed and developing countries between 1975 and 1997. They found that fundamental reason of the crisis was banking system. In the mentioned period, especially because of the bank panic, fragility of banking system raised.

Kregel (2000) asserted that the causes of the 1997 Southeast Asia crisis and the crisis that had been in the Latin American in the 1980's and 1990's were different each other. According to Kregel, fundamental reasons of Latin America crisis were excessive budget deficit and balance of payment problems. He found that Southeast Asia crisis resulted from short run speculative financial movements.

2. How Does The Banking Crisis Happen?

According to monetary theory, excessive debt accumulation during the boomer of firms and countries is on the basis of financial crisis. Correspondingly with this theory, Fisher explains the downward cyclical movements with two fundamental factors. These are deflation and overindeptedness. As a result of an external shock in one of the key sector of the economy, a cyclical upturn is seen. This external shock promotes new investments through price and output rise. While some process increasing the profits, speculation follows this development. In fact this process is fully funded by bank credits so following by deposits and money supply increase (by the help of money multiplier), prices increase. Consequently, turnover of money rise and strengthen the expansion. Increasing of prices compensate the increasing of nominal indebtedness so it reduces the real value of current dept, and encourages more loan. This process continues up to a general indebtedness case (Fisher, 1983).

Mishkin (2001) says that the reasons of the financial crisis are:

- Balance sheet corruption of the firms in the financial sector,
- Increase of interest rate,
- Uncertainty.

Financial crisis is divided into four types as twin crisis, banking crisis, currency crisis, systematic financial crisis and foreign dept crisis (IMF, 1998, p.74).

For any reason whatsoever, as a result of lacking confidence to financial system, savers charge the banks to draw back their deposits. This is the banking crisis case (Seyidoğlu, 2001, p.583).

These attacks to banks, however, are not the cause of the problem but they are the results. The reasons of the banking crisis are unstable macroeconomic structure, inadequacy of legal reforms, absence of active control and supervision mechanism in the banking system (Erdönmez, 2001).

A banking crisis is related to obligatory governmental intervention as large scale support to perform the liabilities of banks that are in bankrupt condition. Banking crisis usually arises from poor quality assets and big price fluctuations of equities and immovable. Moreover, increasing rate of non-performing loan in the credit portfolio of the banks is also a problem. Comparing developing and developed country economies it is seen that developed countries face less currency and banking crisis in the same period (Tapan, 2002, p.701).

In view of the model done by Demirgüç-Kunt and Detragiache (1998); macroeconomic structure has a big importance for banking crisis. In addition to this, most important components that create risk in the banking system are high interest rate and inflation.

These are two main reasons for banking crisis in the developing countries (Aydın and others, 2000, p.4).

The banks carry on a business in an economy which has high mobility in macroeconomic fluctuations and important political regime variations.

There are political interventions to banks.

By spillover effect banking crisis spreads to other sectors. So problem becomes to affect macroeconomic stabilization. Therefore, cost of the crisis is not only a burden to budget for recovering the banking system but also large size effect in the real sense. As it was seen in the last Asia financial crisis, banking and currency crisis arised as twin crisis in the emerging markets due to financial deregulation through international capital flows. This twin crisis affect the world economy following by the effected country (Coşkun, 2003).

3. History of Turkish Banking System

Financial change period in Turkey starts with an austerity package for the economy at 24 January 1980 for being a solution to the crisis that lived in Turkey, following the conjuncture lived in the world during 1970's. There are two important purpose of this package. First one is, reestablishing the internal and external balance, increasing the production and growth due to export, and increasing the currency reserves (Balkan and Yeldan, 2001). Second one is providing to restructure the Turkish economy by transferring wealth to the market and liberalization (Demirgüç and Detragiache, 1998).

The financial market is aimed with the process of transformation of Turkish economy starting with 24th January decree. In accordance with this target, several methods are adopted. Such as financial deregulation, increasing the competitiveness in the financial market and simplifying to enter the sector. Letting banking sector to foreign competition plays important role, too.

Together with these reforms, banking activity is diversified and classical banking system become stronger. But actual locomotive effect happens after 1982 banker crisis. This event gives rise to trust banks and banking system gain political power (Köne, 2003).

In this decree, ceiling application on deposit and credit interest rate was cancelled and deposit certificate application started. At the same time, a tight monetary policy was used to control inflation. As a result, developments in the financial market brought about a huge crisis in 1982 and financial deregulation process was partially set back. Government closed five small bank and assets and liabilities of these banks were undertaken by public banks (Boratav, 2005, p.151). Because of this reason, second phase of liberalization, 1983-1987 periods, was carried out in a more controlled way.

Financial arrangements in this period laid the groundwork for banks to dominate the markets. One of the important examples of this, the banks weren't blocked to have operations in both primary and secondary market. On the contrary, banks were let to be monopolist in some fields such as establishing the mutual funds.

After 1990's, Turkish banking system and financial firms have gone away from production and manufacturing activities. Real sector was ruptured from real sector following financial sector. Thus an important part of the profits of real sector was replaced by non-operating profit (Yenitürk, 2003).

In 1993, just after this accumulation, there has been important fluctuation in the monetary market. After the 1994 crisis, 5 April economic prevention package's aim was reinstating the stable economic ambiance. With the agreed decisions, three banks have been taken to deposit guarantee fund. For having passed the panic of crisis and preventing bank smashes, all deposits were guaranteed. This protected deposits gave rise some unwanted results. Moral risk was the most important one. While investing money, people didn't discriminate banks due to deposits protected by government. However, people chose the bank according to their interest rate and more comfortable facilities. Because of bad controlling system, lack of rivalry by protected deposits and legal spaces; liabilities of banks was transferred to government (Kumcu, 2009).

In 1992 and 1993, large amount of capital inflow comes to Turkey. After 1994 Mexico crisis this global capital flow decreased, but the effects of this crisis passed rapidly. Moreover, global hot money flow increased in 1996. By the help of this flow, sum of 13 billion dollar global capital flew to the Turkey that offered high interest rate occurred in 1997 and 1998 (Somçağ, 2006).

In this period, Turkish banking system had totally different structure due to high interest rates. Bond investment became the most profitable item among the all banking operations. In addition to protected deposits that caused corruption, new bank licenses were given. Moreover, Russia and Asia crisis gave raise the current problems for Turkey.

Especially, after 1994 crisis, public sector borrowing increase in demand caused to increase real interest rate. So, this increased the public sector borrowing requirement. Thus, Turkish banking system tended to finance public deficit instead of providing fund to real economy (Akgüç, 2001).

Early in the 2000, loan talks with IMF were started and details of the programme putted into practice were determined. As a prerequisite to programme, five banks were arrested. Government took the possession of these banks and established a retroactive treasury security tax. Applied programme started to give crisis oriented signals in the fall of 2001. In political and economic instability and anarchy environment weak financial structured banks couldn't carry the system's burden and bank failures started. When Demirbank was taken the possession by government in December, panic increased. Time of the overcoming of the crisis was at the end of 2001 (Kumcu, 2009).

In the light of the above regulations, Turkish banking system has the current structure. As of 2009, 45 banks operate in the system. Three of them are public bank; eleven of them are foreign banks founded in Turkey; eleven of them are private equity; one of them is alienated to saving deposit insurance fund (SDIF); six of them are foreign banks that open branches; three of them are public investment and development banks; six of them are private investment and developments banks and at the end, four of them are foreign investment and development banks (TBB,2009).

4. Evaluation of Turkish Banking System before and After the November 2000 Crisis

Between 1990 and 1999, Turkish banking system was affected from economic progress and decisions taken in this field. In this period, there wasn't any controlling system for banks that lost main functions. The banks undertook only to finance the public and then system went away from the real economy (Toprak, 2001).

In the same period, banking sector preserved to dominate the market. Financial sector is very important for a country's economic structure. Consequently, Turkish banking system often exposed to crisis causes to affect the Turkish economy negatively.

4.1. Advantages and Disadvantages of Turkish Banking System before Crisis

Before the crisis period, the advantages of the Turkish banking sector can be listed like that (Güneş, 1999):

- Positive growth rate of banking system in spite of its small size.
- There was a big investment trend in terms of technology in the banking sector.
- Number of staff and quality of education increased. Before crisis, number of staff was 166.492 in 1998 and 173.988 in 1999. Despite crisis, at the end of the 2000 this number becomes 171.732. This staff increase was interpreted as an indicator of growth of the system.
- Consumer credits became larger and credit card sector progressed rapidly.
- Existence of 80 banks was seen as an important advantage of competitiveness despite of concentration in financial system.
- Albeit the crisis had begun recently, by the end of 2000 number of branches increased. This case was taken into account as a positive trend.

Before the crisis period, disadvantages of the Turkish banking sector can be listed like that (Kandemir, 2006):

- Risk structure of the banking system was in a bad condition. Especially risk of exchange rate was very high.
- Lacking of equity capital in banking sector.
- The owners of the banks were out of the banking system.
- Although there were aggregate corporations apparently, banking system was hold by several hands.

Principle properties of Turkish banking system in the same period (Tunay, 1997, 59-60; Sucu, 2005, 40-44):

- Deposit banking is predominantly in the system.
- There are multiple firms structuring.

- There exists insufficient equity problem.
- Poor asset quality problem.
- Inadequate internal control, risk management and corporate governance.
- Over sensitiveness to market risk.
- Public banks dominate the system.
- Group banking concepts becomes more pronounced in the private banks.
- Oligopolistic market features are observed.

In 2000, negative evaluation was done about existing practice since there were over-valued exchange rates, current account deficit and increased number of open positions in the banking sector.

Most annoying fields of structural adjustment process following by stand-by agreement with IMF were restructuring the banking sector. Before agreement, five banks were alienated to SDIF but this was not only inadequate to strengthen the banking sector but also understood that structure of the system was weak by abroad. With these structural problems, sector maintained foreign borrowing up to November of 2000 crisis. This process went on up to February of 2001 crisis and started to decrease by abandoning the fixed exchange rate (Tunalı, 2007).

Further more, it was understood that some of the rules for auditing the currency and delay mismatches were not applied. The most important example of this, value of open positions exceeded the limit by 2.1 billion dollars at the end of 2000 (Uygur, 2001).

Bayındırbank's Bank BTR (Banco Turco Romana) has gone to bankrupt in Romania in the September of 2000. Since Bayındırbank hadn't got enough fund to save BTR, a public bank, Vakıfbank, provided this fund to BTR. Although this event wasn't regarded as important for the risks of banking system in Turkey, its starting style and backward linkages were considered critic in the abroad. For example, R. Dornbusch defined this event as a first signal of the crisis that would started in Turkish banking system. In the same period, Etibank alienated to SDIF and then economic insecurity fairly increased. Demirbank has lost its credibility and couldn't find any credit neither inside nor outside because it was considered that Demirbank had some share of Etibank (Uygur, 2001).

Above, we summarize the development of the crisis and indicators of it before 2000 depression. However, according to the crisis theories, there are more concrete indicators. This may be useful to evaluate the November 2000 crisis by using these indicators (Uygur, 2002, p.15):

Short run interest rate reflected the ambiguity feature of monetary policy. This rate significantly increased in June, August and September 2000.

Short run foreign dept / currency reserve rate increased during the course of the year.

Current account deficit / currency reserve and current account deficit / GDP rates had continuous increasing movements.

As a result of financial measures and agreement with IMF, fluctuations in the financial markets were steadied, currency reserves of central bank of Turkey increased and interest rates decreased as to crisis environment. It was showed a maximum effort to participate the private sector for solving the problem. In this particular, meetings were

arranged with foreign bank authorized officers not to draw back the funds that had been borrowed by domestic banks. The guarantees, offered by Turkish government about liability of banking system, enabled the agreement, too.

5. February 2001 Crisis and the Results of November 2000 Crisis

After the crisis in November 2000, government announced to sustain the current monetary policy along the 2000 and 2001. There was only one exception of the programme. New targets were considered for the net domestic assets and net international reserves. With the applied programme, Turkey went away from the crisis environment in the January of 2001. Central bank currency reserves were above 25 billion dollar at 5th January. Overnight interest rate, which was 199 % in late December 2000, decreased to 42 % in January (Güneş, 2004).

Because of political controversy between prime minister and president of the republic, all these constructive developments finished. Markets, already sensitive due to November crisis, toppled into a severe currency recession. After people attacked to dollar, the pressure on exchange rate increased. So central bank intervened the market at the risk of disinvesting the reserves. But, then, it had to announce to pass the floating exchange rate system in spite of exchange rate anchor (fixed exchange rate). Currency reserves were 27 billion 943 million dollar at 16th of February. This number decreased to 22 billion 581 million dollar at 19th of February. So Central Bank's currency reserve lost in one day was above the 5 billion dollar (17 and 18th February was the weekend). Under floating exchange rate, dollar appreciated from 688.000 TL to 962.000 TL. High interest rates were used to stop the currency demand. So, overnight interbank interest rate increased to the level of 6200 %, and then it decreased to 103 % at the end of the February (Güneş, 2004).

Fredric Mishkin (2001) clarified the process of the crisis in accordance with moral risk and adverse selection. Experiences lived in Turkish banking system before and during the crisis were best fit do this definition. High moral risk has existed in the system until 2001. Banks started to operate in risky and catchy activities together with increasing probabilities of crisis. Because of these activities, banks' balance structure was failed and bankrupts happened. All these events are parallel to the Mishkin's forecast (Bastı, 2006).

In fact both November and February crisis were resulted from week banking system and over assurance to hot money. Fundamental weakness of banking sector can be listed like that (Sucu, 2005, p.53):

- Structural problems of banks increased and the administration that was far away from active control and banking principles, become widespread.
- Shortening terms to maturity for obtaining dept increased the currency dept and caused asset-liability conflict.
- Because of the increased consolidated budget deficit, banks provided loan only to the public sector despite real sector.
- Non-performing loan volume increased due to the reliability problem.

After the February 2001 crisis as named the deepest recession up to that date, reconstructing the banking sector was applied BRSA (Banking Regulation and Supervision Agency). In fact this programme was arranged in the late 1999. However, it couldn't have been handled comprehensively until 2001 (BDDK, 2001).

Agreed programme in 1999 was overviewed after the recession in 2001. It was known that structural reforms were the corner stone for the success of the programme. So, structural reforms were carried out as soon as possible. Thus, government went through the proper channels to provide the equilibrium of assets and liabilities of the system (Serdengeçti, 2002). Inflation problem could be dealt in the late 2002. This programme was into four major components (Ulusal Program, 2003), (BDDK, 2002):

- Reconstructing the public banks according to their financial and operational structure.
- Closing a deal with problematic banks in SDIF.
- Providing good condition to the private banks affected by the crisis.
- Establishing legal and constitutional regulations to make the sector more efficient and competitive by increasing the regulation and supervision activities.

This regulation in banking field is a radical change for Turkish banking system. This change is accepted as an indicator for lasting the corruption (Duman, 2002).

Relatively crisis years, capital structures of banks are stronger, qualities of assets are better, profitability increases and financial risk decreases by reconstruction (Aktaş, 2006).

After the reconstruction activities, we get these results within the brokerage and banking performance of Turkish banking system (Selçuk, 2006):

- Having success in strengthening the financial structure and capital adequacy is observed. However, banking and brokerage services of the system can not be developed relative to the ex-crisis.
- Especially, public banks'capital adequacy rates show a progress.
- Liquidation of banks in the SDIF and then making stronger their financial structure and reselling them prevent their unwanted effects to the system.
- Capital strengthening studies in the private banks and incentive for foreign constitution to enter the sector provide to increase the capital adequacy rate above the 3 times bigger than EU countries.

6. 2008 Global Crisis and Turkish Banking System and Economy

This global crisis is accepted to be the most destructive one since great depression in 1920. Fundamental point of the global crisis is the nonperforming loans of big investment banks and following this, bank failures. These big investment banks'share certificates are called "toxicant (toxic substance)" and investors who took up the shares felt the effects of crisis. Swiss and England banks which made investment that Yankee bonds deeply felt the first impression. Then Japan followed

them. This means, at the preliminary stage financial crisis outspreaded by the help of this "toxicant" (Saral, 2009).

When crisis begun in late 2006, it was accepted as a temporary shocks. But these optimistic thought gave way to pessimism. When we come to the year 2009, crisis depth, effects to country's economies and precautions are the favorite matters for discussion. Moreover, government intervention and liberalization are coming up against because of this global crisis. So, economic system debates are flamed.

Turkish banking system wasn't affected directly from this global crisis because there wasn't any investment banking and commercial banks didn't make investment that Yankee bonds. However, it was heard that Lehman Brothers made sales to individuals in Turkey. Indirect effects of the global crisis were felt in Turkey in the late 2007.

Although this global crisis origin was financial, real sector in Turkey had a recession period before the financial sector in Turkey. First indicator of this was seen in growth rate. GDP growth rate decreased to -4.5% in 2007. Following a 7.3% decline in agricultural sector, manufacturing industry and building sector declined 5.4% and 5% respectively. While these were happening in Turkey, real sector in the world hadn't got any big problem yet. USA, England and Japan's growth rate was 2% and growth rate for European countries was %2.6. Overall it was 3.7% in the world. USA's three automotive firms'–General Motor, Ford, and Chrysler- problems came into existence in 2008. That is, financial crisis didn't affect real sector instantly. In the same period, developing countries like China (11.9 %), Russia (8.1 %), Argentina (8.7 %), Indonesia (6.3 %) and India (9%) maintained high growth rate by "abundant fund, cheap credit" after 2002. thus, this was the evidence that real sector wasn't directly affected by financial sector. Initially, as it is seen in numbers, none of the countries lost big growth rate like Turkey. This different spillover effects to Turkey caused people suspicion that Turkey had an individual recession independently of USA financial crisis (TEK, 2009).

The inflationist tendencies increased in all over the world in 2008. Like other developing countries, inflation in Turkey also increased. This high inflation resulted from the increased prices of petrol and raw materials (Alantar, 2009).

Between 1980 and 2008, Turkish manufacturing sector had a high foreign dept burden. This dept wasn't dealing with financial part of banking or outside banking (Alantar, 2009).

Financial reforms done after the 2001 crisis strengthened the structure of banking system. In addition, there were no toxicant asset and complex derivative products in the financial sector of Turkey. Turkish banking sector capital adequacy ratio (CAR) was %17.4. So this ratio seemed to be good enough. In Turkey, mostly credit channel was affected. Turkish economy grew by capital inflow and became smaller by capital outflow (Unakıtan, 2008).

Although crisis in Turkey was seen as a real sector crisis and banking system was in a good condition, growing current account deficit gave harm to system. Moreover, Turkey had to give high interest to finance its growth. Interest rate in Turkey was 17-18 % while it is 1.5 % in USA, 4.5 % in England and 3.75 % in EU. While other countries were cutting their interest rates, there was no change in the interest rate in

Turkey. The most important reason for that was the need of foreign capital (Yörükoğlu, 2009).

Between the years 2007 and 2008, emerging developments in Turkish banking system is summarized like that (BDDK, 2008):

- In the late 2008, size of assets in Turkish finance sector grew by %23.1 and reached at a level of 947.8 billion TL.
- Number of banks didn't changed but branches numbers increased by 1187 and staff numbers increased by 14907.
- Interest rate of deposits had an increasing trend in the last two year. This was a negative indicator within the financial intermediation function.
- In the end of 2008, foreign exchange net general position was at zero level and this denoted that exchange rate risk is low. Off balance sheet transactions over total assets was starting to decrease to the level of 2007.
- In 2008, total assets of banking sector increased by 26 % and reached to 733 billion TL according to the last year. The largest contribution was increased credits by 29 %. The deposits also increased by 35 % and maintained its continuity at a level of 70 % out of total foreign resource. According to the total resources, however, shortness of maturity date continued.
- Since banks' total resources increased by 20.6 % according to the previous year, this helped them to maintain their financial structure and to assist the real sector.
- Credits over deposits ratio maintained its same level as 2007. So, mediation function of the banks continued at an equal level.
- In 2008, Consumer loans increased by 23.7 % according to the year 2007. According to 2007, consumption loans and credit card usage, also, increased by 22.6% and 26.3 % respectively.
- Whether deposit weighted resource structure of the banking system continued, deposits formed the 62% of total resources in the late 2008.
- In the late 2008, total off –balance sheet transactions increased by 23.3 % and reached to the level of 476 billion TL. Off –balance sheet transactions composed of the derivative transactions with a level of 41%. This category also increased by 22.4 %.
- Paralleling with the felt effects of global crisis in Turkey, it was observed that credit risk of banking sector rose.
- Turkish banking system' asset and resources profitability ratios were 2% and 16.8 % in the late 2008.

7. Result and Suggestions

It isn't needed any cash precautionary package to the Turkish banking sector after crisis. There is an important reason for that. After the 2001 crisis, Turkey applies a package like that at the cost of a big burden to the capital budget.

Conversion rate of deposits to credit is 80 % and capital adequacy ratio of banking sector is above the legal limits. This limits the effects of crisis to Turkish

banking sector. The scenario analysis made by CBRT (Central Bank of the Republic of Turkey) approximates (10%) the capital adequacy ratio above the 8 % of legal limit. Taking precautions for the financial sector is the fundamental duty of the CBRT. In this point of view, in addition to interest rate decision, some precautions are taken not to face any liquidity shortage. Eventually, additional regulations are done within the currency liquidity. Within this framework, lending interest rates are decreased by extending maturity of currencies that are bought by central bank in terms of USA dollar and euro. CBRT continues to undertake these duties (Yörükoğlu, 2009).

Some of the precautions can be listed like that (Saral, 2009):

- The cost of the credit and lending has to be reduced. In this context, in addition to the interest rate reducing decision of central bank; banking and insurance transactions tax has to be abrogated or reduced at least in some credit types.
- Deposits with letter of credit (DWLC) / worker currencies account has to be used efficiently through the channel of public banks. Moreover, adequate interest rate has to be given to these accounts.
- The cheque law has to be became function. So, bottleneck of goods and service will be surpassed.
- A solution has to be found for the debt of credit cards. An upper limit must be for interest rates.
- The deposit protection has to be taken up to a possible limit for a length of time. To prevent capital outflow, this must be done with high interest rates.
- The banking usage has to become widespread.
- It must be given importance to investment banks.
- Dense capital outflows happen in the crisis periods. Because of this reason, especially developing countries face to liquidity shortage and impel the financial sector to a bottleneck. The banks that have powerful financial structure and capital adequacy will be effected less in the crisis. Therefore, financial risk management has to be effective (BDDK, 2007, p.79).

References

1. Akgüç, Ö., (2001), 'Bankacılık Kesimi Kriz Nedeni mi?*' İktisat Dergisi*, Vol:410-411, pp: 32-33.
2. Aktaş, R., (2006), Bankacılık Sektöründe Yeniden Yapılandırma: Ticari Bankalarda Performans Yönelişleri; Türkiye 2001-2004 Örneği Phd. Thesis, Celal Bayar University, Manisa, Turkey.
3. Alantar, D., (2009), 'Küresel Finansal Kriz ve Türkiye'nin Makroekonomik Göstergelerine Etkileri', *Bütçe Dünyası Dergisi*, Vol:31 pp: 4-12.
4. Aydın, A., İnan, A., Tulay, B., Ataman, P., (2000), 'Bankacılıkta Mali Bünye Sorunları ve Yeniden Yapılandırmada Ülke Uygulamaları', *Bankacılar Dergisi*, Vol:32, pp:37-80.
5. Balkan, E. ve Yeldan, E., (2002), 'Peripheral Development under Financial Liberalization: the Turkish Experience', *In:* N. Balkan and S. Savran (ed),

The Ravages of Neo-Liberalism: Economy, Society and Gender in Turkey, (pp: 2–11), New York: Nova Science.
6. Bastı, E., (2006), Kriz Teorileri Çerçevesinde 2001 Türkiye Finansal Krizi, Ankara: Sermaye Piyasası Kurulu.
7. Batu, T. , Uzuner, M., T. and Yiğit, A., (1997), *Türkiye'de Kamu Bankacılığı ve Sektör Üzerine Etkileri*, Ankara: Ekonomik Araştırmalar Merkezi.
8. BDDK (Bankacılık Düzenleme ve Denetleme Kurumu), (2008), *Finansal Piyasalar Raporu*, Aralık, Ankara.
9. Bortav, K., (2001), '2000-2001 Krizlerinde Sermaye Hareketleri', *İktisat, İşletme ve Finans Dergisi,* Vol:196, pp:7-17.
10. Boratav, K., (2005), *Türkiye İktisat Tarihi 1908-2005*, Ankara: İmge.
11. Coşkun, N., (2003), 'Gelismekte Olan Ülkelerde Bankacılık Krizleri', Obtained through the Internet: http://dergi.iibf.gazi.edu.tr/pdf/3204.lidf8 [accessed 02/04/2009]
12. Delice G., (2003), 'Finansal Krizler: Teorik Ve Tarihsel Bir Perspektif", *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, Vol: 20, pp:57-81.
13. Demirgüç, K., A., and Detragiache, E., (1998), 'The Determinants of Banking Crises in Developed and Developing Countries', *IMF Staff Papers*, Vol:45, pp:81-109.
14. Duman, K., (2002), 'Finansal Kriz ve Bankacılık Sektörünün Yeniden Yapılandırılması', *Akdeniz Üniversitesi İ.İ.B.F. Dergisi,* Year:2002, Vol:4, pp:132-145.
15. Eichengreen, B. and Bordo, M., (1999), 'Is Our Current International Economic Environment Unusually Crisis Prone?' Obtained through the Internet:www.rba.gov.au/PublicationsAndResearch/Conferences/1999/Bordo Eichengreen.pdf [accessed 17/03/2009]
16. Erdönmez, A., P., (2001), '*Sistemik Banka Yeniden Yapılandırmasına Teorik Yaklasım'*, TBB Banking Research Group, Working Paper, Ankara.
17. Güneş, H., (1999), 'Türk Bankacılık Sisteminin SWOT Analizi ve 2000 Yılındaki Muhtemel Profili', In: Enderoğlu T. (ed), *Conference Book Based On The Proceedings Of A Seminar Held in İstanbul / Turkey.* (pp. 35-49), Ankara:İktisadi Araştırmalar Vakfı.
18. IMF (International Monetary Fund), (1998), *'World Economic Outlook'*, Washington, Obtained through the Internet: http://www.imf.org/external /pubs/ft/weo/weo0598/pdf/0598top.pdf [accessed: 17/08/2008]
19. Kandemir, Ş., (2006), *'Finansal Serbestleşme Sürecinin Türk Bankacılık Sistemi Üzerindeki Etkileri'*, Phd. Thesis, Gazi Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
20. Kibritçioğlu, Aykut, (2001), 'Türkiye'de Ekonomik Krizler ve Hükümetler 1969-2001' *Yeni Türkiye Dergisi*, Vol: Eylül- Ekim, pp:174-182.
21. Kregel, J. (2000), 'The Brazilian Crisis: From Inertial Inflation to Fiscal Fragility*'*, *Working Paper 294*, Levy Institute.
22. Krugman, P. (1979), 'A Model of Balance- of- Payments Crises', *Journal of Money, Credit, and Banking*, Vol:11, August, pp: 311-25.
23. Krugman, P. (1997), Currency Crises, Obtained through the Internet: http://web.mit.edu/krugman [Accessed, 15/12/2008]

24. Komulainen, T., Lukkarila J., (2003), 'What Drives Financial Crises in Emerging Markets?', *BOFIT Discussion Paper,* No. 5/2003.
25. Köne, A. Ç., (2003), 'Para-Sermayenin Yeniden Yapılandırılması: Türk Özel Bankacılık Sektörü Örneği", *Doğuş Üniversitesi Dergisi*, Vol: 4(2), pp:233-246.
26. Kumcu, E., (2008), 'G7'de Ekonomi Tartışmaları 2008: Krizler, Para ve İktisatçılar', İstanbul: Remzi.
27. Kunt, A. D., Levine, R., (1997), '*The Financial System and Public Enterprise Reform: Concept and Cases' In*: Financial Development and Economic Growth, (pp:247–286), London: Routledge.
28. Mishkin, F. S., (2001), 'Financial Policies and the Prevention of Financial Crises in Emerging Market Countries' *Policy Research Working Paper Series,*No:2683.
29. Saral, T., (2009), 'Küresel Kriz Üzerine Düşünceler ve Öneriler', *Maliye ve Finans Yazıları,* Year:23, Vol: 83.
30. Selçuk, B., (2006), *'Yükselen Ekonomilerde Kriz Dönemlerinde Bankacılık Sistemlerinin Yeniden Yapılandırılması Ve Türkiye Örneği'*, Phd. Thesis, Marmara Üniversitesi Bankacılık Ve Sigortacılık Enstitüsü, İstanbul, Türkiye.
31. Seyidoglu, H., (2003), "Uluslararası Mali Krizler, IMF Politikaları, Azgelismis Ülkeler, Türkiye ve Dönüşüm Ekonomileri", *Doğuş Üniversitesi Dergisi,* Vol:4, No:2, pp:146-150.
32. Serdengeçti, S., (2002), "Şubat 2001 Krizi Üzerine Düşünceler: Merkez Bankası Bakış Açısından Çıkarılacak Dersler", Obtained through the Internet:http://www.tcmb.qov.ıotr/veni/evds/konusma/tur/2002/METUEvl11.htm. [accessed 11/10/2008].
33. Somçağ, S., (2008), *Türkiye'nin Ekonomik Krizi*, İstanbul:2006.
34. Sucu, İ., (2005), 'Türk Bankacılık Sistemi: Finansal Krizler ve Finansal Kriz Sonrası Uygulanan İstikrar Programlarının Etkileri' Ms. Thesis, Adnan Menderes Üniversitesi, Sosyal Bilimler Enstitüsü, Türkiye.
35. TBB (Türk Bankalar Birliği), Bankalarımız 2000, Obtained through the Internet: http://www.tbb.org.tr/turkce/kitap2000/2000.asp [accessed: 10/08/2008]
36. TBB (Türk Bankalar Birliği), Bankalarımız 2001, Obtained through the Internet: http://www.tbb.org.tr/turkce/kitap2001/2001.asp [accessed: 10/08/2008]
37. TBB (Türk Bankalar Birliği), Bankalarımız 2007, Obtained through the Internet: http://www.tbb.org.tr/turkce/kitap2007/2007.asp [accessed: 10/08/2008]
38. TBB (Türk Bankalar Birliği), 2009, Obtained through the Internet: http://www.tbb.org.tr/ v12/asp/bankalar1.asp [Accessed: 10/03/2009]
39. TEK, '2008 Küresel Krizi ve Türkiye' Obtained through the Internet: http://www.tek.org.tr/dosyalar/calıstay2009/2008.Küresel.Krizi.ve.Türkiye.pdf [accessed 10/02/2009]
40. Toprak, M., (2001), *Küreselleşme ve Kriz,* Ankara: Siyasal.
41. Tunalı, H., (2007), *'Kasım 2000 ve Şubat 2001 Krizleri Sonrasında Türk Bankacılık Sisteminin Yeniden Yapılandırılması'*, Phd. Thesis, İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü, İstanbul, Türkiye.

42. Ulusal Program, (2003), Obtained through the Internet: http://www.belgenet.com/arsiv/ab/up_2003-5.html [accessed, 30/11/2008]
43. Unakıtan, K., (2008), 'Global Financial Crisis and the Effects to Turkey Conference Opening Spech', Bahçeşehir Üniversitesi, İstanbul.
44. Uygur, E. (2001), 'Krizden Krize Türkiye: 2000 Kasım ve 2001 Şubat Krizleri', *Discussion Paper,* Türkiye Ekonomi Kurumu.
45. Yenitürk, N., (2003), Körlerin Yürüyüşü; Türkiye Ekonomisi ve 1990 Sonrası Krizler, İstanbul: Bilgi Üniversitesi.
46. Yılmaz, R., (2001), 'Türk Bankacılık Kesimdeki Yapısal Degisimler Ekseninde Formal ve Formal Olmayan Finans Sektörü', Obtained through the Internet: http://iktisat.uludag.edu.tr/dergi/10/13rasim/ rasim.htm [accessed 23/09/2008]
47. Yörükoğlu, M., (2009), 'Küresel Mali Kriz ve Türkiye Ekonomisi', *TCMB (Türkiye Cumhuriyeti Merkez Bankası) Tebliğleri*, İstanbul.

Economic Convergence and Regional Policy Strategies in UK and Greece: A Comparative Analysis of South Yorkshire and Kentriki Makedonia Region

Alexandros Karvounis

SEERC/University of Sheffield, Kousoula 14, Maroussi Greece, akarvounis@seerc.og

The present paper sets out to contribute to the debate on the causes and underlying factors of regional inequality between core and peripheral regions in countries like UK and Greece. It focuses on two regions – South Yorkshire and Kentriki Makedonia - which are both assisted by structural funds (Objective 1) and have a similar socioeconomic profile but different level technological and human capital development. The purpose is to try to compare the contribution of the structural funds assistance on GDP growth and the factors related to economic convergence.

Keywords – Regional Policy, convergence, disparities, UK, Greece, Objective 1, Kentriki Makedonia, South Yorkshire

1. Introduction

One of the most important policy issues in the European Union is the existence of income disparities among its member states, which resulted into the so-called regional- and cohesion policies. Its main objective is to reduce income disparities by various measures in order to achieve "social and economic cohesion" in the EU terminology.

The present paper aims to contribute to the debate on the level and evolution of regional economic disparities by providing latest evidence from Greece and UK regions. The purpose is to try to reveal if economies that are less developed tend to grow more rapidly than those that are more developed and if economic performance of these regions is growing or reducing the gap among them.

Today, the European Union consists of 27 Member States which form a community and internal market of 493 million citizens. At the same time, the economic and social disparities among these countries and their 268 regions are still great. One region in four has a GDP (Gross Domestic Product) per inhabitant under 75% of the average of the European Union of 27 (European Commission 2007).

Convergence literature, rooted in the theory of economic growth, tries to find an answer to the question if poor regions tend to achieve similar growth levels with rich ones. On the other hand, the literature of new economic geography tries to examine if already rich core attracts further economic activities or whether the poor periphery will also attract new economic centres (Tondl, G. 2001: 16-17).

The methodology adopted consists of a review of the theoretical as well as contemporary empirical evidence of the national and regional disparities across EU member States and especially across UK and Greece, to highlight the evolution of convergence/divergence process between more and less developed countries and regions. The paper will investigate the cohesion problem from the empirical evidence perspective, presenting the latest income disparities across UK and Greek regions. A Comparative performance analysis of two similar profile regions, one from UK and one from Greece, will take place, to explore in more detail the growing or reducing the gap between them. Finally, policy implications will be derived from the previous analysis.

2. Evolution of regional policy in UK and Greece

There is a large literature on the historical evolution of regional policy in UK. Most of the authors (e.g. Armstrong and Taylor 2000, McCallum 1979, Taylor and Wren 1997, Fothergill 2005, Temple 1994) recognise that over its long life of some 80 years, British regional policy has exhibited strongly varying levels of intensity and effectiveness.

On the contrary, development programmes of various types have been conducted in Greece since 1947, but without clear institutional reference, and with weak effectiveness. Regional development planning in Greece was officially introduced only in 1986 with the law 1622/86, which instituted the development of regional programmes. In reality this law was never implemented. Although the official national institutional system of regional development policy was not implemented, a parallel system emerged based on the regulations of the EU's Structural Funds from 1989 onwards. The implementation of what is now called EU cohesion policy replaced in practice the official national Greek regional policy (Economu 1997).

Whilst continuing right through the 1990s and early 2000s to retain a regional policy focused on indigenous development 'from within', British regional policy did begin to subtly change from 1997 onwards with the election of successive New labour governments. A new philosophy was introduced after 2000, with a new model of regional policy being developed in Britain. This so-called 'five drivers' model has come to form the core of both the British government's national instruments of regional policy as well as the various EU Structural Funds programmes within Britain. These five key drivers of growth were set out in Productivity in the UK: The Evidence and the Government's Approach (2000) published alongside the Treasury's 2000 Pre-Budget Report, which describes the UK's productivity performance in a national context. They are:

- skills (human capital)
- investment (physical capital)

- innovation
- enterprise
- competition

During the 1980s, Greek regional policy was connected directly with EU (probably best to stick with 'EU' even though it was not called that then!) regional policy through the introduction of law 1116/81. The main characteristic of the new law was the design of the first special Regional Development Programme 1981-1985, as well as funding from the European Regional Development Fund (ERDF) and the European Investment Bank (EIB). The 5-year Regional Development Programme 1985-1992 that followed was, however, never fully implemented due to changes that occurred during that period (Konsolas et al. 2002:1-2)

Since 1989 there have been four EU part-funded Community Support Frameworks - CSFs (1989-1993, 1994-1999, 2000-2006, 2007-2013), by which Greek Regional Policy has become fully incorporated into the EU's regional policy processes (Economu 1997, Konsolas et al. 2002, Psycharis 2006)

3. Latest regional disparities across UK and Greek regions

Reliable data on convergence in the European Union can be found in the Commission Reports on economic and social cohesion (1997, 2001, 2003 and 2007,2008). All added new elements of judgement to better evaluate the situation. However, all of them have common factors from which a series of conclusions can be extracted (European Commission 2003:69)

The British economy was a strong performer during the later 1990s and through until 2008. Despite this, there remains a substantial divergence in performance between England, Scotland and Wales. There are also significant and persistent differences in economic performance between and within British regions (H.M. Treasury 2001).

All British nations have shown strong GDP growth over the past decade, with England retaining GDP per capita above that of the Britain as a whole. Scotland and Wales have also maintained strong growth, but their gap with England has nevertheless still widened since the mid-1990s (DTI 2006:23).

The weaker economic performance in Wales, the North East and North West is caused by poor labour market performance (especially low participation rates and high unemployment). In addition, Wales and the South West also have unfavourable demographics (with ageing populations and low working-age population shares - H.M. Treasury 2001). In all regions, apart from Wales and Scotland, differences in productivity per worker are the largest contributor to the regional GDP per capita gap with the UK average.

	2002	2003	2004	2005	2006
Tees Valley and Durham	16900	17700	19400	19100	20000
Northumberland, Tyne and Wear	21100	21300	22300	23600	24500
Cumbria	18000	19000	19900	19800	21700
Cheshire	27300	27000	28400	30100	31500
Greater Manchester	22600	23100	25000	25300	26200
Lancashire	20200	20600	21600	22000	22500
Merseyside	18200	18600	19000	19600	20400
East Yorkshire and Northern Lincolnshire	20300	20500	21900	21800	22600
North Yorkshire	22500	23400	24400	23800	24600
South Yorkshire	18700	19100	20500	21100	21900
West Yorkshire	23400	23800	24800	25400	26100
Derbyshire and Nottinghamshire	22000	21900	23400	24300	25000
Leicestershire, Rutland and Northants	23700	25000	26500	26900	28100
Lincolnshire	18700	19300	19700	19000	19500
Herefordshire, Worcestershire and Warks	21400	21600	23500	24100	25600
Shropshire and Staffordshire	19200	19500	20700	20800	21600
West Midlands	23900	24200	24800	25300	25600
East Anglia	22100	22800	24300	24800	26100
Bedfordshire, Hertfordshire	27600	27400	29300	30800	30600
Essex	20000	21500	22500	22500	24100
Inner London	66500	68600	72500	75900	79400
Outer London	22500	22800	24500	24800	25800
Berkshire, Bucks and Oxfordshire	33600	35000	37100	37400	38800
Surrey, East and West Sussex	27000	27000	28000	28100	29500
Hampshire and Isle of Wight	24100	24400	26100	26800	28100
Kent	19900	20400	21700	23000	22600
Gloucestershire, Wiltshire and Bristol/Bath area	28400	29200	30400	30400	30800
Dorset and Somerset	19800	20400	21400	23200	24400
Cornwall and Isles of Scilly	15500	16000	16800	17200	18300
Devon	18500	18600	20700	20800	22000
West Wales and The Valleys	15900	15900	16900	17600	18300
East Wales	24200	25100	26200	26100	26800
Eastern Scotland	24400	25100	26800	27400	29200
South Western Scotland	21600	22200	23500	24600	25300
North Eastern Scotland	:	:	30500	30400	36300
Highlands and Islands	:	:	22700	23500	21400
Northern Ireland	19600	20000	21500	22100	23100

Table 1: Regional gross domestic product (PPS per inhabitant), by NUTS 2 regions

Source : EUROSTAT

It is clear that regional problems persist within Britain. It is proving difficult to remove the remaining regional disparities. Indeed, as has been shown, since the mid-1990s regional disparities have actually widened slightly, a phenomenon also witnessed in many other EU countries (European Commission, 2008 – Periodic Report). There are some long term trends based on latest reports (e.g. Regional Futures: England's Regions in 2030 by Arup et al)) that are expected to continue throughout the period to 2030. The challenge is for policy to affect these trends, raising the levels of economic performance of all regions, and reducing disparities over the longer term. Of course most of the studies have not yet taken account the economic crisis of 2008 which may disturb or/and reverse all future projections. The trend-based prognosis for the economy to 2030 is as follows (Arup et al, 2005):

- Regional disparities are likely to get widener rather than narrow. These widening disparities will occur at least as much within regions as between them.
- Manufacturing employment will continue to decline. Business services will continue to grow mainly in the South but business services employment will increase significantly in some Northern cities and sub-regions, particularly Leeds, Manchester and Edinburgh. Employment in local services will also grow gradually, but more strongly in the wealthiest regions, and the wealthiest parts of regions (Arup et al, 2005).
- Northern and Midlands regions are not likely to be able to catch up with growth rates in the South, at least for the next 10-15 years.

Since the early 1990s, a number of significant structural changes have taken place in Greece. Starting with the international economic environment, the stage-by-stage European Monetary Union (EMU) process was underway through the implementation of stabilizing macroeconomic policies. Subsequently, from 2004 onwards the enlargement of the EU-15 with 10 new member-states profoundly affected the 'cohesion' concept and the regional policy setting of the EU (Michelis et al., 2004). At the national level, Greece and its regions were aided significantly from the Second (1994-1999) and Third (2000-2006) CSFs in terms of physical and human capital investments (Benos and Karagiannis 2007).

Despite the considerable support of Structural Funds in regional development of the country since the early 1990s, especially in the modernization of infrastructure, a number of notable disparities still exist in the Greek regions.

There is a huge literature concerning the issue of regional disparities in Greece. According to Athanasiou et al (1995), regional inequalities widened during the first post-war decades and subsequently declined. Syriopoulos-Asteriou (1998) demonstrated absence of conditional β convergence[1] and evidence of 'dualism'

1 The 'beta' convergence of the neo-classical approach is obtained by a regression analysis estimating the growth of GDP per capita over a certain period of time in relation to its initial level. If the regression coefficient 'beta' has negative sign indicates that GDP per capita of countries with lower initial GDP per capita grow more rapidly than this of countries with higher initial GDP per capita.

between the south and the northern regions in the 1971-1996 period. By contrast, Petrakos-Saratsis (2000) has produced evidence leading to the conclusion that convergence may have occurred during 1971-1991. According to Tsionas (2002), evidence of dualism and non-convergence for 1971-1993 was found, but Michelis et al (2004) believe that there was regional convergence in 1981-1991. Christopoulos-Tsionas (2004) led to convergence in terms of labour productivity during the period 1971-1995. The evidence on β-convergence of GDP per capita since 1971 is therefore rather mixed, although whichever way the evidence leans (i.e. convergence or divergence) it is clear that a major degree of shift has not occurred. This is confirmed by the work of Alexiadis-Tomkins (2004), which reported evidence of no overall convergence and no evidence for the formation of a convergence club for 1970-2000 (Benos and Karagiannis 2008:54) .

	2002	2003	2004	2005	2006
Southern and Eastern	31700	32600	34200	36100	38600
Anatoliki Makedonia, Thraki	12800	13200	13700	14200	14400
Kentriki Makedonia	15000	15400	16400	16800	17900
Dytiki Makedonia	15400	15400	15800	17200	18000
Thessalia	14100	15100	15700	15700	16100
Ipeiros	14000	14600	15500	15700	16700
Ionia Nisia	16200	16400	16300	16900	17500
Dytiki Ellada	12500	13100	13200	13700	14100
Sterea Ellada	18300	18300	20100	21200	22100
Peloponnisos	14500	15200	16000	17100	18200
Attiki	24600	25500	27800	28300	30500
Voreio Aigaio	13000	14300	14400	14800	15900
Notio Aigaio	20100	20900	20900	20900	22700
Kriti	17600	17400	18200	18500	19600

Table 2: Regional gross domestic product (PPS per inhabitant), by NUTS 2 regions

Source : EUROSTAT

Statistics reveal that annual growth in GDP per head in Greece increased from 2.6% in the period 1995-1999 to 4.2% during 2000-03. This trend was followed by most regions, while centrally-located Attica and Central Greece, together with the heavily touristic regions (the South Aegean and Ionian islands) climbed above the national average during that period – 2000-2003 (Benos and Karagiannis 2007).

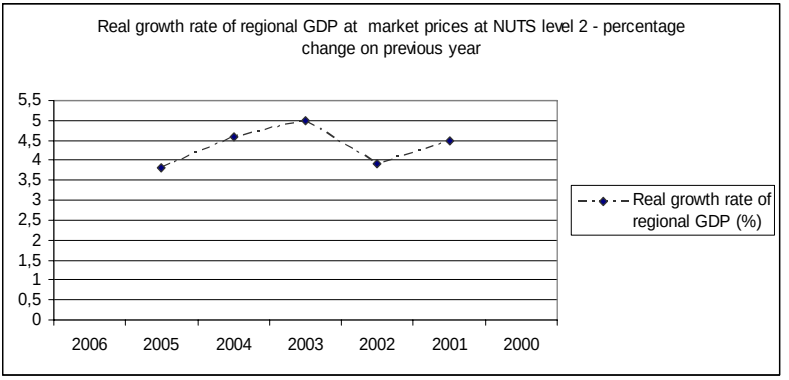


Figure 1. Real growth rate of regional GDP at market prices at NUTS level 2 - percentage change on previous year

Source : Own elaboration, EUROSTAT

4. Reasons for selecting the two case study regions

Selecting and comparing regions with different domestic structures is not an easy task and may seem odd at first sight. It is evident that some regions in the north of Europe (for example UK regions) have performed better than others of the south part (Greek regions). However, there are several factors which make the more dynamic regions exceptional cases and eliminate them as possible case studies. For example, high economic growth in the capital regions, like Inner London and Athens Attica, is almost exclusively driven by their service sector. They are too economically powerful and have attracted too much Foreign Direct Investments to be compared with any of the declining regions. At the other end of the scale, regions with a heavy and largely tourism sector means that their economic trajectory differs widely from that of the rest of the their country.

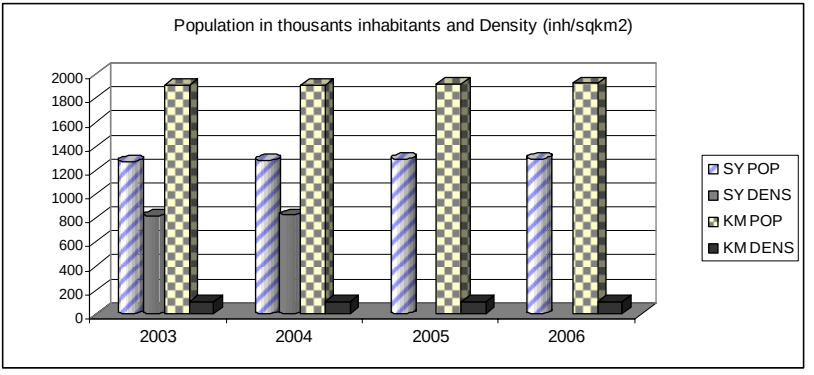


Figure 2. Population and Density

Source : Eurostat, own elaboration

In contrast, South Yorkshire and Kentriki Makedonia are comparable in a number of ways: Both regions are located in the north part of their countries and are medium sized with a similar population level but different density (Figure 2). Both regions have a core-city (Thessaloniki for Kentriki Makedonia and Sheffield for South Yorkshire).

Both assisted (Objective 1 status), these European industrial-based regions are facing declining pressures due to the restructuring of international and local economies in a globalised[2] context. These regions face common problems and common future prospects concerning the next programming period and the Lisbon Agenda. According to the 'new architecture' for EU Cohesion policy, Kentriki Macedonia and South Yorkshire will face the "statistical effect" of enlargement with the new regulations of the next programming period 2007-2013.

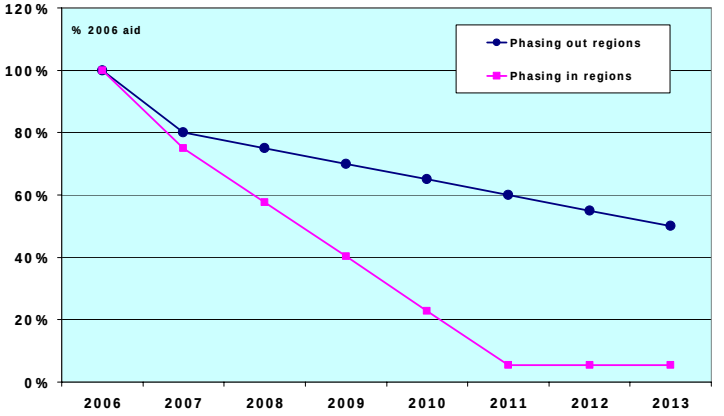


Figure 3. The degressivity profile (as % 2006)

Source: Directorate for Strategy, Planning and Evaluation of Development Programmes , Task Force NSRF 2007-2013

Also, from a structural point of view, South Yorkshire and Kentriki Makedonia are relatively diversified regions. More than half of the total employment in each region is occupied by the service sector (75% for South Yorkshire and 62,3% for Kentriki Makedonia). Both regions have a strong university base and perform better in innovation scoring compared to their basic domestic competitors.

Of course, there are a series of other greater dissimilarities. Accessibility to markets is different. Kentriki Makedonia, which is located in South East part of Europe, has been

2 Both facing globalization pressures – see European Commission (2009), R*egions 2020 - globalisation challenges for European regions*, Brussels,

relatively inaccessible in comparison to South Yorkshire which enjoys a more convenient location for European markets.

Another important difference between the two regions is their former sectoral structure of economies. South Yorkshire was based on the heavy coal industry with a strong coalfield community and Kentriki Makedonia on the manufacturing sector. Both sectors faced tremendous globalisation pressures and declining trends and Objective 1 programmes assisted in supporting the evolution of these sectors.

But perhaps the most important difference is the fact that Kentriki Makedonia presents a handicap compared to South Yorkshire on the technological level, the education attainment, the number of skilled workers, the innovation indicators and mainly their impacts which are related with long term unemployment and unemployment rate.

5. Economic geography and comparative performance of the two regions

This short comparative profile justifies also the selection of the two regions in order to address the introduction's research questions. From the main structural indicators it is evident that the two selected regions present many similarities but important performance variations

5.1 Economic geography of South Yorkshire

South Yorkshire is located within the Yorkshire & Humber region (Figure 4). Sheffield is the core city of the sub-region, the next largest urban centre being Doncaster, followed by Rotherham and the town of Barnsley.

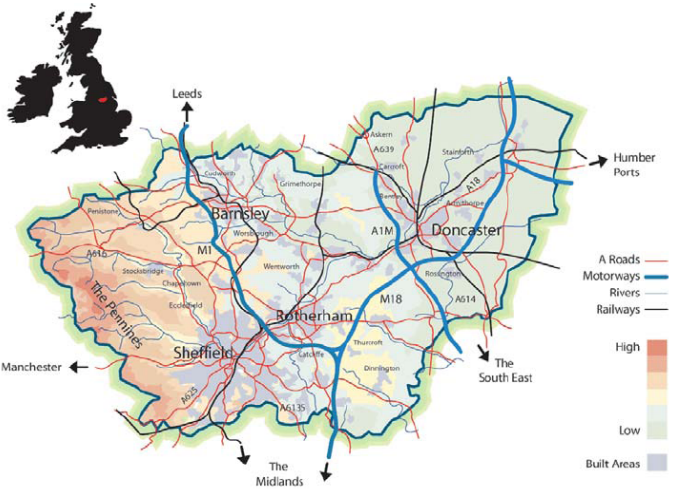


Figure 4 – South Yorkshire Map

Source: South Yorkshire Partnership (2006, p.4)

The population of South Yorkshire is largely urban in character with few small rural towns and villages. There are no really remote rural communities in South Yorkshire but the rural coalfield communities in the area are some of the most significantly deprived in the county. These areas often are not endowed with the environmental assets like other rural areas, which attract more commuters (South Yorkshire Partnership, 2006).

Sheffield is the largest city with a population over half a million people, Rotherham over quarter of a million, Doncaster close to 0.3 million and Barnsley 0.2 million. The total population of South Yorkshire sub-region is therefore over well over one and a quarter million.

Area	Population	Area (Sq km)	Pop. Density (per sq km)
Barnsley	221,000	328	674
Doncaster	289,000	581	497
Rotherham	252,300	283	892
Sheffield	516,100	367	1,406

Table 3. Population of South Yorkshire Districts

Source: Eurostat, National Statistics

In the field of the regional economy, UK per capita GDP (at PPS) has risen steadily against the EU27 average over recent years and had risen to 117% of the EU27 average by 2007. While South Yorkshire has been improving, this has been partly due to the 'pull effect' of the UK national economy. Although that the growth in SY has been faster than in the UK, sub-region was still 7.3 points below the EU27 average GDP per capita in 2005. Despite the continuing gap with the EU27 average, South Yorkshire's progress has meant that it has lost the Objective 1 status for EU Structural Funds in the 2007-13 programming period. It had previously enjoyed Objective 1 status in the 2000-2006 period. Technically, South Yorkshire is one of the 2007-2013 'phasing-in' regions and will therefore continue to enjoy some EU funding through until 2013, but on a much smaller scale than in 2000-2006 (Government Office for Yorkshire and Humber 2006).

Turning to South Yorkshire's sectoral economic structure , the most prominent industrial groups are distribution/hotels/restaurants and banking/ finance/ insurance (service sectors). Compared to the national average, there is a significantly higher proportion of businesses in "distribution, hotels and restaurants" in South Yorkshire, especially Doncaster, and a significantly lower level of activity in the "banking, finance and insurance" sector (although Sheffield's performance is well above that of Doncaster, Barnsley and Rotherham). South Yorkshire also has a higher proportion of manufacturing compared to the region and England – a sector which continues to decline (South Yorkshire Partnership, 2006).

The biggest changes in structure over the six years from 1998 – 2004 have been the decline in distribution, hotels and restaurants (- 4% across South Yorkshire, and almost

6% in Rotherham). There has also been a significant increase in banking, finance and insurance, of over 4% (South Yorkshire Partnership, 2006).

The Regional Survey of Economic Trends (2005) found that 10% of firms in South Yorkshire had research and development (R&D) links to universities, which was found to be slightly above the Yorkshire & Humber regional average (8%). In addition to this, 33% of firms were found to belong to specialist industrial networks. However, this rate was found to be much lower (11%) amongst responding firms from Barnsley (South Yorkshire Partnership, 2006). The sub-region's businesses ranked "Creativity and Innovation" joint fourth alongside "availability of good premises" (after workforce, management skills and flexible working patterns) as key attractive location features. The main investments in funding of innovation and incubation centres across the sub-region are the Advanced Manufacturing Park in Rotherham, the Sheffield's Bio-incubator and the development of Doncaster's Education City, which includes an investment of £30 million in a Digital Knowledge Exchange. £10 million has also been utilised to develop a Digital Media Centre in Barnsley (South Yorkshire Partnership, 2006).

5.2 Economic geography of Kentriki Makedonia

The region of KentrikiMacedonia (RCM) is located in the northern part of Greece (see Figure 5 below). From an economic perspective it is the most dynamic region of Greece after Attica, which is the capital city region.

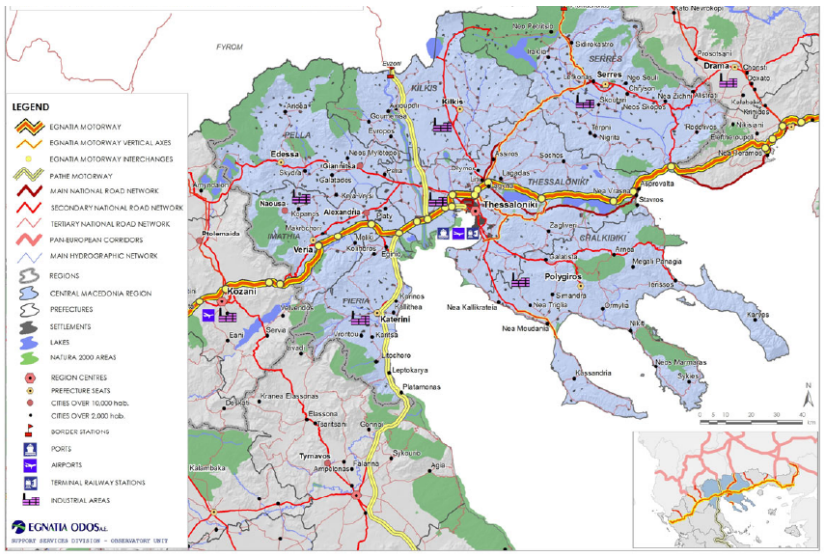


Figure 5. Region of Kentriki Macedonia

Source: Egnatia Observatory (2008)

The region of Kentriki Macedonia plays an essential role within the Greek regional system, as well as playing a significant pole for the wider Balkan region, a gateway of the EU on its South Eastern periphery and a hub for the main axes of the transport, energy and telecommunications networks at the national and International level (Foutakis et al. 2007).

The region of Kentriki Macedonia is considered to be one of the "most important gateways" of the, due to the fact that in shares continental borders with Albania, FYRof Macedonia (FYROM) and Bulgaria and is accessible also by air and sea . It has been awarded as :"the region of the future for 2006/2007", an exceptional honour at a global level country (Regional Authority of Central Macedonia 2008).

As can be seen from the table 4, the population of Kentriki Makedonia has reached around 2 million people by 2006 (Eurostat), comprising about 20% of the total population of the country. In the Balkans context, the population of the Kentriki Makedonia is almost equivalent to that of Slovenia and Croatia while its total GDP is five times as high as that of Albania, six times as high as of FYROM and around half the GDP of Romania (Foutakis et al. 2007).

	Greece	Kentriki Macedonia
Area (km^2)	13.195,743	1,914,616
Population 2006	11.148.500	1.923.600
Population change 2001-2006 (%)	0.36	0.44
GDP per inhabitant, PPS (2005)	21.589	17 921
Unemployment rate (2006)	8,9	9,3

Table 4. KentrikiMacedonia and Greece: main data

Source: EUROSTAT 2006

The impact of the restructuring of the productive system has led to a decline of the primary sector, a relative reduction of the secondary sector (against a background of continued absolute growth in manufacturing) and a significant growth of the tertiary sector (Foutakis et al. 2007).

The composition of production in the Region of Kentriki Macedonia has tended to follow the national pattern. Manufacturing is considerably developed in Kentriki Macedonia, representing 23.5% of the gross regional product (GRP). The region is the centre of a major industrial complex, with 2,700 small and medium-sized enterprises in the branches of food and beverages, textiles and clothing, furniture, and larger companies in the industries of metal production, chemicals and plastics. This new production complex has shown itself capable of internationalisation in the European market, and includes most of the 2,800 Greek firms which have invested in the Balkans. On the other hand, the internationalization process has created job losses inside the region and there has also been a significant increase of the unemployment rate. The number of jobs in Kentriki Makedonia has fallen by more than 3% during 90s, and today many urban centres other than Thessaloniki face severe problems of

decline. The unemployment rate in KM in 2006 reached 9.3 %, according to the National Statistic Service of Greece.

This brief presentation of a number of key characteristics of the region implies that Kentriki Makedonia is currently passing through a structural adjustment phase, with consequences not only for employment but also for intraregional disparities. The productive system is being strongly affected by globalization pressures. Restructuring that has taken place over the last fifteen years in the Balkans has also affected the domestic system by increasing pressures on certain labour-intensive sectors of production in the region and also by creating opportunities for the collaboration and expansion of a number of different productive sectors of the region (Foutakis et al. 2007:142).

5.3 Comparative performance of the two regions

The regional knowledge scorecard (Table 5) compares the two regions across key competiveness indicators; GDP per capita, labour productivity, unemployment, economic activity, innovation (R&D expenditure by businesses and patent applications per inhabitant). These indicators offer a wide-ranging look at both the state of the region's economies and also the level of R&D undertaken.

INDICATORS 2006	SOUTH YORKSHIRE	KENTRIKI MAKEDONIA
Area (.000 km2)	1.547	1.914
Population (1000 Inh.)	1307	1787
GDP per capita (PPS /EU27=100)	89,4	78,7
Productivity GVA	95,1	48
Unemployment rate (%)	6,9	9,3
Employment by sector (%)		
Primary	0,2	13,6
Secondary	20	24,1
Tertiary	75	62,3
Long term unemployment (% of unemployment)	11,14	43,01
Eur. Patent applications Per million of inh.	236	3
R&D expenditure by business (euros per inh.)	105,2	24

Table 5. Regional knowledge scorecard for the year 2006.

Source: Eurostat, own elaboration

GDP Per Capita

Both South Yorkshire and Kentriki Makedonia lag the EU-27 average in GDP per capita with South Yorkshire being closer to the EU-27 average compared to Kentriki Makedonia (Table 5) Latest figures in regional gross domestic product (Figure 6) show a parallel increase of the value of the indicator in both regions, but reveal the remaining of regional disparities which is translated in a substantial performance variation, although the main objective of structural assistance is towards regional convergence.

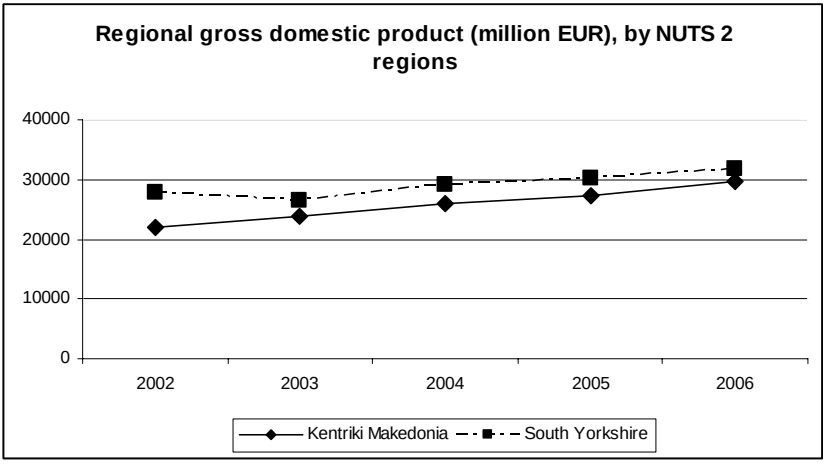


Figure 6. Regional gross domestic product in million Euros

Source: Eurostat

Labour Productivity

Both regions lag the EU-27, but labour productivity is higher in South Yorkshire with an index score of 95.1 (EU-25 = 100), and significantly lower in Kentriki Makedonia, with an index score of 48.

Unemployment and Economic Activity

Unemployment tends to be the most frequently available, most up-to-date and possibly the most sensitive indicator to economic performance. It therefore provides a warning of expected changes in other indicators (South Yorkshire Partnership 2006).

Unemployment in Yorkshire is lower than Kentriki Makedonia, at 6,9 %, reflecting the prolonged period of economic growth in the UK from the mid 1990s. In 2007 it started to fall again after rising in 2005 and 2006 both in the UK and South Yorkshire. On the contrary (Figure 7), unemployment in Kentriki Makedonia started to fall after 2004 (year of great economic growth due to Olympics).

During the 1970s and 1980s South Yorkshire suffered from a catastrophic loss of jobs due mainly to the decline of its traditional base of coal, steel and ancillary engineering industry. This continued until the end of the 1990s losing nearly 200,000 industrial jobs and gaining 100,000 service sector jobs. Since 2000 however the size of the workforce has been increasing and by 2004 rose by over 50,000 employees (Draft Operational Programme 2007-2013 Version 3) The largest employment sector is public administration and health which employs 29.4% of the region's workforce although. Distribution, hotels and restaurants account for another quarter of total employment with banking, finance and insurance (17.0%) and manufacturing (13.6%) accounting for the other larger employment sectors. (Draft Operational Programme 2007-2013 Version 3).

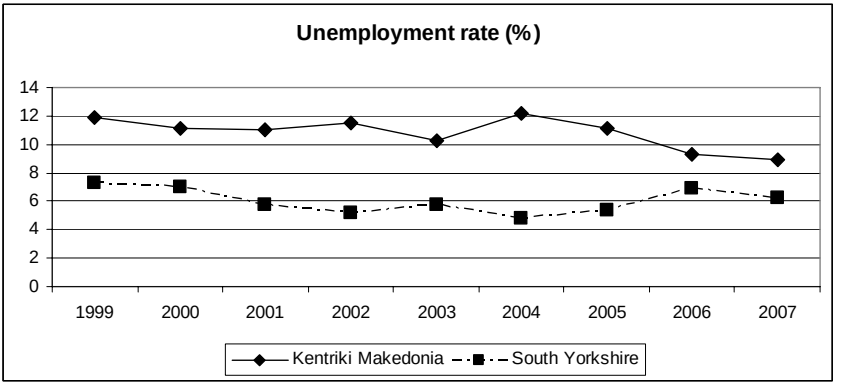


Figure 7. Unemployment rate % in the two regions

Source: Eurostat

Manufacturing is considerably developed in Kentriki Macedonia, representing 23.5% of the gross regional product. The region is the centre of a industrial complex, with 2,700 small and medium-sized enterprises in the branches of food and beverage, textile and clothing, furniture, and larger companies in the industries of metal production, chemicals and plastics. This new productive complex has shown itself capable of internationalisation in the European market, and includes most of the 2,800 Greek firms which have invested in the Balkans[3]. But on the other hand it created job losses inside the region and significant increase of unemployment rate. The number of jobs had fallen by more than 3% between 1996 and 1998, and today many urban centers except from Thessaloniki face severe problems of decline.

In Thessaloniki, the main employment is in the tertiary sector, while in all the other prefectures agriculture is popular, with a wide variety of crops and good prospects for growth. The emerging industries belong to the tertiary sector, including medicine and

3 (http://www.innovating-regions.org/network/whoswho/regions_search.cfm)

health services, software, international retail chains, business services, and tourist services.

Finally, long term unemployment is much higher and persistent in Kentriki Makedonia compared to South Yorkshire and reveals the structural deficiencies of its regional economy.

R&D – Innovation

Two indicators for R&D performance and Innovation within the regions are used. The first, R&D expenditure by business per capita, measures the level of investment in R&D by the private sector in each region. It is evident that in South Yorkshire the amount of investment is four times higher than that of Kentriki Makedonia (105,2 % with 24% respectively. Thus, the firms in Kentriki Makedonia can be regarded as static, they are not changing or developing their products and production processes; a serious problem which curbs economic development.

The second indicator concerns innovation with the European Patent applications per million of inhabitants. The difference between the two regions is tremendous, although Kentriki Makedonia performs better than the rest Greek regions. This shows a national deficit in innovation between Greece and UK, due to the fact that Greek regions invest primarily in the infrastructure development to reduce their gap in accessibility.

6. Existing institutional framework for designing and implementing regional policies in the two regions

There is massive variation in the size, strength, status and coverage of regional governments in Europe. In some countries they are very weak and have few economic responsibilities. In others they are stronger but cover small areas which make little economic sense (Harding et Al. 1996).

According to Millio (2007), regional governments have a set of capacities and the degree to which they exist determines output (the quantitative implementation of resources measured by the expenditure rate). In order to improve output, their administrative capacity needs to be enhanced (first relationship). As soon as they implement the resources, they should produce a result (institutional outcomes) measured in terms of GDP growth (second relationship).

This model (Figure 8) emphasizes the importance of capacity in explaining the policy output (expenditure rate) and ultimately the outcomes (GDP growth).

Unlike in Scotland or Wales there is no elected regional government in UK. A similar picture exists in Greece. In UK the Government Offices represent central government in each region (Government Office of Yorkshire and Humber in the case of South Yorkshire) and their purpose is both to join up the work of individual departments within the regions and to use their local knowledge of the region concerned to influence the development and implementation of Government policy as it affects the area (House of Commons 2007). Each Greek Region contains a Regional Authority which comprises from the General Directorate of the Region (Kentriki Makedonia's in our case) and the Regional General Secretary's Office, which are responsible for the

strategic as well as the economic planning of the region (Table 6). On the contrary, in UK there are separate agencies concerned with economic development called Regional Development Agencies (RDAs): These were created in 1999 to provide a clearer focus for economic development programmes. Although accountable to the government in London, the activities of RDAs (Yorkshire Futures) are governed by a board comprising representatives of business, trade unions and other organisations in each region (Tomaney 2005).

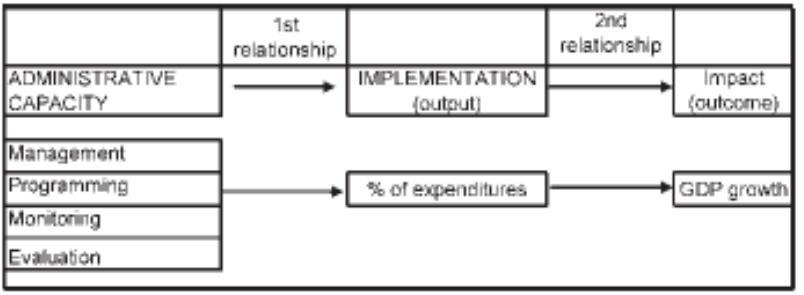


Figure 8. Administrative capacity model

source: Milio S., 2007

In institutional level, Regional Assemblies for each English region (i.e Yorkshire & Humber Regional Assembly) are made up of representatives of local authorities and of business and voluntary bodies, with a small staff and very limited resources. They have a co-ordinating, strategic role and work with the Regional Development Agency, the Government Office, local authorities and other representatives in the region (Tomaney 2005). In Greek regions, the Regional Council is a collective instrument of democratic programming, constituted by the General Secretary, representatives of Local Authority Organizations of the 1st and 2nd degree and representatives of all production powers and entities playing a constructive role in the preparation of integrated Regional development programs.

GOVERNANCE STRUCTURE LEVEL	SOUTH YORKSHIRE	KENTRIKI MAKEDONIA
Strategic	RDA	Secretary General Office
Administrative	Government Office of Yorkshire and Humber	General Directorate of Kentriki Makedonia
Institutional	Regional Assembly	Regional Council
Structural funds management	Objective 1 Directorate of Y&H Government Office	Regional Operational Managing Authority of Kentriki Makedonia

Table 6. Comparative Governance Structures

Source: Own elaboration

Finally, an important role in the preparation of the aforementioned priority guidelines is played by the Kentriki Makedonia's Regional Operational Programme Special Management Services (Managerial Services), which have a specific organization chart and the General Secretary and Chief of the Special Management Services as Head of their operations, with separated tasks and scopes in units consisting of specialized personnel. These units are (Tsiotras G. 2005):

- Programming and Evaluation Unit

- Management and Monitoring Unit

- Control Unit

- Organization and Support Unit

A similar internal structure had the Objective 1 Directorate of Y&H Government Office since 2006. The 2007-2013 ERDF programme for Yorkshire and Humber will be managed from Yorkshire Futures. All these principal agencies lie at the heart of a larger web of bureaucracy in the regions.

7. Comparative analysis of Objective 1 programmes 2000-2006 and contribution to economic performance

For the programming period 2000-2006 the Structural Funds in the UK financed six Objective 1 programmes and 14 Objective 2 programmes for a total contribution of € 11.3 billion and levering additional € 15.9 billion in national match funding, as well 11 URBAN II and 4 Interreg IIIA programmes. The EU Structural Fund programming approach has had an important impact on the effectiveness of regional policy in the United Kingdom. South Yorkshire was a designated Objective 1 region during 2000-2006 (DTI 2006).

The 2000-06 EU Structural Funds budget allocated to Greece amounted to a total of €25.000 billion, compared with €19.271 billion in 1994-99. Thus, the amount available for the current period is 1.1% more each year than in the previous period. The 3rd Community Support Framework (CSF) thus marked a radical shift from the emphasis of the previous two CSFs with their emphasis on general infrastructure investment. CSF priorities in 2000-06 have been focused on the types of investment in physical, human and knowledge capital that are most conducive to increase Greek productivity (G.S.I.D. 2005).

Regarding Kentriki Makedonia, the priorities of its Operational Programme for the period 2000-06 aimed at making the region more competitive and reducing intra-regional disparities in order to create conditions that will ensure long-term development in the region, namely sustainable development and competitiveness [4].

[4] Regional Operational Programme of Central Macedonia 2000- 2006, Single Programming document, Managing Authority 2000.

South Yorkshire 's Objective 1 programme vision[5] was "*to build a balanced, diverse and sustainable high growth economy in South Yorkshire, recognised as a growing European centre for high technology manufacturing and knowledge based services and offering opportunities for the whole community*."

With regard to their structure (Table 7), the Kentriki Makedonia 's Regional Operational Programme[6] included seven Priority 'axes', the seventh of which referred to the Programme's Technical Assistance. Similarly, South Yorkshire's Objective 1 Programme's[7] consisted of seven Key Priorities, with the last one as Programme's Technical Assistance

Both programmes officially closed on the 31st of December 2008, but it is still too early to assess the impact of the policy and the net effects as the Ex Post evaluation has still to be done. Some early estimates[8]. about South Yorkshire 's programme outputs are: 16,500 new gross (or net) jobs, 6,660 SMEs supported - £640m additional sales, 600 new businesses created, 290,000 persons assisted with training/development, 115,000 sq m. new floorspace created, £293m net additional GDP, and £335m Net Safeguarded GDP.

In Kentriki Makedonia, the impact of Regional Operational Programme 2000-06 on employment and environment was projected to be positive, but not so significant in terms of size. New employment is estimated at eventually being around 8,350. This is important, but is still small compared with numbers unemployed in the region (92,608 according to 2001 Census of Population -Kantor–Eurotec 2008). The impact on the unemployment rate will be, at best, small (a decrease of 1%). The contribution of the programme in funding interventions for the protection of environment has been very low, because most of the funding was allocated to the development of spatial plans, and these have had only an indirect impact on the protection of environment (Kantor–Eurotec 2008).

	Financial Allocation Objective 1	Percentage of GDP (million euros for year 2001)
Kentriki Makedonia	1547,198	7%
South Yorkshire	3088,2307	11%

Table 8. Financial Allocation Objective 1 - Percentage of Gdp (million euros for year 2001)

Source: Single Programming Documents, Eurostat, Own elaboration

5 Government Office for Yorkshire and the Humber and Yorkshire (1999), SPD 2000-2006.

6 The draft ROP was approved by the European Commission on 22/03/2001 and the revised ROP was approved on 27/12/2004

7 The programme was formally adopted by the Programme Monitoring Committee on 19 July 2000 and in Doncaster on 25 July 2000.

8 Objective 1 Directorate, (2008), *Objective 1 progress towards targets*, Research and Evaluation Department, Rotherham

SPD OBJECTIVE 1/ 2000-2006		
	SOUTH YORKSHIRE	**KENTRIKI MAKEDONIA**
PRIORITY AXIS 1	**Stimulating the emergence of new growth and high technology sectors**	**Promotion of Thessaloniki's metropolitan role and encouragement of innovation and entrepreneurship**
Total Cost "Million euro"	573,253	378,187
Public Expenditure	350,786	252,973
ERDF Contribution	478,778	189,73
National Public Expenditure	152,818	63,243
Private Participation	222,467	125,215
PRIORITY AXIS 2	**Modernising business through enhancing competitiveness and innovation**	**Protection and promotion of the environment in Central Macedonia**
Total Cost	352,283	136,322
Public Expenditure	210,69	136,322
ERDF Contribution	215,531	102,241
National Public Expenditure	98,21	34,08
Private Participation	141,593	-
PRIORITY AXIS 3	**Building a world leading learning region which promotes equity employment and social inclusion**	**Reduction of interregional disparities in Central Macedonia**
Total Cost	937,462	600,839
Public Expenditure	866,762	544,181
ERDF Contribution	95,542	393,553
ESF Contribution	293,282	14,583
National Public Expenditure	477,938	136,045
Private Participation	70,7	56,658
PRIORITY AXIS 4	**Developing economic opportunities in targeted communities**	**Rural Development**
Overall Cost	292,196	259,996
Public Expenditure	263,795	171,488
EAGGF-G Contribution	45,76	118,55
ERDF Contribution	215,596	6,383
National Public Expenditure	130,215	46,556
Private Participation	28,401	88,508
PRIORITY AXIS 5	**Supporting business investment through strategic spatial development**	**Reduction of unemployment and provision of equal opportunities in knowledge and skills**
Overall Cost	672,8657	78,628
Public Expenditure	471,2457	78,628
ESF Contribution		58,971
ERDF Contribution	672,867	
National Public Expenditure	187,34	19,657
Private Participation	201,62	-
PRIORITY AXIS 6	**Providing the foundations for a successful programme**	**Development of mountainous areas, hinterland zones and disadvantaged / problematic areas**
Overall Cost	224,851	82,352
Public Expenditure	120,121	61,198
EAGGF-G Contribution		25,752
ERDF Contribution	224,851	15,949
ESF Contribution		3,301
National Public Expenditure	33,05	16,196
Private Participation	104,73	21,154
PRIORITY AXIS 7	**Technical Assistance**	**Technical Assistance**
Overall Cost	35,32	10,872
Public Expenditure	35,32	10,872
ERDF Contribution	28,256	5,232
ESF Contribution	7,064	1,149
EAGGF-G Contribution		1,774
National Public Expenditure	17660	2,718
Private Participation		-
	TOTAL	
Overall Cost	3088,2307	1547,198
Public Expenditure	2318,7197	1255,67
Community Expenditure	1221,4887	937,168
National Public Expenditure	1097,231	818,5
Private Participation	769,511	291,6

Table 7. Comparative financial allocation of the two Objective 1 SPDs

Source: Own elaboration

The economic performance of the regions has improved during the 2000- 2006 period (table 9). Concerning Yorkshire & the Humber, the region still remains below the UK average for per capita GDP. The weakest sub-regional economy, that of South Yorkshire, has shown most improvement, due in no small part to the investments made during the Objective 1 Programme. The percentage on GDP change during 2002-2006 rises up to 15%, where the financial assistance of Objecive 1 reached 11% of the 2001 regional GDP (table 8). South Yorkshire needs to continue restructuring and attracting investment to replace lost jobs and meet the employment needs of a growing economy (GoYH, 2007)

Regional gross domestic product (million EUR), by NUTS 2 regions						
	2002	2003	2004	2005	2006	CHANGE 2002-2006
Kentriki Makedonia	21970	23882	25837	27396	29550	35%
South Yorkshire	27683	26589	29026	30228	31807	15%

Table 9. Regional gross domestic product (million EUR), by NUTS 2 regions

Source: Eurostat, Own elaboration

Kentriki Makedonia seems to perform better during this period of time although the financial assistance was lower in relation to GDP. In more detail, Kentriki Makedonia increased its regional GDP around 35% from 2002 till 2006, although Objective 1 assistance amounted around 7% of the 2001 regional GDP. In absolute numbers, Objective 1 assistance was around double in South Yorkshire compared to Kentriki Makedonia, but the last one seems to reduce the gap in regional GDP with South Yorkshire in 2006.

8. Lessons learned

According to Rodríguez-Pose, A. (2000), the success of regional development strategies depends on a series of factors like geography, accessibility, economic and social structure, skills, institutions, politics and culture which regulate the success of development strategies. Furthermore, there is an emerging literature assessing those factors most associated with high rates of regional economic innovation and growth. They highlight supply-side factors dealing with local skills, specialised knowledge and interfirm relations. Also important are technological factors relating to higher education capacities, education-industry links, R&D and technology transfer facilities. Behavioural and cultural factors and 'soft infrastructures' are also stressed, including local traditions of networking and inter personal trust (Harding et Al. 1996)

Hence comparing development strategies in two regions like South Yorkshire and Kentriki Makedonia, which differ in geographical, economic and social terms, makes it difficult to assess to what extent any economic success is the result of the implementation of the policy or of the economic dynamic of the region (Rodríguez-Pose, A. 2000). Current research on programme implementation (EEO Group 2002, Kantor-Eurotec 2008), though still in an early stage, indicates that many of the

deficiencies of previous programmes have not been the result of ignorance or conservatism on the part of the planners. According to Plaskovitis (2006), they rather reflect what is actually demanded on the ground from local populations and programme beneficiaries. Wherever we meet a more ambitious and innovative measure there are huge difficulties in finding users to absorb the allocated funds. In this sense it is probably better to prepare a strategy for changing the wider development culture and stereotypes, rather than pressing for a top down programme modernization.

So is it possible to generalise the experience of these regions in other EU regions? For example, there are tangible effects from Objective 1 funding towards job creation, but there are many doubts about whether the funding has been adequate for the challenges it has faced (Taylor and Wren 1997:845) or is being used optimally (Gripaios and Bishop 2006:950). The impact and effectiveness of the strategies is largely determined by progress in the following key respects: 1) structural reforms in the labour, goods and services markets, 2) sustainable rural development and agriculture, 3) mobilisation of the private sector in all regions, and 4) significant improvement in management capabilities, by respective agencies (G.S.I.D. 2005).

References

1. Alexiadis, S. and Tomkins, J. (2004), *Convergence Clubs in the Regions of Greece,* Applied Economics Letters 11, 387–391
2. Armstrong H. W. (1995), *Convergence among Regions of the European Union, 1950-1990*, Papers in Regional Science, Vol. 74, No. 2, April 1995.
3. Armstrong, H. and Taylor, J. (2000), *Regional economics and policy*, 3rd edition, Blackwell.
4. Armstrong, H., (1998*), What Future for Regional Policy in the UK*? The Political Quarterly, Volume 69, Number 3, July-September 1998, pp. 200-214(15), Blackwell Publishing
5. Arup, Regional Forecasts Ltd and Oxford Economic Forecasting Ltd, (2005), *Regional Futures: England's Regions in 2030*, English Regions Network, RDA Planning Leads Group, Office of the Deputy Prime Minister, Department for Transport
6. Athanasiou, L.A, Kavvadias, P.A, Katachianou, D.N., and Tonikidou, P.I (1995). *Interregional Analysis and Policies and Basic Data per Region and per Department,* Reports 24 (19), Center of Planning and Economic Research, (KEPE), Athens (in Greek).
7. Benos N. and Karagiannis St. (2007), *Growth Empirics: Evidence from Greek Regions*, Paper Nr 96, Center of Planning and Economic Research, (KEPE), Athens (in Greek).
8. Benos N. and Karagiannis St. (2008), *Convergence and economic performance in Greece: evidence at regional and prefecture level*, Review Of Urban & Regional Development Studies, Volume 20, Number 1, March 2008 , pp. 52-69(18),Blackwell Publishing
9. Christopoulos,D.K and Tsionas, E.G. (2004), *Convergence and Regional Productivity Differences: Evidence from Greek Prefectures*, The Annals of Regional Science 38, 387–396.

10. Department of Trade and Industry – DTI, (2006), *National Strategic Reference Framework, EU Structural Funds Programmes: 2007-2013*, Regional European Funds Directorate, London.
11. Economou D. (1997), The planning system and rural land use control in. Greece: A European perspective, European Planning Studies, Vol. 5, No. 4, 1997.
12. Egnatia Observatory (2008), *Development profile of Central Macedonia*, Annual Report, (in greek), Thessaloniki.
13. EKOS Consulting (2004), The impact and added value of EU structural funds in Yorkshire and the Humber, Leeds: EKOS Consulting
14. European Commision (1997,*First cohesion report*, European Commission, Luxembourg.
15. European Commision (1999), Sixth Periodic Report on the Social and Economic Situation and Development of the Regions of the European Union, European Commission, Luxembourg.
16. European Commision (2004), Third progress report on cohesion: Towards a new partnership for growth, jobs and cohesion, European Commission, Luxembourg.
17. European Commision (2005) *The impact of cohesion Policy in Member States*, Evaluation Unit, Luxembourg.
18. European Commision (2007), Fourth report on economic and social cohesion, European Commission, Luxembourg.
19. European Commision (2008), Fifth progress report on economic and social cohesion, European Commission, Luxembourg.
20. European Enterprise Organisation - EEO GROUP (2002), *Ex post evaluation of CSF in Objective 1 regions*, Country report, Greece
21. Fother Gill, S. (2005), A new regional policy for Britain, Regional Studies, Taylor and Francis Journals, vol. 39(5), pages 659-667, July
22. Foutakis D. and Thoidou E., (2007), Development Planning and Territorial Integration Prospects in. South Eastern Europe: A Foresight Exercise in the Region of. Central Macedonia, at Getimis P. and Kaukalas G., Overcoming Fragmentation in Southeast Europe: Spatial Development Trends and Integration Potential, Ashgate Publishing, Ltd
23. General Secreteriat of Investments and Development Finance- GSID (2005), *EU funded Development Programmes in Greece: Policies, Results, Prospects*, Ministry of Finance, Athens
24. General Secreteriat of Investments and Development Finance- GSID (2008), *Greek National Strategic Reference Framework 2007–2011*, Ministry of Finance, Athens
25. Government Office for Yorkshire and Humber, (2004), *South Yorkshire Objective 1 Programme 2000-2006*, Historical Background, Leeds.
26. Government Office for Yorkshire and Humber, (2006), *Draft Operational Programme 2007-2013 Version 3*, Leeds.
27. Government Office for Yorkshire and Humber, (2007), Yorkshire and The Humber Regional Competitiveness and Employment Programme Operational Programme 2007-2013, CCI 2007 UK 162 PO00, Leeds.
28. Gripaios P. and Bishop P. (2006), *Objective One funding in the UK: a critical assessment*, Regional Studies, vol. 40(8), pages 937-951, November

29. Harding Al., Evans R., Parkinson M. and Garside P. (1996), *Regional government in Britain: An economic solution* , Cities, Volume 14, Issue 3, June 1997, Pages 186-187, Policy Press, Bristol
30. H.M. Treasury (2001), Productivity in the UK, No. 3: The Regional Dimension, H.M. Treasury, London.
31. H.M. Treasury (2003a), A Modern Regional Policy for the United Kingdom., H.M. Treasury, London.
32. H.M. Treasury (2003b), Productivity in the UK, No. 4: The Local Dimension, H.M. Treasury, London.
33. House of Commons - Communities and Local Government Committee (2007), Is there a future for Regional Government?: Fourth Report of Session 2006–07, London: The Stationery Office Limited
34. Konsolas N., Papadaskalopoulos A. and Plaskovitis I. (2002), *Regional Development in Greece*, Berlin: Springer.
35. Martin, R. & Tyler, P., (2006), Evaluating the impact of the structural funds on objective 1 regions: An exploratory discussion, Regional Studies, Taylor and Francis Journals, vol. 40(2), pages 201-210, April.
36. McCallum, J.D. (1979), *The development of british regional policy*, in D. Maclennan and J.B. Parr eds., Regional Policy: Past Experience and New Directions, Martin Robertson, Oxford.
37. Michelis, L., Papadopoulos, A.P, and Papanikos, G.T (2004). *Regional Convergence in Greece in the 1980s: An Econometric Investigation*, Applied Economics 36, 881–888.
38. Milio, S. (2007), Can administrative capacity explain differences in implantation performances?, Regional studies, 41 (4). pp. 429-442. ISSN 0034-3404
39. Petrakos G. and Saratsis Y. (2000), "Regional inequality in Greece", Papers in Regional Science, Vol. 76: 57-74
40. Plaskovitis Ilias (2006), *Regional Policy Objectives in Greece - Twenty Years of Regional Development Programmes*, 46th Congress of the European Regional Science Association (ERSA), Volos
41. Regional Authority of Central Macedonia (2008), *Development profile of the Region of Central Macedonia*, Publicity Bulletin, Thessaloniki.
42. Rodríguez-Pose, A. (2000), Economic convergence and regional development strategies in Spain: The case of Galicia and Navarre, EIBPapers , Volume 5 No 1 2000, pp: 89 -115
43. South Yorkshire Partnership, (1999), SOUTH YORKSHIRE OBJECTIVE 1 PROGRAMME 2000-2006 SINGLE PROGRAMMING DOCUMENT, CCI 1999 GB 161 DO 003
44. South Yorkshire Partnership, (2006), *Progress in South Yorkshire 2006*, with funding from Yorkshire Futures, the Regional Intelligence Network
45. South Yorkshire Partnership, (2007), *Progress in South Yorkshire - Light Touch Review 2007*, with funding from Yorkshire Futures
46. Syriopoulos, C. and Asteriou, D. (1998). *Testing for Convergence across the Greek Regions*. Regional Studies 32, 231–238.
47. Taylor, J. and Wren, C. (1997), *UK Regional Policy: An Evaluation*, Regional Studies, Taylor and Francis Journals, vol. 31(9), pages 835-848, December.

48. Temple, Marion (1994), *Regional Economics,* London: The Macmillan Press Ltd; New York:St Martin's Press Inc.
49. Tomaney J. (2005), *The governance of regional policy in the UK*, Paper prepared for the 'International Conference on Regional Innovation' organised by Korea Institute for International Economics and Trade, Daegu, Korea, 5th-7th, October 2005.
50. Tsionas, E.G. (2002), *Another Look at Regional Convergence in Greece*. Regional Studies 36(6), 603–609.
51. Tsiotras G. (2005), *Structural Funds - The Greek Experience "Administrative Skills at a Regional Level*, paper presented in Seminar: «Promotion of Regional Collaboration between Greece and the Czech Republic for the efficient use of European Structural Funds», Zappeion Megaron, Athens, March 4th, 2005

INFORMATION AND COMMUNICATION TECHNOLOGIES

Prion Neural Systems: Why Choosing a Hybrid Approach of Natural Computing?

Luciana Morogan

"Spiru Haret" University, Faculty of Mathematics-Informatics
Ion Ghica 13, Bucharest, Romania
{morogan.luciana}@gmail.com

Abstract. Why choosing a hybrid approach of natural computing to model a controversial biological phenomena? A new direction in computer science may begin starting with a new hybrid system for unconventional biological computation. It is called *prion neural system*. This new biology inspired computational model of the way that an infectious agent, called *prion*, influence and modify the way that brain express itself in the human behavior begins within the pages of this article.

Key words: *PrP - prion protein; PrPc - cellular PrP; PrPSc neural networks; neural computing; evolutionary computing; natural computing.*

1 Biological Inspired Implications as an Introduction

Nervous system. "A gigantic network, building up the "hardware" which carries out the computations underlying the human behavior"[1]. Consisting of **neurons**, it has the particularity of being involved in huge networks by means of the synapses (the contact of the axon - the "output wire" - of a neuron with the dendrites of another neuron). But modeling a gigantic network of neurons is no longer a novelty in the way science already expressed its interests. Natural computing, the field of research that works with computational techniques inspired in part by nature and natural systems, aims to develop new computational tools for solving complex, usually conventionally-hard problems. This often leads to the synthesis of natural patterns, behaviors and organisms, and may result in the design of novel computing systems that use natural media with which to compute. It looks like a grate movie with a big happy-ending. Isn't it? But what about the negative character? What if he/it wins? What if the grate movie will no longer have a happy-ending, as in real life often happens? Maybe it is time for us to model such a phenomena with negative impact on the brain expressing its abnormal functioning into the human behavior.

Prions are infectious pathogens principally composed of abnormal forms of a protein encoded in the host genome. For many years, scientists referred to this infectious pathogens as slow viruses, but the term prion was specifically introduced to distinguish the pathogen from viruses and viroids after it was discovered that preparations enriched for scrapie infectivity were not attenuated by procedures that modify nucleic acids and inactivate viruses but were inactivated by treatments that alter proteins. The human *PrP*

gene, designated *PRNP*, has been found to contain mutations in families with inherited prion diseases. Genetic linkage studies have demonstrated a causal relationship between five different *PRNP* mutations and central nervous system degeneration. The cellular protein, encoded by the chromosomal gene *PRNP* and denoted *PrPc*, is a membrane protein bounded to the external surface of neurons membranes by a glycolipid anchor. Prions are composed exclusively of a *PrPc* isoform, called *PrPsc*. As the highest levels of *PrP* gene expression are found in neurons, *PrPsc* accumulates in structures believed to be secondary lysosomes containing myelin figures (eg. layers of phospholipid-rich membranes of neurons). The *prion concept* explains how a disease can be manifested as a sporadic - which is a very rare event, heritable as well as an infectious illness. We may say that the hallmark common to all of the prion diseases is that they involve aberrant metabolism of the prion protein resulting in the exponential accumulation of the prion protein into the neural cell after the first step of a post-translational conversion of *PrPc* to the seed of *PrPsc* (first protein entered into the cell) appears to be obligatory (Prusiner, 1991).

Genetic predisposition to infection induce the framework of the system behavior. The genetic predisposition must be translated by the necessity of the aminoacids coding sequence (into a protidic chain) homology of the normal cellular protein of the host and the pathogenic form in which the prion has been stabilized. The trigger factor of possible changes into the human neural system behavior - by the subsequent transformations undergoing between *PrPsc* and *PrPc* - is represented by the contamination, equivalent to penetrate the system, of exogenous prion protein. What cause, however, these modifications? The answer is it just seen as *the genetic predisposition factor*. The natural framework can be divided into two distinct categories: one that includes the refractory character of a human host, and the other of propensity to infection. In the first category, either with or without the presence of exogen prion protein inoculated, nothing happens and both the body and the brain will function normally. This will not be considered in our interest of study. In the second case of hosts manifesting the genetical predisposition we find two aspects: if the host has not being inoculated with *PrPsc*, then, again, we will find a normal functioning of the brain, coordinating normal physical and pshycological activities; else in the presence of infectious protein, the host will develop eventually, in time, a prion disease. What triggers in this second case the expression of genetical predispotiotion is one mutation taking place in the DNA sequence in *PRNP* gene of the twentyth human chromosome. There is a finite set of mutations capable of causing the expression of this predisposition to infection such as for each of these mutations, only one strain of prion will be able to manifest the disease. In some sense of natural computing, the whole history of computer science is the history of continuous attempts to discover, study, and, if possible, implement computing ideas or models from the way nature computes. To answer the question already formulated in this article we propose an explanation of why choosing a hybrid system for unconventional biological computation.

2 Why Choosing a Hybrid Approach of Natural Computing?

Both computing inspired by nature (fallowing techniques like evolutionary computation, neural networks, artificial immune systems or swarm intelligence) and the simulation and emulation of nature of various natural phenomena by means of computing using artificial life are main branches of natural computing. The question arising here is which branch our new system belongs to? Or is it that it belongs somehow to both of them? Even at a short glimpse to the biological phenomena we have already presented, we find a new generation of pathological phenomena opening a new chapter in *prion biology*. Why? Because if having its roots in virology, neurology and neuropathology, molecular and cell biology, as well as protein chemistry - have become evident, then the way prions multiply and cause disease opened new windows to biochemistry and genetics.

Remarkably, distinct strains of prions occur despite absence of an agent-specific genome: misfolded proteins themselves may encode strain diversity - with wide implications in biology. Usually, the amino acid sequence of a protein can be easily determined from the sequence on the gene that codes for it. In the vast majority of cases, this primary structure uniquely determines a structure in its native environment. Knowledge of this structure is vital in understanding the function of the protein. A viable general solution to such predictions remains an open problem, but most efforts have been directed towards heuristics that work most of the time. There are exceptions though: **prions**. So what can we say about prediction of *PrPsc* structure which is homologous with the structure of *PrPc*? In evolutionary biology, homology refers to any similarity between characteristics that is due to their shared ancestry. Homology among these proteins and DNA is concluded on the basis of sequence similarity. If two or more genes have highly similar DNA sequences, it is likely that they are homologous. From the genomic point of view, homology is used to predict the function of the *PRNP* gene[1]. In the structural branch of bioinformatics, homology is used to determine which parts of the *PrPsc* protein are important in structure formation and interaction with other protein (*PrPc*). A technique called homology modeling, may use this information to predict the structure of *PrPsc* protein once the structure of its homologous protein, *PrPc*, is known. This currently remains the only way to predict protein structures reliably. Thus a new challenge arises starting with the *hybrid prion neural system* not only to take the neuron itself as a support for computations (and here we have the same ambition as *systems biology* does) but also a synthetic process aimed at creating patterns, forms, behaviors that (do not necessarily) resemble "life-as-we-know-it". Modeling a new system involves the use of simulations of cellular subsystems to both analyze and visualize the complex connections of these cellular processes, but also, following the tradition of neural networks, regarding the process of communication between neurons, we design the system functionality and its configurations of normal and, in time, abnormal behavior. (This will make the subject of a future article.) With *Prion Neural System* new innovating vistas in the natural computing areas are about to be reached by modeling not only the behavior of a healthy neural system, but a more realistic one, a new *hybrid*

[1] In general, if the sequence of gene A, whose function is known, is homologous to the sequence of gene B, whose function is unknown, one could infer that B may share A's function

system for unconventional biological computation. Not using only one of the directions that we already know represent the explicit goal of abstracting a computing model from the structure and the functioning of a living neuron cell as an evolutionary temporal coding computing system characterized by an involution in its functionality and working as a sequential machine at the level of each neuron unit, but in parallel at the level of the hole system. In terms of natural computing, we recreate the artificial system framework into a system supposed to be gifted with the "genetical predisposition", expressed as knowledge of its own neuron anatomical, fisiological and functional structure, transposed in knowledge of its neural states in different moments of time. The time-discrete character of the system behavior will be transposed through a finite set of states, designating each state for each neuron (from the set of all neurons) found in the simulated nervous system, in precisely that moment in time. This will be done by "interrogating" the neural prion system about its normal activity ("health") expressed by sending electrical impulses to all cells in the system. Transposed to the neural system presented below, the fact that a response is generated by one neuron means that the neuronal cell is found in a normal state of functioning. We will consider that a neuron responds to a command/question if and only if, in its turn, it will be capable of firing (sending electrical impulses to all neurons with which it is connected). Negative influences coming from the possibility that a pathogen agent may be inoculated into the system environment, may block the ability of normal functioning, thus inducing a change into the system normal behavior. Let us say that it is found in *an abnormal state of functioning*. This abnormal behavior is not invariant over long periods of time and is found in very different configurations, regarding the changes that it may occurs at system level. We can conclude that corresponding to functionality, at the level of individual neurons, we have only two states. These are the *functional state* and the *not-functional state* (corresponding to the biological *neuron death*). But, at the level of the entire neuronal prion system, we have a finite set of states, corresponding to a continuous evolution (or, actually, involution) in time. In other words, we may say that the dynamics of the virtual evolution in time of our prion system is based on a functional involution (of the neural system) started at the moment of infection with a pathogen agent. So, we are trying to model an evolutionary temporal computing system characterized by an involution in its functionality.

The brain displays such a general behavior that cannot be explained only in terms of neuron interrelationships, but also related to its various anatomical components and their organization. The anatomy determine the behavior of a neuron by dedicating functions to it. The functioning of each neuron assumes chemical, electrical, and informational processing at the same time. One of its most important parts is the membrane, which separates the interior and the exterior both spatially (for the anatomical and physiological point of view) and electrically (regarding its functionality). Sounds familiar in terms of Membrane Computing[2] and somewhat its dual, Brane Calculi[3]? Concluding

[2] Membrane Computing is a framework for devising models with applications based on the fact that many biological processes (especially, at the level of the cell) can be successfully approached in terms of multiset processing

[3] Brane Calculi is working with objects placed on membranes, corresponding to proteins attached to or embedded in the real membranes

this chapter, hypothesis for biological neural systems drove us to a very fruitful interdisciplinary interplay between the experiments done by biologists and all the results obtained so far by means of natural computing which challenged us in trying achieving our goal by designing a more realistic model of a system exhibiting both continuous and discrete dynamic "anatomy" derived behavior of the neurons. Starting with this article, we will use the term of *hybrid dynamic system* to define a new hybrid system that interact with systems biology, neural networks, brain calculi or membrane computing under the framework of DNA computing as the genetic predisposition to infection with a pathogen agent inoculated into the body of a living human host. We benefit this way of encompassing a larger class of systems within the system structure, allowing for more flexibility in modeling the biological dynamical phenomena.

3 Prion Neural Hybrid System

In biological real systems a controversial phenomena of prion infection along with its implication, inspired us in creating a new hybrid computational model of the way that this infectious agent, called *prion*, influence and modify the way that brain express itself in the human behavior. Our system is designed as a hole complex system whose neuroarhitecture and functionality are directly dependent on the anatomical and functional metabolism of each neuronal unit of the assembly. The modeling of the behavior of each of the neuronal units which is directly influenced by a set of chemical reactions (organized into metabolic pathways, in which one chemical is transformed into another by sequences of enzymes) that occur in living neuron in order to maintain life. Before the initiation of the infection with prion pathogen agent, these processes allow the neuron to maintain its normal structure and to respond to its environment.

We model our *neural-like system* structure as parallel distributed communication network of neurons, which relies on both local and global interactions, by placing a finite set of neurons, N, in the nodes of a finite directed graph denoted $N_s = (N, Syn)$ with Syn defining a finite set of directed links called *synapses*. We define $Syn \subseteq N \times N$, a binary relation, such as

$$Syn = \{syn_{ij} = (n_i, n_j) \,|n_i, n_j \in N, i \neq j, i, j \in \{1, \ldots, card\,(N)\}\}\,. \qquad (1)$$

If instead of n_i we have e then the synapse (e, n_j) represents the directed link from the environment to the neuron n_j, respectively, if instead of n_j we have e then the synapse (n_i, e) represents the directed link from the neuron n_i to the environment.

The highly interconnected nature of communications allows neurons to recognize combinations of environmental inputs and also to stabilize the cell signaling pathways by neuronal signaling of spikes transmitted by a neuron through all its synapses. (The cell membrane maintains a voltage difference between the interior and the exterior - *membrane potential*. This is changing if a neuron receives some input via its synapses - manly located at the dendritic tree of a neuron - from other neurons. If enough excitatory input accumulates the membrane potential at the axon hillock eventually reaches a *threshold*. If this event occurs the membrane follows a stereotyped trajectory creating the so called *action potential* or *spike*. This spike travels along the axon to all its terminals which connect the axon to other neurons - *postsynaptic neurons* - via synapses. A

complicated chemical process transforms the spike into a change of the membrane potential of the postsynaptic neuron - *postsynaptic potential* - which comprises the above mentioned input of the next neuron.) The process of neurons communications is called *synaptic transmission*. These processes will be detailed in a future article along with the hole system functionality. As biological central nervous system is composed by different nervous nuclei expressing different voluntary and/or involuntary activities, each of these nuclei being actually an aggregation of neurons, between neurons from the same nervous nuclei (expressing the same functional activities) there is no directed communication (the communication being possible between neurons from different nervous nuclei). Still, the electrical communication between two biological neurons of the same nervous nuclei can be modeled into our hybrid system (there are no arguments that this communication can not be done). A binary relation over $N \times N$ ($\hookrightarrow \subset N \times N$) defines the *reachability* between two neurons. We say that n_j *is reachable* from $\forall n_i$, for $n_i, n_j \in N, i \neq j, i, j \in \{1, 2, \ldots, card(N)\}$ and we write $n_i \hookrightarrow n_j$ if and only if n_i is connected with n_j (there is a sequence of neurons such as there is a synapse from Syn between any two consecutive neurons). The function $\varrho : N \times N \rightarrow \{0, 1\}$, with $\varrho(n_i, n_j) = 1$ if $n_i \hookrightarrow n_j$ and $\varrho(n_i, n_j) = 0$, otherwise, describes the initial *reachability matrix* of the system $R = (r_{i,j})_{1 \leq i,j \leq card(N)}$ such as $r_{i,j} = \varrho(n_i, n_j)$. This leads, though, to a restriction over the system architecture coming from a biological principle on the functionality point of view, consisting in the existence of at least one spare circuit.

At the level of each neuron unit we define its architecture (in order to separate it from the environment surrounding the entire nervous system - like hybrid ensemble and to define the internal compartmentalization) and its biochemistry (its metabolism that coordinates processes and controls the pathways which keep the cell alive; this is done without a high cost of coordination, in a parallel computing manner).

Formally, $N = \{n_1, n_2, \ldots, n_{card(N)}\}$ is the finite set of all the computing neuron-like devices (one may choose $n_i, i \in \{1, 2, \ldots, card(N)\}$ to be an input neuron in the system). For each $n \in N$, we define one *neuron* as a 3-tuple

$$n = (Tar(n), \alpha(n), \beta(n)) \tag{2}$$

where:
- $Tar(n)$ maps to this neuron a characteristic property that includes it in one nervous nuclei. As the system global function is that of controlling the behavior and coordinating the activities of the ensemble, it charges all of its nervous nuclei with activities (each nuclei handles one activity at a time). Each of the nervous nuclei is, this way, labeled with some unique property expressed at the level of its neuron devices found in its component. Suppose we have q nervous nuclei, we will find a set of properties, $\{c_1, c_2, \ldots, c_q\}$ corresponding the q nuclei. We define $Tar : N \rightarrow \{c_1, c_2, \ldots, c_q\}$ such as $Tar(n) = c_j$ with $j \in \{1, 2, \ldots, q\}$, $card(N) \geq q$.
- $\alpha(n)$ defines the tree - like architecture of neuron n. We assume that we have a finite number of r compartments structured as a hierarchical tree arrangement which not only delimit protected compartments as finite spaces, but it also represent supports for chemical reactions of some chemicals embedded inside. We define $\alpha(n) = \langle \alpha_{0,j_0}, \alpha_{1,j_1}, \ldots, \alpha_{r,j_r} \rangle$ the tree architecture, where

– α_{0,j_0} found at the $j_0 = 0$ level of the "tree root" corresponds to the inner finite space of the neuron (containing all the other compartments). This will also be the output region where we will collect the result of the neuron metabolism - like computation, consisting, of some chemical compounds in case of normal activity of n or, in the case of "*prion infection*", into a finite number of the same chemical complex objects - we have already mentioned the pathogen agent that plays the essential role in the final functioning of the hole system. The final prion accumulation (aggregation) inside this region will give the result of the "metabolic" computation represented by its multiplicity;
– $\alpha_{1,j_1}, \ldots, \alpha_{r,j_r}$ are the inner hierarchical arrangement of the neuron such as for $j_1, \ldots, j_r$ natural numbers, not necessarily disjoint, with $j_k \leq r$ for $k \in \{1, \ldots, r\}$, we define the proper depth level of the k - th compartment. We have a maximum of r inner depth levels;

- $\beta(n) = \langle O, T, OC_n, E, O_n, P_n \rangle$ represent the neuron biochemistry that coordinates processes and controls the pathways which keep it alive:

– O is the alphabet of the so called *objects* (to be free of any interpretation in the same manner of membrane computing does), representing chemicals, from the compartments of a neuron cell, swimming in an aqueous solution; the data structure we consider is that of a multiset (a set with multiplicities associated with its elements). $\lambda \in O$ is the empty element;
– $T \subseteq O$ is the alphabet of *terminal* objects (in the case of prion infection, denoted with p, we define the *prion accumulation threshold* the accumulated quantity of prion protein that determines the inevitable death of the neuron cell, a function $\Theta : OC_n \rightarrow R^+$ where R^+ is the set of real positive numbers);
– if we denote by $2^{M(O)}$ the set of all parts of the multiset generated by O and by 2^O the set of all parts generated by O, then we consider the set of all classes of *organic compounds/organic complexes* represented as strings of *objects* from O:

$$OC_n = \bigcup_j \left\{ OC_j | \forall 1 \leq j \leq s, s \leq 2^O - 1, OC_j \subseteq 2^{M(O)} \setminus \{\emptyset\} \right\}$$

OC_j are not necessarily disjoints;
– $E \subseteq O$ is the set of objects placed (or inoculated) in arbitrary many copies in the environment;
– $O_n = \{o_0, \ldots, o_r\}$ with $o_0, \ldots, o_r$ strings over $O \cup OC_n$ representing the initial multisets of objects/objects complexes present in each compartment of $\alpha(n)$ in the *initial configuration* of the neuron n. By the *initial configuration* we define the initial neuronal architecture of n along with the the chemicals and/or the chemical complexes found in each of its compartment regions. Formally, this is represented by $(\alpha_{0,j_0} : o_0, \alpha_{1,j_1} : o_1, \ldots, \alpha_{r,j_r} : o_r)$
– $P_n = P_{b_n} \cup P_{c_n}$ is a finite set of biological processes defined as the union of metabolic process with processes involved in communications among different compartments. The objects (or the object complexes) obtained as results of biochemical reactions will be the subjects transported (by means of communication processes), crossing membranes, from one compartment to another, through protein channels as sites and means of transportation.

The neuron biochemistry (expressed in terms of neuron metabolism) is modeled as a set of *bio-chemical processes* taking place into the biological living neuron $P_{b_n} = (P_0, P_1, \ldots, P_r)$. Each $P_j \in P_{b_n}, j \in \{0, 1, \ldots, r\}$ is, in its turn, defined as a finite set of sub-process of the form $\left\{p_{j,1}, \ldots, p_{j,k_j}\right\}$, for $k_j \geq 1$. These $p_{j,l}$, with $l \in \{1, \ldots, k_j\}$ is ether a bio-chemical reaction rule, ether an $s_{j,l}$ - tuple of bio-chemical reaction rules of the form $\left(r_{j,l_1}, r_{j,l_2}, \ldots, r_{j,l_{s_{j,l}}}\right)$ and $s_{j,l} \geq 1$ - the number of those bio-chemical reaction rules which are specific to each sub-process involved into the neuronal activity at a cellular (or a sub-cellular) level. Because the membranes play a crucial role both in the the cell architecture (from separating it from the environment to defining the internal compartmentalization) and in cell biochemistry, among the processes described we may have processes evolving the membranes delimiting compartments of neurons. In close analogy with what happens in a cell (see, e.g., Alberts et al. (2002)), the bio-chemical reaction rules are applied in a parallel manner, with the objects as reactants of reactions and with the reactions them-selves chosen in a non-deterministic manner. We also define a finite set of processes, P_{c_n}, controlling the pathways of both communications of objects/object complexes trough one region of a compartment, to another and the chemical exchanges between the neuron and the environment. This, called the set of *intra-cellular communication processes*, is composed by two types of processes described as follows:

- $object \nearrow target$, where if for $k_1, k_2 \in \{0, \ldots, r\}$ and $j_{k_1}, j_{k_2} \in \{j_0, \ldots, j_r\}$ we have $object \in \alpha_{k_1,j_{k_1}}$ and $j_{k_1} = j_{k_2} + 1$, then the object will be sent into the surrounding neuronal compartment $\alpha_{k_2,j_{k_2}}$ with the depth level j_{k_2} (or if $j_{k_1} = 1$ then the object will be sent out from the neuron into the environment);
- $object \searrow target$, where if for $k_1, k_2 \in \{0, \ldots, r\}$ and $j_{k_1}, j_{k_2} \in \{j_0, \ldots, j_r\}$ we have $object \in \alpha_{k_1,j_{k_1}}$ and $j_{k_1} + 1 = j_{k_2}$, then the object will be sent into the inner neuronal compartment(s) $\alpha_{k_2,j_{k_2}}$ with the depth level j_{k_2}(or if $j_{k_1} = 0$ then the object will be sent from the environment into the neuron).

A computation halts by reaching a final configuration of the output region α_{0,j_0} and providing a result consisting in a number of chemical compounds. Supposing the case of normal activity of n, in the final configuration we will find the number of chemicals computed by the neuronal device. Else, in the case of "*prion infection*", the final configuration consists into a finite number of the same chemical complex-objects (prions). The multiplicity of the neuron "metabolic" computation expressed in the final prion accumulation (aggregation) inside this region answers, or not, to our question regarding the functionality of each neuron. If the prion accumulation does not rich a specific threshold, the neuron n is still functional and, because all the external stimuli arrived from different others $n_j, n_j \neq n$ neurons such as $(n_j, n) \in Syn$, it will be able to generate a response. Otherwise, explaining in biological terms, the neuronal functionality is abolished and the neuron will no longer live (it dies) being incapable of providing response. As we already mentioned, there may be designed different prion strains for different nervous nuclei targeted. So, at the level of the system, in the case of inoculation with a pathogen agent gifted with the affinity for neurons targets having $Tar(n_i) = c_j$ with $j \in \{1, 2, \ldots, q\}$ for all $n_i \in N$ with $i \in \{1, \ldots, card(N)\}$, the result of the computation in the final system configuration will consist of

– the number of different chemical compounds, for all $n_i \in N$ with $i \in \{1, \ldots, card(N)\}$ such as $Tar\,(n_i) \neq c_j$
– the number of prion protein, for all $n_i \in N$ with $i \in \{1, \ldots, card(N)\}$ such as $Tar\,(n_i) = c_j$

found in the final configuration of each neuron in the system.

Example of a simple Prion Neural System

Lets consider $N_s = (N, Syn)$, with $N = (n_1, n_2, n_3, n_4)$ a prion system of four neurons having $Syn = \{sin_{12}, sin_{23}, sin_{34}, sin_{42}\}$. The initial reachability matrix of our example is $\begin{pmatrix} 0\ 1\ 1\ 1 \\ 0\ 1\ 1\ 1 \\ 0\ 1\ 1\ 1 \\ 0\ 1\ 1\ 1 \end{pmatrix}$. For each $n \in N$ we need to define the system neural structure, by specifying internal architectures and metabolisms, as $n = (Tar(n), \alpha(n), \beta(n))$: (1) $Tar\,(n_1) = 1, Tar\,(n_2) = 2, Tar\,(n_3) = 3, Tar\,(n_4) = 1$. The system is supposed to be genetically predisposed to infection, the prion strain designated for this example was designed to affect all neurons in the system structure for $Tar\,(n) = 3$ (in this example, neuron n_3). (2) For simplifying the firs model example we will define the same tree-like architecture for all neurons $n \in N$ as $\alpha(n) = \langle \alpha_{0,j_0}, \alpha_{1,j_1} \rangle$. (3) Regarding the biochemistry describing the internal processes of each neuron-device, β_n we will commonly define each part of the neuronal devices, except for those ones where some differences may occur. $\beta(n) = \langle O, T, OC_n, E, O_n, P_n \rangle$, where: (i) $O = \{a, b, c, d, p, \lambda\}$ is the alphabet of *objects* representing chemicals from the compartments of a neuron cell - p is the prion (we define the *prion accumulation threshold* $\Theta\,(p) = 3$), c is the cellular normal protein and a, b, d other three different proteins; (ii) the alphabet of *terminal* objects is $T = \{a, b, d, c, p, \lambda\}$; (iii) the class of *organic compounds/organic complexes* represented as strings of *objects* from O: is $OC_n = \{a, b, d, c, p, \lambda\}$ for all $n \in N$; (iv) $p^4 \in E$ is the only chemical complex one throws into the environment of the model, in a concentration of 4 multiplicity units (representing the prion protein that at a desired moment in time will infect some neurons). For avoiding the indeterminism produced by the all the possible different concentrations of prion protein entering into each neuron -and shortening this way our example, we make the choice of exploring only the case of having one concentration of pathogen agent infection unit of prion agent per neuron device. (v) the *initial configuration* of each $n \in N$ is $\left(\alpha_{0,0} : a^2 b, \alpha_{1,1} : c^3\right)$. (vi) $P_n = P_{b_n} \cup P_{c_n}$, where

– for each $n \in N$, the *intra-cellular communication processes* are defined as $P_{c_n} = (c \nearrow \alpha_{0,0}, p \searrow \alpha_{0,0})$
– the definition of the *bio-chemical process*, $P_{b_n} = (P_0)$, differs this time:
  - if $n \in \{n_1, n_2, n_4\}$, the process of coupling of two proteins molecules is modeled as a function $P_0 : OC_n \times OC_n \to OC_n, P_0\,(a, b) = bd$
  - for neuron n_3 we model $P_0 = (p_{0,1}, p_{0,2}, p_{0,3})$ with $p_{0,1}$ defined as the same process of coupling of a and b protein molecules as in the case of neurons n_1, n_2 and n_4, and a function of conversion into a prion protein of the normal cellular protein, defined, here, as $p_{0,2} : OC_{n_3} \times OC_{n_3} \to {OC_{n_3}}^2, p_{0,2}\,(p, c) = p^2$. When the accumulation of the prion protein into $\alpha_{0,0}$ reaches $\Theta\,(p) = 3$, the process $p_{0,3}$ leads to the neuron death making changes into the system structure.

At each step of the computation all the processes/sub-processes are applied in a maximally parallel manner:

Step1. The same set of parallel processes $c \nearrow \alpha_{0,0}$,$p \searrow \alpha_{0,0}$,$P_0\,(a, b) = bd$ applied simultaneously in neurons n_1, n_2, n_4 and in n_3 replacing $P_0\,(a, b) = bd$ with $p_{0,1}$ determines the same

configuration of each neuron in the network: $\alpha(n) = \left(\alpha_{0,0} : abcdp, \alpha_{1,1} : c^2\right)$.

Step2. The same parallel processes $c \nearrow \alpha_{0,0}$ and $P_0(a,b) = bd$ in n_1, n_2, n_4 outputs at the and of this step $\alpha(n) = \left(\alpha_{0,0} : bc^2d^2p, \alpha_{1,1} : c\right)$, for $n \in \{1,2,4\}$; the configuration $\left(\alpha_{0,0} : bcd^2p^2, \alpha_{1,1} : c\right)$, is obtained by the parallel processing in n_3 of $c \nearrow \alpha_{0,0}$,$p_{0,1}(a,b) = bd$ and $p_{0,2}(p,c) = p^2$.

Step3. The process applied in n_1, n_2, n_4 that outputs $\left(\alpha_{0,0} : bc^3d^2p, \alpha_{1,1} : \lambda\right)$ is $c \nearrow \alpha_{0,0}$. In n_3, $c \nearrow \alpha_{0,0}$ and $p_{0,2}(p,c) = p^2$ determines $\left(\alpha_{0,0} : bcd^2p^3, \alpha_{1,1} : \lambda\right)$.

Step4. The computation stops because no rule can be applied in none of the neurons n_1, n_2 or n_4, but in n_3, as the prion accumulation reaches $\Theta(p) = 3$, the neurons dies, leading to a dynamically conformational change into the new system architecture: $\bar{N}_s = \left(\bar{N}, S\bar{y}n\right)$, with $\bar{N} = \{n_1, n_2, n_4\}$ a prion system of three neurons and $S\bar{y}n = \{sin_{12}, sin_{42}\}$. This can be seen by recreating the reachability matrix of the final configuration of the network: $\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$.

4 Conclusions

Beginning with this article, we are trying to create an artificial system gifted with its own capacity of controlling its behavior. Our need of an emerging field expected to have major impact on the future of biological and medical research aims at the system-level of understanding of biological processes, that are not deciphered yet, and thus employing mathematical analysis and computational tools that integrate the information already obtained in experimental biology. For the moment and even in the near future, exploiting a biological area which is not even understood yet, a direction in which many questions have no answers yet, in our opinion has no limitations - at least for the moment.

Acknowledgments. The author wishes to thank professor M. V. Campeanu for all their long discussions and advices on Prion Biology subjects.

References

1. Maass, W.: Networks of spiking neurons: The third generation of neural network models. Neural Networks, Vol. 10, Issue 9, December 1997, pp. 1659-1671.
2. Maass, W., Schmitt, M.: On the Complexity of Learning for Spiking Neurons with Temporal Coding. Information and Computation, Vol. 153, Issue 1, August 1999, pp. 26-46.
3. Maass, W., Natschlager, T.: Associative Memory with Networks of Spiking Neurons in Temporal Coding. Neuromorphic Systems: Engineering Silicon from Neurobiology, World Scientific, 1998, pp. 21-32.
4. Paun, Gh., Rozenberg, G., Salomaa, A.: DNA Computing. New Computing Paradigms. Ed. Springer-Verlag, Berlin Heidelberg, 1998.
5. Paun, Gh.: Membrane Computing. An Introduction. Ed. Springer-Verlag, 2002.
6. Calude, C., Paun, Gh.: Computing with cells and atoms. Chapter 3: Computing with membranes. Taylor and Francis, London, 2000.
7. Paun, Gh.: One More Universality Result for P Systems with Objects on Membranes. International Journal of Computers, Communications and Control, Vol. I, No. 1, 2006, pp. 25-32.

8. Paun, Gh., Perez-Jimenez, M.J.: Recent computing models inspired from biology: DNA and membrane computing. Theoria, No. 18, 2003, pp.72-84.
9. Barbacari, N., Profir, A., Zelinschi, C.: Gene Regulatory Network Modelling by Means of Membrane Systems. Pre-proceedings of the Sixth Workshop on Membrane Computing, Vienna, Austria, July, 2005.
10. Garcia-Arnau, M., Perez, D., Rodriguez-Paton, A., Sosik, P.: Spiking Neural P Systems: Stronger Normal Forms. Fifth Brainstorming Week on Membrane Computing, Sevilla, January 29-February 2, 2007.
11. Ionescu, M., Sburlan, D.: Several Applications of Spiking Neural P Systems. Fifth Brainstorming Week on Membrane Computing, Sevilla, January 29-February 2, 2007.
12. Leporati, A., Zandron, C., Ferretti, C., Mauri, G.: On the Computational Power of Spiking Neural P Systems. Fifth Brainstorming Week on Membrane Computing, Sevilla, January 29-February 2, 2007.
13. Paun, Gh.: Twenty Six Research Topics About Spiking Neural P Systems. Fifth Brainstorming Week on Membrane Computing, Sevilla, January 29-February 2, 2007.
14. Paun, Gh., Perez-Jimenez, M. J.: Membrane computing: Brief introduction, recent results and applications. BioSystems, No. 85, 2006, pp. 11-22.
15. Gutierrez-Naranjo, M. A., Perez-Jimenez, M. J., Romero, F. J.: Simulating Avascular Tumors with Membrane Systems. Electronic format.
16. Morogan, L. M.: A Subliminal Channel in the Framework of P Systems. Infusing Research and Knowledge in South-East Europe, Vol. 2, pp. 79-90.
17. Morogan, L. M.: A model of a one bit subliminal channel in the framework of a tissue-like P system. Workshop on Natural Computing and Graph Transformations, Satellite event to the 4th International Conference on Graph Transformations, Leicester, United Kingdom, September, 2008.
18. DeArmond, S. J., Prusiner, S. B.: Etiology and Pathogenesis of Prion Diseases, American journal of Pathology, Vol. 146, No. 4, Apil 1995.
19. Weissmann, C.: The Ninth Data Lecture. Molecular biology of transmissible spongiform encephalopathies, FEBS Letters, No. 389, 1996, pp. 3-11.
20. Prusiner, S. B.: Molecular biology and genetics of prion diseases. Phil. Trans. R. Soc. Lond., B., No. 343, 1994, pp. 447-463.
21. Dormont, D.: Biology of agents responsible for transmissible spongiform encephalopathies (prions). Revue Neurologique, Paris, No. 154: 2, 1998, pp. 142-151.
22. Prusiner, S. B., Scott, M. R., DeArmond, S. J., Cohen, F. E.: Prion Protein Biology. Cell, Vol. 93, May 1, 1998, pp. 337-348.
23. Weissmann, C.: Molecular Genetics of Transmissible Spongiform Encephalopathies. The Journal of Biological Chemistry. Vol. 274, No. 1, Issue January 1, 1999, pp. 3-6.
24. Jin, T., Gu, Y., Zanusso, G., Sy, M., Kumar, A., Cohen, M., Gambetti, P., Singh, N.: The Chaperone Protein BiP Binds to a Mutant Prion Protein and Mediates Its Degradation by the Proteasome. The Journal of Biological Chemistry, Vol. 275, No. 49, Issue December 8, 2000, pp. 699-704.
25. Campeanu, M. V.: Encefalopatiile Spongiforme Transmisibile. Ed. Viata Medicala Romaneasca, Bucharest, 2001.

Modelling P-systems with Symport/Antiport Rules Using Place Transition Petri Nets

Monica Kere

University of Pitești, Str. Târgu din Vale, No.1,
Pitești, Romania
{monicaiacob11@yahoo.com}

Abstract. Petri nets and p-systems look like having nothing in common. But even if there are many differences between them, Petri nets are useful to describe p-systems. Even if at the first view there are no similarities between p-systems and Petri nets, this paper describes a connection between p-systems and Petri nets. Graphical symbols associated with Petri nets are used to describe the behavior of the p-systems.
There are other articles in the literature with examples of describing p-systems using Petri nets. In this work a p-system with symport/antiport rules is given, using a place transition Petri net. Then it develops an analysis on the properties that are transferred to the p-system from the Petri net that is describing it. A model of the p-system using Petri nets is obtained and studied. Description of an algorithm is done in order to show the similarity of the behavior of the p-system and the Petri net; the Petri nets properties, reachability, boundedness, liveness, reversibility, coverability, persistence and fairness, are defined for the p-systems. The study of the dynamics in Petri nets and p-systems reveals the conclusion that the former is having the same behavior as the one described by the latter.

Key words: Petri nets, token, p-system, symport/antiport rules, reachability, boundedness, liveness, reversibility, coverability, persistence, fairness

1 Introduction

Petri nets were introduced for the first time in Carl Adam Petri paper, in the University of Mathematics and Physics from Darmstadt, West Germany. Petri nets were proposed for a wide variety of applications. This was done because of their generality and permissiveness.

Gheorghe Păun was the one who introduced the concept of p-systems. P-systems represent a computational model that describes in an abstract mode the way of the interactions of the chemical processes of a cell with its membrane.

1.1 Description of Petri Nets

A Petri network is a mathematical representation for the discrete distribute systems. As a modelling language, it can describe graphically the structure of a distributed system

using a direct graph with labels. This network contains nodes which can determine the places, nodes that can determine the transitions and direct arcs that are connecting the places with the transitions.

Definition 1. *A Petri network is a 5-tuple* $PN = (P, T, F, W, M_0)$*, where*
$P = \{p_1, p_2, ..., p_m\}$ *is a finite set of places,*
$T = \{t_1, t_2, ..., t_{m\}}$ *is a finite set of transitions,*
$F \subseteq (P \times T) \cup (T \times P)$ *is a set of arcs,*
$W : F \to \{1, 2, 3, ...\}$ *is the weight function,*
$M_0 : P \to \{0, 1, 2, 3, ...\}$ *is the initial marking,*
$P \cap T = \oslash$ *and* $P \cup T \neq \oslash$*.*

The Petri networks are using graphical symbols to represent states (usually represented by circles), transitions (usually represented like squares) and arrows (represented like arrows) from the states to the transitions and from the transitions to the states $\bigcirc \to \square \to \bigcirc$. The states can be called places or conditions, and the transitions are also referred to as events.

The condition event networks (CEN) represent the base type for the Petri nets. The CEN are formed through *conditions* (states), *events* (transitions) and *connections* (arrows) from the conditions to the events and from the events to the conditions.

A simple extension of the condition event networks is to allow a marking to have more then one token for the conditions. These networks are known as place transition networks (PTN). The conditions are now called *places* and the events *transitions*. In a PTN the places are labelled with a positive number that represent the capacity. This represents the maximum number of tokens that can be in one place. The arrows can be labelled with a positive number representing the weight.

In the colored Petri nets (CPN) there is a difference between the tokens. The term colored is referring to the fact that the tokens are distinct through the value, that is called color. In the CPN any place has associated a colored set that is specifying the set of the colors for that place. A transition may have a sequence of guard expressions that will be evaluated to a Boolean value. The arrows from the places to the transitions and from the transitions to the places are called arcs.

1.2 Description of P-systems with Symport/Antiport Rules

Definition 2. *A p-systems with symport/antiport rules (of degree m) is a construction of the form:*

$$\Pi = (O, \mu, w_1, ..., w_m, E, R_1, ..., R_m, i_0)$$

where:

1. O *is the alphabet of objects,*
2. μ *is the membrane structure (of degree* $m \geq 1$ *labelled with natural numbers 1,2,...,m),*
3. $w_1, ..., w_m$ *are multisets over O and represent objects contained in each membrane 1,...,m in the initial configuration,*

4. *$E \subseteq O$ represents the set of objects in the environment in infinitely many copies,*
5. *$R_1, ..., R_m$ are evolutionary rules associated with each membrane,*
6. *i_0 represents the label of the membrane from μ, that is the output of the system.*

The rules of this system are just symport/antiport rules. In this system the objects can neither be destroyed, nor created, just transported from one membrane to another. The membrane structure is static.

P-systems with symport/antiport rules represent a computational mechanism with m nested membranes that delimitate m regions, and each region contains a set of objects and a set of rules.

The system is having a nondeterministic and maximal parallel evolution. That means that at every step, the actual configuration is changed into a new one by applying of the maximal multiset of rules, in a nondeterministically way.

The formalization of the operations of a p-system with symport/antiport rules, as in [1], can be realized as:

- *(ab, in)* or *(ab, out)* are symport rules, that means that *a* and *b* pass together through a membrane. For the first case *a* and *b* enter into the membrane and in the second case come out.
  This rule doesn't allow *a* or *b* to pass separately through a membrane. For this case it is necessarily another rule like *(a, in)* or *(a,out)* named uniport.
- *(a, in; b, out)* represents an antiport rule, that mean that *a* enters and *b* goes out from the membrane.

The rules from the set R_i are applied for the membrane i. For the case (a, in), the set of objects a enters in the region that is just inside the region in which a was. For the case (a, out), the set of objects a goes out into the region just outside the one it was in. For the case $(a, in; b, out)$, the set of objects a enters in the region just inside the region in which it was and in the same time the set of objects b goes out into the region just outside the one it was in.

In the environment there are objects in infinitely many copies. This is just because in this type of systems the objects are just moved from one membrane to another and this is a method to increase their number for getting multisets as large as possible.

A configuration is represented through a vector with $m+1$ components, the component i is the multiset of objects from the region $i, i = 1, .., m$, and the last component is the multiset of objects from the environment plus the multiset of objects that were sent outside the system, objects different from the ones in E. A transaction represents the passing from one configuration to another by applying all the rules that can be applied in the system at the moment. A computation is represented by a sequence of transactions through the configurations. A computation is successfully finished if it begins with the initial configuration and gets a configuration in which it is not possible to apply any rule for any object. The result of a successfully computation is represented by the objects from the region delimited by the membrane with the label i_0.

The Differences Between P-systems and Petri Nets At the first view, as in [3], there are many differences between the two concepts. More important are:

- Petri nets have no global clock, and the p-systems are synchronous and all regions of the system have a global clock.
- Petri nets have a nondeterministic firing model, but for the p-systems all objects which can evolve should do it, i.e. the choosing of a rule is nondeterministic. The evolution of the system is maximal, because of the maximal parallelism.
- The Petri nets structure is monolithic and static, while the membrane structure of the p-systems is hierarchical.

With all this differences it is not possible for the Petri nets to describe p-systems directly. Even if all the above differences are considered, there are some skillful techniques that can be used to make this modelling possible.

2 The Description of a P-system with Symport/Antiport Rules using Petri Nets

Example 1. Consider the p-system with symport/antiport rules from Figure 1. This system consists of two membranes. For this example the environment is described as follows:

$$\Pi = (\{a, b, c\}, \mu, w_1, w_2, E, R_1, R_2, 2)$$

$\mu = [_1\ [_2]_2,]_1$
$w_1 = \{a^2, b^3, c\}\ \ R_1 = \{r_1 : (b, out), r_2 : (a^3, out), r_3 : (c, in)\}$
$w_2 = \{a, b\}\ \ R_2 = \{r_4 : (ab, in), r_5 : (c, in, a, out)\}$
$E = \{c\}$

In this configuration the rules r_1, r_3, r_4, r_5 are enabled. If one executes the rules from the vector $V = \{r_1, r_1, r_3, r_4, r_5\}$ the new configuration is obtained(a^2c, ab^2c, cb^2) i.e. (objects in membrane 1, objects in membrane 2, objects in the environment).

For modelling this system a place transition Petri net is used. Suppose that the environment contains a finite number of copies of c, a sufficient number for the computations of this system. Suppose that the environment contains 10 copies of c.

Each object was represented as a place and each of the rules as a transition. The places will be graphically represented as cycles and the transitions as rectangles. The number of tokens from one place means the number of appearance of the object. For example, in membrane 1 the evolving of the rule r_1 can be described with a transition r_1. For the rule r_2 three copies of a will be sent, so the weight of the arc will be 3. For the rule r_5 in which c goes inside the membrane 2, and a goes outside, the transition will evolve after the fulfillment of the two conditions (for a and for c). After evolving this transition two objects are moved. The Petri nets obtained is the one from the Figure 2.

2.1 A Generalization of Showing the Correspondence between P-systems and the Petri Nets Used for Description

Because this process has been described in details, [2], [3], here an algorithm for showing the correspondence between p-systems and Petri nets is given. This algorithm gen-

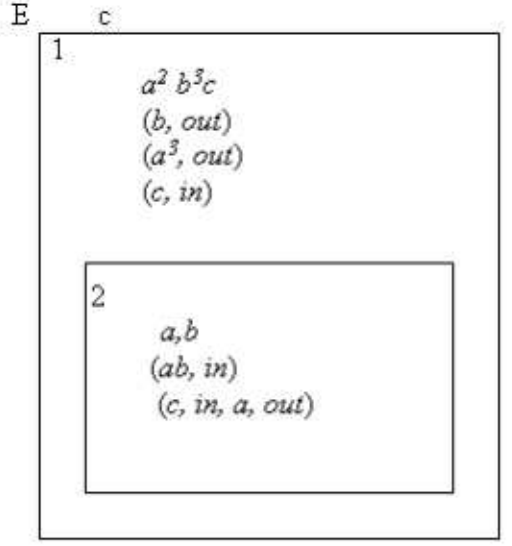


Fig. 1. The p-system with symport/antiport rules

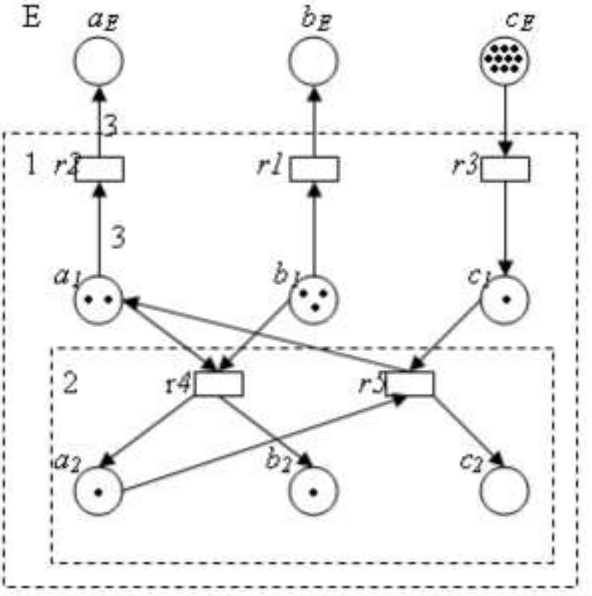


Fig. 2. The place transition Petri net

eralize and schematize the process and it is universal for all non-dynamic types of p-systems. That means it can be used for any non-dynamic type of p-system that it is described with a place transition Petri net. For the dynamic types of p-systems [3], in the algorithm, must be added the steps for the correspondence of the dynamic rules with the transitions in the Petri net. This algorithm is showing how a p-system of any type (without the dynamical types)can be modelled using a place transition Petri net.

Consider a p-system with symport/antiport rules and the Petri net that is used for describing it.

Algorithm:

Step1. Make the correspondence between the objects from each membrane of the p-system and each place of the Petri net;

Step2. Make the correspondence between each rule from the p-system and each transition from the Petri net. The number of copies of an object that are sent coincide with the weight of the arc;

Step3. Make the correspondence between the initial configuration of the p-system and the initial marking of the Petri net;

Step4. Define the relation between the behavior of the p-system and the correspondence Petri net;

Step5. Make the correspondence between the maximum number of steps that are realized for the p-system and the maximum number of steps that are realized for the Petri net.

This algorithm describes the steps to be executed for showing that the p-system and the Petri net that is used for modelling have the same behavior. Each step of the algorithm can be detailed in according with each type of the Petri net used for description.

2.2 The Correspondence of the Petri Nets Properties

To discover the similarities between p-systems and Petri nets several Petri nets properties are studied, trying to formulate them to p-systems.

Reachability is a fundamental property for every dynamic system. This property can be defined for the p-systems as follows: Reachability for one configuration C_n starting from the initial configuration C_0 means to execute the set of transitions (for the computing) to be able to pass from the configuration C_0 to C_n. Because the execution of the rules for a p-system is made in a maximum parallel nondeterministic way, there can be more sequences of transitions for passing from C_0 to C_n.

Boundedness for the p-systems suppose that the number of copies of each object in one region to be less than a number (k). For a p-system to be safe there is necessarily that in every membrane to be at the most one copy of each object. The problem of the p-systems with symport/antiport rules is represented by the environment, in which are objects in infinitely many copies. The property stands because it is given for the regions, that are spaces limited by a membrane. The environment is open, so it can have an infinite number of copies.

Liveness suppose for the p-systems that, without taking into consideration which configuration we get from the initial one, every rule can be realized in one moment in

a computing. This means that in every computing there is a configuration C_i in which that rule can be realized.

Liveness is a ideal property for many systems, thus there are defined some other levels for this property. For the p-systems they are:

- L0-live (dead) if there is one rule r that cannot be executed in any sequence of executions;
- L1-live (potential executable) if r can be executed at least one time in some sequences of executions;
- L2-live if that for any given positive number k, r can be executed at least k times in some sequences of executions;
- L3-live if r appears in an infinite number in some sequences of executions;
- L4-live or liveness if r is L1-liveness for every configuration C from the set of configuration that start from the initial configuration.

Reversibility for p-systems can be defined as follows: Starting from any configuration from the set of configurations that begin with the initial configuration we get the initial configuration. This property, as it is for the Petri nets, can be relaxed using instead of the initial configuration another configuration C' called home state.

Coverability for the p-systems can be defined as follows: Giving a configuration minimum necessary for realizing a rule, that rule is L1-live if and only if the configuration is covered.

Persistence for a p-system suppose that for any two rules that can be executed, the execution of one does not have the effect of not being possible to be executed the other rule. For a persistent p-system, once a rule becomes possible to be executed it is going to remain like this until it will be executed.

For a p-system two rules r_1 and r_2 are in a relationship of bounded-fairness (B-fairness) if the maximum number of times in which one is being executed and the other one is not is bounded. A p-system is a B-fairness system if every two rules from that system are in the B-fairness relationship.

These properties are transferred from Petri nets to p-systems represented with the previous described networks. The p-system is bounded, because there is no region in which the number of the objects to be infinity, more precisely they are 3-bounded. From the point of view of the liveness property, there are no dead rules, so the system is alive (L1-live). The system is reversible, covered and persistent. Taking a look for the Petri net the properties that the net is satisfying are: boundedness (there is no place in which there is an infinite number of tokens), liveness L1-live (there are no dead transitions), reversibility, coverability and persistence.

3 Conclusions

Petri nets are used to graphically describe the structure of the distributed systems which need a representation of their parallel or concurrent activities. It is a modelling language possible to be applied to a variety of types of systems because of their generality and permissiveness.

P-systems represent the description of the membrane cells. A very popular case is establishment by p-systems with symport/antiport rules. This type represents a p-system with pure communication between the membranes and they were related for the first time in [1]. The system is inspired from the biological model, from the cells communication at membranes level.

In this paper a representation of p-systems is made using place transition Petri nets. In this description the rules remain exactly the same. The membranes from the p-systems are not important, they were represented for understanding purposes. The properties for the Petri nets can be described similarly as for the p-systems.

References

1. Păun, A., Păun, Gh.: The power of communication: P systems with symport/ antiport. New Generation Computing, **20** (2002), 295-305
2. Bernardinello, L., Bonzanni, N., Mascheroni, M., Pomello, L.: Modeling Symport/Antiport P Systems with a Class of Hierarchical Petri Nets In: G. Eleftherakis et al. (Eds.): WMC8 2007, LNCS 4860, pp. 124-137, 2007. Springer-Verlag Berlin Heidelberg 2007
3. Zhengwei Qi, Jinyuan You, Hongyan Mao: P Systems and Petri Nets. In: LNCS, vol. 2933, pp. 819-847. Springer, Heidelberg (2004)
4. Bjørner, D.: Software Engineering, 3 vols, Springer, 2006
5. Murata, T.: Petri Nets: Properties, Analysis and Applications, vol 77, pp. 541 - 580, IEEE Computer Society (1989)

Ranking Genes Based on Topics

Nantia Iakovidou

Department of Informatics, Aristotle University of Thessaloniki
54124 Thessaloniki, Greece
niakovid@csd.auth.gr

Abstract. The vast majority of existing methods for ranking gene datasets, do not include or take into account in their exported results other information that might accompany the genes, such as specie or terms. Given a higher order biological data set, we propose a methodology based on multilinear algebra for ranking genes across multiple dimensions. We apply PARAFAC decomposition on a Gene Ontology dataset (GO data) for multiple species and reveal interesting experimental results that provide to the user more information than other consisted methods.

Key words: Parafac decomposition, multi-dimensional biological data.

1 Introduction

Several methods for ranking genes have been proposed in recent years, such as GeneRank [1] and HITS [2] algorithm. All these methods perform ranking across one dimension, that of the totality of the genes and generate only one list of results. In other words they produce one list of scored genes that is consistent to a query gene.

The problem that arises here is how we can rank biological data in which more than one dimensions are involved. This is exactly the case that is investigated in this paper. In particular, we study a drosophila dataset which is a tensor that contains GO terms, species, genes and an integer number that represents the homolog genes within the genome. The question that emerges here is to apply a method so that genes can be ranked across terms and species at the same time. In this way the user would understand better and elicit more information from a higher-order representation of the genes.

The aim of such a task is to analyze genetic diversity and to use this knowledge for discerning the evolutionary relationships among species (i.e. phylogeny reconstruction), comparing different kinds of species and understanding complexities of biological processes (e.g. evolution of genetic regulation).

The solution that we suggest comes from the area of multilinear algebra. The idea that we develop in this paper is to decompose the initial tensor data set to its own distinct dimensions and assign a score to each element. In this way a ranked list of genes will result again, but this time will be followed by its relevant ranked list of terms and ranked list of species at the same time. By using PARAFAC-ALS (Alternating Least Squares) a number of factors are derived and each one of them is associated with a scored list of "hosts" and a scored list of "terms". In this paper by the word hosts we mean genes and

species, while the list of terms is produced from the GO-terms. The results from the aforementioned method are presented and discussed to the related section of this paper.

The remainder of the paper is organized as follows. Section 2 mentions some previous related work. Section 3 describes the method used in this paper. The next section introduces applications and examples that concern the proposed method. Section 5 provides an analytical description of the data set used in the particular study and apposes a discussion about the obtained experimental results. Lastly, we draw some conclusions derived from this work and mention future work that can extend the present paper.

2 Related Work

Authors in [3] describe the Singular Value Decomposition method (SVD) for transforming genome-wide expression data from genes $\times$ arrays space to reduced diagonalized "eigengenes" $\times$ "eigenarrays" space, where the eigengenes (or eigenarrays) are unique orthonormal superpositions of the genes (or arrays). They use this mathematical framework to prove that processing and modelling genome-wide expression data can lead to meaningful results for biology and medicine.

In linear algebra, the singular value decomposition (SVD) is an important factorization of a rectangular real or complex two-way matrix. Applications which employ the SVD include computing the pseudoinverse, least squares fitting of data, matrix approximation, and determining the rank, range and null space of a matrix. On the other hand, in multilinear algebra, there does not exist a multi-way svd that has all the properties of a matrix SVD. A matrix SVD simultaneously computes the orthonormal row/column matrices and a rank-R decomposition. This is generally not possible for multi-way arrays or "data-tensors". Instead, there exist two types of decompositions for multi-way arrays that capture different properties of the matrix svd. One decomposition represents a tensor as sum of rank-1 tensors (Parafac), while the second computes the othonormal spaces associated with the different axes or modes of a tensor. The latter decomposition has been known as HOSVD.

Authors in [4] describe the use of the higher-order singular value decomposition (HOSVD) method for transforming a data tensor of genes into a linear superposition of rank-1 subtensors. By using this framework for analysis of DNA microarray data from different studies, the authors revealed important results about the role of several genes on cell cycle progression.

Kolda et al. proposed a method called TOPHITS which analyzes a semantic graph that combines anchor text with the hyperlink structure of the web [5, 6]. The adjacency structure of the semantic graph is modelled by a three-way tensor containing both hyperlink and anchor text information. Then the authors apply the Parallel Factor (PARAFAC) decomposition, which is a higher-order analogue of the two-way SVD, and produce triplets of vectors with authority and hub scores for the pages as well as topic scores for the terms. This algorithm is an extension of the Kleinberg's HITS algorithm [2] which uses the singular vectors of the hyperlink matrix (a two-way tensor) to produce multiple sets of hubs and authorities.

Several other papers use drosophila datasets and compare GO terms across species as well [7, 8]. Since ontologies are identical, GO terms can be compared across species.

It is worth to mention here that drosophila datasets [9] have been used in very important investigations that study experimental questions such as aging, DNA-damage response, immune response, resistance to DDT and embryonic development [8].

3 Methodology

As mentioned before, in this paper we use a Gene Ontology dataset that contains GO terms, species, genes and an integer number that represents the homolog genes within the genome, which is denoted as frequency. This particular dataset originates from a genus of small flies called drosophila and contains 1473 genes taken from 12 different species of it. It also contains 100 GO terms that are associated with the genes. Since we used a tensor (multidimensional array) to model the aforementioned data it seems inevitable to use in turn multilinear algebra methods that operate tensors, so that we can handle and process the data in a better way.

Here, we focus on PARAFAC (PARAllel FACtor analysis) decomposition method that is common in multilinear algebra. PARAFAC constitutes a generalization of the PCA method to higher orders [10]. In the following, scalars are indicated by lower-case letters, bold capitals are used for two-way matrices and italics capital letters are used for three-way arrays.

3.1 PARAFAC

Parafac is one of several decomposition methods for multi-way data. PARAFAC will decompose a tensor of order N, where $N \geq 3$ into the summation over the outer product of N vectors (a low-rank model). If the order of a tensor is 3 ($N = 3$) then the size of the tensor is for example I by J by K. For instance, given a third-order tensor $X \in R^{I \times J \times K}$ we wish to write it as in (1), where a_{if}, b_{jf}, c_{kf} are elements of the produced matrices **A**, **B** and **C**. Number F represents the number of components of the PARAFAC decomposition [11].

$$X_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} \tag{1}$$

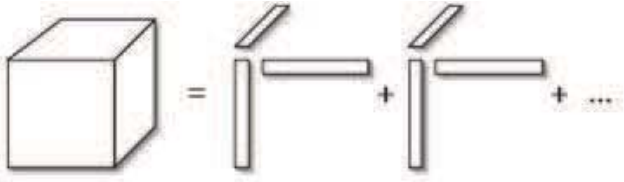


Fig. 1. The PARAFAC method provide a 3-way decomposition that yields terms, genes and species scores.

4 Applications and Examples

GeWare data warehouse system [12] is an application for microarray-based gene expression analysis, which offers high flexibility with multidimensional data models. In these models data is stored in several fact tables which are associated with multiple hierarchical dimensions holding describing annotations (e.g GO annotations) on genes, samples, experiments and processing methods.

GeWare can support algorithms for preprocessing and analyzing multidimensional gene datasets, e.g. to identify lists of interesting genes. The analysis methods are coupled in a simple and powerful way of exchanging experiment groups, gene groups and gene expression matrices.

The GeWare system has been employed in several research projects which study for example the role of the transcription factor IL-6 on the survival of myeloma cells and the factors influencing the binding behavior of sequences on microarrays.

5 Experimental Results

The data set used in this paper [1] is a sparse tensor which contains GO terms, species, genes and an integer number that represents the homolog genes within the genome (denoted as frequency). The size of the sparse tensor is 100 GO terms $\times$ 12 species $\times$ 1473 genes. In general, a GO term consists of a term name (e.g. cell) and a zero-padded seven-digit identifier (or accession number) prefixed by GO: (e.g. GO: 0005623), which is used as a unique identifier and database cross-reference. Species are numbered at Table 1, while genes are represented by a five-digit identifier prefixed by CG (e.g. CG31618).

1	2	3	4	5	6	7	8	9	10	11	12
$dmel$	$dsim$	$dsec$	$dyak$	$dere$	$dana$	$dper$	$dpse$	$dwil$	$dmoj$	$dvir$	$dgri$

Table 1. Drosophila species.

By applying the PARAFAC-ALS decomposition to the initial tensor dataset using the matlab tensor toolbox, three matrices are derived. In each one of them the number of the columns is equal to the number of factors (set by the user) and the number of lines is:

- equal to the number of GO terms for the first matrix
- equal to the number of species for the second matrix and
- equal to the number of genes for the third matrix.

All these matrices represent scored lists of topics and hosts, which are identified with the lists of factors. As mentioned before, in this paper by the word hosts we mean

[1] Downloaded from http://insects.eugenes.org/species/

genes and species, while the list of topics is produced from the GO terms. The results from the aforementioned method are shown in Table 2.

Table 2 presents the parafac decomposition results for the first factor only. It is important to mention here that parafac can compute results, like those depicted in Table 2 for as many factors as the user needs. Here we examine the results of the first factor, since the results as well as the explanation of the rest of the factors is similar to the first.

PARAFAC RESULTS						
——————————————1st factor——————————————						
TOPICS			GENES		SPECIES	
$Score$	$Term$	$GOIDs$	$Score$	$Term$	$Score$	$Term$
0.7123	$nucleosome$	$GO:0000786$	0.5770	$CG31618$	0.5295	1
0.7008	$nucleosomeassembly$	$GO:0006334$	0.5451	$CG31613$	0.3960	5
0.0185	$molecularfunction$	$GO:0005554$	0.4836	$CG31617$	0.3832	3
0.0180	$cellularcomponent$	$GO:0008372$	0.3130	$CG31611$	0.3296	12
0.0132	$biologicalprocess$	$GO:0000004$	0.1133	$CG3379$	0.2898	11
0.0105	$monooxygenaseactivity$	$GO:0004497$	0.0963	$CG13329$	0.2552	6
0.0101	$microsome$	$GO:0005792$	0.0605	$CG5499$	0.2276	10
0.0096	$proteinubiquitination$	$GO:0016567$	0.0383	$CG5825$	0.1986	7
0.0091	$ubiquitin-proteinligaseactivity$	$GO:0004842$	0.0355	$CG7793$	0.1692	2
0.0089	$ubiquitinligasecomplex$	$GO:0000151$	0.0352	$CG3281$	0.1583	9
0.0085	$steroidmetabolism$	$GO:0008202$	0.0352	$CG11290$	0.1133	4
0.0076	$electrontransport$	$GO:0006118$	0.0342	$CG3509$	0.0223	8
0.0017	$peripheralnervoussystemdevelop.$	$GO:0007422$	0.0270	$CG32346$		
0.0017	$SCFubiquitinligasecomplex$	$GO:0019005$	0.0211	$CG8625$		
0.0017	$smoothenedsignalingpathway$	$GO:0007224$	0.0210	$CG5017$		
0.0014	$polysaccharidemetabolism$	$GO:0005976$	0.0210	$CG4236$		
0.0010	$proteinmodification$	$GO:0006464$	0.0206	$CG5330$		
0.0009	$proteinmetabolism$	$GO:0019538$	0.0206	$CG12109$		
0.0009	$transferaseactivity$	$GO:0016740$	0.0187	$CG9383$		
0.0008	$spermatogenesis$	$GO:0007283$	0.0184	$CG3708$		
etc	etc	etc	etc	etc		

Table 2. Ranked list of topics and hosts using the PARAFAC decomposition method.

From Table 2 it is quite obvious that the terms “nucleosome” and “nucleosome assembly” are ranked first in the list of topics and all the rest follow with great divergence. This means that the majority of genes scored on the next column should be mainly or only associated with these two GO terms. If someone takes a look to the initial dataset will see that indeed all the listed genes are associated with either both the first two topics or one of them. Exceptions to this rule constitute the genes CG15440 and CG4299 which, apart from the first topics, are also associated with the terms “protein metabolism” and “spermatogenesis”. The same applies for the list of species, in relation with the other columns. Table 3 reveals the accuracy of the species ranking in Table 2, since it presents the correspondent part from the initial dataset that contains the first two genes and GO terms that are discussed in Table 2. In other words, Table

2 shows that the genes that are highly ranked on the second column are associated with the terms that are also highly ranked on the first column. The same applies with the third column, as well (species column) in relation with the other two columns. In this way, ranking across all these three dimensions is achieved, which was the target of this method. In Table 3 GO terms are indicated by their GO IDs.

Table 3. Part from the initial dataset, before the application of decomposition.

GO IDs	Sp.	Genes	F	GO IDs	Sp.	Genes	F
GO:0000786	1	CG31618	25	GO:0000786	1	CG31613	23
GO:0000786	2	CG31618	10	GO:0000786	2	CG31613	10
GO:0000786	3	CG31618	18	GO:0000786	3	CG31613	16
GO:0000786	4	CG31618	7	GO:0000786	4	CG31613	0
GO:0000786	5	CG31618	22	GO:0000786	5	CG31613	17
GO:0000786	6	CG31618	15	GO:0000786	6	CG31613	14
GO:0000786	7	CG31618	10	GO:0000786	7	CG31613	9
GO:0000786	8	CG31618	0	GO:0000786	8	CG31613	0
GO:0000786	9	CG31618	10	GO:0000786	9	CG31613	7
GO:0000786	10	CG31618	11	GO:0000786	10	CG31613	12
GO:0000786	11	CG31618	14	GO:0000786	11	CG31613	17
GO:0000786	12	CG31618	15	GO:0000786	12	CG31613	20
GO:0006334	1	CG31618	25	GO:0006334	1	CG31613	23
GO:0006334	2	CG31618	10	GO:0006334	2	CG31613	10
GO:0006334	3	CG31618	18	GO:0006334	3	CG31613	16
GO:0006334	4	CG31618	7	GO:0006334	4	CG31613	0
GO:0006334	5	CG31618	22	GO:0006334	5	CG31613	17
GO:0006334	6	CG31618	15	GO:0006334	6	CG31613	14
GO:0006334	7	CG31618	10	GO:0006334	7	CG31613	9
GO:0006334	8	CG31618	0	GO:0006334	8	CG31613	0
GO:0006334	9	CG31618	10	GO:0006334	9	CG31613	7
GO:0006334	10	CG31618	11	GO:0006334	10	CG31613	12
GO:0006334	11	CG31618	14	GO:0006334	11	CG31613	17
GO:0006334	12	CG31618	15	GO:0006334	12	CG31613	20

Sp. stands for Specie and F for Frequency.

6 Conclusions

In cases of multidimensional datasets, existing methods for ranking genes in biological databases cannot help the user analyze and extract useful information from them. We presented a multilinear algebra based methodology which uses the parafac decomposition to rank genes across multiple dimensions of the initial dataset. The proposed scheme can rank GO terms, genes and species at the same time, by providing accurate results that will help a researcher elicit more information from a higher order representation of the data and handle them in a better way.

7 Future Work

This paper can be extended with the insertion of additional similar methodologies to the one presented, as for example the Tucker decomposition method or other. Results from the aforementioned methods can be compared and discussed across species or across multiple datasets.

Acknowledgments.

The following people were kind enough to take a look at earlier versions of this work and offer helpful advice on many fronts: Alexandros Nanopoulos and Yannis Manolopoulos.

References

1. Morrison, J.L., Breitling, R., Higham, D.J., Gilbert, D.R.: GeneRank: Using search engine technology for the analysis of microarray experiments. In: BMC Bioinformatics, vol.6, pp.1–12 (2005)
2. Kleinberg, J.M.: Authoritative Sources in a Hyperlinked Environment. In: Journal of the ACM, Vol.46, No.5, pp.604–632 (1999)
3. Alter, O., Brown, P.O., Botstein, D.: Singular Value Decomposition for Genome-Wide Expression Data Processing and Modeling. In: Proceedings of the National Academy of Sciences 97 (18), pp. 10101–10106 (2000)
4. Omberg, L., Golub, G.H., Alter, O.: A Tensor Higher-Order Singular Value Decomposition for Integrative Analysis of DNA Microarray Data from Different Studies. In: Proceedings of the National Academy of Sciences 104 (47), pp. 18371–18376 (2007)
5. Kolda, T.G., Bader, B.W.: The TOPHITS model for higher-order web link analysis. In: Workshop on Link Analysis, Counterterrorism and Security (2006)
6. Kolda, T.G., Bader, B.W., Kenny, J.P.: Higher-Order Web Link Analy-sis Using Multilinear Algebra. In: ICDM 2005: Proceedings of the 5th IEEE International Conference on Data Mining, pp. 242–249 (2005)
7. Ueda, H.R., Matsumoto, A., Kawamura, M., Iino, M., Tanimura, T., Hashimoto,S.: Genome-wide Transcriptional Orchestration of Circadian Rhythms in Drosophila. In: The Journal of Biological Chemistry Vol. 277, No. 16, Issue of April 19, pp. 14048–14052 (2002)
8. Spellman, P.T., Rubin, G.M.: Evidence for large domains of similarly expressed genes in the Drosophila genome. In: Journal of Biology 1(1):5 (2002)
9. Affymetrix GeneChip Drosophila Genome Array, [http://www.affymetrix.com/products/arrays/specific/fly.affx]
10. Bro, R.: PARAFAC: Tutorial and applications. In: Chemom. Intelligent Lab. Systems., V. 38. pp. 149–171 (1997)
11. Kolda, T.G., Bader, B.W.: Tensor Decompositions and Applications. Technical report Number SAND2007-6702, Sandia National Laboratories (2007)
12. Kirsten, T., Do, H., Rahm, E.: A Data Warehouse for Multidimensional Gene Expression Analysis. Technical Report, Interdisciplinary Centre for Bioinformatics, University of Leipzig (2004)

Service Negotiation Mechanisms in Collaborative Business Processes for Disaster Management

Mihnea Scafeş and Costin Bădică

University of Craiova, Software Engineering Department
Bvd.Decebal 107, Craiova, RO-200440, Romania
{scafes_mihnea, badica_costin}@software.ucv.ro

Abstract. In this paper we present the results of our study of interaction mechanisms between agents involved in collaborative processes occurring in situations of disaster management response. We focus on an analysis of negotiation mechanisms in respect to a set of properties we consider important for service negotiation in critical situations. The analysis consists of: (i) formulation of requirements; (ii) evaluation of three types of negotiation mechanisms against the requirements; (iii) identification of a negotiation mechanism suitable for our problem.
Keywords: Negotiation mechanism, service negotiation, chemical incidents.

1 Introduction

The work reported in this paper was carried out in the context of a project interested in creation of a distributed information system to support population and environment protection against hazards produced by chemical incidents in urban and industrial areas[1]. The system is an agent-based service-oriented infrastructure that will support flexible and distributed collaboration processes between humans and agents engaged in arbitrary problem solving and reasoning tasks, as they are typically required by real situations dealing with disaster management response [1].

Special characteristics make the disaster management domain different from other domains. Among these, we mention: high complexity, the need of a fast and efficient response, collaboration between various organisations, unpredictability. The high complexity of the domain makes the involvement of human experts a must. Agents assist humans in responding to various types of incidents, chemical incidents in our case.

Collaboration between stakeholders during disaster management response is based on interactions mainly directed to information transfer, whenever new information describing the current situation is produced or needed and such that this (request for) information is efficiently routed to interested stakeholders. In our proposal these interactions are the result of providing or running software services. So the key topics that we address in our work include: service creation, service registration, service discovery (including service matching) and service negotiation.

[1]DIADEM project – Distributed information acquisition and decision-making for environmental management: http://www.ist-diadem.eu/

While service registration and discovery problems can be naturally solved using a set of existing software technologies like yellow page agents in multi-agent platforms, many approaches exist for matching service descriptions and service negotiation. In particular, their applicability is heavily dependent on the problem domain – management of chemical incident response in our case. Therefore, the aim of the work is to identify where negotiation can be applied in our problem domain, as well as what negotiation protocols and strategies should be used.

In this paper we will focus on service negotiation mechanisms that can be applied in collaborative business processes between stakeholders during the management of a chemical incident response. These processes are conceptualized as goal-directed distributed workflows that are dynamically created as a response of an incident occurrence, while the threat of a potential disaster is being developed. They are carried out by humans assisted by agents engaged in goal-oriented distributed problem solving. When agents cannot successfully achieve goals by themselves, they need to use capabilities of other agents and must ask these others to adopt their goals – in our approach such a delegated goal is called task. The task contracting process is the subject of a one-to-many service negotiation process that is investigated in this paper.

We are aware of the fact that there are already many studies on negotiation mechanisms. That is why we are first making a survey of negotiation techniques that can be applied in our domain. Among the many negotiation techniques that were proposed in the literature and that were proved to have a practical value, we found relevant for the purpose of our work the following: (i) contract-net protocol with its many extensions; (ii) generalized monotonic concession protocol for multiple agents; (iii) auctions.

We study evaluate these protocols and we try to propose the most appropriate protocol for our work, either by choosing and updating an existing one or by combining features of existing protocols. The theoretical results of our research will be implemented as a negotiation component and validated on real-world usage scenarios of chemical incidents response. We hope that our work will bring noticeable improvements on how a disaster management support system can help human experts in the process of responding to chemical incidents, especially in areas of information flow and task handling.

The paper is structured as follows. We start with related work in the domain of automated negotiation in general and in the field of crisis management in particular, followed by a presentation of the context of our work. We then continue with an introduction into automated negotiation and a presentation of some important results of research in this field. In the next section we identify a set of properties that negotiation mechanisms should have in the field of crisis management and we analyze existing negotiation mechanisms to see if they conform to the properties. Finally, we conclude and point to future work.

2 Related Work

Automatic negotiation has received great attention within the study of multi-agent systems (MAS). The introduction of contract-net (CNET) by Smith [2] has triggered a series of studies of this protocol and the creation of various extensions (for a survey of CNET extensions see [3]). In [4] the authors study the performance of CNET under var-

ious conditions. However, in the original paper of CNET, only the protocol is discussed – the rules for exchanging messages and the negotiation states; there is no study of what decisions should be taken during the negotiation process. In [5], the CNET protocol is applied for web service negotiation. The authors present they work by making use of an insurance scenario. They identify properties of the negotiation subject and negotiation preferences and clear strategies relevant to the chosen scenario.

[6] has a presentation of well known negotiation protocols: (i) *Rubinstein's Alternating Offers* protocol is a negotiation framework in which two agents bargain over an item in turns; (ii) in the *monotonic concession protocol* (MCP), two agents negotiate over a subject in turns and make concessions (a study of this protocol for the multilateral case can be found in [7] and another study for the one-to-many case can be found in [8]); (iii) there is also a simplified version of the monotonic concession protocol, called the one-step protocol.

There are several papers that gather a collection of the existing negotiation methods. In [9], the author discusses, among others, task contracting, coalition formation and argumentation. In [10] the authors discuss various approaches, mentioning heuristics in strategies.

However, choosing the negotiation method depends on the problem domain. There are a couple of papers that present work in the crisis management domain. While negotiation is not emphasized, in [11] the authors present a system for coordination teams of humans and agents for the response to disaster incidents. They mention a situation when there is no need for yellow pages services for service discovery because agents only interact with neighbors. A strategic model of negotiation with relevance to an international crisis (hostage crisis) has been built in [12]. While this work deals with the vast crisis management domain, we are interested in negotiation of services in disaster management, where all agents must cooperate.

3 Context

We are interested in the study of negotiation mechanisms in general and negotiation mechanisms for crisis management (response to chemical incidents) in particular. In this section we make a short presentation of the context that drove this work on negotiation. The DIADEM project aims to be a distributed information system that supports and integrates human experts in the process of responding to chemical incidents. The project is based on a service oriented framework called *DIADEM Process Integration Framework* [13]. DIADEM framework supports complex information workflows composed of processes. Each process is able to: (i) transform and fusion specific types of input information into specific types of output information – this facilitates integration of modules as gas monitoring and detection, multi-criteria decision analysis (MCDA), scenario based reasoning (ii) perform any kind of operation using input data, like automatically executing plans for achieving goals [1]; (iii) transfer control to human experts and support them through user interfaces and integration of other external tools. These processes can be embedded into agents and advertised through a *Yellow Page* type service as *services*.

The DIADEM framework automatically creates communication links between the processes (agents with services) by connecting them using service discovery. Services requiring a specific input data are connected to services that provide as output data that specific input data. Therefore, the inputs and outputs of services are involved in the service registration process. This particular step of communication establishment is of importance to us, as it involves negotiation of services. Whenever a service [2] in the workflow requires data from another service [3] in order to provide its output, a communication link is established between the parent agent and the child agent. From the point of view of the manager agent, the child service is called *task*. The child agent will push the output data provided by its service to the parent agent, thus achieving its task.

The process of contracting a task (the negotiation process) takes care of various parameters, agent-related or service-related. For the rest of this section we will give a concrete example of service negotiation that takes place during the response to a hypothetical chemical incident.

In our previous work (see [1]) we have identified the agents and their goals and plans for a small scenario adapted from the first utilisation scenario in DIADEM User Requirements Document [14]. After an incident emerged and the fire brigade arrived to the location of the incident, it is mandatory that the *Fire Fighter Commander* (FFC) agent marks a safe distance to keep people away from the incident (we assume that: a spill of a chemical substance has leaked from a tank; the substance has already been identified; the substance turns into a gas). The FFC agent cannot compute the safe disance by itself because there are parameters that he doesn't know: (i) how does this gas react to the atmosphere?; (ii) in what measure is this gas harmful to people?; (iii) what are the atmospheric conditions at the location of the incident? Therefore, the FFC agent requires the *Safety distance* service and starts a negotiation for this service with all the agents that are able to provide it.

The location of the incident is important to the negotiation as the safe distance should be computed taking into account the exact location. In many cases the weather conditions cannot be gathered for that exact location but for another location close to the incident. The FFC agent must carefully filter the proposals from other agents taking into account the location parameter. Another important aspect of this negotiation is the precision of the data that will be provided. The FFC agent will surely take this aspect too into account when selecting the service provider agent. Therefore, the FFC agent negotiates about the group of items *(service name, location, precision)*, that is known as negotiation subject (see the next section).

4 Background

Negotiation is the process by which two or more participants eventually reach an agreement over one or more items. Negotiation over a service that one or more participants agrees to provide is called *service negotiation*. We deal with negotiation in multi-agent

[2] the *parent service*, implemented by the *parent/manager agent*

[3] the *child service*, implemented by the *child/contractor agent*

systems and therefore we assume that participants are autonomous software agents. Negotiation theory uses various other theories and methods from mathematics, economics, artificial intelligence and computer science. In MAS, agent preferences are expressed by *utility* functions. Agent's *i* utility function is defined as: $u_i : \mathcal{X} \rightarrow \mathcal{R}$. An agent's utility function is a mapping from a set of states, actions, a set of domain specific attributes or anything else the agent is able to perceive, to a real number. In the negotiation domain, an agent's utility is defined over the set of possible agreements and shows the level of content of the agent in respect to an agreement.

Negotiation in MAS brings together three topics [10] that will be discussed in this paper: *negotiation protocols*, *negotiation subject* and *negotiation strategies*.

A **protocol** is a set of rules that completely guide the negotiation process. This includes description of negotiation states (sending proposal, accepting bids etc), the types of messages that can be sent between participants as well as when to send them and to whom. In other words, a protocol specifies the set of actions an agent is allowed to make when the negotiation has reached a given state. The way agents take decisions is not specified by protocols. Among well known protocols, we mention:

- **CNET**. One agent calls for proposals on an available task and the other agents make bids if they are interested in it.
- **MCP**. Two agents need to reach an agreement. In the first step, each agent proposes the deal it prefers the most. If the proposal of one agent is better than the other agent's current proposal, an agreement has been reached. If not, at the next step each agent has the choice of either making a concession to the other agent or keep his current proposal. Conflict occurs when an agreement has not been reached and no agent accepts to make a concession. An agent makes a concession when it proposes a deal that is prefered more by the other agents than its previous proposal. In [7], the authors introduce a generalisation of MCP for the many-to-many case.
- **auctions**. Auctions are particular negotiation protocols in which agents make bids for tasks and a special agent (*auctioneer*) allocates tasks by evaluating those bids. There are several known auction types like English auction, Dutch auction, first-price sealed-bid and second-price sealed-bid auction [15].

In some situations making proposals and counter-proposals is not enough. Several techniques like argumentation and persuasion have been proposed [10].

A **negotiation subject** describes what is being negotiated between the partners. In service negotiation, the negotiation subject is the service with its parameters - provided, from the contractor's point of view and required from the manager's point of view. One of the preliminary steps of designing a negotiation for a specific problem is to identify the type of the subject: whether it is single-issue (e.g. only the service name) or multi-issue (e.g. service name, time, quality etc). The operations that can be performed over the range of issues must be also identified:

- **fixed-set, fixed-value**: there is an initially identified static set of issues (issues cannot be added or removed from the set) with fixed values. Contractors can either accept or reject the deal. They cannot make counter-proposals because they would change the values in the set.
- **fixed-set, dynamic-value**: the set is fixed but the values in the set can be changed.
- **fixed-set, mixed-value**: the set is fixed, some values in the set cannot be changed and others can.

– **dynamic-set, dynamic-value**: issues can be added or removed from the set, the values in the set can be changed.

Looking back to the example of negotiation we have provided, it is clear that in our domain negotiation over a multi-issue subject is required. In the example we have identified three issues: service name (*Safety distance*), location and precision. In more detail, the subject is of a (fixed-set, mixed-value) type - the set is fixed because the manager agent cannot be attracted by another issue (in the system the reponse to the incident should be directly improved, not an agent's own happiness, although they might be related) and mixed because at least one of the issues cannot be changed: the service name.

A deal that matches the negotiation subject is proposed by the agents following a certain negotiation protocol. They use their utility function to choose deals from the deal space. The utility function may take as inputs the load factor of the agent, the cost in resources of the deal, previous deals, other agent's proposals and preferences (in case of MCP for example, when making concessions to other agents). For the one-to-many service negotiation case, the manager agent must choose the service provider by evaluating the proposed deals. It uses its utility function to compare and rank deals.

A **strategy** is the set of decisions that agents will make during the negotiation. The strategy implemented for a specific negotiation depends on the protocol and subject. An agent applies his strategy when taking actions in the negotiation in order to get the best possible deal. Research in negotiation theory has anayzed possible deals with different outcomes for the agents [6]: *pareto optimal deal* – there is no other deal that all agents prefer it over the pareto optimal deal; the *egalitarian deal* splits the outcome equally between the agents (they have the same utility) and maximizes the sum of utilities for this case; the *utilitarian deal* maximizes the sum of utilities; the *Nash bargaining solution* maximizes the product of utilities. In [7], the author makes a survey of strategies for the case of multilateral MCP by studying how agents propose various types of deals.

An agent can take into account the negotiation history (all the deals that have been proposed) when he applies his strategy. In other words, the strategy maps the negotiation history to a new deal that will be proposed by the agent. Of course that when computing the new deal the agent also takes into account the negotiation subject and the global preferences of the agent (e.g. do not fully occupy resources).

5 Service Negotiation Mechanisms in Disaster Management

We have previously decided that negotiation mechanisms in the system supporting disaster response are of the one-to-many type and require a multi-issue subject. Moreover, they are expected to fulfill the following important requirements:

1. *Negotiation must be guaranteed to terminate*.
2. *Negotiation must always end with an agreement*.
3. *Negotiation must end as fast as possible*.
4. *The outcome of the negotiation must optimize the response to the incident*.
5. *An agreement over all issues in the subject must be reached*.

Any mechanism can fulfill Requirement 5 if the agents are forced to implement a utility function that depends on all subject issues. The manager will always choose the most

appropriate bid that takes into account all issues. For this reason we will not discuss this requirement.

Normally, agents participating in an incident response have a utility function that takes into acount the optimization of the overall response to the incident. Agents must cooperate in order to get the best possible deal for themselves (increase their utility function) and therefore contribute to the optimization of the reponse to the incident. In other words they should search for pareto optimal deals. As resolving this problem involves design of utility functions, we will not discuss Requirement 4 either.

For the rest of this section we check how three known protocols (or group of protocols, in case of auctions) follow these requirements.

CNET

Requirement 1. CNET negotiations are guaranteed to terminate because of the *expiration-time* field in the *task-announcement* message. If the task has not been awarded until the expiration time, the task announcement is not available anymore and the manager must decide what to do.

Requirement 2. Although a CNET negotiation will always end, it is not sure whether it will end in an agreement or in conflict. If the allocated time expires and no suitable contractor has been found, the author of [2] proposes to: (i) award the best bid; (ii) retransmit the task announcement, immediately or after a time delay (iterative CNET). The first option would be appropriate in case possible contractors have been found (agents that are able to provide the service) and their offers can be taken into consideration at least for some preliminary results, while the second one can restart the negotiation in case there are no possible contractors or there are no acceptable offers (conflict). This assures us that CNET will always end in an agreement when there is (or will be, later when the negotiation will be restarted) an agent able to provide the service.

Requirement 3. If there are possible contractors and at least one of them proposes exactly what the manager wants, a CNET negotiation terminates quickly (right after the manager receives the first ideal bid). If there are no possible contractors or the manager simply cannot chose an existing one the negotiation will restart and will take longer. In such a situation, while a negotiation takes place, a new agent that provides exactly the required service could enter the system but it will not be taken into account until the negotiation restarts. A solution to this problem would be to force the manager to continuously send the task announcement. Agents already engaged in negotiation would simply ignore it. The manager agent could use focused addressing to avoid sending useless task announcement when there are no possible contractors. When it requires a service, the manager countinuously searches for possible contractors by querying the YP service and adds them to the negotiation by sending them the task announcement. This is an improvement over the previous solution. Another interesting solution would be to collaborate with the YP server. First, the manager agent tells YP he wants to be continuously informed of all agents that provide the required service (subscribe to YP for service notifications). Then it starts the negotiation by sending the task announcement to existing agents that are able to provide the service. When new agents join the system and register their service at the YP server or when they leave the system, the YP checks their services and informs the subscribed managers accordingly. The managers simply send the current task announcement to the new agents.

MCP

Requirement 1. In case of a finite deal space, 2-agent MCP, multilateral MCP and one-to-many MCP do terminate [6], [7], [8].

Requirement 2. While it will always end, MCP's final result is either an agreement or a conflict. In [8] the authors show that multiple agreements can be reached in the one-to-many MCP and in this case the manager can choose the best one. Knowing that a negotiation will end in an agreement does not mean that the agreement is also known. Even if the agreement can be computed by taking into account the agents' preferences (the best possible deal can be proposed by the agents in the first step – the one-step protocol), this requires knowledge about all the agents' preferences and might also require intensive computations, leading to a centralized approach, which is not desired in MAS. MCP already has a drawback in the fact that utility functions must be publicly available to agents for concession computation.

Requirement 3. Another drawback of MCP is that it can take many steps until an agreement is reached. Forcing the agents to make better concessions leads to elimination of steps from the negotiation, and thus to better negotiation times. One way to impose such concessions would be to make agents aware of negotiation time by adding time as a variable (discount factor) to their utility function. As time passes, agents get worse utilities and they will try to reach an agreement as soon as possible.

Auctions

First-price sealed-bid is similar to one step CNET. The auctioneer (manager) calls for proposals and the participants (possible contractors) submit bids if they are interested in negotiation. If the manager finds one of the bids acceptable, it awards the service contract to the agent that proposed it.

Second-price sealed-bid is unusable in our problem. In a price oriented auction, the auctioneer arranges bids in descending order of the offered prices and the winner agent pays not the price it offered, but the second one. If we used second-price sealed-bid in service negotiation, the autioneer (manager) would arrange received proposals in descending order of the utility score and the winner would contract the service under the conditions proposed in the offer with the second highest utility, this being impossible. Imagine that one agent willing to compute the *Safety distance* for location *X* won the auction but according to the second-price sealed-bid definition, it would have to make the computations for location *Y*, which ranked second in the manager's utility list!

Dutch auction appears similar to one-to-many MCP (in case the manager is the auctioneer) or multiple 2-agent MCP (in case contractors are auctioneers). In the former case, the auctioneer (manager) calls deals for services in descending order of its utility, therefore making concessions that are worse for itself (egocentric concessions), but probably better for the participants and the participants either accept them or not. The first participant to accept a proposed deal is awarded the service. It can be observed that participants do not make concessions, they stick to their internal preferred deal, which is the maximum concession they can make. In the latter case, only contractors make concessions and managers stick to their internal preferred deal.

English auction might look similar to iterative CNET. The first difference is that in English auction the bids are submitted in steps and one bid must be higher than the previous, while in CNET participants send bids in parallel. The condition of increasing

bids can only be respected if participants know each other's bids and this is another difference between the two protocols: in CNET bids are visible only to the manager, while in English auction bids are publicly visible.

6 Conclusions and Future Work

In this paper we analyzed how existing negotiation mechanisms can be used for service negotiation in crisis management, after we had introduced the context of our work and had provided a background in negotiation in MAS. We checked how the mechanisms verify three of five requirements we identified as being critical for our problem. The rest of two requirements impose some restrictions on agents' utility functions and will be studied in subsequent papers.

Preliminary results of our analysis show that: iterative CNET with focused addressing, collaboration with the YP server and selection of bids that are useful for preliminary results follows the three requirements; one-to-many MCP fulfills requiremet 1, implementation of a proper utility function also fulfills requirement 3, but fulfillment of requirement 2 is a challenging task; first-price sealed-bid auction is similar to CNET, English auction has a few similarities to iterative CNET, Dutch auction is similar to MCP in two forms.

Although a negotiation protocol can be easily described at a general level, in practice things get more complicated. In our problem domain the number of possible deals for a negotiation is rather small. Sometimes an agent cannot even make a concession - a static gas detection sensor cannot retrieve data from another location. MCP would not bring any improvement to negotiation processes and probably should not be used if concessions cannot be made. Even if agents can make concessions they must know how to make them, and in order to do that they must know each other's utility functions and the set of deals over which they are defined. In case of one-to-many MCP, the possible contractors only need to know the manager's utility function, but the manager needs to know the utilities of all possible contractors. Making concessions and searching in the deal space is also a time consuming operation that must be avoided when making negotiations in disaster management domain, where a fast response is required. From this preliminary results, we draw the conclusion that using CNET for service negotiation in disaster management is a reasonable choice, but more research is needed for other problems that may appear, such as a contractor agent's commitment to provide the service it has been awarded.

We noticed that there are a lot of negotiation protocols proposed in the agent literature. Moreover, new negotiation protocols are and will be proposed in the context of novel applications. In many cases, theoretical analysis is not sufficient for the study of their qualitative and quantitative properties. That is why negotiation experts need tools to help them define protocols and decision making models, experimentally check negotiations' properties against results of theoretical analysis, gather statistics on various simulated negotiations. We aim to build a negotiation workbench that is able to manage a negotiation described by an arbitrary negotiation protocol. First we must be able to formalize negotiation protocols by establishing a negotiation protocol description language. Second, we want to be able to easily implement various strategies at various

stages in the negotiation process. Therefore, we must successfully merge the protocol description with some decision making models that form the strategy description. And third, we must add to this system a way to define negotiation subjects that will be further used by the strategy models.

Acknowledgments.

Work of Mihnea Scafeş and Costin Bădică was supported by DIADEM project within the framework of FP7 ICT-2007.6.3 call, project reference 224318. We would like to thank all DIADEM team members for their support and fruitful cooperation.

References

1. Scafeş, M., Bădică, C.: Preliminary design of an agent-based system for human collaboration in chemical incidents response. In Ultes-Nitsche, U., Moldt, D., Augusto, J.C., eds.: Proc of 7th Int.Workshop on Modelling, Simulation, Verification and Validation of Enterprise Information Systems (MSVVEIS 2009), INSTICC Press (2009) 53–63
2. Smith, R.G.: The contract net protocol: High-level communication and control in a distributed problem solver. IEEE Trans. Comput. **29**(12) (1980) 1104–1113
3. Bozdag, E.: A survey of extensions to the contract net protocol. (2008)
4. Juhasz, Z., Paul, P.: Scalability analysis of the contract net protocol. In: CCGRID '02: Proceedings of the 2nd IEEE/ACM International Symposium on Cluster Computing and the Grid, Washington, DC, USA, IEEE Computer Society (2002) 346
5. Paurobally, S., Tamma, V., Wooldrdige, M.: A framework for web service negotiation. ACM Trans. Auton. Adapt. Syst. **2**(4) (2007) 14
6. Vidal, J.M.: Fundamentals of Multiagent Systems: Using NetLogo Models. Unpublished (2006) `http://www.multiagent.com/fmas`.
7. Endriss, U.: Monotonic concession protocols for multilateral negotiation. In: AAMAS '06: Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems, New York, NY, USA, ACM (2006) 392–399
8. Bădică, C., Scafeş, M.: One-to-many monotonic concession negotiation protocol. In: Submitted to 4th Balkan Conference in Informatics. (2009)
9. Kraus, S.: Automated negotiation and decision making in multiagent environments, New York, NY, USA, Springer-Verlag New York, Inc. (2001) 150–172
10. Jennings, N.R., Faratin, P., Lomuscio, A.R., Parsons, S., Sierra, C., Wooldridge, M.: Automated negotiation: Prospects, methods and challenges. International Journal of Group Decision and Negotiation **10**(2) (2001) 199–215
11. Schurr, N., Marecki, J., Lewis, J.P., Tambe, M., Scerri, P.: Chapter 1 the defacto system: Coordinating human-agent teams for the future of disaster response (2006)
12. Kraus, S., Wilkenfeld, J.: A strategic negotiations model with applications to an international crisis. IEEE Transactions on Systems, Man, and Cybernetics **23** (1993)
13. Pavlin, G., Wijngaards, N., Nieuwenhuis, K.: Towards a single information space for environmental management through self- configuration of distributed information processing systems. In Hrebicek, J., Hradec, J., Pelikan, E., Mirovsky, O., Pillmann, W., Holoubek, I., Bandholtz, T., eds.: Proceedings of the TOWARDS eENVIRONMENT Opportunities of SEIS and SISE: Integrating Environmental Knowledge in Europe, Masaryk University (2009) In press.

14. Damen, D., Pavlin, G., Kooij, V.D., C., B.C., Comes, T., Lilienthal, A., Fontaine, B., Schou-Jensen, L., , Jensen, J.: (Diadem environmental management requirements document, issue 1.13.1 (2009))
15. Wooldridge, M.: An Introduction to MultiAgent Systems. John Wiley & Sons (2002)

Semantic Web Technologies in Support of Service Oriented Architecture Governance

Dimitrios Kourtesis[1], Iraklis Paraskakis[1] and Anthony J.H. Simons[2]

[1] South East European Research Centre (SEERC),
Research Centre of the University of Sheffield and CITY College
Mitropoleos 17, 54624 Thessaloniki, Greece
{dkourtesis, iparaskakis}@seerc.org
[2] Department of Computer Science, University of Sheffield
Regent Court, 211 Portobello Street, Sheffield S1 4DP, UK
a.simons@dcs.shef.ac.uk

Abstract. As Service Oriented Architecture (SOA) deployments gradually mature they also grow in size and complexity. The number of service providers, services, and service consumers increases, and so do the dependencies among those entities and the various artefacts that describe how services operate, or how they are meant to operate under specific conditions. Appropriate governance over the various phases and activities associated with the service lifecycle is therefore indispensable in order to prevent a SOA deployment from dissolving into an unmanageable infrastructure. The employment of Semantic Web technologies for describing and reasoning about service properties and governance requirements has the potential to greatly enhance the effectiveness and efficiency of SOA Governance solutions by increasing the levels of automation in a wide-range of tasks relating to service lifecycle management. The goal of the proposed research work is to investigate the application of Semantic Web technologies in this context, and propose a concrete theoretical and technological approach for supporting SOA Governance through the realisation of semantically-enhanced registry and repository solutions.

Keywords: Service Oriented Architecture (SOA), Web services, Registry and Repository, Semantic Web technologies, SOA Governance

1 Introduction

Service-oriented computing is emerging as the dominant paradigm for distributed computing and is changing the way software applications are architected, realised, delivered, and consumed [1]. The term Service-Oriented Architecture (SOA) refers to a software architecture perspective where nodes on a network make computational resources available to other network nodes in the form of services. At a conceptual level, there exist three types of actors within a SOA: service providers, service consumers, and service brokers [2]. The prevailing approach for realising SOA in the software industry today is through Web services technology, primarily due to the way in which Web services naturally implement the SOA philosophy of loose coupling

and reusability, and promote interoperability, by leveraging widely accepted XML-based standards such as WSDL, SOAP, and UDDI.

As with any type of software artefact, a service also abides to some form of a lifecycle. Each lifecycle stage is associated with a multitude of service-related activities which are performed by (and may be of interest to) different SOA stakeholders, such as enterprise architects, developers, managers, and others. As SOA deployments gradually mature they also grow in size and complexity [4]. The number of service providers, services, and service consumers increases, and so do the dependencies among these entities. In addition, the number of dependencies among services and the various artefacts that describe how services operate (e.g. interface descriptions, technical documentation), or how services are expected to operate with regard to a specific consumer's requirements (e.g. Service Level Agreements, QoS monitoring logs) also increase at an exponential rate. In the light of these facts, vendors and practitioners started realizing that without appropriate governance over the various phases and activities associated with the service lifecycle a SOA deployment can quickly dissolve into an unmanageable infrastructure [5, 6].

SOA governance is now widely recognised as a precondition for the success and long-term sustainability of a SOA deployment[1], and as a major challenge, from both an organisational and a technological perspective. From an organisational point of view, the challenge lies in establishing an effective and efficient scheme for decision-making with regard to SOA governance and in embedding SOA governance activities within the everyday business processes and working practices of the organisation. From a technological point of view, the challenge lies in providing effective and efficient support for the daily activities of SOA stakeholders, such that SOA governance imperatives (e.g. expressed as policies) can be enforced in a transparent and preferably automated way throughout the service lifecycle. Today's SOA software vendors attempt to address the latter challenge through software tools that integrate service registry and repository functions [7]. However, as will be discussed later in this paper, existing approaches suffer from some inherent limitations that prevent them from reaching full potential in terms of effectiveness and efficiency.

In this paper we explore and try to lay down the initial foundations for investigating the application of Semantic Web technologies in the context of Web service management. The introduction of semantics can assist to overcome some of the aforementioned limitations and advance the state of the art in the area of SOA Governance. In particular, the aim of the proposed research work is to put forward a concrete theoretical and technological approach for supporting SOA Governance through the realisation of semantically-enhanced registry and repository solutions.

The remaining of this paper is organised as follows. Section 2 outlines the characteristics of today's approaches for supporting SOA Governance and points out their limitations. Section 3 discusses the potential benefits that Semantic Web technologies can have in this area. Section 4 sketches the aims and objectives of the proposed research project, and Section 5 summarises the main points from the discussion in this paper.

[1] According to Gartner Group, through 2010, the lack of working SOA Governance arrangements will be the most-common reason for SOA failure (0.8 probability) [5,6].

2 Existing Solutions for SOA Governance and their Limitations

2.1 Service Registry & Repository Systems

Service registries are core components within a SOA deployment since they act as service brokers [2]. In an analogy to the way yellow pages are used, service registries allow providers to advertise the availability of their services through some kind of description concerning what a service does and where it can be reached. These descriptions constitute advertisements which allow prospective consumers to discover services matching their requirements. Depending on the registry technology, the format and content of service descriptions can vary significantly. The service description format may be structured (i.e. be machine-readable) or unstructured, and the contained information may range from coarse-grained to rather fine-grained. This also holds for the way in which service consumers describe the characteristics of the services being requested. Together, the format and content of service advertisements and requests determine the extent to which the procedure of matchmaking between the two can be automated in order to increase effectiveness (in terms of matching precision) and efficiency (in terms of the time required for discovery).

Repositories can be complementary to service registries, as they offer the means for managing the variety of metadata and artefacts that may be associated with the services advertised in registries. As noted above, a single service may relate directly or indirectly with numerous artefacts describing its actual or desired characteristics, and these artefacts may concern different phases and activities within the service's lifecycle. Artefacts may be specific to a service (e.g. comprise functional and non-functional specifications, source code, test data, Service-Level Agreements/contracts, logs, etc), but may also comprise policies and business rules with organisation-wide applicability. Notably, each artefact has its own lifecycle from creation to deprecation. This means that a single change in the state of some artefact may cause significant changes to the states of other artefacts, or to the states of services. Artefacts within a repository should therefore be managed and monitored in a way that allows tracking changes, detecting dependencies, and analysing the impact that a change can have in order to take appropriate measures. As with the service descriptions processed by registries, the format and content of the service-related artefacts determine the extent to which the above functions can be automated.

In summary, the typical functions that an integrated registry & repository tool should support include the following:

- Publishing service advertisements and storing service-related artefacts
- Discovering services that are suitable for reuse based on some search criteria
- Creating and managing contracts (SLAs) among providers and consumers
- Creating and managing policies and associating them with services
- Validating services against policies during design-time
- Enforcing compliance of services to policies and contracts at run-time
- Versioning services and artefacts stored in the registry or repository
- Tracking dependencies among services and artefacts and monitoring change

2.2 Limitations in Today's State of the Art

Today's commercial solutions for SOA Governance support a great number of functions from those listed above. However, there is an important limitation in the way in which service-related artefacts (such as service descriptions, policies, contracts, etc) are represented within these systems, and this essentially prevents SOA Governance solutions from reaching their full potential in terms of effectiveness and efficiency.

The problem lies in that today's SOA Governance solutions rely on service-related artefacts that are encoded based on standards such as WSDL, WSLA, WS-Agreement, WS-Policy, XACML, and others. XML-based standards like those are machine-readable and thus open up the possibility of increasing the levels of automation in SOA Governance, but also have fundamental limitations.

Firstly, their semantics are inherently implicit, and as a result, the information encoded in specifications that conform to these standards is bound to be ambiguous and subject to alternative interpretations. Consequently, this makes the involvement of users necessary for numerous tasks that could otherwise have been fully automated.

Secondly, service-related artefacts lacking formal rigour and machine-understandable semantics are not amenable to processing and reasoning based on formal-logic, which would allow new facts to be inferred from existing knowledge and decisions to be made upon them, thus promoting automation even further.

Thirdly, the information that is encoded in service-related artefacts must be processed in a product-specific manner [8]. The rules by which the information should be interpreted and conclusions should be drawn are not declaratively defined within the artefacts themselves, but is embedded deep within the business logic and source code of the registry & repository tools, thus preventing agility and interoperability.

3 The Potential of Semantic Web Technologies

The lack of machine-understandable semantics within today's Web service specification standards is an inhibitor to providing effective and efficient support for managing the service lifecycle and enforcing SOA Governance requirements. Some even argue that introducing machine-processable semantics in service-related artefacts is indispensable for realising the full potential of SOA [9]. Conversely, we propose that the quality of support that registry and repository tools offer to SOA stakeholders could be significantly enhanced with the incorporation of formal techniques for describing and reasoning about Web service characteristics, through the application of Semantic Web technologies.

Research in this direction, i.e. on combining semantic technologies and Web service technologies, has been mostly taking place under the umbrella term of Semantic Web Services (SWS) [10, 11]. The vision in Semantic Web Services research is to bring formal logic-based semantics into the Web services realm such that the above-mentioned shortcomings of Web services standards can be overcome and service characteristics can be explicated in an unambiguous, computer-

interpretable manner that enables automating a broad range of service design-time and run-time activities. By using formal representation schemes to describe Web service characteristics, service-related artefacts can be automatically processed through logic-based inference and reasoning.

Evidently, the degree of automation that can be achieved depends on the expressiveness and overall capabilities of the semantic representation formalism that is adopted for this purpose. Recent years have seen the development of numerous such formalisms for representing service characteristics, termed Semantic Web Service frameworks [12]. The most prominent Semantic Web Service frameworks, also promoted for standardisation through W3C member submissions, have been OWL-S [13], WSMO [14], WSDL-S [15], and most recently the W3C Recommendation of SAWSDL [16] which evolved from the WSDL-S specification.

Research around these frameworks during the past years has mostly focused on the development of methods, techniques and tools for enabling automated Web service discovery [17, 18, 19], composition [20], and execution [21], while the tasks of service selection [22], monitoring [23], testing [24] and management [25] have received less attention. Recently, researchers have also started to focus on the application of semantic technologies for specifying and managing service contracts (Service-Level Agreements) [26] and service policies [27]. Taken as a whole, the research advancements in these areas constitute a body of work that can serve as the foundation for investigating the application of semantic technologies to the overarching aim of semantically-enriched SOA Governance, through the development of semantically-enhanced registry and repository solutions.

4 Aims and Objectives of the Proposed Research

The aim of the proposed research work is to investigate the application of semantic technologies in the context of service lifecycle management, and propose a theoretical and technological approach for supporting SOA Governance through the realisation of semantically-enhanced registry and repository solutions.

The objectives to be attained in order to fulfil this aim could be formulated as follows:

- Review the theoretical and technological background to the area, focusing on service oriented computing, semantic technologies, policy and compliance management, and existing systems for SOA Governance.
- Analyse key requirements for supporting semantically-enriched SOA Governance, and define them in the form of a generic scheme that can serve as a reference model.
- Investigate methods, techniques and tools that employ semantic technologies to support service design-time or run-time activities, and can be employed for enabling semantically-enriched service lifecycle management.
- Confirm the viability of the proposed approach through the development of a semantically-enhanced registry and repository system prototype that can be shown to satisfy the investigated requirements.

As illustrated in figure 1, there exist three classes of key technologies to be integrated for the development of the envisaged semantically-enhanced registry and repository system.

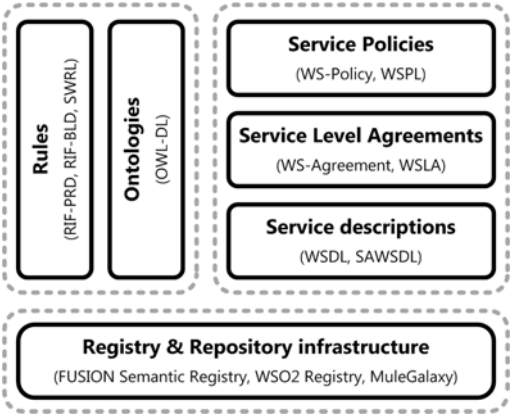


Fig. 1. Enabling technologies for the proposed semantic registry and repository system.

Web Service technology standards: languages and frameworks for the description of service interfaces, service-level agreements and policies – e.g. Web Service Description Language (WSDL), Semantic Annotations for WSDL (SA-WSDL), Web Services Agreement Specification (WS-Agreement), Web Service Level Agreement Language (WSLA), WS-Policy Framework, Web Service Policy Language (WSPL).

Semantic Web technology standards: languages and frameworks for the description of ontologies and rules, including inference rules and production rules – e.g. (Web Ontology Language / Description Logics sublanguage (OWL-DL), Rule Interchange Format Production Rule Dialect (RIF-PRD) and Basic Logic Dialect (RIF-BLD), Semantic Web Rule Language (SWRL).

Registry and repository infrastructure software: open source systems and tools for storing service-related artefacts and managing their lifecycle – e.g. the FUSION Semantic Registry[2], Web Service Oxygen Registry[3], or Mule Galaxy Registry[4].

5 Conclusions

The introduction of automation in Web service management and the support of SOA Governance activities are currently recognized as major challenges in the domain of SOA, as they are instrumental for the long-term sustainability of a SOA deployment. Software vendors have responded with the development of a variety of SOA Governance offerings based on service registry and repository tools. However, the lack of machine-understandable semantics in service-related artefacts (service

[2] http://www.seerc.org/fusion/semanticregistry
[3] http://wso2.org/projects/governance-registry
[4] http://www.mulesource.org/display/galaxy/home

descriptions, policies, contracts, etc) imposes rather restrictive limits to the effectiveness and efficiency such solutions can ultimately achieve.

The incorporation of formal techniques for describing and reasoning about Web service characteristics, based on Semantic Web technologies, can result in significant enhancements in this solution space, can have a positive impact for the industry, and can contribute to the advancement of the state of the art in related academic research areas. Despite its encouraging potential this research topic remains largely unexplored, since the industry has only recently started realising the importance of SOA Governance and since the potential benefits that Semantic Web technologies can bring towards this goal have not yet been fully demonstrated by the Semantic Web Services research community and other related academic communities.

On these grounds, we propose to investigate the application of Semantic Web technologies in the context of Web service management, and to work towards the formulation of a concrete theoretical and technological approach for supporting SOA Governance through the realisation of semantically-enhanced registry and repository solutions.

References

1. M. N. Huhns and M. P. Singh, "Service-Oriented Computing: Key Concepts and Principles," IEEE Internet Computing, vol. 9, no. 1, pp. 75-81, January/February, 2005.
2. T. Erl, Service-Oriented Architecture: Concepts, Technology, and Design. New York: Prentice Hall, 2006.
3. Q. Gu and P. Lago, "A stakeholder-driven service life cycle model for SOA," In Proceedings of the 2nd international Workshop on Service Oriented Software Engineering (IW-SOSWE '07), 2007, pp. 1-7.
4. M. P. Papazoglou and W.J. van den Heuvel, "Web Services Management: A Survey," IEEE Internet Computing, vol. 9, no. 6, pp. 58-64, November/December, 2005.
5. Oracle Corporation, SOA Governance: Framework and Best Practices, White Paper, May 2007.
6. Hewlett-Packard, Transforming Business: Optimizing the Business Outcomes of SOA, White Paper, April 2008.
7. Anne Thomas Manes, The Registry and SOA Governance Market (Burton Group Analyst Report), January 2007.
8. Sun Microsystems Inc, Effective SOA Deployment Using an SOA Registry-Repository, White Paper, September 2005.
9. T. Thangarathinam, "The Roadmap from SOA to Semantic SOA," presented at 2007 Semantic Technology Conference, San Jose, USA, 2007.
10. D. Martin, J. Domingue, M.L. Brodie, and F. Leymann, "Semantic Web Services, Part 1," IEEE Intelligent Systems, vol. 22, no. 5, pp. 12-17, September/October, 2007.
11. D. Martin, J. Domingue, A. Sheth, S. Battle, K. Sycara, and D. Fensel, "Semantic Web Services, Part 2," IEEE Intelligent Systems, vol. 22, no. 6, pp. 8–15, November/December, 2007.
12. L. Cabral, J. Domingue, E. Motta, T. Payne, and F. Hakimpour, "Approaches to Semantic Web Services: An Overview and Comparisons," In Proceedings of the 1st European Semantic Web Symposium (ESWS 04), Lecture Notes in Computer Science (LNCS), vol. 3053, Springer, 2004, pp. 225–239.

13. D. Martin, M. Burstein, J. Hobbs, O. Lassila, D. McDermott, S. McIlraith, S. Narayanan, M. Paolucci, B. Parsia, T. Payne, E. Sirin, N. Srinivasan, and K. Sycara, “OWL-S: Semantic Markup for Web Services,” W3C Member Submission, November 2004.
14. J.d. Bruijn, C. Bussler, J. Domingue, D. Fensel, M. Hepp, U. Keller, M. Kifer, B. Konig-Ries, J. Kopecky, R. Lara, H. Lausen, E. Oren, A. Polleres, D. Roman, J. Scicluna, and M. Stollberg, “Web Service Modeling Ontology (WSMO)”, W3C Member Submission, June 2005.
15. R. Akkiraju, J. Farrell, J. Miller, M. Nagarajan, M.T. Schmidt, A. Sheth, and K. Verma, “Web Service Semantics - WSDL-S”. W3C Member Submission, November 2005.
16. J. Farrell and H. Lausen, (eds.) “Semantic Annotations for WSDL and XML Schema”, W3C Recommendation, August 2007.
17. D. Kourtesis, and I. Paraskakis, "Supporting Semantically Enhanced Web Service Discovery for Enterprise Application Integration," in Semantic Enterprise Application Integration for Business Processes: Service-Oriented Frameworks, G. Mentzas, P. Gouvas, T. Bouras, and A. Friesen, Eds. Hersley: IGI Global, 2009.
18. D. Kourtesis, and I. Paraskakis, “Combining SAWSDL, OWL-DL and UDDI for Semantically Enhanced Web Service Discovery,” in Proceedings of the 5th European Semantic Web Conference (ESWC 2008), Lecture Notes in Computer Science (LNCS), vol. 5021, Springer-Verlag Berlin Heidelberg 2008, pp. 614–628.
19. D. Kourtesis, and I. Paraskakis, “Web Service Discovery in the FUSION Semantic Registry,” in W. Abramowicz and D. Fensel (eds.) Proceedings of the 11th International Conference on Business Information Systems (BIS 2008), Lecture Notes in Business Information Processing (LNBIP), vol. 7, Springer-Verlag Berlin Heidelberg 2008, pp. 285–296.
20. N. Milanovic and M. Malek, “Current Solutions for Web Service Composition,” IEEE Internet Computing, vol. XX, no. YY, pp. 51-59, November/December, 2004.
21. O. Shafiq, M. Moran, E. Cimpian, A. Mocan, M. Zaremba, and D. Fensel, “Investigating Semantic Web Service Execution Environments: A Comparison between WSMX and OWL-S Tools,” Second International Conference on Internet and Web Applications and Services (ICIW '07), p. 31-36, 2007.
22. D. Kourtesis, E. Ramollari, D. Dranidis, and I. Paraskakis, "Discovery and Selection of Certified Web Services through Registry-Based Testing and Verification," in IFIP International Federation for Information Processing, Volume 283/2008, Pervasive Collaborative Networks, L. Camarinha-Matos and W. Pickard, Eds. Boston: Springer, 2008, pp. 473-482.
23. R. Vaculin, K. Sycara, "Semantic Web Services Monitoring: An OWL-S Based Approach," Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008), 2008, pp.313-323.
24. E. Ramollari, D. Kourtesis, D. Dranidis, and A.J.H. Simons, "Leveraging Semantic Web Service Descriptions for Validation by Automated Functional Testing," in Proceedings of the 6th European Semantic Web Conference (ESWC 2009).
25. D. Oberle, S. Lamparter, A. Eberhart, and S. Staab, “Semantic management of Web services,” In Proceedings of the 2005 IEEE International Conference on Web Services (ICWS 2005), 2005, pp. 834-842.
26. A. Paschke, J. Dietrich, and K. Kuhla, “A Logic Based SLA Management Framework,” In Proceedings of the 1st Semantic Web and Policy Workshop (SWPW), 2005.
27. A. Uszok, L.J. Bradshaw, M. Breedy, L. Bunch, P. Feltovich, M. Johnson, and H. Jung, “New Developments in Ontology-Based Policy Management: Increasing the Practicality and Comprehensiveness of KAoS,” In Proceedings of the 2008 IEEE Workshop on Policy, 2008, pp.145-152.

Is Service-Oriented-Architecture Governance a barrier-remover methodology for Healthcare Information Systems Adoption?

Konstantinos Koumaditis

Department of Technology Education and Digital Systems, University of Piraeus, Piraeus, Greece
kkoumaditis@yahoo.com

Abstract. The purpose of this paper is to provide a descriptive study of Service-Oriented-Architecture (SOA) Governance application in the field of Healthcare Information Systems (HIS) adoption. This analysis represents the alignment of a new and cutting age methodology to an existing and growing necessity of integrated Healthcare systems. The study examines Enterprise Integration Architecture (EAI) adoption in the Healthcare sector and introduces SOA Governance methodology solutions to overcome the barriers identified in this process. The research was based on a thorough critical examination and analysis of earlier studies in the field of EAI adoption in Healthcare organisations. The barriers of the adoption were highlighted and the factors overshadowing them reviled. Furthermore, a set of solutions deriving from the SOA Governance methodology applied and explained as means to sustain a better balance between the potential of EAI technology capabilities and the actual benefits of such implementation. The paper provides guidance to IT practitioners in the Healthcare sector to overcome the barriers affecting the adoption of EAI technologies and introduces the new and promising SOA Governance concepts.

Keywords: SOA governance, EAI adoption, Healthcare Information Systems.

1 Introduction

1.1 Background

Today's Healthcare IT organizations are challenged to manage a growing portfolio of system solutions. Healthcare organizations invest significant resources in the development of large and complex Healthcare Information Systems (HIS) that must be modified and extended to respond to changing requirements [1,2]. The cost of acquiring, integrating and maintaining these systems is rising, while the demands of system users are increasing with a significant trend towards web-based applications [3,4].During the past decade, healthcare providers have begun to offer a wide variety

of e-health services to patients and other consumers. These include informational services such as online access to encyclopedic health information and personal health records [5] as well as transactional services including electronic billing and payment [6], computer-mediated communication with physicians and clinical staff [7], interaction with peers and mentors in online support groups [5], and public health reporting [8]. Past HIS adoption efforts have resulted in the employment of disparate, incompatible and heterogeneous systems based on the traditional transaction processing paradigm rather than on supporting business processes. Thus, the communication and collaboration is a challenge faced today by healthcare organizations [1,2]. Moreover, the systems are not only needed to work within the organisation but also accessed outside. The burden of integration usually falls on the users of the system, who are forced often to access many different systems to complete one task.

The research that has been published in HIS adoption remains limited and has mainly focused on Integration Technologies and especially factors (e.g. benefits, barriers etc) [9] that influence the decision making process. Although, the implications of HIS adoption have yet to be assessed, leaving scope for timeliness and novel research. The main contribution of this paper is the identification of the barriers associated with the adoption of HIS and the use of a Service-Oriented-Architecture (SOA) as a barrier-remover.

In this paper, Section (1) is introducing the necessity of Integration systems in healthcare organisations, section (2) analyses the SOA approach to HIS, explains why use Barriers from EAI to understand SOA Governance adoption, section (3) consist of the background research and in more detail of the lack of Governance and SOA Governance, Section (4) presents the conceptual model, and finally section (5) presents the conclusions and future work plans.

1.2 Enterprise Application Integration

Enterprise Application Integration has emerged to piece together inter- and intra-organisational systems and combines a variety of integration technologies such as web services, message and process brokers [10]. It aims at building an integration infrastructure by bridging existing heterogeneous and incompatible applications. EAI refers to the integration of existing Information Systems (IS) rather than the development of new, affects multiple IS and users in organisations, and lacks of established development methodologies [11]. Thus, it is important to understand and analyse these issues before adopting an EAI solution.

1.3 Healthcare Information Systems Integration

Information Systems implementations have been reported to provide significant benefits to healthcare organisations, such as better data quality, availability, accessibility, connectivity, exchange and sharing [14]. Despite the potential benefits of HIS, it has been reported that healthcare systems have a low success rate [15]. Although, the benefits of such systems are evident to healthcare organisations, there

are still many problems relating to their adoption [16]. Among other problems the cost of the development of healthcare integration approaches is high and the level of interoperability remains low [17].

The normative literature has focused on the identification of factors, such as benefits, barriers and costs, affecting the integration technologies adoption[12]. In this paper, due to the research limitations, the authors focus on the analysis of the barriers affecting the EAI adoption in healthcare organisations, as this is a first step of many. Khoumbati et al., [13] has recognized, analysed and identified the barriers of these integration technologies.

The adoption of EAI has been proved to cause barriers to healthcare organizations. These barriers are considered before the organisation proceeds to the adoption of EAI. Khoumbati, [13], classified the barriers in the following: (a) Organisational, (b) Managerial, (c) Strategic, (d) Technical and (e) Organisational. Moreover, Khoumbati, [13] identified and evaluated the most important barriers affecting the EAI adoption in healthcare organisations. Having that in mind Mantzana and Themistocleous [9] developed a new portfolio of barriers and classified them using the same taxonomy evaluated by Themistocleous and Irani [18]. These are displayed in Table 1.

Table 1. EAI Adoption Barriers – Source: Mantzana and Themistocleous [9].

Barriers	Sub-Factors
Operational	Data authentication and consent issues Lack of specific security policy Design of clinical process Lack of common integration standards Lack of EAI skills
Managerial	Lack of standard medical terminologies Resistance to share information Lack of security rules Security and confidentiality concerns
Strategic	Cultural issues Lack of awareness of technology Lack of clinicians' and GPs' willingness Lack of communication between NHS and suppliers
Technical	Lack of EAI adoption benefits' realisation Difficulty in migrating from paper-work to electronic Physicians concerns about monitoring of practices as medical-legal issue
Organisational	High cost of security measures Threats from hackers Loss of autonomy fears of physicians

It is evident, that the barriers hold only a percentage of the factors that influence the adoption of HIS, but due to the scale and timeframe of this research paper they will form the base of the research.

2 Service-Oriented Architecture

2.1 SOA in Healthcare Information Systems

SOA defines a service as an independent unit of work that is self contained and has well-defined, understood capabilities. A unit of work may be an entire process, a function supporting a process, or a step of a business process. With SOA, services directly support business processes as they are "discovered" and orchestrated as a system solution. An interesting opportunity for applying SOA is to increase reuse and standardization that are provided by those functions that are used across systems, departments, and organizations. If system functions are redundant across systems, then the corresponding business processes are likely related and may indicate the need for process sharing as services. SOA provides an environment in which functions can be standardized and used across systems and processes. As SOA is further adopted by the healthcare industry, collections of services as well as specific services will be available for adoption and use, as seen in Figure 1.

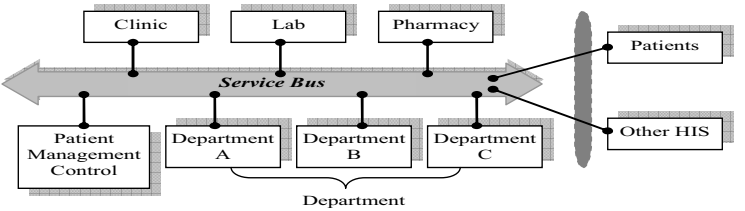


Fig. 1. Service Bus Architecture.

With SOA, system processing is organized and represented as a set of services. Each service is made available to the entire organization through a standard interface. All departments that maintain or access the same information use the same service, making any data and processing redundancies transparent to users. Applications supporting a specific workflow reference one or more services, and each service communicates with the systems to which it is related. Users no longer need to switch between systems to complete a workflow and data is naturally synchronized across processes and supporting systems. Orchestrated services aligned with user workflows enable true interoperability among the healthcare organization's processes and people.

SOA is the next step of system evolution. It builds upon previous architecture approaches while better addressing agility and effective reuse across and outside the organization. SOA provides true interoperability. Most healthcare organizations have a large portfolio of systems with redundant processing and data. SOA allows system

capabilities to be selected and packaged as services that are better focused and available across the entire organization. Organizations can shift their efforts from maintaining a complex data interface strategy to creating service-oriented applications that support interoperability while more closely aligning with healthcare processes.

Data integration and interoperability is a key requirement in healthcare. Medical errors that cost and lives are most often the result from the lack of the right information being available in the right form at the point of care [13]. Worldwide, solving this problem is a key focus area for many governments and associated healthcare institutions. Healthcare Information Systems are the means by which the data integration problems are being addressed. A HIS as a collaboration among the government, hospitals, specialty labs and pharmacies, as well as insurance agencies (payers) can provide a network of data exchange that builds a shared information pathway, data repositories, and application interfaces to rapidly and accurately exchange key health information across a system of healthcare. Figure 2 gives a graphical depiction of the architecture of an SOA-based HIS.

HISs are put in place to support the following main usage models:

- Exchange of patients' electronic medical record between one care provider and another,
- Exchange of referrals between primary and secondary care providers or labs as well as the medical results of those referral visits,
- Electronic pre-authorization of treatment so that it is known quickly whether a treatment or drug is supported by the patient's insurance plan,
- Electronic claims filing and payment to increase the accuracy and speed the cash flow cycle of medical care,
- A means to electronically order and monitor consumption of prescriptions,
- A consolidated data repository of key healthcare information for legally mandated bio-surveillance activities,
- A portal for the patient and healthcare stakeholders means for accessing appropriate data.

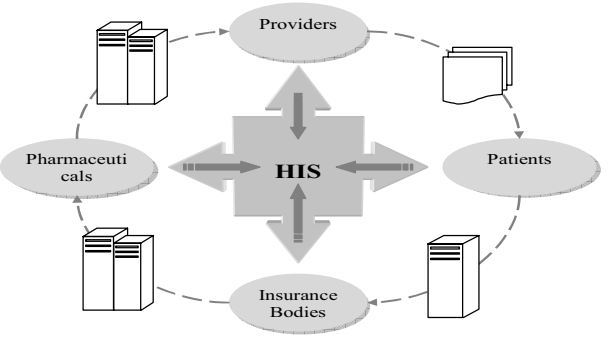


Fig. 2. Healthcare Information System stakeholders.

Using the enterprise service bus technology, both the HIS and each participating organization in the Health Information Network can publish and consume each others services and establish orchestrated workflows to rapidly support new business transactions and interactions among network participants. Additionally, the service

container construct on the service bus architecture allows for existing, in-place clinical and administrative systems within a hospital, lab, pharmacy, or insurance organizations to be "faced" with XML web services and participate in this architecture. This allows for realizing the benefits of SOA in an incremental and iterative fashion thereby leveraging existing technology investments. Figure 3 is a diagram of the service bus architecture.

As seen, using SOA techniques can reduce the costs of implementing a HIS in any scale. Also, SOA delivers features to the community of care as software services that provide a source of ongoing value beyond hosting a simple portal and database of integrated data records on patients.

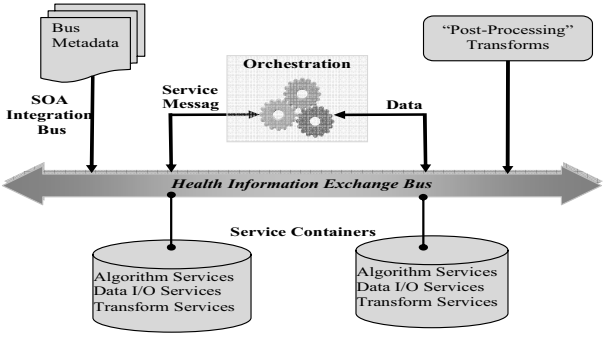


Fig. 3. Service-Oriented Architectur Service Bus.

3 Governance

3.1 Lack of Governance

The primary strategic motivation for most SOA initiatives is business agility, which it defines as being able to respond to changes in the business environment quickly and achieve competitive advantage. However, in the absence of an effective, automated governance strategy, SOA can become brittle, rather than agile [19]. By definition, SOA brings a dramatic increase in the number of interdependent moving parts in the systems environment. In turn, an increase in the number of parts is accompanied by an exponential growth in the number and complexity of interdependencies. Uncontrolled SOA allows services to be developed, invoked, and orchestrated at any time into complicated construction. Rather than creating a platform for effective reuse and responding rapidly to business goals and changing market conditions, uncontrolled SOA leads to redundancy in development and lack of visibility into systems that impact key processes [18,20].

Under this scenario health organizations fail to achieve the return on investment they anticipate, or may even spend more time and money over the long term than they

would have under traditional development. IT may see more redundancy of services and greater infrastructure complexity as a result of poor SOA governance and through the uncontrolled growth of services that are that are difficult to locate or inadequately constructed or understood.

Without governance, expanding SOA development will result in:

- A fragile architecture
- Services that cannot easily be reused because they are unknown to developers or because they were not designed with reuse in mind
- Lack of trust and confidence in services as enterprise assets
- Security breaches that cannot easily be traced
- Unpredictable performance
- And, ultimately, disillusionment with SOA or complete SOA failure.

3.2 SOA Governance

The purpose of SOA governance is to align software governance and business governance including coordinating software development, acquisition, and reuse across domains to achieve maximum agility and economies of scale/scope. In other words, SOA Governance can be characterised as an extension of IT governance that focuses on managing services and the related service-level abstractions. SOA Governance recognizes that services are corporate assets that need to be managed throughout their service lifecycle. This is because services provide a common unit of management for requirements, service-level agreements, business and technical performance, and resource utilization.

The SOA governance policies and processes answer basic questions [21] such as:

- Why? The goals of adopting a SOA,
- Who? The stakeholders and participants in the governance process,
- What? Their roles, responsibilities and activities,
- How? The processes, structures, and meetings that need to be organised,
- When? The timing of the activities.

In general the SOA governance policies and processes define and implement the following processes:

- Service lifecycle processes that govern proposing a new service, initial deployment of a new service, versioning a service, and retiring a service,
- SOA decision-making and issue-resolution process,
- SOA design and development process,
- Monitoring the business performance of services,
- Monitoring the technical performance of services,
- Ensuring that Web services standards are complied with and used appropriately.

The SOA governance policies and processes establishes a governing body within the enterprise that has representation from each service domain and the different business units and from subject matter experts who can speak to the key technological components of the solution. This governing body defines the policies and processes for identifying, implementing, deploying, and versioning services. This can range

from a completely decentralized process (where each department or project is responsible for identifying, implementing, and deploying services that it wishes to make available to the rest of the organization) to a centralized process (where this governing body reviews addition and deletion of services, as well as changes to existing services, before authorizing their implementation and deployment). In either case, it is important that there is some level of standardization across departments and projects to promote service reuse. The SOA governance policies and processes identify mandatory and optional training that members of the SOA team and the larger project teams should complete to effectively implement the SOA processes and use the SOA tools. The SOA governance policies and processes identify the tools that the SOA will use for project management, business modelling, service modelling, data modelling, development, system management, and how these tools will exchange models/information.

The majority of the normative literature focuses on SOA governance in private organisations. An example of this is the model proposed by Marks [21]. Marks' model proposes that an enterprise should refer to can be broken down to four primary classes:

Governance Strategy and Policy Dimensions – This class pinpoints the strategy that provides the backdrop and direction for SOA governance. It refers to the strategies and objectives needed to be addressed. Also, it highlights what resources, decisions or processes must be governed to support this strategy and what business, process, and technology policies are necessary based on the stated SOA strategy and governance philosophy.

Organization, Process, and Roles/Responsibility Dimensions – This class corresponds to the organizational models and processes that support the SOA governance model and enforce the desired policies. Furthermore, it refers to the governance boards and committees integrated with governance processes, checkpoints, and reviews. Moreover, it casts roles and responsibilities that are necessary to achieve SOA governance. In other words, who governs what and how?

Behaviour and Metrics Dimensions – The behaviours and reinforcement processes that are needed to institutionalize governance in the organizational fabric, are placed in this class. In more detail, it focuses in the way the metrics and governance performance management support the behavioural model and how they achieve the results that the organisation is targeting.

SOA Governance Management Dimensions - This category includes SOA funding and budgeting, governance performance management, and principle and policy management processes. SOA funding and budgeting is critical because of its direct impact on shaping the alignment of business unit strategy and goals to the enterprise, as well as on shaping the selection and implementation of key programs and projects that implement corporate strategy. Funding and budgeting are the ultimate governance levers.

4 Discussion

In the previous sections of this research paper the importance of SOA Governance in the adoption of HIS was analysed and presented. Taking as a starting point the barriers presented and analysed in section 1.2 and in conjunction with the model of SOA Governance presented and analysed in section 3.2, this research underlines the crucial role of SOA Governance. A theoretical model that highlights the connection of the SOA Governance methodology and the barrier that influence EAI adoption in healthcare organization can be seen in figure 4.

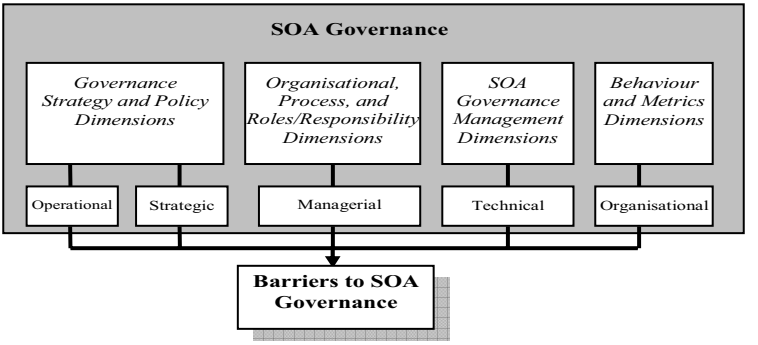


Fig. 4. Barriers to SOA Governance.

In more detail, there is a straight correlation between the EAI adoption barriers and SOA Governance processes. Following the analysis presented in section 2.1, where the advantageous characteristics of implementing SOA architectures to HIS, below a SOA Governance analysis in parallel to EAI adoption barriers is illustrated.

The Governance Strategy and Policy Dimensions class not only pinpoints the strategy and objectives needed to be addressed but also highlights what resources, decisions or processes must be governed to support this strategy and hence reduces the cost to redesign business processes and structure. Furthermore, by position the appropriate business, process, and technology policies at an early stage it decreases the resistance to change and provide ground of cooperation between all partners of the HIS network.

The Organization, Process, and Roles/Responsibility Dimensions class by forming the governance boards and committees responsible for governance processes, checkpoints, and reviews, it identifies in an early stage and resolves the difficulty of finding staff with experience in EAI skills.

The Behaviour and Metrics Dimensions class focuses in the way the metrics and governance performance management support the behavioural model and how they achieve the results that the organisation is targeting and thus resolves cultural and/or political issues that may be influencing the adoption of EAI systems.

The SOA Governance Management Dimensions class critical nature because of its direct impact on the selection and implementation of key programs and projects that implement corporate strategy can resolve problems channelling from the existing systems and other confusing integration technologies.

5 Conclusions and Further Work

The need to improve the healthcare services through the HIS integration has been highlighted and explained in this study. SOA is an emerging technology and although it is widely applied in many sectors, its adoption still holds barriers. Hence, it is evident that this area requires detailed research and utilisation of new approaches. This is of high importance as HIS are critical to human lives and thus information about their integration might be equally significant. In this paper, the author attempted to do so by reviewing the normative literature on barriers affecting the EAI adoption in healthcare organisations and presenting solutions that combine SOA Governance methodology applied as a barrier remover to EAI adoption in the healthcare organisations. The author investigated the barriers related to the EAI adoption and presented and analysed their correlation with SOA Governance.

In doing so, the authors proposed a conceptual model that incorporates barriers reported in previous studies and combines them with the SOA Governance reference model. The proposed model makes contribution at both practical and conceptual level. At a practical level, the model contributes towards a deeper understanding of the SOA Governance application in healthcare. The identification of the barriers is based on the theoretical work conducted by others.

Although the proposed model can be justified from the theoretical background more investigation and validation is needed. A couple of propositions were made for further research. The paper suggests that:

- The additional factors influencing EAI adoption should be included in further research and their connection with SOA Governance identified.
- The model should be validated with empirical data and cases.

One of the limitations of this research is that the model presented herein is based on a theoretical basis. Thus, the conclusion drawn for the importance of SOA Governance methodology as a barrier remover of EAI adoption in the healthcare organisations need to be validated by data and observations. Nonetheless, it is not the intention of this paper to offer conclusive guidelines about SOA Governance importance but rather, describe a new cutting age technology and allow others to relate their experiences to concept reported. Therefore, this paper offers a broader understanding of the integration technologies for healthcare information systems adoption.

References

1. Lenz R. and Kuhn K., “Towards a continuous evolution and adaptation of information systems in healthcare”, International Journal of Medical Information, 73: 75-89. (2004)
2. Huizen Van G and Backman A., “ SOA: First Principles”, SONIC software. (2005)
3. Juneja G. and et. al. , “Service Oriented Architecture Demystified: A pragmatic approach to SOA for the IT executive”, Intel Press
4. IDEA, group Managing Healthcare Information Systems with Web-Enabled Technologies, (2000)
5. Newcomer E. and Lomow G., “Understanding SOA with Web Services”, Addison Wesley Professional December 14. (2004)

6. Altinkemer, K., De, Prabuddha, and Ozdemir, Z. D. Toward a consumer-to-healthcare provider (C2H) electronic marketplace. Communications of the Association for Information Systems 18(19), 413-430. (2006)
7. Wilson, E. V. Asynchronous Health Care Communication. Communications of the ACM, 46(6), 79-84. (2003)
8. Järvinen, O. P. Privacy management of patient-centered e-health. In E.V. Wilson (Ed.), Patient-Centered EHealth.Hershey, PA: IGI Publishing. (2009)
9. Manzana V. and Themistocleous "Benefits and barriers relating to EAI adoption: The case of Healthcare organisation", ACIS, (2006),
10. Linthicum, D. "Enterprise Application Integration", Addison-Wesley, Massachusetts, USA,(1999).
11. Lam, W.. 'Investigating success factors in enterprise application integration: a case-driven analysis', European Journal of Information Systems, 14(2): 175–187.(2005)
12. Themistocleous, M. and Irani, Z. 'Benchmarking the Benefits and Barriers of Application Integration', Benchmarking: An International Journal, 8(4): 317-331. (2001).
13. Khoumbati K., et al., "Evaluating integration approaches benefits adopted by healthcare organisations". ECIS , (2004)
14. Stefanou, C.J . and Revanoglou, A. "ERP integration in a health care environment: A case study", Journal of Enterprise Information Management , Vol.19(1), (2006)
15. Heeks R, Mundy D, Salazar A. "Why healthcare information systems succeed or fail", University of Manchester Web site: <http://www.man.ac.uk/idpm/idpm_dp.htm#isps_wp/>.(1999),
16. Carr and Moore "IHE: a model for driving adoption of standards"
Computerized Medical Imaging and Graphics, 27(3), 137, (2003)
17. Khoumbati, K.. '"Investigating the EAI Adoption in Healthcare Organisations"', School of Information Systems, Computing and Mathematics, Brunel University, London., (2005)
18. Bridges M., "SOA in Healthcare" Health Management Technology (2007)
19. Baer T. "SOA : Building the roadmap" Zapthink White paper, (2007)
20. Josuttis N., "SOA in practise" O'Reilly, 98, (2007)
21. Marks E. (2008), "SOA Governance" John Wiley &sons inc, 5, (2008)

An approach to domain model reverse engineering

Mocanu Mihai

University of Piteşti,
Piteşti, Romania
mocanushtefan@yahoo.com

Abstract. Domain Model is a fundamental notion in software engineering, being a cornerstone in building any non-trivial software application. Despite its importance, the theoretical notions supporting it are close to non-existent, usually domain model being explained through examples rather than definitions. If the notion of domain model is vaguely defined, an attempt to reverse engineer the domain model from the source (or binary) code has never been published, as this is a complicated task even for human operators. In this article, I will try to give an approximation of the domain model based only on the Java source code of an application. The algorithm presented does not fully identify the domain model, but a theoretical basis is established and the way for further refinement of the detection is opened.

1 Examples

Domain model is usually explained through characterization rather than definitions, authors prefer to give examples of such concepts. Actually, domain model is more part of the software development folklore rather than established theoretical concept, art rather than exact science.

Widely used examples of domains are Accountancy, Financial Services, Health Care, Manufacturing, Market, Transportation, and so on. Each of these domains have their own strictly defined terms; moreover, each instantiation of such a domain (for instance an insurance broker from the Financial Services, a hospital from the Health Care domain, a rail company from the Transportation domain) has its own set of rules and relations between its concepts.

For instance, the Accountancy domain usually describe terms like Item, Order, Client, Payment and so on; a Hospital might use terms like Patient, Doctor, Treatment, Receipt, Disease, a rail company might use terms like Track, Lane, Switch, Station and so on.

A software engineer might choose to reflect these concepts into source code as stand-alone entities. In case they are coded as classes in object-oriented programming languages, the classes are called Domain Objects. The set of Domain Objects and the relation between them constitute the Domain Model. Of course, one is not forced to create software constructs to reflect real concepts, but overwhelming success of object-oriented programming languages and the myriad of applications and libraries

developed in this class of programming languages justify the adoption of domain objects.

2 Domain Model

Domain Model concept is in itself fuzzy, as it needs to cover a very broad area of application, making the task of reverse engineering it also a fuzzy task. Empirical evidence points to a set of properties that help constraining the set of Domain Model, which will be further described.

According to Dines Bjørner, a domain is “a universe of discourse (such as an area of human activity - business, industry, service, etc.), characterized by its own set of professional terms, and otherwise pragmatically separable from (but possibly interfaced to (including overlapping with)) other domains”. [1, 2]

Starting from this characterization, a domain model “describes, informally and formally, the domain as it is, without reference to requirements, let alone new software". [1, 2]

There are several approaches in describing the domain model, widely used being through another concept, the Domain Object. The domain objects are direct abstraction at software level of the business concepts the application is meant to represent. In this respect, the domain model is the set of Domain Objects and the relations between them. Although the term “Object” commonly refers to an in-stance of a class in Object-Oriented programming, Domain Objects actually de-scribe classes that part of a certain domain.

Through the article, I am also using another concept, the UML Association, which Robert Martin defines it as “the ability of one instance to send a message to another instance.”[3]

Interpreting the characterizations of the domain object through the directed UML association, I extract a first property of the domain objects:

Property P1.A Domain Object can only have a dependency (in the sense of directed UML association) another Domain Object.

This property is a direct consequence of the domain object characterizations, in particular the fact that domain object do not have references to requirements or other software components. This is a consequence of the fact that domain objects are high-level software patterns, being constructed independent of programming languages and thus offering flexibility regarding implementation.

In modern object-oriented programming, polymorphism is a key concept. It is easy to notice that:

Property P2.If a class is a domain object, then all its subclasses are also do-main objects.

Considering also the above mentioned statements, I am noticing that a class from the Domain Model is either a Domain Object or contains an aggregation [3] of one (or more) Domain Objects, therefore realizing that

Property P3.The Domain Model is included in the closure (with respect to aggregation) of the Domain Objects.

The above property is also a consequence of the previous definitions, and it could be formulated as "domain objects can only aggregate other domain objects, but other classes can also aggregate domain objects". It is indeed common for classes on other layers to make use of domain objects through aggregation, but there are cases, as proven by tests, in which simple association is employed to take use of a domain object's components.

The previous properties are the foundation for the algorithm used to extract the domain objects from the list of an application's classes. It is obvious that domain objects are not the only classes satisfying these properties, resulting in the detection of a superset of an application's domain objects.

3 Algorithm

In the quest for defining an algorithm for detecting the domain model of an application, one has to consider the possibility of cyclic dependencies between domain objects. In real-life applications, it is not forbidden (nor uncommon) to find cyclical dependencies between domain objects, due to the specific of a domain. For instance, one can conceive that, in the domain of accountancy, an Order contains several Items, while an Item is part of one (or more) Orders. This relationship can be modeled in such way a cyclical dependency arises. The idea is to decrease the degree of nodes involved in cyclical dependencies. Using this approach will result in the application reporting even a larger amount of classes as result, but in practice this additional number is not significant.

As a result, an algorithm for approximating the domain model of an application based on the source code is as follows:

1. Create the directed UML association graph
2. Add subclasses into the association graph (Property P2)
3. Eliminate cycles
4. Extract leaves of the resulting graph (Property P3)
5. Output the aggregation closure of nodes resulting at the previous step (Property P1)

Although there could be other chaining sequence of the properties defined previously, this represents the sequence which yields best practical results – deducted on a trial-and-error basis.

4 Testing

In order to test the claims in the article, first I have considered an engineered example: I have implemented a simple application that will allow teachers to store and print the results of students' evaluations electronically – an online catalog. The description is vague, but it allows different implementations in order for the algorithm to be tested in various situations. I have tried to follow the normal process such an implementation requires, but on a more schematic level.

Based on the problem description, I have described some requirements the application must fulfill:

- allow dynamic creation of teachers, students, matters and classes
- allow dynamic addition of students in a class
- allow creation of new evaluations for a specific class and a specific matter. All the students in that class are evaluated for the chosen matter – a student is evaluated with grades between 1 and 10 or marked as absent
- evaluations can be viewed and printed

The next phase is to perform the domain model analysis: starting from the above description and requirements, one isolates concepts – such as Teacher, Evaluation, Enrollment (class), Matter, Result of Evaluation – and implements them as classes in an object-oriented programming languages, in this case – Java. For instance, one could represent these classes as in Fig.1.

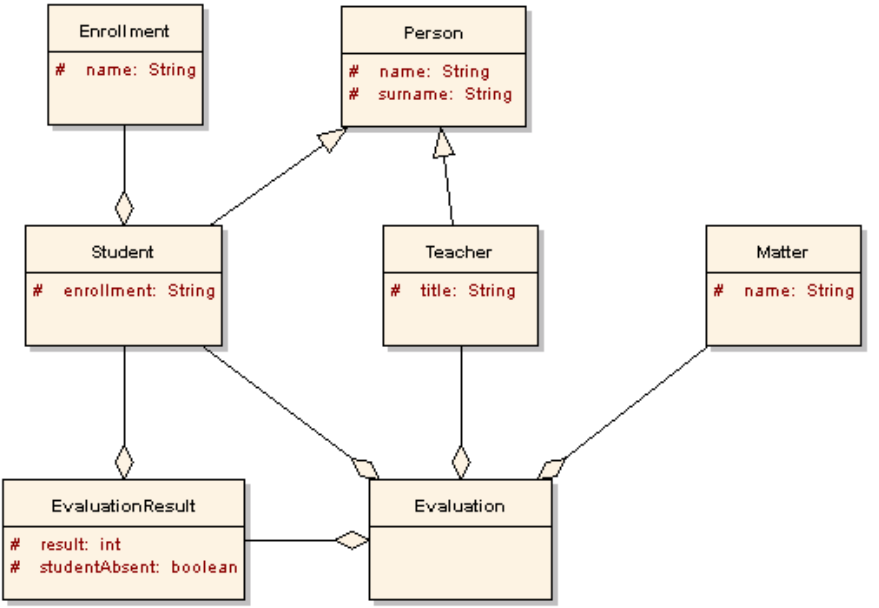


Fig.1. Class diagram for the "Catalog" application.

This represents the core of the application; further, various classes are implemented that will help manipulating the domain objects, resulting the complete application. Various decisions regarding the implementation are taken. For instance, the graphical

user interface type (in this case – a web application), the type of database used (in this case – a database object), the web framework (in this case – a JavaServer Faces implementation), the export library (in this case – a reporting engine). In order to use these additional software components with the given do-main objects, various additional classes need to be implemented, such as:

- User interface classes: these classes help creating a the user interface, such as classes that render menus or which display complex items – like a custom component to let the teacher choose the grade.
- Various services, used for
  - database operations - storing/retrieving data
  - exporting the catalog to PDF format
  - translation utilities
- Service helper classes, which help registering/retrieving services to/from a service repository
- Web server integration classes, for instance special classes used for web session notifications, security filters, etc.
- Business logic: this class helps encapsulating the business decisions (such as creating new domain objects, validating them, saving them to the database, etc.) and is tightly integrated with the web framework.

The domain objects together with the additional classes constitute the final application. It is this final application I am using as input to the reverse engineering application and as desired output is the list of Domain Objects in Fig.1.

Running an implementation of the algorithm on the Catalog application, I obtain the following classes as output: Person, Student, Teacher, Evaluation, Enrollment, EvaluationResult, Matter, BusinessLogic. The output is a superset of the Domain Model, containing additionally the BusinessLogic class, responsible for gluing together the Domain Objects, the database and the user interface. This class is not part of the domain model and is erroneously detected.

Note that this is only a particular example of software design and implementation; other engineers might choose a different domain representation for the same domain, developing an application still respecting the initial specifications. Such decisions are based on various aspects, such as personal style, education, business requirements and so on. For instance, in the previous diagram, an Evaluation aggregates a Matter, but for any Evaluation, a list of Matters needs to be obtained. One might choose to engineer a Matter to aggregate Evaluations, and thus most of the time we find these relations bi-directional. However, this is consistent with Property P3, as we need to have the aggregation in at least one direction.

In a deformed design of the model, one might choose to eliminate all the associations between the entities in the above diagram, and use the concepts as stand-alone entities in the View layer. This is consistent with Property P1, since the View layer will still need to send messages to the Domain Objects.; however, it will not be consistent with Property P3 since the algorithm expects a strong relation between Domain Objects, of the strength of aggregation.

The properties at the foundation of the algorithm are deducted empirically, based on the analysis of many different domain designs and implementations. There is, however, a small percent of extreme cases in which the algorithm does not function properly. For instance, one might choose to use features such as reflection or byte-

code manipulation to perform application logic. The reverse engineering algorithm will fail in most such cases to recognize the Domain Objects, as this is complicated for a human operator also.

There are various well-established tools that are focused on source code reverse engineering, performing some of the steps described in the previous algorithm. However, these tools strip out important information regarding the internal structure of the application. In particular, Property P3 is deducted using information found in class source code, such as collection analysis, for instance analyzing operations performed on java.lang.Iterable implementations[4].

The second step in testing the application is to have it run against real-life applications. For this purpose, I have arbitrarily chosen applications from different domains, applications which have their source code public. I have tested the application against the selected software with the following results:

Name	Version	Type	Total classes	Filtered classes	Domain Object classes	Error
JRoller	3	Blogger	515	120	46	14.40%
Velocity	1.5	Template engine	240	75	24	21.30%
log4j	1.2.15	Logger	260	55	14	15.80%
iText	2.0.6	PDF library	482	192	179	2.70%
Columba	1.4	Mail client	1689	686	272	24.50%
Jtrac	2.1.0	Issue tracker	355	29	14	4.20%
j-spider	0.5.0	Crawler	252	99	71	11.10%
ant	1.7.1	Build system	1137	401	371	2.60%
quartz	1.6.0	Job scheduler	221	74	71	1.40%
Jasper Reports	3.0.1	Report engine	1288	412	330	6.40%
JFreeChart	1.0.11	Chart engine	584	209	209	0.00%
Jforum	2.1.8	Forum software	352	54	40	4.00%
Jdepend	2.9	Dependency analyzer	39	14	8	15.40%

The first column represents the application name, the second is the application version used in test, the 3rd is the application type, the 4th column gives the total

number of classes in the application, the 5th column represents the domain model as outputted by the reverse engineering application – compared to the actual number of domain object classes in the application – 6th column. The last column gives the algorithm's error, expressed as percent of the total application classes the algorithm incorrectly identified as Domain Objects.

On each run, the actual domain model was always contained in the result- plus some additional classes, erroneously detected as being part of the domain model, but the algorithm never missed to include an actual domain object into the output.

5 Uses of Domain Model Reverse Engineering

The process of reverse engineering the domain model of an application is justified by several reasons:

- Software development usually takes place having the domain model as a frontispiece; actually the whole software application is built to support operations on the domain model. During the development phase, some of the initial specifications might be disregarded, wrongly implemented or otherwise not respected. Re-verse engineering at key steps in the software development process would help respecting the initial specifications, or can help in case initial documentation is scarce.
- Domain model reverse engineering could also help assess the quality of the final product. Information from the domain objects will eventually be used in all the application's layers, and using tools to analyze the information flow can provide valuable information regarding the application's quality.
- Domain model reverse engineering highlights the "special" classes of an application, so automated tests can be focused on these classes or test cases can be created that target these classes.
- Also, Use Case reverse engineering might have domain objects as the starting point, as eventually use cases describe particular situations involving domain objects and their relation with internal and external systems.

5 Conclusions and Further Directions

Tests performed on real-life application show promising results. As noted, the algorithm outputs a superset of the domain model. There are errors in this approximation, but further, additional analysis can be performed in order to refine this approximation.

For instance, one additional property that can be harvested is the fact that do-main objects are entirely contained in a single software layer; however this requires more advanced source code analysis.

Source code for the article application as well as other details can be found at http://research.mocanu.info/

References

1. Bjørner, Dines: Software Engineering 1: Abstraction and Modeling. (Springer, 2006)
2. Bjørner, Dines: Software Engineering 2: Specification of Systems and Languages. (Springer, 2006)
3. Martin, Robert: UML for Java Programmers. (Prentice Hall, 2003)
4. Stefan, Mocanu: A new approach to Software Pattern Detection, Proceedings of the 3rd Annual South East European Doctoral Student Conference (2008)

A New Data Structure for Frequent Itemsets Mining: TS-tree

Predrag Stanisic and Savo Tomovic

University of Montenegro,
Dzordza Vasingtona bb, Podgorica, Montenegro
{predrag_stanisic,savica}@cg.yu
http://www.pmf.ac.me

Abstract. In this paper we present new algorithm called TS-Growth algorithm that takes a pattern-growth approach [1] and Rymon's set enumeration framework [2] for frequent itemsets mining. The algorithm encodes the dataset using a compact data structure called TS-tree and extracts frequent itemsets directly from this structure. We will assume that input dataset is represented in horizontal data layout [3]. TS-Growth algorithm uses a single database scan to create the TS-tree, which is a compressed representation of the input data. The TS-tree will contain all itemsets from the input dataset with its support. Because of that we called this tree a Total Support Tree or TS-tree. Mining TS-tree is simple and very efficient: it is sufficient to pass the tree just once level by level and produce all frequent itemsets (itemsets with sufficiently large support).

Key words: frequent itemset mining, Total Support Tree, TS-Growth

1 Introduction

The motivation for discovering association rules has come from the requirement to analyze large amounts of supermarket basket data. A record in such data typically consists of the transaction unique identifier and the items bought in that transaction. Items can be different products which one can buy in super-markets or on-line shops, or car equipment, or telecommunication companies services etc.

A typical supermarket may well have several thousand items on its shelves. Clearly, the number of subsets of the set of items is immense. Even though a purchase by a customer involves a small subset of this set of items, the number of such subsets is very large. For example, even if we assume that no customer has more than five items in his shopping cart, there are $\sum_{i=1}^{5} \binom{10000}{5}$ possible contents of this cart, which corresponds to the subsets having no more than five items of a set that has 10,000 items, and this is indeed a large number!

By association rule mining one can discover rules which present customer habits. For example, the fact that shopper who purchases milk also tends to purchase bread at the same time can be represented by association rule $milk \Rightarrow bread$ [support=20%, confidence=85%]. The rule represents facts:

- 20% of all transactions under analysis contain milk and bread;

– 85% of the customers who purchased milk also purchased bread.

Discovering all such rules is important for planning marketing campaigns, designing catalogues, managing prices and stocks, customer relationships management etc. For example, a shop may decide to place bread close to milk because they are often bought together, to help shoppers finish their task faster. Or the shop may place them at opposite ends of a row, and place other associated items in between to tempt people to buy those items as well, as shoppers walk from one end of the row to the other.

The result of association analysis is strong association rules, which are rules satisfying a minimal support and minimal confidence threshold. The minimal support and the minimal confidence are input parameters for association analysis.

The problem of association rules mining can be decomposed into two sub-problems [4]:

– Discovering frequent or large itemsets. Frequent itemsets have support greater than the minimal support;
– Generating rules. The aim of this step is to derive rules with high confidence (strong rules) from frequent itemsets. For each frequent itemset *l* one finds all not empty subsets of *l*; for each $a \subset l \wedge a \neq \varnothing$ one generates the rule $a \Rightarrow l - a$, if $\frac{support(l)}{support(a)} > minimal\ confidence$.

We do not consider the second sub-problem in this paper, because the overall performances of mining association rules are determined by the first step.

In this paper we present a method which, in single database scan, reorganises transactional database into a simple tree structure and extracts frequent itemsets directly from this structure in just one pass.

The paper is organized as follows. Section 2 provides formalization of frequent itemsets mining problem. Section 3 describes TS-Growth algorithm for TS-tree generation and mining. Finally Section 4 outlines some concluding remarks.

2 Preliminaries

Suppose that *I* is a finite set; we refer to the elements of *I* as items.

Definition 1. *A transaction dataset on I is a function T:* $\{1, ...,n\} \rightarrow$ *P(I). The set T(k) is the* k^{th} *transaction of T. The numbers 1,...,n are the transaction identifiers (TIDs).*

Given a transaction data set T on the set I, we would like to determine those subsets of I that occur often enough as values of T.

Definition 2. *Let T: 1, ..., n*$\rightarrow$*P(I) be a transaction data set on a set of items I. The support count of a subset K of the set of items I in T is the number* $suppcount_T$*(K) given by:*

$$suppcount_T(K) = |\{k|1 \leq k \leq n \wedge K \subseteq T(k)\}|. \quad (1)$$

The support of an item set K (in the following text instead of an item set K we will use an itemset K) is the number:

$$supp_T(K) = suppcount_T(K)/n. \quad (2)$$

The following rather straightforward statement is fundamental for the study of frequent itemsets.

Theorem 1. *Let T: 1, ..., n→ P(I) be a transaction data set on a set of items I. If K and K' are two itemsets, then* $K' \subseteq K$ *implies* $supp_T(K') \geq supp_T(K)$.

Proof. The previous theorem states that $supp_T$ for an itemset has the anti-monotone property. It means that support for an itemset never exceeds the support for its subsets. For proof, it is sufficient to note that every transaction that contains K also contains K'. The statement from the theorem follows immediately.

Definition 3. *An itemset K is* μ*-frequent relative to the transaction data set T if* $supp_T(K) \geq \mu$*. We denote by* F_T^μ *the collection of all* μ*-frequent itemsets relative to the transaction data set T and by* $F_{T,r}^\mu$ *the collection of* μ*-frequent itemsets that contain r items for* $r \geq 1$ *(in the following text we will use r-itemset instead of itemset that contains r items).*

Note that $F_T^\mu = \bigcup_{r\geq 1} F_{T,r}^\mu$. If μ and T are clear from the context, then we will omit either or both adornments from this notation.

3 TS-Growth Algorithm

This section presents new algorithm called TS-Growth algorithm that takes a pattern-growth approach [1] and Rymon's set enumeration framework [2] to discovering frequent itemsets. The algorithm encodes the dataset using a compact data structure called TS-tree and extracts frequent itemsets directly from this structure. The details of this approach are presented next. We will assume that input dataset is represented in horizontal data layout [3].

3.1 TS-tree representation

A TS-tree is a compressed representation of the input data. It is constructed by reading the dataset one transaction in a time and mapping each transaction onto several paths in the TS-tree. As different transactions can have several items in common, their paths may overlap. The more the paths overlap with one another, the more compression we can achieve using the TS-tree. If the TS-tree can be fit into main memory, this will allow us to extract frequent itemsets directly from the structure in memory instead of making repeated passes over the data stored on disk.

We will explain the algorithm for TS-tree generation.

Definition 4. *Let S be a set and let* $d : S \to N$ *be an injective function. The number d(x) is the index of* $x \in S$*. If* $x \in S$*, the view of x is* $view(d, x) = \{ s \in S | d(s) > d(x)\}$.

Let *T: 1, ..., n→P(I)* be a transaction dataset on a set of items *I*. For the set *T* we create a corresponding TS-tree with the following properties:

- initially, TS-tree contains only the root node represented by NULL symbol;
- each node in the tree contains the label of an item from set *I* along with a counter that shows the number of transactions mapped onto the given path, we will call this number path's support;
- the children of a node *P* are nodes from *view(d, P)*.

The algorithm for TS-tree construction is the first and the main part of the TS-Growth algorithm and it is presented next.

Algorithm: TS-Growth, part I
Input: A transaction dataset *T*
Output: The TS-tree representation of the dataset *T*
Method:

1. create the root R of a TS-tree and label it as NULL
2. **for** each transaction *t* in *T* **do**
3. let the items in *t* be $[p|P]$
4. call **insert_tree**$([p|P], R)$
5. if $P \neq \emptyset$, call **insert_tree**(P, R)
6. **end for**

procedure insert_tree(*Trans [p|P], TS-tree R*)

1. **if** *R* has a child *N* such that *N.label=p* **then**
2. increment *N's* count by 1
3. **else** create a new node *N* as child of *R*, and let its count be 1
4. **if** $P \neq \emptyset$, call **insert_tree**(P, N)

Let us explain the basic idea. The dataset is scanned once to create TS-tree. The structure shown in Fig. 1 demonstrates how itemsets contained in a transaction can be systematically enumerated, i.e., by specifying their items one by one, from the leftmost to the rightmost item. Actually, when mapping transaction on paths in the TS-tree, at each node in the tree algorithm uses the hash function *h(x)=d(x)*, to determine which branch of the current node should be followed next. At the end, paths from the TS-tree will represent all itemsets from the input data set while path's support is support of the itemset represented by that path.

The time complexity for the TS-tree generation can be estimated by the following considerations. Each transaction of length $|t|$ produces $2^{|t|}$ itemsets. This is also the effective number of TS-tree traversals performed for each transaction. Because of that, the cost for TS-tree generation is $O(n \cdot 2^{\omega} \cdot \alpha)$, where *n* is the number of transactions in the given dataset, ω is the maximum transaction width and α is the cost for updating the support count of a node in the TS-tree.

Storage of the complete tree of a subsets of *I*, of course, has a requirement of order $2^{|I|}$, which will in general be infisible. However, for large $|I|$ it is likely that most of the subsets *i* will be unrepresented in the database and will therefore not contribute to the support count summation. The TS-Growth algorithm builds the tree dinamically

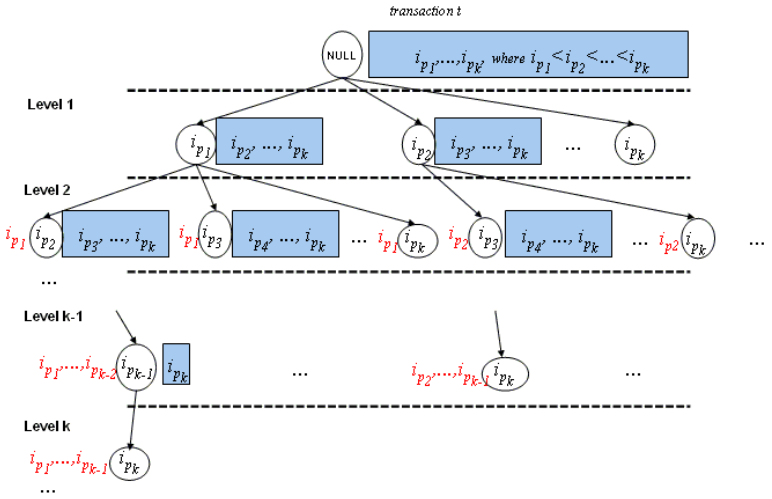


Fig. 1. Mapping transaction onto TS-tree

as transactions are processed, storing supports only for itemsets which appear in the database. Nodes are created only when a new itemset (subset) i is encountered in the database. Building the tree dinamically implies a storage requirement of order n (number of transactions in the database) rather than $2^{|I|}$. This will be reduced further, perhaps substantially, if the database contains a high incidence of duplicates.

We will present the TS-Growth algorithm's steps on the concrete example. Consider a dataset from Table 1 that contains four transactions and five items. The algorithm starts with TS-tree with root node labeled with NULL symbol.

Table 1. Dataset with four transactions and five items

TID	$Items$
1	i_1, i_3, i_4
2	i_2, i_3, i_5
3	i_1, i_2, i_3, i_5
4	i_2, i_5

First transaction to be read is $\{i_1, i_3, i_4\}$. For each item in the first transaction, the algorithm creates nodes at the first level in the tree and set their support to 1. At the first level of the TS-tree the first item is fixed. For instance, $i_1[i_3, i_4]$ represents all itemsets that begins with item i_1, followed by the most two items chosen from the set $\{i_3, i_4\}$. In our example, the algorithms creates two new nodes i_3 and i_4 with support 1 at the second level as children nodes of the node i_1 from the first level of the tree. The second level of the TS-tree represents the number of ways to select the second item. For

example, $i_1 i_3[i_4]$ corresponds to itemsets that begin with prefix i_1, i_3 and are followed by item i_4. The node becomes a leaf when no more options are left or when suffix is empty set. Non-empty suffix at the second level has only node i_3, the suffix is $[i_4]$. Because of that, new node i_4 with support 1 is created at the third level as child node of the node i_3 from the second level. The TS-tree after processing the first transaction $\{i_1, i_3, i_4\}$ is shown in Fig. 2.

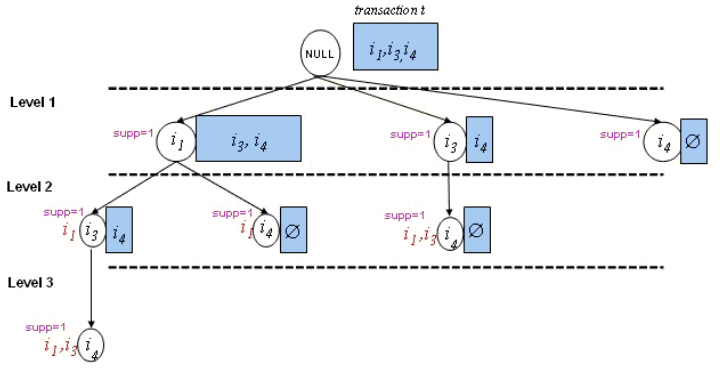


Fig. 2. The TS-tree for dataset from Table 1 after processing the first transaction

The TS-tree in Fig. 2 presents all itemsets contained in the first transaction with their supports. Note that 1-itemsets are situated at the first level of the tree. At the second level all 2-itemsets are placed: they are encoded as prefix + node label. The only one 3-itemset is in the third level. Note that there is no need to store prefix with the node label and support, because the prefix is exactly path from the root to the current node. In general, the level *k* contains all *k*-itemsets; *k*-itemset is ecnoded by path from the root to the particular node at the level *k*; *k*-itemset support is path's support.

The second transaction is $\{i_2, i_3, i_5\}$. A new set of children nodes of the root node is created for items i_2 and i_5 with support 1. Node for item i_3 has already been created (when the first transaction has been processed), so it is sufficient to increment its support. At the second level, nodes i_2 and i_3 contains non-empty suffix and we create children nodes for them: two new nodes for items i_3 and i_5 are created as children of the node i_2; the node i_3 from the first level has suffix i_5, so new node i_5 is created as child of the node i_3; support for all new nodes is 1. Finally, at the third level, just one new node will be created: node i_5 with support 1 at the end of the path $NULL \rightarrow i_2 \rightarrow i_3 \rightarrow i_5$ which corresponds to the second transaction. Fig. 3 shows mapping of the second transaction on the TS-tree.

The TS-tree after reading all four transactions is shown in Fig. 4.

3.2 Frequent Itemset Mining from TS-tree

TS-Growth algorithm extracts frequent itemsets from the TS-tree. It is sufficient to traverse the tree just once, level by level. Mining of a TS-tree is performed by calling the

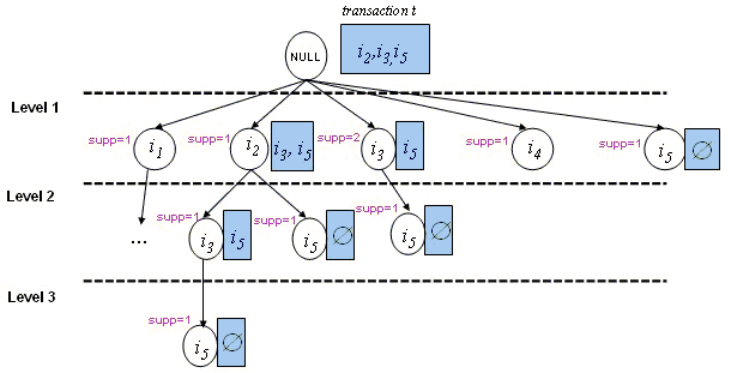


Fig. 3. The TS-tree for dataset from Table 1 after processing the second transaction

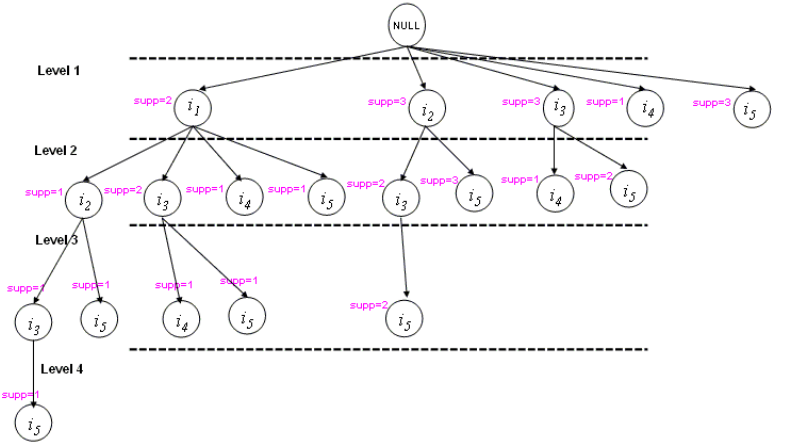


Fig. 4. The TS-tree for dataset from Table 1

second part of the TS-Growth algorithm, which is implemented as follows (we use the QUEUE data stucture for implementation).

Algorithm: TS-Growth, part II
Input: The root *R* of a TS-tree, minimum support threshold μ
Output: F_T^{μ}
Method:

```
1. next = R
2. INSERT(Q, next)
3. while not QUEUE-EMPTY(Q) do
4.    next = DELETE(Q)
5.    if supp_T(next.label) ≥ μ then
6.       F_T^μ += next.label
7.    end if
8.    for each child of next do
9.       set child.label = next.label + child.label
10.       INSERT(Q, child)
11.   end for
12. end while
```

Let us consider concrete example: the TS-tree shown in Figure 4. The TS-Growth algorithm will extract $\frac{1}{2}$-frequent itemsets.

At the **Level 1** all 1-itemsets from the input dataset along with their $suppcount_T$ are contained. The first visited node is i_1 with $suppcount_T(i_1) = 2$, which implies $supp_T(i_1) = suppcount_T(i_1)/4 = \frac{1}{2}$ and i_1 is $\frac{1}{2}$-frequent itemsets. Similarly i_2, i_3 and i_5 are $\frac{1}{2}$-frequent itemsets while i_4 is infrequent ($supp_T(i_4) = suppcount_T(i_4)/4 = \frac{1}{4}$).

At the **Level 2** the algorithm extracts $\frac{1}{2}$-frequent 2-itemsets. First the algorithm follows path $NULL \rightarrow i_1$ and visits node i_2 with supp = 1. This number is actually the number of transactions that contain 2-itemset $\{i_1, i_2\}$ which is encoded by the path $NULL \rightarrow i_1 \rightarrow i_2$ in the TS-tree. The itemset $\{i_1, i_2\}$ has $supp_T(\{i_1, i_2\}) = suppcount_T(\{i_1, i_2\})/4 = \frac{1}{4} < \frac{1}{2}$ and $\{i_1, i_2\}$ is not $\frac{1}{2}$-frequent 2-itemsets. According to the Theorem 1 there is no $\frac{1}{2}$-frequent itemset that contains $\{i_1, i_2\}$, so the algorithm prunes subtree of the node i_2. The next node in the Level 2 that the algorithm visits is i_3 with supp = 2. Accordingly, the itemset $\{i_1, i_3\}$ has $supp_T(\{i_1, i_3\}) = suppcount_T(\{i_1, i_3\})/4 = \frac{1}{2}$ and is declared as $\frac{1}{2}$-frequent 2-itemsets. Two remaining itemsets on the path $NULL \rightarrow i_1$ are $\{i_1, i_4\}$ and $\{i_1, i_5\}$ which do not have sufficient support to be $\frac{1}{2}$-frequent 2-itemsets.

Now, the algorithm follows path $NULL \rightarrow i_2$. There are two itemsets to check: $\{i_2, i_3\}$ and $\{i_2, i_5\}$. Both itemsets have $supp_T \geq \frac{1}{2}$ and are declared as $\frac{1}{2}$-frequent 2-itemsets.

The next path to follow is $NULL \rightarrow i_3$. There are two itemsets to consider: $\{i_3, i_4\}$ and $\{i_3, i_5\}$, but just itemset $\{i_3, i_5\}$ has sufficient support to be $\frac{1}{2}$-frequent 2-itemsets. At this moment, the Level 2 of the TS-tree is completely visited.

At the **Level 3** the algorithm extracts $\frac{1}{2}$-frequent 3-itemsets. Recall that the subtree

of the path $NULL \rightarrow i_1 \rightarrow i_2$ is pruned so the algorithm first follows path $NULL \rightarrow i_1 \rightarrow i_3$. There are two 3-itemsets to be checked: $\{i_1, i_3, i_4\}$ and $\{i_1, i_3, i_5\}$. Both itemsets have insufficient support and can not be $\frac{1}{2}$-frequent 3-itemsets. The last path is $NULL \rightarrow i_2 \rightarrow i_3$ and just one 3-itemset $\{i_2, i_3, i_5\}$ with $supp_T(\{i_2, i_3, i_5\}) = suppcount_T(\{i_2, i_3, i_5\})/4 = \frac{1}{2}$, which means the itemset $\{i_2, i_3, i_5\}$ is $\frac{1}{2}$-frequent 3-itemset.

The complete tree is traversed and the algorithm will output the collection: $F_T^{\frac{1}{2}} = \bigcup_{i=1}^{3} F_{T,i}^{\frac{1}{2}} = \{i_1, i_2, i_3, i_5, i_1i_3, i_2i_3, i_2i_5, i_3i_5, i_2i_3i_5\}$.

4 Conclusion

In this paper we presented TS-Growth algorithm for mining frequent itemsets. The algorithm uses TS-tree as memory resident structure to represent input dataset. The TS-tree is generated in one pass over the input dataset and when the tree is generated we need to traverse the tree just once to find all frequent itemsets, while:

- in the FP-Growth algorithm [1] for FP-tree generation two passes are required and algorithm uses additional data structure called item header table, which contains frequent 1-itemsets sorted in decreasing support count. FP-tree is mined in recursive fashion using divide-and-conquer strategy;
- in the Apriori TFP-algorithm [5] [2] memory resident P-tree is created in one pass over the input dataset. Frequent itemsets are organized in T-tree, which is generated by traversing the P-tree level by level: for each node in the P-tree algorithm update one portion of T-tree. Number of passes of the P-tree can not exceed 1 + size of maximal frequent itemset.

References

1. Han, J., Pei, J., Yu, P.S., Mining Frequent Patterns without Candidate Generation, Proceedings of the ACM SIGMOD Conference on Management of Data, 1-12 (2000)
2. Coenen, F.P., Leng, P., Goulbourne, G., Tree Structures for Mining Association Rules, Journal of Data Mining and Knowledge Discovery Vol. 8, No. 1, 25-51 (2004)
3. Tan., P.N., Steinbach, M., Kumar, V., Introduction to Data Mining, Addicon Wesley (2006).
4. Agrawal, R., Srikant, R., Fast Algorithms for Mining Association Rules, Proceedings of VLDB-94, 487-499, Santiago, Chile (1994)
5. Coenen, F.P., Leng, P., Ahmed, S., Data Structures for Association Rule Mining: T-trees and P-trees, IEEE Transactions on Data and Knowledge Engineering, Vol. 16, No 6, 774-778 (2004)
6. Coenen, F.P., Leng, P., Ahmed, S., T-Trees, Vertical Partitioning and Distributed Association Rule Mining, Proceedings ICDM-2003, 513-516 (2003)
7. Goulbourne, G., Coenen, F., Leng, P., Algorithms for Computing Association Rules Using a Partial-Support Tree, Journal of Knowledge-Based Systems Vol. 13, 141-149 (1999)
8. Grahne, G., Zhu, J., Efficiently Using Prefix-trees in Mining Frequent Itemsets, Proceedings of the IEEE ICDM Workshop on Frequent Itemset Mining Implementations (2003)
9. Simovici, A. D., Djeraba, C., Mathematical Tools For Data Mining (Set Theory, Partial Orders, Combinatorics), Springer-Verlag London Limited (2008)
10. Stanisic, P., Tomovic, S., Apriori Multiple Algorithm for Mining Association Rules, Information Technology and Control Vol. 37, No. 4 (2008)

A Competitive Collaboration between Machine Learning Techniques for Hepatic Fibrosis Detection

Smaranda Belciug[1] and Monica Lupsor[2]

[1] University of Craiova and University of Pitesti,
Pitesti, Romania
smaranda.belciug@gmail.com,
[2] University of Medicine and Pharmacy,
Cluj-Napoca, Romania

Abstract. The aim of this paper is to present a competitive/collaborative automatic diagnosing system, based on three machine learning models: Bayes classifier, k-nearest neighbors and probabilistic neural network. Moreover, the system is applied to a real database, proving its effectiveness.

1 Introduction

Hepatic fibrosis is an indicator in the progression of chronic liver disease. The last stage of liver stiffness leads to cirrhosis and hepatocellular carcinoma. Chronic liver diseases and cirrhosis are important causes of morbidity and mortality in many countries (i.e. Egypt - 14,610 deaths, United States -14,003 deaths, Mexico -10,887 deaths, Romania -10,023 deaths, Japan -9,849 deaths (reference: World Health Organization, January 2004)).

For 60 years the only way that the liver fibrosis progression could be assessed almost for sure was by liver biopsy. However, liver biopsy is a painful and invasive procedure and also carrying a small, but significant life-threatening complication risk.

One of the latest technological discovery worldwide in the evaluation of hepatic fibrosis is the Fibroscan (Echosens, Paris, France), a specially adapted ultrasound device using the principle of the one-dimension transient elastography (TE) for the assessment of liver stiffness.

Modern medical methods provide diagnosis accuracy for detecting hepatic fibrosis ranging from 22% to 71% [8]. One of them [7], concerning the "one against all" technique, used the ROC (Receiver Operating Characteristic) analysis [4] to provide thresholds between different stages of the liver fibrosis, by repeatedly grouping them into two sub-groups (multi-step procedure). This is a pretty slow technique since it is based on randomly splitting patients into two initial groups and then refining step-by-step the procedure until the right stage is correctly identified.

Significantly different from previous medical procedure, we want to directly identify the fibrosis stage by using some well-known intelligent computing models. Thus, the aim of this paper is to provide a single-step diagnosis procedure by managing a competitive collaboration between different machine-learning (ML) techniques, and second to provide a better diagnosis procedure than each ML technique could accomplish

alone. Thus, using this computer-aided diagnosis, we can help the medical personnel establish the correct medical decision. The competitive part plays an important role in establishing the hierarchy of the ML techniques, depending on their diagnosis accuracy. Thus, only the best competitors will be considered to make the final decision. The collaborative part is based on a weighted voting system (WVS) applied to the selected ML techniques [11].

2 Intelligent Computing System

2.1 Competitive Phase

In this phase, each machine learning technique is applied on a database consisting of 722 consecutive patients with chronic HCV infection, examined at the 3rd Medical Clinic, University of Medicine and Pharmacy Cluj-Napoca, Romania, between May 2007 and August 2008. Table 1 presents an example of the parameters and the diagnosis classes (Metavir F), taken into consideration in our study for three patients.

Table 1. Examples of medical parameters and diagnosing classes (3 patients)

Medical parameters and diagnosing classes							
Medical parameters	P1	P2	P3	Medical parameters	P1	P2	P3
Metavir F	1	4	2	Tqs (tocopheryl quinones)	15.2	25.4	15
Stiffness	5.3	27	6	INR (prothrombin time ratio)	0.98	1.85	0.95
Sex	2	1	1	Prol activ part throm-boplastin time	30.9	30.9	28.4
Age	56	46	31	Haematids (erytthrocytes)	4.91	5.11	4.81
BMI (body mass index)	24.3	30.1	34	Hemoglobin	14.5	15.2	14.9
Glycemia	110	84	96	Hematocrit	40.9	45.2	42.6
Triglycerides	54	93	154	Medium eritrocity volume	83.2	88.5	88.6
Cholesterol	133	152	197	Average eritrocitary hemoglobin	29.5	29.7	31
Aspartate aminotrans-ferase	62	105	46	Avg conc of hemoglobinin	35.6	33.5	35
Alanin aminotransferase	61	167	117	Leukocytes	83.2	88.6	88
Gama glutamyl transpep-tidase	19	187	41	Thrombocytes	29.5	29.7	31
Total bilirubin	1.2	1	0.9	Sideraemia	35.5	33.5	35
Alkaline phosphatase	283	246	248	Urea	24	28	32.7
Prothrombin index	102	42.7	107	Creatinim	0.75	1.24	1.36

The machine learning techniques that were applied on this dataset are:

(a) The Naive Bayesian classifier [3];

(b) k - nearest neighbor (k-NN) [2];

(c) Probabilistic neural network (PNN) [10].

Next, we shortly describe the three algorithms subsequently used. Supposing that there are q classes $\Omega_1, ..., \Omega_q$, with m_i training samples in Ω_i, and each data point is represented by an n-dimensional feature vector $x = \{x_1, ..., x_n\}$ depicting n-measurements made on the sample from n attributes, we have to assign to an unknown data sample x a certain class.

Naive Bayesian classifier algorithm

Step 1. Compute the prior probabilities, using the formula $\mathrm{P}(\Omega_i) = \frac{n_i}{n}$, where n_i is the number of training samples of class Ω_i , and n is the total number of training samples.

Step 2. Compute the probabilities $P(x_k|\Omega_i)$ using the training sample **x**, for all x_k. If A_k is categorical variable, then $P(x_k|\Omega_i) = \frac{n_{ik}}{n_i}$, where n_{ik} is the number of training samples of class Ω_i having the value x_k. If A_k is numerical attribute, then $P(x_k|\Omega_i) = \frac{1}{\sqrt{2\pi}\sigma_{\Omega_i}} \exp -(\frac{(x_k - m_{\Omega_i})^2}{2\sigma^2_{\Omega_i}})$, where m_{Ω_i} is the mean, and σ_{Ω_i} is the standard deviation. Compute $P(x|\Omega_i) = \prod_{k=1}^{m} P(x_k|\Omega_i)$. Compute $P(x|\Omega_i) \cdot P(\Omega_i)$, for each class Ω_i.

Step 3. The label of sample **x** is maxim$(P(x|\Omega_i) \cdot P(\Omega_i))$.

k - nearest neighbor algorithm

Step 1. Compute the distance to all the other training records.

Step 2. Identify k nearest neighbors.

Step 3. Taking the majority vote from the k - nearest neighbors (class labels), determine class label for sample **x**.

Probabilistic neural network

Step 1. For each class Ω_i compute the (Euclidian) distance between any pair of vectors and denote these distances by $d_1, d_2, ..., d_{r_i}$, where $r_i = C^2_{m_i} = \frac{m_i!}{2!(m_i-2)!}$. For each class Ω_i compute the corresponding average distances and standard deviations $D_i = \frac{\sum_{j=1}^{r_i} d_j}{r_i}$, $SD_i = \sqrt{\frac{\sum_{j=1}^{r_i}(d_j - D_i)^2}{r_i}}$. For each class Ω_i compute the corresponding confidence intervals $I_{\Omega_i} = (D_i - 3SD_i, D_i + 3SD_i)$ for the average distances.

Step 2. For each decision class Ω_i consider the decision functions $f_i(x) = \frac{1}{(2\pi)^{n/2}\sigma_i^n} \cdot \frac{1}{m_i} \cdot \sum_{j=1}^{m_i} \exp(-\frac{d(x,x_j)^2}{2\sigma_i^2})$. Assign (σ_i, D_i), $i = 1, 2, ..., q$.

Step 3. In each decision class Ω_i (randomly) choose a certain vector x_i^0, fix it and compute all distances $D_{\Omega_i,j} = d(x_i^0, x_j)$, $j = 1, 2, ..., m_i$. Compute $f_i(x_i^0) = \frac{1}{(2\pi)^{n/2} D_i^n} \cdot \frac{1}{m_i} \sum_{j=1}^{m_i} \exp(-\frac{D^2_{\Omega_i,j}}{2D_i^2})$.

Step 4. *(Bayes decision rule)* Compare $f_i(x_i^0)$ and $f_j(x_i^0)$, for all $i \neq j$, following the algorithm: "IF $l_i h_i f_i > l_j h_j f_j$ (for all $j \neq i$) THEN $x_i^0 \in \Omega_i$ ELSE IF $l_i h_i f_i \leq l_j h_j f_j$ (for some $j \neq i$) THEN $x_i^0 \notin \Omega_i$". For each (fixed) decision class Ω_i consider the 3-valued logic: TRUE - if $l_i h_i f_i > l_j h_j f_j$ (for all $j \neq i$) UNKNOWN - if $l_i h_i f_i = l_j h_j f_j$ (for some $j \neq i$)and FALSE - otherwise. l_i, h_i are the model parameters (costs and prior probabilities).

Step 5. Repeat step 3 for another choice for x_i^0 in Ω_i until all of them are chosen. Repeat step 3 for all vectors x_j^0 in Ω_j for all $j \neq i$. Obtain the classification accuracy in percentage.

Step 6. *(Estimating optimal smoothing parameter)* Divide each confidence interval I_{Ω_i} by N dividing knots into $(N + 1)$ equal sectors with length $\Delta_i = 6SD_i/N$. Repeat step 3 by assigning $\sigma_i = D_i + k\Delta_i$ and $\sigma_i = D_i - k\Delta_i$, $k = 1, 2, ..., N$. If the current value of σ_i exceeds I_{Ω_i}, then STOP. Obtain the corresponding classification

accuracy. Compute the maximum value MAX of the variable corresponding to TRUE out of the N cases.

Step 7. The smoothing parameters σ_i, $i = 1, 2, ..., q$ of each class, corresponding to MAX represent the optimal values of the smoothing parameters σ 's for each decision category Ω_i,$i = 1, 2, ..., q$.

The results obtained using the above ML models were statistically evaluated in order to put them in a hierarchical order, depending on their diagnosing accuracy. The aims of statistically analyzing the diagnosis performances of the three ML models used in this paper are the following:

- *Data screening*, that is the preliminary data analysis to examine the suitability of the data for the type of statistical analysis that is intended;

- *Hypotheses testing*, involving the comparison between training and testing performances of the three NN models;

Data screening consisted in verifying both the normality and the equality of variances for the testing performances, hypotheses necessary for applying the t-test for independent samples. Recall that the t-test can be used even if the sample sizes are small enough as long as the variables are normally distributed within each group and the variation of scores in the two groups is not significantly different. We have used both the Kolmogorov-Smirnov & Lilliefors test, which is applicable when the mean and the standard deviation are computed from the actual data, and the Shapiro-Wilk W test [1]. Next, the equality of variances has been verified using the Levene's test [6]. Let us notice that, in practice, inequality of variances of two independent samples is less problematic when the samples have the same number of observations. Technically, the statistical evaluation involved a benchmarking process consisting in two statistical comparison tests concerning the testing performances of the three ML algorithms, that is the classification performances obtained when using new patients after the training phase is completed. The hypotheses testing referred to:

- t-test for independent samples to compare differences in means;

- Comparison of average testing performances (two-sided z-test to compare proportions).

The t-test for independent samples is the most commonly used method to evaluate the differences in means between two independent groups of observations [1]. Basically, with independent groups of observations, we are interested in the mean difference between the two groups, focusing also on the variability between observations. On the other hand, many studies involve testing differences between two population proportions, problem that can be solved using the classical z-test statistic. Technically, the two-proportion z-test for equal variances is given by:

$|z| = \sqrt{\frac{N_1 \cdot N_2}{N_1+N_2}} \cdot \frac{|p_1-p_2|}{\sqrt{p \cdot q}}$, where $p = \frac{p_1 \cdot N_1 + p_2 \cdot N_2}{N_1+N_2}$, and $q = 1 - p$, N_1 and N_2 representing the samples sizes, and p_1 and p_2 the corresponding proportions [9].

Previously to using the above comparison tests, we applied a power analysis technique (*two-tailed* null hypothesis type) to determine the appropriate sample size to achieve adequate power. Statistical power tells us, in probabilistic terms, the capability of a test to detect a significant effect, that is how often a correct interpretation about the effect is reached, if we were to repeat the test many times [1]. Concretely, a sample size of 110 different computer runs for each NN model is considered, providing a statistical

power with a minimum of 99% (with type I error α = 0.05) for the comparison tests subsequently used.

2.2 Collaborative Phase

After establishing the hierarchy in the first phase, we retained the best classifiers only. Generally speaking, when using an initial number of n classifiers in the first phase, only a certain number of k (best) competitors will be retained for the collaborative phase. The number k is chosen by the user, depending on the competitive phase result, on the concrete problem to solve etc. Since we used three ML models only, we have kept all of them. The WVS applied in this paper used the weights directly proportional to the testing performance. Finally, applying the WVS for this computing system, the overall automatic diagnosis can be established for a new patient[11].

2.3 Results

Resampling techniques enable us to estimate the performance of a classifier, such that it is guaranteed that approximately the same level of performance will be achieved with a new dataset of the same domain. For, it is at least necessary to use different sets of data for training and testing. Recall the holdout method, consisting in dividing the existing dataset into one training set and one testing set (e.g. 3/4 and 1/4 of the dataset, which are mutually exclusive). A simple modification of this method, the *k-fold cross-validation* technique, consists in dividing the whole dataset into k equally sized parts, and each part is used as testing set for the classifier trained with the remaining dataset. Finally, the observed performances for the k different runs are averaged. In this study, the 4-fold cross-validation is used as a verification method. Hence, the classification accuracy - training/testing - is computed four times, each time leaving out one of the four sub-samples from the computation and using that sub-sample as a test sample for cross-validation, so that each sub-sample is used 4 times as training sample and just once as testing sample. The ML correct classification rates, computed for each of the four runs of the ML models, are then averaged to give the 4-fold estimate of the classification accuracy. Technically, 542 cases were randomly selected each time for training and the remaining 180 cases for testing. This procedure is repeated four times to complete a cross-validation cycle, consisting of the four runs of each ML model. Computationally, each ML model is run 110 times in order to asses a minimum of 99% statistical power (with type I error α = 0.05) in a complete cross-validation cycle consisting of 75% training samples and 25% testing samples, randomly chosen [5]. For each complete cross-validation cycle, the correct testing classification rate (i.e. testing performance equalling the proportion of cases that are correctly classified in the testing phase) was recorded, since it represents the real diagnosing performance of each model. Thus, three samples, one for each ML model, with equal size (110) were selected for the subsequent comparisons. The results of the diagnosing performance of the three ML models, in terms of mean, standard deviation (SD) and 95% confidence intervals, averaged over the 110 computer run of a complete cross-validation cycle are displayed in Table 2. The corresponding graph illustrating the mean testing performances and the corresponding standard deviations are depicted in Figure 1.

Table 2. ML diagnosing performances (averaged over 110 computer run)

ML diagnosing performances		
ML model	Mean/SD testing performance (%)	95% conf. int. test. perf. (%)
Naive Bayes	52.10 / 5.12	(50.46, 53.74)
PNN	50.40 / 4.96	(48.81, 51.99)
k-NN	43.35 / 5.27	(41.66, 45.04)

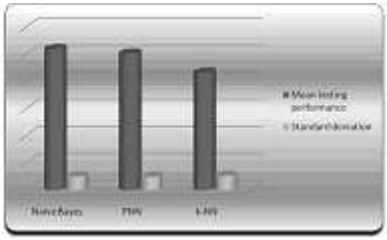


Figure 1. Mean and standard deviations of the testing performances

From the above table it can be seen that the testing performances depend on the ML type, the poorer performance being obtained by k-NN (43.35%), followed by the PNN model (50.40%) and Nave Bayes as the best classifier (52.10). Obviously, this result and corresponding ML hierarchy is decisively determined by this particular database. The testing performances of the three ML models will be further statistically investigated.

Next, in order to apply the t-test for independent samples, we checked up both the normality and the equality of variances for the testing performances. The Kolmogorov-Smirnov & Lilliefors test [1] was performed to test the normality of the testing performance measure. The conclusion was that the testing performances for all the three ML models is Gaussian (P-level > 0.20). Equality of the variances was tested using the Levene's test [1]. Thus, Bayes vs. PNN provided a P-level equaling 0.96, Bayes vs. k-NN provided a P-level equaling 0.98, and PNN vs. k-NN provided a P-level equaling 0.94, that is there is no difference in variances. Applying the t-test for independent samples, we obtained the following result, displayed in Table 3.

Table 3. Comparing testing performances (t-test for independent samples)

Comparing testing performances	
NN type	NN type t-test for independent samples (t-value / P-level)
Naive Bayes vs. PNN	1.51 / 0.14
Naive Bayes vs. k-NN	7.53 / 0.00
PNN vs. k-NN	6.16 / 0.00

From Table 3 we see that there is a significant difference in means between Naive Bayes and k-NN and between PNN and k-NN, while Naive Bayes and PNN are not at all different.

An alternative way for comparing the three ML models is to compare the average proportions of correctly classified cases that are the average testing performances displayed in Table 2. Table 4 below illustrates the results of this comparison using two-sided z-test and shows that there is no significant difference (P-level > 0.05) for all the three ML models average testing performances, in spite of the fact that the t-test for independent samples showed similarity for Naive Bayes and PNN only.

Table 4. Comparing average testing performances

Comparing average testing performances	
Variable	p-level (two-sided)
Naive Bayes vs. PNN	0.80
Naive Bayes vs. k-NN	0.20
PNN vs. k-NN	0.27

Finally, the collaborative phase will use a system of weights w_i, $i = 1, 2, 3$, directly proportional to the diagnosis accuracy of the three ML models, that is $w_1 = 0.36$ for Naive Bayes, $w_2 = 0.34$ for PNN and $w_3 = 0.30$ for k-NN. Next, applying the WVS with the above parameters, the overall automatic diagnosis can be established for a new patient. A simulation consisting in three different cases is displayed in Table 5.

Table 5. Simulation of the competitive/collaborative diagnosing system

Simulation of the competitive/collaborative diagnosing system				
Naive Bayes	PNN	KNN	WVS	Diagnosis
1	1	2	$0.36 \cdot 1 + 0.34 \cdot 1 + 0.3 \cdot 2 = 1.3$	1
4	3	2	$0.36 \cdot 4 + 0.34 \cdot 3 + 0.3 \cdot 2 = 3.06$	3
3	3	0	$0.36 \cdot 3 + 0.34 \cdot 3 + 0.3 \cdot 0 = 2.88$	3

2.4 Conclusions

In this paper we demonstrated the suitability of the ML methodology, used in a both competitive and collaborative way for the diagnosis of liver fibrosis. The classification results were consistent with some of the highest results obtained by using sophisticated and expensive modern medical techniques. Future work would use more ML techniques in order to assess the potential of this Artificial Intelligence domain in medical decision support.

References

1. Altman DG. Practical statistics for medical research. Chapman and Hall, (1991).
2. Dasarthy B.V. Nearest Neighbor Classification Techniques. IEEE Press, Hoboken(NJ), (1990).
3. Domingos P., Pazzani M. Beyond independence: Conditions for the optimality of the simple bayesian classifier. Proceedings of the Thirteenth International Conference on Machine Learning (Morgan Kaufmann Publishers, Inc; 105-112 (1996)
4. Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters; 27: 861-874 (2006).
5. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence; 2 (12): 1137-1143 (1995).
6. Levene H. Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling (I. Olkin et al. eds.). Stanford University Press, 278-292 (1960).
7. Lupsor M., Badea R., Stefanescu H., Grigorescu M., Sparchez Z., Serban A., Branda H., Iancu S., Maniu A., Analysis of histopathological changes that influence liver stiffness in chronic hepatitis C. Results from a cohort of 324 patients. J Gastrointestin Liver Dis. 17(2): 155-163 (2008).

8. Nunez M., Serum Markers for Assessment of Liver Fibrosis in HIV-HCV Coinfected Patients, 57 Annual Meeting of the American Association of Liver Diseases, October 26-31, 2006, Boston, MA (2006).
9. Sheskin D. Handbook of Parametric and Nonparametric Statistical Procedure. 3rd ed. Chapman & Hall/CRC, (2004).
10. Specht DF. Probabilistic neural networks. Neural Networks; 3: 109-118 (1990).
11. Tannenbaum P. Excursions in Modern Mathematics (6 Ed.). Prentice Hall, (2006).

Computer Games as Instructional Tools: Teaching History with Europa Universalis III

Maria Mavrommati

University of Macedonia, Department of Applied Informatics
Egnatia 156, Thessaloniki, 54006, Greece
mmavrom@uom.gr

Abstract: Using computer games as instructional tools is a new practice, emerging from the interest in their constructive and interactive character. Following the design- based research paradigm and using Activity Theory, which examines knowledge construction as a social phenomenon, as a theoretical tool, I examine how learning occurs when a historical strategy pc game, Europa Universalis III, is used in an informal learning environment. Knowledge about the context in which learning occurs when a computer game enters the learning process is useful to educational researchers who study the learning process in general, those who examine it in reference to computer games, as well as educational game designers.

Keywords: Game-play, educational software, digital game-based learning.

1 Purpose of Study

1.1 Introduction

The term Computer Games refers to a wide variety of informatics applications, which share enjoyment, player participation, interaction, role play and multimedia use as common elements. One of the most comprehensive definitions describes computer games as "rule-based systems, or structures for play", identifying the existence of rules and structured environment as their basic features [1]. The variety of game genres and the associated skills developed by game-play beget a variety of educational purposes which can be set if computer games are used as educational tools, a research field in continuous development [2]. Digital game-play has received wide acceptance by teenagers, younger children and young adults alike. It is estimated that for the year 2001 computer games sales were up to 19 billion dollars [3], making them one of the most profitable products of mass entertainment industry. The plenitude of educational software that embody game elements, in addition to mere computer games used in educational contexts, has led to the formation of a new study area under the broad field of Games Studies, which focuses on the educational aspects of game-play, blooming in the USA and in countries of Northern Europe [4].

Children's interest in computer games urges educational researchers and instructors to explore the potential of using computer games as instructional tools [5]. Such a turn seems to be also supported by cognitive neuroscience, which has been suggesting that the use of digital tools of learning and information presentation, such as multimedia and the internet, especially during development, can change the way brain operates. This hypothesis calls for the development of new learning methods and instructional material [6]. Digital games are a relatively newly discovered area of instructional development. According gamers, computer games seem to succeed full immersion and continuing motivation to achieve higher scores and abilities, mobilization of imagination and critical thinking, and development of decision making skills, all necessary for game-play to advance. The diversity of learning goals that could be set if games were used as instructional tools emerges from the wide variety of applications and contents of digital games [2]. The adjustment of children's cognitive skills in the new technological applications opens the door for the use of digital games in education, since the preference they show to games as entertainment tools and the possibilities games offer call for their use as instructional tools as well.

On the other hand, history is commonly referred to as the most "boring" subject of the curriculum. History teachers have been trying to promote student interest by using new technology applications in the history classroom with positive outcomes, achieving deeper understanding of the past [7]. Such initiatives have been successful in various ways in the past, and historians have used the internet, historical documentary editing and multimedia- based activities for their teaching. Over the last decade, the interest of history instructors on the use of constructive game environments to teach history has provided teachers and game developers with thoughtful insight on the process of understanding history through such historical simulations [8]. Following this scientific tradition, my research will focus on understanding the processes with which children construct knowledge through game-play, specifically in the subject of history.

1.2 The Aim of the Study

Modern educational theories and research approach learning as a dynamic system of interactions between the learners and the learning material. Learning is considered an activity that takes place in a constantly developing system, consisting by the students, the learning material, the instructor and the learning goals. Each part of the system evolves and changes through interaction with the others, thus affecting the outcomes of the whole learning process. More and more teachers today, especially in the USA and countries of northern Europe, use commercial computer games as a tool to teach social sciences and especially history [9]. Nevertheless, the actual activity taking place in the classroom has not been widely studied, leaving the relevant efforts with no empirical evidence of how exactly learning occurs in such settings. My research has this specific aim, namely to understand the ways learning happens in a system where computer games made for commercial use hold the role of educational material. Historical computer games such as the *Civilization* series or, in our case, *Europa Universalis*, are used by history educators with the hope that they may rise

questions regarding the nature of social and historical phenomena, and thus create "academic" interest, urging the players to understand and explain how history is "made" [10]. This provides educators with a strong tool for history instruction, although the ways such instruction occurs successfully have not been studied or explained. Therefore, the proposed study aims at explaining the cognitive and social processes taking place during game-play, which may hopefully assist teachers who wish to integrate computer games in their teaching and also educational software designers who try to enhance learning through the application of computer games characteristics. More particularly, I am interested in examining the activity taking place when a computer game is used as an educational tool, in terms of interaction between the students and their classmates, the learning material and the teacher. Viewing computer games as a tool for learning through the construction of a personal understanding of the game environment, the research's aim is to find out how learners achieve understanding and knowledge construction when playing a historical strategy game, *Europa Universalis III*.

1.3 Purposes of the Study

The present research aims at understanding how children learn through playing computer games, a process which includes role- playing, decision making, the development of strategic skills and researching fundamental facts concerning the historical era during which the game supposedly takes place. More extensively, the study's purposes are formed as follows:

1. to examine the context in which learning occurs, when a computer game is used as an educational tool; namely what kind of activity is manifested when computer games enter the classroom;
2. to explore the interaction occurring between students and the learning material, i.e. the computer game, as well as interactions between students as cooperative learners and students and the teacher in the context of knowledge construction;
3. to examine whether students, led by play motivation and personal interest and enjoyment, develop a genuine interest and an understanding of basic historical notions and facts.

2 Methodological Approach

The study will be a naturalistic case study of a design experiment, following the principles of case study research posed by Stake [11]. In accordance with Stake's methodology for educational research, my aim is to understand the particularities and the complexity of the case in question, approaching it as a system of interactions. Data will be collected through observation accounts, interviews, questionnaires, and document analysis, in order to form narrative accounts for the whole classroom, consistent with case study methodology. The experiment will follow the design based

research paradigm, according which the researcher "engineers forms of learning and systematically studies them within the context defined by the means of supporting them" [12] (cited in [13]). This means that the researcher does not study only one variable and the consequences of altering it, but designs learning environments combining them "with the empirical exploration of our understanding of those environments and how they interact with individuals" [14]. Thus, design experiments investigate cognition in the context it occurs, without any attempt to isolate any variable of the learning process, involving the creation of a learning environment and the exploration of the ways cognition occurs in it. Design- based research is a holistic educational intervention that sees the context of learning as a crucial factor of the learning process. Stake's methodology is unique in this sense, for it allows the researcher to respond to the circumstances that rise during the intervention. This kind of research involves the design and experimentation with the set of the components of learning, including curriculum aims, the learning material and the activity developed in the context of learning, making the intervention itself an outcome of the process [15]. Design- based research considers the act of teaching and learning dynamic, making the effort to understand it a multidimensional one. Nevertheless, the most important aspect of design- based research is that theories of learning and the design of learning evolve during the process of the intervention in authentic learning environments. At the same time, all the parts of the intervention are constantly reshaping in order to follow the behaviors that emerge during the learning process. Therefore, the research is conducted through the development of technological tools, curriculum and theories that explain the learning process [13]. This means that the experiment is designed according to the social contexts, applying the needs of the learning environment to the design of teaching material and methods. Such interventions aim at the development of innovative learning structures that respect the needs that arise inside the learning context.

Looking on knowledge construction as an active process formed by context, calls for a theoretical approach that understands cognition as a social phenomenon. In this context, Activity Theory, a theoretical framework which understands all human activity as a set of actions mediated by tools (in our case the computer game) and the cultural context in which it develops (in our case the classroom culture), will be used as the theoretical lens through which we will interpret the knowledge construction process, the actions taking place in the context of learning [9, 16, 17, 18, 19]. Tracing back to Kantian, Hegelian, Marxist and Engelian philosophy, Activity theory is a development of the cultural- historical Vygotskian psychology [20]. In this sense, Activity theory sees all human behaviour as a system consisting of the subject, the object, the goal and the context created through this interaction between the abovementioned parties. With this theoretical approach in mind, it seems natural to approach each learning condition as a unique one, but also multidimensional, consisting of many variables that shape the specific educational setting. Especially in the field of educational sciences, Activity theory is used as a framework in order to understand the activity taking place in a constantly evolving environment, taking into account situations that rise from real- life contexts, making thus educational research fundamentally contextual. As a consequence, it is not enough to isolate one variable in order to understand the consequences, but rather study the behaviours holistically inside this lively context.

The design of this experiment will involve the production of a curriculum based on the *Europa Universalis III* pc game, which deals with European history from the 14th to the 18th centuries. *Europa Universalis III* is a historical strategy game, which gives the players the opportunity to experiment with the historical data in order to create a personal understanding of historical development. Each player chooses a kingdom or a territory and advances it using diplomacy, technological advancement, economy and cultural development, through the course of its history, which expands for more than 350 years, from 1453 to 1789. Players have the opportunity to explore the relationships between different expressions of human civilization, such as war, economy, cultural achievements and social needs, and the outcomes of every decision made in these areas in the course of history. More conventional means of teaching history rarely give such wide opportunities to explore the dynamics of historical progress. Players expand their territory by developing strategies, building their empire and by learning basic facts of European history through game-play. Similar experiments on historical strategy games, such as Kurt Squire's, have shown that students' engagement in the game often lead to experimentation with variables of historical development. This means that in some cases students found game-play an interesting way to perform "hypothetical history" [8].

Using design- based methodology requires flexibility in order to face emergent behaviors during classroom sessions. Therefore the first stage of a design experiment requires a general design of the curriculum, which will be formed and redesigned through game-play and the emergent queries of the students. I consider game-based activities as ones that promote question-rising and examination of various factors, which will be the moving forces of the development of the teaching strategies and the curriculum. Designing a curriculum using digital games as educational tools is an area which still needs to be studied. Nevertheless, a general framework must be developed in this early stage, which will only serve as the best guess of what might make an effective game-based history unit.

An outline of the curriculum for the study of history using *Europa Universalis III* is formed as follows:

Table 1: Outline of the curriculum

Days	Activity	Purposes
Week 1 (days 1-3)	set up of the game- tutorial- providing background information.	This section's purpose is to provide basic information regarding the historical period in question as well as the game's functions.
Week 2 (days 4-6)	start playing in simple scenarios- generate questions- search for answers through the internet and discussion with	This section's purpose is to generate an inquiry culture in the classroom, which is expected to enhance students' curiosity for the game.

	the teachers and the classmates.	
Week 3 (days 7-10)	meaningful game-play-examination of the outcomes of the decisions made during game-play, preparation of presentations.	This section's purpose is to promote connections between knowledge of actual history and the outcomes of the players decisions regarding the development of their chosen nation.
Week 4 (days 11-15)	discussion of students' understandings, comparison of findings of different games.	The purpose of the last session is to generate assertions regarding the development of the nations through game-play and the actual historical developments. These assertions must have an underlying understanding of historical structures, ie how the development of industry in the game advanced a nation's economy, and how this actually happened in real history.

My special interest is to explore the possibilities games offer in constructing historical past, in the context of a learning environment whose components are examined holistically, as a system. The class in which the experiment will be conducted has not been specified yet, although my purpose is to perform it in a first grade of Greek Lyceum class (ages 15-16). I hope that I will be able to realise this intervention outside school hours, in an informal environment, because students seem to lose interest in such experiments when they are performed with the instructional goals of the official curriculum. My purpose is to study the learning contexts and processes when computer games are used in accordance with, but not as a formal part of, the official school curriculum. This decision also emerges from the specific environment formed by the Greek examination system, which requires full attachment to the curriculum's superficial goals in order to achieve University admission. More precisely, I believe that no students, parents or teachers will allow such an experiment to be conducted in school hours, because this will be considered a loss of time. The research will be conducted in an informal learning setting for at least 4 weeks. The participation of the students will be voluntary. Mixed groups of students, (ie Greeks and foreigner nationals) will be preferred. Since the object of the study is the interactions occurring during the game, participation does not consider prior knowledge of the subject taught as a prerequisite.

3 Research Limitations/Implications

Replaying history with the purpose to entertain the gamer, as happens with historical strategy games, has specific limitations to start with. The simulative and entertainment nature of such games leads to oversimplification of historical issues and phenomena that might grow up in misconceptions. Regarding the understanding of historical notions, one of the main concerns arises from the danger of oversimplification emerging from the commercial and not educational nature of the game. This danger suggests that the curriculum designed to accompany game-play has to be aligned with the specific historical features and the ways of explaining and reconstructing the past we want to focus on. Aims and goals of the intervention must be identified, and the design of teaching must be closely attached to those. More particularly, careful design means to appreciate the specific learning outcomes we want to achieve in each part of the process and follow them closely.

One of the limitations of the research is expected to be the difficulty of the game, especially with regard to students with no prior experience in computer game play. Therefore, it is expected that the time needed to explain the structures and the guidelines of the game, combined with the amount of time gamers usually spend on playing such open- ended games (i.e. with no specific winning results) may be more than that initially estimated. Other important limitations are expected to emerge from the technological nature of the experiment, such as the limited number of PCs and the possible technological problems that may surface during the experiment, which I hope to surmount with the help of experienced Informatics professionals attending the play sessions and the careful design of the experiment context [21]. Finally, I hope to surmount the language barrier posed by the English speaking game, by providing the students with translations of the words they will be encountering through game play and constantly helping them with any unknown words or meanings.

4 Practical Implications

Educators have been advocating the use of computer games as educational tools, but research regarding what actually happens in the classroom, seen as a learning system, when computer games are used, has been insufficient. Computer games seem to enhance motivation, imagination, decision making skills and immersion, which are necessary conditions in order to achieve meaningful learning. Modern education needs to learn from the commercial success of computer games, and integrate those features that make them such an immersive medium in applications of online learning, interactive learning environments etc. I hope that researching the topic of knowledge construction and transition through computer game play will provide educators and researchers with substantial information regarding the design of games with educational character. Educational technologists could profit from the understanding we can gain from efforts like the present research, in designing educational software that will embody elements of game-play that enhance learning and motivation. Explaining and understanding how such alternative learning tools achieve learning, and what kind of learning outcomes they have, can lead to better understanding of the

learning process in such contexts and promote the design of software that embodies more entertaining elements, engaging thus the students in learning. The second implication my research is expected to have is to inform all interested parties about the context of game play and how it affects knowledge construction. Computer games are a very popular medium, on which children may spend hundreds of hours. It is very important to understand how knowledge occurs in the contexts created by this medium, and use this knowledge in order to promote learning in ways that interest the children; namely use the features of game-play that make it such a fun way to spend their time on. Alternative ways to approach children in order to teach them are necessary to develop, if we want to keep children to school. Substantial learning occurs only if we achieve internal motivation to advance to higher levels of understanding, and this seems to be a lesson computer games can give us, since they achieve deep immersion and profound interest. One other implication is the understanding of how history is conceived in popular culture today, going into relevant efforts of contemporary historians [22]. Especially since history seems to interest people only after they finish school, in the form of a hobby, and one of the contexts we learn history in is historical simulation gaming, it is of great interest to see how history is constructed during game-play.

5 Originality/Value

The need to apply modern education methods and use contemporary material for educational purposes has been acknowledged once again after the realization of the excessive use of computer games and ICT in general as entertainment tools [23, 24]. Game studies scientists have been advocating the use of computer games in the classroom, in order to achieve constructivist learning. Nevertheless, very few educational researchers and pedagogists have researched the process of knowledge construction through playing open- ended games [9], although the need to address educational process as a constructive one has been widely acknowledged. Hopefully, my research will contribute to a deeper understanding of the processes through which learning occurs. This knowledge could be of great assistance in the areas of instructional game design and history teaching.

References

1. Burn, Andrew & Carr, Diane: "Defining Game Genres", in *Computer Games: Text, Narrative and Play*, Diane Carr, David Buckingham, Andrew Burn and Gareth Scott Eds. Polity Press (2006), 14--29, p. 17
2. O' Neil, Harold, Wainess, Richard, Baker, Eva. L.: "Classification of learning outcomes: evidence from the computer games literature", *The Curriculum Journal*, 16:4, 455--474 (2005)
3. Squire, K.: "Video Games in Education", (2003). [Online]. Available: http://www.cyberfest.us/Education/Video_Games_in_Education-MIT_Study.pdf [Accessed: 13 February 2009]

4. Buckingham, David: "Studying computer games", in *Computer Games: Text, Narrative and Play*, Diane Carr, David Buckingham, Andrew Burn and Gareth Scott Eds. Polity Press (2006), 1--13
5. Gee, J.P.: *What Video Games Have to Teach Us About Learning and Literacy,* New York : Palgrave Macmillan (2003)
6. Prensky, M.: "Digital Natives, Digital Immigrants part II: Do they really think differently?", *On the Horizon*, MCB University press, 9 (December 2001), Available: http://www.marcprensky.com/writing/Prensky%20-%20Digital%20Natives,%20Digital%20Immigrants%20-%20Part2.pdf [Accessed 13 February 2009]
7. Taylor, J. A., Duran, M.: "Teaching Social Studies with Technology: New Research on Collaborative Approaches", *History Teacher*, 40: 1, 9--25 (2006)
8. Squire, K.: "Replaying history: Learning world history through playing Civilization III", Unpublished PhD thesis, Indiana University (2004)
9. Squire, K.: "Cultural Framing of Computer/ Video Games", *Game Studies*, 2:1, 2002, [Online]. Available: http://gamestudies.org/0102/squire/ . [Accessed 13 February 2009]
10. McMichael, A.: "Pc games and the teaching of history", *The History Teacher*, 40:2, 203--218 (2007)
11. Stake, R.: *The art of case study research*, Thousand Oaks, CA: Sage Publications. (1995)
12. Cobb P., Confrey, J., DiSessa, A., Lehrer, R., Schauble, L.: "Design Experiments in Educational Research", *Educational Researcher,* 32: 1, 9--13 (2003)
13. Barab S. and Squire, K.: "Design- Based Research: putting a stake in the ground", *Journal of the Learning Sciences*, 13:1, 1--14 (2004), p. 2
14. Hoadley, C.: "Methodological alignment in Design- Based Research", *Educational Psychologist*, 39:4, 203--212 (2004), p. 205
15. The Design- Based Research Collective: "Design- Based Research: An emerging paradigm for educational inquiry", *Educational Researcher*, 32: 1, 5--8 (2003)
16. Barab, S. and Plucker, J.: "Smart people or smart contexts? Cognition, ability and talent development in an age of situated approaches to knowing and learning", *Educational Psychologist*, 37:3, 165--182 (2002)
17. Barab, S., Barnett, M., Yamagata- Lynch, L., Squire, K., Keating, T.: "Using activity theory to understand the systemic tensions characterising a technology- rich introductory astronomy course", *Mind, Culture and Activity*, 9:2, 76--107 (2002)
18. Brown K. & Cole, M.: "Cultural Historical Activity Theory and the Expansion of Opportunities For Learning After School", (n.d.), [Online]. Available: http://lchc.ucsd.edu/People/MCole/browncole.html [Accessed 13 February 2009]
19. Kaptelinin, V.: "Activity Theory: implications for Human- Computer Interaction", revised version of a paper that appeared in *Proceedings of NATO Advanced Study Institute on Basics of Man-Machine Communication for the Design of Education Systems*, M. Brouwer-Janse and T. Harrington Eds. Amsterdam: Springer-Verlag, (1993). [Online]. Available: http://www.ics.uci.edu/~corps/phaseii/nardi-ch5.pdf [Accessed 13 February 2009]
20. Kuuti, K.: "Activity theory as a potential framework for human- computer interaction research", in *Context and consciousness: activity theory and human computer interaction*, B. Nardi Ed. Cambridge: MIT Press, 17--44 (1995). [Online]. Available: http://www.dwr.bth.se/kari_kuutti%20Nardi_book.pdf [Accessed 8 April 2009]
21. Egenfeldt- Nielsen, S.: "Practical barriers in using educational computer games", *On the Horizon*, 12: 1, 18--21 (2004)
22. De Groot, J.: *Consuming history, Historians and heritage in contemporary popular culture*, Routlege, London (2009)
23. Prensky, M.: "Digital Natives, Digital Immigrants", *On the Horizon*, MCB University Press, 9:5 (2001). [Online] Available: http://www.marcprensky.com/writing/Prensky%20-

%20Digital%20Natives,%20Digital%20Immigrants%20-%20Part1.pdf [Accessed 13 February 2009]
24. Shaffer, D.W., Squire, K., Halverson, R., Gee, J. "Video Games and the Future of Learning" (n.d), University of Wisconsin-Madison and Academic Advanced Distributed Learning Co-Laboratory, [Online]. Available: http://www.discoverproject.net/italy/images/gappspaper1.pdf [Accessed 13 February 2009]

Circular Fuzzy Iris Segmentation

Nicolaie Popescu-Bodorin

Department of Mathematics and Computer Science, Spiru Haret University
Ion Ghica 13 , Bucharest 3, Romania
bodorin@ieee.org

Abstract. This paper proposes a new approach to iris segmentation. In the classical iris segmentation procedures (Wildes, Daugman), pupillary and limbic boundaries are identified by solving three-dimensional optimization problems in order to find a radius and two center coordinates via gradient ascent or by using the Hough Transform or by iterating active contours. Here, iris boundaries are found by solving one-dimensional optimization problems. The proposed Circular Fuzzy Iris Segmentation procedure is designed to guarantee that similar segmentation results will be obtained for similar eye images, despite the fact that the degree of occlusion may vary from an image to another. Pupil finding and limbic boundary approximation are both fuzzy approaches based on k-Means and Run Length Encoding. The result of the proposed segmentation procedure is a circular iris ring (concentric with the pupillary boundary) which approximates the actual iris segment. When two similar images of the same eye are compared, the detected circular iris rings are pointing to the same physical support, possibly occluded by eyelids, eyelashes, specular and lighting reflections.

Key words: iris segmentation, circular fuzzy iris segmentation

1 Introduction

It is currently accepted that finding one iris boundary (pupillary or limbic) means to identify a collection of pixels as a solution of a three-dimensional optimization problem [1]:

- filtering the edges until the finding of a (nearly) circular contour minimizing the integro-differential Daugman operator [2];
- filtering the edges to find circles approximating the iris boundaries using Hough transform [3];
- iterating active contours to find optimal positions matching the iris boundaries [4].

Details about the current segmentation techniques can be found in the refered papers. The iris segmentation procedure proposed here is an attempt to reconsider the iris segmentation not as an explicit optimization problem (which typically has a high computational complexity) but as an implicit one, which can be solved by parsing and classifying chromatic values.

A notion that is used very often in this paper is the *k-means equipotential chromatic map*. It was introduced in [5] together with the *Fast k-Means (Image) Quantization*

(FKMQ) algorithm whose properties were further detailed in [9]. Also, it have to be said that the current proposed solution for finding the pupil has its roots in an attempt to design a real-time version for a computer-assisted diagnosis tool which detects the response to hepatitis viral infection in an image representing a patient serum [6], [7].

A basic idea exploited in this paper is that a target signal can be enhanced against unwanted noise if a suitable filter can be designed for this purpose.

The second section of this paper is dedicated to pupil finding. The proposed procedure is meant to enhance the pupil against its surroundings and against the occlusions that may occur in its area (specular and lighting reflections or eyelashes). There are two main operations at this step:

- a k-means quantization of the eye image is applied to obtain its *k-means equipotential chromatic map* in which the actual pupil is the most circular solid object contained in the lowest level (this lowest level will be referred further to as the *pupil cluster*);
- a fuzzy membership assignment of the pixels within the *pupil cluster* to the actual pupil is constructed using Run Length Encoding (RLE) and requantizing run length coefficients in unsigned 8-bit integer domain. The result is a *Run Length Quantization* of the *pupil cluster*. The defuzzification is achieved by running FKMQ once again to determine a threshold above which the membership of a pixel to the actual pupil is guaranteed. As a result, a pupil indicator is computed. From each of its pixels a flood-fill operation can be started to determine the available pupillary boundary which can be further approximated through a circle or an ellipse.

The finding of the most representative circular ring of the iris is presented in the third section and it is achieved by building three fuzzy membership assignment functions, each of them mapping all the circles concentric with the pupil to the actual pupil, to the actual iris segment and to the actual non-iris area. The defuzzification is achieved by counting the votes that each circle receives from each of those membership functions.

Benchmark details, comments and conclusions are given in the fourth section in order to illustrate the degree of robustness and the range of applicability of the proposed iris segmentation procedure.

2 Computational Routines

Run Length Encoding (RLE) is one of the most simple and popular data compression algorithms [8]: for a given array, any subarray of redundant values is coded as a pair representing its histogram, as shown in the following example:

if $\mathbf{V} = [1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1]$ is the vector to be encoded, then the run length encoding of $\mathbf{V}$ is: $rle(\mathbf{V}) = [(1, 4), (0, 2), (1, 8)]$.

Run Length Quantization for Binary Images is defined here as a procedure to replace all ones within a binary image with the corresponding run length coefficients re-quantized in unsigned 8-bit integer domain by some custom quantization function, as illustrated in the following example:

$$[1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1] \Rightarrow [(1, 4), (0, 2), (1, 8)] \Rightarrow$$
$$[(4, 4), (0, 2), (8, 8)] \Rightarrow [(128, 4), (0, 2), (255, 8)] \Rightarrow$$
$$[128, 128, 128, 128, 0, 0, 255, 255, 255, 255, 255, 255, 255, 255]$$

The (re-)quantization function used in the above example (and further in this paper) is:

$$\mathrm{rqf}(\mathbf{V}(\mathbf{n})) = \min(255, \max(1, \mathrm{round}(255 * \mathbf{V}(\mathbf{n}) / \max(\mathbf{V}(\mathbf{n}))))) \,,$$

where $\mathbf{V}$ is the vector to be quantized and $\mathbf{n}$ is the index of its non-zero components.

Run Length Quantization procedure encodes a morphological property of the input image into a new signal (re-quantized image) by giving the same chromatic meaning to all of those white pixels sharing the same *Run Length* coefficient.

Fast k-Means Image Quantization (FKMQ) [5] is a variant of k-means algorithm designed for fast chromatic clustering in unsigned 8-bit integer domain. It transforms the input image in an equipotential chromatic map [9] with k levels by replacing each chromatic value with the closest centroid. A suboptimal (incomplete) variant of FKMQ can be easily derived by imposing termination in a small number of iterations while resetting the first centroid to the minimum available value (or to zero). In this way, the input image is forced to return a handler to that area covered by lower chromatic values. This is particularly useful in detecting the pupil in those eye images in which the pupil is the darker zone.

3 Finding the Pupil. Fast Pupil Finder Algorithm

For an eye image like that in the Fig.1.a (0006-L-0008.j2c file from Iris Database ©University of Bath), one can compute the *k-means equipotential chromatic map* by applying the FKMQ algoritm (Fig.1.b). The lowest level (cluster) on this map (Fig.1.c) is the *pupil cluster* (PC).

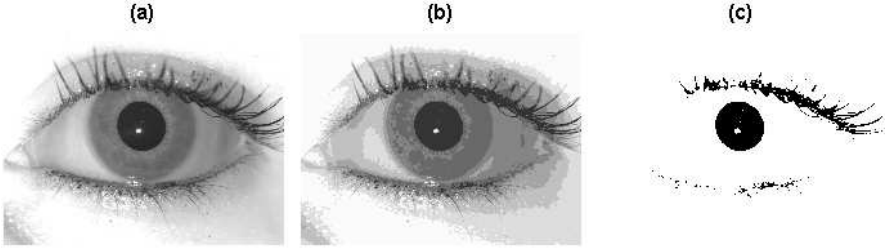


Fig. 1. The original eye image (a), its 8-means quantization (b), and the pupil cluster PC (c).

The first problem to be solved at this stage is finding at least one pixel in PC which belongs for sure to the actual pupil (finding a suitable pupil indicator). By computing the Haar wavelet decomposition pyramid for the pupil cluster it can be seen that finding a suitable pupil indicator means to erode the pupil cluster up to such a scale above which the details come only from the actual pupil. In such a stage, the subsampled frame of the wavelet pyramid is a suitable pupil indicator itself. As a consequence, defining an adaptive erosion procedure is the key for finding a suitable pupil indicator: for each

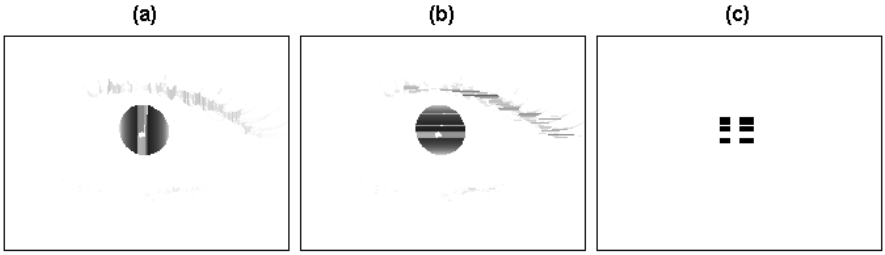


Fig. 2. Vertical Run Length quantization of the pupil cluster (a); Horizontal Run Length quantization of the pupil cluster (b); The pupil indicator PI (c).

pixel within the pupil cluster, the structuring element of the erosion is adaptively determined by the morphological context in the neighborhood of that pixel, meaning that the vertical and horizontal Run Length Encoding is used to requantize (in unsigned 8-bit integer domain) the run length coefficients corresponding to that pixel (the degree of membership of that pixel to a shorter or a longer continuous segment lying in the pupil cluster). The requantized vertical/horizontal run length coefficients encode the degree of membership of each pixel to the actual pupil. The argument is fact that being (or containing) the most circular solid object within the pupil cluster, the actual pupil is the most resilient set to erosion [10] to be found in the pupil cluster.

Let us consider the matrices RLV (Fig.2.a) and RLH (Fig.2.b) as being the vertical and horizontal Run Length quantization of the pupil cluster PC, respectively. The following computational procedure results in finding a pupil indicator for a given pupil cluster:

```
function [k,PI] = getpi(RLH, RLV);
  k=16; PI = 0*RLH;
  While PI is the null matrix do:
    Compute the k-means quantization of RLV and RLH:
       RLHQ = fkmq(RLH, k);
       RLVQ = fkmq(RLV, k);
    Select the logical index of the highest cluster
    within RLHQ and RLVQ, respectively:
       LIH = ( RLHQ == max(RLHQ(:)) );
       LIV = ( RLVQ == max(RLVQ(:)) );
    Compute the binary matrix PI as logical
    conjunction of LIH and LIV:
       PI = LIH & LIV;
    k = k -1;
  EndWhile;
END.
```

Each pixel within the pupil indicator can now be used as a starting point for a flood-fill operation. The most accurate pupil segment available in the pupil cluster is identified

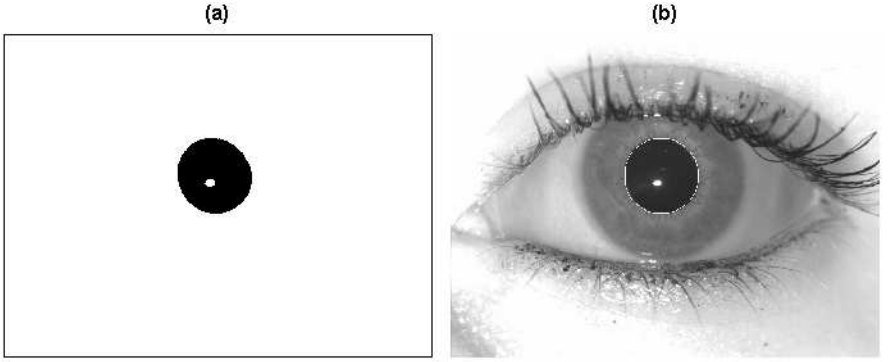


Fig. 3. Available pupil segment (a); Ellipse approximating the pupillary boundary (b).

(Fig.3.a) this way. Further, the specular lights are filled using Run Length Encoding once again. The result is then fitted into a rectangle and approximated by an ellipse (Fig.3.b).

Summarizing all the operations described above, the proposed Fast Pupil Finder algorithm can be stated as follows:

Fast Pupil Finder Algorithm (N. Popescu-Bodorin):

```
INPUT: the eye image IM;
 1.Extract the pupil cluster:
     PC = fkmq(IM,16);
     PC = (PC == min(PC(:)));
 2.Compute horizontal and vertical
    Run Length quantization of PC:
     RLV(:,,j) = vrleq(PC);
     RLH(j,:) = hrleq(PC);
 3.Compute the pupil indicator PI:
     [k, PI] = getpi(RLH, RLV);
     PI = find(PI == 1);
     PI = PI(1);
 4.Extract available pupil segment
    through a flood-fill operation:
     P = imfill(PC, PI);
 5.Fill the specular lights:
     P = rlefillsl(P);
 6.Approximate the pupil by an ellipse;
OUTPUT: The ellipse approximating the pupil;
END.
```

It can be seen in [10] that the most frequent causes of false rejection are the occlusions and the segmentation errors. This is the reason why accurate (pupil / iris) segmentation is a must in order to achieve a certain degree of performance in recognition.

In this context, a big issue is the fact that numerical segmentation results (numerical values identifying pupil centers and iris radii) for Bath University Iris Database are not publicly available at this time. That is why the accuracy of the iris segmentation procedures proposed in this paper is to be directly proven by the iris recognition results further presented in [11] (Gabor Analytic Iris Texture Binary Encoder).

4 Circular Fuzzy Iris Segmentation

The Fast Pupil Finder procedure guarantees an accurate pupil localization and enables us to unwrap the eye image (Fig.4.a - 0009-L-0007.j2c file from Iris Database ©University of Bath) in polar coordinates (Fig.4.b) and also to practice the localization of the limbic boundary in the rectangular unwrapped eye image (Fig.4.c), obtaining an iris segment as in Fig.4.e.

The proposed Circular Fuzzy Iris Segmentation procedure is currently implemented using Matlab and calibrated for University of Bath Iris Image Database (the free version [12]) and it can be stated as follows:

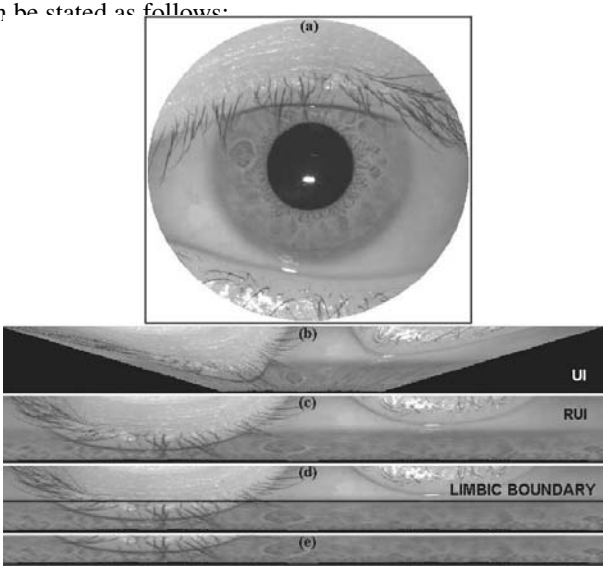


Fig. 4. Iris segmentation stages

Circular Fuzzy Iris Segmentation Procedure (N. Popescu-Bodorin):

```
INPUT: the eye image IM;
 1.Apply the Fast Pupil Finder procedure
    to find pupil radius and pupil center;
 2.Unwrap the eye image in polar
    coordinates (UI - Fig.4.b);
 3.Strech the unwrapped eye image UI
    to a rectangle (RUI - Fig.4.c);
```

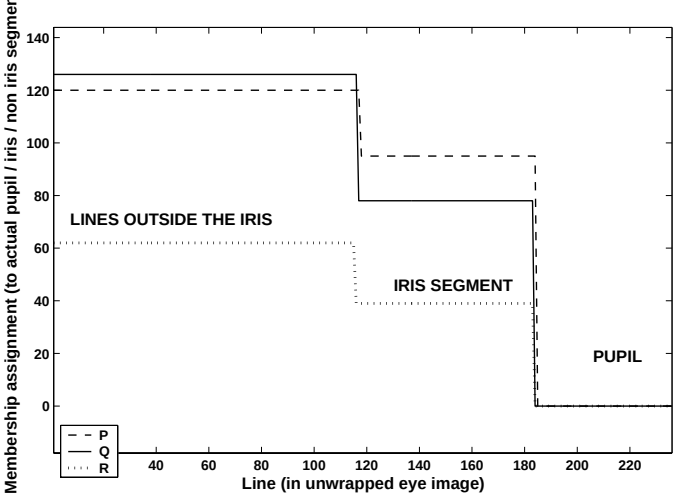


Fig. 5. Iris segmentation procedure: Line assignment (step 5)

```
 4.Compute three column vectors: A, B, C,  where:
    A contains the means of the lines within UI matrix;
    B contains the means of the lines within UI matrix;
    C contains the means of the lines within [A B] matrix;
 5.Compute P, Q, R as being 3-means
    quantization of A, B, C, respectively (Fig.5);
 6.For each line of the unwrapped eye image, count
    the votes given by P,Q and R. All the lines receiving
    at least two positive votes is assumed to belong
    to the actual iris segment;
 7.Find limbic boundary and extract the iris segment
    (Fig.5, Fig.4.d, Fig.4.e);
OUTPUT: pupil center, pupil radius, index of the line
  representing limbic boundary and the final iris segment;
END.
```

5 Results and discussions

The accuracy of the iris segmentation algorithm is very good for all the images in the tested database (1000 images / 50 unique eyes). For illustration, a demo program (Matlab) is available for download [13]. The current demo version enables pupil localization in 12 frames per second and limbic boundary localization in 5 frames per second, for eye images of dimension 240x320 pixels.

When a single detector is used for finding the limbic boundary, wrong segmentation results are observed for eight images. To avoid these errors, the demo program uses two limbic boundary detectors (complementary to each other), both of them being variants of the above Circular Fuzzy Iris Segmentation procedure.

An important feature of the proposed iris segmentation approach is the fact that the average time spent finding the limbic boundary is just two times greater than the time

spent finding pupillary boundary. The speed of the computation is achieved by formulating pupil finding and limbic boundary approximation as one-dimensional optimization problems and also by using fast computational procedures: Fast k-Means variants [9] and Run Length Encoding. K-Means and Run Length Quantization are used to encode the meaningful information, avoiding any unnecessary traversal of the input image.

Circle finding is usually achieved by solving an optimization problem with three parameters (two center coordinates and radius) varying within a rectangular parallelepiped in $\mathbf{R^3}$. Here, the computation of the pupil indicator depends on a single parameter (a threshold for the requantized horizontal and vertical run length coefficients computed for the pupil cluster, threshold above which the membership of a pixel to the actual pupil is guaranteed). On the other hand, the limbic boundary is determined searching for just a line number identifying the first iris line (see the first iris circle in Fig.6, or the first line of the unwrapped iris in the same figure, or the limbic boundary in Fig.4.d).

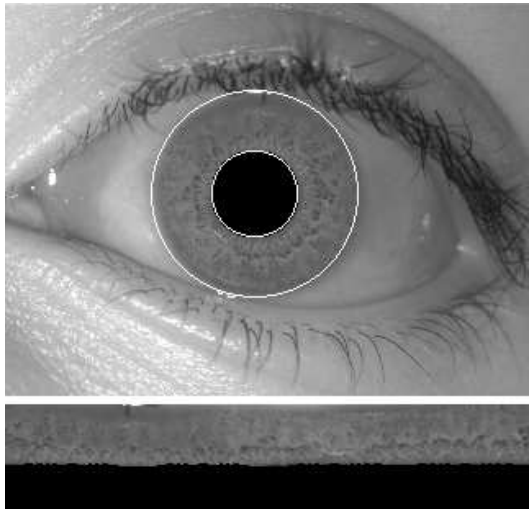


Fig. 6. Circular Fuzzy Iris Segmentation Demo Program

The range of applicability of the proposed segmentation procedure is limited to that class of images which satisfy two working hypotheses:

- the correlation between the area morphology and chromatic variation is sufficiently strong, or else, the initial k-means quantization of the eye image will fail to return an accurate pupil cluster;
- the pupil is (or contains) the biggest, most circular and darker object, or else, other object will prove a higher resiliency to erosion than the actual pupil and a false pupil indicator will be computed in consequence.

In short, the efficiency of the proposed algorithm proves that iris segmentation can be treated as being a one-dimensional optimization problem if there is enough accurate morphological information stored as chromatic variation.

Acknowledgments. The author wishes to thank Professor Luminita State (University of Pitesti, RO) for the comments, criticism, and constant moral and scientific support during the last two years. The author also wishes to thank Professor Donald Monro (University of Bath, UK) for granting the access to the Bath University Iris Database. Without their support this iris recognition study could not have been possible.

References

1. Bowyer, K.W., Hollingsworth, K., Flynn, P.J.: Image Understanding for Iris Biometrics: A survey, Elsevier CVIU, 110 (2), 281-307, May 2008.
2. Daugman, J.G.: High Confidence Visual Recognition of Persons by a Test of Statistical Independence, IEEE TPAMI, Vol. 15, No. 11, November 1993.
3. Wildes, R.P.: Iris Recognition - An Emerging Biometric Technology, Proceedings of the IEEE, Vol. 85, No. 9, September 1997.
4. Daugman, J.G.: New Methods in Iris Recognition, IEEE TSMC, vol. 37, no. 5, october 2007.
5. Popescu-Bodorin, N., State, L. , Optimal Luminance-Chrominance Downsampling through Fast K-Means Quantization, DSC 2008, Proceedings of, vol. 2, pp. 111-125, SEERC, June 2008.
6. State, L., Paraschiv-Munteanu, I., Popescu-Bodorin, N.: Blood corpuscles classification schemes for automated diagnostic of hepatitis, Scientific Bulletin, No.14/2008, University of Pitesti, `http://fmi.spiruharet.ro/bodorin/articles/hdiag1.pdf`, 2008.
7. State, L., Paraschiv-Munteanu, I., Popescu-Bodorin, N.: Blood corpuscles classification schemes for automated diagnosis of hepatitis using RLE and ISODATA algorihtm, Scientific Bulletin, No.16/2009, University of Pitesti, `http://fmi.spiruharet.ro/bodorin/articles/hdiag2.pdf`, 2009.
8. Salomon, D.: Data Compression: The Complete Reference, p. 15, Springer-Verlag, 2006.
9. Popescu-Bodorin, N.: Fast K-Means Image Quantization Algorithm and Its Application to Iris Segmentation, Scientific Bulletin, No.14/2008, University of Pitesti, `http://fmi.spiruharet.ro/bodorin/articles/fkmqis.pdf`, 2008.
10. Ma, L., Tan, T., Wang, Y., Zhang, D.: Efficient Iris Recognition by Characterizing Key Local Variations, IEEE TIP, vol. 13, no. 6, June 2004.
11. Popescu-Bodorin, N.: Gabor Analytic Texture Encoder, accepted in DSC 2009 Conference, SEERC, Thessaloniki, Greece, July 2009, `http://fmi.spiruharet.ro/bodorin/articles/gaitbe.pdf`, 2009.
12. University of Bath Iris Database, http://www.bath.ac.uk/elec-eng/research/sipg/irisweb/, June 2009.
13. Popescu-Bodorin, N.: Circular Fuzzy Iris Segmentation Demo Program, `http://fmi.spiruharet.ro/bodorin/arch/cffis.zip`, 2009.
14. Pegden, W.: Sets resilient to erosion, Dept. of Mathematics, Rutgers University, `http://people.cs.uchicago.edu/~wes/erosion.pdf`, June 2009.

Parallel Visualization on GPU with CUDA

Cosmin Marian Poteras[1], Mihai Gosa[1] and Mihai Mocanu[1]

[1] Faculty of Automatics, Computers and Electronics, Software Engineering Department, University of Craiova,
Bvd. Decebal 107, 200440, Craiova, Dolj, Romania
{poteras_cosmin, gosamihai, mocanu_mihai}@software.ucv.ro

Abstract. Developing applications able to do parallel visualization and to adapt to a continuously evolving hardware platform is a challenging task. One of the applications of this type, volume rendering, can reach optimal speedup through careful programming. There are already some well established approaches in volume visualization that exploit its high degree of intrinsic parallelism, i.e. ParaView from Kitware Inc.[1]. In the past few years one could also notice an impressive development of Graphical Processing Units (GPU) which became highly parallel systems. Due to this development Nvidia released Compute Unified Device Architecture guiding GPUs on their way to general purpose programming. Merging these two threads of development (CUDA and volume rendering) might prove to be an interesting idea. Since our work is still in progress we will avoid presenting concrete performance results. Instead we will resume a few important guidelines which need to be considered when trying to render volumes using CUDA.

Keywords: Parallel visualization, Volume Rendering, GPU, CUDA.

1 Introduction and Related Work

Visualization of very large datasets is a challenging visualization problem that is often suitable for schemes in which multiple threads of control can be employed to work in parallel on the same image. In the past few years the computing industry has witnessed to a very fast and impressive development of GPUs (Graphical Processing Units) which became highly parallel, multithreaded processors, and, as a consequence, the manufacturers began to develop APIs to allow GPUs to find their way through the general purpose processing. As GPUs have been designed to do very intensive arithmetic computations, exactly what computers graphics is all about, they become very well-suited for applications that require data-parallel processing, offering a new breath in fields where very large data sets need analyzing and processing (medical visualization, weather forecasting, real time systems).

Ray casting a volume is a very tedious computational task. Many scientific efforts concentrated on improving performance when dealing with such a task. Kitware Inc. developed a CPU implementation of ray casting with multithreading support in their well known Visualization Toolkit [1]. Trying to run this implementation on a high-scale volume data set led us to the conclusion that CPU is unable to ray cast a volume

and maintain a high level of interactivity. To overcome this drawback, researchers have focused on using the GPU whose computational horse power and many-core parallelism are continuously increasing. At first the GPU could be programmed to perform custom tasks by using shader languages [7], making the implementation uncomfortable and losing performance by using the GPU to perform a task it was not designed for. NVIDIA opened new opportunities by releasing the Compute Unified Device Architecture (CUDA) which came up as a C language extension, making the implementation a lot easier, and improving the performance since the instruction set has been designed to use the hardware behind in an efficient way. We employ in our investigation the CUDA (Compute Unified Device Architecture) model, released in November 2006 from NVIDIA to facilitate general purpose parallel processing on GPU devices. We intend to fully integrate our software with the Visualization Toolkit developed by Kitware Inc [1].

We will point out in this paper the important gaps and performance criteria which arise when dealing with CUDA volume rendering. Almost any approach to solving the volume rendering problem by using the GPU used to be based on texture-based volume rendering. Mainly, ray casting concept employed a stack of parallel slices for sampling. Based on the samples, the pixels values were blended in a back-to-front or front-to-back manner in the frame buffer.

The success of those approaches comes from the parallel character of the pixel processing during rasterization. Based on this simple method, there have been proposed many optimizations like: skipping empty space, early ray termination when the resulting pixel has reached its full opacity, and others. In what follows we will shortly present a few existing GPU implementations of ray casting algorithm.

In [2] the authors studied the various bottlenecks found in the graphics hardware when performing GPU volume rendering and they propose a strategy for balancing the load without compromising the image quality. As an optimization they also use early ray termination. The performance was very promising. Paper [3] presents an extensible framework which allows the programmer to load different shaders for volume rendering. Together with their framework they present sample shaders for Maximum Intensity Projection, refraction, reflection, isosurface, and others. The hardware platform they addressed is the NVIDIA GeForce 6800, which could become a drawback in case of a different hardware platform. The performance results they presented against the slice-based volume rendering don't show big improvements. In [4] is addressed the issues which arise when dealing with virtual endoscopy, mainly the perspective ray casting since there is no near clipping plane, and the camera should be able to enter the volume. The authors propose a framework that allows rendering to interactive frame rates for almost every possible dataset, as well as quality settings and rendering modes. The solutions presented above are based on a compromise between programmability and speed. This is because the hardware they use allows only a fixed set of graphics operations. The parts that needed more flexibility were usually addressed by the CPU, until recently. In order to put together flexibility and programmability NVIDIA released the Compute Unified Device Architecture [5] whose main role was to allow a high level of programmability and flexibility at the same time. In section 3 we will examine the advantages of CUDA but also the constraints and performance criteria which need to be considered in order to achieve maximum performance and correctness.

2 Basic Algorithms

In this section we will offer an overview of the ray casting algorithm. Ray casting is a popular approach for volume rendering, based on the concept of sending rays into the volume and collecting color information along the ray. Considering the camera position and also that the viewer always sees the scene through a screen, one can send rays parallel to the viewing direction, through each pixel, against the volume. If the ray intersects the volume, color samples are extracted at fixed intervals. The resulting color associated to the current pixel will be computed according to a ray function. We can resume the ray casting algorithm to three main phases:

1. Send rays against volume.
2. Extract sample values.
3. Combine the extracted samples and assign output color to current pixel.

Depending on the chosen ray function, steps 2 and 3 might be combined. The samples' colors can be extracted by using different techniques ranging from the nearest neighbor's color from a more complex tri-linear interpolation. There are few commonly used ray functions. We will only mention three of them: Maximum Intensity Projection whose output color is the one with maximum intensity, Average Intensity Projection which computes an average of extracted samples and a more complex one, Compositing which combines all samples encountered, leading to the output color. Compositing ray cast functions extract samples' colors by employing a transfer function which points out the actual color and opacity information. The input to the transfer function might be the voxel's gray color itself or a gradient computed in that point. The advantage of using a transfer function is that one can visualize the same volume using different color and opacity schemes. This algorithm is suitable for highly parallel implementations, as it can be observed, since the same computation needs to be performed for each ray and the number of rays is usually very high. Section 4 will present our parallel design of the ray casting algorithm.

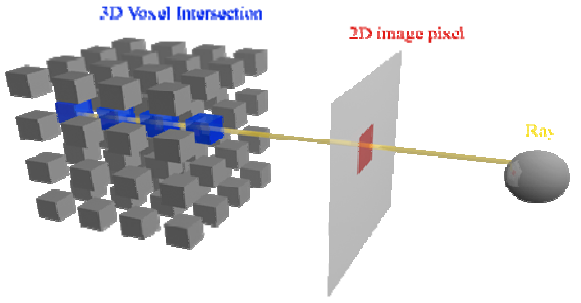


Fig. 1. Ray Casting model.

3 Compute Unified Device Architecture

Having as its core a complex threads hierarchy, a diverse memory hierarchy and synchronization capabilities, CUDA came up as a powerful extension of the C programming language, allowing programmers already familiar with C to adapt very fast and write parallel applications without difficulties [5]. In what follows we will examine the most relevant features of CUDA which makes it very suitable for applications which require a highly parallel data processing, as it's the case for our ray casting algorithm.

3.1 The CUDA Model

The CUDA hardware model, illustrated in Fig.2 is based on a set of SIMD multiprocessors. Each multiprocessor contains three types of memory (shared memory, texture memory, constant memory) and a set of processor which benefit of a small set of register. Each of the processors execute at every clock cycle the same instruction on it's own data, allowing this way a high number of threads to execute concurrently. The CUDA device is assimilated as a co-processor to the CPU (called host), optimized for computationally intensive tasks. The boundary between the CPU and the GPU is called a kernel which is nothing more than a function which can only be called from the host device (CPU) and executes on the device. When invoked, a kernel starts it's execution on the GPU on as many threads as specified in the launch configuration. The kernel launch configuration instructs the device on how to allocate and organize the threads. CUDA transforms the GPU into a highly parallel co-processor of the CPU whose execution and resources are managed from the host computer through CUDA runtime and CUDA libraries.

The CUDA architecture might also be thought of as Single Instruction – Multiple Thread supported by mapping each thread on its own core and by employing a flexible threads architecture which lead to a powerful hardware implementation independence. Fig.3 illustrates the CUDA threading model. Mainly the CUDA threads are grouped together in bi/tri-dimensional blocks which in turn are grouped in one/bi-dimensional grid. Each of the threads is assigned a unique identifier based on its block's position inside the grid as well as on its own position inside the block of threads. Based on this identifier one can determine a set of data which will be processed along the thread's execution. The threads inside the same block can share a reduced segment of shared memory and they can synchronize their execution by using the barrier synchronization. The blocks are allocated and executed independently of one another, offering the possibility of writing scalable code.

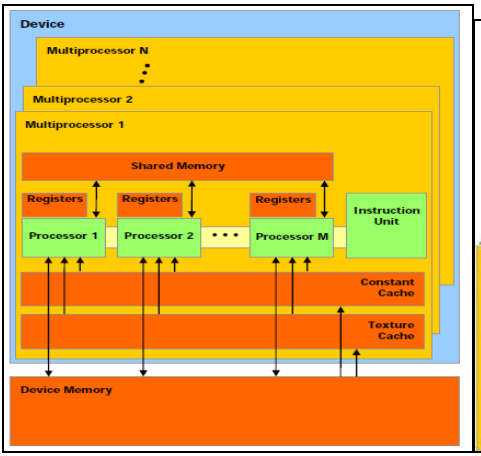


Fig. 2. CUDA Hardware Model.

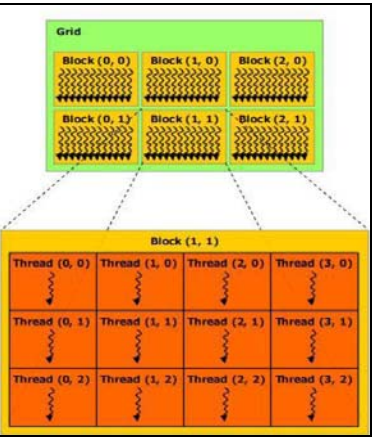


Fig. 3. CUDA Threading Model

3.2 Performance Considerations

Even though we can execute almost any C statement inside a kernel, a high performance will only be achieved if we take special care when dealing with CUDA. In this section we will discuss the most important performance considerations.

As mentioned above, due to the organization of threads, CUDA allows writing scalable code. When defining the launch configuration, one should make sure there are allocated enough blocks whose execution could be overlapped with memory accesses of other blocks and should also consider the future trends of GPUs development since the future devices might support the execution of a higher number of blocks concurrently.

The memory accesses are the most costly. Efforts need to be done so that each thread contains a high number of arithmetic instructions which could overlap with memory accesses. A good approach to achieve this functional criterion is called data pre-fetching [6] and consists of loading before the current processing step, the data set necessary to the next processing step. This way every processing step will be ready for execution without having to wait for a costly memory data transfer.

Depending on device's capabilities the global memory's data might require special alignments and coalescing in order to benefit of special memory transfer instructions [5]. Shared memory is a high speed memory but the amount available is far from being sufficient, which leads to a compromise between scalability and speed, since new blocks can not be allocated if there is not enough shared memory to satisfy the block's need.

4 Implementation and Integration

In order to achieve our goal we should have developed a complete volume visualization framework, which is a tedious work. Instead of doing this we decided not to implement such a framework from scratch, but rather to extend an existing one. Our choice is represented by the Visualization Toolkit (VTK) [1], an open source toolkit, developed at Kitware Inc. VTK already integrates a volume rendering framework, which is entirely implemented in software, making it unsuitable for very large dataset. We will present in this section the main elements of VTK's volume rendering framework together with the changes that need to be done to integrate our CUDA implementation.

Fig.4 illustrates the part of VTK's class hierarchy which contains the volume rendering pipeline. Their roles are further presented:

- vtkImageData – contains image data elements as loaded from disk.
- vtkVolumeRayCastFunction – base class for particular ray cast functions (vtkVolumeRayCastCompositeFunction, vtkVolumeRayCastMIPFunction, a.o.) which compute the ray samples and combine them, resulting the pixel's color
- vtkPiecewiseFunction – gray mapping piecewise function.
- vtkColorTransferFunction – maps the gray value to a color
- vtkCamera – represents the camera together with the viewing parameters of the scene
- vtkInteractorStyle – represents main user interaction handling style.
- vtkVolumeRayCastMapper – This is a software ray caster for rendering volumes in vtkImageData
- vtkVolumeProperty – contains the properties of a volume
- vtkVolume – contains the volume data, its properties and all operations which can be applied on it
- vtkRenderer – represents the rendering process objects.
- vtkRenderWindow – represents the window to render in
- vtkRenderWindowInteractor – provides mouse and keyboard interaction
- vtkVolumeRayCastStaticInfo – class used for collecting information needed for casting all rays.
- vtkVolumeRayCastDynamicInfo – class used for collecting information needed for casting a particular ray.

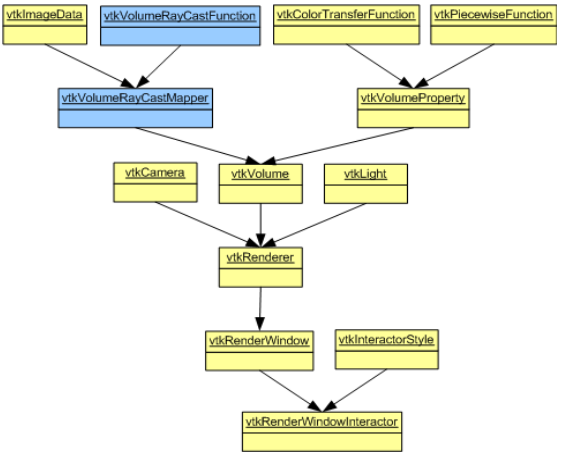


Fig. 4. VTK Volume Rendering classes

By examining this hierarchy and also considering the tasks requiring tedious work load, we identified a subset of the classes presented above, which should be replaced with new ones containing CUDA kernels instead of original processing methods. This subset is represented by the following classes:

- vtkVolumeRayCastFunction – casting rays, as well as data sampling and computing the resulting color according to the chosen ray cast function have been rewritten using CUDA. For each of it's derived classes there has to be implemented a new class using CUDA.
- vtkVolumeRayCastMapper – manages the calls to CUDA kernels, so it needs to be adapted.
- vtkVolumeRayCastStaticInfo and vtkVolumeRayCastDynamicInfo – needs to be altered in such a way that they don't contain references to any of the other VTK classes. Instead they will contain only independent data.

5 Conclusions and Future Work

Our approach for the implementation requires porting all existing code, to CUDA architecture, and applies successively the performance criteria suitable. Although the work is still in progress we can share few remarks related to the implementation of CUDA ray casting, and those are related to the performance criteria. It is very important to run on the device only the code which requires tremendous processing; otherwise the speed will be seriously affected. Also, conditional statements need to be avoided whenever possible since they might cause thread divergence. Considering the big difference in bandwidth terms between memory space, but also the amount of memory available one should make a compromise when deciding which memory space to use for each of the variables. By using a big amount of shared memory for

each block, memory access is very fast, but allocating too much shared memory, diminishes the number of block that can be allocated by the GPU.

For each of our software improvements, measurements will be taken in order to evaluate the impact of each of the performance criteria over the ray casting algorithm. We intend to continue the work and extend the CUDA implementation to other volume rendering techniques, like splatting or shear-warp.

References

1. Kitware, Inc. - VTK Textbook, 4th edition (December 2006), www.kitware.com.
2. Daniel Ruijters, Anna Vilanova – Optimizing GPU Volume Rendering, WSCG2006
3. Simon Stegmaier, Magnus Strengert, Thomas Klein, Thomas Earl – A Simple and Flexible Volume Rendering Framework for Graphics-Hardware-based Raycasting, Institute for Visualization and Interactive Systems, University of Stuttgart, Volume Graphics 2005
4. Henning Scharsach – Advanced Raycasting for Virtual Endoscopy on Consumer Graphics Hardware, Technische Universitat Wien, April 2005
5. NVIDIA – CUDA Programming Guide, www.nvidia.com
6. Wen-Mei, W. Hwu, David Kirk - University of Illinois, Spring 2009 Syllabus course
7. Klaus Engel, Martin Kraus, Thomas Ertl, High-Quality Pre-Integrated Volume Rendering Using Hardware-Accelerated Pixel Shading, Visualization and Interactive Systems Group, University of Stuttgart, Germany

Contrast Enhancement in Space Frequency Domain for Active Contouring of CT Data Sets Using Wavelet Transform

Ioana Cristina Plajer[1] and Detlef Richter[2]

[1] Transilvania Universiy of Brasov, Bulevardul Eroilor Nr. 30,
RO-500007 Brasov, Romania
iplajer@yahoo.com
[2] Universiy of Applied Sciences of Wiesbaden, Kurt-Schumacher-Ring 18,
D-65197 Wiesbaden, Germany
richter@informatik.fh-wiesbaden.de

Abstract. This paper describes two algorithms for local contrast enhancement of tumor boundaries in CT images. The enhancement is done in space frequency domain using the dyadic wavelet transform. The modulus maxima of the detail components at different scales are processed in the modulus image by a special designed operator. In the image obtained after applying the inverse wavelet transform, using the processed detail components, the tumor is contoured by an active contour algorithm. The results of the active contour algorithm of the not enhanced and of the enhanced image are compared. A second algorithm which uses boundary information is proposed in order to enable a better detection of the tumor, if the gray levels along the boundary are varying significantly.

Keywords: Wavelets, 3D CT-data, modulus maxima, local contrast enhancement

1 Introduction

In recent years medical imaging has become an important tool in many fields like diagnosis, surgery or medical research. A special focus is on cancer treatment, where medical image processing enables radiologists to precisely define the target volumes to avoid irradiating healthy tissues. The research area in this field is still very large, as different kinds of tumors in different tissues require specific solutions of the segmentation problem. In 3D CT-data sets segmentation is done separately in each slice to construct the target volume. At present, no fully automatic method exist to segment medical images efficiently. Therefore semiautomatic methods, using input from expert radiologists, have been designed. In this paper the results of contour detection are produced by an active contour algorithm based on the minimization of an energy function [1].

One of the main problems in correct boundary detection is the quality of the image data or the grey-level distribution in different tissues. Therefore some preprocessing of the image, like denoising or/and contrast enhancement are often necessary in order to obtain correct results. This paper presents two methods for contrast enhancement in the space frequency domain using the wavelet transform to enhance local contrast of different tissue regions in 3D CT data sets. The dyadic wavelet decomposition presented in this paper prepares the image data for efficient edge detection by active contours. The algorithms were mainly designed and tested on 3D-CT-data sets containing lung tumors. Some more tests were done for prostate delineation.

2 Methods and Materials

Modeling tumor volumes by active contours implies to define interactively an initial approximate contour in the region of the tumor, which will be iteratively adapted to its actual contour. Adaptation will be mostly carried out correctly in cases of compact lung tumors surrounded by lung tissue. In cases of tumors partly adherent to the thorax wall no distinct contours are visible. They are partly represented by large Hounsfield unit differences to the surrounding lung tissue, or by small differences to the thorax tissue. Therefore they are difficult to detect automatically. In certain situations the active contour in the vicinity of small gradients of the tumor surface is captured by strong edges e.g. by bones. If the small contrast of these weak contours of the tumor could be enhanced locally, an active contour detection will be more successful.

A good approach to local contrast enhancement is the use of wavelets that permit localization in the space frequency domain, thus enabling the manipulation of local features. In comparison to other traditional enhancement techniques, wavelet decomposition enables higher decomposition levels to discriminate real and false edges generated by noise.

2.1 General Description of the Dyadic Wavelet Transform

Wavelet analysis uses appropriate scaling and wavelet functions to decompose the images into an approximation component and a series of detail components. It is appropriate for this application to use the dyadic wavelet transform [2], [3], which results in undecimated components because of its translation invariance.

The 2D dyadic wavelet transform uses two wavelet functions $\psi^1(x,y)$ and $\psi^2(x,y)$ and a scaling function $\varphi(x,y)$. In practice the wavelet functions are described by a horizontal and a vertical filter having the same coefficients. It is shown in [2], [3] that for the purpose of edge detection the wavelet functions should be given by the derivative in x respectively y direction of a smoothing function $\theta(x,y)$.

For a discrete image $f(x,y)$ the decomposition by the dyadic wavelet transform [2] is done from scale of 1 up to a certain maximum scale 2^J, resulting in an approximation component, a set of horizontal and a set of vertical detail components

$$\left\{S_{2^J} f, \left(W^1_{2^j} f\right)_{1\le j\le J}, \left(W^2_{2^j} f\right)_{1\le j\le J}\right\}.$$

At each scale 2^j the horizontal and vertical detail component form the gradient vector of the convolution of the image with the smoothing function θ(x,y) are given by:

$$\begin{pmatrix} W^1_{2^j} f(x,y) \\ W^2_{2^j} f(x,y) \end{pmatrix} = \begin{pmatrix} f * \psi^1_{2^j}(x,y) \\ f * \psi^2_{2^j}(x,y) \end{pmatrix} 2^j \begin{pmatrix} \frac{\partial}{\partial x}\left(f * \theta_{2^j}\right)(x,y) \\ \frac{\partial}{\partial y}\left(f * \theta_{2^j}\right)(x,y) \end{pmatrix} = 2^j \vec{\nabla}\left(f * \theta_{2^j}\right)(x,y),$$

in which $\psi^k_{2^j}(x,y) = \frac{1}{2^{2j}} \psi^k\left(\frac{x}{2^j}, \frac{y}{2^j}\right),\ k = 1,2$.

The approximation component at scale 2^J is

$$S_{2^J} f(x,y) = f * \varphi_{2^J}(x,y).$$

For the reconstruction the same scaling function $\varphi(x,y)$ is used together with two reconstruction wavelets $\chi^1(x,y)$ and $\chi^2(x,y)$ [2].

For the detail components the modulus of the gradient vector is given by:

$$M_{2^j} f(x,y) = \sqrt{\left|W^1_{2^j} f(x,y)\right|^2 + \left|W^2_{2^j} f(x,y)\right|^2},$$

while the angle between the horizontal and the vertical component is given by:

$$A_{2^j} f(x,y) = \arctan\left(\frac{W^2_{2^j} f(x,y)}{W^1_{2^j} f(x,y)}\right).$$

Edge points represent sharp gray scale variations, which in the frequency domain are represented by high frequency components. That means an edge point at position (x,y) in the space domain corresponds to a modulus maximum of the detail components at that position in the direction given by the angle $A_{2^j} f(x,y)$. Therefore the edges can be described using the modulus maxima of the detail components at different scales. For larger scales, i.e. lower resolution, false edges generated by noise disappear.

In practice the dyadic wavelet transform of an image is done by convolution with 1D filter masks. Details and relationships between the coefficients of the filter masks can be found in [2], [3]. The wavelets used in this paper are those as proposed in [2]. Other classes of dyadic wavelets can be found in [3].

2.2 Local Contrast Enhancement using Gaussian Operator

To achieve contrast enhancement an appropriate operator can be applied on the detail components or on the modulus of the detail components at the different scales (2^j)

with 1≤j≤J for which a modified image by the inverse dyadic wavelet transform can be obtained.

Some examples of linear operators can be found in [4], [6]. A nonlinear operator designed for mammographic images which enables contrast enhancement of low contrast features is presented in [5]. In [7] a nonlinear operator designed to enhance local contrast regions can be found. Most of the existing operators have the drawback that contrast enhancement is not limited to a given contrast range.

This paper aims to construct an operator which enhances high-frequency edges in a certain contrast range, while leaving the low frequency structures unchanged. This operator $\Omega_G(m)$ is based on the Gaussian function and is given by:

$$\Omega_G(m) = m + a\, m\, e^{-\frac{(m-b)^2}{\sigma^2}}.$$

The contrast range and values of the operator is determined by the parameters b and σ and the grade of enhancement by the gain parameter a. The operator is applied on the modulus maxima of the modulus image of the detail components. The modified modulus image is projected onto the horizontal and vertical components and then the inverse dyadic wavelet transform is performed. In the resulting image, contrast enhancement is obtained at edges in a certain contrast range, while outside this range of interest the alteration is insignificant.

A graphical representation of the dependency of the operator characteristic curve upon m, the modulus maxima, is given in figure 1.

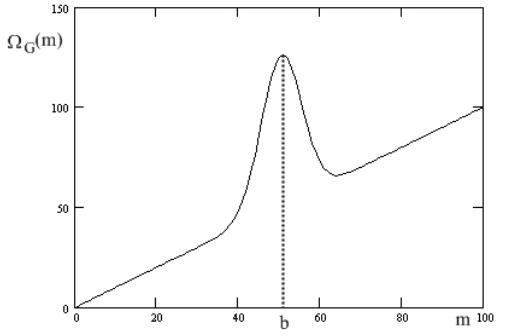


Fig. 1. Contrast enhancement operator $\Omega_G(m)$ with $a = 1.5$, $b = 50$ and $\sigma = 5$.

The only edge information used by this operator is given by the position of high frequency components. The enhancement is done for modulus maxima in a certain range, determined by the parameters of the operator. Edge points of the same edge, which are not contained in this contrast range, remain unchanged. To overcome this problem a second algorithm is proposed.

2.3 Local Contrast Enhancement using Edge Following

The idea of this algorithm is to start from each edge point $p(x,y)$ in a certain contrast range and enhance all the edge points which are on the same contour line of p .The algorithm is based on the idea of the Canny edge detector, using two thresholds T_1 and T_2 which define the contrast range. All the points, for which $M_{2^j} f(x,y) \in [T_1, T_2]$ is a modulus maximum, are enhanced to a given value C. Further, for each such a point, a chain of modulus maxima is constructed as follows. Considering that $MaxM_{2^j}(x,y)$ is defined as

$$MaxM_{2^j}(x,y) = \begin{cases} 0, & M_{2^j}(x,y) \text{ not a modulus maximum} \\ M_{2^j}(x,y), & \text{else} \end{cases}$$

For each element (x_0,y_0) with $T_1 \leq MaxM_{2^j}(x_0,y_0) \leq T_2$ the set of all points $\{(x_i,y_i)\}_i$ which can be connected by non zero modulus maxima with (x_0,y_0) is constructed. For all the elements of this set the modulus value is modified by:

$$M_{2^j} f(x_i,y_i) = \begin{cases} \max(C, M_{2^j} f(x_i,y_i)), & \text{if } M_{2^j} f(x_i,y_i) \geq T_1 \\ \Omega_G(M_{2^j} f(x_i,y_i)), & \text{else} \end{cases}$$

3 Measurements

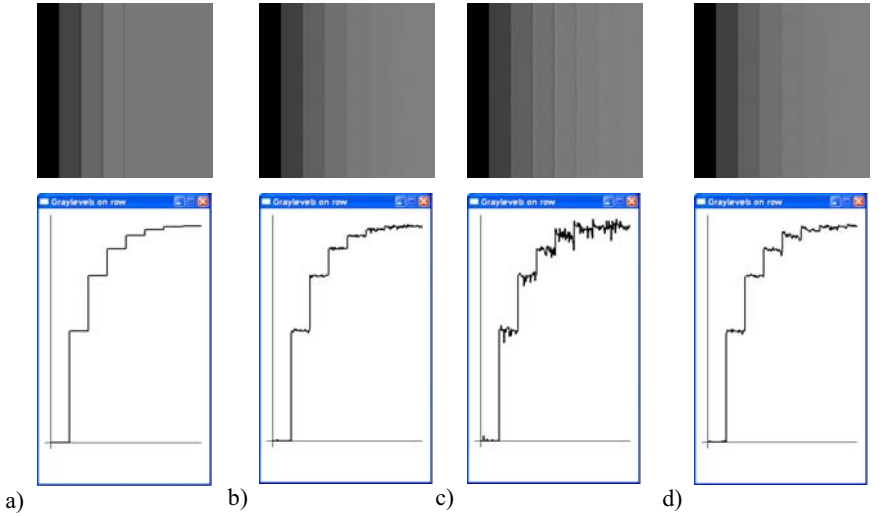
To study the results of enhancement using the operator $\Omega_G(m)$ some synthetic generated test images were used. The following questions should be answered:

- How contrast enhancement depends on the number of decomposition levels.
- How image noise is affected by contrast enhancement.
- Which parameters of the operator influence the results of contrast enhancement.

The measurements on the test images showed that using more than 3 levels of decomposition does not modify significantly the results. Therefore only the first 3 decomposition levels are taken into account.

As noise dominates the first decomposition level, enhancement should be performed only on the levels 2 and 3. To reduce this effect, enhancement should be performed only in the following levels. The result of the enhancement operator applied only on levels 2 and 3 is presented in figure 2 d) and is less affected by noise than the result in figure 2 c).

By changing the values of the parameter σ the width of the enhanced contrast range is modified. The parameter b controls the position where the maximum enhancement is performed, which is dependent on the features of interest.



a) b) c) d)

Fig. 2. a) Original test image; b) Image altered with Gaussian noise; c) Image after contrast enhancement using decomposition levels 1-3; d) Image after contrast enhancement using decomposition levels 2-3.

4 Results

The algorithms described before were tested on a series of medical CT-image data. In figure 3 are presented three different types of CT-image data. The results of contrast enhancement on each one using the first and the second algorithm together with the result of the active contouring before and after the enhancement are represented. The first column of figure 3 presents the original image data, the second column presents the result of contrast enhancement with the first algorithm and in the third column presents the result of contrast enhancement with the second algorithm. The first CT-data set represents a lung tumor adherent to the thorax wall. In this case the difference between the tumor tissue and the tissue of the thorax wall is not significant, so that contours are hard to detect. The second image represents a tumor embedded in lung tissue, where the contrast is varying along the surface. The third CT-data represents an axial cut through a prostate.

It can be seen in figure 3, that the contours obtained by active contours after contrast enhancement were closer to the boundaries of the specific structures, respectively the tumors and the prostate. In the case of the tumor image data, the white contour was manually defined by an expert radiologist.

In the case of the prostate, the dark grey contour obtained after enhancement with the second algorithm is nearer to the specific structure than the light grey contour obtained after enhancement with the first algorithm. For the lung tumors the two algorithms give similar results.

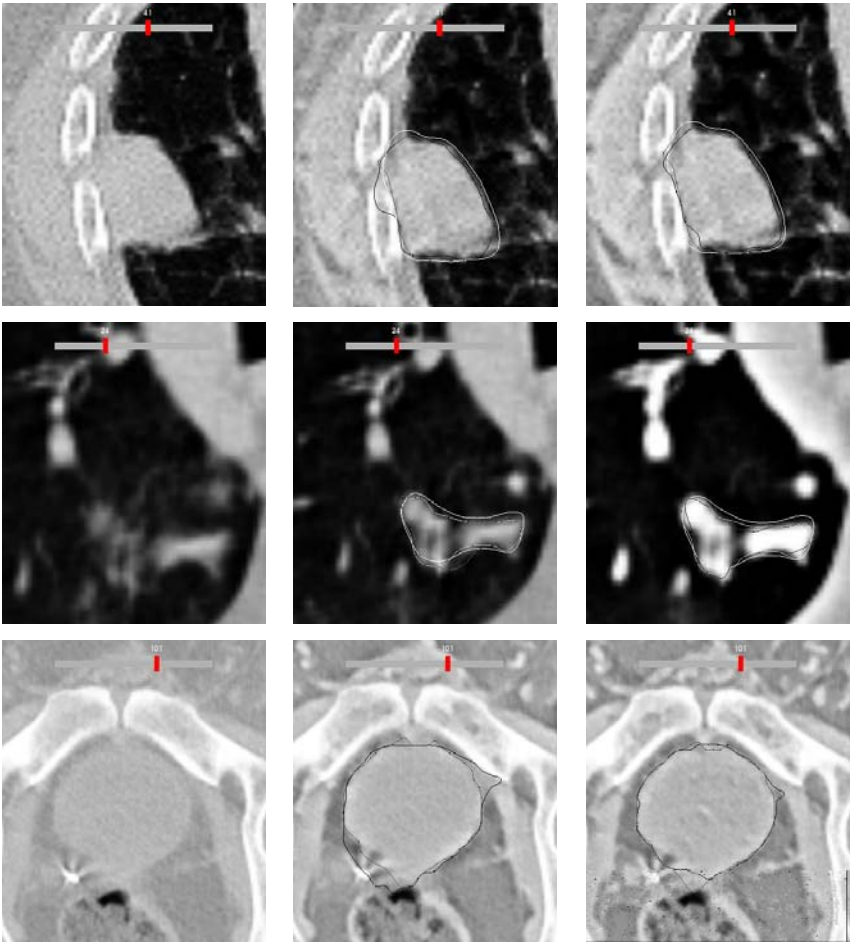


Fig. 3. The left column represents the original image data. In the middle column the white contour was manually delineated by an expert radiologist, the dark gray contour was obtained without enhancement, the light gray contour was obtained after contrast enhancement with the first algorithm. In the right column the light gray contour is the same as in the middle one, while the dark grey contour was obtained after enhancement with the second algorithm.

5 Conclusions and Discussions

Both algorithms were tested on different slices of around eight different 3D-CT data sets. The test results demonstrated that the placement of the active contour was more accurate after image enhancement in some cases and similar to the unenhanced case in others. For mostly circular structures with little variance in contrast along the border of the structure, the placement of the active contour was similar before and after enhancement.

Noticeable improvements were made in the enhanced case, where the structures to detect had low contrast boundaries in the vicinity of other high contrast structures, such as the prostate in figure 3. Likewise, improvements were made after enhancing structures having variable contrast levels along their boundaries. However, for noncircular, irregular ramified structures, no satisfactory contour placement was achieved either in the unenhanced or enhanced images.

It is necessary to make more tests to find the best parameters for the enhancement operator and to obtain standard values. In the second algorithm the selection of the thresholds and the treatment of the modulus maxima which are outside these thresholds could also be realized in different manners, which can produce varying results. More tests are necessary to establish the optimal variants. In future work an active contour algorithm based on wavelets could be designed, integrating the contrast enhancement step.

Acknowledgments.
ICP was the recipient of a DAAD scholarship what she gratefully acknowledges.

References

1. T. K. Nguyen-Pham, Semiautomatische Konturierung von Lungentumoren mit Hilfe aktiver Konturen in 3D-CT Aufnahmen für die Hochpräzisionsbestrahlung, Diploma-Thesis, University of Applied Sciences of Wiesbaden, Germany 2009.
2. S.Mallat, S Zhong, Characterization of Signals from Multiscale Edges, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, No. 7, pp. 710–732 (1992).
3. B. Chiu, A New Segmentation Algorithm for Prostate Boundary Detection in 2D Ultrasound Images, Master-Thesis, Waterloo, Ontario, Canada (2003).
4. A. Laine, J. Fan and W. Jang, Wavelets for Contrast Enhancement of Digital Mammography, IEEE Engineering in Medicine and Biology, pp.536–550 (1995).
5. P. Sakellaropoulos, L. Costaridou and G. Panayiotakis, A Wavelet-Based Spatially Adaptive Method for Mammographic Contrast Enhancement, Physics in Medicine and Biology, vol. 48, pp. 787-803 (2003).
6. A. Mencattini, M. Salmeri, R. Lojacono and F. Caselli, Mammographic Image Enhancement and Denoising for Microcalcification Detection using Dyadic Wavelet Processing, in Proc. IMTC-Instrumentation and Measurement Technology Conference, Sorrento, Italy, April 24-27, 2006, pp. 49-53.
7. X. Zong, A.Laine, "De-noising and Contrast Enhancement via Wavelet Shrinkage and Nonlinear Adaptive Gain", in Wavelet Applications III, Proc. of SPIE, Orlando, Florida, USA, April 8-12, 1996, vol. 2762, pp. 566-574.

GPU Collision Detection

Goşa Mihai[1], Poteraş Cosmin[1], Tănasie Razvan[1] and Mocanu Mihai[1]

[1] University of Craiova, Faculty of Automatics, Computers and Electronics, Software Engineering Department, Bvd. Decebal 107, 200440, Craiova, Dolj, Romania.
gosamihai@yahoo.com, {poteras_cosmin, tanasie_razvan, mocanu_mihai}@software.ucv.ro

Abstract. We describe a novel approach for fast collision / interference detection between multiple polygonal based objects in any type of environment using graphics hardware. Our algorithm uses a GPU based algorithm for finding the colliding objects and makes use of the rather recent NVidia CUDA platform [1] in order to accelerate the results transfer from the GPU back to the CPU. It is a screenspace method and therefore it involves no pre-computation, with the only restriction being that all objects must be closed polygonal based models. It is particularly suitable for objects with very large numbers of triangles, for which we can compute the exact points of collision using the parallel nature of the GPU with many orders of magnitude faster than what has been possible until now on the CPU.

Keywords: Collision detection, GPU, CUDA.

1 Introduction and Related Work

Collision detection is a fundamental task in applications like games, robotics, haptic rendering or CAD/CAM. It is the subject of a lot of ongoing research and it still remains one of the most time consuming tasks in an application's main loop. The concept of doing GPU based collision detection is a relative recent one, increasingly coming into the spotlight as graphics processors become more and more powerful and programmable. Beside last generation computer games, very few other applications take advantage of the full power of the GPU, leaving its great potential to rest. However, throughout the last years, GPU manufacturers began to expose more hardware functions to the user, increasing their programmability to the point where Intel announced, in 2007, the first true general purpose graphics processor, Larabee, which is capable of running x86 instructions while retaining the current GPUs strengths, like the highly parallelized architecture and vector processors. But even now, before Larabee's market debut, the GPUs are programmable enough to be able to resolve some of the most time consuming tasks, such as collision detection.

Over the last years there is also a growing demand for interactive collision detection between objects that can deform. Typical applications include surgery simulators, cloth simulation, virtual sculpting, and free-form deformations. Interactive collision detection between deforming objects is complicated because it requires

frequent updates of the data structures commonly used to accelerate the detection process. Even more important, geometric changes are increasingly performed on programmable graphics hardware using vertex programs, access to displacement textures and geometry shaders. In this case, the changes in geometry might not even be known to the application program, which makes it difficult to maintain a data structure that appropriately represents the modified geometry. The implications thereof with respect to collision detection are dramatic: As large parts of the geometry will permanently be modified and created on the GPU, CPU algorithms relying on the explicit knowledge of the object geometry can no longer be used.

As stated before, typical CPU collision detection algorithms usually imply a lot of pre-calculations, extra-memory and per application scene management and tuning. Our method proposes an image-space algorithm, which detects collisions/interferences as they are drawn to the screen, requiring no per application management and pixel perfect results accuracy. The GPUs parallel architecture is a perfect match for detecting collisions/interferences between multiple bodies at once. Using the GPU programmability exposed by APIs such as DirectX [2] or OpenGL with per fragment/primitive user written programs and NVidia's proprietary Compute Unified Device Architecture (CUDA) system, we will describe a novel collision detection algorithm, which comes with great performance gain over the CPU implementations and previous work by using the aforementioned technologies. We perform a virtual ray-tracing algorithm (by using the stencil and depth buffers of the graphics card) over all the scene's objects to find the intersecting ones. The result is outputted to the screen color coded in each pixel, and gets read by the CPU in order to perform collision responses. We try to overcome the limitations and bottlenecks of previous work by using the CUDA technology for accelerating the results download from the GPU with operations only possible on this new platform. Scatter operations and reading/writing at the same time from/to a texture will make a huge improvement in the download step and reduce memory requirements and execution time.

The exposure of the processing power of GPUs (now called GPGPUs) to the desktop user is a relatively new, in development, technology. General applications and algorithms are just starting to take advantage of this power and the research in this field is just beginning. Even though GPU collision detection algorithms have been proposed before, none of them has yet taken advantage of the new hardware capabilities exposed in CUDA or the new DirectX 10.

2 Related Work

Collision detection is a topic that has been extensively covered in many books and articles and still is the subject of ongoing research. Standard CPU implementations can vary depending on the object complexity, scene organization and what information is needed for the collision response. Having very complex objects can quickly become a bottleneck for CPU implementations and the majority of algorithms usually use approximations for the objects' shapes.

GPU implementations are rather recent [3] since they only became possible when the manufacturers started offering increased programmability to the users.

Implementing collision detection on GPUs came as no surprise given the highly parallel nature of most collision detection algorithm. For many applications, collision detection is applied only for the objects that can be viewed on-screen by the user, as an optimization process. Having the objects already on-screen comes as a plus for image-space collision detection algorithms, since they are applied directly to the scene being drawn. Also, collision response can be more accurate than CPU implementations since we no longer need approximated models for collisions and every contact point is evaluated with pixel perfect accuracy. The algorithm our method is based on [4] came as a big performance increase over a standard CPU implementation and made use of standard graphic processor operations as well as the programmable pixel pipeline. However, some drawbacks exist for all the GPU implementations up to this point, drawbacks that can't be overcome with the functionality exposed by APIs like DirectX or OpenGL. Using the NVidia CUDA model, we managed to solution those problems, gaining significant performance advantage over CPU and recent GPU implementations.

3 Methods and Algorithms

Image space algorithms work by running the calculations directly on the resulting image on-screen, after all the geometry transformations have been made. This makes it especially suitable for deforming geometry, and with the recent capabilities of graphics processors, deforming or adding detail on the GPU is a common task. This makes it very hard for the CPU to keep up and requires the redundant task of making an extra copy of the object (with all it's time consuming transformations) in the system memory. Image space collision detection algorithms overcome that limitation, but come with their own problems. Precision loss due to limited viewport size occurs when multiple collisions overlap at the same pixel. Another major problem is the required task of downloading the results back to the CPU for processing, which is a major bottleneck due to the limited bandwidth from the GPU to the CPU.

Our algorithm improves on a multi pass method for detecting collisions/interferences between closed polygonal objects with the addition of recent technologies in order to overcome previous limitations. The algorithm makes use of the depth and stencil buffers in order to perform ray tracing operations on closed objects and identify on-screen pixels over which two or more such objects intersect. We color code the pixel with the objects identifiers and also which part (polygon edge) of an object intersects the other. After the algorithm is run, the results will be on-screen, color coded pixels with the objects identifiers located in the exact places where collisions have occurred in image space. In the next step, screen pixels are downloaded from video memory and processed by the CPU in order to perform collision responses and other operations. Given two closed polygonal objects, we say that an intersection has occurred between the two only if one or more of the edges of an object intersect one or more of the other object's faces. This translates in a problem of detecting if any pixels from an object's edges are inside another object. For this we will use a hardware ray tracing method, using the graphic card's depth buffer and depth operations to trace the ray and the stencil buffer to account for ray hits. For a

ray originating outside of a closed polygonal object and directed towards the object, it will intersect its' faces an even number of times before it comes out through the other side of the object.

Using the stencil buffer to count each intersection and tracing rays through each object onscreen, we have a method of knowing when an object finds itself inside another by checking the stencil buffer values for odd numbers, meaning that a ray entered an object but it never came out, hitting something inside that object. For each odd value in the stencil buffer we draw the objects in the color buffer, using identifiers coded in the RGBA color channels.

The method is split in two main parts: detecting the penetrated objects and the penetrating objects. For the penetrated objects, we first render the entire scene using line primitives, filling the depth buffer with the objects' edges. Then, we draw again the whole scene, this time drawing only filled front facing polygons. We set the stencil function to increase the stored value if the depth test passes. To finish the face counting, we now need to draw the back facing polygons, decreasing the stencil buffer if the depth test passes. At the end of these operations, we will have a non zero value in the stencil buffer for where an edge occluded the ray, meaning we have a collision on that pixel. To identify the penetrated objects we simply render the objects again, with the color set as their identifier, but drawing only those pixels for which the stencil buffer value is positive.

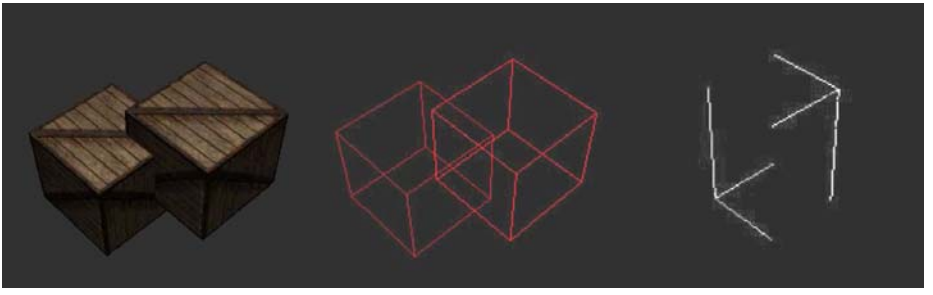


Fig. 1a. Two overlapping boxes.

Fig. 1b. Scene with the object identifiers drawn as colors (notice the different shades of red).

Fig. 1c. Stencil buffer content after the second pass.

Identifying the penetrating objects works in the same manner, only this time we need to draw each object at a time, using the same counting mechanism as described before. For each object, we will have positive stencil buffer values for where edges of other objects intersect it. We draw the object's coded color only where stencil values are positive and in another color channel than before, resulting in that pixel's value having identifiers for both the penetrating and the penetrated object in that point (Fig. 2a, 2b, 2c).

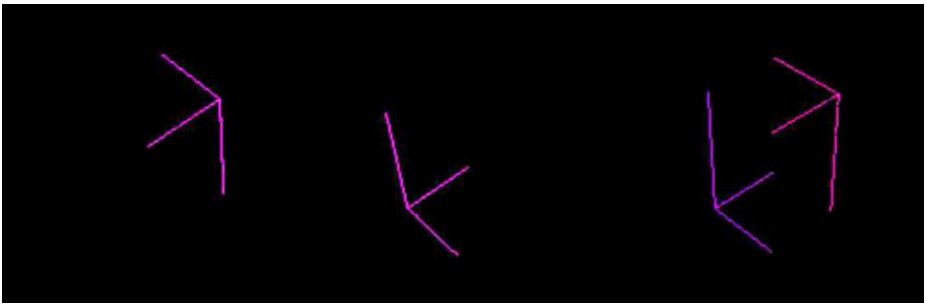


Fig. 2a. First penetrated object.

Fig. 2b. Scene with the object identifiers drawn as colors (notice the different shades of red).

Fig. 2c. Resulting scene with the color coded penetrating and penetrated objects.

The sumed up steps of the algorithm are as follows :

Detecting the penetrated objects:

- Draw scene only in the depth buffer, filling the Zbuffer with each pixel's depth value. This is the first and only time the depth buffer is modified, all subsequent render passes writing either in the stencil or the color buffer.
- Render the scene using the stencil buffer counting mechanism, increasing the value in the stencil buffer for those pixels where intersections occur. In this pass, no color or depth writes occur.
- Draw scene only for pixels with a stencil buffer other than zero, this time writing the penetrated objects' color in the "Red" channel of the color buffer.

Detecting the penetrating objects :

- Draw the scene using the stencil buffer counting mechanism, this time for each object at a time.
- Draw each object at a time only for those pixels where the stencil value is other than zero, using the objects' identifier as a value in the "Blue" color channel.

4 Implementation and Results

Encoding the objects' identifiers in the color buffer is done by separately using the "Red" and "Green" color channels for identifying the penetrated objects and the "Blue" and "Alpha" channels for identifying the penetrating objects. Having eight bits per color channel on a standard 32 bits per pixel RGBA color buffer leaves us with 16 bits for coding the object number. Having a maximum of 2^16 objects in a scene is more than enough, but if additional objects are required a second texture can always

be used. However, we also needed at least another 16 bits of data in order to store the penetrating edge color identifier, which would give us more information about where the collision took place and less CPU power for the final step of collision processing. For this to work, we make use of a render-to-texture technique, which uses a custom color buffer into which we can draw just as before. By using a 16 or 32 bits per pixel texture format (as opposed to standard 8 bits per pixel color buffer), we can now color code the penetrated object, penetrating objects and the penetrating edge identifiers into three channels, leaving us with one color channel left to spare. One of the problems that needs special attention in order to not to interfere with the correctitude of the results, is the lack of precision from the hardware's Z-buffer. Having only 24 bits available for the depth calculations (8 bits being used for the stencil buffer) can result in the same object, rendered two different times with different primitives – as it is required in the algorithm, be registered as laying at different depths, therefore written in the stencil buffer as colliding with itself. These are called self-interferences and can generate a lot of non existent collisions. We can overcome this problem with the cost of a post-processing step which can discard the selfinterferences. Self interferences will be registered in the color buffer as the same identifier in the penetrated object channel as in the penetrating objects' one. These malformed collision pixels can be removed with the cost of a post-processing step, drawing the screen results into another texture buffer and discarding each pixel for which the "Red" channel has the same value as the "Blue" channel. For this operation a custom pixel program was created, which reads from the results texture, compares the R and B values and writes to the new color buffer the old value or zero in case of an equality. As an optimization step, care needs to be taken that we don't use dynamic branching inside the pixel shader, even though recent graphics hardware is becoming better at branching, keeping the "if"s in a low count is imperative. That was done by using arithmetical operations and pixel discard instructions. However, this method requires that we make another copy the color buffer, meaning we're doubling the memory requirements. Here, for the graphics cards that support it, the CUDA model can be an important addition, having the possibility of reading and writing from/to a texture at the same time, removing interferences as the algorithm is running. This means we don't need to keep a copy of the screen for filtering out the interferences. The next step is reading back the results to the CPU. This proved to be the major bottleneck, having to copy the entire screen at once to system memory. Even then, the task of searching for collisions in each pixel of the screen is real time consuming. A convolution filter can help, reducing the size of the screen, but may result in a lot of collisions being lost if many intersections take place in a small screen area. The solution lays in re-arranging the valid pixels in a tight texture area, reading only the required amount of data from the GPU [5]. However, the packing of sparse pixels is not a trivial matter on current GPU's with the functionality exposed by the standard graphic APIs like DirectX or OpenGL. Scatter operations (writing at any giving pixel position) are not possible, as we can only write in the current pixel being drawn. Algorithms exist [6], [7] but they require too many render passes to be efficient. For this problem, the simplest and most efficient solution comes with using the CUDA platform. CUDA allows accessing the texture data as an array, therefore writing in any pixel position of a texture, indexing its data as we would with any data array. To compress the collision data, we create a texture with a memory count of double the

number of objects in the scene, acting as a two dimensional array into which we can write the collision pairs. Still, it's not an optimal solution, many of the pixels remaining empty since we can't theoretically have all objects colliding with every other object in the scene. With the bandwidth gain being more than enough, we chose to leverage performance versus memory requirements and skip further improvements which would certainly require more GPU power.

For testing and results, a custom test scene was used for the algorithm's validation (Fig. 3, 4). With the test scene containing over 20.000 triangles, the cost of transforming the objects in imagespace and performing virtual ray-tracing over the scene (with all redundant pipeline steps disabled) is almost non-existent when taking into account the processing power of a last generation graphics card. The outputted result locations of the collision points are pixel perfect, with the exact number of collision detected depending on the current resolution. The resulting performance gain over a CPU implementation is of several orders of magnitude, with much room left for improvement (Fig. 5).

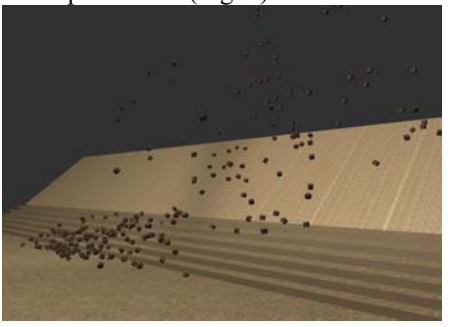


Fig. 3. Test scene.

Fig. 4. Resulting image with color coded collisions after running the algorithm.

The platform used for testing was a PC with an Intel Quad Core 2.66 GHz CPU and a NVidia GeForce 9800 GPU.

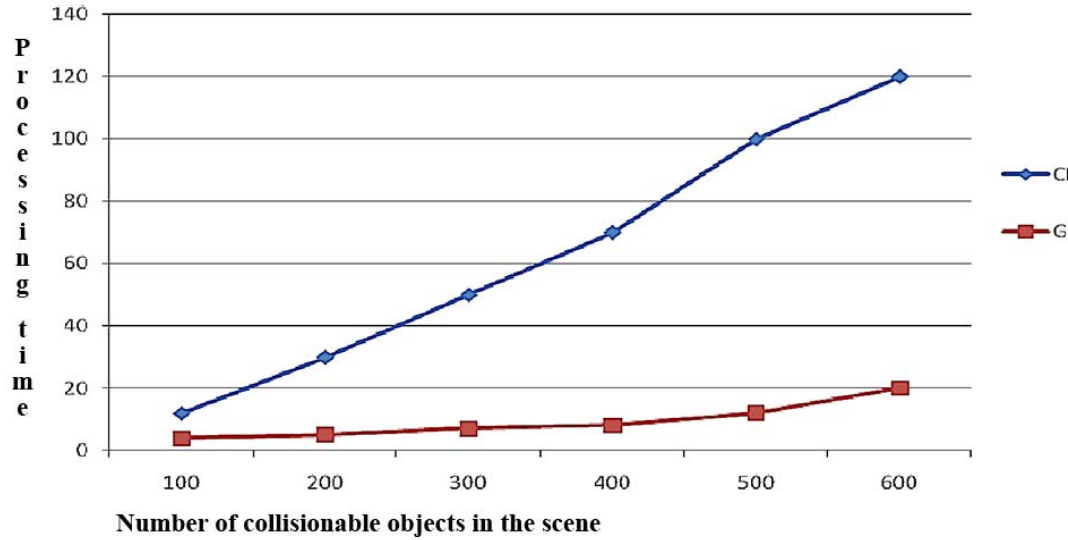


Fig 5. Processing time vs. number of objects for the CPU and the GPU implementations.

5 Conclusions and Future Work

There are certain drawbacks to image-space collision detection algorithms that we haven't been addressed and remain only partially solved, like screen resolution dependant accuracy. Since a pixel can only contain a single collision pair, collisions that are located at different depths but are overlapping in image-space are lost, with only the closest one to the viewer being registered. Larger resolution means increased accuracy but is also very bandwidth/memory demanding. We managed to overcome some of the previous limitations for GPU image-space collision detection algorithm, like high memory requirements due to multiple screen buffers needed for filtering the self-interferences and managed to efficiently lower the very high bandwidth requirement for reading back the results from the GPU with no precision loss, all with the addition of new GPU programmability offered by the CUDA model. Even as now, CUDA allows us to make a complete collision detection and collision response step on the GPU, having no need to download the results back to the CPU for processing. All the information needed for collision response (like polygon normals, object acceleration and other physic properties) can be stored in texture data and accessed directly from the video card memory, with the collision response coded as a CUDA or pixel shader program.

References

1. NVidia : CUDA, http://www.nvidia.com/object/cuda_home.html
2. MSDN : Direct3D Reference, http://msdn.microsoft.com/enus/library/bb172964(VS.85).aspx
3. Baciu, Wong, et al. : Recode: An image-based collision detection algorithm (1999)
4. M. Malmsten and S. Klasen : Practical collision detection on the GPU – A case study using CInDeR
5. A. Greß, G. Zachmann : Object space interference detection on programmable graphics hardware (2004)
6. J. Georgii, J.Kruger and R. Westermann : Interactive GPU-based collision detection, 2007
7. D. Knott : Collision and interference detection in real time using graphics hardware (2003)

Gabor Analytic Iris Texture Binary Encoder

Nicolaie Popescu-Bodorin

Department of Mathematics and Computer Science, Spiru Haret University
Ion Ghica 13 , Bucharest 3, Romania
bodorin@ieee.org

Abstract. The present paper proposes a new method for generating binary iris codes using *Hilbert Transform*. The *strong analytic signal* associated with the chromatic iris sequence is used to recover the phase information which it contains. A meaningful synthetic example is given in order to illustrate the iris binary code extraction. The reasons for canceling the discovery of the radial iris features and for choosing to encode iris features only in the angular direction are also explained here. *Gabor Analytic Iris Texture Binary Encoder* is introduced to prove that accurate recognition of similar iris images can be achieved comparing the binary iris codes that the encoder will generate for the similar circular iris rings (segments) previously extracted through *Circular Fuzzy Iris Segmentation* procedure. A new approach to iris recognition based on *Circular Fuzzy iris Segmentation* and *Gabor Analytic Iris Texture Binary Encoder* is proposed and tested here. Experimentally results obtained using the *Bath University Iris Database* are also presented.

Key words: iris segmentation, iris recognition, Hilbert Transform, strong analytic signal, Circular Fuzzy Iris Segmentation, Gabor Analytic Iris Texture Binary Encoder, Bath University Iris Database

1 Introduction

The strong analytic signal was introduced by Gabor [1] for extracting phase information content from a finite, discrete signal given in time domain i.e. for recovering both the carrier wave and the message modulated on it from the given signal. Since the considered signals are discrete, the name "analytic" is not a direct reference to a connection with the notion of analytic functions but is yet justified through the fact that each of these signals could be viewed as being obtained by sampling some continuous, finite, periodic (hence analytic) signal.

The form of a strong analytic signal is:

$$y = x + j * H(x)\,, \tag{1}$$

where $H(x)$ is the Hilbert Transform of the finite, discrete signal x given in time domain, and j is the complex unit.

The *analytic image* [2] is the 2-dimensional version of the strong analytic signal (heaving the same form as mentioned above where x is an image instead) and can be used for iris recognition as in [3]. Inspired and motivated by this work, the currently

proposed Gabor Analytic Iris Texture Binary Encoder is a simpler and more robust approach to iris binary code extraction based on the discovery of phase information available in the iris texture.

The main reason for working with the one-dimensional strong analytic signal instead of using the analytic image is that the critical information which decides the similarity or non-similarity between two iris rings is mainly stored as chromatic variation in the angular direction. On the other hand, accurate iris movement equations should be available in order to trace and to match chromatic variations along radial direction. Until knowing such motion laws, the chromatic variation along the radial direction will be, without any doubt, an important source of disagreement between those circular iris rings (segments) representing the same iris in different pupil dilatations. The essence of such a disagreement is that it is not "reconcilable" through an elastic deformation. As a practical example one can consider an iris image in which pupil dilatation is sufficiently strong to cause the iris area closest to the pupil to "disappear" or to change dramatically.

A question related to the use of analytic signals in iris recognition is whether the uniqueness of the iris is encoded in the carrier wave or in the carried message or in both of them? This paper confirms a partial answer to this question: the uniqueness of the iris is encoded sufficiently strong in the wave carrier component (i.e. in the phase information) of the analytic signal associated with the chromatic variation along the angular direction.

2 Hilbert Transform and the Strong Analytic Signals. Basic Properties

For a continuous time-domain signal $f(t)$, its Hilbert Transform is defined as follows:

$$\hat{f}(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{f(\tau)}{t-\tau}\, d\tau\,, \tag{2}$$

when the integral exists.

A strong analytic signal is the complex continuous time-domain signal $f(t)$ having the following property:

$$\hat{f}(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{f(\tau)}{t-\tau}\, d\tau = -jf(t)\,. \tag{3}$$

If the strong analytic signal $f(t)$ is separated into its real and imaginary parts:

$$f(t) = g(t) + jh(t)\,, \tag{4}$$

then:

$$\hat{f}(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{g(\tau)}{t-\tau}\, d\tau + j\frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{h(\tau)}{t-\tau}\, d\tau = h(t) - jg(t)\,, \tag{5}$$

and therefore:

$$\hat{g}(t) + j\hat{h}(t) = h(t) - jg(t)\,, \tag{6}$$

hence:

$$\hat{g}(t) = h(t)\,; \hat{h}(t) = -g(t)\,, \tag{7}$$

i.e.:

$$\widehat{Re(f(t))} = Im(f(t))\,. \tag{8}$$

As a consequence, any signal $y = x + j * H(x)$ is a strong analytic signal (Gabor analytic signal associated with x).

The basic properties of the Hilbert Transform are the following:

1. Hilbert Transform of a time-domain signal $f(t)$ is another time-domain signal $\hat{f}(t)$. 2. Hilbert Transform of a real-valued signal $f(t)$ is also a real-valued signal $\hat{f}(t)$.

3. Hilbert Transform of an even function is odd and conversely.

4. Hilbert Transform of a constant function is null.

5. If H denotes the Hilbert Transform, then the inverse Hilbert Transform is $-H$:

$$\widehat{\hat{f}}(t) = -f(t), i.e. : H^2 = -I\,. \tag{9}$$

6. Hilbert Transform commutes with temporal derivative:

$$\frac{d}{dt}\hat{f} = \widehat{\frac{d}{dt}f}\,. \tag{10}$$

7. Hilbert Transform is a convolution [4]:

$$\hat{f}(t) = \frac{1}{\pi t} \star f(t)\,, \tag{11}$$

$$F(H(f(\omega))) = -jsign(\omega)F(f(\omega))\,, \tag{12}$$

$$\hat{f}(t) = F^{-1}(-jsign(\omega)F(f(\omega)))\,, \tag{13}$$

where $\hat{f}$ and H denote Hilbert Transform and F is the Fourier Transform.

8. Hilbert Transform preserves the energy of the signal:

$$E_f = \int_{-\infty}^{\infty} |f(t)|^2\, dt = \int_{-\infty}^{\infty} |\hat{f}(t)|^2\, dt = E_{\hat{f}}\,. \tag{14}$$

9. Any signal with finite energy and its Hilbert Transform are orthogonal:

$$\langle f(t), \hat{f}(t)\rangle = \int_{-\infty}^{\infty} f(t)\hat{f}^{\star}(t)\, dt = \int_{-\infty}^{\infty} F(\omega)\hat{F}^{\star}(\omega)\, d\omega$$

$$= \int_{-\infty}^{\infty} F(\omega)[-jsign(\omega)F(\omega)]^{\star}\, d\omega = \int_{-\infty}^{\infty} j|F(\omega)|^2 sign(\omega)\, d\omega = 0$$

3 Recovering the Phase Information from Gabor Analytic Signal

In the iris recognition context, the most important property of the Hilbert Transform is that it preserves the signal energy. Hence f and $\hat{f}$ have the same energy but also

$$\frac{d}{dt}\hat{f}, \widehat{\frac{d}{dt}f}, \frac{d}{dt}f$$

share the same energy.

When f is assumed to be a line within the unwrapped iris, the meaning of the above fact is that the iris features in the angular direction is encoded with the same fidelity both in f and $\hat{f}$.

Now let us consider the real time-domain signal $f(t)$, its Hilbert Transform $\hat{f}(t)$ and the corresponding Gabor analytic signal:

$$z(t) = f(t) + j\hat{f}(t)\,, \tag{15}$$

expressed in polar form:

$$z(t) = A(t)e^{j\phi(t)}\,, \tag{16}$$

where:

$$A(t) = \sqrt{f^2(t) + \hat{f}^2(t)}\,, \tag{17}$$

is the instant amplitude and:

$$\phi(t) = arctan\left(\frac{\hat{f}(t)}{f(t)}\right)\,, \tag{18}$$

is the instant phase. Instant phase can be used further to recover from the real time-domain signal $f(t)$ its carrier wave and its carried message.

4 A Synthetic Example on Extracting the Iris Binary Code

Let us consider that the signal $f(t)$ contains a carrier wave and a carried message:

$$f(t) = a(t)cos(\phi(t))\,, \tag{19}$$

where $\phi(t)$ is determined in the previous section.

As a synthetic example, let us consider the following Matlab code:

```
td = [0:0.1:4*pi+0.2];
f = 2*cos(td)+cos(pi/4+2* td);
gf = hilbert(f);%(Gabor analytic signal associated with f)
hf = imag(gf); %(Hilbert Transform of f)
```

Let **C** be the carrier wave within **f**, **M** - the carried message and **S** - the sign of the phase of the Gabor analytic signal **gf**. All of these signals are presented in figure 1.

Remarks[1]**:**

[1] See the figure 1

- The information contained in the carried message **M** is relatively poor when compared to the carrier wave **C**.;
- The local extrema of the carried message **M** are encoded anyway in the bipolar iris code **S**;
- The bipolar iris code **S** is syncronized with the monotony intervals of the carrier wave **C**;
- The jumps in the bipolar iris code are syncronized with the inflexion points of both the real and imaginary parts of the Gabor analytic signal;
- The real part and the imaginary part of the Gabor analytic signal are in fact orthogonal: $\langle f, \hat{f} \rangle = -1.5765e-014$;
- The real part and the imaginary part of the Gabor analytic signal are strongly uncorrelated: $corr2(f, \hat{f}) = -1.5965e-016$;
- The bipolar synthetic iris binary code **S** encodes the separation between the most correlated and the most uncorrelated parts of f and $\hat{f}$:

$$corr2(f(S == -1), hf(S == -1)) = -0.6509$$

.

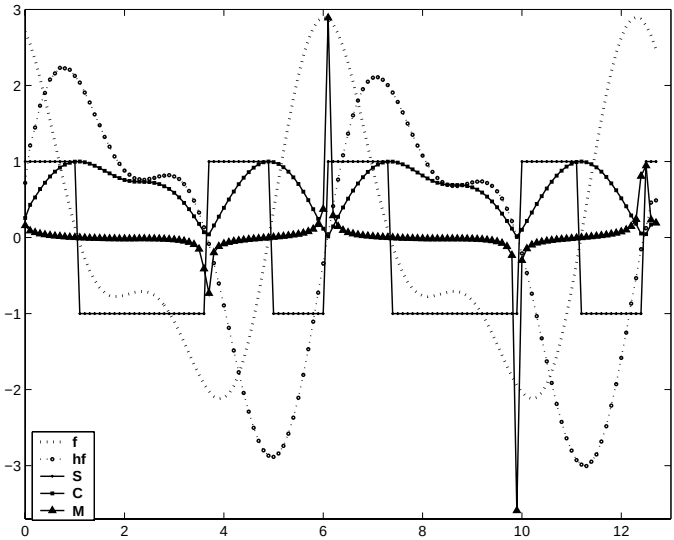


Fig. 1. f - original signal; hf - Hilbert Transform of f; S - bipolar iris code; C - carrier wave; M - carried message.

5 Encoding the Real Iris

The unwrapped iris is obtained by applying the Circular Fuzzy Iris Segmentation procedure [5]. Each line within the unwrapped iris is associated with a corresponding Gabor analytic signal. The bipolar iris code of that line is obtained as described in the previous

section and it is binary encoded becoming a line in the binary iris code on which the iris recognition is based.

As a practical example, the iris codes extracted from two images within the Bath University Iris Database are presented in figure 2 (similar iris codes obtained for similar iris images: 0001-L-0001.j2c, 0001-L-0003.j2c). The Hamming similarity measure for the two iris codes in figure 2 is 0.7346.

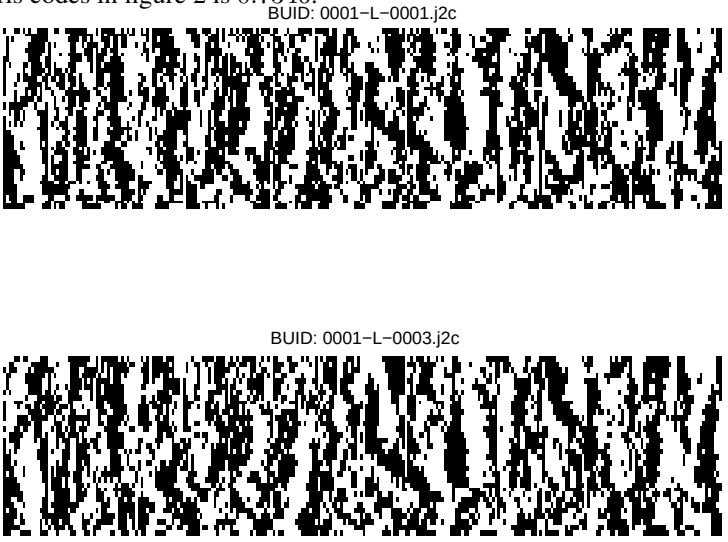


Fig. 2. Two similar iris codes obtained for two similar iris images.

Another example is given in the figure 3 in which three corresponding zones from two different images of the same iris are presented in some detail in order to illustrate the accuracy of both the segmentation and the recognition procedures.

It can be seen in the figure 3 that Circular Fuzzy Iris Segmentation is not the perfect segmentation procedure in terms of finding edges. This is why the circular iris ring is touching the sclera in the top-middle zones of each image. But most important, the segmentation procedure returns similar iris rings for similar eye images. Besides, even these zones are recognizable (as it can be seen in the corresponding zones of the binary iris codes). The Hamming similarity measure for the two iris codes shown in figure 3 is 0.7596.

6 Conclusions and Experimentally Results

The figures (4, 5) show all statistic data collected when running the proposed iris recognition procedure on Bath University Iris Database. The database contains 1000 images taken for 50 unique eyes.

Figure 4 presents the statistic data obtained when using the regular Hamming Distance. For the data in figure 5, a custom modified Hamming-Correlation distance was used. As a consequence the distance between the mean of the two classes of scores

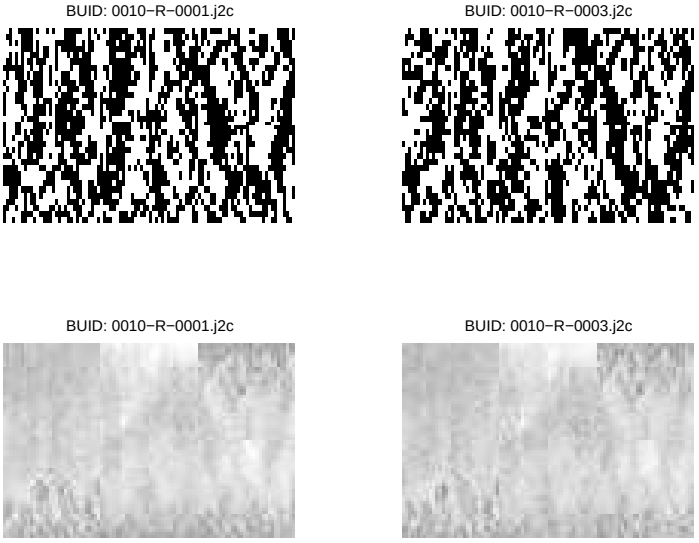


Fig. 3. Two iris codes in some detail.

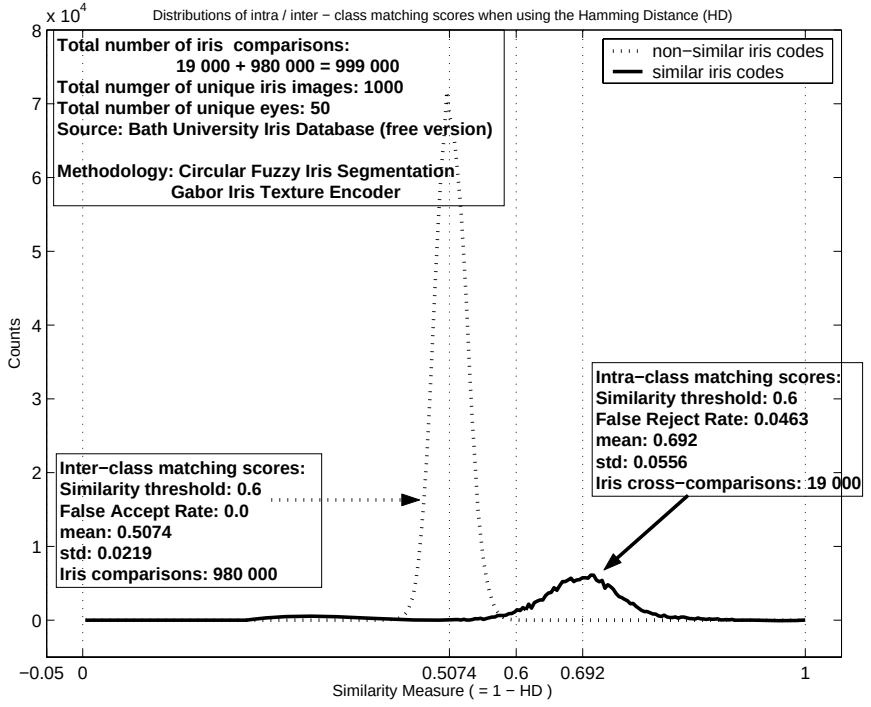


Fig. 4. Distributions of intra / inter - class matching scores when using Hamming Distance.

increases. But this better separation comes at a greater computational cost. Since the False Reject Rate remains nearly unchanged (the False Accept Rate is zero in both cases), modifying the similarity measure proves to be just an exercise.

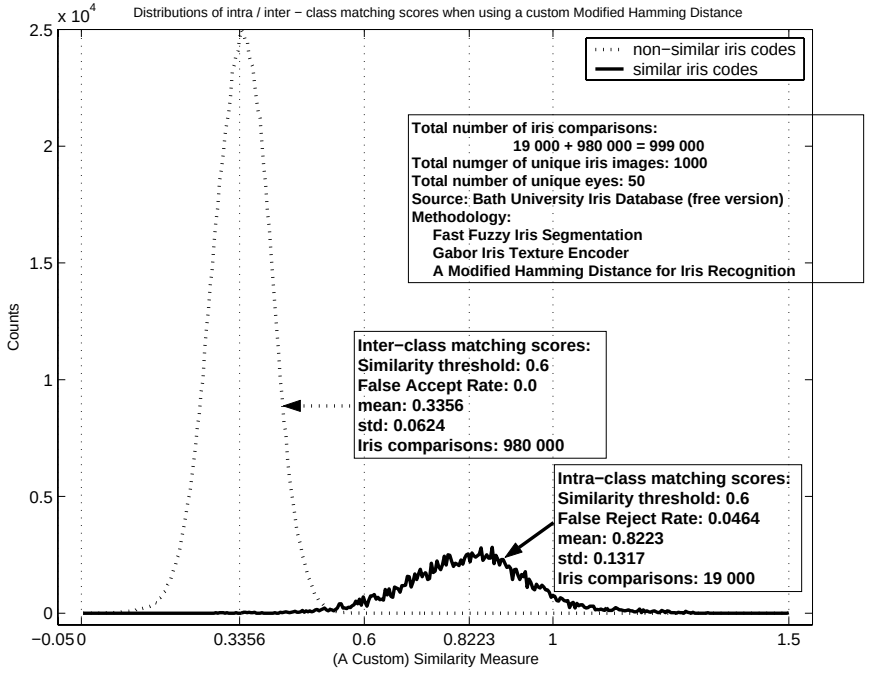


Fig. 5. Distributions of intra / inter - class matching scores when using a custom Modified Hamming Distance.

The most important thing regarding these results is the fact that they come to confirm the quality of the segmentation and recognition procedures: Circular Fuzzy Iris Segmentation and Gabor Analytic Texture Encoder. The False Reject Rate (5%) proves that the accumulated effects of segmentation and recognition errors are very small indeed.

The quality of the segmentation was evaluated also by running two limbic boundary detectors complementary to each other. A segmentation result was considered credible if it was confirmed by both detectors with a maximal error of 8 pixels. Very few exceptions occurred. This proves that the false matching rate is caused mainly by noise, rotated iris, very dilated/contracted pupil. There are two different images of the same iris having the Hamming similarity measure exactly 1 (0008-R-0005.j2c, 0008-R-0001.j2c).

At last, but not the least, the results presented in the figures (4, 5) prove the efficiency of the newly iris recognition approach proposed in this paper. Further developments of the proposed approach are under way and the results will be reported in the near future.

Acknowledgments. The author wishes to thank Professor Luminita State (University of Pitesti, RO) for the comments, criticism, and constant moral and scientific support during the last two years. The author also wishes to thank Professor Donald Monro (University of Bath, UK) for granting the access to the Bath University Iris Database. Without their support this iris recognition study could not have been possible.

References

1. Gabor, D., London, J., Born, M., Wolf, E.: Principles of Optics, pp. 557562, 7th ed., Cambridge University Press, Cambridge, 1999.
2. Havlicek, J.P., Havlicek, J.W., Bovik, A.C.: The analytic image, ICIP (2), pp. 446-449, 1997.
3. Tisee, C., Martin, L., Torres, L. , Robert, M.: Person identification technique using human iris recognition, On-line Proceedings of The 15th International Conference on Vision Interface (May 27-29, 2002), Calgary, Canada,
`http://www.cipprs.org/vi2002/`
4. Kschischang, F. R.: The Hilbert Transform, Department of Electrical and Computer Engineering, University of Toronto,
`http://www.comm.toronto.edu/frank/papers/hilbert.pdf`
5. Popescu-Bodorin, N.: Circular Fuzzy Iris Segmentation, accepted in The 4rd Annual South East European Doctoral Student Conference (), July 2009,
`http://fmi.spiruharet.ro/bodorin/articles/cfis.pdf`, 2009.
6. Bowyer, K.W., Hollingsworth, K., Flynn, P.J.: Image Understanding for Iris Biometrics: A survey, Computer Vision and Image Understanding 110 (2), pp. 281-307, May 2008.
7. Daugman, J.G.: High Confidence Visual Recognition of Persons by a Test of Statistical Independence, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 15, No. 11, November 1993.
8. Daugman, J.G.: New Methods in Iris Recognition, IEEE Transactions on Systems, Man, and Cybernetics - part B: Cybernetics, vol. 37, no. 5, october 2007.
9. Ma, L., Tan, T., Wang, Y., , D.: Efficient Iris Recognition by Characterizing Key Local Variations, IEEE Trans. on Image Processing, vol. 13, no. 6, June 2004.

Agents as Tools for Solving Complex Control Problems

Alketa Hyso[1] and Betim Çiço[2]

[1] Department of Computer Science and Electrical Engineering, University of Vlora, Vlora, Albania
[2] Computer Engineering Department, Polytechnic University, Sheshi "Nene Tereza", Tirana, Albania

Abstract. Agents and multi-agent systems are becoming a new way to analyze, design and implement complex software systems, since the focus of an agent-based approach is on goals, tasks, communication and coordination. The aim of this paper is to present agent as tools that enhance design for solving complex control problems. The agent-based design provides tools for the structuring of complex control problems in terms of partial control problems and their interdependencies, and for integrating partial solutions into an overall solution. A case study will be given to demonstrate this design method. Our purpose is to design a power intelligent management system that is able to fulfill the user comfort and minimize the consumption of the fuel of the generator. It is based on the concepts of an agent and a multi-agent system

Keywords: control agent; multi-controller systems; control agency

1 Introduction

Transportation or infrastructure networks, like power distribution networks, traffic and transportation systems, water distribution networks employee intelligent control techniques between them multi-agent systems.

'When adopting an agent oriented view of the world, it soon becomes apparent that most problems require or involve multiple agents; to represent the decentralized nature of the problem, the multiple loci of control, the multiple perspectives or the competing interests' [1]. Algorithms based on Multi-Agent Systems are nowadays used in several areas such as Computer science or Automatic Control [3, 5, 14]. The developments of solutions based on Multi-Agent Systems, well suited to solve spatially distributed and opened problems, permit to imagine an intelligent multi-agent control system.

An interesting issue of solving practical control problems is that they are generally not solved by using one technique. So, in general, they are solved by using multiple, heterogeneous models and multiple heterogeneous design techniques, while taking into account multiple control objectives. Agents have been proposed as enhanced controllers with features useful for fulfilling the new flexibility, availability and changeability requirements [2, 8, 12]. Agents have been considered as goal-oriented, semi-autonomous controllers in a distributed control system. They are expected to

coordinate control operations both in normal and abnormal situations. In control functions the overall role of agents has usually been proposed to be decision making concerning actions in one controller and coordination of these decisions with other controllers.

This paper represents a qualitative analysis to provide why agent-based systems are well suited for solving complex control problems. It proceeds from the standpoint of using agents as tools for designing multi-controller systems. The paper is structured as follows. Section 2 discusses the essence of a controller agent. Section 3 makes the case as to why agent-oriented approach is well suited to complex control systems and highlights the potential advantages of adopting an agent-oriented approach. Section 4 presents architecture of control agent and section 5 - the application. Finally, the paper presents some conclusions and highlights the future work.

2 The Essence of a Controller Agent

A special role in the theory and tools for solving complex control problems is attributed to the concept of an agent [10, 4, 7]. An agent represents an abstract entity that is able to solve a particular (partial) problem. Conflict between agents, which naturally arise in such systems due to the dependencies between the partial problems the agents solve, are handled by properly coordinating the agents' activities. Agents can be combined into a multi-agent system, such that the overall multi-agent system is able to solve a more complex problem. Combining the concepts of a local controller and an agent has resulted in a so called controller-agent. 'A controller-agent is a local controller that is responsible for the initialization and finalization of its state variables, has knowledge about its operating regime and has an interface to coordinate its behavior with other controller-agents' [16].

Two different ways can be imagined to combine controllers and agents. The first way is to design a controller for the sense-think-act mapping of a particular agent. The controller becomes the architecture of the agent. Another way is to use agents for execution of control algorithms. A controller would consist of several agents, each becoming active and producing control signals under particular operating conditions of the controlled plant. The control structure that consists of one or more control agents can differ:

- The access that the control agents have to the sensors and actuators in the network.
- The communication that the control agents have among one other.
- The way in which the control agents process sensor data to obtain actions
- The authority relations between the control agents.
- The beliefs, desires, and intentions of the control agents. [9]

Intelligent control agents control systems whose state space cannot be completely enumerated through traditional discrete and /or continues means, necessitating the use of artificial intelligence reasoning techniques. Control agents within autonomous software agents; provide an affective mechanism for autonomous control of distributed sensor/actuator systems. Control agents are autonomous in that they do not require human supervision or monitoring in order to function. Control agents employ

a cyclic process to accomplish the desired behavior for the system being controlled [11].

3 The Advantages of Adopting an Agent-Oriented Approach

The most convincing argument that could be made for adopting an agent-oriented approach to multi-control systems over a range of other techniques is related with the superiority of the agents. The agent-oriented approach advocates decomposing problems in terms of autonomous agents that can engage in flexible, high-level interactions. By considering an agent as "an entity that solves a problem" it becomes useful for modeling distributed problems. Each partial problem is represented by a separate agent. Complex control problems are distributed in nature. Their solution is based on multiple modeling techniques and resources needed to design and implement a control system can be distributed in nature. There are several good reasons to use MAS: distributed problem, robustness, scalability, simpler implementation. MAS are an attractive concept to model particular types of distributed problems [13].

A significant part of any design process is finding the right models for viewing the problem. In the case of control complex systems the problem to be characterized consists of subsystems, subsystem components, interactions and organizational relationships. There is a clear and strong degree of correspondence between the notions of subsystems and agent organizations [6].

Software agents exhibit both reactive and deliberative behavior. They are reactive in their ability to reconfigure the systems within their control in the context of an existing plan and deliberative in their ability to create a plan in response to observed states and defined goals. These agents are environmentally aware, and behavior of the environment is predictable through each agent's model.

4 The Architecture of a Controller-Agent

When constructing a multi-controller an important organizational design issue is to determine the entity or functional unit that will be responsible for each of these functions - the local control algorithm to calculate the control signals, the local operating regime of the local control algorithm in order to decide when to (de)activate the local controller, and initialization and finalization functions to initialize and finalize state variables of the local control algorithm.

Agent theory, however, suggests a different organization, [15] i.e., to include all functions into an autonomous entity. The interface of a controller-agent is made up of its inputs and outputs, and its activation request and acknowledges signals. A controller-agent behaves either as being "active" or "inactive". Whether a controller-agent is active or inactive depends on its intentions, and on the intentions of the remaining controller-agents. These intentions are expressed by the activation request signal.

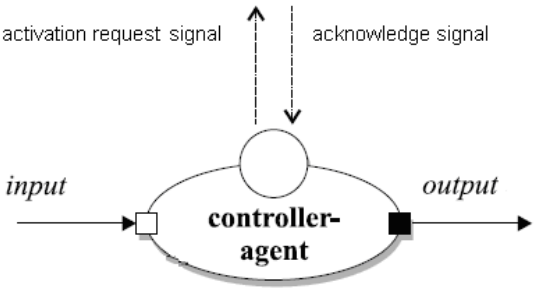


Fig.1 . The architecture of a controller – agent.

To coordinate several controller-agents, a mechanism - coordination object is needed that (in) activates them based on their intentions (i.e., activation request signal). It calculates acknowledge signals by using a so called decide function. So, a coordination object must take the decision and hence still must solve conflicts, deadlock, bumpless transfer and shattering.

5 The Application

We shall take into consideration an agent-based design for multi-controller systems, such that individual local controllers can be added, modified or removed from the overall multi-controller without redesigning the remaining system. We'll use agents for execution of control algorithms. A controller would consist of several agents, each becoming active and producing control signals under particular operating conditions of the controlled plant. We shall analyze the design of a power intelligent management system in a vessel.

This system allows having acclimatization, some light bulbs, and a hot water heater. A generator furnishes power for the vessel. All the equipments are independent and distributed. This system is open and the number of equipments in it is changeable. We can add or remove equipment in this system. The generator will supply with power to all these equipments. Our purpose is to design a power management system that is able to fulfill the user comfort and minimize the consumption of the oil of the generator.

For this reason, we follow the methodology: First of all, during an analysis of the overall goals and the process characteristics, we identify global control strategies which optimally run the process. In a second step, control strategies are decomposed into single control tasks which can be executed locally. These control tasks are then grouped and assigned to the agent. After that, we are faced with the task of coordinating their operations.

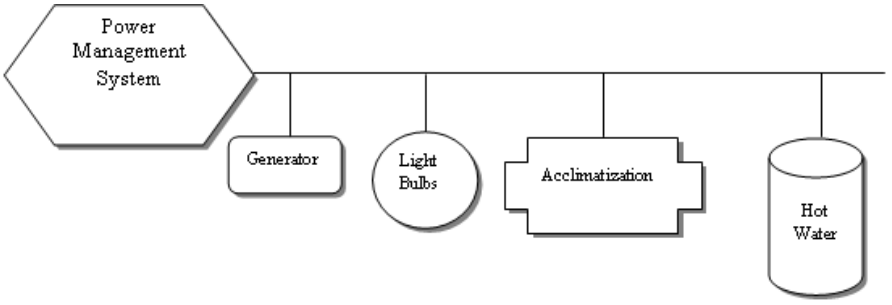


Fig. 2. Power Management System, generator, and consumption equipments

Firstly, the partial control problems should be defined. We need to control the generator consumption so as to control this equipment let's embed an agent (or control agency) in it. The generator is supplied with a flow of a fuel and can generate electrical energy. It has an electrical output terminal or node which can be connected directly to electrical user appliances and which also be connected in parallel with an external electrical energy generation and distributed network. The controller agent regulates the electrical power generated by the generator by controlling the rate of rotation and the injection of fuel in the internal combustion engine. The agent controls the operation of the generator in accordance with predetermined procedures in a mode in which it is disconnected from the distribution network or in a mode in which it is connected, and controls automatic changing from one to the other of the above mentioned operative modes without interrupting the supply of energy to the electrical use appliances.

Also we have to control the light bulbs, (their light intensity must be variable according the time of the day). We need an agent (control agency) which is able to control the light bulbs. This lighting control system has to minimize energy consumption and therefore minimize the cost of energy required. The light agent controls some fluorescent light fixtures. Each fluorescent light includes a communication receiver for receiving control signals from the light agent. Each fluorescent light fixture has the ability to adjust the power level according to control signals that come from the light agent. The agent controls the power delivered to the lamps. If a power level is below 25% or some other given level is requested, then the lamps are turned OFF. The light agent is connected with sensors such as occupancy sensors, light level sensors that can sense conditions within the control zone. These sensors "look at a window to see" if there is sufficient natural light available. The agent operates on the signals from the sensors and applies delay times; calibration values, and controls set points that are appropriate for the zone and for the sensors. The agent can monitor the operating conditions within the zone. So, this agent provides valuable information regarding maintenance scheduling by logging the effective operating hours for the lamps in the zone. The light agent can provide direct control of the power level of the fixture.

Also the acclimatization needs to adjust the room temperature according to the user requirements. It's preferable that the room temperature ranges between 22^{o}C-25^{o}C. If the ambient temperature is lower than 22^{o}C we need to heat the rooms, otherwise we

need to fresh them, when the room's temperature is 22°C-25°C we want to keep it stable. We need a controller agency to control the acclimatization equipments. This controller agent is able to choose through the heat generator or coolant compressor depending on the temperature of the rooms. Energy demand signals are calculated from the agent which defines the duration of and spacing between the closing of the thermostat switches. The agent uses this timing information to control switching signals in accordance with the duration of energy demand.

Another control problem is to control the temperature of the hot water. So a control agency is needed for this purpose. Suppose we have a tank less water heater for heating water passing there through. The controller agent communicates with the temperature sensors positioned to detect water temperature proximate the inlet and outlet portion. Agent also communicates with a flow meter positioned proximate the inlet portion which detect fluid volume. The agent receives the signals from the sensors and flow meter and decides for a proportional amount of electric current to the heating elements distributed on the tube.

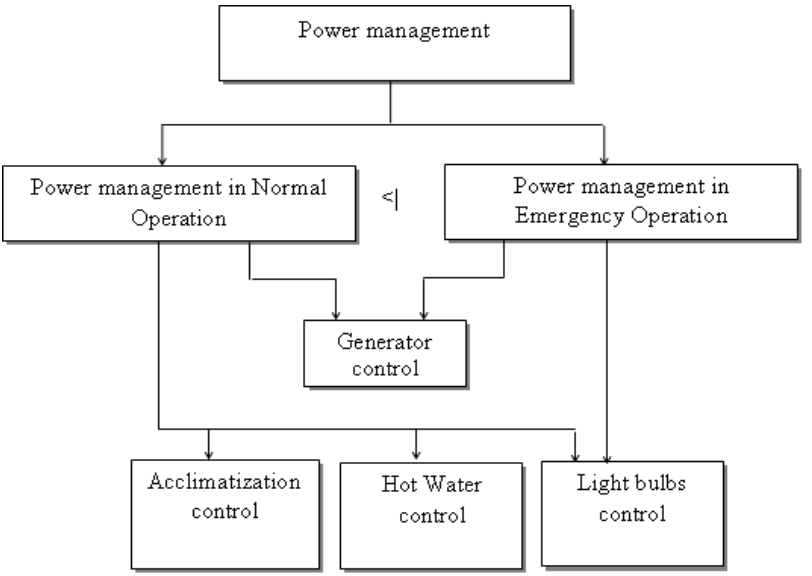


Fig. 3. Structure diagram of power management system control problem

The control algorithm is decomposed into two complementary mechanisms: an emergency mechanism and a normal-Operation mechanism. Emergency mechanism, which is a real time one, will be triggered when the level of fuel in the generator's deposit is lower than a reference level. During the emergency process only the request of the light bulbs control agency will be accepted. If the level of the fuel is higher than the reference level, the normal-Operation mechanism will be operated. During this phase, all the requests that come from all the control agencies will be taken in consideration. The priority of the emergency mechanism is higher than the normal-

Operation mechanism one. Also the priority level of the light bulbs control agency is higher than the priority of any other consumption equipment. Fig.3. shows this decomposition of the control algorithm into two complementary mechanisms

These embedded control agencies need to communicate and coordinate their operations. Every control agency has a coordinated mechanism, which analyze the request signal to be active or inactive that comes from the agents of the control agency. It takes into consideration the priority level of the agents of this agency, and then it transmits this request signal to the central coordination mechanism. This mechanism sends an acknowledge signal to the control agency with higher priority level.

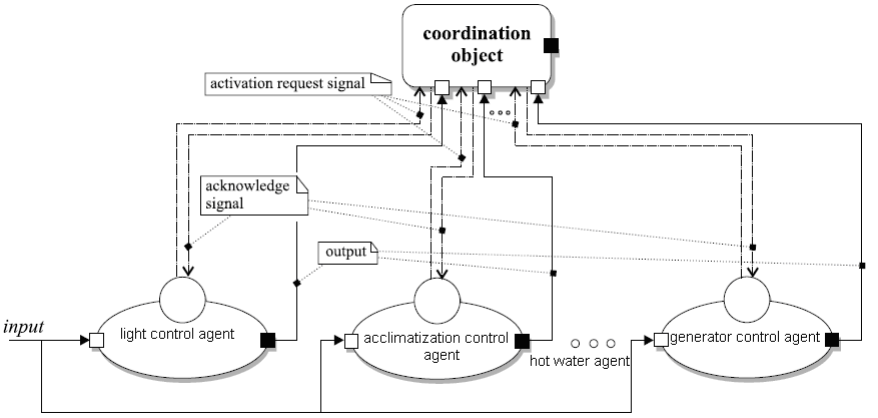


Fig. 4. The coordination object and the agents in the hierarchy

The main coordinator mechanism initializes the negotiation by asking the coordination mechanisms of the control agencies of the consumption equipments to send to it their power needs, with the purpose to reach a satisfactory function provided by this agent. It coordinates the operations of all the agents in the hierarchy, also makes decisions as accept or refuse the request signals that the other agents send. The coordination object acts like a supervisor and decides which controller-agent to (in) activate based on measured information and the controller-agent's intentions. Fig.4. shows the communication of the coordination object with all other agents.

All the information about the current status of the controlled equipments is coming back to their embedded agents. So the agents can consider this information in their future plan.

6 Conclusions and Future Work

This paper has sought to justify the claim that agent-based design has the potential to significantly improve our ability to model, design and build complex, distributed software systems. In making this claim, a series of qualitative arguments were

developed to highlight the high degree of match between the requirements of complex system development paradigms on the one hand and the key concepts and notions of agent-based design on the other.

The agent-based design method presented in this paper helps the designer to solve complex control problems, by offering concepts to structure the problem and to organize the solution. The design method encourages to develop local solutions and to reason about their dependencies. It offers the coordination mechanisms to deal with these dependencies.

This paper has outlined that using the agent-based design method, allows that individual controllers can be designed, implemented and tested separately.

Our paper aims to show a part of our work in theory level, which will be extended in the future in application level. We demonstrate the use of an agent-based design technique for multi-controller systems. It is our future intent to continue the work and to implement step by step one of the control agencies.

References

1. Bond,A.H., and Gasser L. (Eds.). (1998). Readings in Distributed Artificial Intelligence, Morgan Kaufmann, San Mateo, CA.
2. Chan, P. K., Fan, W., Prodromidis, A. L., and Stolfo, S. J. (1999). Data mining to predict aircraft component replacement. IEEE Intelligent Systems 14 (6), pp. 67-74.
3. Cherkassky, B. V., Goldberg, A. V., and Radzik, T. (1994). Shortest paths algorithms: theory and experimental evaluation. In SODA '94: Proceedings of the fifth annual ACM-SIAM symposium on Discrete algorithms, pp 516-525, Philadelphia, PA, USA. Society for Industrial and Applied Mathematics.
4. Ferber, J. (1999). Multi-agent Systems – An Introduction to Distributed Artificial Intelligence. Addison-Wesley.
5. Ha, D. L., Ploix, S., Zamai, E., and Jacomino, M. (2006). A home automation system to improve household energy control. In The 12th IFAC Symposium on Information Control Problems in Manufacturing.
6. Jennings, N.R.. (2000). Artificial Intelligence 117. p. 277–296, Elsevier Science B.V.
7. Jennings, N.R., and Wooldridge, M. (1998). Agent Technology. Applications of Intelligent Agents, Eds. N.R. Jennings and M. Wooldridge. Springer.
8. Marik, V., Fletcher, M. and Pechoucek, M. (2002). Holons & agents: recent developments and mutual impacts. In: V. Marik, O. Stepankova, H. Krautwurmova, M. Luck (eds.), Multi-Agent Systems and Applications II, Springer, Germany, , pp. 233-267.
9. Negenborn, R.R., (2007). Multi-Agent Model Predictive Control with Applications to Power Networks. ISBN 978-90-5584-093-9, Published and distributed by: R.R. Negenborn.
10. Russel, S., and Norvig, P. (1995). Artificial Intelligence - a modern approach. Prentice-Hall International, Inc., Englewood Cliffs, New Jersey.
11. Scheidt, D.H., and Hopkins, J., (2002) apl technical digest, volume 23, number 4, pp. 383-395.
12. Seilonen,I., Appelqvist, P., Vainio, M., Halme, A., and Koskinen, K. (2002). A concept of an agentaugmented process automation system. 17th IEEE International Symposium on Intelligent Control (ISIC'02), Vancouver, Canada.
13. Stone, P., and Velose, M,. (2000). Multiagent systems: a survey from a machine learning perspective. Autonomous Robotics, 8(3).

14. Tichy,P., Slechta,P., Maturana,F., and Balasubramanian S. (2002) Industrial MAS for planning and control. In: V. Marik, O. Stepankova, H. Krautwurmova, M. Luck (eds.), Multi-Agent Systems and Applications II, Springer, Germany, pp. 280-295.
15. Van Breemen, A., (2001). Agent-Based Multi-Controller Systems. ISBN 9036515955. Publisher: Twente University Press. pp. 52-104.
16. Van Breemen, A., and DeVries, T., (2000). Proceedings of the Fifth International Conference on The Practical Applications of Intelligent Agents and Multi-Agent Technology, Manchester, U.K., 10th-12th April 2000, p219-235, ISBN 1 902426 07 X, edited by J. Bradshaw and G. Arnold.

An Intelligent Agent Based Navigation System: Simulation and Application

Eva Cipi[1] and Betim Cico[2]

[1] Department of Computer Science and Electrical Engineering, Vlora University, Vlora, Albania
evacipi@univlora.edu.al
[2] Computer Engineering Department, Polytechnic University, Tirana, Albania
betim.cico@gmail.com

Abstract: This work aims to extend the agent based paradigm in industrial applications. A system and its control are termed intelligent if the system can self determine its decision choices based upon the simulation of needed solutions or upon experience stored in the form of rules in its knowledge base. The main idea is to provide an intelligent system for the navigation of an object along the best real time path in a dense environment taking complete control over the task of reaching a dynamic target without collisions. The method by which the system accomplishes its objectives is presented here. Developing a simulation on a real time map provided from a database agent and executing a motion plan on line, as result, it products an automated controller which makes intelligent decisions on the next step in a sea navigation system.

Keywords: agent based system, intelligent system, control agent, dynamic environment, visual sensor, motion planner, simulation off line.

1 Introduction

Many technical applications in the real world, which involve different software systems require more autonomous system behaviors to provide some functionalities regarding to the efficiency and safety. Finding new solutions and designing approaches that compete the human behavior, it is one of the main challenges today in the artificial intelligence area. Along the way we discuss the use of the agent based technique on the control systems illustrating it by an intelligent ship navigation system. We show a general system architecture using the style of decomposition in some intelligent subsystems with influences which may be discrete or continuous in nature. It is clear the manual control is simpler and less expensive, but the human operator would likely discover how difficult is to control a device in a not completely known environment. Control agents are autonomous in that they do not require human supervision or monitoring in order to function**.** Here we find the motivation to study and design possible agent oriented solutions as the best way for a better performance in real time. The work is divided in two areas:

First, we give an overview of an agent based system control, which are the main processes that complete a cycle of control and its appropriate architectures. In the second part simulation extends to the application of the sea navigation domain.

The model is based on simulating the navigation using a computer assisted control system in a virtual space. The system includes agent modules as passive and active entities making intelligent decisions to complete pre-selected tasks; control agent only is able to exhibit goal-directed behavior taking the initiative. The control method is a agent based software system with a main control agent which receives repetitively input data sending from the other information agents that manage different state of the system[1]. Interfacing and sharing information with these agents, the control agent accumulates required knowledge and incorporates unique feedback methods for planning a new path. Scenarios is a virtual sea map where an object named agent should find an other object (target) that moves randomly in an unknown direction.

2 About a Control Agent

The required level of intelligence for a system agent based depends on how fulfilled its knowledge is about its environment. Rational agents are a certain class of computer system capable of independent autonomous action in order to meet their design objectives [2]. We call them rational because they make good decisions about what to do.

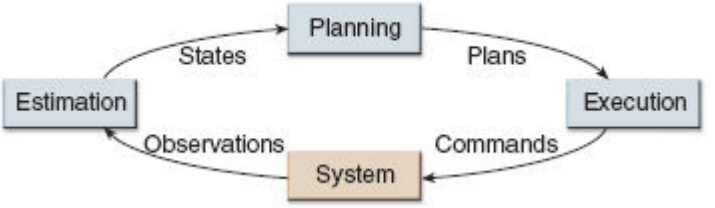


Fig.1. The control cycle.

The intelligent **control agents** built for the centralized or distributed control of the subsystems are based on a few key technologies that combine to provide autonomy, distributed intelligence, and real-time performance[3]. To complete its tasks, an control agent makes a control cycle (Fig. 1) which consists in three steps.

1. Estimation uses observations to compute a belief function over the state space of the system under control. A belief function describes the possible component states within the system that are consistent with current and historical observations. The state space of a system is the set of possible states for the system.
2. Planning uses this belief function, in conjunction with system goals and constraints, to produce a plan of action.
3. Execution translates an action plan into actuator commands

Hierarchical control architectures decompose the control along the lines of component abstraction, providing effective control of complex systems that have multiple hierarchical goals, multiple sensors, and a need for robustness[4]. Model-based reasoning is an effective reasoning technique for providing intelligent control on these components. For increasing the level of the intelligent decisions and actions, software agents are used to provide autonomy and support distribution. These architectures decompose control into two or more layers (Fig. 2).

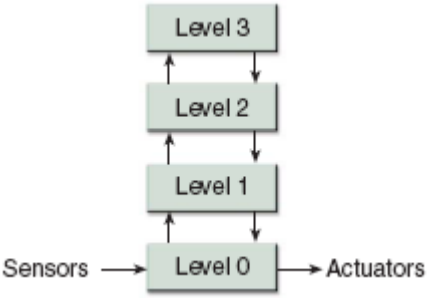


Fig. 2. A hierarchical control architecture.

To synthesize a completely autonomous computer-assisted control system operating in dynamic environment, we use concepts, ideas, and tools from artificial intelligence, simulating a software system agent based as a part of complete control system which is able to lead successfully a system in partially known environment.

3 An Intelligent Control System: Case Study

Knowledge about the surrounding environment determines the structure and methodologies used to control and coordinate the system, which leads to an increase of the intelligence of the individual system components[5]. Because of the rather high level of this knowledge, all the planning activities can be performed off-line, and only task execution needs to be done on-line. When little or no information about the environment is available, a different approach is needed. In this situation, a multi agent system that shows no clear grouping of components is better suited to develop plans and to react to changes in a dynamic environment[6]. All the calculations have to be done on-line. This requires more processing power and faster algorithms than the organized structure, where only the operations in the execution phase have to be computed in real time[7].

We are going to show a part of an intelligent navigation system for a vessel. However, each unit of design and synthesis for system control needs different mathematical and system engineering tools such as graph searching, optimization, neural computing, fuzzy decision making, simulation of discrete dynamic systems, and event-based system methods. These control structures are known in technical applications [8].

Our system should synthesize and execute two phases:

1. Planning and simulation off-line
2. On-line simulation based monitoring and intelligent control

In the first phase a simulation model of a static environment with virtual objects is created. The map database agent offers a static virtual map if the control agent calculates that $\lambda_{o,}$ $\lambda_{t,}$and $\theta_{o,}$ θ_t (respectively object and target coordinates) are in the same real map area. In this phase the navigation is monitored. Control agent is a knowledge-based decision support system which need to act as an adviser. It uses a discrete simulation because it is event -oriented and based on the concept of a complex discrete events system. The geometrical interpretations of object path are obtained from a motion planner.

We need a block of information agents that can give retrieving information from external information sources in response to one shot; requests for periodic information; monitoring external information sources for the occurrence of given information patterns, called change monitoring requests [9].

In the second phase the simulation model is modified when the states of the real object change and current real states are introduced into the model. The focus of planning in our system is on explicating the basic information flow relationships between tasks and other relationships that affect control-flow decisions. Most control relationships are *derivative* of these more basic relationships[10]. All the agents know the goal or destination and they all aim to achieve it. Cognition, perception, action, and learning are all essential components of such systems and their integration into real systems of different levels of complexity should help to clarify the nature of artificial intelligence. This control system will be equipped with:

- A map database of possible motion environment.
- Visual sensors to detect the obstacle's presence and to sense the distance.
- A virtual GPS antenna which gives coordinates every time to system to make a control of its position on the map.

4 The Application

This section aims to analyze the simulation of the orientation process using an algorithm. The simulation shows a rational decision –making behavior of model. We have build a simple algorithm. We see the problem of planning from a practical point of view finding the solutions on the path of motion in a virtual map which must show in real time the dynamic changes of the real map. This is a very important moment because the simulation should demonstrate intelligent behaviors of the system that plans the next path according to the peripheral pixels around the obstacle. Explaining the Figure 4, we are going to represent an application of agent based system that will operate in a two dimensional virtual world map, where the obstacles are coast areas, island areas and other objects positioned statically or dynamically in the map. The agent starts the motion executing a path build in a static digital map. The control system observes if there are changes by comparing the new agents coordinates and the computed coordinates in the simulation off line. Figure 5 presents a view of the computer assisted ship motion when there are not notifications for direction changes by the information agents on line.

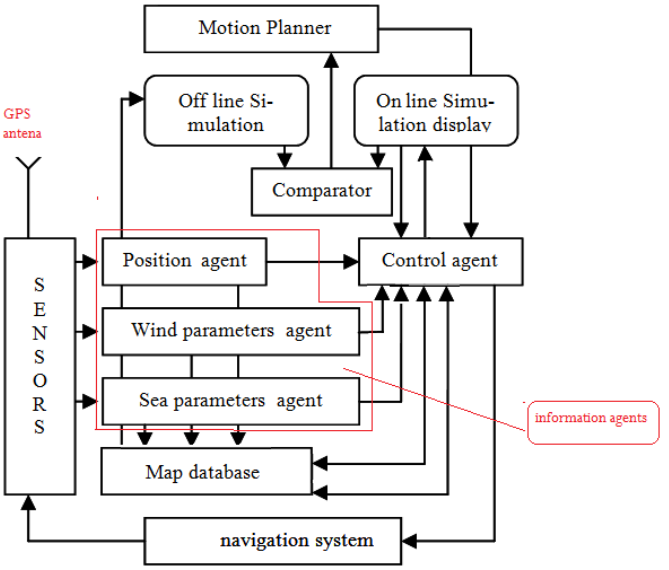


Fig.3. A simple structure of control system

A module of control system products a new direction minimizing the direction angle error made from other conditions of the motion in a real world (Figure 6). When the adaptation process finishes with the result " no changes", the control system allows the motion offered by the path planning of line. If the agent step is not possible, the control agent requires a dynamic new path from a motion planner.

For convenience, we assume that the ship-agent enters on the left from a selected position and its task is to find a target that moves randomly in the same environment. The object controls and detects the target using its sensors and a simple algorithm which is based on dividing the area of simulation in visible areas and invisible areas. This will be done through a software application based on the graph visibility algorithm that we have constructed for this purpose [11].

When the target is not detected in the first area, this state products a motion process of the object following the line of sight which divides the two areas. After a selected time interval, the object controls for a new state, senses, plans and decides to act for collecting more information or to achieve successfully the goal.

5 Expected Results

The application shows a rational decision –making behavior of model. This has been done through a simulation that we have constructed for this purpose. The required level of intelligence for a system agent based depends on how fulfilled its knowledge is about its environment. The intelligent control agent should be able to self-determine the direction of the ship motion avoiding collisions with obstacles or other ships.

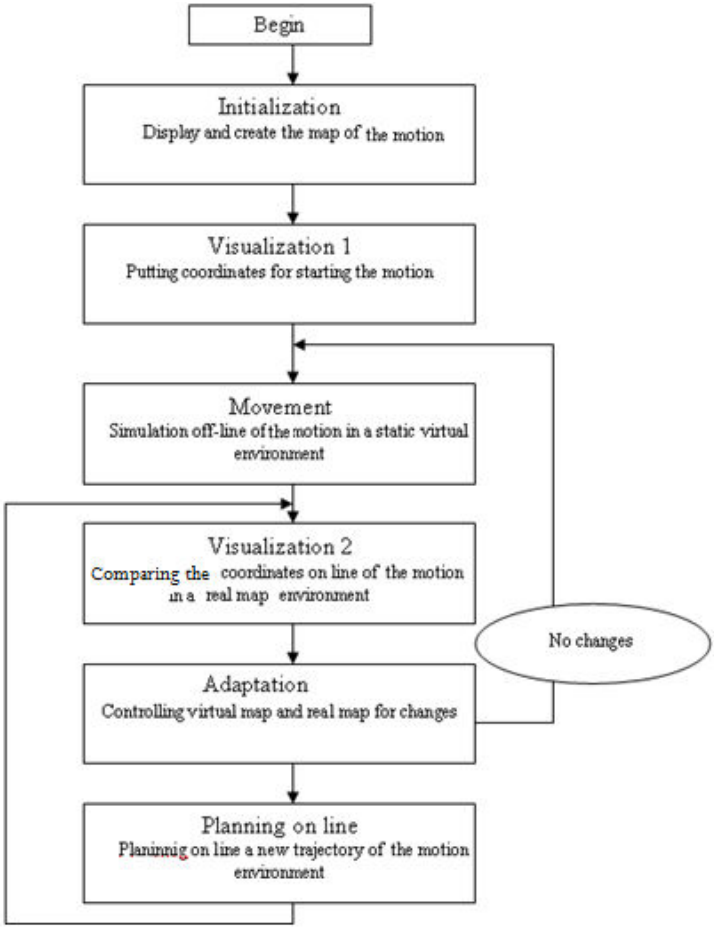


Fig .4. Algorithm of the ship motion

To synthesize a completely autonomous computer-assisted control system operating in a dynamic environment, we have used concepts, ideas and tools from artificial intelligence, simulating a software system agent based as a part of completed sea navigation system which is able to lead successfully a ship in partially known environment.

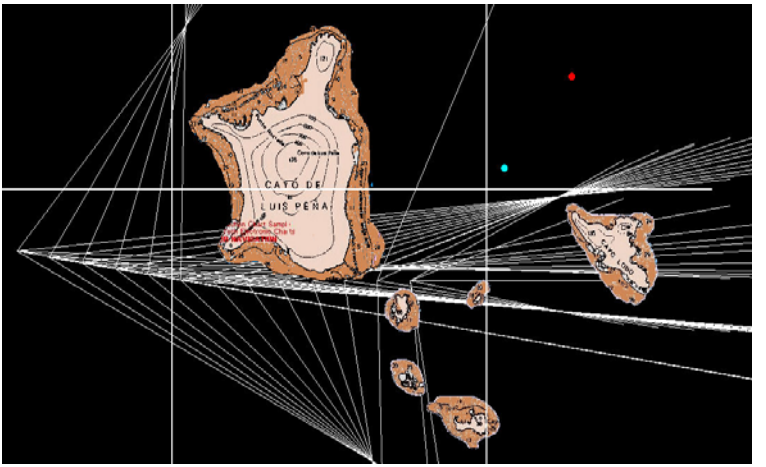


Fig. 5. View of the simulation off line

There are some practical implications. First we had difficulties in simulating a visual sensor which could sense the distance. Then the simulation became more complicated because of the presence of an unexpected obstacle over the line of navigation. The system should perform an intelligent behavior using a simple reflex agent model. In this case , the system combines some behaviors. Changing a state and choosing the appropriate state when there is a selected step during the planning process are two of the main challenges that we tried to overpass.

6 Conclusions

Designing an intelligent system virtually was an experience that required solutions for some problematical situations.

Practically we have realized only the first part of the work – an intelligent control which can plan a trajectory of motion based on the virtual map chosen by the map database agent at the begining of the vessel mission.

The control agent makes decisions in real time in a second phase selecting a range of reactive actions proveded by the GPS agent in presence of uncertenntly obstacles adapting to the changes of the environment.

We are going to represent an application of agent based system that will operate in a two dimensional virtual world map, where the obstacles are coast areas, island areas and other objects positioned statically or dynamically in the map. The system is a set of agent modules as passive and active entities making intelligent decisions to complete pre-selected tasks; control agent is able only to exhibit goal-directed behavior taking the initiative.

The model is based on simulating the navigation using a computer assisted control system in a virtual space. We treat theoretically the control method as an agent based software system but we do not give a complete solution.

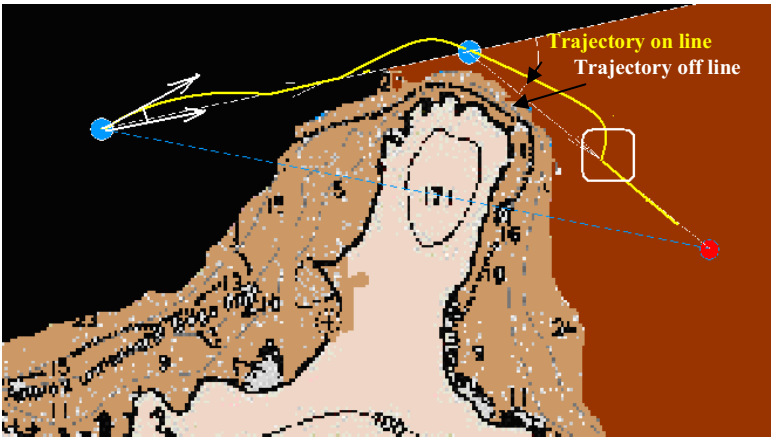


Fig. 6. Visible and not visible area in the map simulation

The path is decided off line. We do not consider the calculations on line made from the other agents and the simulation does not offer a complete view of the complex intelligent control system.

The control agent needs to interface the other agents. It should accumulate requested knowledge. We do not treat this process in this case but it is our future objective.

References

1. Decker, K., Sycara, K., Williamson,M. (1997) Intelligent Adaptive Information Agents. Journal of Intelligent Information Systems, Vol 9 :239-260
2. Wooldridge, M., Jennings,N.R.(1995). Intelligent Agents: Theory and Practice, Wooldridge, M. (2000). Reasoning about Rational Agents. The MIT Press Cambridge,
3. Luger, G. F., Stubblefield, W. A.,(2005) Artificial Intelligence Structures and Strategies for Complex Problem Solving Third Edition:272-277
4. Brooks, R.(1986). A Robust Layered Control System for a Mobile Robot. J. Robotics and Automation : 14–23
5. Penn, A.,Turner, A., (2002). Space Syntax Based Agent Simulation, Springer-Verlag, Berlin, Germany: 99-114.
6. Russell, S., Norvig, P. (1995). Artificial Intelligence: A Modern Approach, Prentice–Hall:: 31-33.
7. Macal, Ch. and NorthMacal, J. C.M.(2006) Tutorial on agent based modeling and simulation part 2 : how to model with agent , Simulation Conference, WSC 06. Proceedings of the Winter, Monterey, CA, Publication :73-83
8. Witold, J(2005). Intelligent robotic systems, design, planning and control, 123-146
9. Decker,K., Sycara,K., Williamson,M. (1997) Intelligent Adaptive Information Agents. Journal of Intelligent Information Systems, Vol 9 :239-260
10. Williams,B.,Nayak,P. (1997), A reactive planner for a model –based executive, Proc. Int.Joint Conf on Articial Intelligence: 1178-1185

11. Russell, S. Norvig, P. Artificial Intelligence. (1995). A Modern Approach. Prentice–Hall, 228-230.