

**J**OURNAL   
of Applied Economic Sciences

**Quarterly Journal**

**Volume V  
Issue 2(12)  
Summer 2010**

## Editorial Board

### Editor in Chief

Laura **Ștefănescu**

### Managing Editor

Mădălina **Constantinescu**

### Executive Editor

Mihaela **Galiceanu**

### International Relations Responsible

Pompiliu **Constantinescu**

### Proof – readers

Camelia **Firică** – *English*

Ana-Maria **Trantescu** - *English*

### Redactors

Cristiana **Bogdănoiu**

Sorin **Dincă**

## Editorial Advisory Board

Claudiu **Albulescu**, University of Poitiers, **France**, West University of Timișoara, **Romania**

Huseyin **Arasli**, Eastern Mediterranean University, **North Cyprus**

Cristina **Barbu**, Spiru Haret University, **Romania**

Christoph **Barmeyer**, Universität Passau, **Germany**

Amelia **Bădică**, University of Craiova, **Romania**

Gheorghe **Bică**, Spiru Haret University, **Romania**

Ana **Bobîrcă**, Academy of Economic Science, **Romania**

Anca Mădălina **Bogdan**, Spiru Haret University, **Romania**

Elena **Doval**, Spiru Haret University, **Romania**

Jean-Paul **Gaertner**, École de Management Strasbourg, Université de Strasbourg, **France**

Shankar **Gargh**, Editor in Chief of Advanced in Management, **India**

Ioan **Gâf-Deac**, Spiru Haret University, **Romania**

Eugen **Ghiorghită**, Spiru Haret University, **Romania**

Emil **Ghiță**, Spiru Haret University, **Romania**

Dragoș **Ilie**, Spiru Haret University, **Romania**

Arvi **Kuura**, Pärnu College, University of Tartu, **Estonia**

Ion Viorel **Matei**, Spiru Haret University, **Romania**

Constantin **Mecu**, Spiru Haret University, **Romania**

Piotr **Misztal**, Technical University of Radom, Economic Department, **Poland**

Marco **Novarese**, University of Piemonte Orientale, **Italy**

Russell **Pittman**, International Technical Assistance Economic Analysis Group Antitrust Division, **United States of America**

Ion **Popescu**, Spiru Haret University, **Romania**

Kreitz Rachel **Price**, École de Management Strasbourg, Université de Strasbourg, **France**

Peter **Sturm**, Institut National de Recherche en Informatique et Automatique, **France**

Andy **Ștefănescu**, University of Craiova, **Romania**

Laura **Ungureanu**, Spiru Haret University, **Romania**

Hans-Jürgen **Weißbach**, University of Applied Sciences - Frankfurt am Main, **Germany**

## Publisher:



European Research Center of Managerial Studies in Business

Administration Brazda lui Novac Street, no 4 Craiova, Dolj, Romania

Email: [jaes\\_secretary@yahoo.com](mailto:jaes_secretary@yahoo.com)

## ***Journal of Applied Economic Sciences***

**Journal of Applied Economic Sciences** is a young economics and interdisciplinary research journal, aimed to publish articles and papers that should contribute to the development of both the theory and practice in the field of Economic Sciences.

The journal seeks to promote the best papers and researches in management, finance, accounting, marketing, informatics, decision/making theory, mathematical modelling, expert systems, decision system support, and knowledge representation. *This topic may include the fields indicated above but are not limited to these.*

*Journal of Applied Economic Sciences* be appeals for experienced and junior researchers, who are interested in one or more of the diverse areas covered by the journal. It is currently published quarterly with three general issues in Winter, Spring, Summer and a special one, in Fall.

The special issue contains papers selected from the *International Conference* organized by the *European Research Centre of Managerial Studies in Business Administration* and *Faculty of Financial Management Accounting Craiova* in each October of every academic year. There will prevail the papers containing case studies as well as those papers which bring something new in the field. The selection will be made achieved by:

- Conference Session Chairs
- Editorial Board Members

**Journal of Applied Economic Sciences** is indexed in CEEOL [www.ceeol.org](http://www.ceeol.org), SCOPUS [www.scopus.com](http://www.scopus.com), EBSCO [www.ebsco.com](http://www.ebsco.com), RePEc [www.repec.org](http://www.repec.org) and in *IndexCopernicus* [www.indexcopernicus.com](http://www.indexcopernicus.com) databases.

The journal will be available on-line and will be also being distributed to several universities, research institutes and libraries in Romania and abroad. To subscribe to this journal and receive the on-line/printed version, please send a request directly to [jaes\\_secretary@yahoo.com](mailto:jaes_secretary@yahoo.com)




# *Journal of Applied Economic Sciences*

ISSN 1843-6110

## Table of Contents



	Alessio Emanuele <b>BIONDO</b> <i>A Growth Rate for a Sustainable Economy</i>	... 7
	Maria <b>BOBROVA</b> , Arndt <b>KÜMPEL</b> <i>Reasoning on Evolution of Culture and Structure</i>	... 21
	A.B. <b>BONACHE</b> , J. <b>MAURICE</b> , K. <b>MORIS</b> <i>A Best Evidence Synthesis on the Link between Budgetary Participation and Managerial Performance</i>	... 34
	Ginters <b>BUSS</b> <i>Forecasts with Single-Equation Markov-Switching Model: An Application to the Gross Domestic Product of Latvia</i>	... 48
	Lisi <b>GAETANO</b> <i>The Unemployment Volatility Puzzle: The Role of the Underground Economy</i>	... 59
	Giuseppe <b>GAROFALO</b> , Patrizio <b>MORGANTI</b> <i>The Financing of R&amp;D Investments: Effects on Growth and Financial Structure</i>	... 70
	Ivan O. <b>KITOV</b> , Oleg I. <b>KITOV</b> <i>Dynamics of Unemployment and Inflation in Western Europe: Solution by the 1-D Boundary Elements Method</i>	... 94
	Evgenia <b>MOTCHENKOVA</b> , Daniel <b>LELIEFELD</b> <i>Adverse Effects of Corporate Leniency Programs in View of Industry Asymmetry</i>	... 114
	Rajesh K. <b>PILLANIA</b> <i>Indo-China Trade: Trends, Composition and Future</i>	... 129
	Georg <b>QUAAS</b> <i>Was the Worldwide Asymmetry in Current Accounts Caused by the Macroeconomic Policy of the Global Economy's Leader?</i>	... 138

## A GROWTH RATE FOR A SUSTAINABLE ECONOMY

Alessio Emanuele BIONDO  
University of Catania, ITALY  
Faculty of Economics  
[ae.biondo@unict.it](mailto:ae.biondo@unict.it)

### Abstract

*A sustainable growth rate – i.e. a growth rate which allows economy to expand without compromising the equilibrium of the natural system – is one of the most important and stimulating topics in growth literature. In this paper two findings will be presented. First of all, a brief discussion of both concepts – growth and development – is presented. A new sight for their relationship is suggested. The usual distinction between quantitative and qualitative variables is shown to be unsatisfactory. Growth and development must fit in a sustainability framework and therefore, progress should be based on steps of sustainable economic growth in order to have higher development levels. Secondly, a two-sector-closed-economy model is presented to demonstrate the existence of a positive sustainable growth rate for the GDP.*

**Keywords:** economic growth, sustainable growth, development, sustainability

**JEL Classification:** O100, O400, O490, Q010, Q500, Q560

### 1. Introduction

The distinction between growth and development has usually been conducted at level of the difference between quantitative and qualitative variables, in related literature. Coherently with this view, according to Shearer (1961), economic growth refers to an increase in the output of goods and services (being therefore a quantitative concept), while economic development implies a more general and qualitative concept, including personal and social values. Nonetheless, this apparently simple distinction has not always been sufficiently clear according to Sen (1992), easily recognizable as the father of the capability approach, who stated clearly that the idea of development must not be confused with the increase in quantity of goods available within an economy. Pearce and Warford (1993) stress that being development a process that leads to improvement or progress, a society which follows this process of economic development will obtain a combination of three effects: first of all, an advance in utility (in terms of per capita income, quality of the environment, and general social well-being); secondly, advances in education, health, and quality of life (in this exposition they use the same classification adopted by Goulet, 1971); and third, growth of self-esteem and self-respect, which leads to independence and capacity of choice.

Economic systems are made by several variables, and they can be defined as developed when they have some attributes<sup>1</sup>. The more a region is developed, the greater and the deeper are components of its welfare. It means that after primary needs, there comes a series of other and higher qualifications which express goodness of quality of life. Then, economic development appears clearly as a much wider concept than economic growth, being the latter defined as just an increase in the level of the per capita gross national product over time. The way in which the progress is pursued puts in evidence the match against the environmental constraint; sustainability assumes a crucial relevance because considered the world today, and given characteristics of mankind and nature, it may not be possible anymore to define development without placing it within a sustainability framework.

It is difficult to give a unique definition of sustainability and/or of sustainable development because of the availability of alternatives that it is possible to find in the literature. Pearce *et al.* (1989) offer more than thirty possible definitions. These definitions can be divided into two groups (Beltratti

---

<sup>1</sup> For example the *World Bank Development Indicators* include the eradication of extreme poverty and hunger, the achievement of universal primary education, the promotion of gender equality, the reduction of children mortality, the improvement of maternal health, the struggle against HIV/AIDS, malaria and other diseases, and the environmental sustainability. According to some other authors (Pearce, Barbier, Markandya 1989), development would involve an increase in real income per capita, improvements in health, nutrition, and education, a fairer distribution of resources and income and an increase in what they call “basic freedoms”.

1996): the first group refers to sustainability in physical terms, looking at the limitations imposed by scarcity of natural resources on the growth process; the second group describes sustainability on the basis of the comparison among utility levels of different generations. Summing up, a broad definition of sustainability would include the preservation of human wellbeing by the maintenance of natural, social, and economic systems. Following Kunte *et al.* (1998), it is possible to define wealth as the per capita stock of assets.

$$W = \frac{K_m + K_n + K_h}{POP}$$

where  $K_m$  is the stock of man-made capital,  $K_n$  is the stock of natural capital,  $K_h$  is the stock of human capital, and  $POP$  is the population. This composition allowed the distinction between strong and weak sustainability: the idea of weak sustainability, based on the studies of Solow (1974) and Hartwick (1977), allows the substitution between natural capital and man-made capital because it is based on the assumption that welfare is not dependent on a specific form of capital and that there is a near perfect substitutability between man-made capital and natural resources. If such a substitution is possible, an economy is recognized as sustainable even if it draws down its stock of natural capital, provided that it creates enough manufactured capital to compensate for the loss of natural capital so that the constancy of the total stock of capital is ensured (Neumayer 1999). Instead, the strong sustainability criterion requires maintaining different kinds of capital intact separately; therefore it refers to the case in which substitutability is not allowed<sup>2</sup>: according to the strong sustainability view, at least some natural capital is non-substitutable and should be maintained at or above some threshold levels. Natural capital that is not substitutable by any other form of capital is called critical natural capital<sup>3</sup> and its preservation assumes a great relevance. Determination of criticality depends on ecological, as well as economic, political and social criteria (Mac Donald *et al.* 1999) and critical levels depend not only on ecological standards, but are also related to standards of living and relative affluence of a particular group, region or nation (Pearce, and Warford 1993), in the sense that the degree to which a function is considered important (i.e. critical) may vary from place to place, from population to population.

From what has been said, sustainability is not only an attribute for development, but also a qualification for the path which allows reaching it. If development were a ladder, growth would be the progress that a system does from the first to the higher steps. For this reason it is possible on one hand to agree with the part of the literature which refers to growth as a path made by stages (see, for example,

---

<sup>2</sup> Some authors do not admit either substitution or compensation (Sen 1982; Barry 1991; Spash 1994; Azar 2000); furthermore, there are some resources so important that substitutability is not even practicable, as for example the ozone layer. Muradian (2001) puts in evidence how surpassing some critical levels in the depletion of some resources can imply unknown and important transformations. Finally, Page (1983) states that opportunities of the next generations will not be threatened if they will inherit the same resource base that the present generation has inherited; while Bromley (1989) finds the intergenerational justice in the case in which every generation can receive undiminished resource stock and environmental quality.

<sup>3</sup> In literature, several definitions for ‘critical’ natural capital have been given:

“Vital parts of the environment that contribute to life support systems, biodiversity and other necessary functions denoted as ‘keystone species and processes’” (Turner 1993).

“Critical natural capital consists of assets, stock levels or quality levels that are: (1) highly valued; and (2) essential to human health; or (3) essential to the efficient functioning of life support systems; or (4) irreplaceable or non-substitutable for all practical purposes (e.g. because of antiquity, complexity, specialization or location)” (English Nature 1994).

“That set of environmental resources which performs important environmental functions and for which no substitutes in terms of human, manufactured or other natural capital currently exist” (Ekins 2003).

Mac Donald *et al.* (1999) underlines how more comprehensive definitions of critical natural capital focus on two main aspects: the “functional” aspect of the natural capital referred to those ecological assets that are essential to human wellbeing or survival (*see also*: Pearce, and Warford 1993) and the “primary” aspect of ecosystems for general biosphere functioning which requires to maintain population or resource stocks within bounds thought to be consistent with ecosystem stability and resilience (*see also*: Turner 1993). Thus, the latter preserves the eco-centric stability (in the sense of maintaining environmental integrity looking mainly at the maintenance of “habitat functions”); and the former, based on an anthropocentric perspective, basically defends those natural equilibria that are indispensable to human survival and cannot be substituted.

Rostow 1952), and on the other hand to deepen the strength of sustainability inside of the meaning of economic growth. This will lead to development, whose definition has no more sense whether it is not sustainable, given that all the dimensions of the sustainability concept enter into utility functions of humans and therefore contribute to define welfare. This approach is pursued by Costantini (2006) who deepens the analysis about quality in economic systems also referring to the *Human Development Index* (HDI), establishing a relationship between human development and sustainability. Her conclusion shows that sustainability is a multidimensional concept based on economic growth, wealth, and natural capital. Coherently also with Biondo (2004), she underlines how a sustainable growth path can be followed starting from a sufficient human development level – which evidently implies satisfaction of basic needs. The higher is the starting-level of development, the easier is sustainable growth.

There are a lot of contributions in growth literature dealing with sustainability and how economic growth (production, technology, and population) influences the environment, receiving at the same time constraints from it. It is not possible to survey them all here, but it is possible to assess a rapid overview at least to recall some of recurring topics: first of all, at least implicitly, literature almost always refers to well-known environmental functions (highlighted by Pearce and Turner, 1990): support to life, input to production, and waste recovery. For example, Dasgupta and Heal (1974) is perhaps the first rigorous consideration of the optimal path for consumption in a representative agent single-commodity model, where the agent maximizes discounted utility and production uses inputs of capital and environmental non-renewable resource; many other models used that framework since then. The environment is seen as input also in Stiglitz (1974), where the author seeks the optimal rate of utilization for resources and builds up a model where environment is a necessary input for production. This kind of analysis opens another important stream of literature, in terms of intergenerational opportunities, as in Solow (1974), who tried to establish proof of intergenerational equity finding the optimal depletion of natural resources. The intergenerational issue, present also in Howarth and Norgaard (1992), who focused on the topic of justice between generations and proposed intergenerational transfers, is strictly connected to the idea of sustainability. A number of papers dealt with sustainability concept, and Pearce *et al.* (1989) listed more than thirty different definitions of it. Other important contributions came from Redclift (1992), Pezzey (1992 and 1997), Barbier and Markandya (1990), and Lele (1991); Pearce and Atkinson (1993) tried to give a measure of sustainability, as in Hamilton *et al.* (1998), and in Hamilton and Clemens (1999) where the path for sustainability passes through stimulating savings. Weitzman (1997) looked at technical progress in relation to sustainability, and similarly Jaffe *et al.* (2000) presented an analysis focused on the impact of technology on the environmental economic field. Seeking for “sustainability rules”, many papers gave important ideas; for example, Hartwick (1977) showed that constant consumption is warranted for an economy with a constant returns to scale Cobb-Douglas production function with capital and non-renewable resource as inputs, equalling resource depreciation to investment in reproducible capital, and defined and generalized his homonymous rule for an optimal resource use in subsequent contributions; Beltratti *et al.* (2000) proposed the Green Golden Rule; Chambers *et al.* (2000) built an indicator for sustainability; Smulders (2000) defined his concept of balanced growth; Pittel (2002) surveyed endogenous growth theory in sight of sustainability. In between streams of intergenerational and sustainability issues, Farmer and Randall (1997) analyzed sustainability through an overlapping generations framework while Pezzey (2004) traced a distinction between environmental policy and sustainability policy referring to intergenerational equity. More broadly, Beltratti (1996) deepened the matter of inserting environment in growth models, and van den Bergh and Hofkes (1999) wrote an interesting review of sustainable-development economic models. Another stream of literature deals with human development, as for example Anand and Sen (2000) where human development is related with sustainable development, while, less specifically, Ranis (2004) reported the evolution of development debate.

The trade-off between growth and environment is not easy to be described, because it is multidimensional; Grossman (1995) suggests that growth affects the environment through three effects: the “scale effect”, the “composition effect”, and the “technique effect”. The “scale effect” simply refers to the augmented quantity of produced output that leads per se to a greater exploitation of the environment, in terms of both resource consumption and polluting emissions. The “composition effect” is the consequence of higher income on the economic activity and life of the system, in the sense that the more the income increases, the more cleaner activities and less pollutant technologies will be preferred: in Grossman’s view, the composition effect is oriented toward the supply side, where industries



substitute agriculture at first, and then the service sector substitutes the industrial one. At the same time, this composition effect will automatically stimulate the “technique effect”, because innovation and progress assume more relevance: in fact, the cleaner the process of production, the stronger must be the innovation and the capacity of the R&D sector at social level. Not only does innovation improve efficiency, but also it reduces usage of environment and allows natural capital to be maintained and, if necessary, reintegrated over time. Only if innovation reduces the use of environmental inputs, a weak sustainability framework can be possible. This point is particularly important in sight of the model presented here. It constitutes a theoretical perspective which successfully describes economic environmental interaction, making the important differentiation between expenditure in innovation and expenditure in recovery and substitution of natural capital; furthermore, it derives the sustainability condition for economic growth (taking into account the grade of consciousness of people), and finally it gives the theoretical structure for future quantitative analysis. Evidently, whether economic growth is good or not for the environment mostly depends on the presence of adequate policies: as Arrow *et al.* (1995) underline, given that all the activity of the economic system depends on the environmental resource base and that every misuse of these resources may reduce the capacity for generating material production for the future, there exists the need for the creation of institutions in order to pursue right policies and completeness of the markets. But this is a higher level of the problem; policies are possible if they are based on instruments which can actually tune variables in the economic system; therefore, the analysis of the relationship between growth and sustainability must at first pass through the exposition of a simple model which will try to show conditions for the economy can grow in a sustainable way. Section two will present the model, section three will conclude.

## 2. The Model

The model presented here, will derive a condition for a positive sustainable growth rate; in doing this attempt, it will represent a very simple two-sector-closed-economy framework, in which only one good is produced and consumed: one sector will be devoted to production of the unique good, while the other sector is used by policy-maker for environmental purposes as it will be explained later. An extensive literature has coped with environmental elements of production and consumption: there are models based on a dynamic optimization problem in which the utility of an infinitely lived representative agent is maximized within the framework of the optimal control theory, and models based on the endogenous approach with increasing returns to scale for the production function. Few examples of the first stream of models are in Tahvonen and Kuuluvainen (1993), where pollution enters both the production and the utility function, and in Lopez (1994), where environment plays directly the role of productive factor; whereas examples of the second kind of models are in Bovenberg and Smulders (1995), where an aggregate stock of environmental services can be found in both production and utility, and Stokey (1998) where pollution is a function of output and enters the utility function. As in Bovenberg and Smulders (1995), the model here will consider an aggregate set of services that the environment delivers to social and economic activities: it simply shows that in order to produce and to consume the unique good  $X$ , the economic system uses environmental resources (for example in terms of raw materials, waste disposal, recreational reasons, just to mention a few); as a result, this “production–consumption cycle” of  $X$  depletes environment and generates polluting wastes ( $W$ ) which can be reasonably assumed as a growing function of the amount of produced–consumed output. Let the relation between  $W$  and  $X$  be expressed by the following function:

$$W = \alpha(\bullet)X \tag{1}$$

At this stage,  $\alpha$  will be considered simply as a positive parameter, but later this assumption will be relaxed and it will be described as a function, in order to explain elements which can affect it; however will be assumed<sup>4</sup> that  $\partial W/\partial X > 0$ . Put in this way,  $\alpha$  represents how much each unit of  $X$  is polluting both to be produced and to be consumed, and the greater it is, the more  $X$  pollutes. Over time,

---

<sup>4</sup> The basic idea that must be taken in mind is that the positive variation of  $X$  implies a positive variation in  $W$ . With the assumption of  $\alpha$  as a simple positive parameter, the equation (1) could be written as:

$$W = \alpha X, \alpha > 0 \tag{2}$$

while production and consumption go on, wastes follow an accumulation process which results in the total amount of pollution  $P$ , as in Tahvonen and Kuuluvainen (1993), who basically built up on Brock (1977), considering recycling capacities of environment, which absorbs part of pollution and wastes and converts them into resources again:

$$\dot{P} = W - \beta P \quad 0 < \beta < 1 \quad (3)$$

Such an equation of motion is quite standard in models which deal with sustainability; it is worth to notice, Smulders (2000) uses a very similar one, but he refers just to production in his accumulative process which explains environmental quality evolution, whereas here  $W$  includes consumption contribution to pollution. In fact, not only consumption of any good implies creation of wastes, but environment itself is consumed for health, fun, and lots of recreational issues by agents. As it is easily understandable,  $\beta$  represents the capacity of the environment to absorb pollution and wastes; it is  $\beta \in (0, 1)$ , because it can be hypothesized that the “carrying capacity” of the environment may vary between two theoretical extremes: the impossibility to absorb ( $\beta = 0$ ) and the capacity to recycle all the pollutants ( $\beta = 1$ ).  $\beta$  could be assumed as a function of natural capital stock (NK):  $\beta = \beta(NK)$ ,  $d\beta/dNK > 0$ . Many times in literature the capacity of environment to absorb pollution is depicted as a natural renewable resource (see for example: Pearce, and Warford 1993).

For any value of  $\beta$ , it is important to underline here that the model describes a path which links unavoidably continuous production and consumption of  $X$  to growing pollution and wastes (asymptotically infinite if the realistic assumption made here, that  $\alpha > \beta$ , holds true<sup>5</sup>).

Considering now the growth rate of eq.(1)<sup>6</sup>, the growth rate of wastes results:

$$\hat{w} = \hat{\alpha} + \hat{x} \quad (4)$$

In order to demonstrate sustainability, the model must account for what will happen whenever the absorption capacity of the environment is completely saturated. In that moment,  $\beta = 0$  in (3) and the economy reaches the maximum quantity of pollution and wastes which saturates nature totally. In this extreme case, the idea of environmental sustainability arises clearly: for the system to be sustainable, since that moment on, pollution cannot increase anymore. That is: if

$$\beta = 0 \Leftrightarrow \dot{P} = W \quad (5)$$

Then sustainability implies:

$$\hat{p} \leq 0 \Leftrightarrow W \leq 0$$

That, in turn implies:

$$\hat{w} = \hat{\alpha} + \hat{x} \leq 0 \quad (6)$$

The production–consumption of  $X$  will be sustainable if it grows at a rate which can satisfy the condition in (6). In order to let production–consumption continue at a positive rate, it is necessary that  $\hat{\alpha}$  is negative (and greater than or at least equal to  $\hat{x}$ , in absolute value). In other words, R&D expenditure in technology and innovation would be compulsory to ensure correspondent reductions of

---

<sup>5</sup> In a certain way,  $\alpha > \beta$  is *obvious* in sight of model, because the case in which  $\alpha \leq \beta$  would imply a production-consumption process whose wastes were completely absorbed by the environment without increasing pollution; therefore the interest in studying the environmental sustainability of such a production-consumption cycle would be extremely scarce.

<sup>6</sup> All of the growth rates will be indicated with the same name of the corresponding variables but in lower case and with a superscript.

$\alpha$  : this should lead to a new (higher) level of development in which  $W$  and  $x$  are linked by a different function – i.e.  $\partial W/\partial X$  decreased. Define now:

$$NI = NI(\pi; X) \text{ and } \pi = \pi(NK) \quad (7)$$

as the investment expenditure to obtain environmental-related technical progress; in other words, it is the R&D expenditure devoted to discover new technologies which can save natural capital ( $NK$ ). This function is particularly relevant in sight of the meaning of this model; in fact, its role is twofold: by one side it represents the R&D expenditure for innovative investments ( $NI$  depends positively on  $X$  because the more the good is produced and consumed, the more R&D expenditure is required, therefore  $\partial NI/\partial X > 0$ ), and by the other it indirectly indicates the amount of natural services required by production<sup>7</sup> and the quality of the actual consistency of natural capital.  $\pi$  represents here the consciousness of the policy-maker about environmental conditions and expresses therefore the preference for innovative investment in terms of expenditure in R&D focused to reduce wastes and polluting emissions. It depends negatively on  $NK$ , because the more  $NK$  is exploited, greater is the importance of the environmental problem, the more  $NI$  is necessary: therefore  $d\pi/dNK < 0$  and  $\partial NI/\partial \pi > 0$ <sup>8</sup>.

Let now  $NC$  be the total amount of other expenditures and costs (different from R&D) necessary to clean nature and recover natural capital. This function behaves exactly like  $NI$ : everything has been explained for  $NI$  holds for  $NC$ . Then it can be written:

$$NC = NC(\rho; X) \text{ and } \rho = \rho(NK) \quad (8)$$

where  $\rho$  has the same role that  $\pi$  does<sup>9</sup>, but referred to cleaning and recovery expenditures. The model, then, takes in account the “weak sustainability” perspective in a certain way, but underline a very important difference between  $NI$  and  $NC$ : they differ in terms of timing, amount and applicability. Environmental protection does not rely solely on substitution between natural and man-made capital ( $NC$ ), but especially on innovative expenditure which actually reduces usage of environment.

The motion of  $NK$  is then given by:

$$N\dot{K} = NI + NC - P \quad (9)$$

The accumulation process for  $NK$  in (9) shows the source of the idea that policy should deal with maintaining a non-declining natural capital; it was first developed in Pearce and Turner (1990) and gives in this model the idea of what sustainability can be at the point where  $\beta=0$ : in order to have non-decreasing (critical) natural capital,  $(NI + NC) \geq P$  must hold.

The rationale underneath the model is that the “generic” good  $X$  is exactly the domestic product of the economy. In this broader perspective, the problem is revealed: the sustainability of further production–consumption of  $X$  is the problem of the sustainability of economic growth. Given elements which have been used into the model, and in the shape of the important part of literature dealing with new national accounting systems<sup>10</sup>, the traditional fundamental closed-economy income equality,

<sup>7</sup> This will be immediately clarified when  $\alpha$  will be presented as a function.

<sup>8</sup> It is easily understandable why  $\pi$  is function of  $NK$ . It is evidently related to sensitivity of people to the environmental problem and to actual wealth: briefly it can be said that the more the system approaches the critical level of natural capital, the more  $\pi$  will increase, and this will in turn imply increasing  $NI$ . The definition of a proper function form for  $\pi$  is not the main goal of this paper, but it is extremely interesting to deepen just one aspect about its elasticity: the consciousness of the environmental problem is not identical everywhere. Then, the function will behave differently from place to place, according to preferences of people, being more rigid where the environmental impact is not deeply considered, and more elastic where people is more sensitive.

<sup>9</sup> Then it is:  $\partial NC/\partial X > 0, d\rho/dNK < 0, \partial NC/\partial \rho > 0$ .

<sup>10</sup> Particularly relevant in this field are: UN (1992, 1993), where first UNSTAT proposals for satellite accounts can be found, Lutz (1993), Bartelmus (1994), Wackernagel and Rees (1997) for a simple introduction to

$$X = C + I + G$$

Adding terms here presented, can be rewritten as:

$$Y = X + NI + NC \quad (10)$$

This more complete definition of GDP includes the “traditional” components ( $X$ ) plus  $NI$  and  $NC$  as they have been just defined: (10) reveals that the model presents a second sector,  $N=NI+NC$ , dedicated to environment, whose activities are aimed both to develop new technologies in order to substitute natural capital with man-made capital and to recover natural system; the  $N$  sector is not pollutant by virtue of assumption. On the basis of what has been expounded up to now, it is finally possible to go back to  $\alpha$ , to deepen its determination and factors which influence it<sup>11</sup>:

$$\alpha = h(NI; NC) \quad (11)$$

Coherently with all the rest of the model it must be assumed that:

$$\frac{\partial \alpha}{\partial NI} < 0 \quad \text{and} \quad \frac{\partial \alpha}{\partial NC} < 0 \quad (12)$$

Differentiating (11) w.r.t. time, one obtains:

$$\begin{aligned} \dot{\alpha} &= \frac{\partial \alpha}{\partial t} = \frac{\partial \alpha}{\partial NI} \frac{\partial NI}{\partial t} + \frac{\partial \alpha}{\partial NC} \frac{\partial NC}{\partial t} = \\ &= \frac{\partial \alpha}{\partial NI} \frac{\partial NI}{\partial t} \frac{NI}{NI} + \frac{\partial \alpha}{\partial NC} \frac{\partial NC}{\partial t} \frac{NC}{NC} \end{aligned}$$

Then

$$\hat{\alpha} = \frac{\dot{\alpha}}{\alpha} = \frac{\partial \alpha}{\partial NI} \frac{\partial NI}{\partial t} \frac{NI}{NI} \frac{1}{\alpha} + \frac{\partial \alpha}{\partial NC} \frac{\partial NC}{\partial t} \frac{NC}{NC} \frac{1}{\alpha}$$

Rearranging

$$\hat{\alpha} = \frac{\partial \alpha}{\partial NI} \frac{NI}{\alpha} \frac{1}{NI} \frac{\partial NI}{\partial t} + \frac{\partial \alpha}{\partial NC} \frac{NC}{\alpha} \frac{1}{NC} \frac{\partial NC}{\partial t}$$

And, defining

$$\eta_{\alpha,NI} = \frac{\partial \alpha}{\partial NI} \frac{NI}{\alpha} \quad \text{and} \quad \eta_{\alpha,NC} = \frac{\partial \alpha}{\partial NC} \frac{NC}{\alpha} \quad (13)$$

It results

ecological footprints which describes methods of calculation, Vitousek *et al.* (1997) for a review of lots indicators, Neumayer (1999), Hamilton (2000) who reviewed theory of genuine savings, and Field (2001).

<sup>11</sup> In presenting next relationship, (12), it is necessary to focus on technology. Building up from a distinction made by Pemberton and Ulph (2001), two cases must be distinguished, according to the way by which technology could enter the model. As a first solution, technology can be considered completely endogenous: in this way it is considered implicitly in R&D activities, and the model still preserves capacity to take it in account; the second solution would imply an exogenous technical progress which does not arise from R&D: it would add in the model a time-dependent production possibility set. For a matter of simplicity the first solution has been used here.

$$\hat{\alpha} = \eta_{\alpha,NI} \hat{NI} + \eta_{\alpha,NC} \hat{NC} \tag{14}$$

The growth rate of  $\alpha$  is the weighted sum of the growth rate of  $NI$  and  $NC$ ; weights are  $\eta_{\alpha,NI}$  and  $\eta_{\alpha,NC}$  which are, as shown in (13), elasticity coefficients of the  $\alpha$  function respectively to  $NI$  and to  $NC$ . It is easy to notice that these elasticity coefficients are both negative given that, by virtue of assumption, conditions (12) hold. Recalling now (6), it is possible to write:

$$\hat{x} \leq -\hat{\alpha} \tag{15}$$

This means that for any positive growth rate of  $X$  there must be an (at least) equal reduction rate in  $\alpha$ . It follows immediately from (14) and (15) that:

$$\hat{x} \leq -\eta_{\alpha,NI} \hat{NI} - \eta_{\alpha,NC} \hat{NC} \tag{16}$$

Which can be finally written, as

$$\hat{x} \leq H_{NI} \hat{NI} + H_{NC} \hat{NC} \tag{17}$$

Posing

$$H_{NI} = -\eta_{\alpha,NI} \quad \text{and} \quad H_{NC} = -\eta_{\alpha,NC} \tag{18}$$

Inequality (17) gives the first result of the model. In order to be sustainable, the growth rate of the economy must be less than (or at maximum equal to) a weighted sum of the growth rate of  $NI$  and  $NC$ , representing weights the efficacy of environmental expenditure in R&D and cleaning in reducing  $\alpha$ . This brings to a weak sustainability idea implicitly: when the environment goes to collapse, the model explains that  $NI$  and  $NC$  substitute and rebuild destroyed natural capital. The effectiveness of this passage depends mostly on the level of development; in fact in developed economies elasticity coefficients will be higher and while  $X$  can be produced at acceptable growth rates, substitution and recovery of environment will be easier than in less developed economies. One important topic is to check whether  $\hat{x}$  can be always positive; in looking for demonstrating it, consider that because of conditions in (12), and given the (16), the R.H.S. of (17) is always positive, as  $H_{NI} > 0$  and  $H_{NC} > 0$ .

In order to show that  $\hat{x}$  can be always positive, it is necessary to analyse values of elasticity coefficients in (13) as done in Table 1, where the first column brings values for  $H_{NI}$ , the second column associates to each possible value for  $H_{NI}$ , every possible value for  $H_{NC}$ , the third column shows consequent results for R.H.S. of (17), and finally the fourth column shows sign of  $\hat{x}$ ,  $\forall \hat{NI}, \hat{NC} > 0$ :

**Table 1.** The analyse of values of elasticity coefficients

	$H_{NI}$	$H_{NC}$	$H_{NI}\hat{NI} + H_{NC}\hat{NC}$	$\hat{x}$
a.		< 1		
b.	< 1	= 1	> 0	> 0
c.		> 1		
d.		< 1		
e.	= 1	= 1	> 0	> 0
f.		> 1		
g.		< 1		
h.	> 1	= 1	> 0	> 0
i.		> 1		

As it can be seen,  $X$  can continue to grow: therefore  $\hat{x}$  is the wanted positive sustainable growth rate for the variable “production and consumption”. Once the policy-maker uses (17) there will be the possibility for the economy to grow sustainably; of course, there is no need to wait the saturation of environment to apply it: (5) was an important hypothesis needed to obtain (6) which leads to (17), but government should apply it before the moment when  $P = P_{MAX}$ , saving a lot of natural capital and

increasing wealth for citizens. In particular, it is useful to check how high  $\hat{x}$  can be, in the sense that a sustainable economy would choose to produce and consume as much as it can, looking for the highest wellbeing; in this sense (17) holds as an equality because  $\hat{x}$  will be pursued at its highest bounding value and following inequalities will hold:

1. in case a., if  $H_{NI} \leq H_{NC} \Rightarrow \hat{x} \geq \hat{NI} + \hat{NC}$
2. in case b.,  $\hat{x} > \hat{NC}$
3. in case c.,  $\hat{x} > 0$
4. in cases d., g., and h.,  $\hat{x} > \hat{NI} + \hat{NC}$
5. in case e.,  $\hat{x} = \hat{NI} + \hat{NC}$
6. in case f.,  $\hat{x} > \hat{NI}$
7. in case i., if  $H_{NI} \geq H_{NC} \Rightarrow \hat{x} \geq \hat{NI} + \hat{NC}$

The case 3 is the worst,  $\alpha$  is rigid to  $NI$  and  $NC$ , but production and consumption can still grow; all of other cases show that  $\hat{x}$  is at least higher than either  $NI$  or  $NC$ .

Finally, given that  $\hat{x}$  is an environmentally sustainable growth rate, and that it has been assumed that the  $N$  sector ( $NI+NC$ ) is not pollutant in any sense, recalling (10), it can be written that the sustainable growth rate for the two-sector-closed economy of the model will be

$$\hat{y} > 0 \tag{19}$$

That ensures a positive growth rate of GDP which is environmentally sustainable.

### 3. Concluding remarks

This paper introduced a qualitative perspective in approaching economic growth rate determination. The establishment of sustainability priority as an unavoidable ingredient of contemporary world's progress definition claims the check of possibilities which the actual system can pursue. Economic growth cannot remain just a quantitative expression and must share the qualitative fashion usually related solely to the development idea. The presented model demonstrates the existence of an always positive growth rate for the economy which can ensure however the sustainability of the progress through a weak-sustainability approach. Innovation and technological progress are presented in terms of the variable  $NI$ , i.e. the investment expenditure in R&D to discover new technologies in order to either improve environmental compatibility of the system or reduce the amount of natural resource exploited<sup>12</sup>. This conceptual difference between  $NI$  and  $NC$  is extremely important: innovation is not a simple substitution between natural capital and man-made capital. It is something more. Not only because the substitution is not always possible, but more clearly because innovation plays a unique role in reducing usage of environment in all of its forms.  $NC$  rebuilds, repairs, replaces natural capital, in the usual and well-known "weak sustainability" approach, but  $NI$ , i.e. innovation and discoveries in technology, also when not directly addressed to environmental protection, may mean improvement in environmental conditions if they imply increases in efficiency of the production function and therefore allow saving natural resources. Evidently, the model counts on equation (11) and on conditions (12). If the possibility for  $NI$  and  $NC$  to reduce  $\alpha$  is removed from assumptions, admitting the existence of R&D expenditure and/or other expenditures to recover natural capital without reduction of  $\partial W / \partial X$ , then the model will not give same conclusions. This is the basis for a key role for the policy-maker, in terms of R&D expenditure and in terms of exogenous constraints which could be posed, as laws and regulations. For example, in order to warranty a constant effort by all agents, the government could establish a compulsory innovative expenditure per year. In this way a positive growth rate of  $NI$  and might be ensured. Mowery and Rosemberg (1989) confirm that policies voted to encourage R&D expenditure for

---

<sup>12</sup> The model does not investigate about the natural resources management problem. Nonetheless, this matter has been considered in a way: the recycling capacity of nature and the transformation made by the environment of wastes in resources again is undoubtedly a renewable resource. This appears in the model giving the main starting step.

innovation are extremely diffuse in actually all of industrialized countries. The model here hypothesizes that the  $N$  sector is an instrument for the policy-maker; this does not mean implicitly to accept that it must be (partially or exclusively) financed by the public sector. There are some contributions in literature which find successful public intervention in financing R&D as Cohen and Noll (1991) and others which do not emphasize government expenditure, as Jaffe (1998); however a mixed system can easily be considered, where that the government can have a key role to enhance technological research. Policies could be addressed to reduce costs after-tax of R&D expenditure, or to provide subsidies to researchers, firms, and consumer, to respectively induce more innovation, adopt it, and choose innovative products. There is part of literature which successfully deals with these important topics, such as Hall and Van Renssen (2000), Klette *et al.* (2000), and Romer (2000). Moreover, private operators may have strong incentives to pursue innovative investments, firstly because Government could levy higher taxes for polluters and secondly because of competition, image marketing, actual saving in production, as highlighted by Dasgupta and Stiglitz (1980), Spence (1984), Levin and Reiss (1988), just to mention a few. Therefore the main force of the model here presented is not only to replace natural capital with man-made capital, but exactly the Schumpeterian incentive to reduce usage of environment. This will lead to sustainable growth, as stated by the model. Of course, the efficacy of  $NI$  and  $NC$  in reducing  $\alpha$  is very difficult to measure, and depends on different factors. First of all, it depends on the degree of development in which the research is conducted (in this sense, all the model view can vary, being  $\alpha$ ,  $\pi$ ,  $\rho$ , and their functions, possibly dissimilar in different contexts, *see* footnotes: n.9 and n.10); secondly it would probably rely very strongly on the cooperation among researches and on diffusion on innovations. Many authors wrote about economic diffusion of technology, studying its dynamic, its costs, and positive externalities arising from it; here it is very useful just recall the role that policy can have in enhancing also this particularly important process. Government could be the main actor in providing the most efficient patent protection, (giving the possibility to distribute knowledge) and in distributing information in order to reduce obstacles as uncertainty, which endogenously characterizes economic life.

Further research will be conducted in studying explicit functional form for equations of the model. These analyses will allow studying  $\alpha$ ,  $NI$ ,  $NC$ , and all of their determinants; even if data are missing and not always available, econometric estimation could be able to set important results to give actual application to this theoretically meaningful model.

## References

- [1] Anand, S., and Sen, A. 2000. Human Development and Economic Sustainability. *World Development* Vol. 28 (12):2029-2049.
- [2] Arrow, K. *et al.* 1995. Economic Growth, Carrying Capacity and the Environment. *Science* 268: 520-521.
- [3] Azar, C. 2000. Economics and distribution in the greenhouse. *Climatic Change* 47:233-238.
- [4] Barbier, E.B., and Markandya, A. 1990. The Conditions for Achieving Environmentally Sustainable Development. *European Economic Review* 34: 659-669.
- [5] Barry, B. 1991. *Liberty and Justice: Essay in Political Theory* 2. Clarendon Press. Oxford.
- [6] Bartelmus, P. 1994. *Environment, Growth, and Development: The concepts and Strategies of Sustainability*. Routledge. London.
- [7] Beltratti A. 1996. *Models of Economic Growth with Environmental Assets*. Kluwer Academic Publishers. Dordrecht.
- [8] Beltratti, A., Chichilnisky, G., and Heal, G. 2000. Sustainable Growth and the Golden Green Rule. *NBER working paper* n. 4430.
- [9] Biondo, A.E. 2003. *Economic Growth and Sustainable Development*. The University of Manchester.
- [10] Biondo, A.E. 2004. *Growth and Development: the Sustainable Perspective*. WP 3/2004 DEMQ, University of Catania.

- [11] Biondo, A.E. 2006. The Sustainable Perspective between Growth and Development. *Sistemaeconomico* n. ½: 25-45.
- [12] Bovenberg, A.L., and Smulders, S.A. 1995. Environmental Quality and Pollution-augmenting Technological Change in a Two-sector Endogenous Growth Model. *Journal of Public Economics* 57:369-391.
- [13] Brock, W.A. 1977. *A Polluted Golden Age*. Economics of Natural Environment Resources.
- [14] Bromley, D.W. 1989. Entitlements, missing markets, and environmental uncertainty. *Journal of Environmental Economics and Management* 17: 181-194.
- [15] Chambers, N., Simmons, C., and Wackernagel M. 2000. *Sharing Nature's Interest: Using Ecological Footprint as an Indicator of Sustainability*. Earthscan.
- [16] Cohen, L., and Noll, R. 1991. *The Technology Pork Barrel*. Brookings Institution. Washington DC.
- [17] Costantini, V. 2006. Sviluppo sostenibile e sviluppo umano: un approccio integrato. *Rivista di Politica Economica* Vol. 96: 129 – 178.
- [18] Dasgupta, P., and Heal, G.M. 1974. The Optimal Depletion of Exhaustible Resources. *Review of Economic Studies*, Symposium, May, 3-28.
- [19] Dasgupta, P., and Stiglitz, J. 1980. Industry Structure and the Nature of Innovative Activity, *Economic Journal* 90: 266-293.
- [20] Dasgupta, P., and Stiglitz, J. 1980. Uncertainty, Industry Structure and the Speed of R&D, *Bell Journal of Economics* 11:1-28.
- [21] Ekins, P. 2003. Identifying critical natural capital. Conclusions about critical natural capital. *Ecological Economics* 44: 277 -/292.
- [22] English Nature. 1994. Sustainability in Practice. Issue 1: Planning for Environmental Sustainability. *English Nature*, Peterborough, UK.
- [23] Farmer, M.C., and Randall, A. 1997. Policies for Sustainability: lessons from an overlapping generations model. *Land Economics* 73 (4):608-622.
- [24] Field, C.B. 2001. Sharing the Garden. *Science*, 294: 2490-2491.
- [25] Georgescu-Roegen, N. 1976. *Energy and Economic Myths*. Institutional and Analytical Economic Essays. Pergamon Press, New York.
- [26] Goulet D. 1971. *A New Concept on the Theory of Development*. New York Atheneum Press.
- [27] Grossman, G.M. 1995. *Pollution and growth: what do we know?* in: I. Goldin and L. A. Winters (Editors) *The Economics of Sustainable Development*. Cambridge University Press, Cambridge:19-47.
- [28] Hall, B., and van Renssen, J. 2000. How Effective are Fiscal Incentives for R&D? A review of the Evidence, *Research Policy*, 29: 449-469.
- [29] Hamilton, K. 1995. Sustainable Development, the Hartwick rule and Optimal Growth, *Environmental and Resource Economics* n.5: 393-411.
- [30] Hamilton, K. 1996. Pollution and Pollution Abatement in the National Accounts. *Review of Income and Wealth* Vol. 42 (1):13-33.
- [31] Hamilton, K. 2000. Genuine Savings as a Sustainability Indicator. *Environmental Economics Series Paper n. 77*, World Bank, Washington, DC.
- [32] Hamilton, K., Atkinson, G., and Pearce, D. 1998. Genuine Savings as an Indicator of Sustainability, *CSERGE Discussion paper*, London.
- [33] Hamilton, K., and Clemens, M. 1999. Genuine Savings Rates in Developing Countries. *World Bank Economic Review* 13:333-356.



- [34] Hartwick, J.M. 1977. Intergenerational equity and the investing of rents from exhaustible resources. *American Economic Review* 67:972-974.
- [35] Hartwick, J.M. 1978. Substitution among Exhaustible Resources and Intergenerational Equity. *Review of Economic Studies* 45:347-354.
- [36] Howarth, R.B., and Norgaard, R.B. 1992. Environmental Valuation under Sustainable Development, *American Economic Review* 82:473-477.
- [37] Jaffe, A.B. 1998. The Importance of Spillovers in the Policy Mission of the Advanced Technology Program. *Journal of Technology Transfer* Summer 1998: 11-19.
- [38] Jaffe, A.B., Newell, R.G., and Stavins, R.N. 2000. Technological Change and the Environment. *NBER working paper n.7970*, Cambridge, MA, USA.
- [39] Klette, T., Møen J., and Griliches Z. 2000. Do Subsidies to Commercial R&D Reduce Market Failures? *Microeconomic Evaluation Studies, Research Policy* 29: 471-495.
- [40] Kunte, A., Hamilton, K., Dixon, J., and Clemens, M. 1998. *Estimating National Wealth - Methodology and Results*. The World Bank, Washington DC.
- [41] Lele, S.M. 1991. Sustainable Development: a Critical View. *World Development* 19:607-621.
- [42] Levin, R.C., and Reiss P.C. 1998. Cost-Reducing and Demand-Creating R&D with Spillovers. *Rand Journal of Economics* 19: 538-556.
- [43] Lewis, W.A. 1955. *Theory of Economic Growth*. Homewood. Ill. Richard Irwing Inc. 420-423.
- [44] Lopez, R. 1994. The Environment as a Factor of Production: The Effects of Economic Growth and Trade Liberalization'. *Journal of Environmental Economics and Management* 27 (2):163-184.
- [45] Lutz, E. (ed.) 1993. *Toward Improved Accounting for the Environment: an UNSTAT*. World Bank Symposium, World Bank, Washington, DC.
- [46] Mac Donald, D.V., Hanley, N., and Moffatt, I. 1999. Applying the concept of natural capital criticality to regional resource management. *Ecological Economics* 29:73 -87.
- [47] Meadows, D., Meadows, D., Randers, J., and Behrens, W. 1972. *The Limits to Growth*. Universe. New York.
- [48] Mowery, D., and Rosenberg, N. 1989. *Technology and the Pursuit of Economic Growth*. Oxford University Press.
- [49] Muradian, R. 2001. Ecological Thresholds: a survey. *Ecological Economics* 38:7-24.
- [50] Neumayer, E. 1999. *Weak versus Strong Sustainability. Exploring the Limits of Two Opposing Paradigms*. Edward Elgar, Cheltenham.
- [51] Page, T. 1983. *Intergenerational justice as opportunity*. In: MacLean, D., Brown, P.G. (Eds.). Energy and the Future. Rowman and Littlefield. Totowa, NJ.
- [52] Pearce, D., Markandya, A., and Barbier, E.B. 1989. *Blueprint for a Green Economy* Earthscan Publications Ltd. London.
- [53] Pearce, D, and Turner, K.R. 1990. *Economics of Natural Resources and the Environment*. The John Hopkins University Press, Baltimore.
- [54] Pearce, D., and Atkinson, G. 1993. Capital Theory and the Measurement of Sustainable Development: An Indicator of Weak Sustainability. *Ecological Economics* 8:103 -/108.
- [55] Pearce, D., and Warford, J.J. 1993. *World without End: Economics, Environment and Sustainable Development*. Oxford: Oxford University Press.
- [56] Pemberton, M., and Ulph, D. 2001. Measuring Income and Measuring Sustainability. *Scandinavian Journal of Economics* Vol. 103 (1):25-40.

- [57] Pezzey, J.C.V. 1992. *Sustainable Development Concepts: An Economic Analysis*, Environment, Paper n.2, World Bank, Washington D.C.
- [58] Pezzey, J.C.V. 1997. Sustainability Constraints versus “Optimality” versus Intertemporal Concern, and Axioms versus Data. *Land Economics* 73, 448-466.
- [59] Pezzey, J.C.V. 2004. Sustainability Policy and Environmental Policy. *Scandinavian Journal of Economics* 106:339-359.
- [60] Pittel, K. 2002. *Sustainability and Endogenous Growth*. Cheltenham, UK and Northampton, MA, US: Edward Elgar.
- [61] Ranis, G. 2004. *The Evolution of Development Thinking: Theory and Policy*. Yale University.
- [62] Redclift, M. 1992. The meaning of Sustainable Development. *Geoforum* 23(3):395-403.
- [63] Romer, P. 2000. Should the Government Subsidize Supply or Demand for Scientists and Engineers? *NBER working paper* n. 7723.
- [64] Rostow, W.W. 1952. *Process of Economics Growth*. NY, W.W. Norton & Co.
- [65] Sen, A.K. 1982. *The choice of discount rates for social benefit-cost analysis*. In: Lind, R.C. (Ed.). *Discounting for Time and Risk in Energy Policy*. Resources for the Future, Washington DC: 325–352.
- [66] Sen A.K. 1992. *Risorse, valori e sviluppo*. Bollati Boringhieri, Torino.
- [67] Shearer 1961. Concept of Economic Growth. *Kyklos* 4: 499.
- [68] Smulders, S. 2000. *Economic Growth and Environmental Quality*. In H. Folmer and L. Gabel (eds.). *Principles of Environmental Economics*, 2nd ed., Edward Elgar, Cheltenham.
- [69] Solow, R. 1974. Intergenerational Equity and Exhaustible Resources. *Review of Economic Studies* (Symposium Issue) 29-46.
- [70] Solow, R. 1986. On the Intergenerational Allocation of Natural Resources. *Scandinavian Journal of Economics* 88(1):141-149.
- [71] Solow, R. 1992. *An Almost Practical Step toward Sustainability*. Resources for the Future. Washington DC.
- [72] Solow, R. 1993. An Almost Practical Step toward Sustainability. *Resources Policy* September: 162-172.
- [73] Spash, C.L. 1994. Double CO2 and Beyond: Benefits, Costs and Competition. *Ecological Economics* 10:27-36.
- [74] Spence, A.M. 1984. Cost Reduction, Competition and Industry Performance. *Econometrica* 52: 101-121.
- [75] Stiglitz, J.E. 1974. Growth with Exhaustible Natural Resources: Efficient and Optimal Growth Paths, *Review of Economic Studies* 41, Symposium on the Economics of Exhaustible Resources, 123-137.
- [76] Stokey, N.L. 1998. Are There Limits to Growth? *International Economic Review* 39:1-31.
- [77] Tahvonen, O. and Kuuluvainen, J. 1993. Economic Growth, Pollution, and Renewable Resources. *Journal of Environmental Economy and Management* 24: 101-118.
- [78] Turner, R.K. (Ed.). 1993. *Sustainable Environmental Economics and Management. Principles and Practice*. Belhaven Press, London.
- [79] UN. 1992. *Integrated Environmental and Economic Accounting*. United Nations, New York.
- [80] UN. 1993. *Earth Summit: Agenda 21, the United Nations Programme of Action*. United Nations Department of Public Information, New York.

- [81] UN. 1993. Handbook of National Accounting: Integrated Environmental and Economic Accounting. *Studies in Method* (Series F, No. 61), Department for Economic and Social Information and Policy Analysis, Statistical Division, United Nations, New York.
- [82] Van der Bergh, J.C.J.M., and Hofkes, M.W. 1999. *Economic Models of Sustainable Development*. In Van der Bergh, J.C.J.M., (ed.), *Handbook of Environmental and Resource Economics*, Edward Elgar, Cheltenham.
- [83] Vitousek, P.M., Mooney, H.A., Lubchenco, J., and Melillo, J.M. 1997. Human domination of Earth's Ecosystems, *Science* 277: 494-499.
- [84] Wackernagel, M. and Rees, W.E. 1997. Perceptual and structural barriers to investing in natural capital: economics from an ecological footprint perspective. *Ecological Economics* 2: 3-24.
- [85] Weitzman, M.L. 1976. On The Welfare Significance of National Product in a Dynamic Economy. *Quarterly Journal of Economics* 90:156-162.
- [86] Weitzman, M.L. 1997. Sustainability and Technical Progress. *Scandinavian Journal of Economics* 99:1-13.

## REASONING ON EVOLUTION OF CULTURE AND STRUCTURE

Maria **BOBROVA**, Arndt **KÜMPEL**  
West East Institute Berlin, **GERMANY**  
[bobrova@westost.eu](mailto:bobrova@westost.eu), [kuempel@westost.eu](mailto:kuempel@westost.eu)

### **Abstract**

*The purpose of this paper is to work toward developing evolutionary reasoning in the social sciences. Along with that, we argue to overcome the artificial divide of natural and social science for the sake of understanding behaviour. We make the case for an evolutionary and culturally sensitive view on long-surviving institutions and its base - individual behaviour. By taking into consideration the unsatisfying answers in the debate on structure and agency, we emphasize the importance of resonance for evolution and stability. We use case studies to make the point for an evolutionary understanding of institutions and to reflect on institutional path dependency.*

**Keywords:** self-organization, path dependency, institutionalization

**JEL Classification:** A13, Z13, B25

### **1. Introduction**

The course of the paper is as follows: Our starting point is the fundamental question of evolutionary epistemology for the reason and degree of consistence of recognition and real categories. We point out the importance of a comprehensive and evolutionary view that integrates vertical (aggregated levels) and horizontal (different sciences) fields into a consistent frame of logic and marries social with natural sciences. Taking the debates in social science about structure and agency as well as macro- and micro-level focus into account, we proceed with an analysis of successful and unsuccessful examples of institution building in different cultural environments. Following on from that, we discuss the case studies in the light of the introduced interpretive schemes and draw up promising questions for future research. We start our quest at the level of meta-physics due to its principal significance for the consistency of reasoning.<sup>13</sup> And we consider reality as a network-like complex system consisting of open coupled-systems characterized by thresholds, buffers, back-coupling, repair mechanisms, multifunctional and function-changing features as well as optimization strategies within self-organized hierarchies (Vollmer 1995, 23). By introducing the idea of universal evolution (Vollmer 1995, 62), we use evolutionary epistemology<sup>14</sup> as a frame to gain higher consistency in our argumentation across sciences. By doing so, we also argue for a re-naturalized reasoning about human behaviour, culture and institutions.

Among the research questions that precede the process of argumentation are the following. In a modern complex world where different levels of functioning of society co-exist and interact in an intricate way, to what extent is an institution-building an open-end, ad-hoc creation process or a “path-dependant” one? Which factors are responsible for both structurization and stabilisation of institutions and their practices? What is the heuristic value of social performance to a researcher? Are traditional sociological dichotomies able to fully encompass and elucidate the dynamic nature of social behaviour? Is there any key to resolving this categorial tension and reconciling the general with the particular, at least at the theoretical level? Could the consideration of a specific case of cross-institutional interaction give us any answers to the questions raised? What are the social lessons one can learn when analysing such phenomena? Do they teach us to better understand, predict and guide our own evolutionary path?

---

<sup>13</sup> “... metaphysics investigates the most general aspects of reality, it is the discipline to which it falls to supply key presuppositions of the special sciences; which are, therefore, based on – though not derivable from – the underlying metaphysics” (Haack 2005, 5).

<sup>14</sup> Evolutionary Epistemology makes specific “that it subscribes to the idea that cognition is to be understood primarily as a product of biological evolution. What does this mean exactly? Biological evolution is regarded as the precondition of the variety of cognitive, cultural, and social behavior that an organism, group or species can portray. In other words, biological evolution precedes (socio-)cultural (co-)evolution. Conversely, (socio-)cultural (co-)evolution originates as a result of biological evolution.” (Gontier 2006)

## 2. Theoretical Framework

Starting from the idea of coincidence and evolution, C.S. Peirce emphasized the need to have a distinct idea of continuation and proposed the term synechism<sup>15</sup>. Synechism does not only connect processes, it also rejects the idea of duality! And his insistence on agapism (love as a sole moral imperative) can be considered to be complementary to synechism, since it provides individual sense and motivation to embody one's beliefs and ideas in selected individual successors by transferring information and resources (Bauer 2007). The very place for this individual capacity is inextricably linked to the way how human brain has evolved. The leeway, that one has to be able to perceive somebody as a member of a functional community with variable borders of sense, has a neural equivalent that enables one to handle the resulting challenges, i.e. the risk to be cheated, the necessity to find common symbols (e.g. common language and spiritual symbols), and the ability to emphasize (and finally even love) another one. The "biological base of sympathy" (Rizzolatti, and Sinigaglia 2008) was found to be in mirror neurons within a reflexive and self-organizing (active and passive) disposition towards inward- (the observer in the brain) (Singer 2002) and outward processes.

Coupled with the high correlation of group size of primates with their respective cortex volume it suggests a functional relationship between both (Ploog 1997, 235). The maximum group size results from the maximum number of group members, among which social relations still can be maintained by individual contact.

By exploring the biological base of social behaviour we approach the question of the evolution and nature of social norms and what kind of behaviour these norms trigger in individuals. Evidence also suggests that the mechanism of the neural base of behaviour is comparable to the way we communicate and interact with others. It is the "symphony of the living" (Cramer 1998) with resonance as a core principle.<sup>16</sup> There seem to exist behavioural correlations (Singer 2003) of neural synchronizations and therefore good reason to include these findings in an analytical framework for individual and group behaviour (as shown in Figure 1 below). In this picture social structures appear to follow function. And we have to interpret the function of social structures in the light of their evolution, which is after all a "gene-culture co-evolution" (Boyd, and Richerson 2008).

The long debate of the duality of structure and action in social science has neither so far resulted in convincing empirical evidence nor has it increased the explanatory power to support this divide. Based on the dualism of Descartes, it argues along the line of Descartes' "*res cogitans – res extensa*" (Weick 2006), with (social) structure being either mind or matter (Calef 2005). And fatally enough, it creates the so called body-mind problem and impedes a reflexive, consistent evolutionary view!

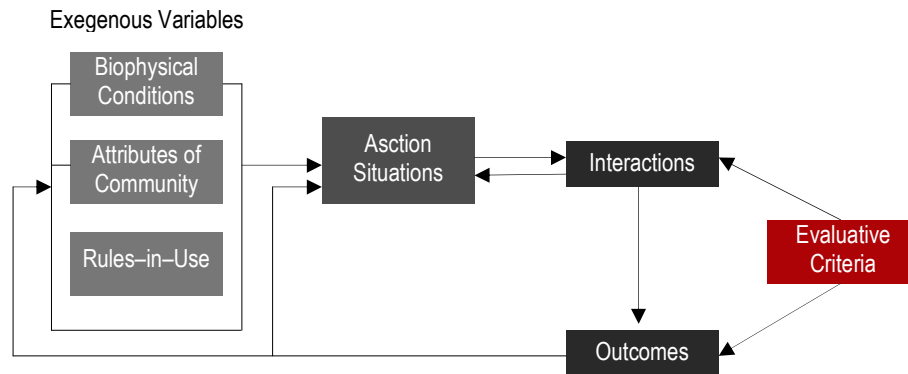
As opposed to this, we assume that a continuous world requires a consistent and comprehensive view that marries all aggregation levels and all dimensions without neglecting the fact that we need a great deal of "*Demut der Wahrnehmung*", understood as "humility of recognition" (von Weizsäcker 1991, 53) in our (non)scientific quest of sense-making.

Firstly, because of the fact that we know much, at the same time our non-knowledge is unlimited. Secondly, we should reflect on our conclusions because we are intertwined with the world in a reflexive way. And, thirdly: we are a product of nature. Therefore the rules that govern us cannot be fundamentally different from the rules which have created us. As Ostrom pointed out recently, community spirit, biophysical conditions and rules-in-use have to be included into a feedback-like framework to explain the permanently changing action situations and the respective social interaction (Ostrom 2005).

---

<sup>15</sup> Synechism as a synthesis of tychism and pragmatism, and "tendency of philosophical thought which insists upon the idea of continuity as of prime importance in philosophy". For more see <http://www.helsinki.fi/science/commens/terms/synechism.html>

<sup>16</sup> Resonance as an integrative and cooperative strategy of the central nervous system can be considered evidenced. (Cramer 1998: 137). If a certain area of neurons gets in resonance with each other, there emerges an innovative operational structure for the solution of a concrete cognitive task. This reassembly by neural synchronization is called *Use-dependent-Plasticity of Assembly Forming Connections* (Singer 1995).



**Figure 1.** A framework for institutional analysis

Endogenous and exogenous biophysical conditions and their respective relative interdependency set the option space for a more or less consistent individual and group behaviour. The degree of consistency is a twofold one: since social systems necessarily start from individuals, the conditions for consistent individual behaviour constitute the social capacity for consistent decision-making and action. And so the (endogenous) consistency within individuals, their energetic and social dynamic equilibrium is a necessary precondition for sustainable consistent collective action. (Relatively) consistent collective action reflects the conscious or unconscious fulfilment of the feedback requirements of individuals for sufficiently confirmed cognition patterns (exogenous consistency). This might be the rules-in-use or, on a less aggregated level, the believable attributes of the community (*see* Figure 1).

As evidence shows, the endogenous consistency itself, however, rests on the dynamic synchronization of specific cortical areas. They strongly suggest that the synchronization of neural responses correlates with cognitive segregation and that neurons, which represent specific features of an object, recognize each other by synchronizing their activity. The chronological coherence of the participating neurons can be considered to be the signature of these neural assemblies (Singer 2006, 37). As studies demonstrate, the frequency of such assembly specific synchronization processes is app. 25 milliseconds (Singer 1999). The time-critical consistency between perceived cognitive stimulus (object) and meta-representation within the brain might constitute a cascade-like mechanism to confirm the neural assembly and therefore the neural stability of the representation. This suggests a neural function of consistency of meta-representations as feedback mechanism.

From this point of view, believability of social symbols and ideas has a more or less stable neural equivalent. The structure of representations here is as follows: the extent, to which certain features of a perceived object match the meta-representation within the brain by object specific neural assemblies, is reflected by the amplitude of the neural responses to the perception. The more features match preferences of neurons to assemble, the higher the amplitude of the neural response to perception (Singer 2003, 315). The cascade of neural representation of cognitive impulses finally results in a more or less confirmed, easier to iterate and therefore more or less believable (convincing) set of self-referential meta-representations. The precisely-timely relations between electrostatic discharge of neurons play an important role within the process of processing cognitive results (Singer 2003, 321). This cognition might be self-generated from the “observer in the brain” with his “inner eye” within a permanent process of reflection, re-reconstruction and reprocessing the results of observation (Singer 2006, 45).

This understanding emphasizes at least two things: Firstly, it reflects the objective need to use our given neural capacity to assemble and resemble neurons in a process of creating meta-representations of the world outside. The inner representations require a sufficient feedback from cognition to be iterated and confirmed in a cascade-like fashion. And, secondly, the activity profile of neural “meta-representation architectures” indicates their function: They enable us to communicate, to draw up a concept of “self” and “others”, both essential constituents for and within a “theory of mind”. We need the “other” to construct the “self”, and this might even justify the assumption that the “others” create, to

a large extent, our “self”, assigning concepts of a “self” the ontological status of social reality (Singer 2006, 49).

Therefore, it is reasonable to argue that communication is a core feature of the emergence of the “theory of mind”, the evolution of the “self” and our relation to the “others”. One of the results of this confirmed relation and its long-term stability might be the community spirit and a sense of belonging as a relatively stable set of repeatedly confirmed meta-representations that correspond to the social capacity of the “self” in “others”.

This reasoning also highlights the meaning of consistent communication for the stability of meta-representations and their function within the construction of culture and stable relations between the “self” and the “others”. And from here it is only a short step towards the discussion on the set of expectations that emerge in the interaction between the “self” and the “others”. These sets of role expectations are institutions! This challenges our understanding of the biological base of social behaviour, our theories about the social function and structure of culture and the explanatory power of values.

In search for an appropriate think tool to represent the dynamic relation between actors and their social constructs, which shape their behaviour and decisions, the “actor centred institutionalism” has proved to be viable (Kümpel 2005), although it is a research heuristics rather than a theory (Mayntz, and Scharpf 1995, 39). It gives an actor a specific analytical value and connects actor, action and concrete institutional settings as framings for actors’ cognition on a rational choice basis. It often gives sufficient explanations for institutional choices in which actors rationally apply reconstructed, institutionally-constituted strategies, and also those in which they do not.

However, discourse can be considered to be the missing link in actor centred institutionalism, (Schmidt 2003). The process of communication as an analytical object seems to have been undervalued. But discourse introduces a reflexive component to the analysis and links decision and behaviour to the results of communication! And even more, communication is an important source of utility in a contemporary economic concept of individual well-being, but so far it still seems to be a blind spot in economic theory.

As Bruno S. Frey has been continuously emphasizing, the process towards certain outcomes can be considered itself a source of utility and, therefore, economic value. This might also, at least partly, explain why humans act with no- or disproportional output. It is likely to be the failure of mainstream economic theory to disregard the procedural utility of processes in general and communication processes in particular. This is even true for individual preferences measured by price, since price acceptance can be traced back to procedural fairness of pricing (Frey, and Pommerehne 1993). Fair procedures contribute to individual happiness (Frey, and Stutzer 2002), since procedural utility might be associated with procedural fairness, as outcome utility might be associated with output fairness (Frey, Benz, and Stutzer 2002, Benz, 2005). As Benz points out with reference to Frey, procedural utility matters a lot for institutional sustainability and has an enormous impact on effectiveness of a wide range of communication- and cascade-like institutionalization processes. The consistency of process with process outcome plays, thereby, an essential role for the stability of the outcome! Benz presents empirical evidence for work and employment, consumption, political participation, the provision of public goods, taxes, organizations and law (Benz 2005, 11).

Process-oriented thinking holds great potential for understanding aspects of (self) organization, such as complexity, stability and change. It provides a think tool on how things become, and therefore a way to understand evolution better. But the contingency of the evolutionary process itself requires not to be too certain about its features. Mainly for this reason we need a think tool that reflects this fundamental insight, i.e. we need a methodology that avoids the pitfalls of induction being aware of the “Black Swan”<sup>17</sup>. The low level of scientific prediction of the breakdown of the so called communist system in the 80’s of the 20<sup>th</sup> century raises the question why one should attribute the label “consistency” to those assumptions and theories which were, in their explanatory power, at best weak, and, in fact, simply proven wrong to reality.

---

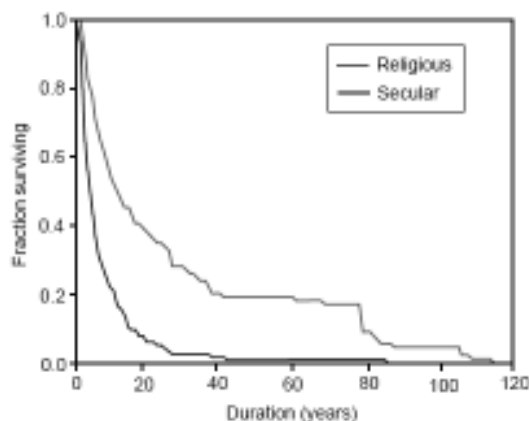
<sup>17</sup> As Socrates, Galilei, Hume, and Albert Einstein all rejected the idea of induction, Karl Popper makes the case with his criticism on induction with his “Black Swan” argument. For the induction problem see: Zalta, E.N. (2010)

The breakdown of Communism seems to be like a “Black Swan”-event for social science and a major blow to the respective theoretical frameworks. Interestingly enough, the financial market crash in 2008 proves the fundamental fragility of knowledge (Talib 2007) and is based on the same “overconfidence bias”, including a major blow to standard economic theory, that generalizes behaviour and therefore excludes major features of humans (Talib 2008). And the common characteristic of both events results from the fact, that both were sudden realizations of the potential created by a misguided understanding of human needs and nature.

However, they eventually share common metaphysical commitments, which might result in the reproduction of the same errors (Weick 2006). And it is noteworthy that economic theory failed to seize the instability of the economic base of the communist bloc and hence any signal for its historically sudden disappearance. Although on a system-logical rather than empirical justification level, it was the system theory of Luhmann that predicted the breakdown of the dysfunctional socialist system, (Luhmann 1984, 11) it could not, however, generate a timeframe for this (von Beyme 1994).

One of the most striking features of the transition in the CEE and the CIS is the recurrence of religion as a base of faith and community spirit. As the Marxist ideology was discredited, many people were in search of faith and spiritual sense of being. They either never had truly lost their faith in religion, or they rediscovered it as a source of relief (Burgess 2009) or still had lost the memory of belonging to church.

Fundamentally, there is a meta-physic demand for relief in situations which put extreme stress on individuals.<sup>18</sup> Since religion provides an independent source of procedural and cognitive utility for a person, it makes one cognitively less sensitive towards economic shocks (Lelkes, 2002). As to be seen in Figure 2, Sosis (Sosis 2000, 77) provides preliminary, but strong evidence for a significant contribution of religion to the sustainability of communities and the respective institutions.



**Figure 2.** Proportion of communes surviving as a function of time since the founding of the commune

With the Polish case we provide evidence for the impact of religious beliefs on social structures and institutions.

### 3. Country Cases

Our concern is the process of structural development and institutionalization in the sphere of cooperation between state and non-state actors, in particular voluntary sector organizations. If one seeks to understand the complex nature of relations between those organizations and the state, one needs to look at historical patterns of development of the social sphere which is typically defined as different from the market and the state, i.e. the civil sphere. The consideration of this sphere shows one whether and how state and non-state actors approach each other, whether and how they seek each other’s support and whether cooperation between them facilitates or undermines the activity of voluntary actors. In this respect it is interesting to compare the situations in the countries of Western and Eastern Europe.

---

<sup>18</sup> Heinsohn points out, that the regular reduction of stress can be reached much more effectively within procedures to celebrate common beliefs. It started with the creation of gods. See: Heinsohn, G. (1997)



The importance of the contextual analysis in studying the third sector has been emphasised by various researchers.<sup>19</sup> Relationships with the state will be expressed in a specific form depending on whether the prevailing institutional culture of the country is that of corporatism or pluralism (Smillie 1994). In a corporatist context, voluntary organisations and governments come into closer contact with each other and jointly discuss important policy issues, as is the case with Scandinavian countries. Whereas in the US, a country with strong pluralist traditions, “fear of socialism and a fundamental mistrust of government action ... encouraged wealthy industrialists ... to establish philanthropic foundations as an alternative to government. ... American voluntarism was based on a rejection of government control and a resistance to consensus.” (Smillie 1994, 161).

The case of Denmark is a typical example of a Scandinavian corporatist context that facilitates formation of healthy institutions and institutionalisation processes. Nordic third sector organisations have historically played an important role in the making of the Scandinavian welfare state (Klausen, and Selle 1996). Governments in these countries were not viewed as the enemy to be opposed to. The state was perceived as a “rational arrangement to solve collective needs as long as it struck the golden mean and allowed private initiatives” (Meldgaard 2001, 15).<sup>20</sup> Considering the Danish case, Meldgaard stresses that “a direction from below and a direction from above have met in a type of society that has created a ‘big government’ and a no smaller civil society.” (Meldgaard 2001, 18). The Danish people have created an enormous law and circular letter machinery, but at the same time personal freedom is a holy cow, and the civil initiative is praised and supported by politicians”.<sup>21</sup> Within this context, high governmental subsidies do not deprive NGOs of their independence and ability to criticise government. Their role of both welfare services providers and democratic institutions is still significant. Situations may slightly vary in some countries. However, as Henriksen (Henriksen 1996) indicates on the basis of the Danish case, voluntary and governmental organizations seem to be co-actors actively engaged in creating new paradigms and practices in social policy. The issue of possible competition between NGOs themselves in the described situation is diminished due to a well-established tradition and capacity to negotiate with the government.

The same question of dependency on governmental funding and, at the same time, independence from the government will have a different answer outside the Scandinavian context. For example, in the US, whose corporatist status is regarded as the lowest among Western countries, high levels of governmental funding have significantly reduced the independence of action and voice (Smith, and Lipsky 1993). A similar situation has been observed in the UK where the governmental practice of contracting and financing NGOs made the latter “ beholden to government, less willing to pioneer new ideas which might endanger their receipt of government grants, less able to stand bulwarks against the concentration of political power, and less capable of serving as schools of citizenship” (Green 2000, 2). Green estimates this outcome as “the unintended result” of the “contract culture” and admits that the British social structure had been weakened by state centralism. Taking the country’s pluralist context into account, however, one could call this the expected consequence of coming too close to the state. Thus, governments could look upon NGOs as convenient and inexpensive delivery mechanisms. A number of organisations, to protect their independence, try to reduce the reliance on the government.<sup>22</sup>

When discussing the case of Eastern European countries one needs to take into account three significant factors. First, all those countries are going through the process of transition and share a set of common “socialist” or “communist” cultural characteristics. Second, historical and geographical settings are unique for each country. Third, Western bilateral and international non-state actors play an essential role in catalyzing the process of voluntary initiatives’ mobilization and their further institutionalization.

---

<sup>19</sup> Italy has practically become a classical example of how historical development of particular regions and welfare politics in general resulted in mutual accommodation with the state. *See*: Putnam, R. D. (1992) and Ranci, C. (1994).

<sup>20</sup> Describing the situation in Denmark, Meldgaard mentions that the state supports a lot of civil voluntary work, and interference mostly takes the form of auditing the accounts. The basis for relations is trust: “The Danes don’t necessarily like each other, but with the state building process they have succeeded to act as if they all belong to the same family” (Meldgaard 2001, 16).

<sup>21</sup> Governmental contribution to social well-being in Denmark is essential. Approximately one third of the national product is spent on social services and close to one-half of all public revenue is used in the social sector (Bjorkoe 2009, 98).

<sup>22</sup> In the US, Oxfam relies almost exclusively on private funding. *See*: Oxfam (2010).

As is the case with Western countries, specific cultural, historical and geopolitical conditions are responsible for how and what institutions will emerge and develop. The state monopoly on the independent articulation of civil concerns, or, in other words, suppression of civil society and its free development by the state was directly or indirectly challenged by the people in Eastern Europe. In each country, for various reasons, the striving for, and the movement of, the civil society towards emancipation from the state takes its original form.<sup>23</sup> This specific content/form affects the relationship of non-state actors with the state.

Talking about Poland, for example, one should bear in mind that the Polish civil society's continuous negotiations with the state administration were a feature of the country's public life during the Soviet rule. It is not surprising that during the earliest stage of transformation of the old regime non-state actors were actively involved in state policy decision-making. Civil society associations, in particular trade unions, had an important impact on "the emerging constitutional order in the country in 1990" (Skąpska 1997, 217). The first law on privatisation, for example, underwent considerable amendments "as a result of negotiations between trade unions, employers associations, and representatives of the state, within the framework of the so called Tripartite Commission" (Skąpska 1997, 217). Poland was among the first former Warsaw-bloc countries to initiate and legitimize the Charter on co-operation of non-state actors with the state, thus emphasizing the necessity of the governmental support for social programmes and independence of non-state actors in pursuing their missions.

At the same time, the role of more traditional institutions such as the Catholic Church has been counterproductive to this cause of independence. Stanosz (1993) points out the pressure the Church has been exercising on the Polish government after the collapse of the Warsaw bloc to ensure that, constitutionally, there is no clear division between the Church and the State. She finds any compromise concerning the relations between the State and the Church very dangerous for democracy in the country.<sup>24</sup> Historically Catholicism played an important role in the process of civil society mobilisation in Poland. The topic of Catholicism united workers and peasants as well as Polish intelligentsia who traditionally associated themselves with the Catholic faith and looked to the Church as the symbol of national unity and, later, anti-communist protest. The Church, for centuries "the sole organisation through which the Polish nation could hope to survive" (Grabska, quoted from a conversation with Potel (1982, 85) had the advantage, as Potel suggests, of "stressing the solid bond that existed in Poland between nationalist sentiment and the religious institutions. It is in this bond, reinforced by the ups and downs of the nationalist struggle in the course of more than a century, that one must look to find one of the deepest reasons for the hold of religion" (Potel 1982, 85).

The Church was successfully used by Solidarity intellectuals and workers as a source of support and a mediator between their interests and those of the state. Under the changed circumstances, when it is possible to relate to the state directly on a legitimate basis, attempts are made by independent organisations to get rid of the formerly helpful mediator that starts dictating the rules of the game to others. Michnik notes in this respect that the Catholic Church has a problem with Polish democracy, and democracy has a problem with the Polish Catholic Church (Michnik 2001). It is not surprising that, on the disintegration of the Warsaw bloc, the Church did not confine its role to a mere religious practising institution, but has remained actively engaged in both social and political life (Millard 1997).

One can compare the situation in Poland to those of other Warsaw bloc countries, in particular the USSR that had 15 different Republics as its constituent states. Under the Soviet regime, e.g. in the Republic of Russia, the institution of the Church, in spite of an official independence of the state, suffered irreparable 'human resources' damage and had to peacefully co-exist with the institutions of the state. Any proselytising activity was forbidden and fraught with dangers of repression. Popular participation in religion ceremonies was considered to be a direct threat to the regime. Contrary to the Polish case, the Church was not looked to and used as an institution that could mediate between the civil society concerns and the interests of the state.

---

<sup>23</sup> For the details of our approach see: Bobrova, M. (2004).

<sup>24</sup> Stanosz admits that in the parliamentary election, for the first time in her life, she voted for the left-wing coalition (coalition of former communists) just because they were the only ones who "declared that there is a necessity to end the Church's domination in Polish public life" (Stanosz 1993).

After the collapse of the USSR, religious institutions enjoyed both authority and trust in society. However, as an institution which had gradually regained power, as we mentioned before when discussing the recurrence of religion, the Church began to directly or indirectly exercise its influence on political life in the country. Thus, during the election campaign in 1996, the Orthodox Church unambiguously supported the acting president despite its declaration of non-preference principles in relation to any state order or political doctrines. Widespread is its participation in the ceremony of consecration of various state institutions (e.g. the White House consecration). Among its growing spheres of influence are the army and educational institutions.<sup>25</sup>

If one turns to other cases of state/church actors' interaction in the USSR and considers, for example the Republic of Armenia where the role of the Church was similar to that in Poland<sup>26</sup>, one encounters yet another unique configuration of cultural and historical factors that shape the process of this interaction. Traditionally non-state social protection activity in Armenia has been associated with the institution of the Church. After the adoption of Christianity in the early 4<sup>th</sup> century Armenian religious representatives committed themselves to the principles of Christian morality and under the head of the Church, Catholicos Nerses the Great, asylums for lonely and disabled people were founded adjacent to churches. Under the Soviet regime, the institution of the Church was not allowed to carry out religious proselytising work among population and enjoyed only a nominal autonomous status.

It can be noted, however, that the Armenian Gregorian Church enjoyed a greater freedom than, e.g. its Orthodox counterpart in Russia (Bourdeaux and Rowe 1980). Historically perceived as crucial for national survival the Church still performed an important function in a nation unification process.<sup>27</sup> Church attendance, even if for a number of people its symbolic meaning was reduced to a mere 'candle lighting ritual' was widely practiced. Having lit a candle and seeing hundreds of other candles burning around one could experience a moving feeling of being part of a bigger unity. Armenian Soviet authorities, though engaged in anti-religious propaganda, viewed the phenomenon likewise and, unlike their counterparts in Russia, did not regard going to church as a direct threat to the regime. The post-Soviet institutional history of the Armenian Apostolic Church and its relation to the state is similar to its Russian counterpart.

When discussing the process of emancipation of the civil society from the state and the role the International and Western voluntary actors play in this process it would be interesting to look at the Yugoslavian case as an example typical of certain sectoral space of the former Warsaw bloc. As has been noted above, non-state actors play a significant role in catalyzing the process of NGO-formation. They have their distinctive influence on their emergence and functioning, on the one hand, and forming the culture of conducting a dialogue and establishing co-operation with the government sector, on the other. One can, however, raise a question of the nature of this process, the future viability of such organizations and whether they will be able to adhere to their original mission, especially when the financial support is no longer ensured by indigenous organizations themselves, in the way which is counterproductive to the original grass-roots practices.

Thus, Large considers the example of the two Yugoslavian organisations 'Suncokret' and Centre for Peace in Osijek (Large 1995) Suncokret was founded in 1992 by students and peace activists in response to the needs of refugees and was characterised by social radicalism and an anti-war attitude. After three years of functioning, however, it transformed into a well-managed, UNHCR-affiliated organisation engaged in rebuffing public opposition to Operations 'Lightning' and 'Storm'. The Osijek Centre for Peace had to take a pause in its activity in 1995 when it realised that the change in its original role and status had been taking place. Commenting on the situation, Large notes that even if "lamenting the loss of a community/volunteer support base, they [the mentioned NGOs] could boast USAID funding and centre members having received training in NGO development and management in the USA, Sweden and Germany" (Large 1995, 122). The further analysis of the situation in former Yugoslavia, the consideration of the role played by organisations like the EU, their involvement in Western NGOs activity, geographical factors in those NGOs placement and their impact on local

---

<sup>25</sup> In mass media and sermons one can hear priests talking about granting the Orthodox Christianity the status of a state religion. For forms of interweaving of politics and religion see: Mchedlov, M. (2000).

<sup>26</sup> Contrary to Poland, however, it was not the forum for anti-communist protest.

<sup>27</sup> Thus, concerned about the national unity and survival agenda, the Armenian Gregorian Church supported the cause of peaceful action in Nagorno-Karabakh conflict which broke out after the collapse of the USSR.

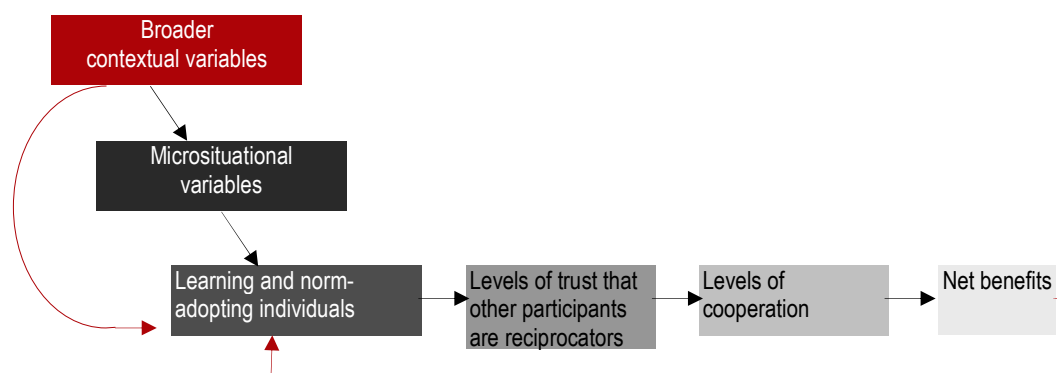
voluntary organisations leads Large to a rather gloomy conclusion: “If there was a ‘voluntary sector’ in ex-Yugoslavia it existed briefly in 1991 and 1992. As the war moved ‘Next Door’... the emphasis shifted on social reconstruction; self-chosen or donor-led” (Large 1995, 124). In this context, the term NGO itself has been perceived to be imbued with negative overtones: “What had been a peace group was now a fully fledged NGO” (Large 1995, 125). Thus the case considered can make one raise the question of possible institutional convergence in the process of state/non-state sector interaction when the latter is aided by well-established foreign NGOs. This does not mean however that the nature of such interaction is considered to be universally malignant.

#### 4. Conclusions

It is important to look at the history and culture of relationships between the state and non-state actors in each country in order to draw adequate conclusions as to whether non-state actors’ independence is influenced by proximity to the state. In the countries where corporatist patterns of relationship prevail, financial assistance from the state does not interfere with their independence in tasks and methods of project implementation. In the pluralist context, financing from the government reduces the independence of voluntary action and is thus counterproductive to healthy institutional performance.

When judging the nature of state and voluntary actors’ interaction in the countries of Eastern Europe, it is important to look at how civil society’s concerns were articulated throughout the history of each country. This perspective allows one to explore how voluntary initiatives are formed and channelled in the process of the emancipation of civil society from the all-inclusive sphere of the state and further affect the interaction with the state. In the process of this interaction such factors as cultural-historical legacy and familiar patterns of institutional behaviour begin to loom large.

Understanding institutional evolution and its diversity requires, therefore, a sound view on the evolution of values, discourse (communication) and institutions. Tabellini (2008) points out that consistence of individual values and “generalized” morality is a feature of well-functioning institutions (Tabellini 2008, 27). The institutional effect of individual values and the trustworthiness which one attributes to another is, hence, complementary, because only consistence of values of many group members leads to a stabilizing group behavior and therefore to stable institutions as a set of role expectations (*see* Figure 2). There is striking evidence that trust is constitutive for the efficiency and effectiveness of institutions (Fehr 2009). However, Fehr highlights the importance of risk and social preferences as well as individual beliefs about other people’s trustworthiness for trusting behavior (Fehr 2009, 262). Complementarily, Poteete et. al. (Poteete, Janssen, and Ostrom 2010) stress the importance of the level of trust for cooperation levels and, in turn, for the net benefits of collective action (*see* Figure 3 below).



**Figure 3.** Microsituational and broader context of social dilemmas affects levels of trust and cooperation

History also shows the relative advantage of religious communities over secular ones.<sup>28</sup>

Luhmann points out that religion reduces complexity and, hence, facilitates the orientation for individual behavior (Luhmann 2000) This is complemented by the joint procedures which create a procedural utility. Additionally, since common religious beliefs create a more or less strong sense of

<sup>28</sup> See Figure 2.

community spirit, group members attribute trustworthiness to other group members easier, thus enhancing mutual cooperation. This in turn creates trust and decreases institutional costs (e.g. control costs as transaction costs). Tabellini even claims a causal effect from shared values and institutional outcomes under certain identifying assumptions (Tabellini 2008).

It shows that the stability of communities heavily influences the path which institutions choose in their evolution. Institutional choice is based on coherent individual behavior, which is mainly driven by an innate capacity to cooperate. This potential is best realized for a common goal of action, since it provides procedural utility and net benefits for individuals and reinforces the social norms which require that behavior. Long surviving institutions obviously accomplish to maintain a critical level of community spirit that enables their members to act constructively based on faith and trust. It also includes a substantial degree of self-reference: The more corporatist an individual is, the more he/she might benefit procedurally from a corporatist mode of action, as pluralists gain from pluralist modes. The “sensitivity under starting conditions” can be considered a major source of path dependency, since it sets the primary marker for the self-reference of the group and their respective institutions.<sup>29</sup>

Along this evolutionary path, declined were those institutions which did not fit the socio-ecological context, i.e. their activity not result in a solution for their context- specific, functionally differentiated task for collective action to handle risk, scarcity and conflict.

However, whatever the specific mode might be, collective action is based on a neural/ genetic capacity that was itself subject to evolutionary selection. Besides, culture evolves much faster than genes and can faster track changing environments (contexts). This is the reason why we see plenty of space for future research here. It could be promising to investigate the co-evolutionary interaction between culture and genes in the context of the evolution of institutions. Another question in our context is what impact migration has on social change and respective social institutions. Finally, since the discourse dimension is largely undervalued in institutional analysis, an investigation into the relation between discourse and path dependency of institutional development would be highly promising.

## References

- [1] Bauer, J. 2007. *Prinzip Menschlichkeit. Warum wir von Natur aus kooperieren*. 5<sup>th</sup> edition, Hamburg: Hoffmann und Campe.
- [2] Benz, M. 2005. The Relevance of Procedural Utility for Economics, *IEW Working Paper 256*, University of Zurich, accessed at: <http://www.iew.uzh.ch/wp/iewwp256.pdf>
- [3] Bjorkoe, J.A. 2009. *Starting from the Heart*, Copenhagen: Kofoed’s School Publications.
- [4] Bobrova, M. 2004. The Issue of Civil Society. Pre-Collapse Practice and Post-Collapse Civil Awareness in Eastern Europe. *Kofoed’s School International Newsletter* No.6, 2004: 17-30, accessed at: [http://www.kofoedsskole.dk/\\_media/8877b68e-faab-4843-aa04-4ad5b4c5209f-Newsletter%206.pdf](http://www.kofoedsskole.dk/_media/8877b68e-faab-4843-aa04-4ad5b4c5209f-Newsletter%206.pdf).
- [5] Boyd, R., and Richerson, P. J. 2008. *Gene-Culture Coevolution and the Evolution of Social Institutions*, In: Engel, C., and Singer, W. (eds.), *Better than Conscious? Decision Making, the Human Mind, and Implications for Institutions*, Cambridge MA: 305. Accessed at: <http://www.sscnet.ucla.edu/anthro/faculty/boyd/BoydRichersonBetterThanConsciousness08.pdf>.
- [6] Burgess, J.P. 2009. *Orthodox resurgence: Civil religion in Russia*. Accessed at: [http://www.georgefox.edu/academics/undergrad/departments/soc-swk/ree/Burgess\\_Orthodox\\_May\\_2009.pdf](http://www.georgefox.edu/academics/undergrad/departments/soc-swk/ree/Burgess_Orthodox_May_2009.pdf).
- [7] Green, D.G. 2000. *Civil Society. The Guiding Philosophy and Research Agenda of the Institute for the Study of Civil Society*, Trowbridge: The Cromwell Press.
- [8] Bourdeaux, M. and Rowe, M. (eds). 1980. *May One Believe - in Russia? Violations of Religious Liberty in the Soviet Union*, London: Darton, Longman and Todd.

---

<sup>29</sup> See for a respective comparative case study on China and Europe: Greif, A. and Tabellini, G. (2009).

- [9] Calef, S. .2005. *Dualism and Mind*, in: Internet Encyclopedia of Philosophy. Accessed at: <http://www.iep.utm.edu/evo-epis/>
- [10] Cramer, F. 1998. *Symphonie des Lebendigen. Versuch einer allgemeinen Resonanz-theorie*, 2<sup>nd</sup> Edition, Frankfurt a. M.: Insel.
- [11] Fehr, E. 2009. On the Economics and Biology of Trust. *Journal of the European Economic Association* April–May 2009 7(2–3):235–266, accessed at: [http://www.iew.uzh.ch/institute/people/fehr/publications/FEHR\\_Economics\\_and\\_Biology\\_of\\_Trust.pdf](http://www.iew.uzh.ch/institute/people/fehr/publications/FEHR_Economics_and_Biology_of_Trust.pdf).
- [12] Frey, B.S., and Pommerehne, W.W. 1993. On the Fairness of Pricing - An Empirical Survey among the General Population. *Journal of Economic Behavior and Organization* 20(3): 295-307.
- [13] Frey, Bruno S., Benz, M., and Stutzer, A. 2002. Introducing Procedural Utility: Not only What, but also How Matter, *IEW Working Paper* 129, University of Zurich, accessed at: <http://www.iew.unizh.ch/wp/iewwp129.pdf>
- [14] Frey, Bruno S., and Stutzer, A. 2002. *Happiness and Economics: How the Economy and Institutions Affect Human Well-Being*. Princeton: Princeton University Press.
- [15] Gontier, N. 2006. *Evolutionary Epistemology*, in: Internet Encyclopedia of Philosophy. accessed at: <http://www.iep.utm.edu/evo-epis/>.
- [16] Greif, A., and Tabellini, G. 2009. *Cultural and Institutional Bifurcation: China and Europe Compared*, accessed at: <ftp://ftp.igier.unibocconi.it/wp/2009/357.pdf>
- [17] Haack, S. 2005. *Not Cynicism, but Synechism: Lessons from Classical Pragmatism*, accessed at: <http://www.as.miami.edu/phi/haack/SYNECHIS.pdf>
- [18] Heinsohn, G. 1997. *Die Erschaffung der Götter. Das Opfer als Ursprung der Religion*. Reinbek bei Hamburg: Rowohlt.
- [19] Henriksen, L.S. 1996. Voluntary organisations and local government: lessons from a Danish case study, *Voluntas* 7(2):177-194.
- [20] Klausen, K.K., and Selle, P. 1996. The third sector in Scandinavia, *Voluntas* (7)2:99-122.
- [21] Kümpel, A. 2005. *On crystallization of democratic consolidation: 74*. accessed at: [http://www.szf.lu.lv/files/petnieciba/publikacijas/04-114\\_Negotiating\\_Futures\\_17.pdf](http://www.szf.lu.lv/files/petnieciba/publikacijas/04-114_Negotiating_Futures_17.pdf)
- [22] Large, J. 1997. *The War Next Door: a study of second track intervention during the war in ex-Yugoslavia*, Stroud: Hawthorn Press.
- [23] Lelkes, O. 2002. *Tasting freedom: Happiness, religion and economic transition*. Accessed at: [http://www.euro.centre.org/data/1196258294\\_82387.pdf](http://www.euro.centre.org/data/1196258294_82387.pdf).
- [24] Luhmann, N. 1984. *Soziale Systeme: Grundriss einer allgemeinen Theorie*. Frankfurt a.M.: Suhrkamp.
- [25] Luhmann, N. 2000. *Die Religion der Gesellschaft*, Frankfurt a.M.: Suhrkamp.
- [26] Mayntz, R., and Scharpf, F.W. 1995. *Der Ansatz des akteurszentrierten Institutionalismus*, in: Mayntz, R.; Scharpf, F. W. (eds.): *Gesellschaftliche Selbstregelung und politische Steuerung*, Frankfurt a. M., New York: Campus: 39-72.
- [27] Mchedlov, M. 2000. *Политика и религия Исторические, правовые и догматические факторы*. accessed at: [http://www.nasledie.ru/politvnt/19\\_13/article.php?art=65](http://www.nasledie.ru/politvnt/19_13/article.php?art=65).
- [28] Meldgaard, O. 2001. State and civil society, *Kofoed's School International Newsletter* No.4: 15-18.
- [29] Michnik, A. 2001. Church and State in Eastern Europe, *East European Constitutional Review*, 7(2), accessed at: <http://www.law.nyu.edu/EECR>.

- [30] Millard, F. 1997. The Influence Of The Catholic Hierarchy In Poland, 1989-96, *Journal of European Social Policy*, 7(2):83-99.
- [31] Ostrom, E. 2005. *Understanding institutional diversity*. Princeton, N.J., in: Ostrom, E.: Beyond Markets and States: Polycentric Governance of Complex Economic Systems. Nobel Price Lecture 2009, p.15, accessed at: [http://nobelprize.org/nobel\\_prizes/economics/laureates/2009/ostrom-lecture-slides.pdf](http://nobelprize.org/nobel_prizes/economics/laureates/2009/ostrom-lecture-slides.pdf).
- [32] Oxfam. 2010. accessed at: <http://www.oxfamamerica.org/whoweare/pages/faq>
- [33] Ploog, D. 1997. *Epilog: Das soziale Gehirn des Menschen*, in: Meier, H.; Ploog, D. (eds.): *Der Mensch und sein Gehirn – Die Folgen der Evolution*, 2nd Edition, München, Zürich: Piper.
- [34] Poteete, A., Janssen, M., and Ostrom, E. 2010. *Working Together: Collective Action, the Commons, and Multiple Methods in Practice*. Princeton, N.J., in: Ostrom, E.: Beyond Markets and States: Polycentric Governance of Complex Economic Systems. Nobel Price Lecture 2009, p.28, accessed at: <http://nobelprize.org/nobelprizes/economics/laureates/2009/ostrom-lecture-slides.pdf>
- [35] Potel, J.-Y. 1982. *The Summe Before The Frost. Solidarity in Poland*, London: Pluto Press.
- [36] Putnam, R.D. 1992. *Making Democracy Work. Civic Traditions in Modern Italy*, Princeton: Princeton University Press.
- [37] Ranci, C. 1994. The third sector in welfare policies in Italy: the contradictions of a protected market, *Voluntas* (5)3:247-271.
- [38] Rizzolatti, G., and Sinigaglia, C. 2008. *Empathie und Spiegelneurone. Die biologische Basis des Mitgeföhls*. Frankfurt a. M.: Suhrkamp.
- [39] Schmidt, V. 2003. *The Boundaries of "Bounded Generalizations": Discourse as the Missing Factor in Actor-Centered Institutionalism*, in: Mayntz, R.; Streeck, W. (eds.): *Die Reformierbarkeit der Demokratie – Innovationen und Blockaden*: 318, Frankfurt a. M., New York: Campus, accessed at: [http://www.mpi-fg-koeln.mpg.de/pu/mpifg\\_book/mpifg\\_bd\\_45.pdf](http://www.mpi-fg-koeln.mpg.de/pu/mpifg_book/mpifg_bd_45.pdf).
- [40] Singer, W. 1995. Development of plasticity of cortical processing architectures, *Science* 270:758-764.
- [41] Singer, W. 1999. Neural synchrony: A versatile code for the definition of relations? *Neuron* 24(1999):49-65.
- [42] Singer, W. 2002. *Der Beobachter im Gehirn. Essays zur Hirnforschung*. Frankfurt a. M.: Suhrkamp.
- [43] Singer, W. 2003. *Die Evolution der Kultur*, in: Fischer, E.P.; Wiegandt, K. (eds.): *Evolution. Geschichte und Zukunft des Lebens*. pp.301, Frankfurt a. M.: Fischer.
- [44] Singer, W. 2006. *Vom Gehirn zum Bewusstsein*, Frankfurt a.M., Suhrkamp.
- [45] Skapska, G. 1997. Learning to be a Citizen: Cognitive and Ethical Aspects of Post-communist Society Transformation, *Democratization*, special issue on *Civil Society: Democratic Perspectives*, 4(1):145-160.
- [46] Smillie, I. 1994. Changing partners: Northern NGOs, Northern governments, *Voluntas*, 5(2):155-192.
- [47] Smith, S., Lipsky, M. 1993. *Nonprofits for Hire: The Welfare State in the Age of Contracting*, Cambridge: Harvard University Press.
- [48] Sosis, R. 2000. Religion and intra-group cooperation: preliminary results of a comparative analysis of utopian communities. *Cross-Cultural Research* Vol.34, No.1:70-87, accessed at: <http://www.anth.uconn.edu/faculty/sosis/publications/Sosiscommunes CCR2000.pdf>.
- [49] Stanosz, B. 1993. Poland: Emerging democracy or religious state? *International Humanist News* 1(4), accessed at: <http://www.iheu.org>.



- [50] Tabellini, G. 2008. Institutions and Culture. *Journal of the European Economic Association* April – May 2008 6(2–3):255–294, accessed at: <http://www.mitpressjournals.org/doi/pdf/10.1162/JEEA.2008.6.2-3.255?cookieSet=1>.
- [51] Talib, N.N. 2007. Black Swans and the Domains of Statistics. *The American Statistician* August 2007, Vol. 61, No. 3:1, accessed at: <http://www.fooledbyrandomness.com/TAS.pdf>.
- [52] Talib, N.N. 2008. *The Black Swan: The Impact of the Highly Improbable*, London: Penguin Books.
- [53] Vollmer, G. 1995. *Biophilosophie*, Stuttgart: Reclam.
- [54] von Beyme, K. 1994. Der Zusammenbruch des Sozialismus und die Folgen für die sozialwissenschaftliche Theorienbildung. *WZB-Mitteilung* 63, März 1994:9, accessed at: <http://www.olaf-eitner.de/QUELLEN/SOWIGEPO/zusammen.htm>.
- [55] von Weizsäcker, C.F. 1991. *Bewußtseinswandel*, München: DTV.
- [56] Weick, E. 2006. Working relationship. A meta-view on structure and agency. *Theory and Science* Vol. 7, Summer 2006, Issue 1, accessed at: <http://theoryandscience.icaap.org/content/vol7.1/weik.html>.
- [57] Zalta, E.N. (ed.). 2010. *The problem of induction*, in: The Stanford Encyclopaedia of Philosophy, accessed at: <http://plato.stanford.edu/entries/induction-problem/>.



# A BEST EVIDENCE SYNTHESIS ON THE LINK BETWEEN BUDGETARY PARTICIPATION AND MANAGERIAL PERFORMANCE

A.B. BONACHE. J. MAURICE

Montpellier Research in Management CREGOR/ERFI Université de Montpellier, France

K. MORIS

LEG-FARGO, Université de Bourgogne, FRANCE

[bonache@rip.ens-cachan.fr](mailto:bonache@rip.ens-cachan.fr)

## Abstract

*Using the best evidence synthesis method (Slavin, 1995), we find out an accurate summary on the link between budgetary participation (BP) and managerial performance (MP). The use of selection criteria allowed to decrease the heterogeneity. Our results report the presence of the heterogeneity by cultural and industrial contingencies. American surveys reveal a significant positive link but suffer from heterogeneity of the sample. Under the sample homogeneity principle, Australian surveys reveal a non-significant negative link and only Taiwanese surveys reveal a positive and significant link. This last result has to encourage researchers to continue the study of publicly traded firms in the Taiwan Stock Exchange to study the causal link between the two variables with a Granger test and to study the evolution of this link over time in other countries.*

**Keywords:** best evidence synthesis, budgetary participation, managerial performance, meta-analysis, subgroup analysis

**JEL Classification:** M49, C89

## 1. Introduction

“A few writers refer to comparing or combining apples and oranges, but the meta-analytic mixtures are usually too heterogeneous to be described with only two fruits. Other writers, with lower levels of enthusiasm or reverence, talk about rotten fruits or even less savoury substances” (Feinstein 1995, 72). Feinstein’s citation is helpful to understand that selectivity is much more attractive than combining heterogeneous papers into a standard meta-analysis that lacks the scientific precautions offered by individual results from randomised trials.

Meta-analysis is a quantitative method of combining the results of independent studies and of synthesizing all the summaries and the conclusions usable to evaluate notably effectiveness of a managerial practice. This type of syntheses differs from traditional reviews of literature using a narrative format to summarise the results of studies on a topic to draw conclusions or inform theory.

In accounting literature, meta-analysis was used to aggregate results in numerous fields (Ahmed, and Curtis 1999, Hay *et al.* 2006, Trotman, and Wood 1991, among others). To the best of our knowledge, two meta-analyses (Derfuss 2009, Greenberg *et al.* 1994) and many reviews of literature were written about the link between budgetary participation and managerial performance (*see for example:* Chalos, and Poon 2001, Shields, and Shields 1998).

*Budgetary participation* (BP hereafter) is usually defined as “a process in which a manager is involved with, and has influence on, the determination of his or her budget” (Shields, and Shields 1998, 49). A *budget* is an expression of company expectations presented in economic terms for a future time period (Samuelson 1973, 31). An usual budgetary participation assessment is Milani’s scale which measures the perceived influence of a “budgetee” on a budget (Milani 1975).

Employee’s *performance* has been defined as “the degree to which successful role achievement is accomplished” (Ferris 1977, 610). The usual *managerial performance* (MP hereafter) questionnaire is based on the results of a survey conducted by Mahoney *et al.* (1963, 1965) that measures eight performance dimensions (planning, investigating, coordinating, evaluating, supervising, staffing, negotiating, representing). This questionnaire provides an overall measure of performance. Mahoney’s and Milani’s scales were used by most of studies investigating the link between BP and MP.

Derfuss (2009) found that BP and MP are significantly and positively linked. Moreover, the link between BP and MP seems contingent on industry differences. Nevertheless, Derfuss’ meta-analysis on this link includes heterogeneous results and only papers published in English language. Following Feinstein (1995), it could be interesting to combine only the quantitative results based on randomised samples.

Our research question is the following one: **is Derfuss' meta-analysis result valid when the only trials based on randomised samples are combined?** By selecting studies with “randomised sample” criterion, we did a “best evidence synthesis” (Slavin 1995).

Best evidence synthesis is “a response to concerns about misleading conclusions from meta-analyses” (Slavin 1995, 11). Following Slavin, if a literature contains some studies high in internal and external validity, thus lower quality studies had to be excluded from the combination of the results. According to Feinstein (1995), studies using randomised samples are more homogeneous and could be aggregated in a meta-analysis.

The rest of the paper is organized as follows. Section 2 describes the data and the used method. Section 3 presents the empirical results. Section 4 discusses them and section 5 concludes.

## **2. Data and method**

### **2.1. Process of studies collection**

Firstly, we collected papers from existing reviews of literature and meta-analyses (Banović 2005, Shields, and Shields 1998, Chalos, and Poon 2001, Derfuss 2009). Then, the first draft was published in the *Muenchen RePec* base of working papers. Thus, our paper appeared in *scholar.google.com* and related papers were collected. These new papers were included in the first draft. This procedure has been iterated until stability of our base of papers on the link between BP and MP.

Finally, we based our synthesis on the list of papers which appears in Appendix. Seventy-six papers were gathered. Some papers have not statistical results and others have unusable results in a meta-analysis because of lack of precisions. Our meta-analysis considers more papers than the one by Derfuss (2009) and our method differs.

### **2.2. Methods and criteria**

To ensure reproducibility of our results, our best evidence synthesis used fixed-effect model of meta-analysis (Hedges, and Olkin 1985). Their statistical procedure is recognised in many scientific fields. The result of our first draft was computed in a spreadsheet. Then, a triangulation of the results was done by using “*rmeta*”: an R package for meta-analysis.

We excluded the following papers from our best evidence synthesis:

- Studies which do not use Milani's measure of budgetary participation (Milani 1975) and Mahoney's one of managerial performance (Mahoney *et al.* 1963, 1965). This exclusion avoids the combination of papers which use different measure scales;
- Laboratory experiments that have low external validity.

After filtering with these two criteria, our base of papers gathered forty-four trials. For the best evidence synthesis, following Feinstein (1995), we used a criterion to exclude the studies which are not based on a randomised sample. Thus, the best evidence synthesis is only based on fifteen randomised trial results coming from the papers summed up in Table 1. Randomised-sample results that do not use Milani's and Mahoney's scales (Breux 2004, Chong *et al.* 2006, Dunk 1995, Kobori 2006) or which presentation of the results is not enough clear to be used (Abdullah 1998, Chong, and Chong 2002) are excluded from the analysis.

Sometimes, the use of selection criteria cannot eliminate heterogeneity between individual studies. If the heterogeneity test rejects the homogeneity null hypothesis, one will use subgroup analysis or one will assess the quality of trials. Nevertheless, evaluation of the methodological quality of a study is a difficult burden (Cho, and Bero 1994). Moreover, the use of quality score is highly criticized in literature (Moher *et al.* 1995 among many others). Thus, subgroup analysis seems to be a better research strategy.

**Table 1.** Papers used for the best evidence synthesis

Randomised-sample papers	Firm description/ Sector	Country	Use of Milani (1975) and Mahoney <i>et al.</i> (1963, 1965) scales
Brownell, and Dunk (1991)	Manufacturing companies	Australia - Sydney	x
Chalos, and Poon (2001)	Listed companies	USA (supposed)	x
Chong, and Bateman (2000)	Manufacturing companies	Australia	x
Dunk (1990)	Manufacturing companies	UK	x
Dunk (1993)	Manufacturing companies	Australia - Sydney	x
Hoque, and Brosnan (2007)	Coal mining companies	Australia	x
Kren (1992)	Manufacturing companies	USA	x
Lau, and Buckland (2000)	Mining companies	Norway	x
Lau, Low, and Eggleton (1995)	Manufacturing companies	Singapore	x
Lau, and Tan (1998)	Financial institutions	Australia/ Singapore	x
Ni, and Su (2001)	Large and Small companies	Taiwan	x
Ni, Su, Chung, and Cheng (2005)	Manufacturing companies	Taiwan	x
Quirin, O'Bryan, and Donnelly (2004)	Large companies	USA	x
Su, and Lin (2007)	Manufacturing companies	Taiwan	x
Subramaniam, and Ashkanasy (2001)	Food manufacturing companies	Australia	x

The presence of cultural contingencies was studied in the literature (Frucot and Shearon 1991, Lau *et al.* 1995, Tsui 2001). The link between BP and MP depends on cultural variables. Thus, if the homogeneity null hypothesis is rejected, it will be useful to make an analysis of differences between cultural subgroups in order to study the causes of the heterogeneity.

The heterogeneity reduction could help to see the impact of other variables on the relationship between BP and MP and to make some recommendations for further research.

### 3. Results

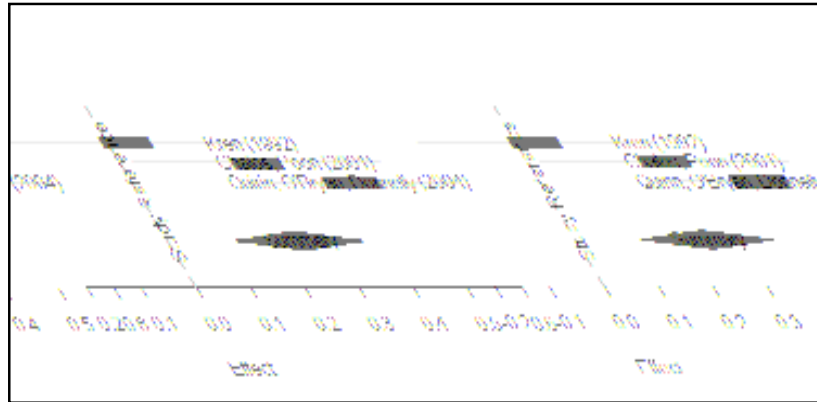
Our final sample is then of 15 results. The summary effect is about -0.0165 (95% confidence interval = [-0.0345, 0.0015]). One cannot rely on this result because of the presence of heterogeneity among results (estimated heterogeneity variance: 0.013, p-value = 0; test for heterogeneity:  $\chi^2(13) = 58.76$ , p-value = 0). The results of these heterogeneity tests has to be compared to the ones without the measurement scale criterion (estimated heterogeneity variance = 0.011, p-value = 0; Test for heterogeneity:  $\chi^2(18) = 67.51$ , p-value = 0). This criterion decreases the heterogeneity, but the homogeneity of the individual results is not significant.

Following Frucot and Shearon (1991), Lau *et al.* (1995) and Tsui (2001), a cultural subgroup analysis was computed. The subgroups are the following: Australian managers (5 randomised-sample results using Milani's and Mahoney *et al.*'s scales), American ones (3) and Taiwanese ones (3). The studies based on survey of managers from other countries were excluded from the subgroup analysis, because of the lack of studies from some political territories.

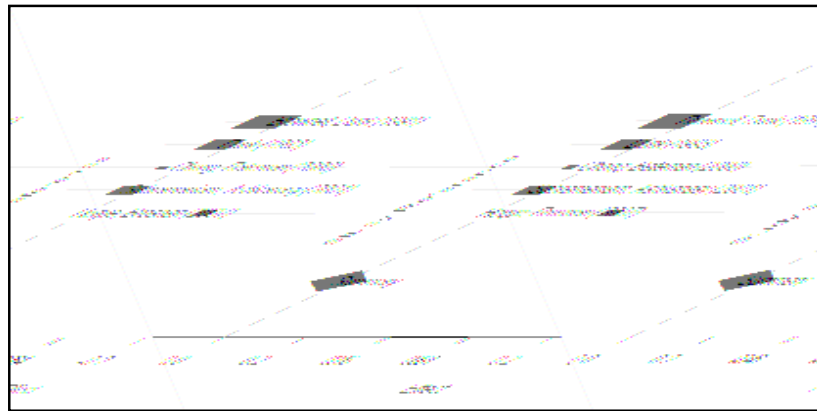
From the Australian subgroup synthesis, it appears a non-significant negative link between BP and MP (summary effect = -0.0184 with 95% CI = [-0.0865, 0.0496]). One can rely on this subgroup synthesis because of non rejection of the homogeneity null hypothesis (estimated heterogeneity variance = 0.0022, p-value = 0.263; test for heterogeneity:  $\chi^2(4) = 5.24$ , p-value = 0.2633). The result seems to be more homogeneous. Hoque and Brosnan's paper increases the heterogeneity (Figure 1): sectoral and industrial contingencies could explain this.

From the American subgroup synthesis, it appears a significant positive link between BP and MP (summary effect = 0.242 , 95% CI = [0.12, 0.364]). One cannot rely on this subgroup best evidence synthesis because of rejection of the homogeneity null hypothesis at the 5% level of significance (estimated heterogeneity variance = 0.025, p-value = 0.044; test for heterogeneity:  $\chi^2(4) = 6.26$ , p-value = 0.0437). The synthesis plot shows a tendency (Figure 2).

But because of the lack of homogeneity between these individual results, one cannot infer something about this temporal tendency.

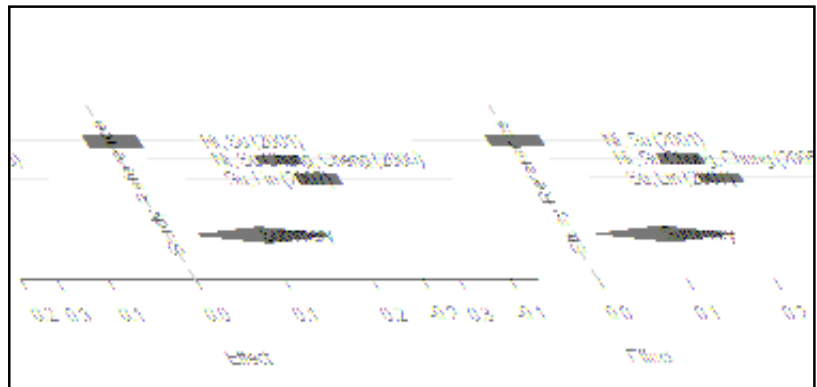


**Figure 1.** Best evidence synthesis of the Australian results



**Figure 2.** Best evidence synthesis of the American results

From the Taiwanese subgroup synthesis, it appears a significant positive link between the studied variables (summary effect = 0.109, 95% CI = [0.0301, 0.188]). One can rely on this subgroup result because of non rejection of the homogeneity null hypothesis at the 5% level of significance (estimated heterogeneity variance: 0.0093, p-value = 0.057; test for heterogeneity:  $\chi^2(2) = 5.72$ , p-value = 0.0571). The synthesis plot shows the same tendency (Figure 3) as the American subgroup one (Figure 2). The relation between BP and MP evolves positively over time.



**Figure 3.** Best evidence synthesis of the Taiwanese results

#### 4. Comments and further research

From these results, it is possible to highlight and to comment three of them. Firstly, the Australian synthesis shows a non-significant negative link between BP and MP that sheds lights on a cultural particularity in Australia: participating in budget determination has no effect on managerial performance. Secondly, the synthesis of the Taiwanese subgroup is really interesting because the sample is the same through studies: publicly traded firms in the Taiwan Stock Exchange. Thus, the time dependency of the link between BP and MP is shown when the effects of cultural and industrial contingencies are under control. Thirdly, Derfuss' results cannot be rejected following our analyses. In fact, in the best evidence synthesis of American subgroup, the lack of control on industrial contingencies could be linked to higher heterogeneity than in the Australian and Taiwanese subgroup.

Based on these comments, it is possible to give a recommendation for further research. The time dependency of the link between BP and MP was inferred from a limited field: the publicly traded firms in Taiwan Stock Exchange from 2001 to 2007. One should examine the robustness of this result:

- in the long-run, by surveying annually traded firms in the Taiwan Stock Exchange during twenty years or more. Then, one will be able to show the evolution over time of the causal link between BP and MP with a Granger test of causality;
- between countries, studying the same firm populations over time (in different political territories).

#### 5. Conclusion

Finally, after having shown that meta-analysis based on the selection of homogeneous individual results is better than "meta-analytic mixtures (...) usually too heterogeneous" (Feinstein 1995, 72), we justified the use of some selection criteria. Moreover, if the combined results are still significantly heterogeneous, it will be justified to combine papers by cultural subgroups.

The best evidence synthesis using "randomised-sample" and "same measurement scales" criteria is heterogeneous. Thus, we have analysed cultural subgroup syntheses. On the base of our subgroup syntheses, it seems that cultural and industrial contingencies are highly plausible. Whereas the Australian subgroup analysis exhibits a negative but non-significant link between BP and MP, the Taiwanese subgroup analysis exhibits a positive and significant one. Both results are based on homogeneous studies. Moreover, the synthesis based on survey of managers of publicly traded firms in Taiwan Stock Exchange from 2001 to 2007, is significantly positive and homogeneous. From this subgroup synthesis, it appears that the link between BP and MP is time dependent.

This time dependency has to be confirmed in further research. One could use the Taiwanese Stock Exchange as a basis to observe the long-run evolution and to test the causal link between BP and MP with Granger's causality test or, if expectations play a role, Sims' one (Granger 1969, Sims 1980). One could replicate the Taiwanese synthesis result by studying this link on the same populations of companies over time in different political territories.

Last but not least, Derfuss' results cannot be rejected following our subgroup analysis. In fact, in the best evidence synthesis of American subgroup, the lack of control on industrial contingencies could be linked to higher heterogeneity than in the Taiwanese and Australian subgroups.

#### References

- [1] Abdullah, S.N. 1998. The role of motivation in the link between budgetary participation and managerial performance. *Analisis* Vol. 5, No 1&2: 1-17.
- [2] Ahmed, K., and Curtis, J.K. 1999. Associations between corporate characteristics and disclosure levels in annual reports: a meta-analysis. *The British Accounting Review* Vol. 31, No 1: 35-61.
- [3] Banović, D. 2005. Evolution and critical evaluation of current budgeting practices. *Master's thesis*, Ljubljana University.
- [4] Breaux, K.T. 2004. The effect of program commitment on the degree of participative congruence and managerial performance in a budgeting setting, *Thesis of the Louisiana State University and Agricultural and Mechanical College*.

- [5] Brownell, P., and Dunk, A.S. 1991. Task uncertainty and its interaction with budgetary participation and budget emphasis: some methodological issues and empirical investigation. *Accounting, Organizations and Society* Vol. 16, No 8:693-703.
- [6] Chalos, P., and Poon, M. 2001. Participative budgeting and performance: A state of the art review and re-analysis. *Advances in Management Accounting* Vol. 10:171-204.
- [7] Cho, M.K., and Bero, L.A. 1994. Instruments for assessing the quality of drug studies published in the medical literature. *Journal of the American Medical Association* Vol. 272:101-104.
- [8] Chong, V.K., and Bateman, D. 2000. The effect of role stress on budgetary participation and job satisfaction-performance linkages: A test of two different models. *Advances in Accounting Behavioral Research* Vol. 3:91-118.
- [9] Chong, V.K., and Chong, K.M. 2002. Budget goal commitment and informational effects of budget participation on performance: A structural equation modelling approach. *Behavioral Research in Accounting* Vol. 14, No 1:65-86.
- [10] Chong, V.K., Eggleton, I.R.C., and Leong, M.K.C. 2006. The Multiple Roles of Participative Budgeting on Job Performance. *Advances in Accounting* Vol. 22:67-95.
- [11] Derfuss, K. 2009. The Relationship of Budgetary Participation and Reliance on Accounting Performance Measures with Individual-Level Consequent Variables: A Meta-Analysis. *European Accounting Review* Vol. 18, No 2:203-239.
- [12] Dunk, A.S. 1989. Budget emphasis, budgetary participation and managerial performance: a note. *Accounting, Organizations and Society* Vol. 14, No 4:321-324.
- [13] Dunk, A.S. 1990. Budgetary participation, agreement on evaluation criteria and managerial performance: a research note. *Accounting, Organizations and Society*, Vol. 15, No 3:171-178.
- [14] Dunk, A.S. 1993. The effect of budget emphasis and information asymmetry on the relation between budgetary participation and slack. *The Accounting Review* Vol. 68:400-410.
- [15] Dunk, A.S. 1995. The joint effect of participative budgeting and managerial interest in innovation on departmental performance. *Scandinavian Journal of Management* Vol. 11, No 1:75-85.
- [16] Feinstein, A.R. 1995. Meta-analysis: statistical alchemy for the 21st century. *Journal of Clinical Epidemiology*, Vol. 48, No 1:71-79.
- [17] Ferris, K.R. 1977. A test of the expectancy theory of motivation in an accounting environment. *The Accounting Review* Vol. 52, No 3:605-615.
- [18] Frucot, V., and Shearon, W.T. 1991. Budgetary participation, locus of control, and Mexican managerial performance and job satisfaction, *The Accounting Review* Vol. 66, No 1:80-99.
- [19] Granger, C.W.J. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* Vol. 37, No 3:424-438.
- [20] Greenberg, P., Greenberg, R.H., and Nouri, H. 1994. Participative budgeting: a meta-analytic examination of methodological moderators. *Journal of Accounting Literature*, Vol. 13:117-141.
- [21] Hay, D.C., Knechel, W.R., and Wong, N. 2006. Audit fees: A meta-analysis of the effect of supply and demand attributes. *Contemporary Accounting Research* Vol. 23, No 1:141-191.
- [22] Hedges, L.V., and Olkin, I. 1985. *Statistical methods for meta-analysis*, Academic Press Orlando.
- [23] Hoque, Z., and Brosnan, P. 2007. Aligning Industrial Relations Risk, Budgetary Participation, and Budgeting Measures of Performance: Impact on Managerial Performance. *SSRN Working paper*.
- [24] Kobori, K. 2006. The effect of budgetary participation on performance: role ambiguity and organizational commitment as an intervening variable, *National Cheng Kung University Department of Business Administration, Master Thesis*.
- [25] Kren, L. 1992. Budgetary participation and managerial performance: The impact of information and environmental volatility. *The Accounting Review* Vol. 67, No 3:511-526.

- [26] Lau, C.M., and Buckland, C. 2000. Budget emphasis, participation, task difficulty and performance: the effect of diversity within culture. *Accounting and Business Research* Vol. 31, No 1:37-55.
- [27] Lau, C.M., Low, L.C., and Eggleton, I.R.C. 1995. The impact of reliance on accounting performance measures on job-related tension and managerial performance: additional evidence. *Accounting, Organizations and Society* Vol. 20, No 5:359-381.
- [28] Lau, C.M., and Tan, J.J. 1998. The impact of budget emphasis, participation and task difficulty on managerial performance: a cross-cultural study of the financial services sector. *Management Accounting Research* Vol. 9:153-183.
- [29] Mahoney, T.A., Jerdee, T.H., and Carroll, S.J. 1963. *Development of Managerial Performance: A Research Approach*, South-Western Publishing, Co., Oxford England.
- [30] Mahoney, T.A., Jerdee, T.H., and Carroll, S.J. 1965. The job(s) of management. *Industrial Relations* Vol. 4, No 2:97-110.
- [31] Milani, K. 1975. The relationship of participation in budget-setting to industrial supervisor performance and Attitudes: a field study. *The Accounting Review* Vol. 50, No 2:274-284.
- [32] Moher, D., Jadad, A.R., Nichol, G., Penman, M., Tugwell, P., and Walsh, S. 1995. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Controlled Clinical Trials* Vol. 16:62-73.
- [33] Ni, F.Y., and Su, C.J. 2001. The Effect of Organizational Size and Motivation on the Effectiveness of Participative Budgeting Systems. *Commerce & Management Quarterly* Vol. 2, No 3:299-318.
- [34] Ni, F.Y., Su, C., Chung, S., and Cheng, K. 2005. The role of trust in supervisor in budgeting system. *The 17<sup>th</sup> Asia-Pacific Conference on International Accounting Issues Wellington, New Zealand, Nov. 22.*
- [35] Quirin, J.J., O'Bryan, D., and Donnelly, D.P. 2004. A nomological framework of budgetary participation and performance: a structural equation analysis approach. *Advances in Management Accounting* Vol. 13:143-165.
- [36] Samuelson, L.A. 1973. Effektiv budgetering:En analys av beslut om budgeteringens utformning, *Ekonomiska Forskningsinstitutet vid Handelshögskolan i Stockholm.*
- [37] Shields, J.F., and Shields, M.D. 1998. Antecedents of participative budgeting. *Accounting, Organizations and Society* Vol. 23, No 1:49-76.
- [38] Sims, C.A. 1980. Macroeconomics and Reality. *Econometrica* Vol. 48, No 1:1-48.
- [39] Slavin, R.E. 1995. Best evidence synthesis: An intelligent alternative to meta-analysis. *Journal of Clinical Epidemiology* Vol. 48, No 1:9-18.
- [40] Su, J.T., and Lin, W.L. 2007. The effect of managerial performance in participative budgetary setting: information asymmetry as an intervening variable, *International Conference on Business and Information, Tokyo, Japan.*
- [41] Subramaniam, N., and Ashkanasy, N.M. 2001. The effect of organisational culture perceptions on the relationship between budgetary participation and managerial job-related outcomes. *Australian Journal of Management* Vol. 26, No 1:35-54.
- [42] Trotman, K.T., and Wood, R. 1991. A meta-analysis of studies on internal control judgments. *Journal of Accounting Research* Vol. 29, No 1:180-192.
- [43] Tsui, J.S.L. 2001. The impact of culture on the relationship between budgetary participation, management accounting systems, and managerial performance An analysis of Chinese and Western managers. *International Journal of Accounting* Vol. 36, No 2:125-146.

**Existing literature on the link between BP and MP (no randomised samples – experiments – no use of Milani and Mahoney’s scales)**

- [44] Agbejule, A., and Saarikoski, L. 2006. The effect of cost management knowledge on the relationship between budgetary participation and managerial performance. *The British Accounting Review* Vol. 38:427-440.
- [45] Alam, B., and Mia, L. 2006. Need for Achievement, Style of Budgeting and Managerial Performance in a Non Government Organization (NGO): Evidence from an Oriental Culture. *International Journal of Business Research* Vol. 6, No 3:35-43.
- [46] Aranya, N. 1990. Budget instrumentality, participation and organizational effectiveness. *Journal of Management Accounting Research* Vol. 2, No 1:61-11.
- [47] Arifin, J. 2007. Pengaruh Karakteristik Gaya Penyusunan Anggaran Terhadap Efisiensi Biaya. *Ka Jian Bisnis Dan Manajeme*, Vol. 9, No 1:23-35.
- [48] Bass, B.M., Leavitt, H.J. 1963. Some Experiments in Planning and Operating. *Management Science* Vol. 9, No 4:574-585.
- [49] Bento A., and White L.F. 2006. Budgeting, performance evaluation, and compensation: a performance management model. *Advances in Management Accounting*, Vol. 51:51-79.
- [50] Brownell, P. 1981. Participation in budgeting, locus of control and organizational effectiveness. *The Accounting Review* Vol. 56, No 4:844-860.
- [51] Brownell, P. 1982a. The Role of Accounting Data in Performance Evaluation, Budgetary Participation, and Organizational Effectiveness. *Journal of Accounting Research*, Vol. 20, No 1:12-27.
- [52] Brownell, P. 1982b. A field study examination of budgetary participation and locus of control. *The Accounting Review* Vol. 57, No 4:766-777.
- [53] Brownell, P. 1983. Leadership style, budgetary participation and managerial behaviour. *Accounting, Organizations and Society* Vol. 8, No 4:307-321.
- [54] Brownell, P. 1985. Budgetary systems and the control of functionally differentiated organizational activities. *Journal of Accounting Research* Vol. 23, No 2:502-512.
- [55] Brownell, P., and Hirst, M. 1986. Reliance on accounting information, budgetary participation, and task uncertainty: tests of a three-way interaction. *Journal of Accounting Research* Vol. 24, No 2:241-249.
- [56] Brownell, P., and McInnes, M. 1986. Budgetary participation, motivation, and managerial performance. *The Accounting Review* Vol. 61, No 4:587-600.
- [57] Brownell, P., and Merchant, K.A. 1990. The budgetary and performance influences of product standardization and manufacturing process automation. *Journal of Accounting Research* Vol. 28, No 2:388-397.
- [58] Chalos, P., and Haka, S. 1989. Participative Budgeting and Managerial Performance. *Decision Sciences* Vol. 20, No 2:334-347.
- [59] Chenhall, R.H., and Brownell, P. 1988. The effect of participative budgeting on job satisfaction and performance: Role ambiguity as an intervening variable. *Accounting, Organizations and Society* Vol. 13, No 3:225-233.
- [60] Cherrington, D.J., and Cherrington, J.O. 1973. Appropriate reinforcement contingencies in the budgeting process. *Journal of Accounting Research* Vol. 11:225-253.
- [61] Eker, M. 2009. The impact of budget participation and management accounting systems on performance of turkish middle level managers. *Akdeniz İ.İ.B.F. Dergisi* Vol. 17:105-126.
- [62] Frucot, V., and White, S. 2006. Managerial levels and the effects of budgetary participation on managers. *Managerial Auditing Journal* Vol. 21, No 1&2:191-206.



- [63] Godener, A., and Fornerino, M. 2009. La participation des managers au contrôle de gestion et leur performance managériale, *Proceeding of 30<sup>th</sup> Association Française de Comptabilité Conférence*, Strasbourg, France.
- [64] Govindarajan, V. 2006. Impact of participation in the budgetary process on managerial attitudes and performance: Universalistic and contingency perspectives. *Decision Sciences* Vol. 17, No 4: 496-516, Blackwell.
- [65] Gul, F.A., Tsui, J.S.L., Kwok, H.Y.L., and Fong, S.C.C. 1995. Decentralisation as a moderating factor in the budgetary participation-performance relationship: some Hong Kong evidence. *Accounting and Business Research* Vol. 25, No 98:107-113.
- [66] Hassel, G.H., and Cunningham, G.M. 1996. Budget Effectiveness in Multinational Corporations: An Empirical Test of the Use of Budget Controls Moderated by Two Dimensions of Budgetary Participation Under High and Low Environmental Dynamism. *Management International Review* Vol. 36, No 3:245-266.
- [67] Hirst, M.K. 1987. Some Further Evidence on the Effects of Budget Use and Budget Participation on Managerial Performance. *Australian Journal of Management* Vol. 12, No 1:49-56.
- [68] Jermias, J., and Setiawan, T. 2008. The moderating effects of hierarchy and control systems on the relationship between budgetary participation and performance. *International Journal of Accounting* Vol. 43, No 3:268-292.
- [69] Kenis, I. 1979. Effects of budgetary goal characteristics on managerial attitudes and performance. *The Accounting Review* Vol. 54, No 4:707-721.
- [70] Lau, C.M., and Lim, E.W. 2002. The effects of procedural justice and evaluative styles on the relationship between budgetary participation and performance. *Advances in Accounting, Incorporating Advances in International Accounting* Vol. 19:139-160.
- [71] Leach-López, M.A., Stammerjohan, W.W., and Lee, K.S. 2009. Budget participation and job performance of South Korean managers mediated by job satisfaction and job relevant information. *Management Research News* Vol. 32, No 3:220-238.
- [72] Leach-López, M.A., Stammerjohan, W.W., and McNair, F.M. 2007. Differences in the role of job-relevant information in the budget participation-performance relationship among US and Mexican managers: a question of culture or communication. *Journal of Management Accounting Research* Vol. 19, No 1:105-136.
- [73] Lindquist, T.M. 1995. Fairness as an Antecedent to Participative Budgeting: Examining the Effects of Distributive Justice and Referent Cognitions on Satisfaction and Performance. *Journal of Management Accounting Research* Vol.7:122-147.
- [74] Merchant, K.A. 1981. The design of the corporate budgeting system: influences on managerial behavior and performance. *The Accounting Review*, Vol 56, No 4:813-829.
- [75] Merchant, K.A. 1984. Influences on departmental budgeting: an empirical examination of a contingency model. *Accounting, Organizations and Society* Vol. 9, No 3/4:291-307.
- [76] Mia, L. 1988. Managerial attitude, motivation and the effectiveness of budget participation. *Accounting, Organizations and Society* Vol. 13, No 5:465-475.
- [77] Mia, L. 1989. The impact of participation in budgeting and job difficulty on managerial performance and work motivation: a research note. *Accounting, Organizations and Society* Vol. 14, No 4:347-357.
- [78] Mia, L., and Patiar, A. 2002. The Interactive Effect of Superior-Subordinate Relationship and Budget Participation on Managerial Performance in the Hotel Industry: An Explanatory Study. *Journal of Hospitality & Tourism Research* Vol. 26, No 3:235-257.
- [79] Ni, F.Y., Su, M., and Su, J.T. 2003. The Impact of Organizational Commitment on The Relationship between Budgetary Participation and Performance: A Cross-Cultural Study, *The 12<sup>th</sup>*

*International Conference on Comparative Management*, College of Management National Sun Yat-Sen University, Kaohsiung, Taiwan.

- [80] Nouri, H., and Parker, R.J. 1998. The relationship between budget participation and job performance: the roles of budget adequacy and organizational commitment. *Accounting, Organizations and Society* Vol. 23, No 5:467-483.
- [81] Orpen, C. 1992. Job Difficulty as a Moderator of the Effect of Budgetary Participation on Employee Performance. *The Journal of Social Psychology* Vol. 132, No 5:695-696.
- [82] Otley, D., Pollanen, R.M. 2000. Budgetary criteria in performance evaluation: a critical appraisal using new evidence. *Accounting, Organizations and Society* Vol. 25, No 4:483-496.
- [83] Parker, R.J., and Kyj, L. 2006. Vertical information sharing in the budgeting process. *Accounting, Organizations and Society* Vol. 31, No 1:27-45.
- [84] Sheely Heath, R., and Brown, J.F. 2007. A re-examination of the effect of job-relevant information on the budgetary participation-job performance relation during an age of employee empowerment. *Journal of Applied Business Research* Vol. 23, No 1:111-124.
- [85] Shields, M.D., Deng, F.J., and Kato, Y. 2000. The design and effects of control systems: tests of direct- and indirect-effects models. *Accounting, Organizations and Society* Vol. 25, No 2:185-202.
- [86] Shields, J.F., and Young, S.M. 1993. Antecedents and consequences of participative budgeting: evidence on the effects of asymmetrical information. *Journal of Management Accounting Research* Vol. 5, No 1:265-280.
- [87] Su, M. 2001. The Influence of Budgetary Participation of the Administrators of Japanese Companies in Taiwan to the Managerial Performance: Organizational Commitment as Intervening Variable. *Master's Thesis, Kaohsiung, Taiwan: National Sun Yat-Sen University*.
- [88] Taylor, D., Abdul-Hamid, F., and Mohd-Sanusi, Z. 2008. The Factors Impacting Managerial Performance in Local Government Authorities within an Islamic Cultural Setting. *Journal of Applied Management Accounting Research* Vol. 6, No 2:73-89.
- [89] Tiller, M. 1983. The dissonance model of participative budgeting: an empirical exploration. *Journal of Accounting Research* Vol. 21, No 2:581-595.
- [90] Tintri, D. 2002. Pengaruh Struktur dan Kultur Organisasional Terhadap Keefektifan Anggaran Partisipasif dalam Peningkatan Kinerja Managerial. *Jurnal Ekonomi & Bisnis* Vol. 2, No 7:59-68.
- [91] Tsamenyi, M., and Mills, J. 2002. Perceived environmental uncertainty, organizational culture, budget participation and managerial performance in Ghana. *Journal of Transnational Management Development* Vol. 8, No 1/2:17-52.
- [92] Wentzel, K. 2002. The influence of fairness perceptions and goal commitment on managers' performance in a budget setting. *Behavioral Research in Accounting* Vol. 14, No 1:247-271.
- [93] Winata, L., and Mia, L. 2005. Information technology and the performance effect of managers' participation in budgeting: evidence from the hotel industry. *International Journal of Hospitality Management* Vol. 4, No 1:21-39.
- [94] Yahya, M.N., Ahmad, N.N.N., and Fatima, A.H. 2008. Budgetary participation and performance: some Malaysian evidence. *International Journal of Public Sector Management* Vol. 21, No 6:658-673
- [95] Yuen, D. 2007. Antecedents of budgetary participation: enhancing employees' job performance. *Managerial Auditing Journal* Vol. 22, No 5:533-548.

## APPENDIX

The literature on the link between BP and MP  
(Some of them are not usable for our analysis as explained)

Studies	Samples	Randomised-sample	Budgetary participation measurement scale	Managerial performance measurement scale
Abdullah, 1998	79 Australian managers	Yes	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Agbejule, and Saarikoski, 2006	83 Finnish managers	No	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Alam, and Mia, 2006	113 Bangladeshi NGO managers	No	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Aranya, 1990	97 Canadian managers	No	Personal scale	Personal scale
Arifin, 2007	44 Indonesian managers	No	Personal scale	Personal scale
Bass, and Leavitt, 1963	3 experiments with 36 managers (USA supposed)	Yes (supposed)	Personal scale	Personal scale
Bento, and White, 2006	64 American managers	No	Chow (1999) adapted from Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Breaux, 2004	197 AICPA members (USA)	Yes	Clinton and Hunton (2001)	Mahoney <i>et al.</i> (1963, 1965)
Brownell, 1981	Experiment with 46 students and 48 managers (USA)	No	Personal scale	Personal scale
Brownell, 1982a	38 American managers	No	Hofstede (1967) and Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Brownell, 1982b	40 American managers	No	Hofstede (1967) and Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Brownell, 1983	46 (supposed Australian) managers	No	Hofstede (1967) and Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Brownell, 1985	61 (supposed Australian) managers	No	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Brownell, and Dunk, 1991	79 Australian managers	Yes	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Brownell, and Hirst, 1986	76 Australian managers	No	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Brownell, and McInnes, 1986	108 (supposed American) managers	No	Hofstede (1967) and Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Brownell, and Merchant, 1990	146 (supposed American) production managers	No	Adapted from Swieriga and Moncur (1975)	Personal scale
Chalos, and Haka, 1989	Experiment with 240 M.B.A. students (USA supposed)	Yes	Personal scale	Personal scale
Chalos, and Poon, 2001	72 (supposed American) marketing managers	Yes	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)

Studies	Samples	Randomised-sample	Budgetary participation measurement scale	Managerial performance measurement scale
Chenhall, and Brownell, 1988	33 (supposed American) managers	No	Milani (1975)	Personal scale
Cherrington, and Cherrington, 1973	Experiment with 230 business students (USA supposed)	No	Personal scale	Personal scale
Chong, and Bateman 2000	79 Australian managers	Yes	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Chong, and Chong, 2002	79 Australian managers	Yes	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Chong, Eggleton, and Leong, 2006	74 Australian managers	Yes	Adapted from Milani (1975)	Personal scale, consistent with Merchant (1981)
Dunk, 1989	26 managers from North UK	Yes	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Dunk, 1990	26 managers from North UK	Yes	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Dunk, 1993	78 American managers	Yes	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Dunk, 1995	78 Australian managers	Yes	Milani (1975)	Adapted from Merchant (1981, 1984) and from Brownell and Merchant (1990)
Eker, 2009	150 Turkish managers	No	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Frucot, and Shearon, 1991	83 Mexican managers	No	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Frucot, and White, 2006	178 American managers	No	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965) and Heneman (1974)
Godener, and Fornerino, 2009	155 French managers	No	Adapted from Godener and Fornerino (2005)	Adapted from Govindarajan and Gupta (1985)
Govindarajan 1986	77 (supposed American) managers	No	Swieringa and Moncur (1974)	Mahoney <i>et al.</i> (1963, 1965)
Gul, Tsui, Kwok, and Fong, 1995	37 managers from Hong Kong	No	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Hassel, and Cunningham, 1996	36 Finnish managers and 31 foreign managers	No	Adapted from Milani (1975)	Adapted from Govindarajan (1984) and from Gupta and Govindarajan (1984)
Hirst, 1987	44 Australian managers	No	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Hoque, and Brosnan, 2007	55 Australian managers	Yes	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Jermias, and Setiawan, 2008	204 Indonesian public managers	No	Adapted from Vroom and Mann (1960)	Adapted from Mahoney <i>et al.</i> (1963, 1965)
Kenis, 1979	169 American	No	Personal scale	Personal scale

Studies	Samples	Randomised-sample	Budgetary participation measurement scale	Managerial performance measurement scale
	managers			
Kobori, 2006	81 Taiwanese managers	Yes	Hofstede (1967)	Mahoney <i>et al.</i> (1963, 1965)
Kren, 1992	80 American managers	Yes	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Lau, and Buckland, 2000	71 Norwegian managers	Yes	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Lau and Lim, 2002	83 Australian managers	No	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Lau, Low, and Eggleton, 1995	112 Singaporean managers	Yes	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Lau, and Tan, 1998	104 Australian managers and 85 Singaporean managers	Yes	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Leach-López, Stammerjohan, and Lee, 2009	71 South Korean managers	No	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Leach-López, Stammerjohan, and McNair, 2007	45 Mexican and 98 American managers	No	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Lindquist, 1995	Experiment with 86 students (USA supposed)	Yes	Personal scale	Personal scale
Merchant, 1981	19 American companies	No	Personal scale	Personal scale
Merchant, 1984	170 American managers	No	Personal scale	Personal scale
Mia, 1988	51 Australian managers	No	Milani (1975)	Personal scale
Mia, 1989	62 New Zealander managers	No	Milani (1975)	Personal scale
Mia, and Patiar, 2002	52 Australian managers	No	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Milani, 1975	81 (supposed American) foremen	No	Milani (1975)	Personal scale
Ni, and Su, 2001	205 Taiwanese managers	Yes	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Ni, Su, Chung, and Cheng, 2005	155 Taiwanese managers	Yes	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Nouri, and Parker, 1998	135 American managers	No	Milani (1975)	Govindarajan and Gupta (1985)
Orpen, 1992	136 Australian managers	No	Milani (1975)	Personal scale
Otley, and Pollanen, 2000	121 Canadian managers	No	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Parker, and Kyj,	70 (supposed	No	Milani (1975)	Mahoney <i>et al.</i>

Studies	Samples	Randomised-sample	Budgetary participation measurement scale	Managerial performance measurement scale
2006	American) managers			(1963, 1965)
Quirin, O'Bryan, and Donnelly 2004	98 American managers	Yes	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Sheely Heath, and Brown, 2007	256 American employees	Yes	Milani (1975)	Fraser (1995)
Shields, Deng, and Kato, 2000	358 Japanese engineers	No	Adapted from Shields and Young (1993)	Personal scale
Shields, and Young 1993	98 American corporate controllers	No	Personal scale	Personal scale
Su, 2001 (published in Ni, Su, and Su, 2003)	194 Japanese and Taiwanese managers in Taiwan	Yes	Hofstede (1967)	Mahoney <i>et al.</i> (1963, 1965)
Su, and Lin, 2007	168 Taiwanese managers	Yes	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Subramaniam and Ashkanasy, 2001	114 Australian managers	Yes	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Taylor, Abdul-Hamid, and Mohd-Sanusi, 2008	81 Malaysian managers from a local public administration	No	Milani (1975)	Adapted from Mahoney <i>et al.</i> (1963, 1965)
Tiller, 1983	Experiment with 150 students in psychology (USA supposed)	Yes	Personal scale	Personal scale
Tintri, 2002	53 (supposed Indonesian) managers	No	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Tsamenyi, and Mills, 2002	89 Ghanaian managers	No	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Tsui, 2001	89 managers from Hong Kong	No	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Wentzel, 2002	74 (supposed American) managers in a large hospital	No	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Winata, and Mia, 2005	74 Australian managers	No	Adapted from Milani (1975)	Personal scale
Yahya, Ahmad, and Fatima, 2008	111 Malaysian managers from the Ministry of Defence	No	Milani (1975)	Mahoney <i>et al.</i> (1963, 1965)
Yuen, 2007	216 Chinese public managers	No	Milani (1975) (supposed)	Mahoney <i>et al.</i> (1963, 1965)

# FORECASTS WITH SINGLE - EQUATION MARKOV - SWITCHING MODEL: AN APPLICATION TO THE GROSS DOMESTIC PRODUCT OF LATVIA

Ginters BUSS

Riga Technical University, LATVIA

Faculty of Computer Science and Information Technology

[ginters.buss@rtu.lv](mailto:ginters.buss@rtu.lv)**Abstract**

*The paper compares one-period ahead forecasting performance of linear vector-autoregressive (VAR) models and single-equation Markov-switching (MS) models for two cases: when leading information is available and when it is not. The results show that single-equation MS models tend to perform slightly better than linear VAR models when no leading information is available. However, if reliable leading information is available, single-equation MS models tend to give somewhat less precise forecasts than linear VAR models.*

**Keywords:** Markov-switching, VAR, forecasting, leading information

**JEL Classification:** C32, C53

**1. Introduction**

To the best of our knowledge, there is no publicly available paper that would discuss the forecasting performance of Markov-switching models used to forecast Latvia's gross domestic product (GDP). Thus, this paper is the first attempt to fill the gap in the forecasting literature by comparing one-period ahead forecasting performance of single-equation Markov-switching (MS) and linear vector-autoregressive (VAR) models using Latvia's macroeconomic data.

We use MS models developed in Hamilton (1989) and in a series of papers by Krolzig (*see* Krolzig 1998, 2000, 2003, among others), written in a GAUSS code by Bellone (2005), and adapted to Scilab environment by Dubois and Michaux (2010) in econometrics toolbox Grocer.

The paper is organized as follows. Section 2 describes the model and its estimation. Section 3 presents the results for the one-period ahead forecasting performance of single-equation MS and linear VAR models expressed in terms of root mean squared forecast error (RMSFE) for both cases when leading information is available and when it is not available. Following the results in Buss (2009) that an extra regular differencing of the data might improve the forecasting precision during the switch of the business cycle phases, we also show results for the case when the data are subject to two regular differences, instead of one. Finally, Section 4 concludes.

**2. Methodology****2.1. The Model**

Consider a single-equation Markov-switching model whose parameters are, at least partly, unconditionally time-varying but constant when conditioned on an unobservable discrete regime variable  $s_t \in \{1, \dots, M\}$ :

$$y_t = x_t \beta_{s_t} + z_t \delta + u_t$$

$$u_t | s_t : N(0, \sigma_{s_t}^2), \tag{1}$$

where  $y_t$  is a scalar dependent variable at time  $t$ ,  $x_t = (x_{1t}, \dots, x_{nt})'$  is an  $(n \times 1)$  vector of exogenous regressors at time  $t$  subject to switching regimes,  $z_t = (z_{1t}, \dots, z_{qt})'$  is a  $(q \times 1)$  vector of exogenous regressors at time  $t$  that are not subject to switching regimes,  $\delta$  is a  $(q \times 1)$  vector of regression coefficients which are regime independent, and  $u_t$  is a Gaussian error term subject to regime changes.

Parameter shift functions  $\beta_{s_t}$  and  $\sigma_{s_t}^2$  describe the dependence of the model's parameters  $\beta$  and  $\sigma^2$  on the regime variable  $s_t$ :

$$\beta_{s_t} = \begin{cases} \beta_1 & \text{if } s_t = 1, \\ \vdots & \\ \beta_M & \text{if } s_t = M, \end{cases} \quad (2)$$

and

$$\sigma_{s_t}^2 = \begin{cases} \sigma_1^2 & \text{if } s_t = 1, \\ \vdots & \\ \sigma_M^2 & \text{if } s_t = M. \end{cases} \quad (3)$$

Let  $I_{t-1} := (y_{t-1}, \dots, y_1, x_{t-1}', \dots, x_1', z_{t-1}', \dots, z_1)'$  be the information set at  $t-1$ . The unobservable realizations of the regime  $s_t \in \{1, \dots, M\}$  are generated by an M-state irreducible ergodic Markov stochastic process that is independent of  $I_{t-1}$  or current exogenous variables,  $x_t$  and  $z_t$ , and is defined by its transition probabilities:

$$\begin{aligned} p_{ij} &= P(s_t = j | s_{t-1} = i, s_{t-2} = k, \dots, x_t, z_t, I_{t-1}) \\ &= P(s_t = j | s_{t-1} = i), \sum_{j=1}^M p_{ij} = 1 \text{ for all } i, j \in \{1, \dots, M\}, \end{aligned} \quad (4)$$

that can conveniently be represented with a transition matrix

$$\mathbf{P} = \begin{bmatrix} p_{11} & \cdots & p_{M1} \\ p_{12} & \cdots & p_{M2} \\ \vdots & \vdots & \vdots \\ p_{1M} & \cdots & p_{MM} \end{bmatrix}. \quad (5)$$

Model (1) can be reduced to such special cases as

- the mean-variance model:

$$y_t = \beta_{s_t} + u_t,$$

- the Markov-switching autoregressive (MS-AR) regime-dependent model:

$$y_t = v_{s_t} + y_{t-1}\beta_{1,s_t} + \cdots + y_{t-n}\beta_{n,s_t} + u_t = (1, y_{t-1}, \dots, y_{t-n})\beta_{s_t} + u_t,$$

- the MS-AR intercept regime-dependent model:

$$y_t = v_{s_t} + y_{t-1}\delta_1 + \cdots + y_{t-q}\delta_q + u_t = \beta_{s_t} + (y_{t-1}, \dots, y_{t-q})\delta + u_t.$$

with  $E[u_t^2 | s_t] = \sigma_{s_t}^2$  or  $E[u_t^2 | s_t] = \sigma^2$  for any of the above models.

## 2.2. Estimation

The model is estimated by the maximum likelihood in the following steps:



1. set the initial values of parameters and estimate the model recursively with the Expectation-Maximization (EM) algorithm;
2. compute smoothed probabilities;
3. forecast observed variables.

The rest of the section gives a more detailed look on the procedure, see also Ch. 22 in Hamilton (1994). Denote  $\xi_t$  a random  $(M \times 1)$  vector whose  $j$ th element is equal to unity if  $s_t = j$  and zero otherwise:

$$\xi_t = \begin{cases} (1, 0, 0, \dots, 0)' & \text{if } s_t = 1, \\ \vdots & \\ (0, 0, 0, \dots, 1) & \text{if } s_t = M. \end{cases} \quad (6)$$

If  $s_t = i$ , then the  $j$ th element of  $\xi_{t+1}$  is a random variable that takes on the value unity with probability  $p_{ij}$  and zero otherwise. Thus, the conditional expectation of  $\xi_{t+1}$  given  $s_t = i$  is:

$$E(\xi_{t+1} | s_t = i) = \begin{bmatrix} p_{i1} \\ \vdots \\ p_{iM} \end{bmatrix}, \quad (7)$$

which implies:

$$E(\xi_{t+1} | \xi_t) = \mathbf{P} \xi_t$$

and from Markov property

$$E(\xi_{t+1} | \xi_t, \xi_{t-1}, \dots) = \mathbf{P} \xi_t, \quad (8)$$

which implies:

$$\xi_{t+1} = \mathbf{P} \xi_t + v_{t+1}, \quad (9)$$

Where:

$$v_{t+1} := \xi_{t+1} - E(\xi_{t+1} | \xi_t, \xi_{t-1}, \dots) \quad (10)$$

is a martingale difference sequence. Let  $\mu_{ij} := E[y_t | s_t = j, x_t, z_t, \mathbf{I}_{t-1}] = x_t' \beta_j + z_t' \delta$  be the conditional expectation of  $y_t$ , and let  $\alpha := (\beta_j', \delta', \sigma_j^2)'$  be the vector of parameters characterizing the conditional density. Then the conditional probability density function is:

$$f(y_t | s_t = j, x_t, z_t, \mathbf{I}_{t-1}, \alpha) = (2\pi)^{-\frac{1}{2}} \sigma_j^{-1} \exp\left(-\frac{(y_t - \mu_{ij})^2}{2\sigma_j^2}\right). \quad (11)$$

If there are  $M$  different regimes, then there are  $M$  different densities represented by (11), that are collected in an  $(M \times 1)$  vector

$$\eta_t = \begin{bmatrix} f(y_t | s_t = 1, x_t, z_t, \mathbf{I}_{t-1}, \alpha) \\ \vdots \\ f(y_t | s_t = M, x_t, z_t, \mathbf{I}_{t-1}, \alpha) \end{bmatrix}. \quad (12)$$

Collect  $\alpha$  and the transition probabilities  $p_{ij}$  governing (11) in a vector of parameters  $\theta$ . Let  $P(s_t = j | \mathbf{I}_t, \theta)$  denote the analyst's inference about the value of  $s_t$  based on data obtained through

date  $t$  and based on knowledge of the population parameters  $\theta$ . This inference takes the form of a conditional probability that the analyst assigns to the possibility that the  $t$ th observation was generated by regime  $j$ . Collect these conditional probabilities  $P(s_t = j | I_t, \theta)$  for  $j = 1, 2, \dots, M$  in an  $(M \times 1)$  vector  $\hat{\xi}_{t|t}$ . Collect the forecasts of such probabilities at  $t+1$  given observations obtained through date  $t$ ,  $P(s_{t+1} = j | I_t, \theta)$ , in an  $(M \times 1)$  vector  $\hat{\xi}_{t+1|t}$ .

Note that it is assumed that  $x_t$  and  $z_t$  are exogenous, that is,  $x_t$  and  $z_t$  contain no information about  $s_t$  beyond that contained in  $I_{t-1}$ . Hence, the  $j$ th element of  $\hat{\xi}_{t|t-1}$  could also be described as  $P(s_t = j | x_t, z_t, I_{t-1}, \theta)$ . The  $j$ th element of the  $(M \times 1)$  vector  $(\hat{\xi}_{t|t-1} \mathbf{e} \eta_t)$  where  $\mathbf{e}$  denotes the element-by-element product, can be interpreted as the conditional joint density-distribution of  $y_t$  and  $s_t$ :

$$\begin{aligned} & P(s_t = j | x_t, z_t, I_{t-1}, \theta) \times f(y_t | s_t = j, x_t, z_t, I_{t-1}, \theta) \\ &= p(y_t, s_t = j | x_t, z_t, I_{t-1}, \theta). \end{aligned} \quad (13)$$

The density of  $y_t$  conditioned on past observables is the sum of the  $M$  magnitudes in (13) for  $j = 1, 2, \dots, M$ . This sum can be written in vector notation as:

$$f(y_t | x_t, z_t, I_{t-1}, \theta) = \mathbf{1}'(\hat{\xi}_{t|t-1} \mathbf{e} \eta_t), \quad (14)$$

where  $\mathbf{1}$  is an  $(M \times 1)$  vector of 1s. If the joint density-distribution in (13) is divided by the density of  $y_t$  in (14), the result is the conditional distribution of  $s_t$ :

$$\begin{aligned} & \frac{p(y_t, s_t = j | x_t, z_t, I_{t-1}, \theta)}{f(y_t | x_t, z_t, I_{t-1}, \theta)} = P(s_t = j | y_t, x_t, z_t, I_{t-1}, \theta) \\ &= P(s_t = j | I_t, \theta). \end{aligned}$$

Hence, from (14),

$$P(s_t = j | I_t, \theta) = \frac{p(y_t, s_t = j | x_t, z_t, I_{t-1}, \theta)}{\mathbf{1}'(\hat{\xi}_{t|t-1} \mathbf{e} \eta_t)}. \quad (15)$$

But recall from (13) that the numerator in the expression on the right side of (15) is the  $j$ th element of the vector  $(\hat{\xi}_{t|t-1} \mathbf{e} \eta_t)$  while the left side of (15) is the  $j$ th element of the vector  $\hat{\xi}_{t|t}$ . Thus, collecting the equations in (15) for  $j = 1, 2, \dots, M$  into an  $(M \times 1)$  vector produces:

$$\hat{\xi}_{t|t} = \frac{(\hat{\xi}_{t|t-1} \mathbf{e} \eta_t)}{\mathbf{1}'(\hat{\xi}_{t|t-1} \mathbf{e} \eta_t)}. \quad (16)$$

To obtain  $\hat{\xi}_{t+1|t}$ , take expectations of (9) conditional on  $I_t$ :

$$E(\xi_{t+1} | I_t) = \mathbf{P}E(\xi_t | I_t) + E(v_{t+1} | I_t). \quad (17)$$

Since  $v_{t+1}$  is a martingale difference sequence with respect to  $I_t$ , (17) becomes:

$$\hat{\xi}_{t+1|t} = \mathbf{P}\hat{\xi}_{t|t}. \quad (18)$$

The optimal inference and forecast for each date  $t$  in the sample are found by iterating on the pair of equations (16) and (18). Given a starting value  $\hat{\xi}_{1|0}$  and an assumed value for the population parameter vector  $\theta$ , one can iterate on (16) and (18) for  $t = 1, 2, \dots, T$  to calculate the values of  $\hat{\xi}_{t|t}$  and  $\hat{\xi}_{t+1|t}$  for each date  $t$  in the sample. The log likelihood function  $L(\theta)$  for the observed data  $I_T$  evaluated at the value of  $\theta$  that is used to perform the iterations is calculated as a by-product of this algorithm from:

$$L(\theta) = \sum_{t=1}^T \log f(y_t | x_t, z_t, I_{t-1}, \theta), \quad (19)$$

Where:

$$f(y_t | x_t, z_t, I_{t-1}, \theta) = \mathbf{1}'(\hat{\xi}_{t|t-1} \mathbf{e} \eta_t)$$

and maximized numerically with respect to  $\theta$ . When the model is estimated, the smoothed probabilities can be calculated. Let  $\hat{\xi}_{t|\tau}$  represent the  $M \times 1$  vector whose  $j$  the element is  $P(s_t = j | I_\tau, \theta)$ . For  $t < \tau$  it represents the smoothed inference about the regime the process was in at date  $t$  based on data obtained on through some later date  $\tau$ . Smoothed inferences are calculated using an algorithm developed by Kim (1993). In vector form, this algorithm can be written as:

$$\hat{\xi}_{t|T} = \hat{\xi}_{t|t} \mathbf{e} \{ \mathbf{P}'[\hat{\xi}_{t+1|T} \% \hat{\xi}_{t+1|t}] \}, \quad (20)$$

where the sign  $\%$  denotes element-by-element division. The smoothed probabilities  $\hat{\xi}_{t|T}$  are found by iterating on (20) backward for  $t = T-1, T-2, \dots, 1$ . This iteration is started with  $\hat{\xi}_{T|T}$ , which is obtained from (16) for  $t = T$ . This algorithm is valid only when  $s_t$  follows a first-order Markov chain, when the conditional density (11) depends on  $s_t, s_{t-1}, \dots$  only through the current state  $s_t$ , and when  $x_t$  and  $z_t$ , the vectors of explanatory variables other than the lagged values of  $y_t$ , are strictly exogenous, meaning that  $x_t$  and  $z_t$  are independent of  $s_\tau$  for all  $t$  and  $\tau$  (see Appendix 22.A in Hamilton 1994). From the conditional density (11) it is straightforward to forecast  $y_{t+1}$  conditional on knowing  $I_t, x_{t+1}, z_{t+1}$  and  $s_{t+1}$ :

$$E(y_{t+1} | s_{t+1} = j, x_{t+1}, z_{t+1}, I_t, \theta) = x_{t+1}'\beta_j + z_{t+1}'\delta. \quad (21)$$

There are  $M$  different conditional forecasts associated with the  $M$  possible values for  $s_{t+1}$ . Note that the unconditional forecast based on actual observable variables is related to these conditional forecasts by:

$$\begin{aligned} & E(y_{t+1} | x_{t+1}, z_{t+1}, I_t, \theta) \\ &= \int y_{t+1} f(y_{t+1} | x_{t+1}, z_{t+1}, I_t, \theta) dy_{t+1} \\ &= \int y_{t+1} \left\{ \sum_{j=1}^M p(y_{t+1}, s_{t+1} = j | x_{t+1}, z_{t+1}, I_t, \theta) \right\} dy_{t+1} \end{aligned}$$

$$\begin{aligned}
 &= \int y_{t+1} \left\{ \sum_{j=1}^M [f(y_{t+1} | s_{t+1} = j, x_{t+1}, z_{t+1}, I_t, \theta) P(s_{t+1} = j | x_{t+1}, z_{t+1}, I_t, \theta)] \right\} dy_{t+1} \\
 &= \sum_{j=1}^M P(s_{t+1} = j | x_{t+1}, z_{t+1}, I_t, \theta) \int y_{t+1} f(y_{t+1} | s_{t+1} = j, x_{t+1}, z_{t+1}, I_t, \theta) dy_{t+1} \\
 &= \sum_{j=1}^M P(s_{t+1} = j | I_t, \theta) E(y_{t+1} | s_{t+1} = j, x_{t+1}, z_{t+1}, I_t, \theta).
 \end{aligned}$$

Note that although the Markov chain itself admits the linear representation,  $\xi_{t+1} = \mathbf{P}\xi_t + v_{t+1}$ , the optimal forecast of  $y_{t+1}$  is a nonlinear function of observables, since the inference  $\hat{\xi}_{t|t}$  in (16) depends nonlinearly on  $I_t$ .

### 3. Results

The dependent variable of the model (1) is Latvia's quarterly GDP series from 1995Q1 till 2009Q3. The explanatory variables are an aggregate of a few components of the GDP from the production side (*cp*), imports (*imp*), net exports (*nx*), and money supply  $M1(m)$ . All series are quarterly, expressed in logs, and once regularly and once seasonally differenced, except  $m$ , that is not seasonally differenced. For comparison, we also run models on series for which one regular differencing of the data is replaced with two regular differences due to the result in Buss (2009) that an extra regular differencing might improve the precision of one-period ahead forecasts during a switch of the business cycle phases. Appendix 1 contains a more detailed description of the data. All calculations are performed in Scilab with the aid of its econometrics toolbox Grocer.

The results are summarized in four tables. Tables 1 and 2 show the results about models imposed on once regularly differenced series, while Tables 3 and 4 - on twice regularly differenced series. Tables 1 and 3 give results about models that use leading information, that is, it is assumed in these tables that variables *cp*, *imp*, *nx* and  $m$  are known one period ahead of the GDP, so the four explanatory variables are used as leading indicators. Note that, in this case, the exogeneity assumption of the explanatory variables no longer holds for single-equation MS model. Nevertheless, we can still analyze the model's forecasting performance, bearing in mind that not all of the model's assumptions are satisfied. On the contrary, Tables 2 and 4 assume there is no leading information, so the timing of the four explanatory variables there coincide with the timing of the GDP.

Table 1 shows root mean squared forecast errors (RMSFE) for the full sample, the first half of the sample (RMSFE1.half) and the second half of the sample (RMSFE2.half) from one-period ahead pseudo real-time forecasts beginning at sample size 19 from single-equation Markov-switching (MS) and linear vector autoregressive (VAR) models with leading information on once regularly differenced series to compare the predictive performance of the two types of models. VAR models are specified by their endogenous variables (first parenthesis) and a lag order (second parenthesis). MS models are specified by the number of regimes (first parenthesis), whether the model is autoregressive (AR) or autoregressive distributed lag (ARDL) model with the explanatory variables other than the endogenous lagged variable specified in parenthesis; next, follows the specification of the lag orders for each explanatory variable along the indication of their dependence (s) or independence (ns) of the regime; finally, it is specified whether the model contains an intercept (c) and whether the intercept is switching (s) or not (ns). The sample is split in halves because the first half of the sample contains a smooth growth whereas the second half of the sample contains a rapid economic downturn, so one can see how the forecasting performance of the models changes along a business cycle. The least RMSFE for each sample space is framed. The forecasting performance of VAR models is shown in the first four rows in Table 1. Model (1) contains two endogenous variables, GDP and *cp*; one can see from models (2) to (4) that an addition of endogenous variable *nx* to the model slightly deteriorates the one-period ahead forecasting

performance of the first model, whereas an addition of *imp* or *m* slightly improves it for the second half of the sample.

**Table 1.** A comparison of one-period ahead pseudo real-time forecasting performance from single-equation MS and VAR models with leading information on once regularly differenced series in terms of RMSFE for the full sample, first half of the sample and second half of the sample.  
The least RMSFE in each sample space is framed.

N <sup>o</sup>	Model	RMSFE	RMSFE1.half	RMSFE2.half
1	VAR(GDP,cp)(2)	0.0186595	0.0142492	0.0222107
2	VAR(GDP,cp,nx)(2)	0.0198866	0.0169021	0.0224782
3	VAR(GDP,cp,imp)(2)	0.0175402	0.0154376	0.0194165
4	VAR(GDP,cp,m)(2)	0.0187341	0.0149037	0.0219047
5	MS(2)-AR(0)-c(s)	0.0289542	0.0165902	0.0374361
6	MS(2)-AR(3)(s)-c(ns)	0.0288561	0.0156118	0.0377043
7	MS(2)-AR(2)(s)-c(ns)	0.0284772	0.0164813	0.0367461
8	MS(2)-AR(2)(ns)-c(s)	0.0324838	0.0174618	0.0424910
9	MS(2)-ARDL(cp)(1,2)(s)-c(ns)	0.0176693	0.0127261	0.0215047
10	MS(2)-ARDL(cp,nx)(1,2,2)(s)-c(ns)	0.0182237	0.0156688	0.0204620
11	MS(2)-ARDL(cp,imp)(1,2,1)(s)-c(ns)	0.0182519	0.0146824	0.0212295
12	MS(2)-ARDL(cp,m)(2,2,2)(s)-c(ns)	0.0207493	0.0174545	0.0235882
13	MS(2)-ARDL(cp,nx)(2(s), {2,2}(ns))	0.0175944	0.0144578	0.0202508
14	MS(2)-ARDL(cp,imp)(2(s), {2,2}(ns))	0.0176390	0.0149478	0.0199708

**Source:** derived by the author.

**Table 2.** A comparison of one-period ahead pseudo real-time forecasting performance from single-equation MS and VAR models without leading information on once regularly differenced series in terms of RMSFE for the full sample, first half of the sample and second half of the sample.  
The least RMSFE in each sample space is framed.

N <sup>o</sup>	Model	RMSFE	RMSFE1.half	RMSFE2.half
1	VAR(GDP,cp)(2)	0.0292156	0.0167268	0.0377798
2	VAR(GDP,cp,nx)(2)	0.0370149	0.0247806	0.0461100
3	VAR(GDP,cp,imp)(2)	0.0341238	0.0196662	0.0440694
4	VAR(GDP,cp,m)(2)	0.0299038	0.0171283	0.0386665
5	MS(2)-AR(0)-c(s)	0.0289542	0.0165902	0.0374361
6	MS(2)-AR(3)(s)-c(ns)	0.0328038	0.0186472	0.0424789
7	MS(2)-AR(2)(s)-c(ns)	0.0282835	0.0162663	0.0365420
8	MS(2)-AR(2)(ns)-c(s)	0.0366473	0.0181612	0.0485410
9	MS(2)-ARDL(cp)(2,2)(s)-c(ns)	0.0283754	0.0161356	0.0367419
10	MS(2)-ARDL(cp,nx)(2,2,2)(s)-c(ns)	0.0322806	0.0195304	0.0412630
11	MS(2)-ARDL(cp,imp)(2,2,2)(s)-c(ns)	0.0318800	0.0257724	0.0369926
12	MS(2)-ARDL(cp,m)(2,2,2)(s)-c(ns)	0.0287118	0.0168035	0.0369646
13	MS(2)-ARDL(cp,m)(2(s), {2,2}(ns))	0.0309477	0.0169541	0.0403495
14	MS(2)-ARDL(cp,imp)(2(s), {2,1}(ns))	0.0323059	0.0151246	0.0431113

**Source:** derived by the author.

**Table 3.** A comparison of one-period ahead pseudo real-time forecasting performance from single-equation MS and VAR models with leading information on twice regularly differenced series in terms of RMSFE for the full sample, first half of the sample and second half of the sample.

The least RMSFE in each sample space is framed.

N <sup>o</sup>	Model	RMSFE	RMSFE1.half	RMSFE2.half
1	VAR(GDP,cp)(3)	0.0208546	0.0189648	0.0226907
2	VAR(GDP,cp,nx)(3)	0.0224267	0.0219535	0.0229187
3	VAR(GDP,cp,imp)(3)	0.0208160	0.0205227	0.0211232
4	VAR(GDP,cp,m)(3)	0.0198595	0.0177855	0.0218484
5	MS(2)-AR(0)-c(s)	0.0331835	0.0264885	0.0390591
6	MS(2)-AR(3)(s)-c(ns)	0.0266697	0.0222837	0.0306497
7	MS(2)-AR(2)(s)-c(ns)	0.0290660	0.0237741	0.0337925
8	MS(2)-AR(2)(ns)-c(s)	0.0323721	0.0273066	0.0370021
9	MS(2)-ARDL(cp)(2,2)(s)-c(ns)	0.0252970	0.0200264	0.0298957
10	MS(2)-ARDL(cp)(3,3)(s)-c(ns)	0.0303869	0.0316822	0.0289470
11	MS(2)-ARDL(cp,nx)(2,2,1)(s)-c(ns)	0.0256729	0.0218496	0.0291915
12	MS(2)-ARDL(cp,imp)(2,2,1)(s)-c(ns)	0.0277238	0.0260182	0.0294278
13	MS(2)-ARDL(cp,m)(2,2,1)(s)-c(ns)	0.0275117	0.0248066	0.0301209
14	MS(2)-ARDL(cp,nx)(2(s),{2,1}(ns))	0.0286470	0.0233144	0.0333926
15	MS(2)-ARDL(cp,m)(2(s),{2,2}(ns))	0.0245432	0.0191248	0.0292193
16	MS(2)-ARDL(cp,nx)({3,3}(s),3(ns))	0.0296731	0.0277234	0.0316133
17	MS(2)-ARDL(cp,imp)(3,3,3)(s)-c(ns)	0.0317678	0.0339501	0.0292714
18	MS(2)-ARDL(cp,m)(3,3,3)(s)-c(ns)	0.0330116	0.0335336	0.0324477

Source: derived by the author.

**Table 4.** A comparison of one-period ahead pseudo real-time forecasting performance from single-equation MS and VAR models without leading information on twice regularly differenced series in terms of RMSFE for the full sample, first half of the sample and second half of the sample.

The least RMSFE in each sample space is framed.

N <sup>o</sup>	Model	RMSFE	RMSFE1.half	RMSFE2.half
1	VAR(GDP,cp)(3)	0.0283351	0.0251617	0.0313568
2	VAR(GDP,cp,nx)(3)	0.0374566	0.0377379	0.0371554
3	VAR(GDP,cp,imp)(3)	0.0351003	0.0318312	0.0382690
4	VAR(GDP,cp,m)(3)	0.0341215	0.0346078	0.0335972
5	MS(2)-AR(0)-c(s)	0.0331835	0.0264885	0.0390591
6	MS(2)-AR(3)(s)-c(ns)	0.0364793	0.0290636	0.0429787
7	MS(2)-AR(2)(s)-c(ns)	0.0289120	0.0233798	0.0338124
8	MS(2)-AR(2)(ns)-c(s)	0.0320115	0.0272778	0.0363722
9	MS(2)-ARDL(cp)(2,2)(s)-c(ns)	0.0291961	0.0228417	0.0346950
10	MS(2)-ARDL(cp,nx)(2,2,2)(s)-c(ns)	0.0339771	0.0292548	0.0383628
11	MS(2)-ARDL(cp,imp)(2,2,2)(s)-c(ns)	0.0317295	0.0266464	0.0363598
12	MS(2)-ARDL(cp,m)(2,2,2)(s)-c(ns)	0.0308232	0.0234447	0.0370879
13	MS(2)-ARDL(cp,m)(2(s),{2,2}(ns))	0.0297117	0.0261638	0.0330669
14	MS(2)-ARDL(cp,imp)(2(s),{2,1}(ns))	0.0313216	0.0245121	0.0372157
15	MS(2)-ARDL(cp)(3,3)(s)-c(ns)	0.0264927	0.0230135	0.0297469
16	MS(2)-ARDL(cp,nx)(3,3,2)(s)-c(ns)	0.0341934	0.0350868	0.0332178
17	MS(2)-ARDL(cp,imp)(3,3,3)(s)-c(ns)	0.0358542	0.0351741	0.0365629
18	MS(2)-ARDL(cp,m)(3,3,2)(s)-c(ns)	0.0281086	0.0238231	0.0320399

Source: derived by the author.

Now, let us discuss the forecasting performance of MS models, starting with MS-AR ones. Model (5) is the mean-variance model since it does not contain any other regressors than a switching constant. Models (6) and (7) are MS-AR models with switching autoregressive coefficient and a non-switching constant. Model (8) is an MS-AR model with non-switching autoregressive coefficients but a switching intercept. It can be seen that these four models, (5)-(8), are doing poorly in the second half of the

sample, although are still competitive in the first half of the sample. Next, consider four MS-ARDL models, (9)-(12), with switching slope coefficients and a non-switching intercept. It can be seen that model (9), which is a close counterpart to model (1), performs slightly better than model (1) in terms of RMSFE in both halves of the sample. Similarly, model (10), which is a close counterpart to model (2), performs slightly better than the latter in all sample spaces. However, models (11) and (12) seem to perform slightly worse than their VAR counterparts, (3) and (4), respectively. Finally, models (13) and (14) exclude a constant and allow for switching coefficients for lagged GDP and  $cp$ , but fix other coefficients. One can see that these two models perform well in both halves of the sample. To summarize information in Table 1, the forecasting performance of the two model types is similar, with the least RMSFE for the first half of the sample obtained by an MS model, and the least RMSFE for the second half of the sample and the full sample – by linear VAR. Next, consider Table 2 that summarizes the forecasting performance of linear VAR and single-equation MS models when no leading information is available. As in the previous table, the first four models are VAR ones with different sets of endogenous variables. Model (5) is the mean-variance model with a switching intercept being the only regressor. Regardless of the simplicity of model (5), it shows a slightly better forecasting performance than any of the VAR models considered in any sample space. Models (6) to (8) introduce non-zero number of lags. It can be seen that introducing two lags of the dependent variable and allowing slope coefficients to be regime-dependent, further improves the forecasting performance of the MS-AR model for all sample spaces. Models (9)-(12) are single-equation MS counterparts to the linear VAR models (1)-(4), respectively. One can see that all four MS models perform better than the respective VAR models for all sample spaces, except model (11) for the first half of the sample. The results for models (13)-(14) show that allowing only the coefficients for lagged dependent variable to switch does not improve the forecasting performance. To summarize Table 2, single-equation MS models tend to perform slightly better than linear VAR models in terms of one-period ahead forecasting performance on once regularly differenced series when no leading information is available.

Following the result of Buss (2009) that, during a switch of the business cycle phases, the short-term forecasting performance might improve if two, instead of one, regular differencing is implemented, Tables 3 and 4 show results for twice regularly differenced data. Table 3 summarizes a comparison of one-period ahead pseudo real-time forecasting performance of single-equation MS and linear VAR models with leading information on twice regularly differenced series. One can see that the performance of MS models lags behind that of VAR counterparts. Table 4 shows the results of single-equation MS and linear VAR models without leading information on twice regularly differenced series. Table 4 shows that, comparing the models with the same variables and lag order, single-equation MS models (15)-(18) tend to give smaller RMSFE than the corresponding linear VAR models for all sample spaces, except for model (17) for the first half of the sample. The results show that, if leading information is available, a second regular differencing of the data does not seem to improve the forecasting precision during a switch of the business cycle phases. However, if no leading information is used, the second regular differencing appears to improve forecasting precision, which is in line with the results in Buss (2009), where no leading information was used.

#### 4. Conclusions

To the best of our knowledge, this is the first publicly available paper that attempts to evaluate short-term forecasting performance of MS models for Latvia's economy. This paper compares one-period ahead pseudo real-time forecasting performance of single-equation MS models compared to linear VAR models with and without leading information. The results show that when leading information is available, the forecasting performance of single-equation MS models is slightly worse than linear VAR models. On the contrary, if there is no leading information at hand, MS models tend to perform somewhat better in terms of one-period ahead forecasts than their linear VAR counterparts.

The results also show that if leading information is available, a second regular differencing of the data does not appear to improve the forecasting precision during a switch of the business cycle phases. However, if no leading information is used, the second regular differencing appears to improve forecasting precision, which is in line with the results in Buss (2009) where no leading information was used.

**References**

- [1] Bellone, B. 2005. Classical Estimation of Multivariate Markov-Switching Models using MSVARlib. *Econometrics* 0508017, EconWPA.
- [2] Buss, G. 2009. Comparing Forecasts of Latvia's GDP Using Simple Seasonal ARIMA Models and Direct Versus Indirect Approach. *MPRA Paper* 16832, University Library of Munich, Germany.
- [3] Dubois, E., and Michaux, E. 2010. Grocer 1.41: an econometric toolbox for Scilab. <http://dubois.ensae.net/grocer.html>.
- [4] Hamilton, J.D. 1989. A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica* Vol. 57, No. 2:357-384.
- [5] Hamilton, J.D. 1994. *Time Series Analysis*. Princeton University Press, Ch. 22: 677-703.
- [6] Krolzig, H.M. 1998. Econometric Modelling of Markov - Switching Vector Autoregressions using MSVAR for Ox. Working paper.
- [7] Krolzig, H.M. 2000. Predicting Markov-Switching Vector Autoregressive Processes. Working paper.
- [8] Krolzig, H.M. 2003. Constructing Turning Point Chronologies with Markov-Switching Vector Autoregressive Models: the Euro-Zone Business Cycle. Working paper.



## APPENDIX

The list of data used in the paper. All national accounts series are chain-priced as of 2000.

Series	Definition	Source
GDP	Gross domestic product	Central Statistical Bureau of Latvia
C	Output in mining and quarrying industry	Central Statistical Bureau of Latvia
D	Output in manufacturing industry	Central Statistical Bureau of Latvia
E	Output in electricity, gas and water supply industry	Central Statistical Bureau of Latvia
F	Output in construction industry	Central Statistical Bureau of Latvia
H	Output in hotels and restaurants industry	Central Statistical Bureau of Latvia
L	Output in public administration and defense, and compulsory social security industries	Central Statistical Bureau of Latvia
D21	Taxes	Central Statistical Bureau of Latvia
cp	Sum of C, D, E, F, H, L, and D21	Derived by the author
exp	Exports	Central Statistical Bureau of Latvia
imp	Imports	Central Statistical Bureau of Latvia
nx	Net exports, exp-imp	Derived by the author
m	Monetary aggregate M1, quarterly average	Bank of Latvia

## THE UNEMPLOYMENT VOLATILITY PUZZLE: THE ROLE OF THE UNDERGROUND ECONOMY

Lisi GAETANO  
University of Cassino, ITALY  
[gaetano.lisi@unicas.it](mailto:gaetano.lisi@unicas.it)

### Abstract

*Relying on the non-negligible role played by the underground economy in the labour market fluctuations, this paper extends the standard matching model à la Mortensen-Pissarides by introducing an underground sector along with an endogenous sector choice for both entrepreneurs and workers. These modifications improve the quantitative properties of the standard matching model, thus providing a possible explanation for the unemployment volatility puzzle.*

**Keywords:** unemployment and vacancies volatility, productivity and job destruction shocks, underground economy, shadow economy, hidden economy, matching models

**JEL Classification:** E26, E32, J23, J24, J63, J64

### 1. Introduction

In an influential paper, Shimer (2005) shows that the standard matching model fails to reproduce the large volatility in unemployment and vacancies during the business cycle. In fact, these variables are much more volatile in the US data than in the calibrated model subject to productivity shocks of a realistic magnitude.

This is a very important setback for the profession: the reason is that the Mortensen-Pissarides model (1994) is the “workhorse” model used by academic and government economists to evaluate various economic policies and to study the problem of unemployment (Hagedorn, and Manovskii 2008). Hence, many and influential improvements have been offered to enrich the quantitative properties of the standard matching model and to challenge the unemployment volatility puzzle (see e.g. Shimer 2004, 2005, Hall 2005, Garibaldi 2006, Uren 2007, Hall, and Milgrom 2008, Hagedorn, and Manovskii 2008).<sup>30</sup>

The current paper belongs to that literature that points out the need to focus on the underground economy. There is in fact substantial economic activity in the underground sector in many developing countries as well as in many transition countries, and the underground sector is also important in some developed economies (Albrecht *et al.* 2009). Since these underground activities make a sizeable contribution to the GDP, it is difficult to understand the business cycles without some knowledge of the underground economy fluctuations. Indeed, introducing an underground sector in a *Real Business Cycle* framework improves the fit of the model to the data (Busato, and Chiarini 2004). Furthermore, Bosch and Esteban-Pretel (2009) build a search and matching model extended to the underground sector which satisfactorily explains most of the cyclical properties found in the Brazil data.

This paper presents a two-sector search and matching model which addresses the lack of amplification in unemployment and vacancy volatility by introducing the sector choice of individuals. Precisely, the model assumes the exogenous existence of firms (posting vacancies) and workers (searching for jobs); however, heterogeneous entrepreneurs sort themselves into official or unofficial statuses according to their ex-ante entrepreneurial ability levels, and heterogeneous workers direct their searches towards one of the two types of firms according to their ex-ante skill levels. Furthermore, entrepreneurs and workers can move from one sector to the other in reply to exogenous changes in policy parameters.<sup>31</sup>

We then calibrate and simulate the model to try to account for the Shimer puzzle. As in Uren (2007), we compare different steady-state equilibria in which the value of the productivity of the match and the job destruction rate are stochastic and are drawn from a bivariate normal distribution constructed

---

<sup>30</sup> See Hornstein, Krusell, and Violante (2005), Yashiv (2006) and Mortensen and Nagypál (2007), for exhaustive surveys of this huge and recent literature.

<sup>31</sup> Bosch and Esteban-Pretel (2009) consider only direct transitions from informal to formal jobs.

to match the stylised business cycle facts of the US economy.<sup>32</sup> The labour market tightness in both sectors is given at the microeconomic level and endogenously determined in the aggregate economy. Indeed, once the aggregate value of vacancies and unemployment has been calculated in reply to productivity and job destruction shocks, the value of labour market tightness in each sector can be obtained. This process triggers off two extra effects on vacancies and unemployment: the first works on the probabilities to find a job and fill a vacancy (*direct effect*), and the second works on the sector choice of individuals, thus modifying the share of entrepreneurs and workers in the two sectors (*indirect effect*).

Furthermore, since in this model the zero-profit condition is no longer used to determine the labour market tightness, the incidence of wage on job creation is mediate and it is thus possible to maintain the Nash bargaining rule.<sup>33</sup>

The main aim of this paper is to show that, once incidence of the underground economy on the business cycle (i.e. the sector choice of individuals) is accounted for, a matching model is able to replicate most of the fluctuations in unemployment and job vacancies caused by productivity shocks of plausible magnitude.

The rest of the paper is organized as follows: section 2 contains a brief review of the related literature; sections 3 and 4 present a two-sector matching model with an endogenous sector choice for both entrepreneurs and workers; section 5 provides a quantitative evaluation of an economy subject to shocks to productivity and job destruction rate; section 6 concludes.

## 2. Related Literature

The literature has mainly responded to the Shimer puzzle by suggesting that the wage setting mechanism in the standard model has to be altered in order to break the close link between wages and productivity (Shimer 2004, 2005, Hall 2005, Hall, and Milgrom 2008). Indeed, according to the literature, a principal reason for the lack of explanatory power of the standard matching model (in which the labour market tightness is obtained by the zero-profit condition) is that the wage implied by the Nash bargaining rule responds to *offset all* the effects of productivity shocks on job creation. Hence, following Binmore, Rubinstein and Wolinsky (1986), Hall and Milgrom (2008) propose a strategic bargaining game in which the default option, i.e. the threat point, is the delay rather than the search for an alternative job. This modification limits the influence of labour market conditions on wages.<sup>34</sup> The insight is the following: if the delay is the only outside option, then the wage agreement will not reflect the value of unemployment and, consequently, the wage is more rigid (i.e. less pro-cyclical) than that implied by the standard sharing rule, i.e. the Nash rule. Nevertheless, as claimed by Mortensen and Nagypál (2007), this solution cannot be sufficient to solve the Shimer puzzle by itself.

Several papers take a different route. In particular, there is a piece of the literature which suggests that the problem lies not in the model itself, but in the way the model is typically calibrated. Hence, Hagedorn and Manovskii (2008) propose a new calibration strategy for the two central parameters of the Mortensen-Pissarides model: the opportunity cost of employment and the worker's bargaining power. Indeed, in Hagedorn and Manovskii (2008), the standard matching model may generate the observed volatility of unemployment and vacancies. However, as claimed by Mortensen and Nagypál (2007), they make use of unrealistic calibration values (i.e. a huge opportunity cost of employment and a tiny workers' bargaining power) to solve the puzzle.

Other recent papers introduce new and interesting features in the standard Mortensen-Pissarides model. Garibaldi (2006) shows that a calibration of the model that explicitly considers hiring freeze and bankruptcy of firms can account for 20 to 35 percent of the variability displayed by the data. Mortensen and Nagypál (2007) find that an augmented matching model, which includes capital costs, countercyclical job destruction rate and a less procyclical wage, explains about 40% of the observed

---

<sup>32</sup> The labour productivity is the key parameter and represents the driving force in business cycle analysis in most macro-models of the labour market. Fluctuations in the separation rate are considered a further potential driving source (Garibaldi 2006). Indeed, according to the Brazil data, the volatility of unemployment is explained in large part by changes in the job separation rate (Bosch and Esteban-Pretel 2009).

<sup>33</sup> Recall that in the standard framework (see Pissarides 2000), the *Job Creation Condition* depicts a negative relationship between labour market tightness and wage, since the zero-profit (or free entry) condition directly relates the two variables.

<sup>34</sup> Using a simple rigid wage, the response in the vacancy-unemployment ratio to productivity shocks is still too small. Furthermore, the rigid wage assumption is difficult to justify (Mortensen, and Nagypál 2007).

volatility of the vacancy-unemployment ratio. Finally, Uren (2007) displays that the simple introduction of the endogenous decision of an individual to either become entrepreneur or worker can substantially amplify the impact of productivity shocks upon the level of unemployment and vacancies; such a model in fact generates fluctuations in unemployment and vacancies that are roughly 3 and 5 times greater than those produced by the standard model as evaluated by Shimer (2005).

### 3. The economy

The economy is populated by a continuum of entrepreneurs and a continuum of workers.<sup>35</sup> Each entrepreneur is born with a specific entrepreneurial ability  $x$ , which is drawn from a known distribution  $F : [x_{min}, x_{max}] \rightarrow [0, 1]$ . Similarly, each worker is endowed with a different level of skill  $q$ , with known distribution  $G : [q_{min}, q_{max}] \rightarrow [0, 1]$ . The economic environment is characterised by a labour market with search frictions and wage bargaining. Each entrepreneur can either operate regularly or against the tax regulations. Further, our notion of underground employment is one of low productivity jobs (see e.g. Boeri and Garibaldi, 2002, 2006).<sup>36</sup>

The matching of vacancies ( $v$ ) and unemployed ( $u$ ) per unit of time is regulated by an aggregate matching function:  $m_i = m(v_i, u_i)$  with the restrictions  $m(v_i, 0) = m(0, u_i) = 0$  and  $m(v_i, u_i) \leq \min(v_i, u_i)$ , where  $i \in \{r, s\}$  denotes the sector (with  $r = regular$ ,  $s = shadow$ ). As usual, the matching function is non-negative, increasing, concave in both arguments and homogenous of degree one (Pissarides, 2000; Petrongolo and Pissarides, 2001), so that the vacancy-filling rate,  $f(\theta_i) \equiv m(v_i, u_i) / v_i = m(1, \theta_i^{-1})$ , and the job-finding rate,  $g(\theta_i) \equiv m(v_i, u_i) / u_i = m(\theta_i, 1)$ , are both functions of the vacancy-unemployment ratio,  $\theta_i = v_i / u_i$ . Standard technical assumptions are assumed:

$$f'(\theta_i) < 0, f''(\theta_i) > 0, g'(\theta_i) > 0, g''(\theta_i) < 0;$$

$$\lim_{\theta_i \rightarrow \infty} f(\theta_i) = \lim_{\theta_i \rightarrow 0} g(\theta_i) = 0, \lim_{\theta_i \rightarrow 0} f(\theta_i) = \lim_{\theta_i \rightarrow \infty} g(\theta_i) = \infty, \text{ with } i \in \{r, s\}.$$

Following Garibaldi (2006), each individual takes as given  $\theta_i$ , and in the aggregate economy the labour market tightness depends on the aggregate number of vacancies and the stock of unemployed.

The Bellman equations used to find infinite horizon steady-state solutions are:

$$\begin{aligned} rJ_r &= p - w_r - \tau + \delta \cdot [V_r - J_r], & rJ_s &= \kappa \cdot p - w_s + (\delta + \rho) \cdot [V_s - J_s] \\ rW_r &= w_r + \delta \cdot [U_r - W_r], & rW_s &= w_s + (\delta + \rho) \cdot [U_s - W_s] \\ rV_i &= -c_i + f(\theta_i) \cdot [J_i - V_i], & rU_i &= z_i + g(\theta_i) \cdot [W_i - U_i] \quad \text{with } i \in \{r, s\} \end{aligned}$$

where  $J_i$  is the value of a filled job;  $W_i$  is the value for being employed;  $V_i$  is the value of a vacancy;  $U_i$  is the value for seeking a job;  $r$  is the discount rate;  $p$  is the productivity of the match;  $w_i$  is the wage;  $\delta$  is the job destruction rate;  $c_i$  is the vacant job cost and  $z_i$  is the opportunity cost of employment. In the regular sector firms pay a production tax  $\tau$ , while in the underground sector this tax is evaded and there is a monitoring rate  $\rho$ . Conditional on being monitored in the underground sector, the irregular job is destroyed. Furthermore, as in Bosch and Esteban-Pretel (2009), evading taxation implies that irregular firms can only take advantage of a fraction  $\kappa \in (0, 1)$  of the productivity of the match. The key payoffs for the entry into the labour market, i.e. the value of a vacancy and the value for seeking a job, can be expressed as single-valued functions of market tightness  $\theta_i$ , with  $\partial U_i / \partial \theta_i > 0$  and  $\partial V_i / \partial \theta_i < 0 \quad \forall i$  (see Appendix 1).<sup>37</sup> Finally, wages are assumed to be the outcome of a Nash bargaining problem:

<sup>35</sup> Time is continuous, and individuals are risk neutral and infinitely lived. We neglect the endogenous decision of individuals to either become entrepreneurs or workers because it is widely discussed in the matching literature (see Fonseca, Lopez-Garcia, and Pissarides, 2001, Pissarides, 2002, and Uren, 2007). This framework can thus be seen as a following step.

<sup>36</sup> We neglect possibilities of moonlighting, so workers can perform only one activity at a time.

<sup>37</sup> Intuitively, this is straightforward to understand since the greater  $\theta_i$  the smaller the probability of filling a vacancy for the firm, and the greater  $\theta_i$  the higher the probability of finding a job for the worker.

$$w_i = \arg \max \left\{ (W_i - U_i)^{\beta_i} \cdot (J_i - V_i)^{1-\beta_i} \right\} \Rightarrow (W_i - U_i) = \frac{\beta_i}{(1-\beta_i)} \cdot (J_i - V_i) \quad \text{with } i \in \{r, s\}$$

where  $\beta_i \in (0, 1)$  is the surplus share for labour. Simple manipulations thus yield the formulae of wages:

$$w_r = (1 - \beta_r) \cdot rU_r(\theta_r) + \beta_r \cdot (p \cdot y_r - \tau - rV_r(\theta_r)) \quad (1)$$

$$w_s = (1 - \beta_s) \cdot rU_s(\theta_s) + \beta_s \cdot (p \cdot y_s - rV_s(\theta_s)) \quad (2)$$

with  $w_i'(\theta_i) > 0 \quad \forall i$ , since  $V_i'(\theta_i) < 0$ , and  $U_i'(\theta_i) > 0 \quad \forall i$ .

#### 4. Entrepreneurs and workers

To start up a regular business an entrepreneur has to pay an extra cost  $h$ , which measures the burden of overall constraints in the official sector (the so-called *barriers to entrepreneurship*).<sup>38</sup> Entrepreneurial ability  $x$  influences this start-up cost, but as in Pissarides (2002) it does not affect the job productivity. Therefore entrepreneurs whose  $x$  satisfies the following inequality enter the regular sector:

$$V_r(\theta_r) - x \cdot h \geq V_s(\theta_s) \quad (3)$$

hence, regular (good) entrepreneurs have a low  $x$ , irregular (poor) entrepreneurs have a high  $x$ . It follows that there is a *reservation entrepreneurial ability*  $x = R$ , with  $R'(\theta_r) < 0$  and  $R'(\theta_s) > 0$ , such that an entrepreneur enters the official sector if  $x \leq R$ , otherwise he/she starts a business in the underground sector. Therefore, a fraction  $F(R)$  of the entrepreneurs (either posting a vacancy or producing) are regular, while the complementary fraction  $1 - F(R)$  are irregular. The evolution of vacancies in each sector is thus given by:

$$\dot{v}_r = \underbrace{[F(R) - v_r]}_{\text{inflow}} \cdot \delta - \underbrace{v_s \cdot f(\theta_r)}_{\text{outflow}}$$

$$\dot{v}_s = \underbrace{[1 - F(R) - v_s]}_{\text{inflow}} \cdot (\delta + \rho) - \underbrace{v_s \cdot f(\theta_s)}_{\text{outflow}}$$

steady-state implies  $\dot{v}_r = \dot{v}_s = 0$ , so that:

$$v_r = \frac{F(R) \cdot \delta}{f(\theta_r) + \delta} \quad (4)$$

$$v_s = \frac{[1 - F(R)] \cdot (\delta + \rho)}{f(\theta_s) + (\delta + \rho)} \quad (5)$$

From the supply side, workers can be either high-skilled or low-skilled, thus achieving two different levels of productivity, depending on whether they choose to invest in education properly. Indeed, formal education enhances the worker's skill (Laing *et al.* 1995). As in Dulleck *et al.* (2006), workers with higher ability (i.e. in the specific instance with lower  $q$ ) have lower costs of higher education  $e$ . Therefore workers whose  $q$  satisfies the following inequality enter the regular sector:<sup>39</sup>

$$U_r(\theta_r) - q \cdot e \geq U_s(\theta_s) \quad (6)$$

as a result, there will be a *cut-off* value  $Q$ , with  $Q'(\theta_r) > 0$  and  $Q'(\theta_s) < 0$ , below/equal which workers are regular/high-skilled and above which they are irregular/low-skilled. Then,  $G(Q)$  and  $1 - G(Q)$  are the share of high-skilled and low-skilled workers, respectively. Under these assumptions, the unemployment rate adjusts according to the following *law of motion*:

$$\dot{u}_r = [G(Q) - u_r] \cdot \delta - u_r \cdot g(\theta_r)$$

$$\dot{u}_s = [1 - G(Q) - u_s] \cdot (\delta + \rho) - u_s \cdot g(\theta_s)$$

in steady-state, unemployment is thus given by:

<sup>38</sup> This can be explained by higher entrance barriers into the official sector or access costs to legality associated with excessive regulations, administrative burdens, licence fees, bribery (see e.g. Bouev 2005), but also money protection if the firm copes with a context where organized crime operates.

<sup>39</sup> Indeed, there is a strong negative correlation between informal-sector employment and education level within countries (cf. Albrecht *et al.* 2009).

$$u_r = \frac{G(Q) \cdot \delta}{g(\theta_r) + \delta} \quad (7)$$

$$u_s = \frac{[1 - G(Q)] \cdot (\delta + \rho)}{g(\theta_s) + (\delta + \rho)} \quad (8)$$

In the aggregate economy, the equilibrium value of  $\theta_i$  must be consistent with the value assumed at the microeconomic level. Formally, the aggregate definition of  $\theta_i$  is:

$$\theta_r = \frac{v_r}{u_r} \Rightarrow \theta_r = \frac{F(R) \cdot \delta}{f(\theta_r) + \delta} \cdot \frac{g(\theta_r) + \delta}{G(Q) \cdot \delta} \quad (9)$$

$$\theta_s = \frac{v_s}{u_s} \Rightarrow \theta_s = \frac{[1 - F(R)] \cdot (\delta + \rho)}{f(\theta_s) + (\delta + \rho)} \cdot \frac{g(\theta_s) + (\delta + \rho)}{[1 - G(Q)] \cdot (\delta + \rho)} \quad (10)$$

which are non-negative, by construction, since  $v_i, u_i \in [0,1]$ . Rearranging them, the equilibrium value of  $\theta_r$  and  $\theta_s$  must satisfy the following equations:

$$[g(\theta_r) + \theta_r \cdot \delta] \cdot G(Q) - F(R) \cdot [g(\theta_r) + \delta] = 0 \quad (9')$$

$$[g(\theta_s) + \theta_s \cdot (\delta + \rho)] \cdot [1 - G(Q)] - [1 - F(R)] \cdot [g(\theta_s) + (\delta + \rho)] = 0 \quad (10')$$

since  $g(\theta_i) = \theta_i \cdot f(\theta_i)$ . A unique equilibrium exists if there is a unique value of  $\theta_r$  and  $\theta_s$  that solves equations (9') and (10'), respectively. It is possible to verify that,

**Proposition 1.** *A steady state equilibrium exists and is unique (see Appendix 2 for the proof). Furthermore,*

**Proposition 2.** *Given the assumption of search frictions in both sectors, the equilibrium is characterized by an interior solution (see Appendix 3).<sup>40</sup>*

This model differs from the standard Mortensen-Pissarides framework because the effects of exogenous changes to parameters on vacancy-unemployment ratio depend on the variation of the threshold values. Indeed, in this model the labour market tightness in each sector is given by the aggregate level of vacancies and unemployment. The comparative static results are generally intuitive and straightforward, since the effects of higher monitoring, taxes and labour market regulations are common in the literature:

$$\frac{\partial R}{\partial \rho} > 0; \quad \frac{\partial R}{\partial \tau} < 0; \quad \frac{\partial R}{\partial c_r} < 0; \quad \frac{\partial R}{\partial p} > 0;$$

furthermore, this last result implies that in boom regular firms take more advantage of the increase in productivity (irregular jobs are less productive than regular ones since  $\kappa < 1$ ).

## 5. Quantitative evaluation of the model

In order to evaluate the quantitative properties of the model, a numerical simulation is performed. The model calibration is reported in Table 1. To take the sectoral differences into account, we make use of Boeri and Garibaldi's (2006) calibrations (see Table 2).

As in Uren (2007), we compare different steady states as the value of the productivity ( $p$ ) and the job destruction rate ( $\delta$ ) vary. Consistent with the data presented in Shimer (2005), the productivity ( $p$ ) and the job destruction rate ( $\delta$ ) are stochastic and are drawn from a bivariate normal distribution constructed to match the stylised business cycle facts of the US economy. Hence, the mean value of  $p$  is normalised to 1 and the standard deviation is set to 0.019, whereas the mean of  $\delta$  is set to 0.1 and the standard deviation is set to 0.075. Finally, the correlation between these variables is -0.592, by construction.

The sector choice is also solved analytically by the variation of the threshold values  $R$  and  $Q$ . Both the cumulative distribution functions are assumed to be negative exponential.<sup>41</sup>

<sup>40</sup> This is anything but a trivial result. Indeed, as claimed by Bouev (2005), the most important weakness in the underground economy theory is represented by the restrictive assumptions required to find an interior equilibrium in which the underground sector coexists with the regular one in the long run.

Table 1. Model calibration

Matching function	$m_i = v_i^\eta \cdot u_i^{1-\eta}$		Petrongolo, and Pissarides (2001)
Parameter	Notation	Value	Source / Target
Job-finding rate elasticity with respect to market tightness	$\eta$	0.4	Uren (2007)
Matching function elasticity with respect to unemployment rate	$1 - \eta$	0.6	Uren (2007)
Productivity loss in the irregular sector	$\kappa$	0.81	Bosch, and Esteban-Pretel (2009)
Discount rate	$r$	0.012	Shimer (2005); Uren (2007)
Bargaining power of workers	$\beta$	0.50	Uren (2007); Boeri, and Garibaldi (2006)
Barriers to entrepreneurship (US Index)	$h$	1.236	OECD (2009)
Opportunity cost of employment	$z$	0.4	Shimer (2005)
Higher education cost (US Index on OECD average)	$e$	1.896	OECD (2009)
Vacant job cost	$c$	0.215	Uren (2007)
Exponential distribution parameter	$\lambda$	1	Boeri, and Garibaldi (2006)
Variable			
Vacancies	$v$	endogenous	<i>law of motion</i> (steady state solution)
Unemployment	$u$	endogenous	<i>law of motion</i> (steady state solution)
Wage	$w$	endogenous	Nash bargaining rule
Entrepreneurial ability threshold	$R$	endogenous	<i>cut-off condition</i>
Worker's ability threshold	$Q$	endogenous	<i>cut-off condition</i>
Key parameter			
Productivity of the match	$p$	stochastic	simulation
Job destruction rate	$\delta$	stochastic	simulation

Table 2. Sector calibration\*

	Notation	Regular	Shadow
Production tax	$\tau$	0.20	–
Monitoring rate	$\rho$	–	0.06
Labour market tightness	$\theta$	2.70	0.16
Job-finding rate	$g(\theta)$	0.82	0.28
Vacancy-filling rate	$f(\theta)$	0.30	1.75

Source: Boeri, and Garibaldi, 2006.

<sup>41</sup> A negative exponential distribution is used by Boeri and Garibaldi (2006) for the distribution of labour productivity.

\* As in Boeri and Garibaldi (2006), we assume that  $\beta$  and  $\eta$  are identical in the two sectors. Note that the values of labour market tightness, job-finding rate and vacancy-filling rate reported in Table 2 are the starting exogenous values of these variables.

The simulation works as follows: given  $p$  and  $\delta$  the corresponding values of  $w_r$ ,  $w_s$ ,  $R$ ,  $Q$ ,  $v_r$ ,  $v_s$ ,  $u_r$  and  $u_s$  are calculated. Furthermore, once the aggregate value of vacancies and unemployment has been calculated, we get the “updated” value of labour market tightness in each sector. In short, this process triggers off three effects on vacancies and unemployment, thus amplifying their volatility. First, the productivity and job destruction shocks work on the sector choice of individuals, thus modifying the share of entrepreneurs and workers in the two sectors (*starting effect*). Second, the aggregate value of labour market tightness in each sector works by modifying the probabilities to find a job and fill a vacancy (*extra direct effect*). Finally, the aggregate value of labour market tightness in each sector also affects the sector choice of individuals (*extra indirect effect*).

This process is repeated 10,000 times and the correlations and the standard deviations of the key variables are calculated. The simulation results are shown in Table 3.

Table 3. Results

Standard deviation	0.019	0.075	0.112	0.103	0.098	0.084
Correlations	$p$	$\delta$	$v_r$	$v_s$	$u_r$	$u_s$
$p$	1					
$\delta$	-0.596	1				
$v_r$	0.259	0.293	1			
$v_s$	-0.215	0.341	-0.014	1		
$u_r$	-0.219	0.352	-0.095	0.165	1	
$u_s$	0.186	0.323	0.185	-0.149	-0.221	1

Aggregate Statistics (standard deviation)	$\sigma_u$	$\sigma_v$
Our calibration	0.128	0.145
Uren’s calibration (2007)	0.124	0.087
Standard calibration (Shimer, 2005)	0.031	0.011
Volatility observed in the data by Shimer	0.190	0.202

We can summarize them as follows:<sup>42</sup>

- *Aggregate Unemployment.* The standard deviation of aggregate unemployment ( $u = u_r + u_s$ ) explains about 70% of the unemployment volatility observed in the data by Shimer (2005);
- *Aggregate Vacancies.* A two-sector matching model explains more than 70% of the observed volatility of job vacancies;
- *Beveridge Curve.* The model succeeds in correctly predicting a downward sloping vacancy-unemployment relationship in both sectors;
- *Cyclical Behaviour.* The counter-cyclical behaviour of the underground economy is caught by the negative correlation between  $p$  and  $v_s$ .

## 6. Conclusions

Relying on the non-negligible role played by the underground economy in the labour market fluctuations, this paper extends the standard matching model *à la* Mortensen-Pissarides by introducing an underground sector along with an endogenous sector choice for both entrepreneurs and workers. These modifications improve the model’s implications for the amplification of shocks, thus providing a possible explanation for the unemployment volatility puzzle. Indeed, a two-sector matching model explains about 70% of the unemployment volatility observed in the data by Shimer (2005), and more than 70% of the observed volatility of job vacancies.

<sup>42</sup> The statistics for job vacancies and unemployment are those reported in Shimer (2005) for an Hodrick-Prescott (HP) filter with a smoothing parameter of  $10^5$  instead of 1600 as followed by Andolfatto (1996) and Merz (1995). However, as claimed by Hornstein *et al.* (2005), the choice of smoothing parameter has no impact on the unemployment and vacancy statistics.



**References**

- [1] Albrecht, J., Navarro, L., and Vroman, S. 2009. The Effects of Labour Market Policies in an Economy with an Informal Sector. *Economic Journal* 119(539):1105-1129.
- [2] Andolfatto, D. 1996. Business Cycles and Labor-Market Search. *American Economic Review* 86(1):112-132.
- [3] Binmore, K., Rubinstein, A., and Wolinsky, A. 1985. The Nash Bargaining Solution in Economic Modelling. *The Rand Journal of Economics* 17(2):176-188.
- [4] Boeri, T., and Garibaldi, G. 2002. Shadow Activity and Unemployment in a Depressed Labour Market. *CEPR Discussion Papers*, 3433, June.
- [5] Boeri, T., and Garibaldi, G. 2006. Shadow Sorting. *Fondazione Collegio Carlo Alberto Working Paper Series*, 10, May.
- [6] Bosch, M., and Esteban-Pretel, J. 2009. Cyclical Informality and Unemployment. *CIRJE Discussion Papers*, 613, February.
- [7] Bouev, M. 2002. Official Regulations and the Shadow Economy: A Labour Market Approach. *William Davidson Institute Working Paper*, 524, December.
- [8] Bouev, M. 2005. State Regulations, Job Search and Wage Bargaining: A Study in the Economics of the Informal Sector. *William Davidson Institute Working Paper*, 764, April.
- [9] Dulleck, U., Frijters, P., and Winter-Ebmer, R. 2006. Reducing Start-up Costs for New Firms: The Double Dividend on the Labor Market. *Scandinavian Journal of Economics*, 108(2): 317-337.
- [10] Garibaldi, P. 2006. Hiring Freeze and Bankruptcy in Unemployment Dynamics. *IZA Discussion Paper Series*, 2263, August.
- [11] Hagedorn, M., and Manovskii, I. 2008. The Cyclical Behaviour of Equilibrium Unemployment and Vacancies Revisited. *American Economic Review*, 98(4):1692-1706.
- [12] Hall, R.E. 2005. Employment Fluctuations with Equilibrium Wage Stickiness. *American Economic Review*, 95(1):50-65.
- [13] Hall, R.E., and Milgrom, P.R. 2008. The Limited Influence of Unemployment on the Wage Bargain. *American Economic Review*, 98(4):1653-1674.
- [14] Hornstein, A., Krusell, P., and Violante, G.L. 2005. Unemployment and Vacancy Fluctuations in the Matching Model: Inspecting the Mechanism. *Economic Quarterly*, 91(3):19-51.
- [15] Laing, D., Palivos, Th., and Wang, P. 1995. Learning, Matching and Growth. *Review of Economic Studies*, 62(1):115-29.
- [16] Merz, M. 1995. Search in the Labor Market and the Real Business Cycle. *Journal of Monetary Economics*, 36(2):269-300.
- [17] Mortensen, T.D., and Pissarides, Ch.A. 1994. Job Creation and Job Destruction in the Theory of Unemployment. *Review of Economic Studies*, 61(3):397-415.
- [18] Mortensen, T.D., and Nagypal, E. 2007. More on Unemployment and Vacancy Fluctuations. *Review of Economic Dynamics*, 10(3):327-347.
- [19] OECD. 2009. *Education at a Glance 2009: OECD Indicators* (www.oecd.org).
- [20] Petrongolo, B., and Pissarides, Ch.A. 2001. Looking into the Black Box: A Survey of the Matching Function. *Journal of Economic Literature*, 39(2):390-431.
- [21] Pissarides, Ch.A. 2000. *Equilibrium Unemployment Theory* (2<sup>nd</sup> Edition), The MIT Press.
- [22] Pissarides, Ch.A. 2002. Company Start-Up Costs and Employment. *CEP Discussion Papers*, dp0520, March.

- [23] Shimer, R. 2004. The Consequences of Rigid Wages in Search Models. *Journal of the European Economic Association*, 2(2-3):469-479, 04/05.
- [24] Shimer, R. 2005. The Cyclical Behavior of Equilibrium Unemployment and Vacancies. *American Economic Review*, 95(1):25-49.
- [25] Uren, L. 2007. Entrepreneurship and Labour Market Fluctuations. *Economics Bulletin*, 10(13):1–11.
- [26] Yashiv, E. 2006. Search and Matching in Macroeconomics. *CEPR Discussion Paper*, 5459, January.
- [27] Wölfl, A., Wanner, I., Kozluk, T., and Nicoletti, G. 2009. Product Market Regulation in OECD Countries 1998-2008. *OECD Economics Department Working Paper*, forthcoming.

## APPENDIXES

### Appendix 1. Value functions

From the Bellman equations very simple algebra gives:

$$[J_r - V_r] = \frac{p \cdot y_r - w_r - \tau + c_r}{r + \delta + f(\theta_r)}; \quad [J_s - V_s] = \frac{p \cdot y_s - w_s + c_s}{r + \delta + \rho + f(\theta_s)}$$

$$[W_r - U_r] = \frac{w_r - z_r}{r + \delta + g(\theta_r)}; \quad [W_s - U_s] = \frac{w_s - z_s}{r + \delta + \rho + g(\theta_s)}$$

Hence, it is straightforward to get:

$$rV_r = \frac{f(\theta_r) \cdot (p \cdot y_r - w_r - \tau) - c_r \cdot (r + \delta)}{r + \delta + f(\theta_r)} \quad (1.1)$$

$$rV_s = \frac{f(\theta_s) \cdot (p \cdot y_s - w_s) - c_s \cdot (r + \delta + \rho)}{r + \delta + \rho + f(\theta_s)} \quad (1.2)$$

$$rU_r = \frac{g(\theta_r) \cdot w_r + z_r \cdot (r + \delta)}{r + \delta + g(\theta_r)} \quad (1.3)$$

$$rU_s = \frac{g(\theta_s) \cdot w_s + z_s \cdot (r + \delta + \rho)}{r + \delta + \rho + g(\theta_s)} \quad (1.4)$$

with  $\partial V_i / \partial \theta_i < 0$  and  $\partial U_i / \partial \theta_i > 0$ ,  $i \in \{r, s\}$ . Further,

$$\begin{aligned} \lim_{\theta_r \rightarrow 0} rV_r &= py_r - w_r - \tau & \lim_{\theta_s \rightarrow 0} rV_s &= py_s - w_s; & \lim_{\theta_i \rightarrow 0} rU_i &= z_i \\ \lim_{\theta_i \rightarrow \infty} rV_i &= -c_i; & \lim_{\theta_i \rightarrow \infty} rU_i &= w_i & \text{with } i \in \{r, s\} \end{aligned}$$

### Appendix 2. Existence and uniqueness of the equilibrium

To prove the existence and uniqueness of the equilibrium, let us define:

$$\Gamma(\theta_r) = [g(\theta_r) + \theta_r \cdot \delta] \cdot G(Q) - F(R) \cdot [g(\theta_r) + \delta] \quad (2.1)$$

$$\Gamma(\theta_s) = [g(\theta_s) + \theta_s \cdot (\delta + \rho)] \cdot [1 - G(Q)] - [1 - F(R)] \cdot [g(\theta_s) + (\delta + \rho)] \quad (2.2)$$

and note that,

$$\lim_{\theta_r \rightarrow 0} \Gamma(\theta_r) = -F(R) \cdot \delta; \quad \lim_{\theta_s \rightarrow 0} \Gamma(\theta_s) = -[1 - F(R)] \cdot (\delta + \rho);$$

$$\lim_{\theta_r \rightarrow \infty} \Gamma(\theta_r) = +\infty \text{ and } \lim_{\theta_s \rightarrow \infty} \Gamma(\theta_s) = +\infty, \text{ by the } l' \text{ H\^opital rule.}^{43}$$

Furthermore, for given  $R$  and  $Q$ , we get:

$$\Gamma'(\theta_r) = [g'(\theta_r) + \delta] \cdot G(Q) - F(R) \cdot g'(\theta_r) > 0,$$

$$\text{if } G(Q) \cdot \left[ 1 + \frac{\delta}{g'(\theta_r)} \right] > F(R);$$

$$\Gamma'(\theta_s) = [g'(\theta_s) + \delta + \rho] \cdot [1 - G(Q)] - [1 - F(R)] \cdot g'(\theta_s) > 0$$

$$\text{if } [1 - G(Q)] \cdot \left[ 1 + \frac{\delta + \rho}{g'(\theta_s)} \right] > [1 - F(R)].$$

More importantly, when  $R$  and  $Q$  are handled as variables the previous result is obtained without any restriction. Indeed, since  $R'(\theta_r) < 0$ ,  $R'(\theta_s) > 0$ ,  $Q'(\theta_r) > 0$  and  $Q'(\theta_s) < 0$ , when  $\theta_r$

<sup>43</sup> Note that, in order to apply the *l' H\^opital rule*, we can rewrite (2.1) and (2.2) as follows:

$$[g(\theta_r) + \delta \cdot \theta_r] \cdot G(Q) \cdot \left[ 1 - \frac{[g(\theta_r) + \delta] \cdot F(R)}{[g(\theta_r) + \delta \cdot \theta_r] \cdot G(Q)} \right];$$

$$[g(\theta_s) + (\delta + \rho) \cdot \theta_s] \cdot [1 - G(Q)] \cdot \left[ 1 - \frac{[g(\theta_s) + \delta + \rho] \cdot [1 - F(R)]}{[g(\theta_s) + (\delta + \rho) \cdot \theta_s] \cdot [1 - G(Q)]} \right].$$

rises,  $F(R)$  decreases and  $G(Q)$  increases. Similarly, when  $\theta_s$  rises,  $[1 - F(R)]$  decreases and  $[1 - G(Q)]$  increases. Furthermore, according to the model, the labour market tightness in the two sectors cannot move in the same direction. Hence, their effects on  $R$  and  $Q$  cumulate and strengthen. As a result, equations (2.1)–(2.2) are strictly increasing and continuous functions of  $\theta_r$  and  $\theta_s$ , respectively. Therefore, the *intermediate value theorem* implies existence of a solution to (9<sup>2</sup>)-(10<sup>2</sup>) and the *monotone nature* of  $\Gamma(\theta_r)$  and  $\Gamma(\theta_s)$  guarantees uniqueness.

**Appendix 3. Interior solution**

Note that we can rewrite the aggregate definition of  $\theta_r$  and  $\theta_s$  as the ratio of entrepreneurs-workers in each sector:

$$\frac{\theta_r \cdot [f(\theta_r) + \delta]}{g(\theta_r) + \delta} = \frac{F(R)}{G(Q)} \quad (3.1)$$

$$\frac{\theta_s \cdot [f(\theta_s) + (\delta + \rho)]}{g(\theta_s) + (\delta + \rho)} = \frac{[1 - F(R)]}{[1 - G(Q)]} \quad (3.2)$$

it is straightforward to prove that equations (3.1) and (3.2) are equal to zero (or infinite) if  $\theta_i$  is equal to zero (or infinite). But as shown in Appendix 1, when  $\theta_i = 0$  or  $\theta_i = \infty$  the value functions  $rV_i$  and  $rU_i$  are no longer functions of labour market tightness. Indeed, when  $\theta_i = 0$  the vacancies are immediately filled and when  $\theta_i = \infty$  the job-seekers are immediately employed. As a result, a corner solution can exist only in absence of search frictions in one of the two sectors.

# THE FINANCING OF R&D INVESTMENTS: EFFECTS ON GROWTH AND FINANCIAL STRUCTURE\*

Giuseppe GAROFALO, Patrizio MORGANTI  
Tuscia University of Viterbo, ITALY  
[garofalo@unitus.it](mailto:garofalo@unitus.it)  
[pat.morganti@gmail.com](mailto:pat.morganti@gmail.com)

## Abstract

*The financing of innovation impacts on economic growth. What's, among bank-based or market-based system, the best organizational form? The paper supports the view that both forms have to be merged into a unique approach (law and finance view) embracing not only the overall quality of the financial services' supply (financial services view) but also the related legal aspects.*

*In particular, the paper agrees with the thesis that the financial structure of the economy emerges endogenously depending both on the firms' R&D investment choices and on their initial wealth level.*

*The paper presents also an empirical validation of the financial services and, more specifically, of the law and finance view, carried out over the period 1980-2008. A focus is dedicated to the Chinese case, which shows strong economic growth and increasingly efficient financial sector, in despite of the weakness of the legal system.*

**Keywords:** R&D investment, finance and growth, financial system and institutions

**JEL Classification:** D82, E44, G20, K20, O32, P48

## 1. Introduction

The creation of a production technology is connected to firms' financial ability to invest on R&D. If firms are unable to self-support the investment, they have to raise funds from external finance (privates, banks or government). This type of fund-raising is embedded in endogenous growth models, which merge the typical real sector of the economy with the financial sector. Generally, finance and growth models do not assign to specific financial variables a driving role for growth; rather they are focused in identifying the channels through which finance can affect growth. Figure 1, taken from (Levine 1997), can help describing the finance and growth nexus.

Recent literature tends to distinguish among finance and growth models, depending on the:

- type of endogenous growth;
- type of financial structure;
- treatment of asymmetric information.

Main finance and growth models' features are summarized in Table 1.

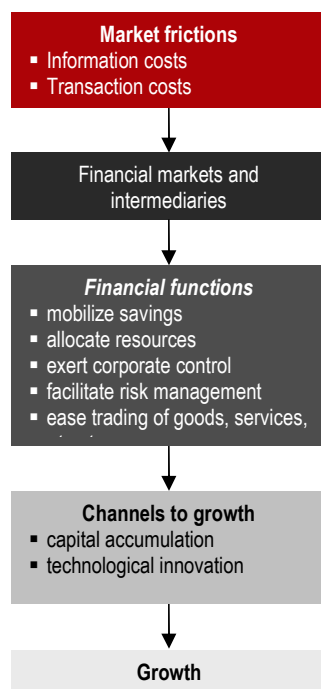
Finance and growth literature is still characterized by the classical dichotomy opposing the *bank-based view* to the *market-based view*.<sup>44</sup> The first argues that banks hold a leading role in stimulating economic growth as they provide more efficient services, such as savings mobilization and the reduction of agency problems, while the latter highlights financial markets' virtues (particularly large, diffused and liquid) in fostering risk management, corporate governance and in spreading market signals to investors.<sup>45</sup> Nevertheless, recent empirical studies have shown that these two traditional views have been overtaken by alternative *financial-services* and *law and finance* approaches, which emphasize the efficiency of a country's financial and legal institutions rather than focusing on its financial structure.

---

\* The authors would like to thank two anonymous referees for helpful comments.

<sup>44</sup> Allen and Gale, 2001 and Levine, 2002.

<sup>45</sup> Sau, 2008.



**Figure 1.** A theoretical approach to finance and growth

**Source:** Levine, 1997.

The financial-services view, introduced by Merton and Bodie (1995) and Levine (1997), stresses the abilities of financial institutions to provide efficient services and sets aside the *bank-based vs market-based systems* debate by sustaining that both systems should complement each other. The law and finance view, introduced by (La Porta *et al.* 1998, hereafter LLSV)<sup>46</sup> is a special case of the financial-services view that also considers the efficiency of legal institutions: by using investors protection and law enforcement indexes, the authors show that a country founded on efficient legal institutions and deeply rooted on rule of law can avoid agency problems and boost economic growth.

This work is organized as follows. In Sections 2 we introduce a theoretical model that specifies the finance and growth nexus, coherently with (Chakraborty, and Ray 2006). The main assumptions are:

- overlapping generations, with population consisting of households and firms;
- endogenous growth with Hicks-neutral technological spillovers (*following* Romer 1986);
- mixed financial structure;
- moral hazard treated as deceit.

We show that the financial structure of the economy emerges endogenously depending both on the firms' R&D investment choices and on their initial wealth level, and argue that the link between financial and real variables of the economy depends primarily on the financial institutions ability to deliver efficient services and to solve agency problems rather than on its financial structure.

Section 3 provides an econometric analysis where real GDP per-capita growth rates is regressed on macroeconomic, institutional and financial variables for a sample of 49 countries over the period 1980-2008, using the STATA software. The results are consistent with Levine's (2002) findings and support the financial-services and the law and finance view. Compared to Levine's analysis, which covers the period 1980-1995, our study is extended to 2008 and includes China, a controversial counter-

<sup>46</sup> LLSV study based on 49 countries (which were grouped by *common, civil, german and scandinavian* legal origins) analyses the quality of the legal system and law enforcement, and shows that common law countries provide the best investors protection and the highest law enforcement level, while civil law countries provide the worst investors protection and the lowest law enforcement quality. (Countries with german and scandinavian legal origins are placed in between both situations).

example that highlights a strong economic growth and an increasingly efficient of the financial sector, in despite of a weak legal system. Section 4 concludes.

**Table 1.** Core features of the main finance and growth models

Authors	Production externality Source of endogenous growth	Finance mechanism	Information problem
Bencivenga, and Smith (1991)	Production externalities	Insurance market and entrepreneurship	Exogenous liquidity shock
Saint (1992)	Production externalities	Capital market	Exogenous productivity shock
King, and Levine (1993)	Vertical innovation	Entrepreneurial funding, heterogeneous agents	Adverse selection ( <i>screening</i> )
Bose, and Cothren (1996)	Production externalities	Contract or screen heterogeneous entrepreneurs	Adverse selection ( <i>screening</i> )
De la Fuente, and Marin (1996)	Orizontal innovation	Funding and monitoring entrepreneurship	Moral hazard ( <i>effort adersion</i> )
Aghion, and Bolton (1997)	Capital accumulation	Capital market and income inequalities	Moral hazard ( <i>monitoring</i> )
Blackburn, and Hung (1998)	Horizontal innovation	Entrepreneurship, project appraisal and risk diversification	Moral hazard ( <i>deceit</i> )
De Gregorio, and Kim (1998)	Human capital accumulation	Credit markets	None
Harrison, Sussman, and Zeira (1999)	Production externalities	Banks and monitoring entrepreneurship	Moral hazard ( <i>monitoring</i> )
Morales (2003)	Vertical innovation and human capital accumulation	Entrepreneurship and screening	Moral hazard ( <i>effort adersion</i> )
Chakraborty, and Ray (2006)	Production externalities	Capital market, banks and monitoring entrepreneurship	Moral hazard ( <i>deceit</i> )
Aghion, Howitt, and Mayer-Foulkes (2005)	Vertical innovation	Entrepreneurship and credit constraints	Moral hazard ( <i>deceit</i> )
Blackburn, Bose, and Capasso (2005)	Production externalities	Entrepreneurship, markets and banks	Adverse selection and moral hazard
Trew (2005)	Human capital accumulation	Entrepreneurial funding, heterogeneous agents	Moral hazard ( <i>effort adersion</i> )

**Source:** Our elaboration of (Trew 2005).

**2. A model on the finance and growth nexus**

A small open economy consisting of a final sector, a traditional sector, a R&D sector and a financial sector it has been considered. Perfect capital mobility across borders assures free access to the international capital market: every agent can invest at the constant world gross rate return  $R^*$ .<sup>47</sup>

Time is discrete, continues forever and is indexed by  $t=1,2,\dots,\infty$ . A continuum of two-period lived agents (youth and old age) is born every period according to an overlapping generations' structure. Population does not grow with time. Each generation is composed by two kinds of agents:  $\mu$  households and  $(1-\mu)$  entrepreneurs.  $\mu$  is normalized to 1/2 whereas the number of households and entrepreneurs is normalized each to one. In the first period of life each agent is endowed with one unit of labour,  $z=1$ , which it supplies inelastically to the final sector (if agents are households), or to the traditional or R&D sectors (if agents are entrepreneurs). A generation- $t$  household gains a wage income of  $w_t$ , entirely saved and used for consumption in old age,  $c_{t+1}$ , with  $S_t = w_t$ . By supposing that households do not

<sup>47</sup>  $R^* = (1+r^*)$  is obtained from the international capital market. Agents take it as given.  $R^*$  is the return for both bank deposit and firm securities.

leave bequests to their offspring, a generation- $t$  household's lifetime utility depends only upon the old age consumption,  $U_{t+1}^F = u(c_{t+1}^F)$ , where  $c_{t+1}^F = R^* w_t$  is the consumption at date  $-t+1$  which equals the wage saved and invested at  $R^*$ .

Final sector is perfectly competitive and produce, by combining capital ( $K$ ) and labour ( $N$ ) inputs, a homogeneous consumption good by the following production technology:

$$Y_t = A_t K_t^\alpha N_t^{1-\alpha} \quad 0 < \alpha < 1 \quad (1)$$

Where  $A_t = A k_t^{1-\alpha}$  is the technological spillover (following [Arrow (1962)] and [Romer (1986)]),  $k_t = K_t/N_t = \left[ \int_{j \in E_t} K_t^j dG_t \right] / N_t$  denotes capital per worker and  $E_t$  is the set of entrepreneurs who supply capital goods at date- $t$ . The technological spillover enters the (1) in a Hicks-neutral way and firms know its entity only after they have fixed their inputs optimal set. The equilibrium rental ( $\rho$ ) and wage ( $w$ ) rates in the final sector are given by:

$$\alpha A = \rho_t \quad (2)$$

$$(1 - \alpha) A k_t = w_t \quad (3)$$

Contrary to households, entrepreneurs are altruistic and leave a part of their old age wealth to their offspring. A generation- $t$  entrepreneur's  $j$ , with  $j \in (0,1)$ , lifetime utility depends upon old age consumption,  $c_{t+1}$ , and bequests made,  $b_{t+1}$ . We formalize this altruism motive in the following “warm-glow”<sup>48</sup> utility function:

$$U_t^j = \beta u(c_{t+1}^j) + (1 - \beta) u(b_{t+1}^j) \quad \beta \in (0,1)$$

Where  $\beta$  is an objective parameter which denotes the entrepreneurial altruism. Given logarithmic preferences and denoting by  $z_{t+1}^j = c_{t+1}^j + b_{t+1}^j$  the entrepreneur's wealth at date- $t+1$ , entrepreneur- $j$  bequeaths a constant proportion of her realized old age income:

$$b_{t+1}^j = (1 - \beta) z_{t+1}^j \quad (4)$$

Introducing a wealth cumulative distribution function  $G_t(b)$ , which denotes the proportion of a generation- $t$  entrepreneurs with wealth less than  $b$ , eq. (2.4) tracks the wealth distribution through time given an initial wealth distribution  $G_0$  and the entrepreneurial income  $\{z_{t+1}^j\}_{t=0}^\infty$ . Hereafter we consider entrepreneurs as individual firms.

In youth, firms can invest their own resources in a traditional technology or in a modern R&D technology, depending on their initial wealth ( $b_t^j$ ). Traditional technology allows firms to produce, with a one-period lag, the same final good for an amount of:

$$x_{t+1} = a b_t^\delta \quad a > 0 \quad \text{And} \quad \delta \in (0,1) \quad (5)$$

Which is entirely self-consumed and it is not included in the national income accounts ( $a$  and  $\delta$  are productivity parameters).

---

<sup>48</sup> See Barro 1974, Galor, and Zeira 1993, and Amegashie 2006.



Research firms invest an amount  $q_t > b_t^j$ , raising the difference  $q_t - b_t^j$  from the financial sector.<sup>49</sup> A moral hazard problem arises from the fact that the success of the investment depends on an unobserved action taken by the firms. Firms can choose between three types of research projects: only the one with an efficient technology assure the success of the investment, while the others two, endowed of inefficient technology, have a  $0 < \pi < 1$  probability of success and offer private benefits to firms equal to  $V$  and  $v$ , which denote, respectively, the private benefit of the high and the low moral hazard project. By assuming  $0 < v < V < 1$ , firms clearly prefer the high moral hazard project over the low one because it requires a lower use of resources, equal to  $q_t - Vq_t$ , and implies an higher private benefits, equal to  $Vq_t$ . If firms succeed, they produce a capital amount of:

$$K_{t+1}^j = q_t \tag{6}$$

While if they fail, they produce nothing.

Initial wealth allows firms to operate in R&D by gaining access to:

- direct or market finance, made through the purchase of tradable securities (like corporate bonds and equities<sup>50</sup>) by households;

- mixed or indirect finance, made through the funding among banks and markets.

Direct finance has no monitoring activity<sup>51</sup>, firms are thus willing to invest in the high moral hazard project which gives them a higher private benefit. In the mixed finance banks are endowed with a monitoring technology which allows them to observe the borrower action and thus reduce the agency problem by removing the high moral hazard project. Financial contracts are drawn up if and only if incentive and participation constraints are simultaneously satisfied. An optimal contract is defined as the contract which induce firms to behave diligently (accept the project which succeed for sure) and neither party gains nothing if the research project fails.

The *direct finance optimal contract*, signed by households and firms, is formalized as follows:

$$\theta_{t+1}^E + \theta_{t+1}^U = \rho_{t+1} K_{t+1}^j \Rightarrow \theta_{t+1}^E + \theta_{t+1}^U = \rho_{t+1} q_t$$

where  $\theta_{t+1}^E > 0$  denotes the date- $t+1$  firm's rent,  $\theta_{t+1}^U > 0$ <sup>52</sup> denotes households' return at date- $t+1$  and  $\rho_{t+1}$  is the capital return rate at time  $t+1$  by renting out  $K_{t+1}$  to the final sector. A firm- $j$  will be diligent (choose the not risky project) if the following incentive constraint is satisfied:

$$\theta_{t+1}^E \geq \pi \theta_{t+1}^E + Vq_t \Rightarrow \theta_{t+1}^E \geq \frac{V}{(1-\pi)} q_t \tag{7}$$

Similarly, households will accept the contract if the following participation constraint is satisfied:

$$\left[ \rho_{t+1} - \frac{V}{(1-\pi)} \right] q_t \geq R^* [q_t - b_t^j] \tag{8}$$

Since only internal solutions of the model are relevant for this study, eq. (8) is solved for  $b_t^j$  by forcing the equal sign:<sup>53</sup>

<sup>49</sup> We omit the  $j$ -superscript by supposing that a generation- $t$  firms, who supply capital goods, invest the same amount  $q_t$  in R&D.

<sup>50</sup> We assume that corporate bonds and equity shares give the same return  $R^*$  and that does not exists a secondary securities market because capital fully depreciate after one period.

<sup>51</sup> Households have not a monitoring technology because it is too expensive and it needs specific technical skills for its application.

<sup>52</sup> Superscript  $U$  stands for households as '*uninformed investors*' because they have imperfect information. On the contrary, banks are the '*informed investors*' because they have the monitoring technology.

$$b_t^H(q_t) = \left[ 1 - \frac{1}{R^*} \left( \rho_{t+1} - \frac{V}{1-\pi} \right) \right] q_t \quad (9)$$

Where  $b_t^H$  is the minimum amount of initial wealth that firms need to have access to direct finance.<sup>54</sup>

The *mixed finance optimal contract*, formalized in the expression  $\theta_{t+1}^E + \theta_{t+1}^U + \theta_{t+1}^B = \rho_{t+1} q_t$  with  $\theta_{t+1}^B$  as the bank return, it is drawn up if and only if are simultaneously satisfied:

- firms incentive constraint,  $\theta_{t+1}^E \geq \pi \theta_{t+1}^E + v q_t \Rightarrow \theta_{t+1}^E \geq \frac{v}{(1-\pi)} q_t$ ;

- banks incentive constraint given by  $\theta_{t+1}^B - R^* \gamma q_t \geq \pi \theta_{t+1}^B \Rightarrow \theta_{t+1}^B \geq \frac{R^* \gamma}{(1-\pi)} q_t$ , where  $\gamma$  is the

constant and unitary cost of monitoring;

- households participation constraint, given by  $\left[ \rho_{t+1} - \left( \frac{v + R^* \gamma}{1-\pi} \right) \right] q_t \geq R^* M_t^j$ , where

$M_t^j = (q_t - b_t^j - L_t^j)$  is the balance that households must supply to firms if they take part to the contract and  $L_t^j = q_t \gamma \frac{\pi}{1-\pi}$  is bank credit to firm- $j$ .

We assume:

$$\gamma \leq \frac{(1-\pi)}{\pi} \quad (\text{Assumption 1})$$

So that the monitoring cost is upper bounded; on the contrary, high value of  $\gamma$  does not allow banks to solve moral hazard.

By supposing competitive banks, with profits equal to  $\Pi_{t+1}^B = R_{t+1}^L L_t - R^* D_t$ , the first order condition of the bank's optimization problem is given by:

$$R_{t+1}^L = \frac{R^*}{\pi} > R^* \quad (10)$$

Which equals the loan rate charged to borrowers,  $R_{t+1}^L$ , to the deposit rate  $R^*$  weighted for  $1/\pi$ . Combining the previous expressions we obtain the minimum level of initial wealth that firms need to have access to mixed finance:

$$b_t^L(q_t) = \left[ 1 + \gamma - \frac{1}{R^*} \left( \rho_{t+1} - \frac{v}{1-\pi} \right) \right] q_t \quad (11)$$

It is reasonable to assume that  $b_t^H(q_t) \geq b_t^L(q_t)$ , otherwise there will be no demand for financial intermediation (monitoring would be too costly to be socially useful). This condition is also realistic because bank credit is more accessible than market finance. Matching (9) and (11) one can obtain:

<sup>53</sup> This procedure will be also extended to next constraints.

<sup>54</sup> Eq. (9) can be seen as the critical level of initial wealth necessary to get access to the stock exchange.

$$b_t^H(q_t) \geq b_t^L(q_t) \Leftrightarrow \frac{V-v}{1-\pi} \geq R^* \gamma \quad (\text{Assumption 2})$$

Always true because of  $0 < v < V < 1$  and  $0 < \pi < 1$ .

Therefore, depending on the initial wealth level, a firm- $j$  income in its second period of life is given by:

$$z_{t+1}^j(b_t^j | q_t) = \begin{cases} a(b_t^j)^\delta, & \text{if } b_t^j \in [0, b_t^L(q_t)), \\ [\rho_{t+1} - R^*(1+\gamma)]q_t + R^*b_t^j, & \text{if } b_t^j \in [b_t^L(q_t), b_t^H(q_t)), \\ (\rho_{t+1} - R^*)q_t + R^*b_t^j, & \text{if } b_t^j \in [b_t^H(q_t), \infty). \end{cases} \quad (12)$$

The properties of optimal loan contracts and external financing choices are summarized in the following:

**Proposition 1.** Given  $q_t > b_t^j$  and the initial wealth distribution  $G_t(b)$ ,

(i) firms with  $b_t^j < b_t^L(q_t)$  are unable to obtain any external finance and operate in the traditional production sector;

(ii) firms with  $b_t^j \in [b_t^L(q_t), b_t^H(q_t))$  have access to mixed or indirect finance: they borrow from banks an amount  $L_t^j = q_t \gamma \frac{\pi}{1-\pi}$  at the loan rate  $R_{t+1}^L$ , agree to being monitored, and raise the remaining funds  $M_t^j = (q_t - b_t^j - L_t^j)$  directly from households at the rate  $R^*$ . Optimal contracts guarantee these firms incentive compatible payments such that they behave diligently (invest on the efficient technology);

(iii) firms with  $b_t^j \geq b_t^H(q_t)$  borrow only from households (direct finance), paying them a return of  $R^*$ . Likewise the previous point, incentive compatible payments to these firms ensure that investments succeed for sure;

(iv) firms income in each case is given by (12).

Firms which invest in R&D are financially constrained and have to rely on external finance. By defining  $q^*$  as the minimum investment size for firms wishing to operate in R&D,<sup>55</sup> the number of these firms is endogenously determined by their access to outside finance and so by their capacity to overcome the R&D's financial constraints. The level  $q^*$  also defines a minimum level of initial wealth,  $b_t^*$ , that a firm needs to qualify at least for bank credit. This level is obtained by combining (2) and (11):<sup>56</sup>

$$b_t^* = b_t^* = \frac{q^*}{R^*} \left[ R^*(1+\gamma) - \rho_{t+1} + \frac{v}{1-\pi} \right] = \frac{q^*}{R^*} \left\{ \frac{v}{1-\pi} - [\alpha A - R^*(1+\gamma)] \right\} \quad (13)$$

A firm- $j$  with  $b_t^j < b_t^*$  must work in the traditional sector earning an income  $z_{t+1}^j = a(b_t^j)^\delta$  (depending on (12)), and bequeaths a  $(1-\beta)$  fraction of his income to the next generation. Substituting this income into (4), the asset accumulation for such firms is given by  $b_{t+1}^j = (1-\beta)a(b_t^j)^\delta$  and it follows that their long-run wealth level converges to:

<sup>55</sup>  $q^*$  denotes the entry barrier of the R&D sector.

<sup>56</sup> The amount  $q^*$  guarantee the access to the minimum bank credit (the wealth cut-off is lower than that of the direct finance) for working in R&D.

$$\frac{b_{t+1}^j}{b_t^j} = b_C = [a(1 - \beta)]^{\frac{1}{1-\delta}} \quad (14)$$

By defining with  $G_0(b_L^*)$  the fraction of generation-0 firms with assets less than  $b_L^*$ , it can be stated that:

**Proposition 2.** *Given the parameters set  $(\gamma, v, V, R^*)$ , the initial wealth distribution  $G_0(b)$  determines which firms can work in the research sector. A constant fraction of them, given by  $G_0(b_L^*)$ , will never be able to obtain external finance, engaging in traditional sector activities instead: in the long run their wealth converges to the steady-state level  $b_C$  given by (14).*

Hereafter  $G_0(b_L^*)$  denotes the set of financially constrained firms which have not access to external finance.

The access to the financial system is granted to those firms which have a level of initial wealth higher than  $b_L^*$ . Given optimal contracts arrangement, the optimal investment size  $q_t$  is chosen by a firm as to maximize his income. Substituting (2) into (11) and into (9) one can obtain the minimum amount of internal funds required to qualify firms for mixed and direct finance.<sup>57</sup>

$$b^L(q_t) = \frac{q_t}{R^*} \left\{ \frac{v}{1-\pi} - [\alpha A - R^*(1+\gamma)] \right\} \quad (15)$$

$$b^H(q_t) = \frac{q_t}{R^*} \left[ \frac{V}{1-\pi} - (\alpha A - R^*) \right] \quad (16)$$

In Figures 2 and 3, which explain firms investment decision, equations (15) and (16) are shown by two increasing lines passing through the origin. Since  $b^H(q_t) > b^L(q_t)$  for any  $q_t$  (Assumption 2), the  $b^H(q_t)$  line is steeper than  $b^L(q_t)$ . The intersection points of these lines with  $q^*$  are labelled  $b_H^*$  and  $b_L^*$  respectively.

---

<sup>57</sup> Note that (15) and (16) are similar to (11) and (9) but formally differ for the meaning of  $q$ : in (11) and (9)  $q^*$  is the minimum investment size that a firm must bear to operate in the research sector, while (15) and (16) denote the optimal investment decision and so the amount of resources that a firm wants to invest in R&D (this amount is obviously equal or higher than  $q^*$ ). Conditions from footnotes 19 and 20 (*infra* Section 2.1) guarantee that  $b^L(q_t) > 0$  and  $b^H(q_t) > 0$  for any  $q_t$ .

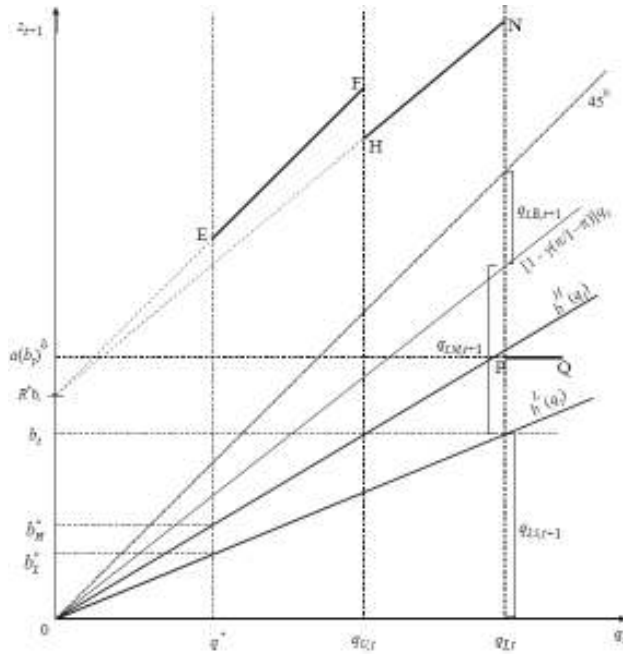


Figure 2. Investment choice in a bank-based system

Source: Chakraborty, and Ray 2006.

Description:

$q_{I,B}$  – bank loans in mixed finance;

$q_{I,M}$  – market funding in mixed finance;

$q_{I,S}$  – self-financing in mixed finance.

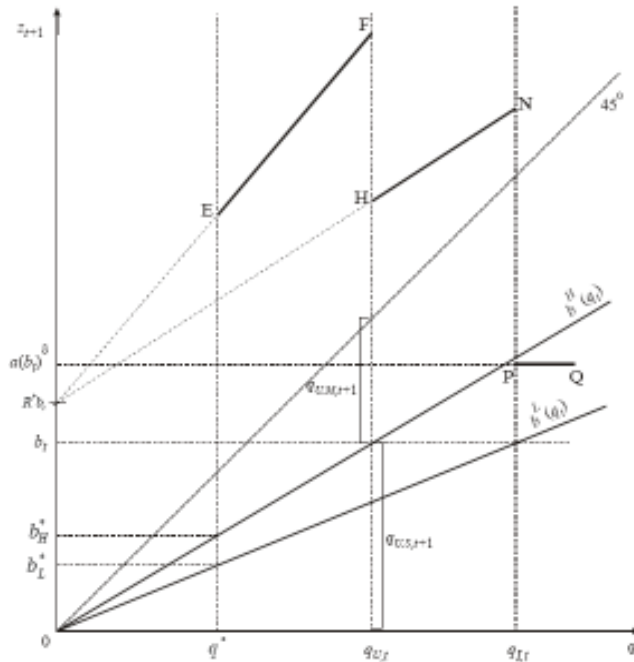


Figure 3. Investment choice in a market-based system

Source: Chakraborty, and Ray 2006.

Description:

$q_{U,M}$  – market funding in direct finance;

$q_{U,S}$  – self-financing in direct finance.

Omitting the  $j$ -superscript, consider a generation- $t$  firm with initial wealth  $b_t > b_t^*$ . In both figures  $q_{I,t}$  e  $q_{U,t}$ <sup>58</sup> are given by the intersection points of  $b^L(q_t)$  and  $b^H(q_t)$  with the firm's initial wealth  $b_t$ .

From (12), firms have three possible choices of investment:

1. If  $b_t < b^L(q_t)$  for any  $q_t > q_{I,t}$ , firms desiring to invest resources more than  $q_{I,t}$  do not have access to external finance because they are completely rationed from credit market and cannot convince households to invest in their securities. They have to work in the traditional sector and earn an income  $z_{t+1} = ab_t^\delta$ , represented by the horizontal line PQ in both figures.

2. Likewise,  $b^L(q_t) \leq b_t < b^H(q_t)$  for any  $q_{U,t} < q_t < q_{I,t}$ . If firms choose an investment size in this range, they can convince households to fund her project only if they simultaneously borrow from banks. Their income is  $z_{t+1} = R^*b_t + [\alpha A - R^*(1 + \gamma)]q_t$ , shown by the HN line with intercept equal to  $R^*b_t$  and slope coefficient equal to  $[\alpha A - R^*(1 + \gamma)]$ .

3. Firms with initial wealth  $b_t \geq b^H(q_t)$ , only for any  $q^* \leq q_t \leq q_{U,t}$  can funding directly from the financial markets without requiring any bank finance. Their income, given by  $z_{t+1} = R^*b_t + [\alpha A - R^*]q_t$ , is represented by EF line, with the same intercept of HN line but more steeper.

To summarize, when a firm chooses her optimal level of investment, her earnings are given by:

$$z_{t+1}(q_t|b_t) = \begin{cases} R^*b_t + (\alpha A - R^*)q_t, & \text{if } q_t \in [q^*, q_{U,t}] \\ R^*b_t + [\alpha A - R^*(1 + \gamma)]q_t, & \text{if } q_t \in (q_{U,t}, q_{I,t}] \\ ab_t^\delta, & \text{if } q_t \in (q_{I,t}, \infty) \end{cases} \quad (17)$$

Eq. (17) differs from (12) because in the first firms choose their optimal level of investment, given their initial wealth; while in the latter firms determine their minimum investment size. A firm will choose  $q_t$  as to maximize  $z_{t+1}(q_t)$ . This optimization problem can be graphically solved by using Figures 2 and 3.

In both figures, income is given by linear schedule EF, HN and PQ. Two possibilities arise:  $q_{I,t}$  is the investment choice when the height of the point N is greater than that of point F (Figure 2), whereas  $q_{U,t}$  is chosen when the opposite holds (Figure 3).

Closed-form solutions for these investment levels can be obtained using (15) and (16):

$$q_{I,t} = \left\{ \frac{R^*}{v/(1-\pi) - [\alpha A - R^*(1 + \gamma)]} \right\} b_t, \quad q_{U,t} = \left\{ \frac{R^*}{v/(1-\pi) - (\alpha A - R^*)} \right\} b_t \quad (18)$$

Since  $z_{t+1}(q_t)$  is strictly increasing in the range  $q_t \in [q^*, q_{U,t}]$ , maximal earning occurs at  $q_t = q_{U,t}$  which, substituted in the first eq. of (2.17), is given by:

<sup>58</sup> Subscripts  $I$  e  $U$  stand for mixed finance (*intermediated*) and direct finance (*unintermediated*).

$$z_{t+1}(q_{U,t}) = \left\{ \frac{[V/(1-\pi)]R^*}{V/(1-\pi) - (\alpha A - R^*)} \right\} b_t \quad (19)$$

Likewise, since  $z_{t+1}(q_t)$  is strictly increasing in the range  $q_t \in (q_{U,t}, q_{I,t}]$ , the maximum earning in this range occurs at  $q_t = q_{I,t}$  which, substituted in the second eq. of (17), and is given by:

$$z_{t+1}(q_{I,t}) = \left\{ \frac{[v/(1-\pi)]R^*}{v/(1-\pi) - [\alpha A - R^*(1+\gamma)]} \right\} b_t \quad (20)$$

The type of financial system emerges endogenously depending on entrepreneurial preferences for external finance access. Firms prefer to apply to mixed finance if and only if:

$$\begin{aligned} z_{t+1}(q_{I,t}) &\geq z_{t+1}(q_{U,t}) \\ \Leftrightarrow \frac{\alpha A - R^*(1+\gamma)}{v} &\geq \frac{\alpha A - R^*}{V} \end{aligned} \quad (21)$$

Which does not depend upon firm characteristics.

Firm-financing decisions can thus affect the financial structure of the economy. As long as (21) holds, every firm applies to indirect finance (except for the fraction  $G_0(b_L^*)$ , which is completely credit-rationed)<sup>59</sup>: this is a *bank-based financial system* (Figure 2).

On the other hand, if (21) does not hold, firms earn a higher income in a *market-based* regime (Figure 3). Despite this, one group of firms has to rely upon bank-finance at least in the short-run. To see this, consider a firm with  $b_L^* \leq b_t < b_H^*$  initial wealth: with the new investment level  $q_{U,t}$  and  $q_{I,t}$ , one would have  $q_{U,t} < q^* < q_{I,t}$ . Since (21) is not satisfied, this firm's new income would be maximized for the investment level  $q_{U,t}$ ; but such level would not be permissible since  $q_{U,t} < q^*$ , hence the only one solution is to invest the  $q_{I,t}$  amount.

Therefore, if (21) is not satisfied, firms in the range  $b_t \in [b_L^*, b_H^*)$  have to rely upon mixed finance (at least in short-run), whereas those with  $b_t \geq b_H^*$  use only market finance. This reliance on mixed finance is, however, temporary because wealth of firms with  $b_t \in [b_L^*, b_H^*)$  grows at rate  $g_I$ <sup>60</sup> and can eventually cross  $b_H^*$ : thereafter their wealth grows at rate  $g_U$  of the market finance.<sup>61</sup> In long-run all firms with  $b_t \geq b_L^*$  have to rely only upon direct finance (Figure 3).

From (21) and depending on the parameters set  $(\alpha, A, \gamma, v, V, R^*)$ , a bank-based or market-based financial system may thus emerge endogenously. Given the importance of results, condition (21) can be rewritten as follows:

**Proposition 3.** *The financial structure of an economy is bank-based if and only if*

$$1 - \frac{\gamma R^*}{(\alpha A - R^*)} \geq \frac{v}{V}$$

<sup>59</sup> *Supra* Proposition 2.

<sup>60</sup> *Infra* eq. (23).

<sup>61</sup> *Infra* eq. (25).

Otherwise it present a market-based structure without any dependence on indirect finance in the long-run, except for those firms with low initial wealth ( $b_t \in [b_L^*, b_H^*]$ ) who have to rely, in the short-run, upon intermediated finance even in a market-based system.

One can see that the financial structure is likely to be bank-based whenever the cost of monitoring is low and whenever the private benefit of low moral hazard project (under bank monitoring) is low relatively to the private benefit of high moral hazard project (absence of monitoring).

### 2.1. Long-run economic growth

This Section provides the dynamic equilibrium of the model by determining the balanced growth paths of wealth, capital, GDP and consumption. Starting from a bank-based economy, wealth at date- $t+1$  for a credit-unconstrained firm- $j$  ( $b_t^j \geq b_L^*$ ) is obtained by substituting (20) into (4):

$$b_{t+1}^j = (1 + g_I) b_t^j \quad (22)$$

where  $g_I$  is the wealth constant growth rate for firms who have to rely upon indirect or mixed finance, which is given by:<sup>62</sup>

$$1 + g_I = (1 - \beta) \left\{ \frac{[v/(1 - \pi)]R^*}{v/(1 - \pi) - [\alpha A - R^*(1 + \gamma)]} \right\} \quad (23)$$

Likewise in a market-based economy, combining (19) and (4) we have:

$$b_{t+1}^j = (1 + g_U) b_t^j \quad (24)$$

Where  $g_U$  is the wealth constant growth rate for firms who apply to direct finance, which is equal to:<sup>63</sup>

$$1 + g_U = (1 - \beta) \left\{ \frac{[V/(1 - \pi)]R^*}{V/(1 - \pi) - (\alpha A - R^*)} \right\} \quad (25)$$

Firms optimal investment choices are linear in their wealth as (18) shows. The aggregate capital produced in the economy at date- $t+1$  depends upon the volume of investments undertaken by research firms at date- $t$ , so one can define:

$$\bar{q}_{I,t} = \int_{b_L^*}^{\infty} q_{I,t}^j dG_t \quad \text{And} \quad \bar{b}_t = \int_{b_L^*}^{\infty} b_t^j dG_t$$

Respectively as the total amount of R&D investment at period- $t$  and the wealth of all credit-unconstrained firms at period- $t$  (which apply at least to indirect finance). Since optimal contracts assure firms to behave diligently, using (18) and (6) the aggregate and per-capita capital produced in a bank-based and a market-based system is given by:

$$k_{I,t+1} = \bar{q}_{I,t} = \left\{ \frac{R^*}{v/(1 - \pi) - [\alpha A - R^*(1 + \gamma)]} \right\} \bar{b}_t$$

$$k_{U,t+1} = \bar{q}_{U,t} = \left\{ \frac{R^*}{V/(1 - \pi) - [\alpha A - R^*]} \right\} \bar{b}_{U,t}$$

<sup>62</sup> To assure  $g_I > 0$  we assume:  $[\alpha A - R^*(1 + \gamma)] \leq v/(1 - \pi) \leq [\alpha A - R^*(1 + \gamma)]/[1 - R^*(1 - \beta)]$ .

<sup>63</sup> To assure  $g_U > 0$  we assume:  $(\alpha A - R^*) \leq V/(1 - \pi) \leq (\alpha A - R^*)/[1 - R^*(1 - \beta)]$ .



Combining previous expressions with, respectively, (22) and (24), one can obtain that capital per-capita grows at the same wealth rate for each financial regime:

$$k_{I,t+1} = (1 + g_I)k_{I,t} \quad \text{And} \quad k_{U,t+1} = (1 + g_U)k_{U,t}$$

GDP growth rate, either for a bank-based or a market-based system, mimes exactly the capital accumulation process because the intensive production function, given by  $Y_t/N_t = A(K_t/N_t) \rightarrow y_t = Ak_t$  is linear in capital per-capita. Per-capita GDP grows at the same  $g_I$  or  $g_U$  rate depending on the financial structure of the economy:

$$y_{I,t+1} = (1 + g_I)y_{I,t}, \quad y_{U,t+1} = (1 + g_U)y_{U,t}$$

Consider now consumption paths for each type of agent. For households, the equilibrium wage rate is linear in capital per-capita, as shows (3), while their second period consumption is given by  $R^*w_t$ : hence per-capita consumption grows at the same capital rate ( $g_I$  or  $g_U$ ) for each type of financial structure. On the other hand, the conditions that allow firms to rely upon external finance create differences among their consumption and wealth profiles. A fraction  $G_0(b_L^*)$  of them works in the traditional sector, with a wealth convergent to the steady-state level of (14) and a per-capita consumption also convergent to the steady-state value of  $\beta a(b_C)^\delta$ <sup>64</sup>. The remainder firms, which have access to external finance, enjoy the  $g_I$  or  $g_U$  per-capita consumption growth rate.

The dynamic equilibrium of the model is thus characterized by constant growth rates of GDP, capital per-capita and consumption; furthermore, since the composition of population is constant, GDP and capital per-capita grow at the same rate, while per-capita consumption does not. This is connected to firms' sectorial investment choices: households and research firms have the same consumption growth rate, while traditional firms do not. Income levels of the two entrepreneurial classes diverge along the balanced growth path.

This theoretical model attempts to specify the finance and growth nexus aiming to the complementarity and to the efficiency of financial structures in order to solve agency problems and ease savings mobilization. Growth rates, defined by (23) and (25), depend negatively by the parameters  $\nu$ ,  $V$  and  $\gamma$ , which denote the intensity of moral hazard and bank monitoring. To conclude, the whole model supports the financial-services view even if it has been demonstrated that bank-based systems foster better wealth distribution and structural changes of the economy in comparison than market-based systems do.

### 3. An empirical analysis

Levine's (2002) study, which is based on a sample of 48 countries over the period 1980-1995, is surely one of the most interesting works supporting the financial-services and law and finance view: the author regresses real GDP per-capita growth rates on macroeconomic, institutional and financial variables. Combining financial variables, Levine sets up structure and development indexes. The first assess the countries prevailing orientation (bank or market-based) while the latter denote the overall efficiency of the financial sector.

By adopting the same methodology, an extension of the analysis is proposed by:<sup>65</sup>

- widening the time period, from 1980 to 2008;
- including China in the sample;
- constructing new macroeconomic and institutional indicators from alternative sources, due to the lack and heterogeneity of data.

Tables 2, 3 and 4 respectively report financial variables and relative structure and development indexes.

<sup>64</sup> *Supra* eq. (12).

<sup>65</sup> We exclude *structure-aggregate* and *finance-aggregate* because we do not know the estimate method adopted by Levine. *Structure-regulatory* is taken from (Levine 2002), except for China, whose value is taken from (Allen *et al.* 2005).

Two information sets are distinguished for macroeconomic and institutional variables (Table 5):<sup>66</sup>

- *simple set*, consisting of per-capita GDP and schooling in 1980;
- *full set*, consisting of simple set plus the macroeconomic subset which contains inflation rate, black market premium, government expenditure and international trade, and the institutional subset which contains indicators of bureaucracy efficiency, political stability and corruption.

About Tables 6 and 7, which contain the ranked financial indexes of our analysis, potential anomalies have to be pointed out. First of all some countries can register high scores of financial structure either because they have well developed financial markets or because they have an inefficient banking system. On the contrary, other countries can register low scores of financial structure either because they have a wide and efficient banking sector or because they have very small and under-developed financial markets.

**Table 2.** Financial Variables

Variables	Description	Source
<i>Bank credit ratio</i> (PRIV_new)	Average, over the period 1980-2007, of the value of deposit money banks credits to private sector as a share of GDP.* GDP is adjusted for the consumer price index (CPI).	Beck, Demirgüç-Kunt, and Levine (2008)
<i>Private credit ratio</i> (PRIVO_new)	Average, over the period 1980-2007, of the value of financial intermediary credits to the private sector as a share of GDP.* GDP is adjusted for CPI.	Beck, Demirgüç-Kunt, and Levine (2008)
<i>Market capitalization ratio</i> (MCAP_new)	Average, over the period 1980-2007, of the value of domestic equities listed on domestic exchanges as a share of GDP. GDP is adjusted for CPI.	Beck, Demirgüç-Kunt, and Levine (2008)
<i>Total value traded</i> (TVT_new)	Average, over the period 1980-2007, of the value of domestic equities traded on domestic exchanges as a share of GDP.** It is usually used to measure markets liquidity.	Beck, Demirgüç-Kunt, and Levine (2008)
<i>Overhead costs</i> (OVERHEAD_new)	Average, over the period 1980-2007, of the overhead costs of the banking system to total banking system assets ratio. This measure indirectly reflects the efficiency of the banking sector: high values stand for high inefficiencies.	Beck, Demirgüç-Kunt, and Levine (2008)

\* This measure excludes credits to public sector (central and local government as well as public enterprises).

\*\* Since the value of domestic shares traded on domestic exchanges is computed over the year and since GDP is also computed over the year, it is not necessary to make the same deflation adjustment for the CPI.

**Table 3.** Financial Structure Indexes

Index	Description
Structure-Activity (struct_act_new)	Ln (TVT/PRIV). It measures the activity of stock markets relatively to that of banks. Higher values of the index imply a more market-based financial system.
Structure-Size (struct_size_new)	Ln (MCAP/PRIV). It measures the size of stock markets relatively to that of banks. Higher values of the index imply a more market-based financial system.
Structure-Efficiency (struct_eff_new)	Ln (TVT*OVERHEAD). It measures the efficiency of stock markets relatively to that of banks. Higher values of the index imply a more market-based financial system.
Structure-Regulatory (structure_regulatory)	It is an aggregate index which measures the regulatory restrictions on commercial bank activities with regard to: <ul style="list-style-type: none"> <li>▪ allowing bank to engage in securities, insurance and real estate activities;</li> <li>▪ allowing bank to own and control nonfinancial firms.</li> </ul>

<sup>66</sup> The complete dataset is available on request.

**Table 4.** Financial Development Indexes

Index	Description
Finance-Activity (finance_act_new)	Ln (TVT*PRIVO). It measures the activity of stock markets and intermediaries. Higher values of the index signify a greater activity of the financial sector.
Finance-Size (finance_size_new)	Ln (MCAP+PRIVO). It measures the size of stock markets and intermediaries. Higher values of the index signify a greater size of the financial sector.
Finance-Efficiency (finance_eff_new)	Ln (TVT/OVERHEAD). It measures the financial sector efficiency. Higher values of the index signify a greater efficiency of the financial sector.

**Table 5.** Macroeconomic and institutional variables

Variables	Description	Source
Trade (LTRADE_new)	Average, over the period 1980-2003, of Ln ((exports+imports)/GDP).	Global Development Network Growth Macro Time Series, DRI, NY University
Corruption (CORRUPT_new)	Average, over the period 1995-08, of the <i>Corruption Perceptions Index</i> . It ranges from 0, high corruption, to 10, low corruption.	Transparency International
Investor rights (OUTRIGHTS_new)	An aggregate index given by the sum of the creditor rights index and antidirector rights index.	DMS (2007) for creditor rights and DLLS (2007) for antidirector rights
Bureaucratic efficiency (BUREAU_new)	It measures the quality of the public bureaucracy, and it ranges from 0, low efficiency, to 10, high efficiency (we changed the scale from its original range going from -2.5 to +2.5). Data, taken from the <i>Government Effectiveness</i> index of the <i>Worldwide Governance Indicators</i> (WGI), are averaged over the period 1996-07.	WGI
Initial schooling (LSCHOOL80)	Ln (years of schooling in the working age population) in 1980.	Barro, and Lee (1996)
Initial GDP (LRGDPSH)	Ln (real per capita GDP) in 1980.	Levine (2002)
Black market premium (LBMP_new)	Ln (1+ black market premium averaged over the period 1980-1999).	Global Development Network Growth Macro Time Series, DRI, NY University
Rule of law (LAW_new)	Average, over the period 1996-2007, of the <i>Rule of law</i> index taken from WGI. It is an assessment of a country's legal and order tradition. It ranges from 0, low legal and order tradition, to 10, high law tradition (we changed the scale from its original range going from -2.5 to +2.5).	WGI
Government expenditure (LGOV_new)	Logarithm of the average (over the period 1980-04) of government expenditures as a share of GDP.	Penn World Tables 6.2.
Stabilità politica e assenza di violenza/terrorismo (STABILITY_new)	Average, over the period 1996-2007, of the <i>Political Stability &amp; Absence of Violence/Terrorism</i> index taken from WGI. It ranges from 0 to 10 (we changed the scale from its original range going from -2.5 to +2.5). Higher values of the index imply a better stability and a low degree of crime.	WGI
Average GDP growth rate (NGDPRPC80-08)	Average growth rate (over the period 1980-08) of per capita GDP at constant prices (NGDPRPC).	World Economic Outlook 2009
Inflation (LPI_new)	Logarithm of the average end of period consumer prices (PCPIEPCH) over the period 1980-2008.	World Economic Outlook 2009

Table 6. Ranked Structure Indexes

Structure Activity		Structure Size		Structure Efficiency		Structure-Regulatory	
United States	1.00	South Africa	1.06	Switzerland	-2.59	New Zealand	4
Taiwan	0.77	Zimbabwe	1.01	United States	-3.00	Austria	5
Turkey	0.46	Ghana	0.91	U.K.	-3.49	Germany	5
Sweden	0.33	Jamaica	0.89	Taiwan	-3.52	Switzerland	5
Switzerland	0.19	United States	0.75	Spain	-3.70	United Kingdom	5
Pakistan	0.15	Trin. & Tob.	0.47	Sweden	-3.74	France	6
Finland	0.13	Sweden	0.45	South Africa	-3.93	Netherlands	6
India	0.04	Finland	0.33	Australia	-4.01	Argentina	7
U.K.	0.01	Argentina	0.28	Turkey	-4.06	Canada	7
Spain	-0.07	Malaysia	0.25	Finland	-4.20	Finland	7
Netherlands	-0.19	Australia	0.25	Germany	-4.20	Philippines	7
Australia	-0.22	U.K.	0.25	France	-4.23	Spain	7
South Africa	-0.22	Chile	0.22	Canada	-4.25	Sri Lanka	7
Canada	-0.51	Switzerland	0.21	Netherlands	-4.28	Australia	8
Norway	-0.52	Belgium	0.18	Italy	-4.38	Cyprus	8
France	-0.61	Mexico	0.15	Denmark	-4.39	Denmark	8
Malaysia	-0.63	Philippines	0.14	Brazil	-4.57	Ireland	8
Italy	-0.63	Peru	0.07	Japan	-4.70	Norway	8
China	-0.75	Canada	0.02	Norway	-4.72	Panama	8
Denmark	-0.78	India	0.01	Israel	-4.85	Peru	8
Israel	-0.82	Turkey	-0.01	Malaysia	-4.87	South Africa	8
Germany	-0.83	Egypt	-0.15	Pakistan	-4.88	Belgium	9
Mexico	-0.86	Taiwan	-0.16	India	-5.00	Greece	9
Greece	-0.88	Israel	-0.17	Greece	-5.10	Honduras	9
Ireland	-0.94	Netherlands	-0.18	China	-5.12	Portugal	9
Japan	-0.94	Kenya	-0.19	Thailand	-5.14	Sweden	9
Brazil	-1.06	Ireland	-0.20	Cyprus	-5.22	Thailand	9
Thailand	-1.14	Colombia	-0.30	Mexico	-5.29	Trin. & Tob.	9
Philippines	-1.15	France	-0.33	Philippines	-5.58	Brazil	10
Zimbabwe	-1.23	Denmark	-0.34	Belgium	-5.62	Colombia	10
Belgium	-1.23	Greece	-0.34	Portugal	-5.70	India	10
Egypt	-1.34	Brazil	-0.35	Zimbabwe	-5.73	Italy	10
Cyprus	-1.54	Norway	-0.41	Ireland	-5.83	Kenya	10
New Zealand	-1.73	Sri Lanka	-0.44	Austria	-5.86	Malaysia	10
Argentina	-1.80	Spain	-0.45	New Zealand	-5.86	Pakistan	10
Peru	-1.83	Pakistan	-0.57	Chile	-5.93	Chile	11
Portugal	-1.90	Japan	-0.59	Argentina	-5.98	Ghana	12
Chile	-1.99	Italy	-0.62	Peru	-6.15	Jamaica	12
Jamaica	-2.09	New Zealand	-0.62	Jamaica	-6.21	Mexico	12
Austria	-2.32	Cyprus	-0.68	Egypt	-6.48	Taiwan	12
Sri Lanka	-2.43	Thailand	-0.87	Colombia	-6.76	Turkey	12
Honduras	-2.46	Germany	-0.98	Honduras	-6.98	United States	12
Ghana	-2.52	China	-0.99	Trin. & Tob.	-7.20	Egypt	13
Colombia	-2.52	Panama	-1.05	Sri Lanka	-7.22	Israel	13
Trin. & Tob.	-2.71	Ecuador	-1.07	Kenya	-7.31	Japan	13
Kenya	-2.83	Honduras	-1.10	Ghana	-7.87	Zimbabwe	14
Tunisia	-3.75	Portugal	-1.41	Ecuador	-7.96	Cina	16
Ecuador	-3.97	Austria	-1.53	Tunisia	-8.13	Ecuador	ND
Panama	-4.77	Tunisia	-1.58	Panama	-9.00	Tunisia	ND

Table 7. Ranked Finance Indexes

Finance Activity		Finance Size		Finance Efficiency	
Switzerland	0.99	Switzerland	1.20	Taiwan	4.85
Taiwan	0.65	South Africa	0.93	Netherlands	4.10
United States	0.64	United States	0.90	Switzerland	3.77
Netherlands	0.31	Netherlands	0.88	United States	3.68
U.K.	0.03	U.K.	0.84	Japan	3.59
Japan	-0.17	Japan	0.83	Malaysia	3.59
Sweden	-0.24	Malaysia	0.82	Finland	3.54
Spain	-0.37	Canada	0.64	U.K.	3.53
Malaysia	-0.64	Sweden	0.63	Ireland	3.33
Canada	-0.65	Taiwan	0.56	China	3.28
Finland	-0.79	Finland	0.41	Spain	3.26
Germany	-0.82	Ireland	0.36	Sweden	3.24
South Africa	-0.88	Australia	0.36	Canada	2.84
France	-0.98	France	0.35	France	2.64
China	-1.08	Cyprus	0.35	Australia	2.63
Australia	-1.16	Spain	0.35	Thailand	2.59
Norway	-1.21	Germany	0.32	Germany	2.55
Thailand	-1.41	Norway	0.23	Norway	2.52
Ireland	-1.43	Thailand	0.22	India	2.40
Denmark	-1.58	New Zealand	0.20	South Africa	2.32
Italy	-1.61	Chile	0.19	Israel	2.28
Cyprus	-1.75	Belgium	0.16	Pakistan	2.26
Israel	-1.76	China	0.15	Italy	2.14
Portugal	-2.12	Israel	0.14	Denmark	2.03
New Zealand	-2.22	Denmark	0.14	Belgium	1.90
Greece	-2.48	Portugal	0.11	New Zealand	1.84
Belgium	-2.49	Austria	0.08	Greece	1.74
Austria	-2.56	Italy	-0.06	Cyprus	1.68
India	-2.64	Trin. & Tob.	-0.09	Portugal	1.68
Pakistan	-2.76	Panama	-0.19	Egypt	1.60
Turkey	-3.11	Jamaica	-0.22	Turkey	1.40
Chile	-3.22	Greece	-0.26	Austria	0.98
Brazil	-3.26	Tunisia	-0.33	Philippines	0.73
Egypt	-3.41	Egypt	-0.41	Chile	0.60
Philippines	-3.51	Philippines	-0.42	Brazil	0.22
Mexico	-4.37	Zimbabwe	-0.51	Mexico	0.01
Zimbabwe	-4.73	Brazil	-0.56	Tunisia	-0.62
Trin. & Tob.	-4.74	India	-0.64	Trin. & Tob.	-0.67
Tunisia	-4.87	Kenya	-0.77	Zimbabwe	-0.69
Jamaica	-4.99	Colombia	-0.92	Sri Lanka	-0.73
Honduras	-5.16	Argentina	-0.98	Honduras	-0.83
Argentina	-5.44	Mexico	-0.99	Jamaica	-0.87
Peru	-5.44	Pakistan	-1.01	Peru	-1.17
Sri Lanka	-5.52	Honduras	-1.01	Argentina	-1.28
Colombia	-5.59	Sri Lanka	-1.05	Kenya	-1.44
Kenya	-5.62	Peru	-1.08	Panama	-1.53
Panama	-5.75	Turkey	-1.10	Colombia	-1.83
Ecuador	-6.94	Ecuador	-1.18	Ghana	-2.75
Ghana	-8.11	Ghana	-1.54	Ecuador	-2.97

For example, taking the indexes of structure-activity and structure-efficiency, one can notice that, except for Taiwan, Switzerland, Sweden, USA and UK which are prevalently market-based because of their active markets, other countries like Turkey, Spain and Ghana are identified as market-based because they have a low developed and very inefficient banking sector. From the structure-regulatory index it emerges that countries like Austria, Germany, France and USA confirm their tradition (bank-based for the first three countries, with few restrictions on the bank activity, and market-based for USA,

with tight regulations) whereas others like the UK and New Zealand have surprising high scores compared to the three European countries, even if they identified as market-based.

Financial development indexes proxy for the degree to which national financial system provide financial services (easing savings mobilization, consider good investment opportunities, reduce agency problems and ease risk management, etc.) and so they denote its development and efficiency. These indicators can rectify the classification of some countries distorted by anomalies of structure indexes as they consider the activity, size and efficiency of the overall financial system and therefore, by showing the true qualities, they return a more realistic classification. A country with high finance scores enjoys of a more efficient financial system than others either because it has a mixed financial structure (banks and markets) or because the development of a particular structure balances the inefficiency of the other one. For example, Switzerland, Taiwan, Japan, USA and UK confirm their ranking of structure indexes, whereas Turkey and Ghana (in the first position of structure indexes), being in the medium-low part of the classification, show distorted structure scores.

By using STATA software, average growth rates were regressed on each indicator of structure and development for both information sets using OLS estimation with heteroscedasticity-consistent standard error. Results are reported in Tables 8 and 9.<sup>67</sup>

**Table 8.** Financial structure and economic growth

Dependent variable: NGDPRPC80-08	Simple Set					Full Set				
	Coeff	S.E.	t-stat.	P-value	R <sup>2</sup>	Coeff	S.E.	t-stat.	P-value	R <sup>2</sup>
Structure Activity	0.328	0.177	1.86	0.070	0.168	0.204	0.183	1.11	0.274	0.464
Structure Size	-0.836	0.476	-2.23	0.031	0.223	-0.616	0.392	-1.57	0.125	0.512
Structure Efficiency	0.271	0.195	1.39	0.171	0.151	0.078	0.207	0.38	0.710	0.454
Structure Regulatory	0.154	0.129	1.19	0.240	0.185	0.209	0.010	2.10	0.043	0.540

**Table 9.** Financial development and economic growth

Dependent variable: NGDPRPC80-08	Simple Set					Full Set				
	Coeff	S.E.	t-stat.	P-value	R <sup>2</sup>	Coeff	S.E.	t-stat.	P-value	R <sup>2</sup>
Finance Activity	0.419	0.133	3.16	0.003	0.347	0.343	0.201	1.71	0.096	0.509
Finance Size	1.081	0.503	2.15	0.037	0.237	0.296	0.713	0.41	0.680	0.457
Finance Efficiency	0.482	0.125	3.86	0.000	0.408	0.448	0.188	2.38	0.022	0.548

Except for structure-size in the simple set, for structure-regulatory and finance-activity in the full set, the results are consistent with Levine and support the financial-services view because:

- structure indexes are not significant at 5% level;
- finance indexes are significant at 5% level (finance-activity and finance-efficiency for the simple set also at 1% level), while finance-size is not significant at 5% level<sup>68</sup>.

The significance of structure-size at 5% level (but not at 1%) seems to support the bank-based view, because the estimated coefficient is negative and equals -0.84, which multiplied by negative values of the index<sup>69</sup> causes a positive (hence, increasing) effect on the growth rate. From Table 3, structure-size is given by the logarithm of the market capitalization ratio divided by the bank credit ratio: its significance could be connected with the growing and the continuous recourse, especially in recent years, to traditional and structured finance forms, which, by increasing the market capitalization and the

<sup>67</sup> We only report the results for financial structure and development indicators.

<sup>68</sup> The result of finance-size is consistent with (Levine 2002) and (Levine and Zervos 1998), who sustain that the market capitalization ratio is not a robust predictor of economic growth.

<sup>69</sup> The lower are the values of structure indexes, the more bank-based is an economy.

volume of bank credit, make structure-size able to explain cross-country differences in economic growth.

Structure-regulatory significance at 5% level, but not at 1%, is strictly connected to the inclusion of China in the sample<sup>70</sup>: by excluding the Country, the estimate becomes not significant.<sup>71</sup> This result is consistent with Allen *et al.* (2005, 2007, and 2008) who extend the analysis of La Porta *et al.* (1998) and show that China is an important counterexample of the finance and growth nexus because it mixes a weak institutional and financial environment with a strong economic growth.

Finance-activity may not be significant because of strongly correlated variables in the full set, which can create a multicollinearity problem that returns more erratic statistic tests; moreover, by using OLS standard estimation, the coefficient of finance-activity has a p-value of 0.044 and thus is significant at 5% level.

Finally the consistency of our results were tested with the law and finance view, by using the two least squares (2SLS) regression. Investor rights and rule of law indicators are used as instrumental variables. All estimates are significant at 1% level and hence show that the component of overall financial development determined by a country's legal and institutional efficiency is positively and robustly linked with economic growth (Table 10).

**Table 10.** Financial development, economic growth and instrumental variables

Dependent variable: NGDPRPC80-08	Simple Set					Levine 2002, Table VII				
	Explanatory variables	Coeff.	S.E.	t-stat.	P-value	R <sup>2</sup>	Coeff.	S.E.	t-stat.	P-value
Finance Activity	0.695	0.212	3.28	0.002	0.288	0.858	0.297	2.892	0.006	
Finance Size	1.703	0.628	2.71	0.010	0.239	1.704	0.566	3.010	0.005	
Finance Efficiency	0.750	0.247	3.04	0.004	0.267	0.876	0.326	2.687	0.011	
Instrumental Variables	Investors rights (OUTRIGHTS_new) and rule of law (LAW_new).					Investors' rights and rule of law.				

### 3.1 The “chinese exception”

As anticipated, China is a well known important exception on the finance and growth literature because it combines a weak institutional and financial environment with a strong economic dynamic. The inclusion of the Country in the original sample improves the estimates of all structure and finance indexes.

Following the method of Allen *et al.* (2005, hereafter AQQ), we compare China's legal (Table 11) and financial system (Table 12) with those of Levine countries grouped by legal origins<sup>72</sup>, by using data and indicators of our analysis. Except for creditor rights index, whose value is higher than countries with civil and scandinavian legal origin, China registers scores below the average of other countries, especially for shareholders rights and corruption, demonstrating that it has a weak legal system that lacks of a solid and efficient political-institutional structure.

<sup>70</sup> China's *structure-regulatory* value is 16, taken from Allen *et al.* (2005).

<sup>71</sup> For the simple set the coefficient of *structure-regulatory* is 0.023, p-value 0.779 and R<sup>2</sup> 0.036, while for the full set the coefficient is 0.101, p-value 0.172 and R<sup>2</sup> 0.429, consistent with Levine. In general China improves all estimates of structure and development indexes.

<sup>72</sup> We use Levine's (2002) dummy variables LEGOR\_FR (*french*), LEGOR\_GE (*german*), LEGOR\_SC (*scandinavian*) e LEGOR\_UK (*common*), which are also included in our dataset.



**Table 11.** A comparison of legal systems

Countries	Rule of law	Corruption	Political stability and absence of violence	Bureaucratic efficiency	Creditor rights	Share holder rights
China	4.20	3.21	4.53	4.96	2	1
Common law countries average	5.99	5.33	4.80	6.34	2.41	4.18
Civil law countries average	5.48	4.58	4.89	5.88	1.30	3.05
German law countries average	8.47	7.88	7.30	8.46	2.25	3.38
Scandinavian law countries average	8.79	9.28	7.56	9.17	1.75	3.63

**Table 12 –** A comparison of financial systems

	Common law countries average	Civil law countries average	German law countries average	Scandinavian law countries average	China
PRIV	0.52	0.46	1.22	0.61	0.85*
OVERHEAD	0,04	0,05	0.03	0.03	0.01
TVT	0.33	0,19	0.73	0.53	0.40
MCAP	0.61	0,32	0.81	0,65	0.31
<b>Structure indexes</b>					
Structure-activity	-1.07	-1.63	-0.98	-0.21	-0.75
Structure-size	0.11	-0.39	-0.72	0.01	-0.99
Structure-efficiency	-5.36	-5.79	-4.17	-4.26	-5.12
Structure-regulatory	9.26	9.06	7	8	16**
<b>Finance indexes</b>					
Finance-activity	-2.65	-3.49	-0.64	-0.96	-1.08
Finance-size	0.00	-0.37	0.61	0.35	0.15
Finance-efficiency	1.46	0.59	2.72	2.83	3.28

\* Beck *et al.* (2008) dataset has only data from 1987 to 1997.

\*\* This value is taken from Allen *et al.* (2005).

Structure indexes support the Country's *bank-based* orientation. Indeed, Chinese financial system is still characterized by a large banking sector (resulting of a long reform process begun in 1979) previously dominated by *State-owned Commercial Banks* (SOCB) which held the majority of the Country's total financial activities.<sup>73</sup>

Some indexes in Table 12 may differ from AQQ analysis because, besides from being more recent<sup>74</sup>, they are calculated on the Levine sample and not on that of LLSV, but others, like finance indexes (particularly finance-efficiency), show that China's overall financial system has reached a satisfactory level of efficiency due to the reduction of its past system distortions and weaknesses. This improvement is mainly related to:

<sup>73</sup> Called the "Big Four", the *Bank of China* (BOC), the *People's Construction Bank of China* (PCBC), the *Agriculture Bank of China* (ABC) and the *Industrial and Commercial Bank of China* (ICBC), have controlled since 1979 the Chinese banking sector. The ICBC, born in 1984 like a spin-off of the central *People's Bank of China* (PBOC) and considered in 2004 among the worst chinese banks due to the high amount of bad loans (over 20% share of total loans), has been liberalized and recapitalized in 2006 and recently reached the five place in world rankings by market capitalization with about USD 150 billions. See Sau (2008), Tabella 1, Bortolotti, and Beltratti (2007).

<sup>74</sup> Data of AQQ's (2005) analysis are of 2002.



- reform of the banking sector<sup>75</sup>, which has achieved a greater efficiency by reducing overhead costs (OVERHEAD score is the lowest compared to other countries);
- the growing role of financial markets (structure-activity score is lower only in comparison to countries with scandinavian legal origin).

On the first point it must be highlighted that the Chinese banking sector is burdened by a high amount of *non performing loans* (NPL) even compared to the major emerging economies of the area (AQQ 2007, Iannini 2008): for a long time Chinese public banks, wholly State controlled, did not follow market dynamics and by easily providing credit to public enterprises (*State-owned Enterprises – SOE*) they were not interested to the soundness and to the profitability of the investments they funded, by resulting, with time, in massive insolvencies. The opening of the banking reform in 1984 introduced many significant changes<sup>76</sup> which produced robust results only in the last decade: official data from 2007 *China's Financial Stability Report* provides a seemingly positive situation with regards to the reduction of NPL and points out the need of a further strengthening by in particular encouraging the development of small banks in rural areas.<sup>77</sup>

Financial markets, developed mainly since 1990 with the opening of Shanghai and Shenzhen stock exchanges, show a good performance confirmed by the *World Bank's Beijing Section Report*: in 2007 the stock exchanges set together reached the third place in world rankings by market capitalization (with USD 4.480 billions, about 150% share of GDP) (Sanacuore 2008, 21). Institutions have played a key role in this rise by reducing the inefficiencies arising from a lack of regulation and a strong speculative environment (presence of many small capitalization companies, absence of blue-chips, presence of non tradable shares, high turnover and high prices' volatility), and by improving markets stability. One of the major moves has been the turning of *non tradable shares* (NTS) into *tradable shares* (TS), which positively affected the firms governance<sup>78</sup>, the implementation of the privatization process (increasing of private shares) and the market liquidity. Chinese authorities are also considering the opportunity to enhance national Exchanges by merging with the Hong Kong Stock Exchange (at the sixth place among stock exchanges ranking by market capitalization) (Sau 2008, Tabella 2-B), which could benefit the:

- prestige and reputation of listed companies;
- attraction of more foreign capital;
- undifferentiated listing of companies throughout China's territory, by diversifying entrepreneurial activities, first among all the ITC sector.<sup>79</sup>

Obviously, such integration should be carried out by carefully monitoring the correct working of the markets and by allowing the impressive flow of chinese savings, and not just by allowing an inflow of foreign financial resources<sup>80</sup> (almost 50% of GDP) (Sanacuore 2008, 25). This virtuous process should be contextual to the gradual reduction of firms' government interference (finishing the reform of NTS) and to a further strengthening of the banking sector efficiency: this should stabilize and continue to feed an already strong economic growth.

So far private entrepreneurs, basing their economic relationships on implicit contractual relations based on reputation, trust and reciprocity (companies owned by founder members, families and friends),

---

<sup>75</sup> In the last decade banking reform has mainly concerned: banks restructuring and recapitalization (by reducing bad loans and increasing the capitalization rate) and financial liberalization (introduction of market practices, interest rates liberalization and opening to foreign competition).

<sup>76</sup> We mention in particular, the reduction of government interference on credit allocation (reduction of NPL) and the abolishment, further to China's WTO admission, of constraints which prevent foreign banks settlement.

<sup>77</sup> From 2005 microcredit starts in some Chinese regions. See Sanacuore (2008), Figure 4, p. 10.

<sup>78</sup> Allen *et al.* investigate on performances and corporate governance mechanisms of firms belonging to, respectively, public sector, listed sector and private sector, and they sustain that Chinese economic growth held up particularly on the development of the private sector both at local and national level. Governance mechanisms of listed and public companies are weak and inefficient (clash of interests of the administrative and audit organs, concentration of participations, strong state control). (AQQ 2005), from p. 86 to p. 92.

<sup>79</sup> Shanghai and Shenzhen SE have are particularly relevant for their own respective provinces: Yangsee for the first one and Guendong and Hunnan for the second one.

<sup>80</sup> China is one of the most attractive country for FDI.

managed to minimize transaction costs<sup>81</sup> and overcome legal system inefficiencies (investor rights and law enforcement). It is assumed that in the future the continuation of the growth process will require improvement also in the Country's legal system.

#### **4. Conclusion**

In this paper it has been pointed out how the financing of innovation is closely linked to economic growth. By examining agents' choices of a small open economy (consisting of four sectors: final, traditional, R&D, and financial) it is shown how the firms financial capability, which allows them to access to external finance (banks or markets) in accordance with their initial wealth level, affects the dynamic equilibrium of the economic system. The thesis argued is that the link between financial and real variables of the economy depends primarily on the financial institutions ability to deliver efficient services (first of all the savings mobilization) and to solve agency problems (through monitoring activities) rather than on its financial structure (bank or market-based). The paper also highlights the importance of the legal and normative background of a country in the sense of support the law and finance view, which explains more adequately cross-country differences of growth rates compared to those of the financial-services view.

The econometric research, realized by regressing real GDP per-capita growth rates on macroeconomic, institutional and financial variables for a sample of 49 countries over the period 1980-2008, confirms the above theory. The main outcome is that economic growth is positively and robustly linked with financial development indexes. China, which could appear an exception of the finance and growth nexus, confirms the outcomes by considering recent structural changes and possible institutional developments.

#### **References**

- [1] Aghion, P., and Bolton, P. 1997. A Model of Trickle-Down Growth and Development. *Review of Economic Studies* 64:151-72.
- [2] Aghion, P., Howitt, P., and Mayer-Foulkes, D. 2005. The Effect of Financial Development on Convergence: Theory and Evidence. *The Quarterly Journal of Economic* 120(1):173 - 222.
- [3] Allen, F., and Gale, D. 2001. *Comparative Financial Systems: A Survey*. MIT Press, Cambridge.
- [4] Allen, F., Qian, J., and Qian, M. 2005. Law, Finance and Economic Growth in China. *Journal of Financial Economics*, No. 77:57-116.
- [5] Allen, F., Qian, J., and Qian, M. 2007. *China's Financial System: Past, Present and Future, China's Economic Transition: Origins, Mechanism, and Consequences*, edited by L. Brandt and T. Rawski, Cambridge University Press. Available to the following link: [http://finance.wharton.upenn.edu/~allenf/download/Vita/AQQ\\_BookChapter\\_328\\_2007.pdf](http://finance.wharton.upenn.edu/~allenf/download/Vita/AQQ_BookChapter_328_2007.pdf).
- [6] Allen, F., Chakrabarti, R., De, S., Qian, J. and Qian, M. 2008. *Law, Institutions and Finance in China and India, Emerging Giants: China and India in the World economy*, edited by Barry Eichengreen, Poonam Gupta, and Rajiv Kumar. Available to the following link: <http://finance.wharton.upenn.edu/~allenf/download/Vita/China-India-ICRIER%20draft-feb26.pdf>.
- [7] Amegashie, J. A. 2006. Economics, Gratitude and Warm Glow. *Working Paper*.
- [8] Arrow, K.J. 1962. The Economic Implications of Learning by Doing. *Review of Economic Studies* Vol. 29:155-73.
- [9] Barro, R.J. 1974. Are Government Bonds Net Wealth? *Journal of Political Economy* 82:1095-1117.
- [10] Barro, R. J., and Lee, J. 1996. International measures of schooling years and schooling quality. *A.E.R. Papers and Proceedings* 86:218-223.

---

<sup>81</sup> Transaction costs are generated by political and administrative interferences on bureaucratic procedures.

- [11] Beck, T., Demirgüç-Kunt, A., and Levine, R. 2008. Financial Structure Dataset revised, November 21, 2008, available to the following link [Research - A New Database on Financial Development and Structure \(updated Nov. 2008\)](#) .
- [12] Bencivenga, V.R., and Smith, B.D. 1991. Financial Intermediation and Endogenous Growth. *Review of Economic Studies* 58(2):195–209.
- [13] Blackburn, K., Bose, N., and Capasso, S. 2005. Financial Development, Financing Choice and Economic Growth. *Review of Development Economics* 9(2):135–49.
- [14] Blackburn, K., and Hung, V.T.Y. 1998. A Theory of Growth, Financial Development and Trade. *Economica* 65(1):107–24.
- [15] Bose, N., and Cothren, R. 1996. Equilibrium Loan Contracts and Endogenous Growth in the Presence of Asymmetric Information. *Journal of Monetary Economics* 38(3):363–76.
- [16] Bortolotti, B., and Beltratti, A. 2007. The nontradable share reform in the Chinese stock market. *Fondazione Eni Enrico Mattei Working Papers*, n. 112.
- [17] Chakraborty, S., and Ray, T. 2006. Bank-based versus market-based financial systems: A growth-theoretic analysis. *Journal of Monetary Economics*, Volume 53 (2), 329-350. Available to the following link: <http://www.uoregon.edu/~shankhac/papers/Bank%20Market.pdf> .
- [18] De la Fuente, A., and Marin, J.M. 1996. Innovation, Bank Monitoring and Endogenous Financial Development. *Journal of Monetary Economics* 38(1):269–301.
- [19] De Gregorio, J., and Kim, S.J. 1998. Credit Markets with Differences in Abilities: education, distribution and growth. *International Economic Review* 41(3):579–607.
- [20] Djankov, S., La Porta, R., Lopez-de-Silanes, F., and Shleifer, A. 2007. [The Law and Economics of Self-Dealing](#). Forthcoming, *Journal of Financial Economics*. [Dataset \(Excel\)](#) .
- [21] Djankov, S., McLiesh, C., and Shleifer, A. 2007. [Private Credit in 129 Countries](#). *Journal of Financial Economics* 84, 299-329.
- [22] Galor, O., and Zeira, J. 1993. Income Distribution and Macroeconomics. *Review of Economic Studies* vol. 60:35-52.
- [23] Harrison, P., Sussman, O., and Zeira, J. 1999. Finance and Growth: Theory and New Evidence. Board of Governors of the Federal Reserve System, Finance and Economics Discussion Series 9935.
- [24] Iannini, G. 2008. *Banche e sviluppo in Cina*. Dipartimento di Statistica ed Economia Applicata, “L. Lenti”, Università di Pavia.
- [25] International Monetary Fund. World Economic Outlook Database, April 2009 (<http://www.imf.org/external/pubs/ft/weo/2009/01/weodata/index.aspx> ).
- [26] King, R.G., and Levine, R. 1993. Finance, Entrepreneurship, and Growth: theory and evidence. *Journal of Monetary Economics* 32(3):513–42.
- [27] La Porta, R., Lopez-de-Silanes, F., Shleifer, A., and Vishny, R. 1998. Law and finance. *Journal of Political Economy* 106:1113–1155.
- [28] Levine, R. 1997. Financial Development and Economic Growth: Views and Agenda. *Journal of Economic Literature* 35:688-726.
- [29] Levine, R. 2002. Bank-Based or Market-Based Financial Systems: Which Is Better? *Journal of Financial Intermediation*, 11: 398-428. Data available to the following link: [http://www.econ.brown.edu/fac/Ross\\_Levine/Publication/2002\\_jfi\\_bb\\_or\\_mbfs.xls](http://www.econ.brown.edu/fac/Ross_Levine/Publication/2002_jfi_bb_or_mbfs.xls).
- [30] Levine, R., and Zervos, S. 1998. Stock markets, banks, and economic growth. *American Economic Review* 88:537–558.

- [31] Merton, R.C., and Bodie, Z. 1995. A conceptual framework for analyzing the financial environment. *The Global Financial System: A Functional Perspective* (Dwight B. Crane *et al.*, Eds.), pp. 3-32. Harvard Business School Press, Boston, MA.
- [32] Morales, M.F. 2003. Financial Intermediation in a Model of Growth Through Creative Destruction. *Macroeconomic Dynamics* 7(3):363–93.
- [33] New York University, Development Research Institute. *Global Development Network Growth, Macro Time Series Data 2005* (<http://www.nyu.edu/fas/institute/dri/index.html>).
- [34] Romer, P. 1986. Increasing Returns and Long-Run Growth. *Journal of Political Economy* vol. 92: 1002-37.
- [35] Saint-Paul, G. 1992. Technological Choice, Financial Markets and Economic Development. *European Economic Review* 36(4).
- [36] Sanacuore, M.G. 2008. Paradossi e potenzialità del settore finanziario cinese. Working Paper no. 34.
- [37] Sau, L. 2008. *La Struttura del Sistema Finanziario in Cina*. Department of Economics “S. Cagnetti de Martiis”. Università di Torino, Working Paper Series 07/2008.
- [38] Trew, A.W. 2005. Finance and Growth: A Critical Survey. Center for Dynamic Macroeconomics Analysis, Working Paper Series, July 2005.
- [39] Worldwide Governance Indicators. 2008. available to the following link: <http://info.worldbank.org/governance/wgi/pdf/wgidataset.xls>.

# DYNAMICS OF UNEMPLOYMENT AND INFLATION IN WESTERN EUROPE: SOLUTION BY THE 1- D BOUNDARY ELEMENTS METHOD

Ivan O. KITOV, Oleg I. KITOV  
Institute for the Geospheres' Dynamics  
Russian Academy of Sciences, RUSSIA  
[ikitov@mail.ru](mailto:ikitov@mail.ru)

## Abstract

*Using an analog of the boundary elements method in engineering and science, we analyze and model unemployment rate in Austria, Italy, the Netherlands, Sweden, Switzerland, and the United States as a function of inflation and the change in labor force. Originally, the model linking unemployment to inflation and labor force was developed and successfully tested for Austria, Canada, France, Germany, Japan, and the United States. Autoregressive properties of neither of these variables are used to predict their evolution. In this sense, the model is a self-consistent and completely deterministic one without any stochastic component (external shocks) except that associated with measurement errors and changes in measurement units. Nevertheless, the model explains between ~65% and ~95% of the variability in unemployment and inflation. For Italy, the rate of unemployment is predicted at a time horizon of nine (!) years with pseudo out-of-sample root-mean-square forecasting error of 0.55% for the period between 1973 and 2006. One can expect that the unemployment will be growing since 2008 and will reach ~11.4% [ $\pm 0.6$  %] near 2012. After 2012, unemployment in Italy will start to descend.*

**Key words:** unemployment, inflation, labor force, boundary integral method, prediction, Western Europe

**JEL classification:** J64, J21, J11, E24, E31

## 1. Introduction

Current discussions on the rate of unemployment as an economic phenomenon and those on modern versions of the Phillips curve, where unemployment plays a crucial role as one of defining parameters, in particular have been rapidly growing since the early 1960s. There is an extensive set of empirical findings and models exploring various assumptions on the forces behind unemployment. There was no unique and comprehensive model for developed countries, however, which could explain all or part of observations relevant to the level and evolution of the portion of people in labor force but without job.

We have constructed and tested an alternative model linking inflation and unemployment in developed countries to the change rate of labor force by linear and lagged relationships. Our model is a completely deterministic one with the change in labor force being the only driving force causing all variations in the indissoluble pair unemployment/inflation, i.e. the reaction in unemployment and inflation lags behind the change in labor force. The model is somewhat orthogonal to conventional economic models and concepts. In its original form, the model was revealed and formulated for the United States (Kitov 2006a). After the correction for known breaks in labor force data, a revised model (Kitov 2006b) allowed a significant improvement on the original one with the root-mean-square forecasting error (RMSFE) of inflation at a 2.5 year horizon of 0.8% between 1965 and 2004. Because of well-known non-stationary of all involved variables, the model was tested for the presence of cointegrating relations (Kitov, Kitov, Dolinskaya 2007a). Both, the Engle-Granger and Johansen approaches have shown the existence of cointegration between unemployment, inflation and the change in labor force, i.e. the presence of long-term equilibrium (in other words, deterministic or causal) relations. Because the change in labor force drives the other two variables, it can be a stochastic process.

The first attempt to obtain empirical models for West European countries also provided strong support to the model (Kitov 2007a). In France, it was found that forecasting horizon for inflation is four years, i.e. the change in labor force leads inflation by four years. Unemployment in France also leads inflation by four years, and various cointegration tests showed the existence of long-term equilibrium relations between the three variables (Kitov, Kitov, Dolinskaya 2007b). In Austria, the change in labor force and the pair unemployment/inflation is synchronized in time. For Austria, it was found that the break in units of measurement around 1987 requires the change in coefficients in linear lagged relationships.

In continuous efforts to extend the set of countries demonstrating the presence of a causal link between the change in labor force, inflation and unemployment, we have build empirical models for the second and third largest economies in the world – Japan (Kitov 2007b) and Germany (Kitov 2007d). Surprisingly, the model for Canada (Kitov 2007c), which has the United States as the largest trade partner, is also an accurate and reliable one, demonstrating the independence of unemployment and inflation on external factors.

It is important to use the rate of growth not increment as a predictor in order to match dimension of inflation and unemployment, which are defined as rates as well. An implicit assumption of the model is that inflation and unemployment do not depend directly on parameters describing real economic activity (Kitov 2006a). Moreover, inflation does not depend on its own previous and/or future values because it is completely controlled by a process of different nature.

The principal source of information relevant to this study is the OECD database (<http://www.oecd.org/>) which provides comprehensive data sets on labor force, unemployment, GDP deflator (DGDP), and CPI inflation. In several cases, national statistical sources and the estimates reported according to definitions adopted in the United States are used for obtaining original data on inflation and corroborative data on unemployment and labor force. In some cases, readings associated with the same variable but obtained from different sources do not coincide. This is due to different approaches and definitions applied by corresponding agencies. Diversity of definitions is accompanied by a degree of uncertainty related to methodology of measurements. For example, figures related to labor force are usually obtained in surveys covering population samples of various sizes: from 0.2 per cent to 3.3 per cent of total population. The uncertainty associated with such measurements cannot be directly estimated but it certainly affects the reliability of empirical relationships between inflation and labor force.

We often use the term “accuracy” in this study. When using it, we rather refer to some estimated uncertainty of measurements than to the difference between measured and true values, i.e. to standard definition of accuracy. This uncertainty might be roughly approximated by variations in a given parameter between consequent revisions or between different agencies. Survey reported uncertainties are just a formal statistical estimate of the internal consistency of measurements. However, population related variables could be potentially measured exactly because they are countable not measurable. In any case, the discrepancy between values predicted by models and corresponding measurements has to be considered in the light of the measurement uncertainty.

The reminder of the paper is organized in four sections. Section 1 formally introduces the model as obtained and tested in previous studies. In many countries, the US and Japan among others, the generalized link between labor force and two dependent variables can be split into two independent relationships, where inflation apparently does not depend on unemployment. However, in few countries, a striking example is France, only the generalized model provided an adequate description of the evolution of both dependent variables since the 1960s.

Section 2 introduces the method of cumulative curves for the solution of the model equations. The proposed method is similar to the method of boundary elements in science and engineering because it is based on the conversion of original differential equations into a set of integral equations. The advantage of our method consists in the availability of an exact solution of the problem. It is shown that the cumulative curves method is a superior one to cointegration tests in obtaining long-term equilibrium relations between the studied variables. For example, the difference between measured cumulative curves for inflation in the United States and France and that predicted from the change in labor force, which are both proven  $I(2)$  series, is an  $I(0)$  process! This feature undoubtedly demonstrates that the link between labor force and inflation is a causal one.

Empirical models for the evolution of unemployment and/or inflation in Italy, the Netherlands, Sweden, and Switzerland are presented in Section 3. These countries significantly enlarge the set of West European countries modelled so far: Austria, Germany, and France. The previously considered case of Austria has been revisited using different data sets and extended to 2007, compared to 2003 in the original version. We also update the prediction of unemployment in the United States. Section 4 concludes.

## 2. The model

As originally defined by Kitov (2006a), inflation and unemployment are linear and potentially lagged functions of the change rate of labor force:

$$\pi(t) = A_1 dLF(t-t_1)/LF(t-t_1) + A_2 \quad (1)$$

$$UE(t) = B_1 dLF(t-t_2)/LF(t-t_2) + B_2 \quad (2)$$

where  $\pi(t)$  is the rate of price inflation at time  $t$ , as represented by some standard measure such as GDP deflator (DGDP) or CPI;  $UE(t)$  is the rate of unemployment at time  $t$ , which can be also represented by various measures;  $LF(t)$  is the level of labor force at time  $t$ ;  $t_1$  and  $t_2$  are the time lags between the inflation, unemployment, and labor force, respectively;  $A_1$ ,  $B_1$ ,  $A_2$ , and  $B_2$  are country specific coefficients, which have to be determined empirically in calibration procedure. These coefficients may vary through time for a given country, as induced by numerous revisions to the definitions and measurement methodologies of the studied variables, i.e. by variations in measurement units.

Linear relationships (1) and (2) define inflation and unemployment separately. These variables are two indivisible manifestations or consequences of a unique process, however. The process is the growth in labor force which is accommodated in developed economies (we do not include developing and emergent economies in this analysis) through two channels. First channel is the increase in employment and corresponding change in personal income distribution (PID). All persons obtaining new paid jobs or their equivalents presumably change their incomes to some higher levels. There is an ultimate empirical fact, however, that PID in the USA does not change with time in relative terms, i.e. when normalized to the total population and total income (Kitov 2009b). The increasing number of people at higher income levels, as related to the new paid jobs, leads to a certain disturbance in the PID. This over-concentration (or “over-pressure”) of population in some income bins above its “neutral” long-term value must be compensated by such an extension in corresponding income scale, which returns the PID to its original density. Related stretching of the income scale is the core driving force of price inflation, i.e. the US economy needs exactly the amount of money, extra to that related to real GDP growth, to pull back the PID to its fixed shape. The mechanism responsible for the compensation and the income scale stretching, should have some positive relaxation time, which effectively separates in time the source of inflation, i.e. the labor force change, and the reaction, i.e. the inflation.

Second channel is related to those persons in the labor force who failed to obtain a new paid job. These people do not leave the labor force but join unemployment. Supposedly, they do not change the PID because they do not change their incomes. Therefore, total labor force change equals unemployment change plus employment change, the latter process expressed through lagged inflation. In the case of a “natural” behavior of the economic system, which is defined as a stable balance of socio-economic forces in the society, the partition of labor force growth between unemployment and inflation is retained through time and the linear relationships hold separately. There is always a possibility, however, to fix one of the two dependent variables. Central banks are definitely able to influence inflation rate by monetary means, i.e. to force money supply to change relative to its natural demand. For example, the Banque de France predefines a strict percentage growth rate of monetary aggregate M2 when formulating its monetary policy (Banque de France 2004). Such a violation of natural, i.e. established over decades, economic behavior should undoubtedly distort the partition of the change in labor force – the portion previously accommodated by inflation would be redirected to unemployment, i.e. those who had to get new jobs would fail because of the lack of money in the economy. To account for this effect one should to use a generalized relationship as represented by the sum of (1) and (2):

$$\pi(t)+UE(t)= A_1 dLF(t-t_1)/LF(t-t_1)+B_1 dLF(t-t_2)/LF(t-t_2)+A_2+B_2 \quad (3)$$

Equation (3) balances the change in labor force to inflation and unemployment, the latter two variables potentially lagging by different times behind the labor force change. Effectively, when  $t_1 \neq 0$  or/and  $t_2 \neq 0$ , one should not link inflation and unemployment for the same year. The importance of this generalized relationship is demonstrated in (Kitov 2007a) for France.

One can rewrite (3) in a form similar to that of the Phillips curve, without any autoregressive terms, although:

$$\pi(t) = C_1 dLF(t-t_1)/LF(t-t_1) + C_2 UE(t+t_2-t_1) + C_3 \quad (4)$$

where coefficients  $C_1$ ,  $C_2$ , and  $C_3$  should be better determined empirically despite they can be directly obtained from (3) by simple algebraic transformation. The rationale behind the superiority of the empirical estimation is the presence of high measurement noise in all original time series. In some places, (4) can provide a more effective destructive interference of such noise than does (3). Consequently, the coefficients providing the best fit for (3) and (4), whatever method is used, may be different. In this study we use relationship (4), but one should not consider it as an equation predicting inflation. It is rather a convenient form of the equation balancing inputs of all three variables with the labor force driving the other two. Moreover, inflation may actually lead unemployment, as it is found in the United States. Then inflation defined by (4) actually depends on some future readings of unemployment (Kitov 2009a).

For the USA, there was no need to apply relationship (3) because corresponding monetary policies and other potential sources of disturbance do not change the natural partition of the change in labor force, as observed since the late 1950s. Coefficients in relationships (1) and (2) specific for the USA are as follows:  $A_1=4$ ,  $A_2=-0.03$ ,  $t_1=2$  years (DGDP as a measure of inflation),  $B_1=2.1$ ,  $B_2=-0.023$ ,  $t_2=5$  years.

For Japan,  $A_1=1.31$ ,  $A_2=0.0007$ ,  $t_1=0$  years (DGDP), and  $B_1=-1.5$ ,  $B_2=0.045$ ,  $t_2=0$  years (Kitov 2007b). It is worth noting that  $B_1$  is negative and any decrease in the level of labor force or too weak growth would result in increasing unemployment. The change rate of labor force measured in Japan has been negative since 1999 and the measured inflation, DGDP and CPI, has been negative as well. There is no indication of a recovery to positive figures any time soon if to consider the decrease in working age population and participation rate as has been observed in Japan since the late 1990s.

For Germany, there exists a Phillips curve:  $UE(t-1) = -1.50[0.1]DGDP(t) + 0.116[0.004]$  with inflation lagging unemployment by one year. The goodness-of-fit is ( $R^2=$ ) 0.86 for the period between 1971 and 2006, i.e. during the period where data are available. Coefficients in (1) and (2) are as follows:  $A_1=-1.71$ ,  $A_2=0.041$ ,  $t_1=6$  years (CPI), and  $B_1=2.5$ ,  $B_2=0.04$ ,  $t_2=5$  years (Kitov 2007d). Considering the presence of the same time lags before and after the reunification, but different coefficients, the case of Germany is an outstanding one. It demonstrates how deep are the socio-economic roots of the driving force behind inflation and unemployment. On the other hand, it is really difficult to imagine that the process of the transformation of the change in labor force into inflation takes six years. However, the lag of inflation behind the change in labor force allows a prediction at a six-year horizon with a very small RMSFE.

In Canada,  $A_1=2.58$ ,  $A_2=-0.0043$ ,  $t_1=2$  years (CPI) with  $R^2=0.67$ , and  $B_1=-2.1$ ,  $B_2=0.12$ ,  $t_2=0$  years. Therefore, the change in labor force and unemployment lead inflation by two years allowing a natural forecasting horizon of two years.

In (Kitov, Kitov, Dolinskaya 2007b), we have carried out a formal statistical assessment of the empirical linear lagged relationship (1) for the USA. It has demonstrated that the pseudo out-of-sample RMSFE for CPI inflation at a two-year horizon for the period between 1965 and 2002 is only 0.8%. This value is superior to that obtained with any other inflation model by a factor of 2, as presented (Atkeson, and Ohanian 2001, Stock, and Watson 1999, Stock, and Watson 2005). This forecasting superiority is retained for shorter sub-periods with RMSFE of 1.0% for the first (1965-1983) and 0.5% for the second (1983-2002) segment. In the mainstream models of inflation, the turning point in 1983 is dictated by the inability to describe inflation process with one set of defining parameters. Therefore, special discussions are devoted to statistical, economic, and/or financial justification of the split and relevant change in parameters (Stock, Watson 2005). Our model denies the necessity of any change in the factors driving inflation in the US around 1983 or in any other point after 1960. Each and every inflation reading is completely defined by the change in labor force occurred two years before.

### **3. The boundary elements method in economics**

In (Kitov, Kitov, and Dolinskaya 2007a), we have introduced a simple but effective method to find an appropriate set of coefficients in (1) through (4). This method consists in the search of the best-fit between cumulative values and is similar to the boundary elements method (BEM) in engineering and



science, in its 1D form. The BEM reduces a set of linear (partial) differential equations, e.g. relationships (1) and (2), to a set of integral equations. The solution of the integral equations, as expressed in boundary integral form, is an exact solution of the original differential equations. In the case of relationship (1):

$${}_{t_0} \int^{t_{01}} d[\ln P(t)] dt = {}_{\tau_0} \int^{\tau_{01}} (A_1 d[\ln LF(\tau)] d\tau + {}_{\tau_0} \int^{\tau_{01}} A_2 d\tau \quad (5)$$

The solution of the integral equation (5) is as follows:

$$\ln P \Big|_{t_0}^{t_{01}} = A_1 \ln LF \Big|_{\tau_0}^{\tau_{01}} + A_2 t \Big|_{\tau_0}^{\tau_{01}} + C \quad (6)$$

where  $P(t)$  is the level of price (index) at time  $t$  ( $\pi(t) \equiv dP(t)/P(t) = d\ln P(t)$ );  $t_0$  and  $t_{01}$  are the start and end time of the integration, respectively; and  $C$  is the free term, which has to be determined together with coefficients  $A_1$  and  $A_2$  from the boundary conditions:  $P(t_0) = P_0$ ,  $P(t_{01}) = P_1$ ,  $LF(\tau_0) = LF_0$ ,  $LF(\tau_{01}) = LF_1$ , where  $\tau = t - t_1$  is the time lagged by  $t_1$ , i.e. by the lag of the change in price behind the change in labor force.

For 1-D problems, we have fixed values as boundary conditions instead of boundary integrals. The number of boundary conditions in (6) is complete for calculation (or quantitative estimation, if there is no analytic solution) all involved coefficients, considering that, without loss of generality, one can always set  $P_0 = 1.0$  as a boundary condition. When estimated, these coefficients entirely define the particular solution of (6):

$$\ln [P(t_{01})] = A_1 \ln [LF(\tau_0)/LF(\tau_{01})] + A_2 (\tau_{01} - \tau_0) + C \quad (7)$$

on both boundaries, i.e. at  $t_0$  and  $t_{01}$ , as well as over the entire time domain between  $t_0$  and  $t_{01}$ . (It is presumed that  $LF(t)$  is a function of time known from measurements.) The estimation of all involved coefficients is the essence of numerical solution of 2D and 3D problems by BEM in scientific applications. In this study, a simple trial-and-error method is used as based on visual fit. Therefore, the residual between observed and predicted curves is not minimized in any metrics and a better OLS solution is likely to exist.

For solving problems (1) and (2) using (7) with an increasing accuracy, we can use a series of boundary conditions for subsequent years. As a rule, inflation in developed countries varies in a relatively narrow range and only rarely dives into the zone of negative growth rate. This makes it difficult to obtain reliable estimates of the involved coefficients from short time series, which is a characteristic feature of economic research. Our experience shows, that depending on the dynamics of inflation and the level of measurement noise in a given country, one needs from 30 to 50 readings to get a reliable solution of problems (1) or (2). (Japan is a brilliant example of a country with a long history of deflation, which makes the resolution of cumulative curves possible even for shorter time intervals.) Similar to many physical problems, the wider is the range of inflation change and lower is the noise level (i.e. the higher is signal to noise ratio) the shorter observation period is needed. A proper set of coefficients should make subsequent residuals between observed and predicted cumulative curves to be a stationary time series with decreasing relative (i.e. normalized to the attained cumulative value) errors. This is a consequence of high correlation in measurement errors for any variable with increasing level: over time, each annual reading is characterized by the same absolute error, but the cumulative change over decades, which is much larger than any annual step, is measured with the same absolute and falling in relative terms error. Compare this feature with famous tests for cointegration, where time series are differentiated with a significant decrease in signal-to-noise ratio and corresponding increase in measurement error, both absolute and relative. In physical terms, if a link between two variables does exist one should better use integral not differential approach.

Overall, solution (7) is the basis of the cumulative curve approach to estimate coefficients  $A_1$ ,  $A_2$ , etc., in relationships (1) and (2). In terms of the boundary elements method, the right hand side of (7) is the particular solution of the (ordinary) differential equation (1). Because  $t_1 \geq 0$ , the causality principle holds, and the independent function is known before the dependent one. The only principal difference with the standard BEM used in scientific applications is that the solution (7) is not a closed-form or

analytic solution, as those in fluid dynamics, acoustics, and electromagnetism. It is the change in labor force in a given country, which may follow a quite exotic trajectory as related to demographic, social, economic, cultural, climatic, etc. circumstances. From (7), inflation in the country can be exactly predicted at a time horizon  $t_1$ , and possibly evaluated at longer horizons using various projections of labor force (Kitov, Kitov 2008a). As a logic consequence, there is no alternative way to predict inflation since it is entirely constrained by the change in labor force. At the same time, solution (7) may have infinite number of future trajectories.

It is a requirement that BEM is applicable to problems for which Green's functions can be derived, i.e. the functions describing the solution in the body between the boundaries. For example, fields in linear homogeneous media created by point (Dirac delta-function) sources. What does play the role of the Green function in the problem under consideration? The answer is obvious:  $\ln[(LF(t)/LF(t_0))]$  and  $(t-t_0)$ . A linear combination of these two functions comprises any particular solution of (3). In physical terms, there is a causal link between labor force and the combination of inflation and unemployment. However, the particular solutions of problems (1) and (2) are sensitive to the change in "physical" conditions in a given economy. When a central bank introduces a strict bound on inflation, both coefficients in (1) should change. Unemployment must react to compensate the deviation from the original relationship (4) and both coefficients in (2) must also change. However, the generalized relationship holds as long as the economy reproduces all economic and social links between its agents. Apparently, there are circumstances in which the generalized relationship does not work any more. Moreover, at some point in the past, there was no generalized relationship between labor force, inflation and unemployment. At some point in the future, current relations will not hold any more replaced by some new economic laws.

There is a variety of numerical methods for the estimation of coefficients in boundary problem (7), which are the workhorse of the BEM. For our purposes, even the simplest visual fit between observed and predicted cumulative curves over 40 to 50 years is an adequate method for the estimation. Thus, all coefficients estimated in this study are likely to be slightly biased in sense of OLS, but still provide a much better overall fit than any set of coefficients, which can be obtained by OLS (or even the VAR technique) from dynamic data.

In order to demonstrate the power of the cumulative curve concept we applied the method to Austria and the USA (Kitov, Kitov, and Dolinskaya 2007a, 2007b). Figure 1 (right panel) displays the observed cumulative curve in Austria and that obtained from the particular solution (7) of equation (1):  $\pi(t) = 1.25dLF(t)/LF(t) + 0.0075$ ,  $t > 1986$ , and  $\pi(t) = 2.0dLF(t)/LF(t) + 0.033$ ,  $t \leq 1986$ , where  $\pi(t)$  is the GDP deflator reported by national statistics, with the boundary conditions set for 1960 and 2003 [4]:  $\Pi(1959)=0$ ;  $\Pi(2003)=1.637$ ;  $LF(1959)=2,364,200$ ;  $LF(2003)=3,424,900$ , where  $\Pi(2003)=\sum\pi(t_i)$ ,  $i=1960, \dots, 2003$ .

These boundary conditions evolve with calendar year, but since the predicted cumulative curve is always close to the observed one, coefficients in (4) do not fluctuate much. It is possible to get the best-fit coefficients using the full set of estimations for all possible combinations of the start and end years. However, this is not the purpose of the study, which demonstrates that the concept linking inflation and unemployment to the level of labor force is an adequate one even in its simplest realization.

The boundary integral equation method allows effective noise suppression, both the one induced by the discretization of continuous functions and that related to measurement errors. The superior performance of the boundary integral method is demonstrated by Figure 2, where the differences between predicted and observed time series are shown. As one can judge from the left panel of the Figure, the difference between the cumulative curves does not deviate much from the zero line.

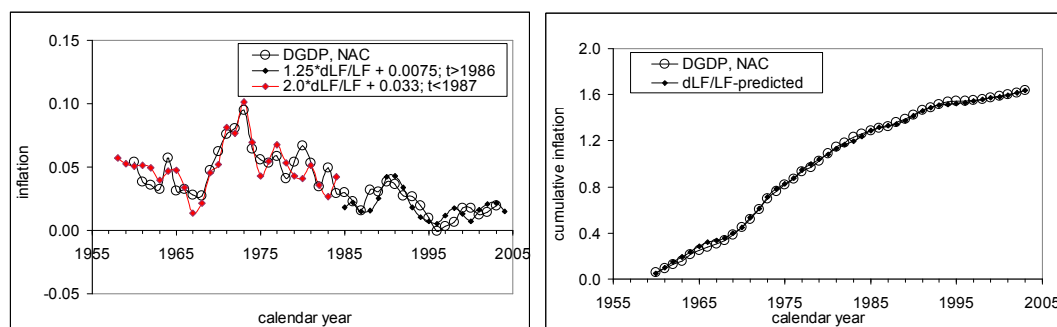
The measured dynamic series of  $\pi(t)$  and that predicted from the  $dLF(\tau)/LF(\tau)$  in the left panel of Figure 1 contain 47 observations and are essentially non-stationary. According to the Phillips-Perron unit root test,  $z(\rho)=-7.1$  for the measured time series and  $z(\rho)=-5.29$  for the predicted one. The 1% critical value is -16.63 and the 5% critical value is -13.17. The augmented Dickey-Fuller test gives  $z(t)=-1.87$  and -1.53, respectively, with the 1% critical values -3.61 and the 5% critical value of -2.94. Therefore, one can not reject the null hypothesis of the presence of unit roots in both time series. As shown below, the first differences of both series have no unit roots, and thus the original series are essentially integrated of order 1,  $I(1)$ , and it is necessary to test them for cointegration before using linear regression analysis.

The Phillips-Perron test shows that the difference between the measured and observed variables is characterized by  $z(p)=-34.37$ , with the 1% critical value of  $-18.63$ . The Dickey-Fuller test gives  $z(t)=-5.41$ , with the 1% critical value of  $-3.61$ . Therefore, both tests reject the presence of unit roots in the difference and, thus, it is integrated of order 0. This finding evidences in favor of the hypothesis that the observed and predicted inflation are cointegrated because the difference is similar to the residual in the Engle-Granger (1987) two-step method. The only distinction is that the residual of a linear regression is replaced with the difference of the measured and predicted time series. This is a insignificant deviation, however, because there exists a linear combination of two  $I(1)$  series, which is proved to be a  $I(0)$  series. For that reason, we do not follow up the original Engle-Granger method in this study. It should give no other result except the absence of unit roots in the residual. Otherwise, it is a flawed method.

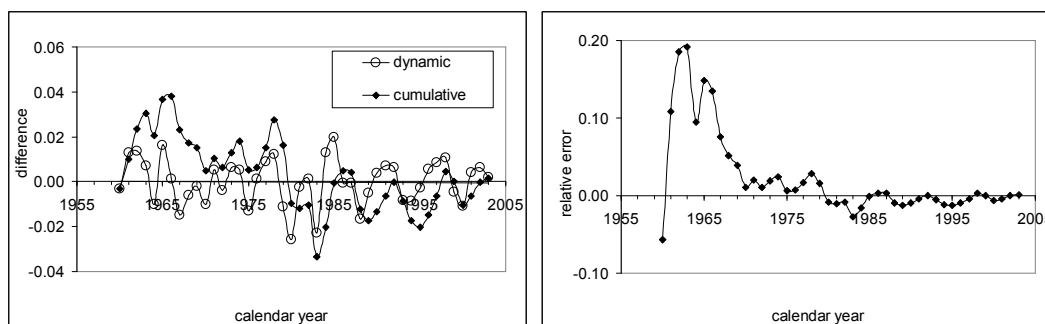
The Johansen procedure tests for cointegration and defines cointegration rank at the same time. For two variables, cointegration rank 1 means the existence of an equilibrium long-term relation. For Austria, cointegration rank is 1 with trace statistics of the Johansen test of 2.80 compared to the 5% critical value of 3.76. Thus, the null hypothesis of the existence of a cointegrating relation cannot be rejected.

From this analysis one can conclude that there exists a cointegrating relation between the observed inflation and that predicted from the change in labor force and unemployment using relationship (4) with the empirical coefficients derived above. Therefore, the case of Austria validates the model obtained by the method of cumulative curves in econometric terms and one can use the long-term causal relationships as empirical relations describing the Austrian economy as a physical system. The coefficients obtained by simple fit of the cumulative curves provide a linear relation between the inflation and the change in labor force with an  $I(0)$  process as the residual. On the contrary, the above tests for cointegration are based on an extraordinary sophisticated assumptions and sets of simulated data.

Now the existence of a cointegrating relation between the observed and predicted inflation is proved and one can use standard statistical methods for the estimation of predictive power of the model. A simple linear regression gives  $R^2=0.82$  and standard (RMS) error of 0.9% for the years between 1960 and 2003. Hence, the predicted series explains 82% of the variability in the measured one and the prediction uncertainty is on the level of measurement uncertainty, with the range of inflation change between 9.5% and -0.4%. It is also instructive to build a VAR model for the observed and predicted inflation in Austria, which uses autoregressive properties of both series. With the largest time lag of 4, the VAR model is characterized by RMSE of 1.1% for the measured series, i.e. by a larger error than that obtained by OLS. In any case, the uncertainty is very low according to the standards dictated by conventional models.



**Figure 1.** Observed and predicted inflation in Austria. Left panel – annual (dynamic) curves. Right panel: cumulative curves.



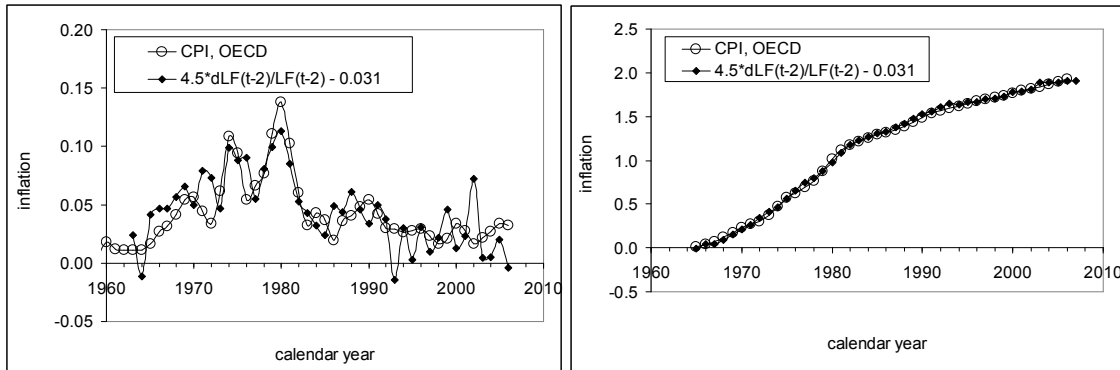
**Figure 2.** Left panel: The difference between the observed and predicted inflation (i.e. the prediction error) presented in Figure 4. The difference presented for the dynamic and cumulative curves. Right panel: The difference for the cumulative curves normalized to the attained level of cumulative inflation. Notice the gradual decrease in the relative error through time.

The cumulative curves have another implication. The long-term behavior of the observed and predicted curves implies that the former can be replaced with the latter without any loss of accuracy. Kitov (Kitov 2006a) demonstrated that high prediction accuracy for both the dynamic and cumulative series is also achievable when only the change in labor force is used. Corresponding RMSE are 0.01 and 0.016 for the period between 1960 and 2003, as shown in Figure 2 (left panel), which also demonstrates that relative deviation between the measured and predicted cumulative curves, i.e. the difference divided by the attained level of cumulative inflation, gradually decreases (right panel). Both differences in the left panel of Figure 2 are integrated of order 0, as the Phillips-Perron (PP) and the Dickey-Fuller (DF) tests show. Specifically, the PP test gives  $z(p)=-50.6$  and  $-19.9$  for the dynamic and cumulative series with the 1% critical value of  $-18.8$ ; and  $z(t)=-6.96$  and  $-3.54$  with the 1% critical value of  $-3.63$  (5% critical value is  $-2.95$ ). The DF test gives  $z(t)=-6.97$  and  $-3.56$  with the 1% critical value of  $-3.63$ . Therefore, one can reject the null the series contain a unit root, i.e. both differences are stationary processes. The difference between the cumulative curves is an  $I(0)$  process and is a linear combination of two  $I(2)$  processes! This is the expression of the power of the boundary elements method in science – integral solutions, when exist, suppress noise very effectively by destructive interference.

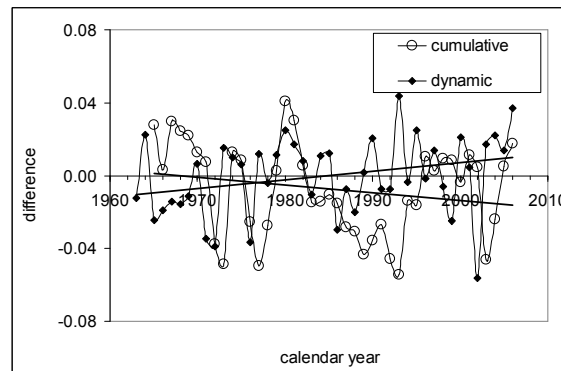
Thus, the replacement of the observed curve with the predicted one is the solution of the original ordinary differential equation. Such a replacement does not compromise the accuracy of cumulative inflation growth, i.e. the overall change in price over longer periods. Imagine, there is no need to measure inflation – just count the number of people in labor force! It is also applicable to the prediction of the price evolution at longer horizons – one has to project the growth in labor force.

We have carried out a similar analysis for the United States. Figure 3 compares dynamic and cumulative curves of measured inflation to that predicted from labor force. The best visual fit of the cumulative curves provides the following relationship:  $\pi(t) = 4.5dLF(t-2)/LF(t-2) - 0.031$ , where  $\pi(t)$  is the CPI inflation. The period covered by this relationship is between 1965 and 2006. Figure 4 depicts the differences between the observed and predicted curves, both the dynamic and cumulative ones, for the same period. Despite the best fit between the cumulative curves there is a constant term in both differences. This is an important specification of the following unit root tests.

Both differences in Figure 4 are integrated of order 0, as the Phillips-Perron (with maximum time lag 4) and the Dickey-Fuller (DF) tests show. Specifically, the PP test gives  $z(p)=-38.44$  and  $-14.71$  for the dynamic and cumulative difference with the 1% critical value of  $-12.54$ ; and  $z(t)=-5.94$  and  $-3.54$  with the 1% critical value of  $-2.63$  (5% critical value is  $-1.95$ ). The DF test gives  $z(t)=-5.95$  and  $-3.10$ , respectively, with the 1% critical value of  $-2.63$ . Therefore, one can reject the null that the series contain a unit root, i.e. both differences are stationary processes. Having the stationary difference between the cumulative curves one can suggest that the dynamic times series, which are obtained from the cumulative curves, are also cointegrated. In other words, there exists a linear combination of the dynamic series which creates an  $I(0)$  process.



**Figure 3.** Observed and predicted (CPI) inflation in the United States. Left panel – annual rate curves. Right panel: cumulative curves.



**Figure 4.** The difference between the observed and predicted inflation presented in Figure 3.

We have analyzed the performance of the cumulative curve method for two countries. Both examples demonstrate that one can easily obtain coefficients of linear lagged relationship between inflation and the change in labor force, which provide excellent prediction for both dynamic and cumulative inflation. As in physics, the success of this method can be only rooted in the existence of a causal link between these macroeconomic variables. And similar to the situation in the hard sciences, there is no ultimate proof of the existence of the link - only statistical evidences obtained from numerous measurements with increasing accuracy. Hence, to make the link more reliable one must extend relevant data set – both over time and across other developed economies. Section 3 adds on four West European countries.

#### 4. Unemployment in European countries

The linear and lagged relationships between inflation, unemployment, and the change in labor force have been demonstrating an excellent performance for the largest (the United States, Japan, Germany, and France) and smaller (Austria and Canada) world economies since the early 1960s. These relationships are expected to be successful for other developed economies with similar socio-economic structure. European countries provide a variety of features related to inflation and unemployment as one can conclude from the economic statistics provided by the OECD. This diversity includes periods of very high inflation accompanied by high unemployment, periods of low inflation and unemployment, and other combinations complicated by transition periods. It is a big challenge for any theory of inflation to explain the diversity of empirical facts.

##### Austria

It is convenient to start with Austria for four reasons. First, there are several alternative data sets for all involved variables, which demonstrate the uncertainty of measurements as related to definition. Second, we have already derived a model for Austria for the period between 1960 and 2003 (Kitov 2007a), which is also discussed in Section 2. Recent readings (between 2004 and 2008) and alternative data sets allow validation of the previous model. Third, the length of data set for Austria is around 50

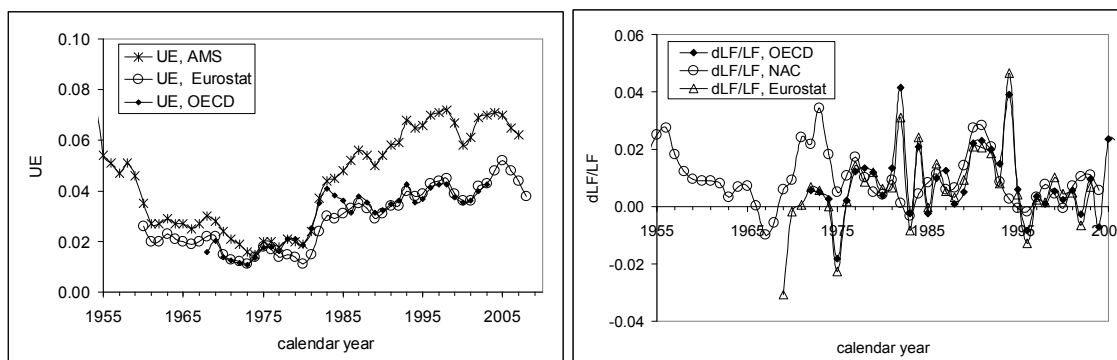
years long what made it possible to test the model for cointegration. Other European countries in this study are characterized by shorter periods of measurement what effectively makes any test for cointegration unreliable. Fourth, it was found that the Austrian data sets have a major break in 1986, when new units of measurements were introduced by revisions to current definitions. The OECD provides the following description of relevant breaks (OECD 2008):

**Series breaks:** Employment data from 1994 are compatible with ILO guidelines and the time criterion applied to classify persons as employed is reduced to 1 hour. Prior to 1994, armed forces were included in the civilian labour force, in services. In 1987, a change occurred in the definition of the unemployed where non-registered jobseekers were classified as unemployed if they had been seeking work in the last four weeks and if they were available for work within four weeks. In previous surveys, the unemployment concept excluded most unemployed persons not previously employed and most persons re-entering the labour market.

Therefore, the model for Austria is an instructive one and evidences that there were no structural breaks in terms of the change in underlying economic processes, but rather new measurement units were used. (Same as a country would shift from miles per hour to kilometers per hour with the same formal speed limit.) Therefore, the relationships derived for the period before 1986 should be scaled to fit new units.

The model and its performance have been described and tested in Section 2 and in several papers. So, there is no harm to skip many formal details in describing the same procedures for individual cases and to focus on empirical results represented in graphical form. Scatter plots and time history curves bring most valuable information on amplitude, timing, scatter of the involved variables and the difference between observed and predicted time series. Generally, graphical representation is more informative than long tables.

Figure 5 displays several time series for unemployment and the change rate of labor force obtained according to different definitions: AMS (Arbeitsmarktservice - <http://www.ams.at/>), NAC (national accounts - <http://www.statistik.at/>), Eurostat, and OECD. A remarkable feature in both panels of Figure 5 is a large difference between amplitudes of the curves, otherwise evolving synchronously. Therefore, it is possible to scale the curves. The difference between the curves could serve as a conservative estimate of the uncertainty in relevant measurements, i.e. as a proxy of the difference between measured and true values of the studied variables. Specifically, one can compare the  $dLF/LF$  curves near 1975, 1985, 1995, and 2005. There are spikes in the OECD and Eurostat curves, which not present in the NAC curve. Usually, such spikes manifest a step-like revision to population controls or the introduction of new definitions. All in all, the modelling of unemployment and inflation using the  $dLF/LF$  curves in Figure 5 has some room for the deviation between predicted and observed curves as associated with measurement errors. Figure 6 depicts three time series with two measures of inflation rate: CPI inflation reported by the national statistical agency and GDP deflator reported by the OECD and Eurostat. The difference between the latter two measures is so large because they count nominal and real GDP in local currency and Euros (using current exchange rate).



**Figure 5.** Unemployment rate (left panel) and the rate of labor force change (right panel) in Austria according to various definitions: NAC and AMS - national definition, Eurostat and OECD. Variations in the variables are a crude approximation of the uncertainty in relevant measurements.

Figure 7 demonstrates the accuracy of the generalized model (4) in prediction of inflation. This model is similar to that described in Section 2 and also has a break in 1986, but it uses different measures of labor force and unemployment (both AMS). As a consequence, empirical coefficients are also different, except fixed one:  $C_2=-1.0$ ;  $C_1=1.1$  and  $C_3=0.068$  before 1986;  $C_1=0.8$  and  $C_3=0.077$  after 1986. These coefficients are obtained using the cumulative curve method, as shown in the right panel of Figure 7.

For Austria, there is no need to distinguish between unemployment and inflation. The predictive power of the generalized model allows an accurate forecast for both macroeconomic variables from projections of labor force. The better is the projection, the higher is the accuracy.

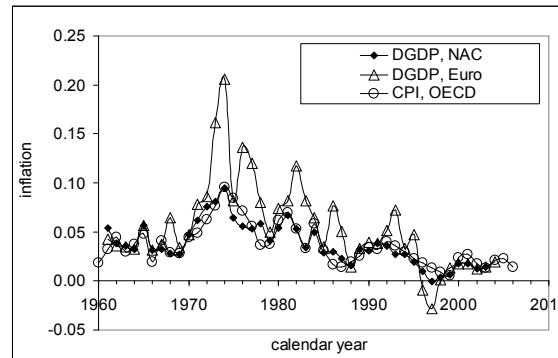


Figure 6. Three definitions of the rate of price inflation in Austria: GDP deflator (DGDP) - national currency and Euro, and CPI inflation.

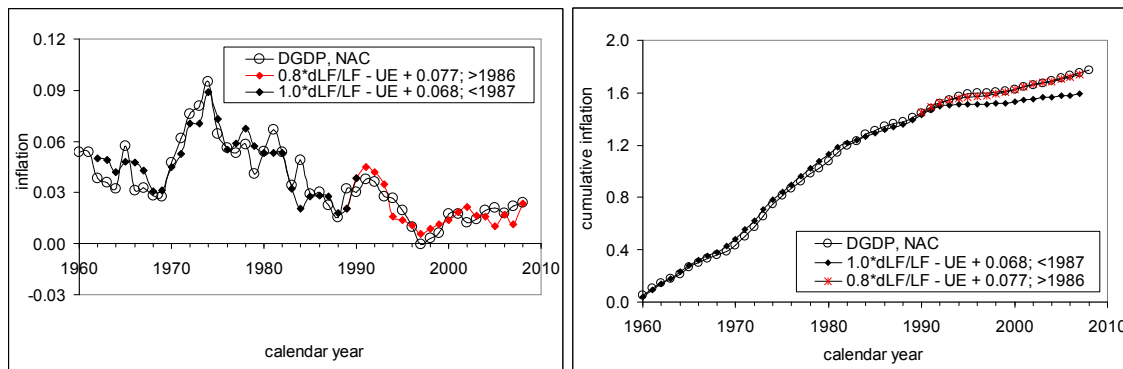


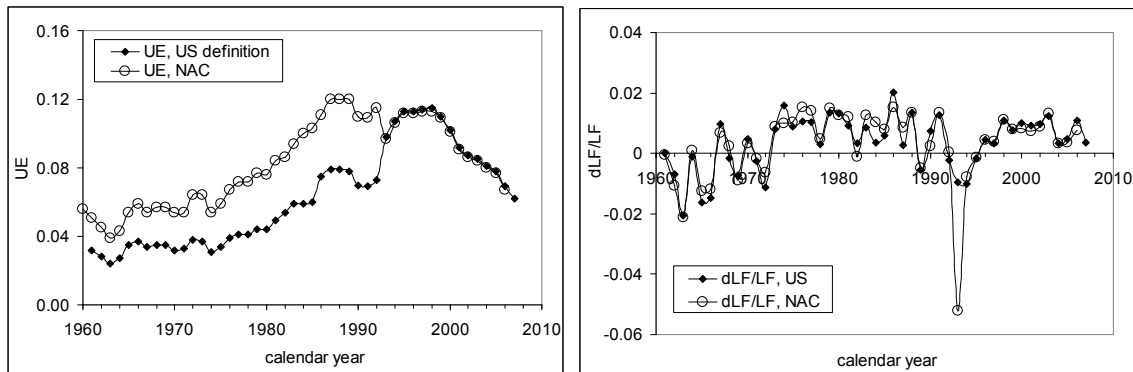
Figure 7. Prediction of GDP deflator in Austria using the generalized model. Left panel: the rate of price inflation. Right panel: Cumulative inflation. (See details in the text.)

### Italy

Retaining the above developed approach to the presentation of data and empirical model, Figure 8 displays two measures of unemployment and the change rate of labor force in Italy. First measure is introduced by national statistics and second is estimated according to the approach developed in the United States. All time series are available through the US Bureau of Labor Statistics (<http://www.bls.gov/data/>). The difference between the curves is obvious. In contrast to Austria, the unemployment curves cannot be so easily scaled, i.e. their cross-correlation is lower. Moreover, it seems that the national definition was replaced by the US one near 1993. This assumption is supported by the change rate of labor force having a large spike in the same year, the spike being a well-known sign of a large revision to definitions (OECD 2008):

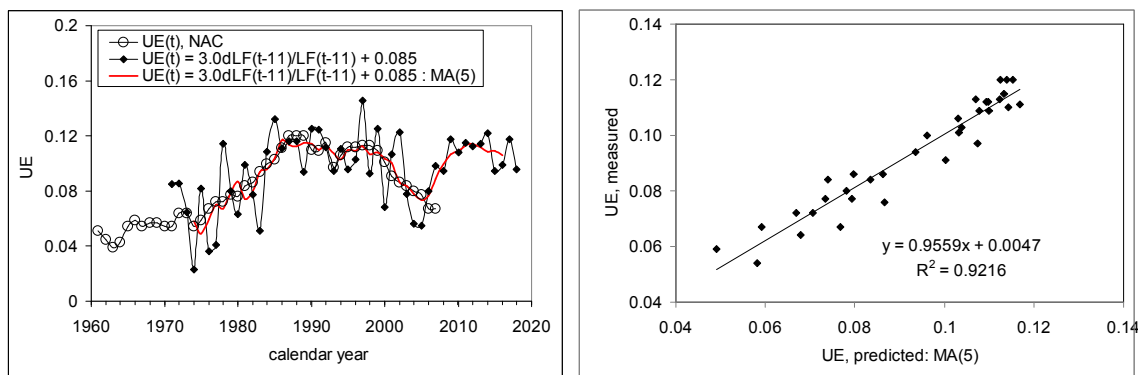
**Series breaks:** In October 1992, changes were introduced in the Household Labour Force Survey concerning the lower age limit of the active population (from 14 to 15 years old), the definition of unemployment, the population estimates, the estimation procedure and the imputation procedure. These changes resulted in a reduction in level estimates for employment and unemployment. In

January 2004, major changes were introduced such as: the passage to a continuous survey, the implementation of CAPI/CATI instead of PAPI for interviews, better adherence to the international definition of employment, the change of the age limit in the definition of the unemployed (74 years old), and the data were revised back till the third quarter 1992.



**Figure 8.** Unemployment rate (left panel) and the rate of labor force change (right panel) in Italy according to national definition (NAC) and the definition adopted in the US.

In Figure 9, the observed unemployment is compared to that predicted from the change in labor force. Using only visual fit between the dynamic curves, we have estimated the coefficients in (2) for Italy:  $B_1=3.0$ ,  $B_2=0.085$ , and  $t_2=11$  (!) years. Hence, an increase in labor force causes a proportional increase in unemployment with a factor of 3. The original time series  $dLF/LF$  is very noisy and thus is smoothed by a five-year moving average, MA(5). The right panel in Figure 9 displays a scatter plot - the measured UE vs. the predicted one, a linear regression line, and corresponding equation. The slope of the regression line is 0.96 (a slight underestimation of the slope results from the uncertainty in the independent variable) and  $R^2=0.92$ . So, the predicted series explains 92% of the variability in the observed one.



**Figure 9.** Observed and predicted unemployment in Italy. The prediction horizon is 11 years. Due to large fluctuations associated with measurement errors in the labor force we are using centered MA(5) for the prediction, which reduces the horizon to 9 years. Goodness-of-fit ( $R^2=$ ) 0.92 for the period between 1973 and 2006, with RMSFE of 0.55%. The unemployment should start to increase in 2008.

As a rule, smoothing by MA(5) is an operation introducing a high bias in time series due to increasing autocorrelation. However, Italy is an exceptional economy with an eleven-year lag of unemployment behind the change in labor force. This lag is twice as large as in Germany and the largest among all studied countries. What kind of social and economic inertia should be involved in such a long delay of reaction? In any case, the smoothing with MA(5) is an adequate and accurate procedure for the prediction of unemployment at 9-year horizon in Italy with RMSFE of only 0.55% for the period between 1973 and 2006. This small RMSFE is associated with a wide range of change in the

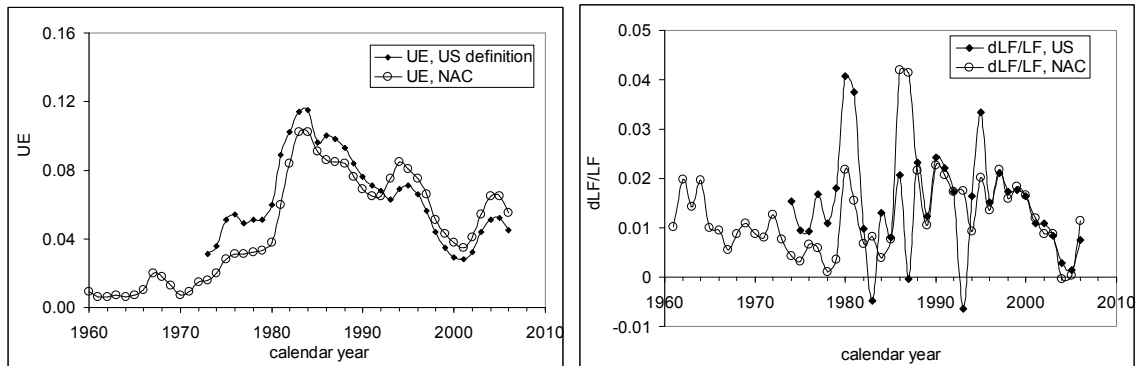


unemployment from 5.4% in 1974 to 12% in 1999. Together with the extremely large forecasting horizon, such uncertainty allows a decisive validation of our model in the next 9 years. One can expect that unemployment in Italy will be growing since 2008 and will reach  $\sim 11.4\%$  [ $\pm 0.6\%$ ] near 2012. After 2012, unemployment in Italy will likely start to descend.

### The Netherlands

For the Netherlands, two different time series are available for unemployment and labor force – according to national (NAC) definition and the US concept. Two measures of inflation are reported by the OECD. Figures 10 and 11 display all the involved series, which have different length but all are in the range between 1960 and 2006. The difference in the measures of labor force is outstanding, especially around 1980 and 1990. One should not tolerate such a discrepancy. Otherwise, no quantitative analysis is possible and economics will never have a chance to join the club of the hard sciences. The OECD (2008) provides the following information on the changes in definitions in the Netherlands:

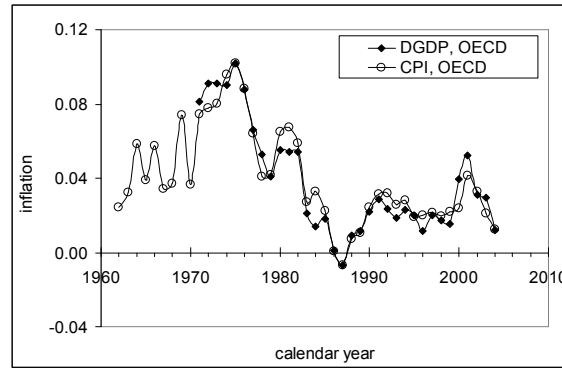
**Series breaks:** The break in 1991-92 is due to the introduction of new definitions in the survey. The implementation of a continuous survey also caused a break in series between 1986 and 1987. The increase in employment figures is due to the survey collecting more information on persons working fewer weekly hours (less than 20 hours a week). Between 1982 and 1983, the break, mainly in the unemployment series, is due to the implementation of the Labour Force Survey.



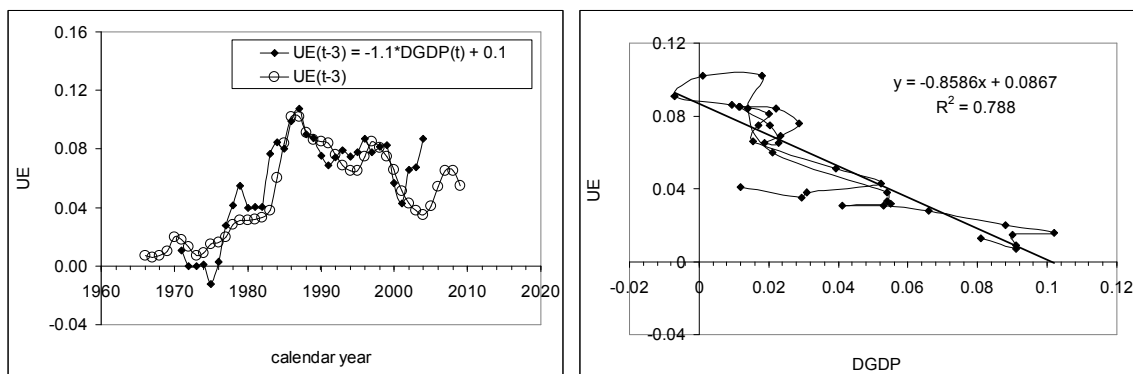
**Figure 10.** Unemployment rate (left panel) and the rate of labor force change (right panel) in the Netherlands according to national definition (NAC) and the definition adopted in the US.

Figure 12 corroborates the presence of a reliable Phillips curve in the Netherlands, at least between 1970 and 2001; the DGDP time series is taken from the OECD data set and the unemployment readings obey the national definition. It is worth noting that the unemployment is three years ahead of the inflation. Therefore, one can predict inflation in the Netherlands using measured unemployment, but not vice versa. Results of linear regression of the unemployment on the inflation are shown in the right panel:  $R^2=0.79$  for the entire period between 1971 and 2004, with RMSFE at a 3-year horizon of 1.6%. The measured unemployment changed from 0.9% (!) in 1971 to 10.2% in 1983. If to exclude three recent points from the unemployment curve,  $RMSFE=1.3\%$  and  $R^2=0.89$ . The year of 2001 is likely to be associated with a break in measurement units as Figures 10 and 11 show, because of the introduction of new statistics.

The unemployment rate and the change in labor force both are three years ahead of the inflation. Thus, it is natural to use the generalized model to predict inflation. In the left panel of Figure 13 we present the observed GDP deflator and that predicted from the labor force for the years between 1971 and 2004. Because of high fluctuations in the labor force time series (see Figure 10) the predicted curve is smoothed with MA(5). Finally, the right panel in Figure 13 evaluates the predictive power of the generalized model. The agreement is excellent with a possibility to predict at a three year horizon, when accurate measurements of unemployment and labor force will be available. In a sense, the prediction from labor characteristics has an uncertainty similar to that associated with the definition of inflation – compare the curves in Figures 11 and 13.

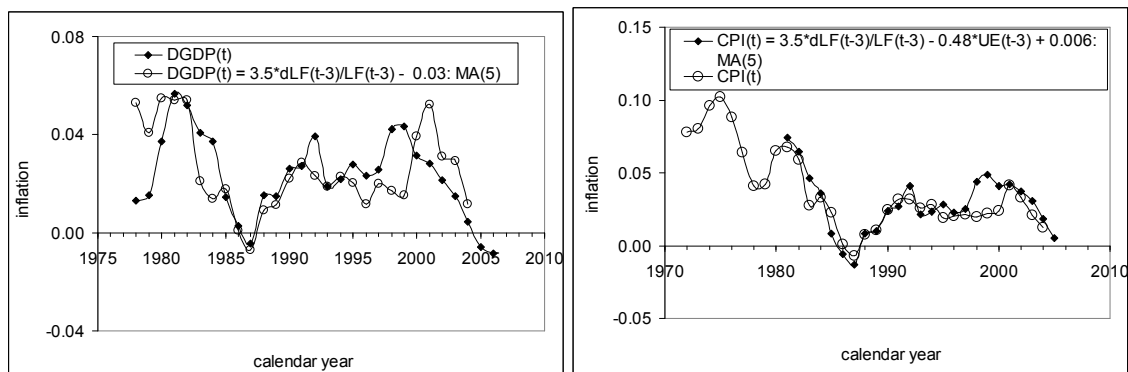


**Figure 11.** Two measures of inflation in the Netherland; GDP deflator and CPI inflation. Both are reported by the OECD.



**Figure 12.** Phillips curve in the Netherlands.

So, one might consider the case of the Netherlands as an evidence in favor of the generalized model. One more developed country reveals the presence of a long-term equilibrium link between inflation, unemployment, and the level of labor force. The goodness-of-fit between the observed and predicted time series in the right panel of Figure 13 is ( $R^2=$ ) 0.77 for the period between 1980 and 2006. We do not conduct tests for cointegration because the samples are small for any robust inferences.



**Figure 13.** Inflation in the Netherlands (DGDP and CPI) as a function of the change in labor force and unemployment.

**Sweden**

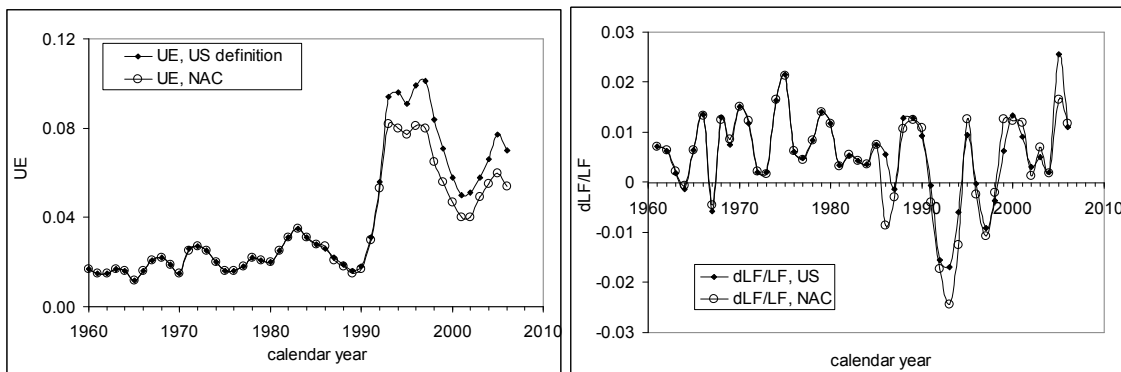
For Sweden, two different versions of unemployment rate and labor force are available – the NAC and US ones. Figure 14 displays the evolution of all series between 1960 and 2006. Surprisingly, the curves are very close with only one visible step in 1993, as reported by the OEDC (2008):

**Series breaks:** In 1993 a new reference week system and new estimation procedures were introduced. Also, the definition of unemployed was adjusted so that it followed the recommendations of the ILO more closely. In the new reference week system, the Labour Force Survey measures all weeks during the year as opposed to two weeks per month in the older system. In 1987 a new questionnaire was introduced resulting in the presentation of additional variables, and in the establishment of dependent interviewing. From 1986 to 2006, data refer to all persons aged 16 to 64 years; previously they referred to all persons aged 16 to 74 years. As from October 2007 the data cover age group 15 to 74 again.

The Phillips curve in Sweden is not a reliable link between unemployment and inflation, as Figure 15 shows. The goodness-of-fit is only 0.67 for the years between 1971 and 2006. The unemployment curve does not match the fluctuations in the inflation curve. It is likely that the definition of unemployment before 1987 was not adequate – it essentially produced a straight line from 1960 to 1980.

It seems that the definition of labor force has the same problem as the definition of unemployment, as the left panel of Figure 16 demonstrates. The observed curve diverge after 2001 from that predicted using labor force only. (This is the same problem as for the Netherlands.) Otherwise, the agreement is good. As in many European countries, the change in labor force leads the change in unemployment; the former is two years ahead of the latter in Sweden.

The generalized model links all three involved macroeconomic variables much better than individual relationships. The right panel of Figure 16 evidences a reliable description of the inflation by the unemployment and the change in labor force. Because of fierce fluctuations in the inflation and labor force moving average MA(5) and MA(7) are applied to the original series. In Sweden, an increase in the level of labor force reduces the rate of unemployment and pushes inflation up, as coefficients in individual relationships between relevant variables shows.



**Figure 14.** Unemployment rate (left panel) and the rate of labor force change (right panel) in Sweden according to national definition (NAC) and the definition adopted in the US.

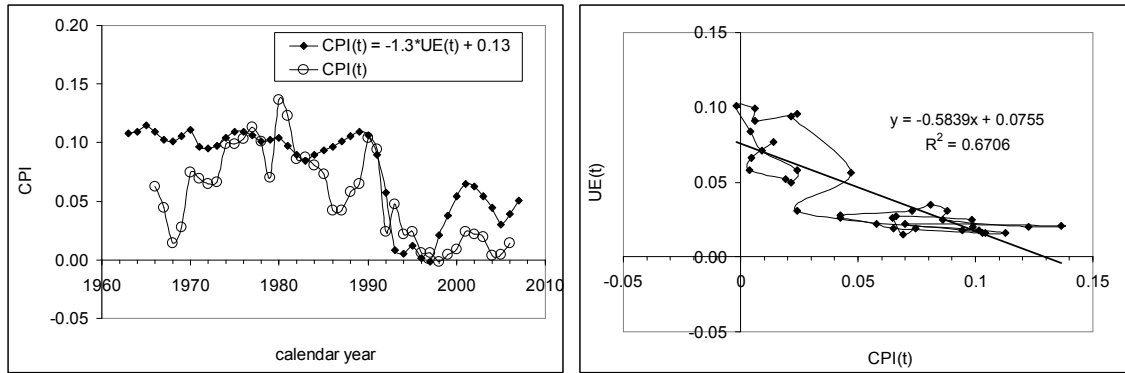


Figure 15. The Phillips curve in Sweden.

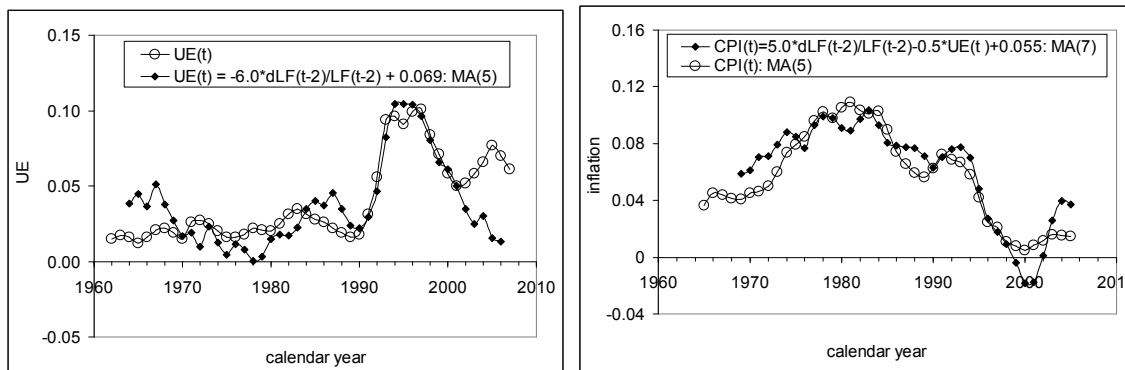


Figure 16. Left panel: Prediction of unemployment by the change in labor force. Right panel: The generalized model in Sweden: inflation as a function of the change in labor force and unemployment).

### Switzerland

The OECD reports measurements of unemployment in Switzerland only from 1991, as Figure 17 depicts. The BLS does not provide unemployment readings for this country at all. This makes the construction of a generalized model meaningless at this point in time. Longer observations of labor force (right panel, Figure 17) and inflation (Figure 18) are available, however. Apparently, the labor force series has two breaks: one in 1974 of unknown nature and one in 1991, as the OECD (2008) informs:

**Series breaks:** From 1998, data are adjusted in line with the 2000 census. Prior to 1991, data refer only to persons who are gainfully employed at least six hours per week.

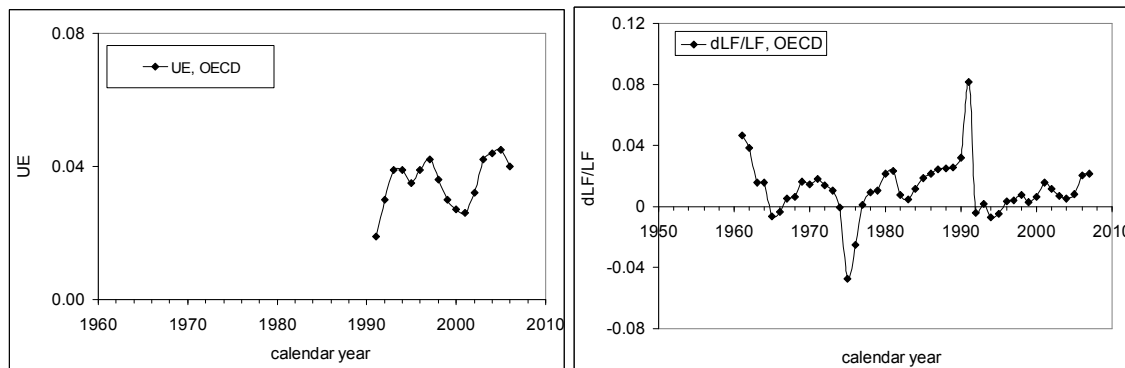
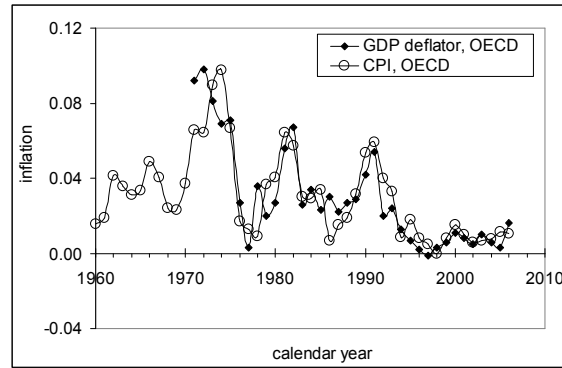
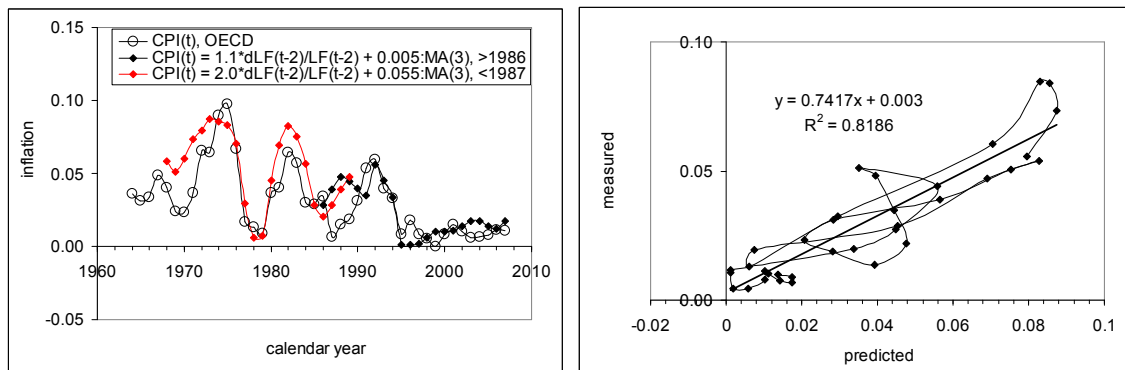


Figure 17. Unemployment rate (left panel) and the rate of labor force change (right panel) in Sweden according to national definition (NAC) and the definition adopted in the US



**Figure 18.** Two definitions of the rate of price inflation in Switzerland: GDP deflator and CPI inflation according to OECD definition.

The link between inflation and labor force also has a break around 1987, as Figure 19 depicts. Linear regression of the observed series on the predicted one is characterized by slope 0.74, free term 0.003, and  $R^2=0.82$ . According to well-know problem with OLS, the slope is underestimated. Otherwise, the agreement is excellent. We did not use the cumulative curves for the estimation of coefficients in (1) for Switzerland since corresponding time series are not long enough to provide a robust estimate. On the other hand, the original inflation curve heavily oscillates, and one needs only to fit the peaks of the oscillations in order to find appropriate coefficients in (1), as shown in the Figure.



**Figure 19.** Left panel: The rate of CPI inflation in Switzerland as predicted by the generalized model with a structural break neat 1987 related to the change in measuring units. Notice that the predicted series is smoothed with MA(3). Right panel: Linear regression of the data in the left panel.

### The United States

In 2006, we presented a prediction of unemployment at a six-year horizon (Kitov 2006a, 2006b). Despite the US is not a European country, this is a good opportunity to extend the prediction by four new readings and check the accuracy of the previous prediction. There are two possibilities to predict unemployment in the USA – using the dependence on the change in labor force and the generalized model. Figure 20 and 21 displays pertinent dynamic and cumulative curves with their equations. As predicted, the measured unemployment has been decreasing since 2004. The accuracy of the forecast increases when the measured and predicted curves are smoothed with MA(5) and MA(3), respectively. This is the effect of suppression of measurement noise by destructive interference. Considering the lag of six years (actually 22 quarters) such smoothing does no harm to the out-of-sample prediction.

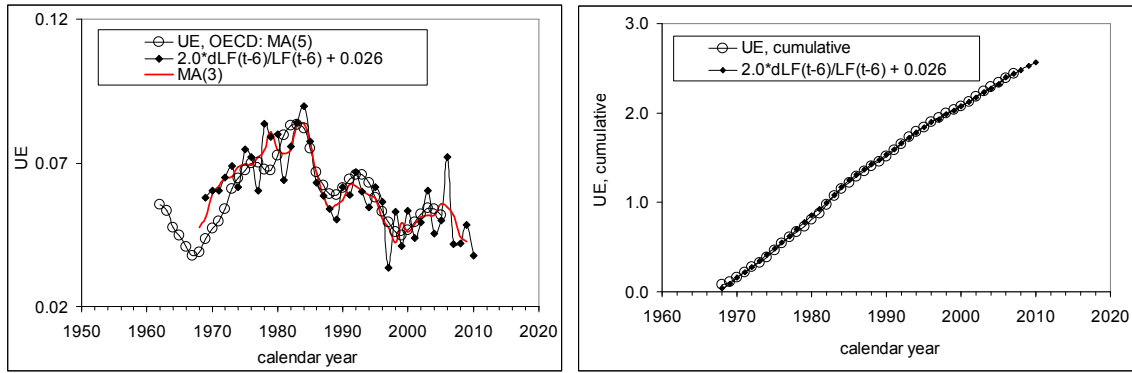


Figure 20. Prediction of unemployment USA from the change in labor force.

The second half of 2008 and the beginning of 2009 are characterized by an unprecedented fast growth in unemployment (not seasonally adjusted), from 5.2 % in May 2008 to 8.9% in February 2009. The predicted curves in Figures 20 and 21 do not show any sign of so high rate in the long-run. Therefore, both models predict that the unemployment should return to the level ~5% in near future.

Previously, we made a prediction of unemployment at a longer horizon using various projections of labor force. It is time to update the forecast and to check whether one should expect the transition to a deflationary period since 2012. Figure 22 presents a number of prediction curves with data borrowed from the Congressional Budget Office (2004). According to our model, deflation in the United States is inevitable.

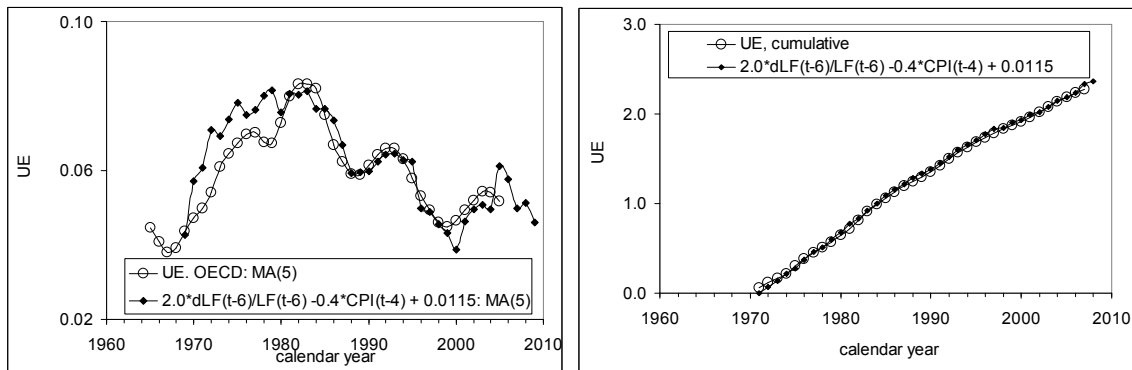


Figure 21. Prediction of unemployment in the USA by the generalized model.

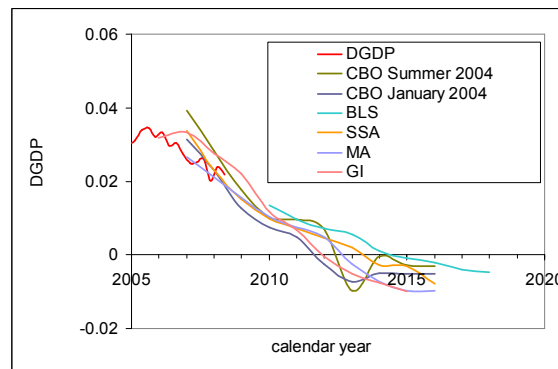


Figure 22. The evolution of GDP deflator predicted using various projections of labor force. Sources: Congressional Budget Office; Department of Labor, Bureau of Labor Statistics (BLS); Social Security Administration (SSA); Macroeconomic Advisers (MA); Global Insight (GI).

## 5. Conclusion

We have presented an empirical model explaining the evolution of inflation and unemployment in developed countries as driven by the change in the level of labor force. The model was previously tested on data from the biggest economies – United States, Japan, Germany, and France. Smaller economies of Austria and Canada also support the existence of the link.

In order to validate the model and facilitate the procedure of the estimation of the model coefficients we have introduced and tested a new technique – the method of cumulative curves. This method is a direct analog of the boundary elements methods extensively used in engineering and science. Using Austria and the United States as examples, we have demonstrated that cumulative curves provide particular solutions of the model equations and guarantee the residual of the model to be a I(0) process. In other words, the difference between measured and predicted cumulative curves, both I(2) processes, is a stationary process. This is strong evidence in favor of the existence of a causal link between labor force and the pair inflation/unemployment.

Instead of repeating major and minor conclusions made in our previous papers we would like to summarize all results obtained so far as a complete list of empirical models derived for developed countries (in alphabetic order):

Austria:

$$\begin{aligned}\pi(t) &= 1.2dLF(t)/LF(t) - 1.0UE(t) + 0.066; 1965 \leq t \leq 1986 \\ \pi(t) &= 0.9dLF(t)/LF(t) - 1.0UE(t) + 0.0074; t \geq 1987\end{aligned}$$

Canada:

$$\begin{aligned}UE(t) &= -2.1dLF(t)/LF(t) + 0.12 \\ \pi(t) &= 2.58dLF(t-2)/LF(t-2) - 0.043 \\ \pi(t) &= 3.8dLF(t-2)/LF(t-2) + 0.79UE(t-2) - 0.098\end{aligned}$$

France:

$$\pi(t) = 4.0dLF(t-4)/LF(t-4) - 1.0UE(t-4) + 0.095$$

Germany:

$$\begin{aligned}\pi(t) &= -1.71dLF(t-6)/LF(t-6) + 0.041 \\ UE(t) &= 2.5dLF(t-5)/LF(t-5) + 0.04 \\ UE(t-1) &= -1.50\pi(t) + 0.116; t > 1971 \\ \pi(t) &= -0.3dLF(t-6)/LF(t-6) + 0.59UE(t-1) + 0.072\end{aligned}$$

Italy:

$$UE(t) = 3.0dLF(t-11)/LF(t-11) + 0.085; t > 1968$$

Japan:

$$\begin{aligned}UE(t) &= -1.5dLF(t)/LF(t) + 0.045 \\ \pi(t) &= 1.77dLF(t)/LF(t) - 0.0035\end{aligned}$$

The Netherlands:

$$\begin{aligned}\pi(t) &= 3.5dLF(t-3)/LF(t-3) - 0.03 \\ \pi(t) &= 3.5dLF(t-3)/LF(t-3) - 0.48UE(t-3) + 0.006\end{aligned}$$

Sweden:

$$\begin{aligned}\pi(t) &= 1.15UE(t) + 0.11 \\ UE(t) &= -6.0dLF(t-2)/LF(t-2) + 0.069 \\ \pi(t) &= 5.0dLF(t-2)/LF(t-2) + 0.044 \\ \pi(t) &= 5.0dLF(t-2)/LF(t-2) - 0.5UE(t) + 0.006\end{aligned}$$

Switzerland:

$$\begin{aligned}UE(t) &= -1.0\pi(t) + 0.04 \\ \pi(t) &= 2.0dLF(t-2)/LF(t-2) + 0.005 (t \leq 1986)\end{aligned}$$

$$\pi(t) = 1.1dLF(t-2)/LF(t-2) + 0.055 \quad (t \geq 1987)$$

The United States:

$$UE(t) = 2.1dLF(t-5)/LF(t-5) - 0.023$$

$$\pi(t) = 4.0dLF(t-2)/LF(t-2) - 0.03$$

These developed countries produce a larger portion of the world's GDP. Most of these countries are situated in Western Europe. They had various economic and social histories in the 20<sup>th</sup> century and in the first decade of the 21<sup>st</sup> century. Nevertheless, they all demonstrate similar links between inflation, unemployment and labor force. The empirical models for Australia, Spain, and the United Kingdom are under construction.

## References

- [1] Kitov, Ivan. 2006a. *Inflation, unemployment, labor force change in the USA*, Working Papers 28, ECINEQ, Society for the Study of Economic Inequality.
- [2] Kitov, Ivan. 2006b. *Exact prediction of inflation in the USA*, MPRA Paper 2735, University Library of Munich, Germany.
- [3] Kitov, Ivan, Kitov, O., Dolinskaya, S. 2007a. *Inflation as a function of labor force change rate: cointegration test for the USA*, MPRA Paper 2734, University Library of Munich, Germany.
- [4] Kitov, Ivan. 2007a. *Inflation, Unemployment, Labor Force Change in European countries*, in T. Nagakawa (Ed.), Hauppauge NY: Nova Science Publishers, *Business Fluctuations and Cycles*: 67-112,
- [5] Kitov, Ivan, Kitov, O., Dolinskaya, S. 2007b. *Relationship between inflation, unemployment and labor force change rate in France: cointegration test*, MPRA Paper 2736, University Library of Munich, Germany.
- [6] Kitov, Ivan. 2007b. *Exact prediction of inflation and unemployment in Japan*, MPRA Paper 5464, University Library of Munich, Germany.
- [7] Kitov, Ivan. 2007d. *Exact prediction of inflation and unemployment in Germany*, MPRA Paper 5088, University Library of Munich, Germany.
- [8] Kitov, Ivan. 2007c. *Exact prediction of inflation and unemployment in Canada*, MPRA Paper 5015, University Library of Munich, Germany.
- [9] Kitov, Ivan. 2009b. *Mechanical model of personal income distribution*, MPRA Paper 13422, University Library of Munich, Germany.
- [10] Banque de France. 2004. *Definition of the single monetary policy*. Retrieved January 10, 2007 from <http://www.banque-france.fr/gb/instit/sebc/1d.htm>
- [11] Kitov, Ivan. 2009a. *The anti-Phillips curve*, MPRA Paper 13641, University Library of Munich, Germany.
- [12] Atkeson, Andrew, Ohanian, Lee E. 2001. "Are Phillips curves useful for forecasting inflation?", *Federal Reserve Bank of Minneapolis Quarterly Review* 25: 2-11.
- [13] Stock, H. James, Watson, W. Mark, 1999. "Forecasting inflation", *Journal of Monetary Economics* 44: 293-335.
- [14] Stock, H. James, Watson, W. Mark. 2005. *Has inflation become harder to forecast?*, manuscript, Harvard University, [http://ksghome.harvard.edu/~JStock/pdf/qepd\\_sw\\_jmcb\\_submission\\_1.pdf](http://ksghome.harvard.edu/~JStock/pdf/qepd_sw_jmcb_submission_1.pdf).
- [15] OECD. 2008. Notes by Country, <http://www.oecd.org/dataoecd/35/3/2771299.pdf>
- [16] Congressional Budget Office. 2004. CBO's Projections of the Labor Force, <http://cbo.gov/ftpdocs/58xx/doc5803/09-15-LaborForce.pdf>



## ADVERSE EFFECTS OF CORPORATE LENIENCY PROGRAMS IN VIEW OF INDUSTRY ASYMMETRY

Evgenia **MOTCHENKOVA**, Daniel **LELIEFELD**  
VU University Amsterdam, **NETHERLANDS**,  
Dutch Antitrust Authority (NMa), **NETHERLANDS**  
[emotchenkova@feweb.vu.nl](mailto:emotchenkova@feweb.vu.nl) , [d.b.liefeld@nmanet.nl](mailto:d.b.liefeld@nmanet.nl)

### Abstract

*This paper studies the effects, leniency programs have on cartel stability and the subsequent abuse of market power. A game-theoretical model, which allows for asymmetry and retaliation, is employed to analyze this problem. We find that a leniency program does not always lead to a breach of trust; in certain industries leniency programs are unable to break collusion. Moreover, they may have the adverse effect in the sense that they strengthen cartel stability or lead to abuse of market power. A relatively large firm can use coercion to remove the option to a smaller firm to self-report. In industries characterized by a certain degree of asymmetry in market shares and high exit costs this is an even more likely scenario. In view of this limitation, policies aimed at the removal of the threat of retaliation need to be considered. This paper's emphasis is placed on leniency programs for cartels, but the line of reasoning may also be extended to corporate whistle-blowing programs.*

**Keywords:** Antitrust Policy, Antitrust Law, Self-reporting, Leniency Programs

**JEL Classification:** K21, L41

### 1. Introduction

Leniency programs form a relatively new feature of antitrust law enforcement. Its main objective is to remove trust between cartel members. Trust is an essential element of every conspiracy. A similar approach is successfully employed in the prosecution of the mafia (the so called “witness protection program”). In practice, contact is established between a member of a conspiracy and the justice department with a proposition to serve as a witness against its co-conspirators. As a reward the witness receives (partial) amnesty from its own misconduct and protection from punishment by the other members of the (former) crime syndicate. In the ‘80’s and ‘90’s of the previous century the witness protection program proved to be a great success. Even though the trust between (family) members of the mafia was relatively strong the program enabled the successful conviction of a great number of criminals. No wonder a similar approach was adopted by the antitrust authorities in the fight against organized infringements of competition law. This paper examines the consequences of the introduction and use of leniency programs in their attempt to remove trust between cartel members. The emphasis of this paper is on leniency programs for cartels and the prosecution of price-fixing agreements. The reasoning can, however, also be applied to corporate whistle blowing programs and witness protection programs.

Several economists have previously expressed their doubts as to the success of leniency program’s application in this new industrial setting. This literature is reviewed in Spagnolo (2008) and includes such contributions like Buccorossi and Spagnolo (2006) or Ellis and Wilson (2001). The main intuition they provide in order to explain the existence of counterproductive effects of leniency programs is that in the framework of repeated interactions leniency may increase the firms’ ability to punish deviations (e.g. threatening to report to the antitrust authority after cheating occurs), thereby stabilizing the cartel agreements by reducing gains from defecting. In this way leniency can reduce the incentives to cheat and make cartels more stable. The current paper adds to these considerations by addressing another issue that has been underestimated in the design of leniency programs. An important element of the witness protection program, which is missing in the design of leniency programs for cartels, is to safeguard the witness from punishment by its former partners in crime. We show that, using the realistic assumption of industry asymmetry, the current design of leniency programs can’t prevent larger firms from using a threat of punishment as a means of coercion, effectively not allowing smaller firms to appeal for leniency. When the antitrust authority is unable to credibly protect leniency applicants from retaliation by convicted cartel members the program is abused by cartels. It actually serves to strengthen trust between its members. The program may have the adverse effect in the sense that it may facilitate cartel stability.

It should be stressed that any type of leniency program should contain the following crucial elements. Firstly, the criminal should provide sufficient evidence on the misconduct by its former partners to the authorities. Secondly, the criminal acts as a witness and receives a, previously agreed to, lenient treatment with regard to the criminal's punishment. This ranges from a reduction to a fine to clemency from a prison sentence and can even entail a reward (*see e.g. Kovacic 2006*). A discretionary approach ensures the incentive is set according to the constraint faced by the proposed witness and no resources are wasted. Thirdly, the witness is protected from punishment by its former co-conspirators. When all policy parameters of these elements are customized to fit the typical case, the proposed witness accepts the offer, the crime syndicate is terminated, its members are convicted, no resources are wasted and, perhaps most importantly, an example is set for those firms or individuals considering the prospect to conspire.

Leniency programs for cartels differ from witness protection programs in two main aspects, first of all the protection aspect and secondly the fact that leniency programs are less customized. Both parts of the policy instrument have, however, proven to be detrimental to its success. In its new design and new environment these parameters have unfortunately remained underexposed. It can be shown that, unless the necessity to protect the reporter from punishment is acknowledged by the authorities, cartels may strengthen their ties by means of the leniency program. Customization of the program with respect to the size of the fine and protection after self-reporting can partially overcome this deficiency in the program. More generally, the leniency program will always need to be accompanied by the traditional law enforcement efforts of the competition authority.

A typical punishment strategy a firm might consider employing involves dumping or aggressive pricing. Not only can a "larger" (more efficient) firm usually establish lower marginal costs, the market share asymmetry also enables such a firm to establish a larger buffer than the smaller firm. When the larger firm employs its aggressive punishment strategy both firms incur losses. These eat away at both firms' buffers. Since the larger firm's buffer is generally greater it will be able to sustain these losses for a longer period of time. Employing aggressive strategy for a sufficient period of time will eventually push the smaller firm into bankruptcy. This set-up resembles the "long-purse story" explanation of predatory pricing (*see e.g. Stigler 1964 or Tirole 1988*).

The next point we want to address is the role of asymmetry for cartel stability. The bigger part of the literature on leniency programs uses undiversified companies and industry symmetry as a starting point. The main reason behind this is the general perception that asymmetry reduces cartel strength. Leading textbooks such as Tirole (1988) state that, based on work by e.g. Chamberlin (1929) and Stigler (1964), the coordination towards a focal price under differentiated costs and products is more difficult. Motta and Polo (2003) argue that asymmetries between large and small firms represent an obstacle for industry wide collusion. In reality firms are rarely truly symmetrical in their cost functions, products or market presence. Asymmetry is the rule rather than the exception. Symmetry would imply that all colluding firms apply for leniency at the same time. This is rarely the case. In general though symmetry increases the incentive to collude, but this scenario is not realistic and the introduction of a leniency program by an antitrust authority might change the incentives for firms. Asymmetry among firms in products, markets and cost functions is the subject of this paper. We model asymmetry by considering firms that have different market shares due to e.g. historic reasons.

The fact that a cartel can be stable under asymmetrical conditions was recently proved by the EU Competition Directorate General that uncovered a beer cartel in Holland and (parts of) Belgium in which small players coexisted with larger firms (*see EC decision 2007, reference IP/07/509*). An illustration of coercion maintained by the threat of retaliation can be found in the leniency application of British Petrol (BP) in the Bitumen Cartel (*see EC decision 2006, reference MEMO/06/324*). During its existence the colluders managed to increase trust between the cartel's members by designing a collective punishment strategy. Each time a cartel member violated the cartel's agreements the other members agreed to punish the perpetrator. The cartel managed to create a threat of retaliation by joining forces, using an asymmetry of power, sustained by formal trust.

The literature on analysis of self-reporting schemes in antitrust starts with the paper by Motta and Polo (2003). They study a two-stage game in which the AA first chooses once and for all its antitrust policy followed by the competition phase in which the firms compete with each other, which is modelled as an infinitely-repeated oligopoly game. The cartel adopts grim-trigger strategies in which cheating on the cartel by either setting a different price or applying for leniency triggers competitive

behaviour forever, while the cartel continues collusion as usual each time it is caught by the AA. Under the optimal antitrust policy, introduction of ex-post leniency programs will increase the chance of the cartel being captured, but ex-ante leniency programs that grant reduced fines are ineffective. As later shown in Spagnolo (2004) and Rey (2003), effective ex-ante leniency programs require substantial rewards. This is also recommended in Kovacic (2006).

Spagnolo (2004) concludes that courageous leniency programs are closest to the optimal. He uses a game theoretical model to relate a first best “courageous” leniency scheme and a “moderate” leniency scheme to a benchmark case of traditional law enforcement. The courageous program is one in which the reporting party is actually rewarded with a part of the fine paid by the other parties besides receiving amnesty. In this way a first best solution is established according to Spagnolo.

The above arguments are closely related to the more general analysis of the optimal structure and design of leniency programs that has been extensively discussed in the literature. *See*, for example, Rey (2003), Hinlopen (2003), Feess and Walzl (2004), Motchenkova (2004), Buccorosi and Spagnolo (2006), Spagnolo (2008), Harrington and Chen (2007), Chen and Rey (2007), Harrington (2008), and Houba *et al.* (2009). The question of optimal design of leniency programs has two main debatable components. They are the number of fine reductions and the size of fine reductions. The research normally comes to the conclusion that the currently applied moderate leniency program could be made more effective by limiting a fine reduction to the first firm to report. Adding more than one possibility to a fine reduction reduces the deterring effect of the scheme. Most of the analysis also concludes that the first self-reporter should be fully exempted from the fine.

The current paper analyzes the effects of leniency programs on the survival of cartels formed by firms of different size (i.e. different market shares). We analyze a setup where the bigger firm can threaten to force the smaller firm to exit the market by employing aggressive strategy in case the smaller firm self-reports.<sup>82</sup> The innovative aspect of the paper relates to considering how cartel members can react to the introduction of leniency programs in order to preserve the continuation of collusion. The punishment strategy for self-reporters is intended to counterbalance the incentives to reveal information to the antitrust authority. We find that a leniency program does not always lead to a breach of trust; in certain industries leniency programs are unable to break collusion. Moreover, they may have the adverse effect in the sense that they strengthen cartel stability or lead to abuse of market power. A relatively large firm can use coercion to remove the option to a smaller firm to self-report. In view of this limitation, policies aimed at the removal of the threat of retaliation need to be considered.

The structure of the paper is as follows. Section 2 contains a formal description of the model. In Section 3, we solve the model and find sub-game perfect equilibriums of the game. Finally, in Section 4 the policy implications are discussed and the analysis is concluded. In the appendix a comparative analysis of the approach to leniency programs used in the United States and in several European countries is provided.

## 2. The Model (Formal Analysis)

We consider two asymmetric firms, which may form a cartel, taking into account the enforcement activity of the antitrust authority. The asymmetry is related only to the size of the firms or their market shares, while it is assumed that firms have identical marginal costs. The antitrust authority commits to a certain enforcement policy, which uses leniency programs. Leniency programs grant either complete or partial exemption from fines to the firms, which reveal the existence of a cartel to the antitrust authority and come up with sufficient evidence. The main innovation of this model is that we consider firms that have different market shares. This implies different accumulated profits during the period of collusive pricing. Hence, unless the antitrust authority (AA) is able to remove any asymmetry in the accumulated profits (buffer) of each individual member of cartel, some “bigger” members enjoy a strategic advantage. A firm with a relatively large buffer will be able to employ the difference in buffer size as a means of coercion, such as the threat of punishment through aggressive pricing in case the rival deviates from cartel agreement by self-reporting. Essentially,  $K$  denotes the costs of aggressive behaviour (through e.g. setting price below marginal costs for some periods) for the bigger firm. These costs are high when firms are more symmetric and, *vice versa*, these costs are low when asymmetries are high.

---

<sup>82</sup> In this paper we adopt the legal definition of aggressive (predatory) pricing, which is characterized as setting price below the marginal cost.

The credibility and impact of this type of pricing strategy depends on the asymmetry in size between firms, such as the difference in market shares. Market shares are denoted by  $\beta$  for “bigger” firm and by  $1-\beta$  for “small” firm, with  $\beta > 1-\beta$  and  $0 < \beta < 1$ .

First, we describe the policy choices of the antitrust authority. Second, we describe the timing of the game. And, finally, we specify the firms’ strategies.

**Enforcement policy:** The main goal of the antitrust authority is to prevent the formation of cartels in the first place. However, if the cartel has already been formed, the antitrust authority aims to break the trust at the lowest possible cost. Here, following the reasoning in Section 1, we restrict the number of fine reductions in case of multiple applications for leniency to one. Only the first reporter gets complete exemption from the fine. This, as explained above, reduces trust among cartel members. This set-up is also motivated by the fact that the structure of leniency programs employed in US allows only for one fine reduction. Moreover, the US scheme also has a longer history than its European counterpart and has proven to be more successful.<sup>83</sup> Following Becker (1968), we distinguish two main parameters of enforcement policy: penalty and probability of detection. Hence, the antitrust policy in the presence of leniency programs can be described by the following parameters:

- The full fines  $F = \alpha\pi$ , which are proportional to illegal gains for firms that were proven guilty and have not cooperated with the antitrust authority, or are not the first to come forward with information about cartel.<sup>84</sup> Here  $\alpha$  is the coefficient of proportional fine.  $\pi$  denotes per period illegal profits from cartel formation. Competitive profits ( $\pi_n$ ) are assumed to be zero for simplicity. So that  $\pi$  can also be viewed as pure illegal gains. Note also that  $\pi_m$  is maximal per period payoff for each firm in case of full collusion (i.e. when firms are able to charge monopoly prices);
- The reduced fine  $f$  specified by the US leniency program is equal to zero.<sup>85</sup> This set-up allows for the strictest adherence to the leniency rules;
- The probability of law enforcement by the antitrust authority equals  $p \in (0, 1]$ . This variable can be thought of as an instantaneous probability that the firm is checked by antitrust authority and found guilty. Contrary to Motta and Polo (2003), we assume that whenever the antitrust authority checks the guilty firm, the violation is successfully discovered. Moreover, we assume that  $p$  is determined by e.g. an exogenous budget of the antitrust authority financed by the government that can be used to promote enforcement, so that  $p$  reflects the costs of efforts of antitrust authority put into law enforcement activities.

**Timing of the game:** Two asymmetrical firms play the two stage game in the presence of antitrust enforcement which incorporates leniency programs.

At time  $t=0$  the antitrust authority sets parameters of the enforcement policy:  $F = \alpha\pi$  and  $p$  and parameters of the leniency program (which allows for only one fine reduction and reduced fine  $f$  equals 0). So, self-reporting becomes an attractive option at this stage. Prior to this stage  $t < 0$  firms may decide to form a collusive agreement. As conventional analysis of super-games (see Tirole 1988) implies, in the absence of the antitrust enforcement, collusion can arise in equilibrium only when the discount factor is large enough, namely,  $\delta \geq \beta = \delta_c$ .<sup>86</sup> So, for further analysis we will direct our attention to the values of the discount factor  $\delta \geq \delta_c$ , which ensures that cartels are stable in the absence of antitrust enforcement and, hence, the first stage of the “revelation-retaliation” game is reached.

Next, the game between the two asymmetrical firms is played. At time  $t=1$  (stage 1 of the game) the small firm moves.<sup>87</sup> It can choose between two actions: self-report or keep cartel secret.

<sup>83</sup> See historical overview and overview of structures of leniency programs in Appendix.

<sup>84</sup> The motivation for this type of structure and example of application of proportional penalty schemes in competition law enforcement was studied in Motchenkova and Kort (2006).

<sup>85</sup> See also Appendix.

<sup>86</sup> In the absence of any antitrust enforcement, i.e. when neither fines nor rate of law enforcement can be used, collusion can be sustained only when the short run gain from an unilateral deviation from collusive agreement by undercutting in prices together with competitive profits thereafter is smaller than the payoff from sustaining collusive strategy forever:  $\beta\pi_m/(1-\delta) > \pi_m + \delta\pi_n/(1-\delta)$  for  $i=1,2$ . Hence, with competitive profits  $\pi_n$  normalized to 0, we have  $\delta \geq 1-\beta$  for “bigger” firm and  $\delta \geq \beta$  for “small” firm. The second constraint is binding since  $\beta > 1-\beta$ . Hence, we have  $\delta_c = \beta$ .

<sup>87</sup> We assume here that incentives for the bigger firm to keep the cartel secret are always higher since it gets higher expected gains from continuation. So, the big firm would either need stronger incentives or will self-report only later in time than the small firm.

Further, at time  $t=2$  (stage 2 of the game) the big firm responds to the action of the small by choosing whether to punish the small firm for reporting the cartel or to abstain from punishment.

Note that the antitrust authority does not take an active part in the game. It only sets policy parameters,  $F, f, p, \alpha$ , and the rules of leniency programs. This complies with the current “one size fits all” setting of the antitrust policy parameters.

Payoffs of both players in each of the four possible cases are described in the following subsection. Each time we refer with “Small” to the smaller firm and with “Big” to the other player.

#### Strategies and Payoffs:

1. Small has decided to report and Big responds by setting a predatory price: Big receives its current share  $\beta\pi_m$  of collusive profits and the monopoly profits forever after (i.e.  $\pi_m(\delta/(1-\delta))$ ). But it has to overcome a loss of size  $K$  (due to aggressive pricing) and a one time fine of size  $\alpha\beta\pi_m$  and there is the risk of a fine when setting a predatory price of  $p\alpha\pi_m(\delta/(1-\delta))$ . The latter is the net present value of the expected fine Big might have to pay, because of the abuse of its dominant position. Small receives its current share  $(1-\beta)\pi_m$  but loses an amount of  $S$  (exit cost), since it has to leave the market. Small cooperated with the antitrust authority, so it is exempted from a fine.

2. Small has decided to self-report and Big decides not to retaliate and simply moves to marginal costs pricing (competitive equilibrium)<sup>88</sup>: Big receives its current share of collusive profits  $\beta\pi_m$  and is fined  $\alpha\beta\pi_m$ . Small receives its current share  $(1-\beta)\pi_m$ , but doesn't make any economic profit forever after. Since it reported to the antitrust authority it isn't fined.

3. Small has decided not to report and Big is inclined to use aggressive strategy. This means that aggressive pricing was attractive strategy already before the antitrust enforcement and leniency programs were introduced ( $t=0$ ). We will rule out this possibility later on when we discuss the solution of the game (it just imposes additional constraint on discount factor ( $\delta \leq \delta^{**}$ , see section below). In this case Big receives its current share of monopoly profits  $\beta\pi_m$  less a loss due to the aggressive pricing ( $K$ ), but after small leaves it will receive the entire (discounted) monopoly profit forever after  $\pi_m(\delta/(1-\delta))$ , though it also faces a risk of detection during the transition stage over its share of profits  $p\alpha\beta\pi_m$  and there is the chance of a fine in every period thereafter, which results in  $p\alpha\pi_m(\delta/(1-\delta))$ . Small receives its current share  $(1-\beta)\pi_m$  but loses its exit cost  $S$ . Since the firm is bankrupt the authorities cannot levy a fine on the firm for its misconduct.

4. Small has decided not to report and Big is inclined to continue the collusive price setting: Big receives its share of collusive profits forever  $\beta\pi_m(1/(1-\delta))$ , but faces the risk of being fined in every period there after. This results in  $p\alpha\beta\pi_m(1/(1-\delta))$ . Similarly, Small receives its current share forever  $(1-\beta)\pi_m(1/(1-\delta))$  but faces the risk of prosecution in every period after  $p\alpha(1-\beta)\pi_m(1/(1-\delta))$ .

It should be stressed that for any  $t > 2$ , decisions of both players do not change, and payoffs obtained at  $t=2$  will be discounted. This is due to the fact that the coefficient of the proportional penalty and the rate of law enforcement are fixed and, hence, the environment does not change. Moreover, we assume that in case of self-reporting trust is broken and firms do not go back to collusion ever again. Therefore, outcomes (1) and (2) are stable by assumption.

We summarize the above description of the game as follows:

**Stage 0:** The Antitrust Authority announces the parameters of the penalty scheme:  $p$  and  $F$ , and the parameters of leniency program:  $f=0$  and the number of fine reductions.

**Stage 1:** The smaller firm decide whether to reveal information about the existence of the cartel to the antitrust authority or not (once and for all decision).

**Stage 2:** The bigger firm observes the decision of the smaller firm and decides whether to punish it for self-reporting or not (once and for all decision).

If no self-reporting is chosen by the smaller firm and the bigger firm decides to continue collusion, then the repeated game, between authority and firms, where authority can discover violation with probability  $p$  in each period, is played till infinity (under assumption that even in case violation is discovered by antitrust authority, firms go back to collusion).<sup>89</sup>

The discount factor is denoted by  $\delta=1/(1+r)$ , where  $r$  is the interest rate. The game tree and players' payoffs are summarized in Figure 1.

<sup>88</sup> We assume here that in case one of the firms self-reported, trust will be broken forever and firms will not go back to collusion anymore.

<sup>89</sup> Similar assumption is employed in Motta and Polo (2003).

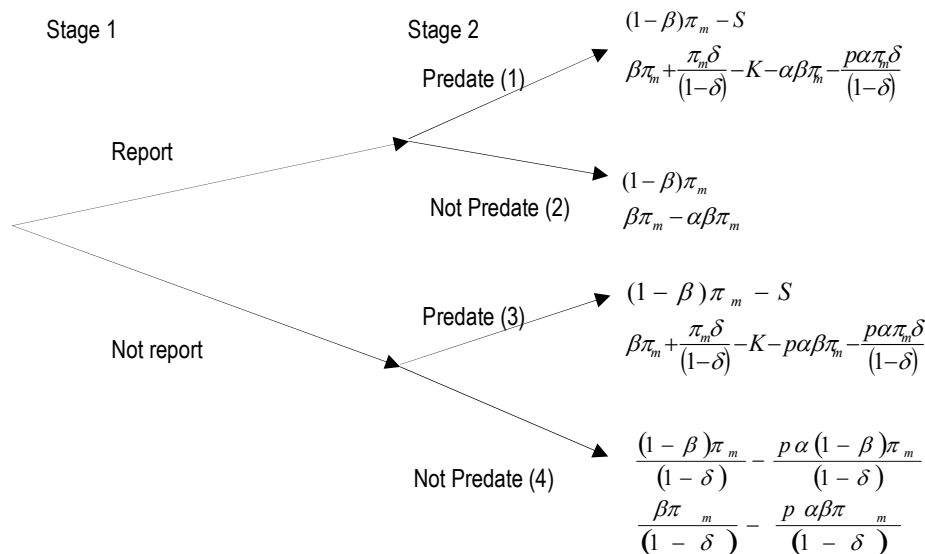


Figure 1. Game tree and players' payoffs.

We now proceed to establish the sub-game perfect equilibriums of the two-stage repeated game, which is described in Figure 1, played by both firms once the policy parameters are set.

### 3. Solution of the Game

To find the sub-game perfect equilibriums of the game we employ backward induction. First we consider the decision of the bigger firm which is taken in Stage 2, and next the decision of the smaller firm which is taken in Stage 1. Three different parts of the game in Figure 1 can be distinguished as sub-games. First, Small has a choice whether to report or not. It will base its decision on the reaction to its choice it expects from the bigger player. Therefore Big faces two possibilities. If Small has reported their common illegal conduct, Big has to decide whether to retaliate or not. In case of no retaliation, Big sets its price equal to marginal or variable costs so that small is “punished” for reporting to the antitrust authority, but keeps operating. In case of retaliation Small firm is driven out of the market. If Small decides not to report, Big faces the same decision. Clearly Small’s initial action depends on the Big’s reaction. Since it knows Big’s position and is aware of the value of all other parameters (e.g. law enforcement and discount value) it will choose to play accordingly.

#### 3.1. Collusion is the preferred strategy before leniency is introduced

We start by considering the choice of the bigger firm between retaliation and collusion (i.e. by comparing outcomes (3) and (4) described in previous section). Note that if outcome (4) is preferred over outcome (3) by the bigger firm, collusion is also the preferred strategy before leniency is introduced (under traditional antitrust enforcement). This happens when  $\beta\pi_m - K + \pi_m(\delta/(1-\delta)) - p\alpha\beta\pi_m - p\alpha\pi_m(\delta/(1-\delta)) > \beta\pi_m - K + \pi_m(\delta/(1-\delta)) - p\alpha\beta\pi_m - p\alpha\pi_m(\delta/(1-\delta))$ . This inequality implies that retaliation is more attractive than collusion for the bigger firm in both situations (with or without the availability of a leniency program) when the discount factor is greater than the following threshold:

$$\delta > \frac{K}{K + \pi_m(1 - \beta)(1 - p\alpha)} = \delta^{**}(K, p, \alpha). \quad (1)$$

So, for any values of the discount rate above the threshold  $\delta^{**}$ , there is no collusion before the leniency program is introduced and the game doesn’t apply. In the remainder of this paper all values of the discount rate over threshold  $\delta^{**}$  are regarded as values for the parameter  $\delta$ , in which equilibrium (3)

(no self-reporting, retaliation) arises. This equilibrium isn't of any interest to answering the questions posed above and will therefore be left out of the analysis.

Expression (1) gives us the first incentive compatibility constraint. It is represented in Figures 2 and 3 by the line  $\delta^{**}$ , which plots  $\delta(K)$  as a function of  $K$  in the  $(\delta, K)$ -plane. In addition, as discussed above, comparative statics of the behaviour of  $\delta^{**}(K, p, \alpha)$  with respect to the main parameters of the model shows that

$$\frac{\partial \delta^{**}(K, p, \alpha)}{\partial K} > 0 \quad \text{if } p\alpha < 1, \quad \frac{\partial \delta^{**}(K, p, \alpha)}{\partial K} < 0 \quad \text{if } p\alpha > 1, \quad \frac{\partial \delta^{**}(K, p, \alpha)}{\partial p} > 0, \quad \frac{\partial \delta^{**}(K, p, \alpha)}{\partial \alpha} > 0. \quad (2)$$

These inequalities show that the likelihood of collusion is increased further and predation is less likely when higher values of  $p$  and  $\alpha$  apply before the introduction of the leniency program. The first derivatives of  $\delta^{**}$  with respect to  $K$ ,  $p$  and  $\alpha$  are positive when  $p\alpha < 1$  (i.e. in the setting with low expected penalty). Meaning that raising either of these parameters will increase the height of this threshold, thus increasing the likelihood of the situation in which collusion is sustainable. Intuitively this makes sense, since retaliation is also illegal and increasing  $K$  implies more symmetry. This complies with general theory on collusion and symmetry (see Motta and Polo 2003).

Finally, in order to ensure consistent behaviour (meaning that collusion is sustainable and there are no incentives to retaliate in the absence of the possibility of self-reporting and subsequent clemency) we will consider only interval  $\delta_c < \delta < \delta^{**}$ , so that outcome (3) is ruled out and collusion is sustainable before the revelation game starts.

Recall from Section 2 (footnote 5) that  $\delta_c = \beta$ . Taking into account (1), this implies that interval  $[\delta_c, \delta^{**}]$  is not empty when  $\delta^{**} > \beta$ , i.e. when  $K > \pi_m \beta (1 - p\alpha)$  with  $p\alpha < 1$  or when  $K > |\pi_m \beta (1 - p\alpha)|$  with  $p\alpha > 1$ . This implies that the issue we are considering becomes especially sharp in industries characterized by relatively low asymmetry (i.e. where  $K$  is high).

### 3.2. Determination of other thresholds for equilibrium intervals

Big's choice between aggressive strategy and simple competitive strategy is based on a comparison of the outcomes (1) and (2). The outcome (2) in the model is the situation in which strategies (*report, not retaliate*) are used by the smaller firm and the bigger firm respectively. Outcome (1) in the model is the situation in which strategies (*report, retaliate*) are used. Big is not inclined to retaliate in case of reporting by Small when Big considers its payoff in equilibrium (2) to be higher than its payoff in equilibrium (1). The condition for equilibrium (*report, not retaliate*) to arise holds when the following inequality is satisfied:  $\beta\pi_m - \alpha\beta\pi_m > \beta\pi_m - K + \pi_m\delta/(1-\delta) - \alpha\beta\pi_m - p\alpha\pi_m\delta/(1-\delta)$ . This inequality implies that competitive pricing is more attractive for the bigger firm than retaliation after the smaller firm applied for leniency if the discount factor is less than the following threshold:

$$\delta < \frac{K}{K + \pi_m(1 - p\alpha)} = \delta^*(K, p, \alpha). \quad (3)$$

Differentiating this expression with respect to  $K$  implies that

$$\frac{\partial \delta^*(K, p, \alpha)}{\partial K} > 0 \quad \text{if } p\alpha < 1. \quad (4)$$

This implies that when  $p\alpha < 1$  (i.e. expected penalty is low) the equilibrium (2) is less likely to occur the smaller the size of  $K$ . Recall that  $K$  is the size of the buffer of Small, since it equals the cost of e.g. driving the smaller firm out of the market. After Small loses its buffer it can't sustain the losses associated with the aggressive behaviour by Big. Intuitively this means that the greater the size difference (asymmetry), the lower  $K$  and therefore threshold  $\delta^*$  will be lower when asymmetry is greater. It also implies that raising the risk of being fined will increase  $\delta^*$ . Intuitively it means that the smaller the asymmetry and the higher the chance of detection and substantial fine, the more likely the perceived discount rate is below the threshold  $\delta^*$ .

Next, we move to Stage 1 and consider the decision of the smaller firm given no retaliation is chosen by Big in the second stage of the game. Outcome (2) is preferred over outcome (4) by Small if the following inequality is satisfied:  $(1-\beta)\pi_m > (1-\beta)\pi_m/(1-\delta) - p\alpha(1-\beta)\pi_m/(1-\delta)$ . This inequality implies that self-reporting is more attractive for Small when the discount factor is lower than the following threshold:

$$\delta < p\alpha = \delta^{***}(K, p, \alpha). \tag{5}$$

This is a clear indication that raising the probability of detection and the proportional fine will make the smaller firm to choose equilibrium (2) over the payoff from equilibrium (4), and will therefore decide to self-report instead of continuing to collude.

Finally, we also have to compare the payoffs for Small in case outcome (1) arises and in case outcome (4) arises. Equilibrium (1) in the model is the situation in which strategies (*report, retaliate*) are employed by the smaller firm and the bigger firm respectively. Equilibrium (4) in the model is the situation in which strategies (*not report, not retaliate*) are used. Now the smaller player is confronted with a choice between being predated on by Big (and the associated exit cost) or going along with Big in the collusive price setting. The latter implies that Small prefers to choose a strategy leading to the collusive price setting over a strategy leading to bankruptcy. This occurs when the payoff of equilibrium (4) is higher than the payoff in equilibrium (1) for Small. I.e. the following inequality is satisfied:  $(1-\beta)\pi_m(1/(1-\delta)) - p\alpha(1-\beta)\pi_m(1/(1-\delta)) > (1-\beta)\pi_m - S$ . This inequality implies that collusion is more attractive for the small firm if discount factor is determined by the following inequalities:

$$\delta > \frac{p\alpha\pi_m(1-\beta) - S}{\pi_m(1-\beta) - S} = \delta^{****}(K, p, \alpha), \quad \text{when } S < \pi_m(1-\beta) \tag{6}$$

$$\delta < \frac{S - p\alpha\pi_m(1-\beta)}{S - \pi_m(1-\beta)} = \delta^{****}(K, p, \alpha), \quad \text{when } S > \pi_m(1-\beta) \tag{7}$$

Closer analysis of expressions (6), (7) and (5) shows the following regularities:

$\delta^{****} > \delta^{***}$  when  $p\alpha > 1$  and  $(1-\beta)\pi_m > S$ , or when  $p\alpha < 1$  and  $S \geq (1-\beta)\pi_m$ . In this case we have also that  $\delta^{****} > 1$ .

$\delta^{****} < \delta^{***}$  when  $p\alpha > 1$  and  $S \geq (1-\beta)\pi_m$ , or when  $p\alpha < 1$  and  $(1-\beta)\pi_m > S$ . In this case we have also that  $\delta^{****} < 1$ .

### 3.3. Derivation of Equilibrium Solutions

Next, once we have determined all the thresholds in terms of the discount factor, we can move to the description of equilibrium outcomes for each possible combination of the parameter values.

Firstly, two cases need to be distinguished: when  $p\alpha > 1$  and when  $p\alpha < 1$ . Inequality  $p\alpha > 1$  corresponds to the case when the expected penalty is already high enough to prevent any misconduct (in a static setting) in the absence of leniency programs. The other inequality corresponds to the situation when traditional antitrust enforcement is not strong enough.

We start our analysis with the discussion of a sufficiently high penalty (the case where  $p\alpha > 1$ ). In this setting two sub-cases depending on the size of  $\delta^{****}(K, S, p, \alpha)$  and on the size of exit costs can arise. When  $p\alpha > 1$  and  $(1-\beta)\pi_m > S$ , we obtain that  $\delta^{****} > 1$ , and the distribution of outcomes can be described as is done in the left hand side of Figure 2. When  $p\alpha > 1$  and  $(1-\beta)\pi_m < S$ , we obtain that  $\delta^{****} < 1$ . Hence, the distribution of outcomes is given by the right hand side of Figure 2.



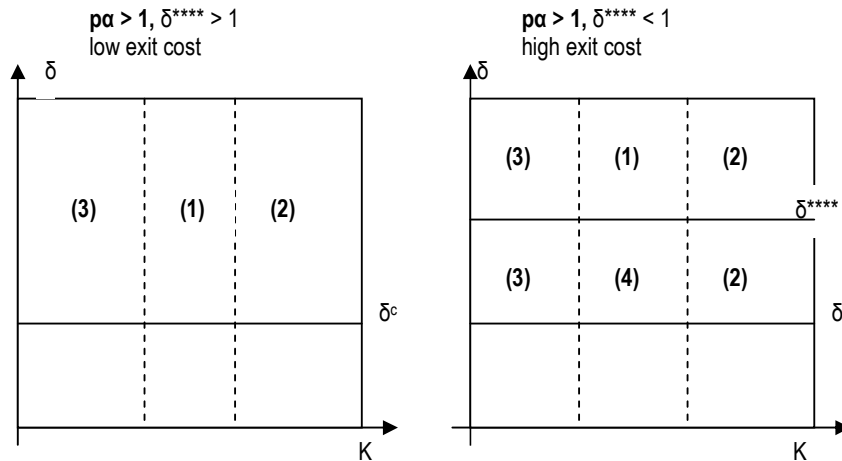


Figure 2. Equilibrium outcomes when  $pa > 1$ .

Figure 2 presents the locuses  $\delta^*$ ,  $\delta^{**}$ ,  $\delta^{***}$ , and  $\delta^{****}$  (derived in previous subsection and given by (3), (1), (5), and (6) respectively) in  $(K, \delta)$ -space. The left panel of Figure 2 implies that, in industries with low exit costs and relatively strong antitrust enforcement, depending on the degree of asymmetry the following three outcomes can arise. When there is high asymmetry ( $K < |\pi_m \beta(1-p\alpha)|$ ), outcome (3) will arise in equilibrium.<sup>90</sup> This means that in this industry retaliation is the most attractive strategy even before a leniency program is introduced. With an intermediate degree of asymmetry (i.e.  $|\pi_m \beta(1-p\alpha)| < K < |\pi_m(1-p\alpha)|$ ), outcome (1) arises in equilibrium.<sup>91</sup> In this case Big (strong) firm will choose to retaliate on a smaller firm after the latter chooses to self-report. This is the outcome the antitrust authority wants to avoid. In these types of industries a greater emphasis needs to be put on the protection part of a leniency program. Perhaps through stricter monitoring after a firm reported to the AA. Finally, when there are low asymmetries (i.e.  $K > |\pi_m(1-p\alpha)|$ ), outcome (2) will arise in equilibrium.<sup>92</sup> This implies that with high penalties in the industries with high  $K$  (or low asymmetries) there is no danger of retaliation or collusion. The first best outcome with self-reporting and competitive pricing afterwards arises. In this setting leniency programs appear to be effective.

The right panel of Figure 2 represents the results of the analysis for industries with relatively high exit costs and relatively strong antitrust enforcement. Here, again depending on the degree of asymmetry between firms, the following outcomes can arise. When there are high asymmetries ( $K < |\pi_m \beta(1-p\alpha)|$ ), outcome (3) will arise in equilibrium. When asymmetries are low (i.e.  $K > |\pi_m(1-p\alpha)|$ ), outcome (2) arises in equilibrium. For an intermediate level of asymmetry both outcome (1) and outcome (4) can arise in equilibrium. So in addition to the possibility of retaliation, there is a small danger of collusion, when exit costs for small firm are too high. This result is quite intuitive, since with high exit costs the threat of a possible retaliation can force small firm to keep the cartel secret and not to apply for leniency. The above analysis can be summarized in the following proposition.

**Proposition 1:** *When traditional antitrust enforcement is strong (i.e.  $pa > 1$ ), after introduction of leniency programs there exists a threat of retaliation and of even stronger collusion in the industries with an intermediate level of asymmetry (i.e.  $|\pi_m \beta(1-p\alpha)| < K < |\pi_m(1-p\alpha)|$ ).*

<sup>90</sup> **Proof:** since  $\delta^{**} < 0$  and  $\delta^* < 0$ , any  $\delta$  in the interval  $(0, 1)$  is higher than  $\delta^{**}$ , hence (3) is played in equilibrium.

<sup>91</sup> **Proof:** since  $\delta^{**} > 1$  and  $\delta^* < 0$ , any  $\delta < \delta^{**}$  and any  $\delta > \delta^*$ . This in turn implies that outcome (4) is preferred over (3) by Big in case Small chooses not to self-report and outcome (1) is preferred over (2) by Big in case Small chooses to reveal information. Next, since any  $\delta < \delta^{****}$ , when  $\delta^{****} > 1$ , taking into account best response of Big, Small will prefer outcome (1) over (4). Hence, (1) is played in equilibrium.

<sup>92</sup> **Proof:** since  $\delta^{**} > 1$  and  $\delta^* > 1$ , any  $\delta < \delta^{**}$  and any  $\delta < \delta^*$ . This in turn implies that outcome (4) is preferred over (3) by Big in case Small chooses not to self-report and outcome (2) is preferred over (1) by Big in case Small chooses to reveal information about the cartel. Next, since any  $\delta < \delta^{****}$ , when  $\delta^{****} = pa > 1$ , taking into account the best response of Big, the smaller firm will prefer outcome (2) over (4). Hence, (2) is played in equilibrium.

To summarize the above discussion it needs to be stressed that, even when penalties are high enough to block the cartel formation in static settings, (i.e.  $p\alpha > 1$ ) there could be adverse effects of leniency programs on the incentives to the firms to collude in a dynamic setting. There could be a threat of retaliation and of enhanced collusion in the industries with an intermediate level of asymmetries (i.e. in industries where  $|\pi_m\beta(1-p\alpha)| < K < |\pi_m(1-p\alpha)|$ ). This implies that, in this kind of industries, a strong emphasis on the protection of leniency applicants needs to be introduced and particular attention should be paid to industries where exit costs are high.

Next, we continue our analysis with the discussion of the case where  $p\alpha < 1$ . In this setting again two sub-cases, depending on the size of  $\delta^{****}(K, S, p, \alpha)$  can arise. When  $p\alpha < 1$  and  $(1-\beta)\pi_m < S$ ) we obtain that  $\delta^{****} > 1$ , and, hence, the distribution of outcomes can be described as is done in the left panel of Figure 3. When  $p\alpha < 1$  and  $(1-\beta)\pi_m > S$ ) we obtain that  $\delta^{****} < 1$ , and, hence, the distribution of outcomes is given in the right panel of Figure 3.<sup>93</sup>

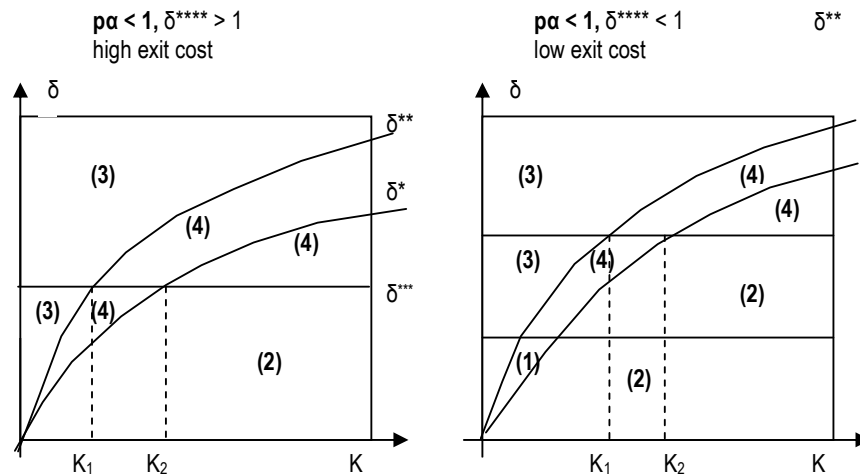


Figure 3. Equilibrium outcomes when  $p\alpha < 1$ .

In the case where  $p\alpha < 1$  the following regularities are satisfied for any parameter values:  $\delta^{****} = p\alpha < 1$ ,  $\delta^* > 0$ ,  $\delta^{**} > 0$ ,  $\delta^{**} > \delta^*$ . Note also that when  $p\alpha < 1$ ,  $(1-\beta)\pi_m > S$ , and  $\delta^{****} < 1$ , we have that  $\delta^{****} - \delta^{***} = S(p\alpha - 1) / ((1-\beta)\pi_m - S) < 0$ . Hence, inequality  $\delta^{****} < \delta^{***}$  holds for any parameter values. This is also depicted in the right panel of Figure 3. Moreover, in both cases described above (namely,  $\delta^{****} < 1$  and  $\delta^{****} > 1$ , with  $p\alpha < 1$ ), we have that  $\delta^{**} - \delta^{****} > 0$ , when  $K > p\alpha\pi_m(1-\beta)$  (denoted by  $K_1$  in Figure 3) and  $\delta^* - \delta^{****} > 0$ , when  $K > p\alpha\pi_m$  (denoted by  $K_2$  in Figure 3).<sup>94</sup>

Finally, based on the above analysis, we conclude that the following proposition holds. It relates four industry types to an environment with rather weak law enforcement (the product of the rate of capture and the coefficient of proportional fine is smaller than one). This situation applies to most European countries and to EU antitrust law as well. The US anti-cartel enforcement seems to be stricter. It only grants a single fine reduction to the first reporter. In addition it generally uses criminal law and includes possibility of imprisonment also for violations of antitrust law. For a more detailed insight into these matters see the appendix. Having established that the environmental condition of the above proposition holds we can move on to the industry characteristics and complement these statements with policy implications.

**Proposition 2:** *When traditional antitrust enforcement is weak (i.e.  $p\alpha < 1$ ), the following four effects of impact of industry structure on incentives to self-report can be found:*

**Proposition 2(1):** *In industries with little asymmetry ( $K$  is high) and low discount rate the first best outcome with self-reporting and competitive pricing afterwards (equilibrium (2)) can be achieved.*

<sup>93</sup> We do not provide detailed proofs of these results here, since they follow the same lines as in case where  $p\alpha > 1$ . Detailed proofs and clarifications are available from authors upon request.

<sup>94</sup> Proofs of these simple regularities are available from authors upon request.

This result shows that a leniency program can have the desired effect, i.e. destabilizing cartel and inducing more competition. However, the aim of the leniency program to establish competition in cartelized industries is only achieved when firms within an industry are more or less of the same size and little emphasis is put on future profits. This result is counterintuitive to the reasoning where symmetry creates a common focal price and facilitates collusion. Clearly asymmetry can be an important aspect for the stability of cartels. By far the largest number of industries can be described along the above lines. The firms within these industries have a strong incentive to start competing healthily, deviate from collusion or report to the AA upon introduction of the program. A new question that can be raised is whether collusion in these types of industries is as harmful to society as collusion is in industries characterized by a higher discount rate.

**Proposition 2(2):** *In industries characterized by a high discount rate ( $\delta > \delta^{**}$ ) retaliation is always the most attractive strategy for any type of firm (regardless of asymmetry). Outcome (3) arises in equilibrium.*

Industries that do put an emphasis on future profits will have a reason to do so. Mostly these industries are comprised of rather larger firms than the previously mentioned type of industry.<sup>95</sup> Natural monopolists are willing to make great investments in networks since they know they will be able to profit from these for many years. The introduction of a leniency program has no effect what so ever in this type of industry since there is usually only a single firm active in the market. In other words there is no collusion in the first place. The prevention and regulation of monopolists lies beyond the scope of this paper.

**Proposition 2(3):** *In industries with low exit costs ( $S < (1-\beta)\pi_m$ ), high asymmetry ( $K < p\alpha\pi_m(1-\beta)$ ), and a low discount rate there is a threat of retaliation on the self-reporting firm. Outcome (1) with self-reporting and retaliation arises in equilibrium.*

The event of fierce competition isn't always an illustration of "healthy" competition (where healthy refers to pricing at or slightly above marginal cost). A war on prices or a period of repetitive advertising on prices might be an indication of dumping or predatory pricing. These practices reduce welfare, since investments are wasted through the destruction of capital. Moreover the consequence of this aggressive behaviour is the exit of firms from the industry and healthy competition is further away than it ever was. Proposition 2(3) shows that the introduction of a leniency program in some cases may lead to aggressive pricing (retaliation can occur as a response to self-reporting) and a loss of welfare. Besides the welfare loss the antitrust authority will also have to spend resources in the future to regulate the newly created monopolists. This scenario is especially likely to occur in asymmetrical industries with a relatively low sunk cost and a greater emphasis on future profits. When sunk costs are low, small firm will not lose much when predated upon, and will therefore be more inclined to report. On the other hand, due to the low discount factor for the big firm the potential fine of collusion will outweigh the future gain from cartel. This will cause retaliation. It is the promise to protect any party to self-report to the antitrust authority that helps to overcome this scenario. This promise needs to be clear and credible, though resources need not be wasted. Therefore a promise to protect should be incorporated in the leniency guidelines.

**Proposition 2(4):** *In industries with high exit costs ( $S > (1-\beta)\pi_m$ ) and discount factor in the range  $\delta^* < \delta < \delta^{**}$  collusive equilibrium (4) is sustainable even after leniency programs are introduced. Moreover, in industries with high exit costs (4) is sustainable for a bigger range of discount factors compared to low exit cost industries.*

The worst effect the introduction of a leniency program can have is the strengthening the stability of cartels. The results of the analysis show this does occur, however. In industries characterized by high sunk costs this scenario is more likely to occur. The reason the stability of the cartel is increased lies in the possibility for large firms to use the leniency program as a means to increase the trust they put in

---

<sup>95</sup> Note also that discount rates do, in fact, change with time for all industries. Thus, according to the model, a leniency policy that has one effect when the interest rate in the economy is low might have a different effect when the interest rate is high. Another interpretation is when higher discount rate is associated with higher industry stability. The discount rate ( $\delta$ ) can be interpreted as probability that the firm will survive in the next period. So that higher  $\delta$  implies higher industry stability. In this sense, industries that are comprised of larger firms, which are stronger and have higher probability of survival, would be more stable and, as a result, can be viewed as industries characterized by the higher discount rate.

other firms not to report to the AA. The larger firm “trusts” the smaller not to report to the AA. It can do so since the smaller firm knows the punishment of the larger firm is severe. This scenario can never be prevented completely. The chance, this scenario develops however, can be lowered by developing a comparable promise to protect the reporting firm as is described in the previous paragraph. Besides this type of policy approach the leniency programme should always go accompanied by an effort of traditional antitrust law enforcement directly aimed at the industries described above. When the threat of a fine due to the more traditional prosecution increases, more firms will choose to abandon the cartel.

#### **4. Conclusions**

The analysis of this paper adds to the current economic literature on leniency programs for cartels and illegal price-fixing activities, but the reasoning can also be applied to corporate whistle-blowing programs and witness protection programs. It reveals a number of adverse effects of the introduction of leniency programs in view of industry asymmetry. The main conclusion is that the introduction of a leniency program, regardless of the size of the fine, might facilitate the stability of cartels in certain industries. This is mainly due to a leniency program’s inability to remove the threat of punishment on a self-reporter (or whistle-blower) by its former partners. After a firm is convicted it remains with sufficient resources to retaliate on the reporting party. It enables some firms to use coercion as a means to increase trust in the cartel. Though the removal of trust is the aim of the program, the introduction of the scheme actually provides colluding firms with the means to stabilize the cartel.

Increasing the size of the fine and limiting the number of fine reductions to the first party to report isn’t sufficient to (fully) overcome the adverse effect of the introduction of the leniency program. The analysis in this paper implies that the program’s effectiveness largely depends on the environment and the type of industry to which it is being applied. Raising the rate of capture (through e.g. limiting the number of fine reductions) and the size of the penalty do help to diminish the adverse effect. The size of the fine can for instance be increased by putting a greater emphasis on aggravating circumstances, such as coercion. It will however not be sufficient to tackle cartels in industries with an intermediate level of asymmetry. When an AA is unable to raise sufficient resources to increase the rate of capture through traditional law enforcement, in this type of industry it should direct its focus on the promise to protect self-reporters from retaliation by former collusive partners. Since the current type of policy approach is sufficiently effective in a great number of industries, diversification of the program can give rise to a more efficient use of resources. Customization of the program, where it comes to protection, size and number of the fine reduction, paralleled by a traditional effort of law enforcement aimed at industries in which the adverse effect is likely to occur will help to make the program more effective.

The analysis of Section 3 implies that, even when penalties are sufficiently high to block the formation of cartels, the leniency program can still withhold firms from self reporting. Because of the threat of retaliation after a cartel is uncovered, trust between cartel members will be stronger in those industries characterized by an intermediate degree of asymmetry and barriers to entry, due to the introduction of the scheme. This implies that in these kinds of industries a strong self-reporter’s protection program should be introduced besides the leniency program.

When penalties are lower (i.e. the product of the rate of capture and the coefficient of the proportional fine is lower than one), which is currently generally the case in most European countries, the effectiveness of leniency programs largely depends on the environment the firms find themselves in and on the type of the industry. In this case the focus of the competition authority should be on those industries characterized by a low to intermediate degree of asymmetry and an intermediate to high discount rate. Since, in these types of industries, regardless of any barriers to entry, chances are that, the introduction of a mild leniency program facilitates collusion. It serves to strengthen trust between colluders, rather than to create a breach of trust.

Another effect of the introduction of a leniency program is reversion to aggressive pricing in order to retaliate on the reporting partner. Though this might at first look like healthy competition it eventually reduces welfare. In an environment of high fines (product is greater than one) this is a more likely scenario and it will occur in industries characterized by an intermediate level of asymmetry and low barriers to entry. However also in an environment of low fines aggressive pricing can be the effect of the introduction of the program, especially when barriers to entry are low. Besides having to spend resources on regulating these new (semi) monopolists, the destruction of capital associated with the retaliation strategy is detrimental to welfare.

To summarize the above analysis, in industries characterized by high barriers to entry/exit (such as high exit costs) and degree of asymmetry leniency programs may be ineffective and give rise to increased cartel strength. In industries with low exit costs leniency programs may be more effective, but retaliation is more likely to occur as a response to self-reporting. Policies aimed at the removal of this threat of punishment through aggressive behaviour need to be considered in order to remove these kinds of hard core cartels. A first means is to employ higher fines in order to remove a bigger part of the illegal gains. Putting more emphasis on aggravating circumstances, such as coercion, in the fining guidelines can also be an effective approach. Another regulatory measure is to introduce the promise to “protect” the reporting party after reporting in the leniency application. In general though a leniency program alone can not be fully effective in its aim to prevent and prosecute all cartels. A certain amount of effort will always need to be directed towards certain industries beside the leniency program.

The currently employed model calls for the decision to collude to be made prior to the decision to introduce a leniency policy. This, to some extent, limits the applicability of the model only to analysis of pre-existing cartels. The natural extension would be to enrich this model by employing a sequential repeated game framework where in each period in the presence of leniency programs the following two stage game is played. In the first stage firms decide whether to form cartel or not, the second stage is the revelation decision with the subsequent retaliation decision. Similar structure is implemented in Houbart *et al.* (2009) with symmetric firms and in the absence of retaliation. However, incorporating asymmetric firms and possibility of retaliation in such an advanced framework may make the model intractable and difficult to solve. In this manuscript we forego excessive technical difficulties and adopt a simplified version of the model in order to concentrate on how possibility of retaliation can influence the effectiveness of leniency programs for already formed cartels.

## References

- [1] Becker, G. 1968. Crime and Punishment: an Economic Approach, *Journal of Political Economy*, 76:169-217.
- [2] Buccirosi, P., and Spagnolo, G. 2006. Leniency Programs and Illegal Transactions, *Journal of Public Economics*, 90(6-7):1281-1297.
- [3] Chamberlin, E. 1929. Duopoly: Value Where Sellers Are Few, *Quarterly Journal of Economics* 3:63-100.
- [4] Chen, J., and Harrington, J. 2007. The impact of the corporate leniency program on cartel formation and the cartel price path, in *The Political Economy of Antitrust*, ed. by V. Ghosal, and J. Stennek. Elsevier.
- [5] Chen, Z., and Rey, P. 2007. *On the Design of Leniency Programs*, IDEI Working Papers 452, Institut d'conomie Industrielle (IDEI), Toulouse.
- [6] D.O.J. 1993. *US Corporate Leniency Policy*, <http://www.usdoj.gov/>
- [7] D.O.J. 2003. *US Sentencing Guidelines* (Chapter 8: Sentencing of Organizations), <http://www.uscc.gov/2003guid/CHAP8.htm>.
- [8] EC 2002. *Commission Adopts New Leniency Policy for Companies which Give Information on Cartels*, Brussels, <http://europa.eu.int/comm/competition/antitrust/leniency/>.
- [9] EC. 2006. Guidelines on the method of setting fines imposed pursuant to Article 23(2)(a) of Regulation No 1/2003, *Official Journal of the European Union*, (2006/C 210/02), Brussels.
- [10] Ellis, C., and Wilson, W. 2001. *Cartels, Price-Fixing, and Corporate Leniency Policy: What Doesn't Kill Us Makes Us Stronger*, mimeo, University of Oregon, Oregon.
- [11] Feess, E., and Walzl, M. 2004. Self-reporting in Optimal Law Enforcement when there are Criminal Teams, *Economica* 71:333-348.
- [12] Harrington, J. 2008. Optimal Corporate Leniency Programs, *The Journal of Industrial Economics*, LVI(2):215-246.

- [13] Hinloopen, J. 2003. An Economic Analysis of Leniency Programs in Antitrust Law, *De Economist*, 151:415-432.
- [14] Houba, H., Motchenkova, E., and Wen, Q. 2009. *The Effects of Leniency on Maximal Cartel Pricing*, Tinbergen Institute Discussion Paper 09-081/1.
- [15] Kovacic, W. 2006. Bounties as inducements to identify cartels, in *European Competition Law Annual 2006*, ed. by C. Ehlermann, and I. Atanasiu. Hart Publishing, Oxford/Portland, Oregon.
- [16] Motchenkova, E. 2004. *Effects of Leniency Programs on Cartel Stability*, CentER Discussion Papers Series 2004-98, Tilburg University, Tilburg.
- [17] Motchenkova, E., and Kort, P. 2006. Analysis of Current Penalty Schemes for Violations of Antitrust Laws, *Journal of Optimization Theory and Applications*, 128(2).
- [18] Motta, M., and Polo, M. 2003. Leniency Programs and Cartel Prosecution, *International Journal of Industrial Organization* 21:347-379.
- [19] OECD. 2002. *Fighting Hard Core Cartels: Harm, Effective Sanctions and Leniency Programmes*, Paris, <http://www.oecd.org/dataoecd/41/44/1841891.pdf>.
- [20] Rey, P. 2003. Towards a Theory of Competition Policy, in *Advances in Economics and Econometrics: Theory and Applications*, ed. by M. Dewatripont, L. Hansen, and S. Turnovsky. Cambridge University Press.
- [21] Spagnolo, G. 2004. *Optimal Leniency Programs*, Centre for Economic Policy Research, Discussion paper series, 4840, <http://www.cepr.org/pubs/new-dps/showdp.asp?dpno=4840>, (revised 2008).
- [22] Spagnolo, G. 2008. Leniency and Whistleblowers in Antitrust, in *Handbook of Antitrust Economics*, ed. by P. Buccirossi. MIT Press.
- [23] Stigler, G. 1964. A Theory of Oligopoly, *The Journal of Political Economy* 72 (1):44-61.
- [24] Tirole, J. 1988. *The Theory of Industrial Organization*. MIT Press, Cambridge.

## APPENDIX

### Historical Overview and Structure of LPs

**Table 1.** Fining Systems and Structure of Leniency Programs

Country	Size of Fine	Limitation of Fine	No of Fine Reductions	Max Fine Reduction first reporter	Max Fine Reduction other reporters
EU	Base level of fine is determined by the gravity and duration	10% of the total annual turnover of the year before conviction	>1	100%	Up to 75%
US	Base level of fine is determined by the gravity, illegal gains and damage to society	No upper bound	1	100%	No fine reduction
UK	Seriousness and relevant turnover form a basis	10% of total UK turnover of the year before conviction	>1	100%	50%

**Sources:** DoJ (2003), EC (2006), EC (2002), and OECD (2002).

**Table 2.** Timing of Introduction of Competition Law and Leniency Programs

Country	General Competition Law		Leniency Program	
	First Initiative	Organization in charge	LP introduction	LP latest version
EU	1959	Commission	1996	2002
US	1890	DoJ	1978	1997
UK	1998	Office of Fair Trade	1998	2004
Netherlands	1998	NMa	2002	2006
Germany	1958	Bundeskartellamt	2002	2006

**Sources:** DoJ (2003), EC (2006), EC (2002), and OECD (2002).

## INDO-CHINA TRADE: TRENDS, COMPOSITION AND FUTURE

Rajesh K. PILLANIA  
Management Development Institute, INDIA  
[r\\_pillania@yahoo.com](mailto:r_pillania@yahoo.com)

### Abstract

*India and China are among the oldest civilizations of the world with long history of interaction and relationship. They are the fastest growing economies among the major economies of the world. Both have made rapid progress after liberalisation. One of the major events in international trade and economics is the recent fast growth in the bilateral trade. India has emerged as one of the top ten trading partners of China whereas China has emerged one of the top three trading partners of India. Due to large size of economies and composition of economies and exports as well as high growth rates and political will from both sides, the bilateral trade will grow further and would have significant impact on global trade and economy.*

**Key words:** India, China, Indo-China trade, trade composition, emerging economies

**JEL Classification:** F14, O53, O57, R11

### 1. Introduction

China and India are the two fastest re-emerging economies (The Economist 2006). Until the last 19<sup>th</sup> century, India and China were the world's two biggest economies but they were outsmarted by the steam engine invention and subsequent industrialisation in Europe. Now the circle seems to be complete as we look at the fast growth of India and China in recent years and various projections for future (Goldman Sachs 2003). China's and India's economic growth rates have outperformed world average growth rates for the most part of last fifteen years. China has grown at an average rate of close to ten percent annually during 1990-2007 whereas India has grown at an average rate of six percent annually as compared to world economic growth rate of three percent (Kowalski 2007). Their rapid growth has had a significant impact on world economy (Srinivasan 2006). The story has moved beyond GDP growth rates... the emergence of China and India gives the world an opportunity to experiment and make sure that there is equitable wealth generation (WEF 2006). Together they constitute a big part of global population and geographic area. The two countries now account for 37.5 percent of world population and 6.4 percent of the world output and income at current prices and exchange rates (World Bank 2007).

Chinese phenomenal growth started in late 1970s whereas Indian growth picked up in the early 1990s. The rapid growth started in both countries after the liberalization process. Liberalization has brought foreign competition in various industries and encouraged the domestic companies to become competitive. With liberalization, both countries have also restructured their trade destinations over the years. In general they have used developed countries' market for their export destinations whereas their dependence has gone up for imports from other developing countries (Mohanty, and Chaturvedi 2005). Whereas China is repeatedly mentioned as the 'World Factory' highlighting the contribution of its manufacturing sector, India is increasingly named as 'World Back office' recognizing increasing outsourcing of global services to India. Various studies have projected that these economies will continue to grow and become economic superpowers in next few decades (Goldman Sachs 2007).

India and China are among the oldest civilizations with long history of association. India and China relationships have a long history. Nathu La Pass was the famous silk route. Trading through Nathu La Pass accounted for eighty percent of total border trade volume between India and China in the early twentieth century (Anonamous 2006). Trade ties suffered after Sino-India War of 1962. Visits of Prime Ministers, starting with Indian Prime Minister Rajiv Gandhi in 1988 followed by Chinese Premier Li Peng's visit in 1991 started positive steps for trust building and trade promotion. In 1984, India and China signed a Trade agreement for providing Most Favoured Nation (MFN) treatment. Border trade resumed in July 1992 after a hiatus of more than thirty years, consulates reopened in Mumbai and Shanghai in December 1992 (Economywatch 2005). Trading through Nathu La Pass was suspended in 1962 after the Indo-China War 1962. This pass was reopened after forty-four years of its closure on July 6, 2006. In recent years, the two countries have put in efforts to improve relations and to foster economic ties (TOI 2006, Peopledaily 2008, Chinese Embassy 2008, Indian Embassy 2008). The bilateral trade between the two countries has shown a phenomenal growth and yet there is a lot of



untapped future potential. This research paper is an attempt to study the Indo-China trade with focus on recent past and explore potential for future. This paper consists of six sections including the introduction part. The second section takes a quick overview of the performance of Indian and Chinese economies. The third section studies the recent trends in Indo-China trade. Fourth section analyses the composition of Sino-India trade. Fifth section deals with future of Sino-India trade and the last section is conclusion.

## 2. Recent performance of indian and chinese economies

China has been performing quite well during the past few decades. Its GDP has crossed USD 2 trillion mark in 2005 and the GDP growth rate has been more than eight percent in the first years of 21<sup>st</sup> century. Chinese exports of mass production items have caused serious concern (Mohanty, and Chaturvedi 2005); and studies have predicted that this might contribute to recession in the US (Palley 2004). The performance of China on selected indicators and the changes are shown in Table 1 and Table 2 respectively.

**Table 1.** China: Selected Indicators (USD billions)

Indicators	Years	2000	2001	2002	2003	2004	2005
Exports of goods and services (% of GDP)		23.3	22.6	25.1	29.6	34.0	..
Foreign direct investment, net inflows (BoP, current US\$)		38	44	49	54	55	..
GDP (current US\$)		1,198	1,325	1,454	1,641	1,932	2,229
GDP growth (annual %)		8.4	8.3	9.1	10.0	10.1	9.9
Imports of goods and services (% of GDP)		20.9	20.5	22.6	27.4	31.4	..
Long-term debt (DOD, current US\$)		133	129	120	120	131	..
Merchandise trade (% of GDP)		39.6	38.5	42.7	51.9	59.8	63.8
Net barter terms of trade (2000 = 100)		100.0	100.9	100.5	97.3	91.8	..
Total debt service (% of exports of goods, services and income)		9.3	7.9	8.3	7.4	3.5	..

**Source:** World Development Indicators 2001, 02, 03, 04, 05, 06, 07.

A big portion of Chinese GDP comes from International Trade and thus the performance of the economy depends to a large extent on the world economy. A slump in world imports shows its effects on Chinese exports and its overall economy. The expansion of international trade has been the key feature of the country's rising prominence in the world economy with average annual growth rates of trade at three times the world rates. Already in 2005 China became the third largest trading nation after the United States and Germany and its contribution to the growth of world merchandise trade over the period 1996-2006 amounted to 20 percent (Kowalski 2007). Looking forward, it is estimated that the China will become the world's top exporter by the beginning of the next decade owing to attractiveness to FDI, a high domestic saving rate, improvements in productivity spurred by reduced internal and external barriers to trade, and a significant surplus of labour (OECD 2005). Overall, Chinese goods exports account for approximately 90% of its total exports, which is substantially higher than the world average at a little over 80%. Its services exports only account for a little less than 10% compared with a world average of 20% (Table 2). All this suggests that China's services exports are still relatively underdeveloped and its integration into the world economy was mainly driven by goods trade (Kowalski 2007).

**Table 2.** China: Selected Indicators (Percentage change over the years)

Indicators	Years					
	2000	2001	2002	2003	2004	2005
Exports of goods and services (% of GDP)	..	-3.11	11.21	17.60	14.87	..
Foreign direct investment, net inflows (BoP, current US\$)	14.87	15.21	11.45	8.51	2.68	..
GDP (current US\$)	2.68	10.54	9.74	12.87	17.72	15.38
GDP growth (annual %)	15.38	-1.19	9.64	9.89	1.00	-1.98
Imports of goods and services (% of GDP)	-1.98	-2.09	10.16	21.25	14.77	..
Long-term debt (DOD, current US\$)	14.77	-3.11	-6.28	-0.08	9.12	..
Merchandise trade (% of GDP)	9.12	-2.79	10.99	21.45	15.25	6.75
Net barter terms of trade (2000 = 100)	6.75	0.85	-0.34	-3.24	-5.66	..
Total debt service (% of exports of goods, services and income)	-5.66	-14.41	4.51	-10.99	-52.87	..

**Source:** World Development Indicators 2001, 02, 03, 04,05,06,07.

Indian economy responded positively to the liberalization process started in 1991 and the growth rates picked up. Table 3 and Table 4 show the performance of India in recent years. Interestingly, the growth in India's economy and trade has been led by services rather than manufacturing. Whereas India has a deficit in balance of trade, it runs a surplus of balance of payments, mainly due to rising services exports. In 1993-2003, India's services exports grew at a compounded annual growth rate (CAGR) of as much as 17.3 percent, which for once is even higher than China's growth rate of 15 percent. Among the major economies of the world, India has the fastest growing services sector exports (Prem Singh 2004).

**Table 3.** India: Selected Indicators (USD billions)

Indicators	Years					
	2000	2001	2002	2003	2004	2005
Exports of goods and services (% of GDP)	13.2	12.7	14.5	14.8	19.0	..
Foreign direct investment, net inflows (BoP, current US\$)	3.6	5.5	5.6	4.6	5.3	..
GDP (current US\$)	461.3	478.3	506.1	600.7	694.7	785.5
GDP growth (annual %)	4.0	5.3	3.6	8.3	8.5	8.5
Imports of goods and services (% of GDP)	14.1	13.6	15.5	16.1	21.0	..
Long-term debt (DOD, current US\$)	95.6	94.8	100.7	107.6	115.2	..
Merchandise trade (% of GDP)	20.4	19.6	20.9	21.4	24.9	28.2
Net barter terms of trade (2000 = 100)	100.0	97.8	88.6	80.6	75.9	..
Total debt service (% of exports of goods, services and income)	14.5	11.7	14.8	18.9	..	..

**Source:** World Development Indicators 2001, 02, 03, 04,05,06,07.

India's recent economic dynamism has led many to compare it with China and to expect a similar dramatic insertion in world markets. However, India's trade expansion is much less impressive. Its share in world trade has first declined steadily since the beginning of the 1970s to around 0.5% at the beginning of 1990s and then rose steadily to just above 1% currently. The compound annual growth rate of India's exports of goods and services for the 1990-2005 period was 14%—well above the world average growth of 6%. In particular, in the last five years Indian exports have increased at around 18-20% per annum—three times the rate of world trade growth. Yet, these significant increases reflect to a large extent a relatively low base; India's contribution to the growth of world merchandise trade over the period 1996-2006 amounted to a mere 2%, as compared to 20% in the case of China (Kowalski 2007). Despite India being relatively abundant in skilled labour and capital, its manufacturing trade is highly concentrated in low-technology goods and the share of high-technology manufactured goods in its total exports has barely changed since the mid-1990s and remains under 5%, as compared to 30% for China (Kowalski 2007).

**Table 4.** India: Selected Indicators (Percentage change over the years)

Indicators	Years					
	2000	2001	2002	2003	2004	2005
Exports of goods and services (% of GDP)		-3.42	13.88	1.79	28.28	..
Foreign direct investment, net inflows (BoP, current US\$)	3.6	5.5	5.6	4.6	5.3	..
GDP (current US\$)	461.3	478.3	506.1	600.7	694.7	785.5
GDP growth (annual %)	13.07	32.30	-31.23	128.11	2.99	0.03
Imports of goods and services (% of GDP)	0.03	-3.41	13.76	3.56	30.64	..
Long-term debt (DOD, current US\$)	95.6	94.8	100.7	107.6	115.2	..
Merchandise trade (% of GDP)	7.07	-3.70	6.62	2.22	16.49	13.31
Net barter terms of trade (2000 = 100)	13.31	-2.17	-9.41	-9.04	-5.89	..
Total debt service (% of exports of goods, services and income)	-5.89	-19.14	27.02	27.28	..	..

**Source:** World Development Indicators 2001, 02, 03, 04, 05, 06, 07.

### 3. Recent trends in Indo-China Trade

Indo-China trade has grown very fast during last few years as shown in Table 5. After studying Indo-China trade over the years few trends emerge. First, after adjusting for partner GDP (i.e. bilateral trade divided by the trading partner's GDP) India's trade with China is greater than that with Japan, the US or the entire world (Gupta 2008). Second, China is one of top three trading partner of India whereas India is one of top ten trading partners of China. Further, China's trade is increasing much faster with India out of its top ten trading partners and same is the case with India. Third, the trade balance is in China's favour consistently over the years, which may be a cause of concerns to policy makers in New Delhi.

**Table 5.** Indo-China Trade (USD million)

Year	India's Export To China	Rate of Growth	India's Imports From China	Rate of Growth	Total Trade	Growth of Total Trade	Trade Balance
1994-95	254.3		761.04		1015.34		-506.74
1995-96	333.2	31.03	813.19	6.85	1146.39	12.91	-479.99
1996-97	615.32	84.67	757.55	-6.84	1372.87	19.76	-142.23
1997-98	718.94	16.84	1120.7	47.94	1839.64	34.00	-401.76
1998-99	427.06	-40.60	1096.47	-2.16	1523.53	-17.18	-669.41
1999-00	539.41	26.31	1288.27	17.49	1827.68	19.96	-748.86
2000-01	830.03	53.88	1494.92	16.04	2324.95	27.21	-664.89
2001-02	955.19	15.08	2043.33	36.68	2998.52	28.97	-1088.14
2002-03	1980.61	107.35	2799.29	37.00	4779.9	59.41	-818.68
2003-04	2962.92	49.60	4063.96	45.18	7026.88	47.01	-1101.04
2004-05	5,615.88	90.04	7,097.98	75.12	12,713.86	81.41	-27,981.49
2005-06	6,759.10	20.36	10,868.05	53.11	17,627.15	38.65	-46,075.19
2006-07	8,287.48	22.61	17,447.01	60.53	25,734.49	45.99	-59,341.43

**Source:** Monthly Statistics of Foreign Trade of India – DGCI&S, Govt. of India and Commerce Ministry, Govt. of India.

#### 4. Composition of Indo-China trade

The composition of Indo-China trade is given in Table 6 and Table 7. Indian major exports to China mainly consist of primary goods as shown in Table 6. The top three Indian exports are iron ore; primary and semi-finished iron and steel; and, plastic and linoleum products.

**Table 6.** Major Indian Exports to China (USD million)

Name of the Commodities	1998-99	1999-2000	2000-01	2001-02	2002-03	2003-04	2004-05
Iron Ore	87.77	81.43	130.16	207.56	409.37	825.74	2084.43
Prim. & semi-fin.iron & steel	4.72	3.59	26.4	11.02	491.02	577.25	488.74
Plastic & linoleum products	6.28	25.36	103.11	141.95	177.92	306.44	438.9
Processed minerals	16.99	11.56	15.24	23.4	87.58	99.99	230.85
Inorg/org/agro chemicals	22.28	39.32	60.01	48.84	88.89	83.96	217.91
Other ores & minerals	39.93	65.47	73.15	113.81	154.3	128.88	192.48
Drugs, pharma.& fine chem.	42.68	46.01	58.56	80.37	93.02	102.53	106.33
Machinery & instruments	9.17	11.72	19.72	11.36	32.47	73.5	98.14
Residual chem. & allied Prod	6.42	8.5	11	16.75	21.45	33.89	76.77
Non-ferrous metals	0.1	0.08	17.99	2.85	21.2	43.53	65.42
Marine products	51.51	87.81	116	85.19	118.39	89.08	65.38
Cotton yarn fab madeups etc	40.73	56.05	71.54	75.06	64.04	74.11	64.92
Electronic goods	4.08	9.24	20.66	15.9	22.82	42.31	44.53
Castor oil	17.19	31.95	13.83	7.08	1.64	8.93	40.89
Oil meals	32.64	9.69	8.29	4.22	3.67	17.28	35.81
Finished leather	3.32	4.02	8.41	12.7	15.84	21.65	30.5
Manufactures of metals	3.79	1.41	10.57	11.5	32.31	27.9	28.95
Ferro alloys	2	0.54	0.15	4.54	3.93	6.2	28.44
Dyes, intermediaries etc	3.41	3.31	8.59	11.29	20.08	66.2	18.63
Gems & jewelry	1.27	0.31	0.17	0.66	2.08	9.73	18.45
All Commodities	427.06	539.41	830.03	955.19	1980.61	2962.92	5344.88

**Source:** Monthly Statistics of Foreign Trade of India – DGCI&S, Govt. of India and Trade and Balance of Payment Statistics – CMIE, July 2005

Chinese major exports to India mainly consist of manufactured goods as shown in Table 7. Top three Chinese exports are electronic goods ; coal, coke and lubricants; and, organic chemicals.

**Table 7.** Major Indian Imports from China (USD million)

Name of the Commodities	1998-99	1999-2000	2000-01	2001-02	2002-03	2003-04	2004-05
Electronic goods	161.46	178.22	244.56	385.19	812.16	1384.44	2069.19
Coal, coke & lubricants	108.2	146.04	261.34	263.43	175.61	221.5	780.19
Organic Chemicals	163.54	178.71	181.02	242.38	326.42	474.06	606.66
Non-electrical machinery	48.92	53.04	60.49	72.14	105.69	184.38	424.43
Electrical machinery	20.6	28.58	46.34	59.98	72.3	101.7	211.23
Medical & Pharm products	71.27	73.05	70.32	105.26	150.25	185.26	192.8
Other tex yarn, fab, madeups		20.5	34.75	46.89	75.97	107.66	172.95
Silk yarn & fabrics		12.81	17.27	32.2	52.23	105.2	156.52
Non-ferrous metals	52.27	49.01	41.89	54.53	46.23	86.82	145.13
Silver		20.16	8.3	94.41	121.21	58.42	138.62
Iron & steel	20.44	30.75	7.53	14.52	9.24	31.65	136.21
Inorganic chemicals	62.72	53.2	48.37	59.66	69.25	104.46	131.22
Silk raw	48.31	87.44	95.66	122.9	106.96	113.7	123.34
Non-metallic mineral mnfs.	7.42	10.24	14.65	30.24	44.42	64.58	121.12
Manmade filament/spun yarn/waste		12.38	9.84	17.32	56.05	82.61	115.78
Metaliferrous ores & metal scrap	8.5	11.26	19.65	14.52	28.46	63.96	102.99
Professional inst, optical goods etc.	18.85	23.37	23.38	55.31	80.49	90.01	99.32
Transport equipment	3.49	8.45	6.62	9.06	22.15	13.16	88.5
Manufactures of metals	12.04	15.38	17.77	31.14	32.69	55.34	88.07
All commodities	1096.47	1288.27	1494.92	2043.33	2799.29	4063.96	6768.92

**Source:** Monthly Statistics of Foreign Trade of India - DGCI&S, Govt. of India, and Trade and Balance of Payment Statistics – CMIE, July 2005

The trade composition as discussed above and shown in Table 6 and 7 supports the presence of strong complementarities between the two economies having similar factor endowments but well differentiated economic structures (Boillot and Labbouz, 2006)

### 5. Future of Indo-China trade

Looking at the economic size and growth rates there is huge scope for further growth in Indo-China trade. The scatter plot of the data reveals that there is an exponential growth of trade in the given time series. Moreover the coefficient of determination for a linear regression equation comes out to be low ( $R^2 = .680$  and adjusted  $R^2 = .654$ ) strongly suggesting a non-linear trend in the time-series. Performing a non-linear regression analysis upon the given data and plotting it shows that there is a better fit of the estimated curve. In addition the  $R^2$  and adjusted  $R^2$  values, both, are considerably high suggesting a good fit of the model. ANOVA statistics also substantiate the appropriateness of the model considered. The trend lines of exports as well as imports also exhibit a similarity in distribution and are given as non-linear regression equation given as equation (2) and (3). Growth rates of exports and imports are highly spiked and import growth rates exhibit a weak linearity ( $R^2=.064$ ), but as the data is highly skewed (skewness = 2.18) normality of the growth rate of imports is doubtful.

Non-linear Regression Equations

$$Y_{\text{Tot\_trd}} = 7624.038681 - 3671.088981X + 392.699657X^2 \dots \dots \dots (1)$$

$$R^2 = .95301 \quad \text{Adjusted } R^2 = .94446 \quad \text{Standard Error } 2652.60028$$

$$F = 111.54171 \quad \text{Signif } F = .0000$$

$$Y_{\text{Exports}} = 1799.83 - 895.02 X + 108.264 X^2 \dots \dots \dots (2)$$

$$R^2 = .979 \quad F = 258.64 \quad \text{Signif } F = .0000$$

$$Y_{Imports} = 5824.21 - 2776.1 X + 284.436 X^2 \dots\dots\dots (3)$$

$R^2 = .925$                        $F = 67.78$                        $Signif F = .0000$

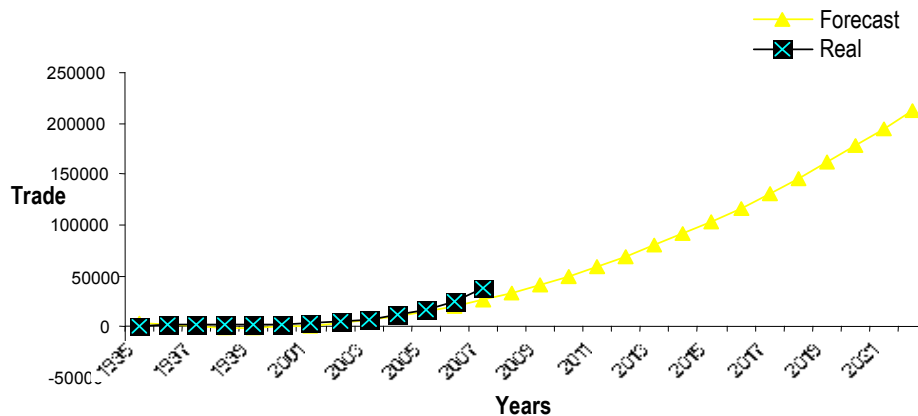


Figure 1. Indo-China Trade

The regression analysis reveals that the total trade of the country will increase in quadratic proportions, and in the years to come one can expect more and more bilateral trade to follow between China and India. The projection for the year 2008 onwards is given in Table 8 and Figure 1. Total trade between India and China can be expected to become 8 to 10 times of the level it was in the year 2007. Even if the growth rate in India-China trade slows down to 25% annually (a conservative projection) from the current rate of over 50%, bilateral trade between them will be almost \$75 billion in 2010 and \$225 billion in 2015, i.e., as large as China-US trade just three years ago (Gupta 2008).

Imports from China are larger than the exports and the growth based on the regression analysis reveals that it would remain higher in the coming years. Differentiating Equation 2 and 3 we get:  
 $dy/dx_{export} = -895.02 + 216.528 X$ ,                       $dy/dx_{import} = -2776.1 + 568.872 X$   
 we can observe that coefficient of X in  $dy/dx_{export}$  is less than  $dy/dx_{import}$ .

India can expect the deficit in balance of trade to accompany the growth in trade. The mean growth rates of exports is higher than that of imports (39.0638 and 34.0215) suggesting that if India can substantially increase the export growth rate the problem of deficit in balance of trade can be taken care of. The total trade can also be expected to fluctuate, as it can be seen in a number of occasions in the given data that there are minor fluctuations as in the year 2002-03. The analysis of the variables under consideration suggests that the growth of trade in the initial years have been low, but in the coming years this can rise sharply. As the larger portion of the total trade is comprises of imports from China, in the long-run this can prove to be detrimental to the balance of trade between the countries.

Table 8. Future India-China Trade (Estimated)

Year	Trade(USD million)
2008	33197.93
2009	40915.13
2010	49417.73
2011	58705.73
2012	68779.13
2013	79637.92
2014	91282.12
2015	103711.7
2016	116926.7
2017	130927.1
2018	145712.9
2019	161284.1
2020	177640.7
2021	194782.7
2022	212710.1

India and China are two fastest growing economies with huge exports growth rates and roughly one third of humanity. India and China are major developing countries and our strategic relationship has global significance (President Hu 2006). As their economies keep growing further and as the current bilateral trade is a small fraction of the trade potential, the trade between these two economies is one of the most important aspect of twenty-first century international trade and it will have significant impact on global trade and economy. It is more and more essential for developed countries to adapt to the emergence of these two Asian giants, and, to do so, the developed countries must target their demand more (Boillot, and Labbouz 2006).

## 6. Conclusion

India and China are the two fastest growing economies among the major economies of the world. Both countries have shown remarkable growth after liberalisation process. One of the recent and more important aspects is the bilateral trade between the two countries. The bilateral trade has grown remarkably in the last six-seven years. Today China is one of the top three trading partners of India whereas India has become one of the top ten trading partners of China. Looking at the size of the economy, composition of economy and exports and growth rates, there is a lot of untapped potential for bilateral trade growth. This bilateral trade will have significant impact on the global trade and economic scenario.

## References

- [1] Anonymous. 2006. *Historic Indo-China Trade Link Revived*, Indiabusinessweek website. <http://indiabusinessweek.com/News/Economy/21.html> (Accessed on 21<sup>st</sup> Feb 2008)
- [2] Boillot, J., and Labbouz, M. 2006. India – China Trade: Lessons Learned and projections for 2015, *Economic and Political weekly*, June 30-July 6:2893-2901.
- [3] Chinese Embassy. 2008. *China – India Trade Keeps on Increasing Fast*, New Delhi, Embassy website. <http://www.chinaembassy.org.in/eng/jjmy/t146698.htm> (Accessed on 23<sup>rd</sup> Feb 2008)
- [4] Chinese Embassy. 2008. *A Survey Before 2003*, New Delhi, Embassy website. <http://www.chinaembassy.org.in/eng/jjmy/zymyw1/jbxx/t61319.htm> (Accessed on 17<sup>th</sup> March 2008)
- [5] Economywatch. 2005. *India China Trade*, Economywatch website. [http://www.economywatch.com/world\\_economy/china/indo-china-trade-relations.html](http://www.economywatch.com/world_economy/china/indo-china-trade-relations.html) (Accessed on 23<sup>rd</sup> Feb 2008)
- [6] Goldman Sachs. 2003. *Dreaming with BRICs: The Path to 2050*, Goldman Sachs website. <http://www2.goldmansachs.com/ideas/brics/brics-dream.html> (Accessed on March 16, 2008)
- [7] Goldman Sachs. 2007. *BRICs and Beyond*, Goldman Sachs website. <http://www2.goldmansachs.com/ideas/brics/BRICs-and-Beyond.html> (Accessed on March 16, 2008)
- [8] Gupta, A. 2008. The future of India-China trade, *The Economic Times*, 14th Jan. [http://economictimes.indiatimes.com/View\\_Point/The\\_future\\_of\\_India-China\\_trade/articleshow/2697720.cms](http://economictimes.indiatimes.com/View_Point/The_future_of_India-China_trade/articleshow/2697720.cms) (Accessed on 17<sup>th</sup> March 2008)
- [9] Hu, J. 2006. India China to double trade to \$40 Bn by 2010, in *Times of India online*, Timesonline, 21Nov. <http://timesofindia.indiatimes.com/articleshow/msid-506140> (Accessed on 21<sup>st</sup> Feb 2008)
- [10] Indian Embassy. 2008. *Bilateral Relations*, Beijing, Embassy website. <http://www.indianembassy.org.cn/> (Accessed on 17<sup>th</sup> March 2008)
- [11] Kowalski, P. 2007. *China and India: A Comparison of Two Trade Integration Approaches*, Conference on India and China's Role in International Trade and Finance and Global Economic Governance, CRIER, New Delhi December 6-7. [http://www.icrier.org/pdf/6dec07/PresentationPrzemek%20Kowalski\\_Session4.pd](http://www.icrier.org/pdf/6dec07/PresentationPrzemek%20Kowalski_Session4.pd) (Accessed on 17<sup>th</sup> March 2008)
- [12] OECD. 2005. *Economic Survey of China*, OECD, Paris.



- [13] Peopledaily. 2008. Singh's China visit a landmark event in China-India relations, *Peopledaily website*. <http://english.people.com.cn/90001/90776/90883/6340429.html> (Accessed on 17<sup>th</sup> March 2008)
- [14] Preamsingh, M. 2004. Where India's creaming China, *Economic Times*, Aug.10.
- [15] Srinivasan, T.N. 2006. China, India and the World Economy, *Economic and political Weekly*, Aug 26-Sept 2:3716-3727.
- [16] TOI. 2006. India China to double trade to \$40 bn by 2010, *Timesonline*, 21Nov. <http://timesofindia.indiatimes.com/articleshow/msid-506140>(Accessed on 21<sup>st</sup> Feb 2008)
- [17] The Economist. 2006. *The new titans, A survey of the world economy*, September 16.
- [18] WEF. 2006. Annual Meeting, Davos, Switzerland 25-29 January. <http://www.weforum.org/pdf/summitreports/am2006/emergence.htm> (Accessed on 21<sup>st</sup> Feb 2008)
- [19] World Bank. 2007. *Dancing with Giants: China, India, and the Global Economy*, The World Bank, Washington, DC.
- [20] World Bank. 2001. *World Development Indicators, 2001*, The World Bank, Washington, DC.
- [21] World Bank. 2002. *World Development Indicators, 2002*, The World Bank, Washington, DC.
- [22] World Bank. 2003. *World Development Indicators, 2003*, The World Bank, Washington, DC.
- [23] World Bank. 2004. *World Development Indicators, 2004*, The World Bank, Washington, DC.
- [24] World Bank. 2005. *World Development Indicators, 2005*, The World Bank, Washington, DC.
- [25] World Bank. 2006. *World Development Indicators, 2006*, The World Bank, Washington, DC
- [26] World Bank. 2007. *World Development Indicators, 2007*, The World Bank, Washington, DC



# WAS THE WORLDWIDE ASYMMETRY IN CURRENT ACCOUNTS CAUSED BY THE MACROECONOMIC POLICY OF THE GLOBAL ECONOMY'S LEADER?

Georg QUAAS  
Universität Leipzig, GERMANY  
[quaas@uni-leipzig.de](mailto:quaas@uni-leipzig.de)

## Abstract

*Some authors sketch the causal chain that produced the current account surplus in China and the current account deficit of the US (as a part of global imbalances) as follows: declining interest rates in the US cause a redirection of capital flows into the periphery, rising capital inflows into China and other Asian countries trigger currency purchases by periphery central banks, and increasing stocks of foreign reserves on the asset side in the central bank balance sheet are matched by a proportional increase of reserve money on the liability side. To keep the exchange rate stable, foreign reserves are accumulated and reserve money expands. The Peoples Bank of China is trying to fight the inflation pressure with several measures, among them higher interest rates. This attracts even more foreign capital to China. Moreover, it cannot solve a problem that originates in the macroeconomic policy of the global economy's leader – a crucial point in this argument is the redirection thesis. The empirical evidence does not support this thesis in several respects – there is no evidence for a redirected capital flow away from the US toward China, and there is no evidence that interest rates controlled by the Federal Reserve are the cause of the capital flow to China.*

**Keywords:** global imbalances, current account surplus, current account deficit

**JEL Classification:** F32

## 1. Introduction

In the search of the causes of the worldwide recession that began in the US in 2007 and reached Germany a year later, the main focus was on the crisis in the financial system with its roots in an overheated real estate and housing market. Another more structural cause is seen in the global imbalances that evolved in the past decade and that are reflected by rising asymmetries of the current accounts of some global players. How is this phenomenon linked to the real estate and financial crisis in the US? Is there a link that is responsible for these asymmetries? Some authors argue that East Asian current account surpluses are responsible for US current account deficit (Dooley/Folkerts-Landau/Garber 2003, Bernanke 2005) because the East Asian countries keep their exchange rate artificially undervalued, for instance by pegging it to the dollar. Their goal is to promote export-led growth. This feeds over-consumption in the US. Other authors argue the other side and are more focused on fiscal and monetary policy. Rising deficits in large countries such as the US are producing rising surpluses in an increasing number of periphery countries, including China (Schnabl/Freitag 2009, 14). According to these authors, fiscal and monetary policies in the US are among the main factors causing the global asymmetries.<sup>1</sup>

## 2. The analytical framework

To avoid the over-simplifying terminology of a *center* and its *periphery*, every country that trades internationally is called a *global player*. A few have large economies with strong international trade and financial markets providing internationally used currencies in the world monetary system. By now and in the foreseeable future, the US dollar is the leading international currency. International trade and capital flows are denominated mostly in the US dollar. Backed by the large size of US goods and financial markets, the dollar is *the* dominant international medium of exchange, unit of account, and store of value in the Americas, Asia, the Middle East, and the Commonwealth of Independent States. There are different reasons why other countries and private international agents accept the dollar as common international money; they include low transaction costs, ubiquity, and higher stability compared to their own currency.

---

<sup>1</sup> With few exceptions, the first paragraphs summarise the author's shared views with Schnabl/Freitag (2009).

Since the euro was created, its international role has been steadily increasing not only in Europe, but also beyond. From this point of view and taking into account the dimension of its international trade, the *European Monetary Union* (EMU) can be regarded as another important actor in the global economy. To mark the difference between the leading country and the other global players we see the US as a global player exerting *hegemony* (Mansfield 1994, 272) and define the US as the *leader*. The rest of the global players are *challengers*. According to the literature, the EMU and its currency can be regarded as one of the challengers of the US's role in the world economy (Chinn, and Frankel 2005). In spite of the fact that China, a large and rapidly growing economy with strong international trade, is pegging its currency to the US dollar, it is a global player that also can be called a challenger because its monetary policy can be changed at any time in the future.

I do not raise the question here of whether there are or have been other challengers in the global economy such as Russia or Japan; for the special purpose of this analysis it is sufficient to focus on the leader and one or two of its challengers. To make the picture more clear cut in terms of a nation-state framework, the analysis treats the EMU as being represented by its economically leading country, Germany. The *European Central Bank* (ECB) acts independently from any member state of the EMU. Therefore in this analysis, the EMU is financially represented by the ECB and economically by Germany. This is a simplification, but it plays no crucial role in the discussion of my leading question: Can the global imbalances be explained by the macroeconomic policy of the US as it is done by Schnabl and Freitag?

### **3. The main actors**

#### **3.1. The United States**

The US is (as is the EMU) a large and—in the view of Schnabl and Freitag (2009, 8)—a comparatively closed economy. As a matter of fact, the trade openness (Suzuki/Krause 2005) in 2004 was 0.31 in the US, 0.65 in China, and 0.71 in Germany. Monetary decisions made by the Federal Reserve (Fed) are based first on domestic targets such as growth and financial and price stability, even though authorities may feel the burden of responsibility connected with having an international currency (see Chinn/Frankel 2005, 7,11-12). By and large, it may be legitimate to classify macroeconomic policies in the US of the past decade as being expansionary (Schnabl/Freitag 2009, 4), but this assessment ignores the period from 2004 to 2006 which is characterized by rising interest rates in reaction to the overheating of the domestic market. External targets of macroeconomic policy such as export competitiveness are regarded as secondary by the Fed and the exchange rate is being left to free floating. With two short exceptions among the past 18 years the two prosperous phases of the US economy are accompanied by a rising current account deficit. What was the cause of the deficit? The authors give a very vague answer to this question: “Low interest rates and buoyant domestic activity are likely to contribute to rising imports and increasing current account deficits” (Schnabl/Freitag 2009, 15-16.) Why did the flourishing economy of the US not lead to rising exports and increasing current account surpluses like in Germany? According to a widely shared point of view the answer is this—exports are a cause of a flourishing economy instead of an effect. But imports can be regarded this way. Therefore, the account deficit of the US economy is mostly homemade, caused by an overheated consumption, in spite of the fact that the staggering demand in the rapidly developing countries may be one of the causes for lagging exports. Another explanation refers to the increasing food and raw material prices that contributed to the rising current account surplus of countries with this kind of exports (Schnabl/Freitag 2009, 18-19), but only from 2008 onward and less valid for trading partners China and Europe.

#### **3.2. Europe**

The euro was introduced in 1999 and has gained importance as a regional international currency. It was backed by the substantial size of European goods and financial markets. The euro is used for payment transactions within the EMU and as a vehicle currency between EMU members and non-EMU countries. Some countries with institutional links to the EMU have redirected their exchange rate strategies to the euro (Lithuania, Bulgaria). Foreign reserves are increasingly held in euro-denominated assets (e.g., Russia). According to ECB (2008), private and public agents have increased the use of the euro for their international transactions. Like in the US, the macroeconomic policy of the EMU is designed to meet domestic targets such as price stability, growing output and financial stability, first, but

these are supra-national targets. External targets such as exchange rate stability and export competitiveness are regarded as secondary. Therefore, exchange rate is left as free floating. Macroeconomic policies in the euro area tended to be more restrictive than the Fed, because of the priority of price stability. The aggregated current account of the euro area has been balanced by and large. Some members of the EMU such as Greece and Spain have experienced increasing deficits. If we focus to Germany as the economic leader of the EMU, her current account has increased since 2001.

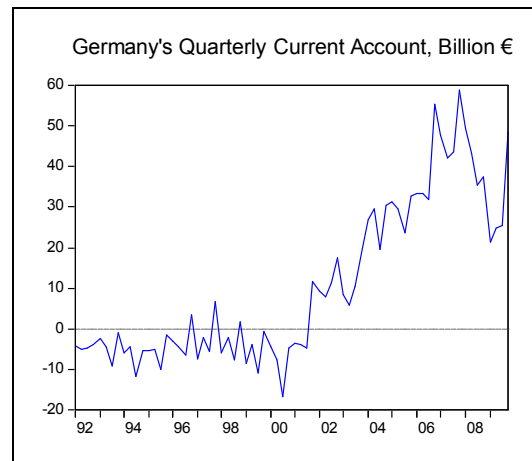


Figure 1. Germany's quarterly current account, in billion €.

### 3.3. China

China has a big and rapidly growing economy with emerging markets and it has tended to run account surpluses since 1996. There are several reasons why China pursues a soft peg of its currency to the dollar. China still has underdeveloped capital markets; pegging to the dollar provides stable conditions for China's export dependent industrial sector, it secures a loss-free recycling of the revenues generated in dollars abroad, and it is attractive to foreign investors. An appreciation of a currency is an appropriate measure in the case of an increasing current account surplus, but this would worsen the conditions of exports and erode the value of international assets in terms of the Renminbi ("the People's currency" with its unit "Yuan"). On the other hand, there is the danger of importing inflation of the dollar. Soft pegs allow for a restricted kind of exchange rate flexibility. A slow process of appreciation of the Yuan has been underway since 2004 (*see* Figure 2).

### 4. The Asymmetries

The exchange rate policies of the three global players, the US, EMU, and China, generates three relationships characterized in two types: the (almost) constant rate between China and the US as a consequence of China's pegging to the dollar is contrasted by a flexible rate between the euro and the dollar and between the euro and the Yuan. To summarize the asymmetries between the current accounts of the three big global players, the international position of the US economy is characterized by a rising deficit, while the EMU (Germany) and China show rising surpluses. This generates increasing pressure on the Yuan being appreciated compared to the dollar, while appreciation of the euro has been steady since its introduction. Looking at the macroeconomic policy behaviour, the US tended to be as expansionary as China, while the EMU had more restrictive fiscal and monetary policies.

#### 4.1. Monetary, fiscal and exchange rate policies

In the Mundell-Fleming framework, flexible exchange rates are dampening the effectiveness of an expansionary fiscal policy due to increases of the interest rate and the tendency to an appreciation of the currency. According to the AS-AD model, fiscal expansion aimed to stimulate a staggering economy is more effective when it is supported by a monetary expansion that keeps domestic interest rates low and softens appreciation pressure. Such a rare coordination of macroeconomic policies is one of the standard examples in macroeconomic textbooks (Blanchard/Ilting 2004, 155-156). It also can be supposed to be the case in China, because there is no clear institutional separation between the government and central

bank. The EMU countries are much more restricted in their scope to carry out discretionary fiscal policies because (i) the ECB is independent from the expectations of the member-states of the EMU and bound to pursue first and foremost price stability; and (ii) the legal limits for government deficits supposed by the Maastricht treaty.

#### 4.2. The redirection thesis

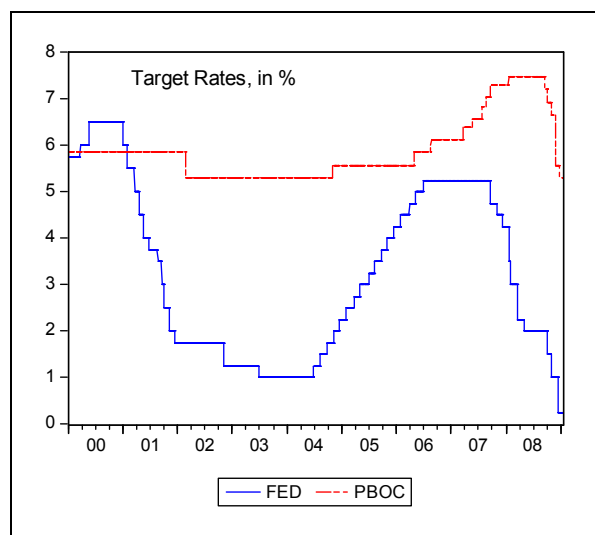
The consequence of dollar or euro pegs is mirrored by the asset side of the central bank's balance sheets. Foreign reserves are the most important item that builds the basis for reserve money creation. "Claims on government and on the private sector play only a marginal role for reserve money creation. From a long-term perspective, when output grows the necessary increase in reserve money is via the accumulation of foreign reserves" (Schnabl/Freitag 2009, 11). This is the background for the hypothesis by the authors in the scheme of a center and periphery of the global economy: "As a result interest rates in periphery countries are dependent on the monetary policy of the centers. If interest rates in the center decline, capital flows are redirected towards the peripheries, and the currency of the periphery country appreciates. To keep the exchange rate stable, foreign reserves are accumulated and reserve money expands" (Schnabl/Freitag 2009, 11).

#### 4.3. Claiming a causal link

The redirection of capital flows from the US to China which is supposed to be triggered by declining interest rates of the Fed is the central argument delivered by Schnabl and Freitag: "...the direction of causality matters. Are the complementary trends in global imbalances driven by the centers or the peripheries?" (Schnabl/Freitag 2009, 13). Contrary to the view of others who see a causality running from East Asia to the US, the authors hypothesise, "We assume a reverse causality: rising deficits of large centers... are assumed to produce rising surpluses in an increasing number of periphery countries" (Schnabl/Freitag 2009, 14.). They identify two types of transmission channels that explain how current account deficits (surpluses) in the center are transformed into surpluses (deficits) in the periphery. The first channel is a link between the macroeconomic policies of the center and the periphery mainly mediated by the exchange rate policies; the second channel consists of relative prices that influence the current accounts of exporting and importing countries. This paper is concerned with the first channel only.

#### 5. The Facts

There is no quarrel over the described asymmetries between the three global players or over the immense accumulation of foreign reserves in China. The question discussed here is this: Can the accumulation of dollars and foreign assets in China be seen as the result of a higher capital inflow caused by low interest rates in the US? A first answer can be found by a simple inspection of the data.



**Figure 2.** Target Rates of the Fed and the PBoC

The interest rates set by the Fed were low compared to those set by the Peoples Bank of China (PBoC) before the crisis broke out. On this background it sounds plausible that American credit conditions (besides those of other countries such as Japan, etc.) fostered capital flows to China. Figure 3 depicts foreign direct investment (FDI) as one indicator of the capital flows to China and the US.

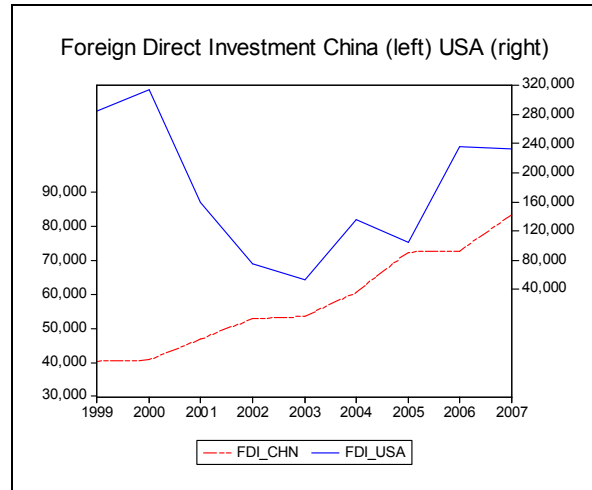


Figure 3. Foreign direct investment, China and USA, in millions \$.

As can be seen in Figure 3, there was indeed a rising capital flow to China, but this is also true for the US, at least since 2003. The thesis disputed here is that capital flows were redirected from the US to China because of the comparatively low interest rates in the US. In reality, the opposite was the case. In 2003 the downward trend of capital flows to the US, which was not caused by declining interest rates, but by the events of September 11, 2001, was inverted and this is the opposite direction of what is asserted. Moreover, the US-FDI was higher than China's all the time. Of course, this is no surprise. Everybody knows that capital flows to the US fuelled the finance and housing bubble that burst in 2007.

There was a steadily growing capital flow to China, too. Looking at Figure 3, growth rates of both flows seem to be inversely linked together. However, the correlation between the two variables of  $-0.16$  is statistically not different from zero. If we compare Figure 2 with Figure 3, there seems to be no sign of a direct causal connection between PBoC's interest rates (as a cause) and the changes of capital inflows to China (as the effect), not to mention between PBoC's interest rates and the capital flow to the US. Although the target rates of the PBoC were comparatively higher, they could not hinder the rising capital flow to the US. In addition, there is no visible influence exerted by the Fed on the capital flow to China. As can be seen by a comparison of the curves of Figures 2 and 3, the swelling capital flow to the US was *not* caused by larger interest rates set by the Fed; instead it was a *reaction* of the Fed to an overheated capital and housing market at home.

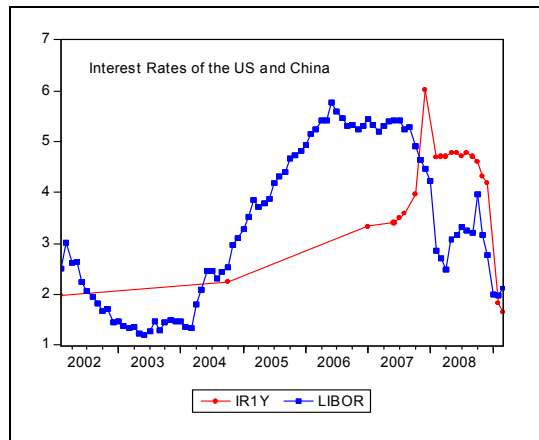
The number of observations is too small to carry out reliable statistical tests. In Table 1 the probabilities are reported of testing the hypothesis that the variable  $x$  does not Granger cause the variable  $y$ . The table may be read in the following way: if there is a higher probability in cell  $(y_i, x_j)$  than in cell  $(y_j, x_i)$ , it is more plausible that  $x_i$  causes  $y_j$  compared to the hypothesis that  $x_j$  causes  $y_i$ .

Table 1. Probabilities of the null hypothesis "x does not Granger cause y," number of observations in brackets

	x:	FDI_CHN	FDI_USA	PBoC	Fed
y					
FDI_CHN		--	0.92 (9)	0.77 (6)	0.86 (6)
FDI_USA		0.95 (9)	--	0.25 (6)	0.12 (6)
PBoC		0.20 (6)	0.22 (6)	--	0.51 (6)
Fed		0.53 (6)	0.75 (6)	0.43 (6)	--

According to these tests, it seems to be the case (i) that the foreign direct investment to China influences the target rate of PBoC rather than the other way round, and (ii) that interest rates set by the Fed influence FDI to the US rather than inverted. The differences of the other comparable cells are too small to make even a tentative assertion. As far as the result (ii) can be taken seriously, it is in line with Schnabl's and Freitag's argument—but the evidence is modest. Similar encouraging results may have convinced the authors, that there is a causal relationship. However, there is no evidence for the other part of the causal chain: Interest rates of the PBoC seem to have no influence on China's FDI.

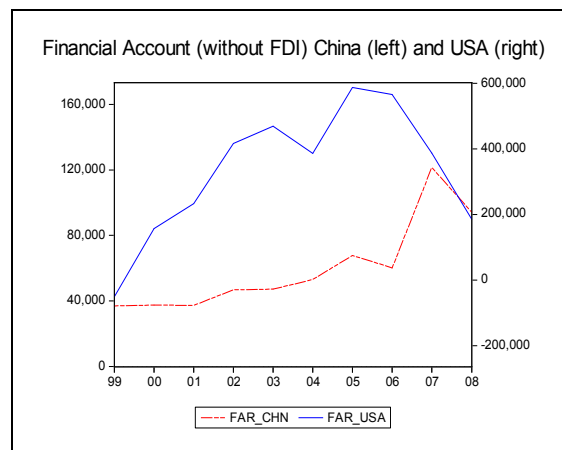
The reason for target rates having only a minor influence on capital flows is simple: target rates of central banks are never a direct cause of capital flows. Capital is attracted by sufficiently higher funds rates on capital markets. This missing link between central banks' monetary policy and the real movement of capital is not mentioned by the authors. It is discussed (among others) by Dooley/Folkerts-Landau/Garber (2003) and by McKinnon and Schnabl (2009).



**Figure 4.** LIBOR and Interest Rates of 1-Y-Deposits China (IR1Y)

McKinnon and Schnabl (2009, 1) assert a refusal of China's industrial corporations and financial institutions to invest abroad, because (i) they expect a further appreciation of the Renminbi and with it a loss of their foreign investments, and because (ii) the US federal funds rate was low. This was true before 2004 and after 2007 (see Figure 4), and the question is how the global imbalances that partially caused the crisis can be explained. Taking into account the interest rates of the US that were higher than those of China from 2004 to 2007, there was *no cause* of a redirection of the capital flow to China, because the US capital market was much more attractive for investors.

As Prof. Schnabl remarked in a personal note to the author, FDI is one of the indicators of a country's capital inflow and only a part of the financial (capital) account comprising other components like foreign-owned assets and other investments. In looking for empirical evidence of what Schnabl and Freitag could have meant, the financial accounts were purged from FDI (see Figure 5).



**Figure 5.** Financial Accounts (IMF definition) without FDI, in millions \$.

There are two periods in which the reduction of financial inflows to the US is connected with an enhanced inflow to China, 2003 and 2006. The downswing 2006 seems to be linked especially to an upswing of capital flows to China. The problem with the redirection thesis is this—the reduction of US\$178 billion may have fuelled the additional capital flow to China of US\$61 billion; but how can this explain the global imbalances that emerged years before?

Therefore, a similar conclusion can be drawn from Figure 5 depicting the financial accounts purged from the FDI in China and in the US and including the other indicators of capital flows. *There were no enhanced capital flows to China on the cost of the US which were caused by higher interest rates in China and which would have caused the global imbalances.* There was no redirection of the capital flows away from the US toward China during the years before the crises broke out. As an explanation of the causes of the crisis, the redirection thesis is useless.

Curious as it is, the authors know these facts, but do not draw the consequences. For instance, instead of a enhanced capital flow to China caused by the allegedly lower interest rates in the US, the periphery central banks, including PBoC, were expanding the “holdings of US and euro area government bonds,” (Schnabl/Freitag 2009, 15) quite the opposite direction of stipulated capital flow.

By the way, the attractiveness of the US capital market was given even in periods when interest rates were lower than those in other countries. Dooley, Folkerts-Landau, and Garber (2003) offer an explanation for this anomaly. Again” the challenger China plays a crucial role—China’s current account surplus is used by the officials to buy US securities without regard to their risk and return characteristics. “Their appetite for such investments is, for all practical purposes, unlimited because their growth capacity is far from its limit” (Dooley/Folkerts-Landau/Garber 2003, 6). In other words, export-led growth in China leads to current account surpluses and to *more* capital flows from China to the US, just the opposite of the disputed proposition. Of course, this is not to say that China is the only one responsible for the global imbalances. Beside China, official sectors of Japan and Taiwan plus private investors in Europe, Canada and Latin America helped finance the US current account deficit. And last but not least, there have been structural and political conditions in the US that fostered over-consumption for decades.

## 6. Once more: The disputed argument

According to the authors, changes in the monetary stance in the US are likely to lead, independent from the exchange rate regime, to lower rates in China (and other countries of the dollar periphery) for the following reasons: declining interest rates in the US are supposed to cause a redirection of capital flows into the periphery; rising capital inflows into China and other Asian countries trigger currency purchases by periphery central banks; increasing stocks of foreign reserves on the asset side of the central bank balance sheet are matched by a proportional increase of reserve money on the liability side (Schnabl/Freitag 2009, 14). What are the consequences of the latter situation related to interest rates whatever the reasons were that led to it?

From the point of view of modern macroeconomic theory, a rising money supply is associated with shrinking interest rates, as long as money demand is constant. This is in line with an overall view in textbooks (Blanchard 2006, 389) and with the authors’ argument that interest rates of the periphery converge toward the rates of the center. In the case of China a rising money supply matched a rising money demand in an emergent market (Dooley, Folkerts-Landau and Garber, 2003; McKinnon and Schnabl 2009, 3), and as a consequence the interest rates were lowered only slightly one time when the Fed’s interest rates were falling. From 2006 to the mid-2008 the PBoC was continuously reacting with adjustments to interest rates to fight inflation.

A look at the monetary policy of the PBoC raises similar questions. If the currency is or shall be tightly pegged to the dollar, there is no other way to react to a rising capital inflow than to *lower* interest rates set by the central bank, according to the authors (Schnabl and Freitag 2009, 15). This measure is thought to *dampen* the capital inflow. On the other side, in front of the danger of an overheating economy accompanied by inflation, the central bank should not *cut*, but *enhance* interest rates. This is exactly what the PBoC was doing.

If central banks do not react to the situation of enhancing capital inflow with an appreciation and try to keep the level of their interest rates, appreciation expectations reinforce capital inflows. To avoid excessive appreciation, interest rates have to change eventually. Under a flexible exchange rate regime, the domestic currency will be appreciated and exports will tend to decline. The central bank will likely

react with interest rate cuts—not to ease inflationary and appreciation pressure caused by capital inflows (Schnabl and Freitag 2009, 15)—but to ease the consequences of worse terms of trade.

Apparently, the core of the authors' argument is the magnetic effect of high interest rates. There is no doubt that there is such an effect, but it is mediated by the funds rates on a capital market. It may well be that lower interest rates attract less foreign investment; on the other hand, they spur the domestic economy and *this* attracts more foreign capital by means of higher funds rates. This was the case in China. At the same time, the capital inflow to the US was waning and waxing—no redirection to China can be observed. If this critique turns out to be correct, other claims made by Schnabl and Freitag which are consequences are questionable; for example, that fiscal consolidation in periphery countries can be seen as the outcome of low interest levels in the center countries (Schnabl and Freitag 2009, 12), that interest rates in the center are directly translated into interest rates changes of periphery countries (Schnabl and Freitag 2009, 14), the latter being true only when taken with a pinch of salt (Figure 2).

## 7. Conclusions

The focus of the former analysis was laid on Schnabl's and Freitag's claim to deliver an explanation for the global asymmetries of current account deficits and surpluses. The hypothesis of a redirected capital flow turned out to be wrong in several respects. After fixing indicators in an appropriate way, it was possible to identify some periods that confirm the thesis, but neither a plausible link to interest rates nor an identified tendency of redirection could be interpreted as one of the significant causes of the crisis in 2007. Of course, the author of this study does not claim to have carried out a proper causal analysis. According to his understanding, this would presuppose sufficient high correlations and a theoretical backed hypothesis referring to a common cause of the variables explored (Saris, and Stronkhorst 1984). Both conditions are not in sight. In spite of this, Schnabl and Freitag claim to have discovered one of the causes of the global imbalances that emerged in the past decade, but regression analysis is not the best method to test causal hypotheses. As long as no empirical evidence can be delivered, the redirection thesis must be regarded as mere metaphysics.

## References

- [1] Bernanke, Ben. 2005. The Global Saving Glut and the US Current Account Deficit, *Sandridge Lecture, Virginia Association of Economics*, Richmond Virginia.
- [2] Blanchard, Oliver. 2006. *Macroeconomics*, Upper Saddle River, NJ.
- [3] Blanchard, O., Illing, G. 2004. *Makroökonomie*, München.
- [4] Chinn, Menzie, Frankel, Jeffrey. 2005. *Will the Euro Eventually Surpass the Dollar as Leading International Reserve Currency?* NBER Working Paper 11510.
- [5] Dooley, M., Folkerts-Landau, D., Garber, P. 2003. *An Essay on the Revived Bretton Woods System*. NBER Working Paper 9971.
- [6] European Central Bank ECB. 2008. *Review of the International Role of the Euro*, Frankfurt a. M.
- [7] Mansfield, Edward D. 1994. *Power, Trade and War*, Princeton, New Jersey.
- [8] McKinnon, R., Schnabl, G. 2009. *China's financial conundrum and global imbalances*. BIS Working Papers No 277.
- [9] Saris, W.E., Stronkhorst, L.H. 1984. *Causal Modelling in Nonexperimental Research. An Introduction to the LISREL Approach*, Amsterdam.
- [10] Schnabl, G., Freitag, S. 2009. *An Asymmetry Matrix in Global Current Accounts*. Working Paper No. 76, University of Leipzig, Faculty of Economics and Business Administration.
- [11] Suzuki, S., Krause, V. 2005. Determinants of civil war in post-colonial Asia, 1950 – 1992, the role of trade openness and economic development, *International Relations of the Asia-Pacific*, Vol. 5.

## Data sources

- [1] Asian Development Bank: Key Indicators for Asia and the Pacific 2009



- [2] Bureau of Economic Analysis: International Economic Accounts
- [3] Leitzinsen\_Org
- [4] MoneyCafe
- [5] Mortgage Information Service
- [6] People's Bank of China: Statistics
- [7] RMB-Guide
- [8] Shibor\_Org
- [9] Statistisches Bundesamt: Volkswirtschaftliche Gesamtrechnungen, Fachserie 18, Reihe 1.2, Inlandsproduktberechnung Vierteljahresergebnisse. März 2010
- [10] WEO World Economic Outlook, April 2009
- [11] United Nations Conference on Trade and Development, April 2009

Journal  
of Applied Research in Finance



## Call for papers Journal of Applied Research in Finance

Published two times a year, the journal is the official publication of *The European Centre of Managerial and Business Studies*, academic organization devoted to the study and promotion of knowledge about financial economics. The journal has been established in year 2009 as a descendant to *Journal of Applied Economic Sciences* (JAES). Two issues are published per volume. All articles and communications are available online for free. Printed copies can be ordered at a cost. The editors maintain classic *double blind peer review* procedure aiming at high academic standards but at the same time emphasize dynamic referee process so that the journal tracks scientific progress in real time.

Currently, *Journal of Applied Research in Finance* is indexed in EBSCO, CEEOL and IndexCopernicus.

The *Journal of Applied Research in Finance* invites paper submissions on issues related but are not limited to:

- Monetary Economics,
- Money and Interest Rates,
- Monetary Policy, Central Banking, and the Supply of Money and Credit,
- Macroeconomic Aspects of Public Finance,
- International Finance,
- Macroeconomic aspects of Finance,
- General Financial Markets,
- Financial Institutions and Services,
- Corporate Finance and Governance,
- Taxation, Subsidies, and Revenue,
- Fiscal Policies and Behavior of Economic Agents,
- Public Finance,
- Behavioral Finance.

Submissions to *Journal of Applied Research in Finance* are welcome. The paper must be an original unpublished work written in English (consistent British or American), not under consideration by other journals.

Invited manuscripts will be due till October 15, 2010, and shall go through the usual, albeit somewhat expedited, refereeing process.

### Schedule

Deadline for Submission of Papers:

15<sup>th</sup> October 2010

Expected Publication Date:

November (e-version) – December (hard-copy) 2010

E-mail: [jarf\\_secretary@yahoo.com](mailto:jarf_secretary@yahoo.com)

Web: [www.jarf.reprograph.ro](http://www.jarf.reprograph.ro)

# AUCO Czech Economic Review



The 660-year-old Charles University in Prague is now undertaking a new publication venture by re-launching a journal *AUCO Czech Economic Review*.

*AUCO Czech Economic Review* (ISSN 1802-4696) presents original, rigorously peer-reviewed research in economics with solid microeconomic grounds. Coverage includes both theoretical and methodological articles (game theory, mathematical methods in economics) as well as empirical articles (political economy, institutional economics, public economics). AUCO encourages also short communications (usually limited to 2,000 words) that provide an instrument of rapid and efficient dissemination of new results, models and methods in above mentioned fields of economic research. One of the primary purposes is to serve as a common ground for economists and political scientists who explore political economy from a formal perspective (positive political economy, public choice and social choice, political economics). Another goal is to attract key contributions

of gifted European junior economists.

*AUCO Czech Economic Review* is published by Charles University in Prague. The journal has been established in year 2007 as a descendant to a traditional Czech-written outlet, *Acta Universitatis Carolinae Oeconomica*. Three issues are published per volume. All articles and communications are available online for free. Printed issues can be ordered at a cost. The editors maintain classic **double blind peer review** procedure at high academic standards but at the same time emphasize **dynamic referee process** so that the journal tracks scientific progress in real time. The journal is indexed in EconLit, EBSCO, RePEc, in Czech Government list of reviewed journals, and recently is considered for indexing in Scopus.

## Submissions

Submissions to *AUCO Czech Economic Review* are welcome. The paper must be an original unpublished work written in English (consistent British or American), not under consideration by other journals. Instruction for authors is available on journal web-site.

Contact [auco@fsv.cuni.cz](mailto:auco@fsv.cuni.cz)

Web: <http://auco.cuni.cz/>

Post address: UK FSV IES, AUCO Czech Economic Review, Opletalova 26, 110 00 Prague 1, Czech Republic

## **Editorial Board**

### **Eitor in Chief**

Laura **Ștefănescu**

### **Managing Editor**

Mădălina **Constantinescu**

### **Executive Editor**

Mihaela **Galiceanu**

### **International Relations Responsible**

Pompiliu **Constantinescu**

### **Proof – readers**

Camelia **Firică** – *English*

Ana-Maria **Tranțescu** – *English*

### **Redactors**

Cristiana Bogdănoiu

Sorin Dincă

## **Editorial Advisory Board**

Claudiu **Albulescu**, University of Poitiers, **France**, West University of Timișoara, **Romania**

Huseyin **Arasli**, Eastern Mediterranean University, **North Cyprus**

Cristina **Barbu**, Spiru Haret University, **Romania**

Christoph **Barmeyer**, Universität Passau, **Germany**

Amelia **Bădică**, University of Craiova, **Romania**

Gheorghe **Bică**, Spiru Haret University, **Romania**

Ana **Bobîrcă**, Academy of Economic Science, **Romania**

Anca Mădălina **Bogdan**, Spiru Haret University, **Romania**

Elena **Doval**, Spiru Haret University, **Romania**

Jean-Paul **Gaertner**, École de Management Strasbourg, Université de Strasbourg, **France**

Shankar **Gargh**, Editor in Chief of *Advanced in Management*, **India**

Ioan **Gâf-Deac**, Spiru Haret University, **Romania**

Eugen **Ghiorghiță**, Spiru Haret University, **Romania**

Emil **Ghiță**, Spiru Haret University, **Romania**

Dragoș **Ilie**, Spiru Haret University, **Romania**

Arvi **Kuura**, Pärnu College, University of Tartu, **Estonia**

Ion Viorel **Matei**, Spiru Haret University, **Romania**

Constantin **Mecu**, Spiru Haret University, **Romania**

Piotr **Misztal**, Technical University of Radom, Economic Department, **Poland**

Marco **Novarese**, University of Piemonte Orientale, **Italy**

Russell **Pittman**, International Technical Assistance Economic Analysis Group Antitrust Division, **United States of America**

Ion **Popescu**, Spiru Haret University, **Romania**

Kreitz Rachel **Price**, École de Management Strasbourg, Université de Strasbourg, **France**

Peter **Sturm**, Institut National de Recherche en Informatique et Automatique, **France**

Andy **Ștefănescu**, University of Craiova, **Romania**

Laura **Ungureanu**, Spiru Haret University, **Romania**

Hans-Jürgen **Weißbach**, University of Applied Sciences - Frankfurt am Main, **Germany**

## **Publisher:**



Spiru Haret University  
Faculty of Financial Management Accounting Craiova



European Research Center of Managerial Studies in Business Administration

Brazda lui Novac Street, no 4 Craiova, Dolj, Romania

Phone: +40 251 598265 Fax : + 40 251 598265

Email: [jaes\\_secretary@yahoo.com](mailto:jaes_secretary@yahoo.com)

Web: [www.jaes.reprograph.ro](http://www.jaes.reprograph.ro)

## **Journal of Applied Economic Sciences**

**Journal of Applied Economic Sciences** is a young economics and interdisciplinary research journal, aimed to publish articles and papers that should contribute to the development of both the theory and practice in the field of Economic Sciences.

The journal seeks to promote the best papers and researches in management, finance, accounting, marketing, informatics, decision/making theory, mathematical modelling, expert systems, decision system support, and knowledge representation. *This topic may include the fields indicated above but are not limited to these.*

Journal of Applied Economic Sciences be appeals for experienced and junior researchers, who are interested in one or more of the diverse areas covered by the journal. It is currently published quarterly with three general issues in Winter, Spring, Summer and a special one, in Fall.

The special issue contains papers selected from the *International Conference* organized by the *European Research Centre of Managerial Studies in Business Administration* ([www.cesmaa.uv.ro](http://www.cesmaa.uv.ro)) and *Faculty of Financial Management Accounting Craiova* in each October of every academic year. There will prevail the papers containing case studies as well as those papers which bring something new in the field. The selection will be made achieved by:

- Conference Session Chairs
- Editorial Board Members

**Journal of Applied Economic Sciences** is indexed in CEEOL [www.ceeol.org](http://www.ceeol.org), SCOPUS [www.scopus.com](http://www.scopus.com), EBSCO [www.ebsco.com](http://www.ebsco.com), RePEc [www.repec.org](http://www.repec.org) and in [IndexCopernicus www.indexcopernicus.com](http://www.indexcopernicus.com) databases.

The journal will be available on-line and will be also being distributed to several universities, research institutes and libraries in Romania and abroad. To subscribe to this journal and receive the on-line/printed version, please send a request directly to [jaes\\_secretary@yahoo.com](mailto:jaes_secretary@yahoo.com)

# *Journal of Applied Economic Sciences*

ISSN 1843-6110

## Table of Contents



	Alessio Emanuele <b>BIONDO</b> <i>A Growth Rate for a Sustainable Economy</i>	... 7
	Maria <b>BOBROVA</b> , Arndt <b>KÜMPEL</b> <i>Reasoning on Evolution of Culture and Structure</i>	... 21
	A.B. <b>BONACHE</b> , J. <b>MAURICE</b> , K. <b>MORIS</b> <i>A Best Evidence Synthesis on the Link between Budgetary Participation and Managerial Performance</i>	... 34
	Ginters <b>BUSS</b> <i>Forecasts with Single-Equation Markov-Switching Model: An Application to the Gross Domestic Product of Latvia</i>	... 48
	Lisi <b>GAETANO</b> <i>The Unemployment Volatility Puzzle: The Role of the Underground Economy</i>	... 59
	Giuseppe <b>GAROFALO</b> , Patrizio <b>MORGANTI</b> <i>The Financing of R&amp;D Investments: Effects on Growth and Financial Structure</i>	... 70
	Ivan O. <b>KITOV</b> , Oleg I. <b>KITOV</b> <i>Dynamics of Unemployment and Inflation in Western Europe: Solution by the 1-D Boundary Elements Method</i>	... 94
	Evgenia <b>MOTCHENKOVA</b> , Daniel <b>LELIEFELD</b> <i>Adverse Effects of Corporate Leniency Programs in View of Industry Asymmetry</i>	... 114
	Rajesh K. <b>PILLANIA</b> <i>Indo-China Trade: Trends, Composition and Future</i>	... 129
	Georg <b>QUAAS</b> <i>Was the Worldwide Asymmetry in Current Accounts Caused by the Macroeconomic Policy of the Global Economy's Leader?</i>	... 138