Journal of Applied Economic Sciences

JAES

Journal of Applied Economic Sciences

Volume IV

Issue 1(7)

Spring 2009

I.S.S.N. 1843-6110

Publisher:

European Research Center of Managerial Studies in Business Administration

Email: jaes_secretary@yahoo.com

Journal of Applied Economic Sciences

***J*ournal of *A*pplied *E*conomic *S*cience** 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.*

***J*ournal of *A*pplied *E*conomic *S*ciences** 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

***J*ournal of *A*pplied *E*conomic *S*cience** is indexed in CEEOL - www.ceeol.org, EBSCO www.ebsco.com and in RePEc www.repec.org 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

Journal of Applied Economic Sciences

ISSN 1834-6110

Table of Contents

Journal of Applied Economic Sciences
Volume IV/ Issue 1(7)/ Spring 2009

ACTIVITY BASED COST FROM THE PERSPECTIVE OF COMPETITIVE ADVANTAGE

Cristiana **BOGDĂNOIU**
Spiru Haret University, **Romania**
Faculty of Financial Accounting Management Craiova
cristina82b@gmail.com

Abstract

Activity-Based Costing (ABC) is a methodology that measures the cost and performance of activities, resources, and cost objects such as products and services to provide more accurate cost information for managerial decision making.

ABC represents an alternative paradigm to traditional cost accounting system and it often provides more accurate cost information for decision making such as product pricing, product mix, and make-orbuy decisions. ABC models the causal relationships between products and the resources used in their production and traces the cost of products according to the activities through the use of appropriate cost drivers.

Keywords: Activity-Based Costing, management, performance, resources

JEL Classification: M11, D7

1. Introduction

In the last several years we can see a dramatically growing importance of the high quality information for the company management decision. The area of the cost management systems used for costing and budgeting is one of the most important areas of company financial management. According to growing competition on the globalized markets, companies need the information about the profitability of a product, customers or markets, about cost consumed by different activities and other different areas where the costs have the important role. If the company wants to keep in touch with the strongest competitors, its costing system has to implement the ability to react to changes in product and activities structure and feature these changes in the product costing. If the costing system does not change and does not conform with process, activities and product structure dynamics, than the costing system will become obsolete and will produce the incorrect information about the company cost.

Modern process and activity based costing systems has been developed in early 1980's and are mostly used by US companies and the European top companies.

Activity – Based Costing (ABC) is a method for developing cost estimates in which the project is subdivided into discrete, quantifiable activities or a work unit. The activity must be definable where productivity can be measured in units (e.g., number of samples versus manhours). After the project is broken into its activities, a cost estimate is prepared for each activity. These individual cost estimates will contain all labor, materials, equipment, and subcontracting costs, including overhead, for each activity.

Each complete individual estimate is added to the others to obtain an overall estimate.

Contingency and escalation can be calculated for each activity or after all the activities have been summed. ABC is a powerful tool, but it is not appropriate for all cost estimates.

2. The Genesis of Activity Based Costing

Activity-Based Costing shows - or at least attempts to show - the impact of changes in the costs and yields of every activity on the results of the whole. Peter F. Drucker

In the business literature, emphasis on business activities and processes can be traced to Michael Porter's (1980) "value chain" framework for analyzing sources of firms' competitive advantage. During the same time period, Porter's colleagues at Hardvard Business School, Robin Cooper and Robert Kaplan, discovered firms that had modified traditional cost accounting systems to better represent the flow of resources to the products and services that consumed them. Traditional costing systems are typically designed to allocate overhead costs from the general ledger to an entity with reporting responsibility (e.g. cost centers, profit centers) and from that entity to the products or services that it produces. The latter allocation of overhead costs to products is tipically accomplished by prorating overhead costs on the basis of direct labor hours, material costs, or another measure of unit variable costs.

The cost systems that Cooper and Kaplan identified bore a strong resemblance to Porter's value chain – with overhead costs traced from the general ledger to a *business activity.* Business activities often cross boundaries of the reporting and responsibility structure of the organization. Moreover, unlike traditional systems, activity costs were subsequently traced to products or services that place demands on the activity using "cost drivers" to assign costs in proportion to the level of demand for the activity. In documenting these systems, Kaplan and Cooper constructed a theory of resource consumption and suggested an approach, known as *Activity-Based Costig* (ABC) for representing resource provision and consumption in cost accounting systems. These representations led to prescriptions for managing the firm's value chain through *Activity-Based Management* (ABM).

Today, most managers are aware of activity based costing as a management information tool and many have direct involvement in the implementation of ABC or have access to ABC data. Technological advances, such as installation of *Enterprise Resource Planning Systems* (ERPs), integration of disparate databases into easily queried "data warehouses", and development of powerful data analysis tools, have facilitated the technical demands that ABC places on organization. However, these advances have generally not dealt with a basic finding of this study – effective ABC systems are often defined, not by the accuracy of the cost data, but by their ability to answering pressing questions.

3. Activity-Based Costing – A Cost Management Tool

Activity-Based Cost (ABC) is a tool for cost management. Activity-Based Management seeks to portray a company as a series of activities which are related to customer desires and cost. Activity-Based Cost is a process for measuring the cost of the activities of an organization. Activities within an organization are identified and an average cost is associated with each activity. The total cost of a product is the sum of the costs of activities required to bring forth, sustain, and retire the product. The cost of an activity for a product is defined as the average cost of the activity times the number of times the activity is required for that product.

Turney (1989) notes that *Underlying ABC is the assumption that activities consume resources and products consume activities. Activities include establishing vendor relations, purchasing, receiving, disbursing, setting up a machine, running the machine, reorganizing the production flow, redesigning the product, and taking a customer order. The performance of these activities triggers the consumption of resources that are recorded as costs in accounts. The activities are performed in response to the need to design, produce, market, and distribute products*.

Appropriately applied, ABC provides a far more accurate portrayal of cost than previous accounting methods. Given a better understanding of cost, management can make far better decisions in terms of competitive advantage. Furthermore, the improved understanding and localization of cost can be used to eliminate low value high cost activites and hence reduce cost. It is thus an aid to business process reengineering

Cooper and Kaplan (1992) show how activity-based costing permits the very important distinction between resource usage and resource spending. The difference is unused capacity. Elimination of this unused capacity permits costs to be reduced. Unfortunately, they fail to note the necessity to take into account the statistical variation within the system which can lead to the inability to supply near peak demands and the necessity to ensure that associated downsizing does not reduce either product or enterprise quality. Failure to meet demand and reduction in product quality can lead to revenue reductions which can generate a net loss in profit over the future. Reduction in enterprise quality can increase the average cost of a number of activities to the point that the net cost reduction is negative. Cost and quality are tightly linked within the dynamic stochastic nature of the business system.

Turney (1992) illuminates important nontemporal links between cost and enterprise quality. He also illustrates important relations between resources, resource drivers, activities, activity drivers, processes, enterprise performance (enterprise quality characteristics), cost drivers, and cost. He also notes that *Cost and non-financial information join forces to provide a total view of the work done ....*

This is the core concept upon which paramatric cost analysis and theoretical cost analysis have been based since at least the early 1960s.

Activity-Based Cost is a special form of function cost analysis where the cost of the functions of the system to bring forth, sustain, and retire the product are measured, as opposed to the functions of the product measured in value engineering.

ABC can also be used as the measurement tool for transaction cost economics by defining the activities to be those providing the desired transactions as outputs and using the number of transactions as the cost driver. Note that activities and transactions form the nodes and links of a network with activity cost and number of transactions being the associated measures. This observation links cost to graph theory and network topology.

The definition of accounting supplied by Zlatkovich, *et al.* (1966) is *the process of identifying, measuring, and communicating economic information to permit informed judgements and decisions by users of the information.*

This definition leaves plenty of room for the application of mathematical techniques within accounting. However, the practice of accounting limits itself to numbers, as opposed to equations and coordinate systems. It only took the application of mathematics, coupled with the scientific experiment, to transform physics from an anecdotal practice into a science.

From a mathematical perspective, we can consider the activity costs to form a coordinate system which describes the means which the enterprise uses to bring forth, sustain, or retire a product. There are also other perspectives of value to management. Following Kaplan (1988) we suggest that one cost perspective is not enough. For example, we place constraints on the organization to ensure timeliness compliance, cost compliance, and quality compliance for the customer. How much do those constraints cost? This triple of compliance costs is a very useful coordinate system. Yoshikawa, Innes, and Mitchell (1994) use function analysis to develop a transformation from an activity coordinate system to the coordinate system associated with this compliance triple. Another common coordinate system is that which defines the cost of each part within a product. The mappings performed to obtain part cost or functin cost from the activities consumed by the part or function are common, but unidentified, coordinate transformations.

The existence of these coordinate systems and transformations between them suggests that the mathematical concept of coordinates could play as big a role in the future of cost as it has played in the development of physics.

ABC is a methodology that measures the cost and performance of activities, resources, and cost objects to provide more accurate cost information for managerial decision making.[1] ABC is not an accounting exercise, but rather a methodology that produces a bill of activities that describes the cost buildup for individual products, services, or customers.[2] By recognizing the causal relationships among resources, activities, and cost objects such as products or customers, ABC allows one to identify inefficient or unnecessary activities and opportunities for cost reduction or profit enhancement.

The basic concept of ABC is that activities consume resources to produce an output (Figure 1). Expenses should be separated and matched to the level of activity that consumes the resources. Speciafically, the expenses that are needed to produce individual units of a particular service or product should be separated from the expenses that are incurred to produce different products or services or to serve different payers.

[1] Turney, Peter B.B., (1992), *What an ABC Model Looks Like*, in: *Journal of Cost Management 5*, no. 4 , Winter, pp. 54

[2] Kaplan, Robert S., (1992), *In Defense of Activity-Based Cost Management*, in: *Management Accounting 74*, no. 5, November, pp. 58

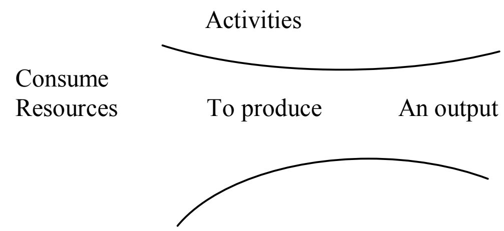


Figure 1. Theory of Resource consumption

Source: Copyrighted Material, Baker, Judith, J., (1998), *Activity Based Costing and Activity – Based Management for Health Care,* Hardcover,

The ABC approach differs from the traditional approach because of its fundamental concentration on activities. An ABC approach uses both financial and nonfinancial variables as bases for cost allocation. A typical ABC approach utilizes more indirect cost pools than does the traditional approach and uses a greater number of cost drivers as cost allocation bases.

As shown in Figure 2, there are two views of ABC: a cost assignment view and a process view.[3] The cost assignment view assigns costs to the significant activities of an organization.

Activities are then assigned to a cost object that uses the activities such as a product or customer. The process view provides operational intelligence about the processes of an organization. A process is a series of activities that are linked together to achieve an objective. The process view provides information about cost drivers and performance measures for each activity or series of activities in a process.

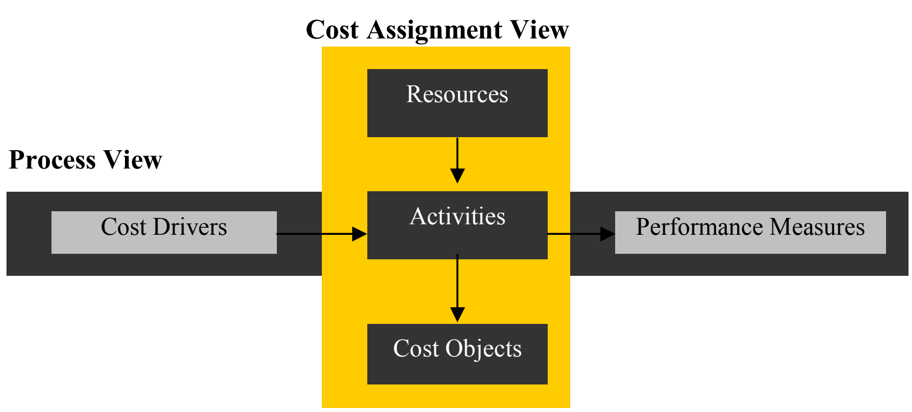


Figure 2. The ABC Model

Source: Turney, Peter B.B., (1991), *Common Cents: The ABC Performance Breakthrough,* Hillsboro: Cost Technology

The cost assignment view is comprised of three building blocks: resources, activities, and cost objects (Figure 2). Resources are economic elements that are the sources of cost. In a logistics operation, resources can include direct labor, direct material, and indirect costs (e.g., overhead and management salaries). Activities are the processes or procedures that produce work. Logistics

[3]Turney, Peter B.B., (1991), *Common Cents: The ABC Performance Breakthrough,* Hillsboro: Cost Technology, pp. 20

activities, for example, can include transportation, distribution, warehousing, order processing, and customer service. Since activities use resources, they are connected to activities via resource drivers that approximate the use of resources by activities (e.g., square footage, percent of effort, etc.). Each resource that is traced to an activity becomes a cost element in an activity cost pool that measures the total cost associated with an activity. This provides a better understanding of why resources are used. The information provided can help identify which activities consume the most resources and where cost reduction opportunities may exist.

The next step after assigning resources to activities is to trace the activities to cost objects. A cost object is typically a product, product line, or customer, so it is the reason why work is performed. Activity drivers measure the use of activities by the cost object, thus linking activities to cost objects. The total cost of the cost object is the sum of all the activity costs used by the cost object. This process provides economic information to help in analyzing decisions such as pricing, product mix, sourcing, product design, and improvement efforts.

As shown in Figure 2, three main building blocks comprise the process view: cost drivers, activities, and performance. Cost drivers determine why and how much work is required to perform an activity or a chain of activities.

A customer order, for example, initiates the order processing chain of activities - the "why". The size of the customer order determines how much work is required - the "effort". Cost drivers include both internal factors related to a specific activity and factors related to prior activities. Each activity in a series is a customer of a prior activity. Activities work together in an internal customer chain to provide value to the external customer.

Cost drivers are important because they reveal opportunities for improvement. A defect part received from a supplier, for example, will require correction activity to correct the problem, thereby expending more effort and resources. A quality certification program could help reduce a supplier's defect rate and thus reduce total costs of both the buyer and supplier.

Performance measures identify how well an activity is performed. Typical performance measures include activity efficiency, time required to complete an activity, and quality of work. Generally, the longer it takes to perform an activity, the greater the resources used and overall costs. Likewise, poor quality usually results in the use of more resources (e.g., scrap and rework in manufacturing organizations) and higher overall costs. The objective is to use this information to help improve performance and increase the value of products and services.

4. Continuous Improvement

The implementation of ABC can make the employees understand the various costs involved. This will then enable them to analyze the cost, and to identify the activities that add value and those that do not add value. Finally, based on this, improvements can be implemented and the benefits can be realiyed. This is a continuous improvement process in terms of analyzing the cost, to reduce or eliminate the non value added activities and to achieve an overall efficiency.

ABC has helped enterprises in answering the market need for better quality products at competitive prices. Analyzing the product profitability an customer profitability, the ABC method has contributed effectively fot the top management's decision making process. With ABC, enterprises are able to improve their efficiency and reduce the cost without sacrificing the value for the customer. Many companies also use ABC as a basisi for a balanced scorecard.

This has also enabled enterprises to model the impact of cost reduction and subsequently confirm the savings achieved. Overall, Activity-Based Costing is a dynamic method for continuous improvement. With Activity-Based Costing any enterprise can have a built-incompetitive cost advantage, so it can continuously add value to both its stakeholders and customers.

The implementation of ABC is not easy – not an ABC.

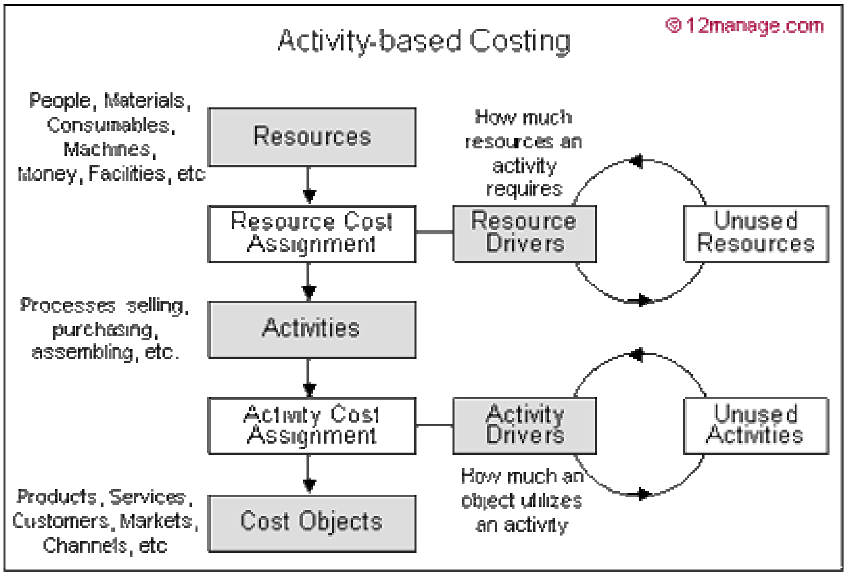


Figure 3. Activity-Based Costing

Source: 12manage

5. Conclusion

Activity-Based Costing (ABC) is a methodology that produces a bill of activities for cost objects such as individual products, services, or customers by measuring the cost and performance of activities and resources. It provides more accurate cost information than traditional cost accounting systems by recognizing the causal relationships among resources, activities, and cost objects.

6. References

[1] Akyol, Derya Eren; Tuncel, Gonca, and Bayhan, G. Mirac, (2005), *A comparative analysis of activity-based costing and traditional costing*, Proceedings of World Academy of Science, Engineering and Technology, Volume 3 January 2005, ISSN 1307-6884.

[2] Anderson, Shannon W.; Young, Mark S., (2001), *Implementing Management Innovations: Lessons Learned from Activity Based Costing in the U.S. Automobile Industry*, Springer, ISBN 0792374371, 9780792374374.

[3] Beyerly, W.E., (1916), *An Introduction to the Use of Generalized Coordinates in Mechanics and Physics*, Ginn and Company, New York NY, republished in 1965 by Dover Publications, New York, NY, USA.

[4] Oprea, Călin; Cârstea, Gheorghe, (2002), *Contabilitatea de gestiune şi calculația costurilor*, Editura Genicod, Bucureşti.

[5] Cooper, R.; Kaplan, R. S., (1992), *Activity-Based Systems: Measuring the Costs of Resource Usage*, Accounting Horizons, September.

[6] Cucui, Ion; Horga, Vasile; Radu, Mariana, (2003), *Control de gestiune,* Editura Niculescu, Bucureşti.

[7] Ebbeken, Klaus; Possler, Ladislau; Ristea, Mihai, (2001), *Calculația şi managementul costurilor*, Editura Teora, Bucureşti.

[8] Kaplan, Robert S., (1992), *In Defense of Activity-Based Cost Management*, in: *Management Accounting 74*, no. 5, November.

[9] Moore, Kevin R., (2000), *Using Activity-Based Costing to Improve Performance: A Case Study Report*, Maxwell Air Force Base, Alabama, April.

[10] Popesko, Boris; Novak, Petr, *Activity – Based Costing aplications in Czech Republic*, in: Lex *et Scientia* No. XV Vol. 1/2008.

[11] Ştefãnescu, Laura; Ungureanu, Laura; Matei, Viorel, (2008), *Reengineering as an Efficient Solution to Redesign Activities and Processes of an Enterprise,* in: *Journal of Applied Economic Sciences (JAES)*, VolumeIII, Issue1(3), Spring 2008.

[12] Turney, Peter B.B., (1991), *Common Cents: The ABC Performance Breakthrough,* Hillsboro: Cost Technology.

[13] Turney, Peter B.B., (1992), *What an ABC Model Looks Like*, in: *Journal of Cost Management 5*, no. 4, Winter.

[14] Yoshikawa, T.; Innes, J.; and Mitchell, F. (1990), *Cost Tables: A Foundation of Japanese Cost Management*, in: *Journal of Cost Management*, Fall.

[15] Zlatkovich, C.T. *et al.* (1966), *American Accounting Association Committee to Prepare a Statement of Basic Accounting Theory*, American Accounting Association, Sarasota, USA.

[16] Varsanyi, Zoltan, (2008), *A Simple Model of Decision Making –How to Avoid Large Errors?*, in: *Journal of Applied Economic Sciences (JAES)*, VolumeIII, Issue3(5), Fall2008.

OVERCONFIDENCE BIAS: EXPLANATION OF MARKET ANOMALIES FRENCH MARKET CASE

Mouna **BOUJELBENE ABBES**, Younès **BOUJELBENE** and Abdelfettah **BOURI**
University of Sfax (FSEG), **Tunisia**
Faculty of Business and Economics
abbes.mouna@gmail.com

Abstract

In this study, we test whether the overconfidence bias explains several stylized market anomalous, including a short-term continuation (momentum), a long-term reversal in stock returns, high levels of trading volume and excessive volatility. Using data of French stocks market, we find empirical evidence in support of overconfidence hypothesis. First, based on a restricted VAR framework, we show that overconfident investors overreact to private information and underreact to public information. Second, by performing Granger-causality tests of stock returns and trading volume, we find that overconfident investors trade more aggressively in periods subsequent to market gains. Third, based on a two GARCH specifications, we show that self attribution bias, conditioned by right forecasts, increases investors overconfidence and trading volume. Fourth, the analysis of the relation between return volatility and trading volume shows that the excessive trading of overconfident investors makes a contribution to the observed excessive volatility.

Keywords: overconfidence, behavioural finance, over (under) reaction, trading volume, volatility

JEL Classification: C52, G12

1. Introduction

Some fascinating events in financial markets, such as a short-term continuation (momentum) and a long term reversal in stock returns, high levels of trading volume and excessive volatility cannot be explained by traditional models based on investors' rationality. A developing strand of the finance literature [Benos, (1998), Daniel *et al*., (1998), and Odean, (1998)] proposes theoretical models built on the assumption of investor overconfidence to account for these observed anomalies.

Daniel Hirshleifer and Subrahmanyam (1998) develop a behavioural model based on the assumption that investors display overconfidence and self-attribution bias. In their model, the informed traders attribute the performance of ex-post winners to their stock selection skills and that the ex-post losers to bad luck. As result, these investors become overconfident about their ability to pick winners and thereby overestimate the precision of their signals. Based on their increased confidence in their signals, they push up the price of the winners above the fundamental value. The delayed overreaction in this model leads to momentum profits that are eventually reversed as prices revert to their fundamentals (reversals).

According to Gervais and Odean (2001) overconfidence is enhanced in investors that experience high returns, even when those returns are simultaneously enjoyed by the entire market. Odean (1998) and Gervais and Odean (2001) suggest that intertemporal changes in trading volume are the primary testable implication of overconfidence theory.

Moreover, overconfidence is proposed as an important reason for excessive price volatility. Benos (1998) proposes a model in which overconfident traders' aggressive exploitation of their profitable information, together with rational traders' conservative trading strategy, leads prices to move too much in one or the other direction. In their model, Daniel Hirshleifer and Subrahmanyam (1998) show that overconfident investors increases prices volatility at the time reception of private signals.

In this study, we attempt to identify the contribution of overconfidence bias to explaining several stylized market anomalous (momentum, reversals, excessive trading volume and excessive volatility). For this end, four hypotheses derived from overconfidence previous theoretical work are tested: H1, overconfident investors overreact to private information and underreact to public information. H2, market gains (losses) make overconfident investors trade more (less) aggressively in subsequent periods. H3, self-attribution bias, conditioned by right forecasts, increases investors overconfidence and trading volume. H4, the excessive trading of overconfident investors makes a contribution to the observed excessive volatility.

Our contribution is to provide the empirical evidence on various implications of the overconfidence hypothesis by focusing on aggregate French market behaviour.

The methodology followed in this study considers various empirical frameworks. First, a Bivariate Vector Autoregression is employed to study the impulse responses of stock returns to private and public information shocks. Second, Granger causality tests are used for testing the relation between returns and trading volume. Then, two GARCH (EGARCH and GJR-GARCH) models are used to evaluate the effect of investors' error forecasts on their overconfidence. Finally, the conditional variance of GARCH models is estimated by introducing two components of trading volume. The first component, due to past stock returns, is related to investors' overconfidence. The second component is unrelated to investors' overconfidence.

The rest of the paper is organized as follows. In section 2, we present the hypotheses and empirical methodology. Section 3 describes empirical data. Section 4 presents the empirical results. Section 5 concludes the paper.

2. Hypotheses and empirical methodology

2.1. Overconfidence and differential reaction to information

DHS (1998) and Gervais and Odean (2001) models predict that overconfident investors overestimate the precision of their own valuation abilities, in the sense that they overestimate the precision of their private information signals As a result, they make investment decisions by relying on their own private signals while they ignore public signals. Based on these theoretical predictions, we derive the first hypothesis, *H1*, written as follows:

H1. Overconfident investors overreact to private information and underreact to public information.

To identify private and public information, the methodology presented by Chuang and Lee, (2006) is considered. A structural VAR (Vector Autoregression) model is employed. Consider a vector γ_t ($\gamma_t = [V_t, r_t]'$) consisting of two stationary variables: trading volume V_t and stock return r_t series. Based on the Wold theorem, the vector γ_t has a Bivariate Moving Average Representation (BMAR) given by the following relation:

$$\begin{vmatrix} V_t \\ r_t \end{vmatrix} = \begin{vmatrix} B_{11}(L) & B_{12}(L) \\ B_{21}(L) & B_{22}(L) \end{vmatrix} \begin{vmatrix} \varepsilon_t^{private} \\ \varepsilon_t^{public} \end{vmatrix} \quad (1)$$

where $\varepsilon_t^{private}$ and ε_t^{public} are respectively the private and the public information shock. $b_{ij}(L)$ represent the effect of shocks on the trading volume and the stock returns. The shocks on private and public information are distinguished by a restriction imposed on the BMAR. That is, private information shock has a contemporaneous impact on trading volume, while public information shock has no contemporaneous impact on trading volume. This restriction is motivated by several theoretical considerations and it can be formally written as follows:

$$b_{12}(k)/k = 0 = b_{12}(0) = 0 \quad (2)$$

On the one hand, the DHS (1998) overconfidence model shows that excessive trading volume is primarily due to investor's overreaction to their private signals and their underraction to public information. In addition, Campbell *et al.* (1993) show that public information does not affect significantly market trading volume [Chuang and Lee, (2006)].

Really, the BMAR is derived by inverting a Bivariate Vector Autoregression (BVAR), given by the following expression:

$$\gamma_t = \begin{vmatrix} V_t \\ r_t \end{vmatrix} = a(L)\gamma_{t-1} + u_t \equiv \begin{vmatrix} a_{11}(L) & a_{12}(L) \\ a_{21}(L) & a_{22}(L) \end{vmatrix} \begin{vmatrix} V_{t-1} \\ r_{t-1} \end{vmatrix} + \begin{vmatrix} u_{1t} \\ u_{2t} \end{vmatrix} \tag{3}$$

The relation between the BVAR model [Eq. (1)] and the BMAR model [Eq. (3)] is described in the Chuang and Lee (2006) study.

Once a restricted BVAR model of trading volume and stock return is estimated, we can analyze the stock return responses to private and public information shocks to see whether the responses are compatible with the prediction of the overconfidence hypothesis (*H1*).

2.2. Overconfidence and trading volume

Several studies consider the proposition that investor overconfidence generate the high trading volume observed in financial markets [Odean, (1998a, 1998b, 1999), Gervais, and Odean, (2001)]. Gervais, and Odean, (2001) and Odean, (1998) theoretical models predict that high total market returns make some investors overconfident about the precision of their information. Although the returns are market wide, investors mistakenly attribute gains in wealth to their ability to pick stocks. Overconfident investors trade more frequently in subsequent periods because of inappropriately tight error bounds around return forecasts. Alternatively, market losses reduce investor overconfidence and trading, although perhaps not in a symmetric fashion. Thus, the second hypothesis of overconfidence predicts a causality running from stock returns to trading volume.

H2. Market gains (losses) make overconfident investors trade more (less) aggressively in subsequent periods.

To identify the relation between stock returns and trading volume, we use a bivariate Granger causality test. Formally, if the prediction of *Y* using past values of *X* is more accurate than the prediction without using *X* in the mean square error sense ($\sigma^2(Y_t/\Omega_{t-1})\ \mathrm{p}\ Y_t/\Omega_{t-1} - X_t)$, where Ω_t is the information set at time *t*), then *X* Granger-causes *Y*.

The specification of used test is as follows:

$$V_t = \alpha_{11} + \alpha_{12}|r_t| + \alpha_{13}MAD_t + \sum_{j=1}^{p}\beta_{11j}V_{t-j} + \sum_{j=1}^{p}\beta_{12j}r_{t-j} + \varepsilon_{1t} \tag{4}$$

$$r_t = \alpha_{21} + \alpha_{22}MAD_t + \sum_{j=1}^{p}\beta_{21j}V_{t-j} + \sum_{j=1}^{p}\beta_{22j}r_{t-j} + \varepsilon_{2t}$$

where, r_t is the market return, $|r_t|$ is the absolute value of r_t and MAD_t denotes the mean absolute cross-sectional return deviation:

$$MAD_t = \sum_{i=1}^{N} w_i |r_{it} - r_t| \tag{5}$$

where, r_{it} is the return of stocks *i.*

We choose the number of lag, *p*, by considering the Akaike Information Criterion (*AIC*) and Schwarz criterion.

The first control variable, $|r_t|$, is based on Karpoff's (1987) survey of research on the contemporaneous volume–volatility relationship. The second control variable, MAD_t is motivated by Ross (1989) intuition that in a frictionless market characterized by an absence of arbitrage opportunities, the rate of information flow is revealed by the degree of price volatility.

To test the overconfidence hypothesis, we focus on the null hypothesis that stock returns do not Granger-cause trading volume. The rejection of the null hypothesis (β_{12j} = 0, for any *j*) authenticating our second hypothesis. Moreover, the rejection of the null hypothesis that trading volume does not

Granger-cause stock returns (β_{21j} = 0, for any *j*) will be an evidence against the market efficiency (trading volume is not a fundamental variable of the firm). The presence of a feedback relation between stock returns and trading volume provides evidence in favor of positive feedback trading.

2.3. Self-attribution and investors' overconfidence

Another central aspect of the overconfidence-related finance literature that we consider is the biased self-attribution; the tendency of the individuals to attribute good outcomes to their own qualities and bad outcomes to bad luck or other factors. The self-attribution bias is considered by some behavioral models that attempt to provide a theoretical framework for the empirical return anomalies documented in the finance literature [DHS, (1998); Gervais and Odean, (2001)]. According to DHS (1998) model, investor overconfidence varies because of biased self-attribution, which means that when investors receive confirming public information, their confidence level increases, but when they receive disconfirming public information; their confidence level falls only modestly.

On the empirical level, biased self-attribution leads investors to become overconfident after a good past performance [Gervais and Odean, (2001)]. Consequently, trading volume is greater positively correlated with past stock returns conditional on investors' right forecasts, than that conditional on their wrong forecasts. Indeed, if investors make a right forecast in that they predict positive stock returns at time *t-1* and realized stock returns are positive at time *t*, then their overconfidence rises significantly and, consequently, they trade more actively in subsequent periods. If, on the other hand, investors make a wrong forecast in that they predict negative stock returns at time *t-1* and realized stock returns are positive at time *t*, then their overconfidence may fall only modestly because they still benefit from market gains [Chuang and Lee, (2006)]. We therefore formally state the third hypothesis as follows:

H3: *Self-attribution bias, conditioned by right forecasts, increases investors' overconfidence and their trading volum*e.

To empirically test this hypothesis, we use two different GARCH-type specifications (EGARCH and GJR-GARCH) taking into account an asymmetric effect in which a negative return shock increases volatility more than does a positive return shock (leverage effect). Equations 6 and 7 represent respectively EGARCH [Nelson, (1991)] and GJR-GARCH [Glosten, *et al.* (1993)] models:

$$\begin{aligned} & r_t = \mu_t + \eta_t \\ & \eta_t / (\eta_{t-1}, \eta_{t-2} ...., r_{t-1}, r_{t-2} ....) : \; GED(0, h_t), \\ & ln\, h_t = \omega + f_1 \left(\frac{|\eta_{t-1}| + \kappa \eta_{t-1}}{\sqrt{h_{t-1}}} \right) + f_2 h_{t-1} \end{aligned} \tag{6}$$

where μ_t and h_t represent respectively the expected return and conditional volatility.

The asymmetric effect in EGARCH model is represented by the volatility parameter κ . If $\kappa <$ *0,* then conditional volatility tend to increase (to decrease) when the standardized residual is negative (positive).

$$\begin{aligned} & r_t = \mu_t + \eta_t \\ & \eta_t / (\eta_{t-1}, \eta_{t-2} ...., r_{t-1}, r_{t-2} ....) : \; GED(0, h_t), \\ & h_t = \omega + f_1(\eta_{t-1}^2) + f_2 h_{t-1} + \theta S_{t-1}^{-}(\eta_{t-1}^2) \end{aligned} \tag{7}$$

where S_{t-1}^{-} = 1 if η_{t-1} p *0* and S_{t-1}^{-} = 0 so not.

The asymmetric effect in the GJR-GARCH model is represented by the volatility parameter θ . If θ f 0 then a negative shock has an impact on conditional volatility superior than does a positive shock.

To allow for the possibility of non-normality of the returns distribution, we assume that the conditional errors of EGARCH and GJR-GARCH specifications follow a Generalized Error

Distribution, *GED*. The two different GARCH-type specifications permit decomposing the stocks returns into expected and unexpected returns.

To test whether the self-attribution bias hypothesis can explain the investors' overconfidence dynamic, we estimate the following regression:

$$V_t = \alpha_0 + \alpha_1 |r_t| + \alpha_2 MAD_t + \alpha_3 r_t + \sum_{j=1}^{p} \beta_j \left(r_{t-j} \times I_{t-j}^{+} \right) + \sum_{j=1}^{p} \gamma_j \left(r_{t-j} \times (1 - I_{t-j}^{+}) \right) + \sum_{j=1}^{p} \lambda_j |\eta_{t-j}| + \varepsilon_t \quad (8)$$

where η_t is the unexpected return (or forecast error) derived from the GARCH specifications and $|\eta_t|$ is the absolute value of η_t. The dummy variable I_{t-j}^{+} takes on a value of one if $\mu_{t-1} \times r_t \succ 0$ in which μ_t is the expected returns derived from the GARCH specifications, and zero, otherwise.

The β_j and γ_j coefficients are designed to measure the effect of self-attribution bias on trading volume. If investors are subject to self-attribution bias, these coefficients will be positive and that $\sum_{j=1}^{p} \beta_j \succ \sum_{j=1}^{p} \gamma_j$. The absolute value of forecast error $|\eta_t|$ is designed to measure the effect of the investors' forecasts precision on their overconfidence. If investors are overconfident, we expect the λ_j coefficients to be negative.

2. 4. Overconfidence and volatility

Overconfidence has been advanced as an explanation for the observed excessive volatility. Odean (1998), Gervais and Odean (2001) show that the volatility is increasing in a trader's number of past success and thereby in a level of investors' overconfidence. Therefore, the fourth hypothesis associated with overconfidence can be written as follows:

H4. The excessive trading of overconfident investors makes a contribution to the observed excessive volatility.

The relation between volatility and trading volume was the subject of many prior researches [Lamoureux and Lastrapes, (1990); Schwert, (1989); Benos, (1998); Albulescu, (2007)]. The objective of testing empirically our fourth hypothesis is to distinguish excessive trading volume of overconfident investors from other factors that affect volatility. In the one stage of the test procedure, the trading volume is decomposed into one component related to investors' overconfidence (*OVER*) and another non-related to the overconfidence (*NONOVER*) and it can be written as:

$$V_t = \alpha + \sum_{j=1}^{p} \beta_j r_{t-j} = \left[\sum_{j=1}^{p} \beta_j r_{t-j} \right] + \left[\alpha + \varepsilon_t \right] = OVER_t + NONOVER_t \quad (9)$$

In the second stage, we include these two components of trading volume into the conditional variance equation of EGARCH and GJR-GARCH models, respectively as follows:

$$\begin{aligned} & r_t = \mu_t + \eta_t \\ & \eta / \left(V_t, \eta_{t-1}, \eta_{t-2}, \ldots\ldots, r_{t-1}, r_{t-2} \right) \sim GED\left(0, h_t\right), \\ & Inh_t = \omega + f_1 \left(\frac{|\eta_{t-1}| + \kappa \eta_{t-1}}{\sqrt{h_{t-1}}} \right) + f_2 Inh_{t-1} + f_3 NONOVER_t + f_4 OVER_t \end{aligned} \quad (10)$$

and

$$\begin{aligned} & r_t = \mu_t + \eta_t \\ & \eta_t / \left(V_t, \eta_{t-1}, \eta_{t-2} \ldots\ldots\ldots, r_{t-1}, r_{t-2} \ldots. \right) \sim GED(0, h_t), \end{aligned} \quad (11)$$

$$h_t = \omega + f_1(\eta_{t-1}^2) + f_2 h_{t-1} + \theta S_{t-1}^{-}(\eta_{t-1}^2) + f_3 NONOVER_t + f_4 OVER$$

The parameter f_4 represents the effect of overconfidence on volatility and the parameter f_3 measures the effect of other factors on excessive volatility.

3. Data and descriptive statistics

Our sample consists of 120 French stocks traded in the Paris stocks exchange from January 1995 to December 2004. Data used in this empirical study are price, trading volume and turnover ratio. We use daily data to construct monthly market variables; notably, trading volume (Vol_t), turnover ratio (Tov_t), and return (r_t). Our choice of monthly variables is justified by the fact that the investors' overconfidence level change on the monthly or yearly horizons [Odean, (1998), Gervais and Odean, (2001) and Statman *et al.* (2003)]. Our focus on aggregate investor behavior is motivated in part by the argument of Odean (1998), DHS (2001), Gervais and Odean (2001) that investor behavior should be observable in market level data, and in other part by the idea of Kyle and Wang (1997), Benos (1998), DHS (1998), Hirshleifer and Luo (2001) and Wang (2001) that overconfident investors can survive and dominate the markets in the long run.

The market return is computed as the equally weighted index return:

$$r_t = \frac{1}{N}\sum_{i=1}^{N} r_{it} \qquad (12)$$

where r_{it} is the i stock return in t and N represents the number of sample stocks.

The turnover is defined as the ratio of the number of shares traded in a month to the number of shares outstanding at the end of the month. The use of trading volume and turnover is justified by the considerable increases of trades' number. Moreover, one problem with using the number of share traded as a measure of trading volume is that it is unscaled and, therefore, highly correlated with firm size [Chordia and Swaminathan, (2000)].

Figures 1(a) and 1(b) present, respectively, monthly turnover ratio and trading volume, from January 1995 to December 2004. Trading volume gradually increased to reach its maximum during the period of November 1999 until November 2002.

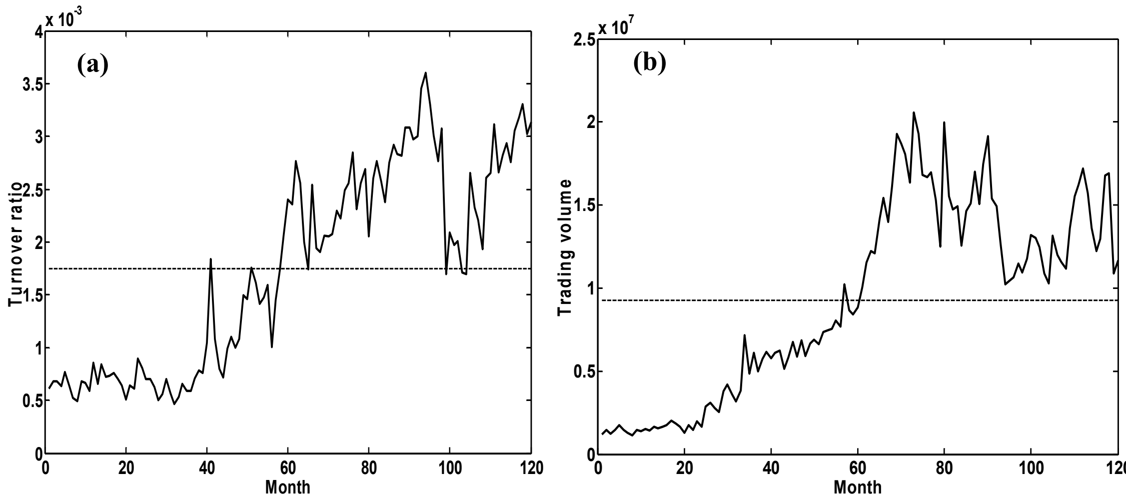


Figure 1. Monthly moving of turnover and trading volume

Source: (own)

Table 1 presents summary statistics on monthly market variables: return, turnover and trading volume. The normality test results show that the three variables distributions are not normal (Skewness ≠ 0 and kurtosis ≠

3). The results of stationary test show that the monthly returns and trading volume series are stationary. About the turnover ratio, Dickey and Fuller statistic (ADF test) is higher than critical value, which means that the series is non-stationary and exhibit a trend. Hodrick-Prescott (1997) algorithm is used to detrending this series.

Table1. Market descriptive statistics

	Return	Turnover	Trading volume
Mean	0.0107	0.0017	9263384
Standard deviation	0.0360	0.0009	5775993
Min	-0.1053	0.0004	1150292
Max	0.0901	0.0036	20563872
kurtosis	3.8096	1.5561	1.7266
Skewness	-0.5889	1.305	0.06813
Jarque and Bera	10.153	10.675	8.2002
ADF test	-8.0367	0.968	-9.98
Critical value (1%)	-4.0407	-3.4885	-3.449

Source: (own)

This table presents market descriptive statistics for monthly return, turnover and trading volume, from January 1995 to December 2004. The table reports, the mean, standard deviation, minimal value, maximum value, Skewness statistic, kurtosis statistic, Jarque and Bera statistic, ADF statistic (Augmented Dickey and Fuller) and critical value of ADF test.

4. Empirical results

4.1. Overconfidence and information differential reaction: Hypothesis 1

In this section, we report the empirical results of the first overconfidence hypothesis. To estimate the BVAR of y_t, we have to choose the number of lags in each equation. Formal overconfidence theories do not specify a time frame for the relationship between returns and trading volume, so we let the data determine the number of monthly lags to include. Specifically, we set five lag ($k=5$) based on both Schwartz Information Criteria (SIC) and Akaike Criteria (AIC). Two cases are considered to measure V_t series: the turnover, (Tov_t) and the trading volume (Vol_t).

Figures 2 and 3 present the impulse-responses of returns r_t to one standard deviation shocks on public and private information, ε_t^{public} and $\varepsilon_t^{private}$, respectively for Vol_t [Figures 2(a) and 2(b)] and Tov_t [Figures 3(a) and 3(b)]. The shocks on public and private information are orthogonalized using Cholesky decomposition. The dynamic responses of r_t are measured by standard deviations of this variable over 15 months. The figures present a conditional band of the standard error, computed with Monte Carlo simulation method, around the mean response.

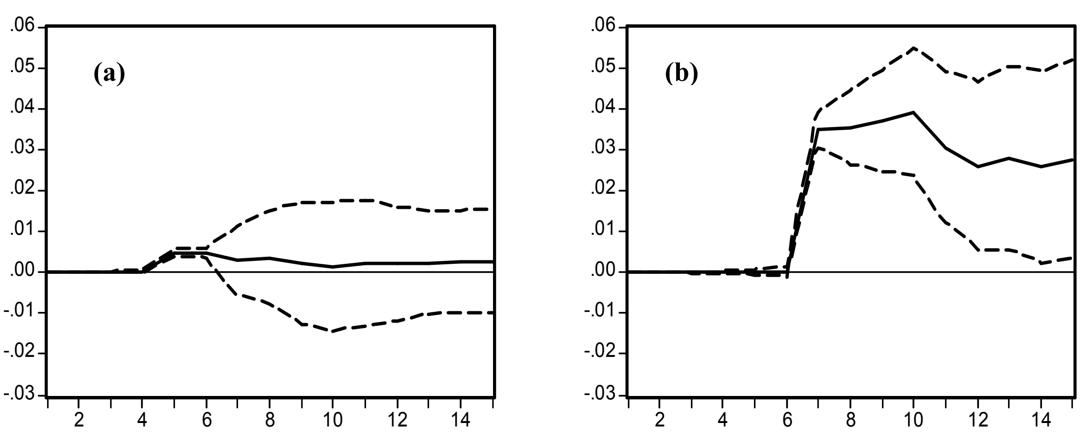


Figure 2. Response of stock returns to private and public information shocks, $V_t = Vol_t$

Source: (own)

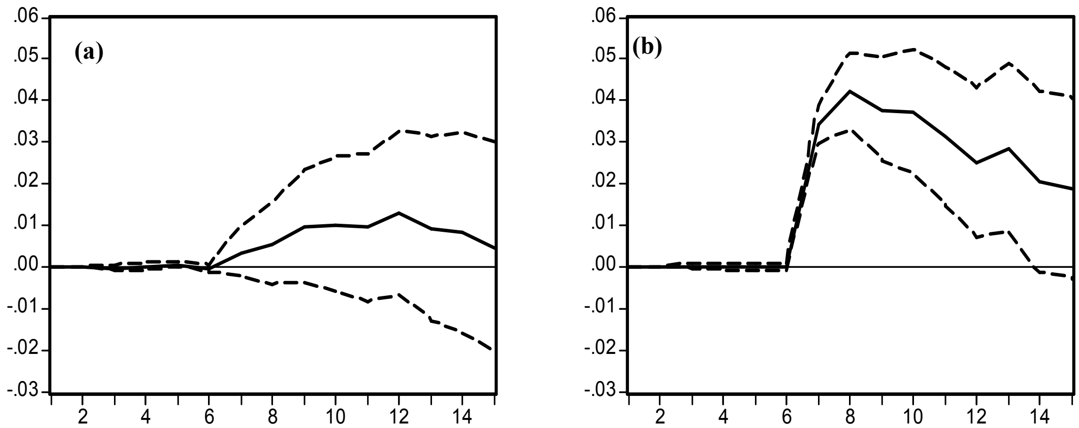


Figure 3. Response of stock returns to private and public information shocks, $V_t = Tov_t$.

Source: (own)

Figures 2(a) and 3(a) illustrate under-reaction of returns to public information shocks. However, an overreaction of returns to private information shocks is showed in Figures 2(b) and 3(b). After an initial under-reaction, stocks prices reach their equilibrium through a correction process. Investors' under- and overreaction to information can work together or independently to generate short-term price continuation returns (*momentum*) and their long-term reversals. Indeed, overconfident investors buy stock which progressed while thinking that market did not sufficiently evaluated it compared to their private information. The returns progress beyond their value suggested by public information. The correction intervenes in long run when public information becomes such as it eclipses the private signals.

4.2. Overconfidence and trading volume: Hypothesis 2

Table 2 summarizes an estimation of Granger causality test in two panels. In Panel A, dependent variables vector is constituted of trading volume and returns $[Vol_t, r_t]'$. In Panel B vector of dependent variables is formed by turnover ratio and returns $[Tov_t, r_t]'$. The estimation results provide confirming evidence that stock returns positively Granger-cause investors' confidence. In addition, the cumulative effect of lagged monthly stock returns on trading volume is positive and significantly different from zero. Moreover, the predictive power in term of $\overline{R}^2$ coefficient is higher for the Tov_t and Vol_t dependent variables compared to that of r_t. This result is consistent with the overconfidence hypothesis suggesting that market gains help to envisage a trading volume increase.

Table 2. Bivariate causality tests of trading volume and stock returns

	Panel A				**Panel B**			
Dependent Variable	Tov_t		r_t		Vol_t		r_t	
Independent Variable	Tov_{t-j}	r_{t-j}	r_{t-j}	Tov_{t-j}	Vol_{t-j}	r_{t-j}	r_{t-j}	Vol_{t-j}
χ^2_1 *(p-value)*		11.3445 (0.029)		8.7431 (0.0679)		11.254 (0.024)		11.254 (0.0185)
$\sum$*coefficients*	0.8519	0.0003	0.117	-17.953	0.993	3.5893	0.089	-0.005

	Panel A				Panel B			
χ_2^2 (p-value)	1.59×10^{-11} (0.000)	10.649 (0.001)	0.275 (0.000)	0.434 (0.509)	1145 (0.000)	5.369 (0.020)	0.1900 (0.664)	2.970 (0.085)
$\overline{R}^2$		0.951		0.145		0.929		0.211
$Q(6)$		2.017		4.442		2.0167		2.017

Source: (own)

This table presents the results of Granger causality test estimate:

$$V_t = \alpha_{11} + \alpha_{12}|r_t| + \alpha_{13}MAD_t + \sum_{j=1}^{p}\beta_{11j}V_{t-j} + \sum_{j=1}^{p}\beta_{12j}r_{t-j} + \varepsilon_{1t}$$

$$r_t = \alpha_{21} + \alpha_{22}MAD_t + \sum_{j=1}^{p}\beta_{21j}V_{t-j} + \sum_{j=1}^{p}\beta_{22j}r_{t-j} + \varepsilon_{2t}$$

where V_t is the market trading volume (Tov_t or Vol_t) and r_t is the market return, $|r_t|$ is the absolute value of r_t and MAD_t represents the mean absolute cross-sectional return deviation. The χ_1^2 test statistic is used to test the double Granger causality. The χ_2^2 test statistic is used to test the null hypothesis that the sum of the estimated coefficients is equal to zero. The p-value is the probability of obtaining the value of the corresponding test statistic or higher under the null hypothesis. $\overline{R}^2$ is the adjusted coefficient of determination. Q(6) is the Ljung-Box Q-statistic used to test the joint significance of the autocorrelations up to 6 lags for the residuals in each regression.

4.3. Overconfidence and self-attribution bias

Results of the EGARCH and GJR-GARCH models estimation are shown in table 3. An asymmetric relationship between returns and volatility is noted. Indeed, negative return shocks of a given magnitude have larger impact on volatility than positive return shocks of the same magnitude. The GARCH estimator parameter f_2 is significantly positive for EGARCH and GJR-GARCH models. Consequently, the current returns variance is strongly related to that of previous period.

Table 3. Univariate EGARCH and GJR-GARCH regression parameters

Model	**EGARCH**		**GJR-GARCH**	
Parameter	*Coefficient*	*Z- statistic*	*Coefficient*	*Z- statistic*
ω	-1.7286	-1.6763	0.0003	1.8947
f_1	-0.0584	-0.3298	-0.1527	-2.0528
κ (θ GJR -*GARCH*)	-0.2297	-2.1077	0.1893	2.2423
f_2	0.7209	4.4962	0.8597	9.3396

Source: (own)

This table presents the results of the EGARCH and GJR-GARCH conditional variance estimation:

$$r_t = \mu_t + \eta_t$$

$$\eta_t / (\eta_{t-1}, \eta_{t-2} \ldots\ldots\ldots, r_{t-1}, r_{t-2} \ldots.) : \ GED(0, h_t),$$

$$\ln h_t = \omega + f_1\left(\frac{|\eta_{t-1}| + \kappa\eta_{t-1}}{\sqrt{h_{t-1}}}\right) + f_2 h_{t-1}$$

and

$$r_t = \mu_t + \eta_t$$
$$\eta_t / (\eta_{t-1}, \eta_{t-2}........, r_{t-1}, r_{t-2}....) : \ GED(0, h_t),$$
$$h_t = \omega + f_t(\eta^2_{t-1}) + f_2 h_{t-1} + \theta S^-_{t-1}(\eta^2_{t-1})$$

where, r_t is the market return, μ_t is the expected return and h_t represents conditional volatility.

In order to show the asymmetric response of volatility to good and bad news, we present in figure 4 the estimated news impact curve. This figure displays positive and negative shocks impact on EGARCH conditional variance. It is clear that bad news (negative shock) tends to increase volatility more than good news (positive shock).

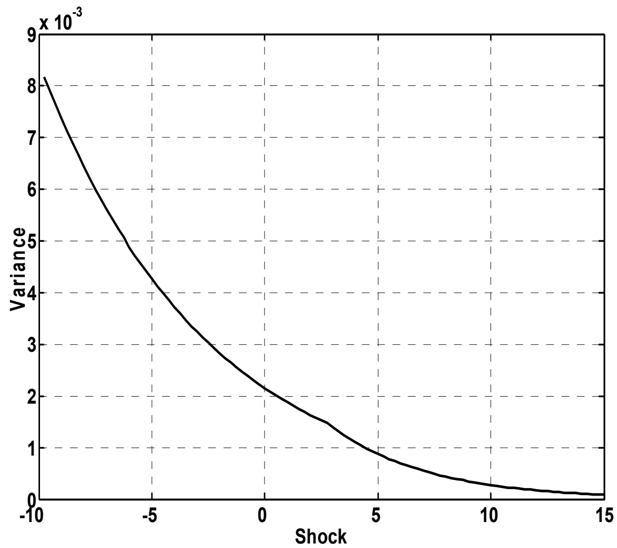


Figure 4. Asymmetric response of volatility to return shocks

Source: (own)

The unexpected return η, derived from GARCH models are then used to study the self-attribution effect on trading volume. Table 4 presents estimation of Eq. (8) in two panels. In panel A, the expected and unexpected returns (η and μ) are derived from the EGARCH model. In panel B, these returns are derived from the GJR-GARCH model. Two types of dependent variables are considered in each panel, turnover, Tov_t and trading volume, Vol_t. The regression is estimated by adopting the lag length of 3. The reason of this choice is due to the fact that our analysis from the bivariate Granger causality tests shows that the most significant positive causal relation between stock returns and trading volume concentrates on the first three months.

The estimation results show that the sum of β_j coefficients which captures the effect of investors' overconfidence on trading volume when they make right forecasts, is positive and the null hypothesis that $\sum_{j=1}^{3} \beta_j = 0$, is rejected. While, the sum of γ_j coefficients that measures the same effect when investors make wrong forecasts is negative and the null hypothesis that $\sum_{j=1}^{3} \beta_j = \sum_{j=1}^{3} \gamma_j$, is rejected. Consequently, the right and the wrong forecasts do not have the same effect on trading volume. The reception of information which confirms investors' forecasts tends to accentuate their overconfidence ($\beta_1 + \beta_2 + \beta_3$>0) while information which contradicts their forecast decreases their confidence ($\gamma_1 + \gamma_2 + \gamma_3$<0). This result suggests that French investors are subject to self-attribution bias. This bias leads them to become more overconfident and trade more aggressively following market gains as they make right forecasts of future stock returns.

Table 4. Relationship between trading volume and stock returns conditional on investor's forecasts

Source of μ and η	Panel A : ARMA (1.1)-EGARCH (1.1)				Panel B : ARMA (1.1)- GJR-GARCH (1.1)			
Dependent Variable	Tov_t		Vol_t		Tov_t		Vol_t	
$\beta_1+\beta_2+\beta_3$	32.776		7.479		40.408		27.036	
$\chi^2_{\beta(1)}$ (p-value)	7.562	(0.006)	1.581	(0.208)	11.072	(0.000)	4.720	(0.029)
$\chi^2_{\beta(2)}$ (p-value)	8.519	(0.036)	2.16	(0.539)	12.231	(0.006)	5.230	(0.155)
$\gamma_1+\gamma_2+\gamma_3$	-0.414		-8.6		-12.033		-9.4	
$\chi^2_{\gamma(1)}$ (p-value)	6.460	0.011)	3.669	(0.055)	7.961	(0.004)	4.626	(0.031)
$\chi^2_{\gamma(2)}$ (p-value)	7.248	(0.064)	4.141	(0.246)	9.075	(0.028)	5.4601	(0.141)
$\lambda_1+\lambda_2+\lambda_3$	-0.557		-0.373		-0.596		-0.146	
$\chi^2_{\lambda(1)}$ (p-value)	4.503	(0.033)	1.943	(0.163)	5.805	(0.016)	3.169	(0.075)
$\chi^2_{\lambda(2)}$ (p-value)	4.509	(0.211)	2.052	(0.561)	5.837	(0.119)	3.234	(0.357)
$\chi^2_{\beta\gamma}$ (p-value)	10.91	(0.000)	10.64	(0.000)	15.198	(0.007)	6.987	(0.008)

Source: (own)

This table presents the results of the following regression estimate:

$$V_t = \alpha_0 + \alpha_1|r_t| + \alpha_2 MAD_t + \alpha_3 r_t + \sum_{j=1}^{p}\beta_j\left(r_{t-j}\times I^{+}_{t-j}\right) + \sum_{j=1}^{p}\gamma_j\left(r_{t-j}\times(1-I^{+}_{t-j})\right) + \sum_{j=1}^{p}\lambda_j\left|\eta_{t-j}\right| + \varepsilon_t$$

where V_t is the market trading volume (Tov_t or Vol_t) and r_t is the market return, $|r_t|$ is the absolute value of r_t and MAD_t represents the mean absolute cross-sectional return deviation. $|\eta_t|$ is the absolute value of the unexpected return (or forecast error) at month *t* derived from GARCH (EGARCH and GJR-GARCH) specifications. The dummy variable I^{+}_{t-j} takes on a value of one if $\mu_{t-1}\times r_t f\ 0$ in which μ_{t-1} denotes the monthly expected return at month *t-1*, and zero otherwise. The $\chi^2_{\beta(1)}$, $\chi^2_{y(1)}$ et $\chi^2_{\lambda(1)}$ test statistics are used to test the null hypothesis that $\beta_1+\beta_2+\beta_3=0$, that $\gamma_1+\gamma_2+\gamma_3=0$ and that $\lambda_1+\lambda_2+\lambda_3=0$, respectively. The test statistics $\chi^2_{\beta(2)}$, $\chi^2_{y(2)}$ and $\chi^2_{\lambda(3)}$ are used to test the null hypothesis that $\beta(j)=0$ for all *j*, $\gamma_{(j)}=0$ for all *j* and $\lambda_{(j)}=0$ for all *j*, respectively. The $\chi^2_{\beta\gamma}$ test statistic is used to test the null hypothesis that $\beta_1+\beta_2+\beta_3=\gamma_1+\gamma_2+\gamma_3$. The *p-value* is the probability of obtaining the value of the corresponding test statistic or higher under the null hypothesis.

4.3. Overconfidence and volatility

Table 5 reports the results from estimating Eq. (10) and Eq. (11). It is seen that the effect of unrelated overconfidence variable on stocks volatility, measured by f_3 parameter, is significantly positive for EGARCH model. Concerning overconfidence effect on volatility, the statically significance of the estimated f_4 parameter, for the EGARCH model, associated with the rejection of null hypothesis that $f_3=f_4$ suggests that overconfidence bias contributes to the return volatility on French securities market.

Table 5. Relationship between conditional volatility and trading volume

	ARMA-EGARCH		ARMA-GJR-GARCH	
ω *(t-stat)*	-0.0531	(-0.556)	0.0357***	(19.425)
f_1 *(t-stat)*	-0.3567***	(-11.440)	0.3511***	(17.784)
κ (θ) *(t-stat)*	-0.1386	(-0.3483)	0.0459	(0.2868)
f_2 *(t-stat)*	0.8671***	(22.572)	0.8754***	(7.4517)
f_3 *(t-stat)*	0.0089**	(1.684)	0.0001	(1.380)
f_4 *(t-stat)*	0.5166***	(3.641)	0.0023	(1.136)
χ^2 *(p-value)*	13.031	(0.000)	51.576	(0.000)

Source: (own)

This table reports the results of conditional variance equation estimate of the ARMA (1.1)-EGARCH (1.1) and ARMA (1.1)-GJR-GARCH (1.1) models.

$$r_t = \mu_t + \eta_t$$

$$\eta_t / (V_t, \eta_{t-1}, \eta_{t-2}, \ldots\ldots, r_{t-1}, r_{t-2}) : \ GED(0, h_t),$$

$$Inh_t = \omega + f_1\left(\frac{|\eta_{t-1}| + \kappa\eta_{t-1}}{\sqrt{h_{t-1}}}\right) + f_2 Inh_{t-1} + f_3 NONOVER_t + f_4 OVER_t$$

and

$$r_t = \mu_t + \eta_t$$

$$\eta_t / (V_t, \eta_{t-1}, \eta_{t-2} \ldots\ldots\ldots, r_{t-1}, r_{t-2} \ldots) : \ GED(0, h_t),$$

$$h_t = \omega + f_t(\eta^2_{t-1}) + f_2 h_{t-1} + \theta S^-_{t-1}(\eta^2_{t-1}) + f_3 NONOVER_t + f_4 OVER$$

where V_t is the market trading volume (Tov_t or Vol_t) and r_t is the market return. $OVER_t$ is the component of V_t related to lagged market returns at month *t*, $NONOVER_t$ is the component of V_t unrelated to lagged market returns at month *t*,

$$V_t = \alpha + \sum_{j=1}^{p} \beta_j r_{t-j} = \left[\sum_{j=1}^{p} \beta_j r_{t-j}\right] + [\alpha + \varepsilon_t] = OVER_t + NONOVER_t$$

The χ^2 test statistic with one degree of freedom is used to test the null hypothesis that $f_3 = f_4$, and the *p-value* is the probability of obtaining the value of the χ^2 test statistic or higher under the null hypothesis.

Note: ***, **, denote significant at the 1%, and 10% levels, respectively.

5. Conclusion

There is a growing literature showing that the overconfidence bias is useful for explaining many asset pricing anomalies. Using French stock market data, this paper provides an evaluation of the overconfidence empirical implications.

The analysis of the returns impulse responses to the private and public information shocks shows that these returns overreact to private information and underreact to public information. Preceded by an initial underreaction, this overreaction is followed by a correction process reaching the stocks prices to the equilibrium. Price behavior in response to private and public information is in favor of our first hypothesis that overconfident investors overreact to private information and underreact to public information.

Granger-causality tests of stock returns and trading volume estimation show that after high returns subsequent trading volume will be higher as investment success increases the degree of overconfidence. This result is consistent with our second overconfidence hypothesis that overconfident investors trade more (less) aggressively in periods subsequent to market gains (losses).

To see whether self-attribution bias causes investors' overconfidence, we investigate the investors' reaction to market gain when they make right and wrong forecasts. Investor's forecasts of future stock returns and forecast errors are derived from two GARCH specifications that allow for asymmetric shocks to volatility. We find that when investors make right forecasts of future returns, they become overconfident and trade more in subsequent time periods. On the other hand, when they make wrong forecasts, their overconfidence may fall modestly. This finding provides empirical evidence in support of our third hypothesis that self-attribution bias, conditioned by right investors' forecasts, increases their overconfidence and their trading volume.

Finally, we study the relation between excessive trading volume of overconfidence investors and excessive prices volatility. The trading volume is decomposed into a first variables related to overconfidence and a second variable unrelated to investors' overconfidence. The analysis of the relation between return volatility and these two variables shows that conditional volatility is positively related to trading volume caused by overconfidence bias. This result is in favor of overconfidence contribution to prices excessive volatility.

Generally, our results provide strong statistical support to the presence of overconfidence bias among investors in French stocks market. This psychological bias constitutes a confirmed explanation of the most stylized market anomalies.

6. References

[1] Akaike, H., (1974), *A new look at the statistical identification model*, in: *Transactions on Automatic Control*, Vol. 19, pp. 716-723.

[2] Albulescu, C.T., (2008), *The Impact of the New Financial Products on the Volatility of the Economic Growth,* in: *Journal of Applied Economic Sciences,* Vol. 2, No. (2), pp.12-19.

[3] Benos, A.V., (1998), *Aggressiveness and survival of overconfident traders*, in: *Journal of Financial Markets*, No. 1, pp. 353-383.

[4] Bessembinder, H., Chan, K., Seguin, P. J., (1996), *An empirical examination of information*, *differences of opinion, and trading activity*, in: *Journal of Financial Economics*, No. 40, pp. 105-134.

[5] Campblell, J.Y., Grossman, S.J., Wang J., (1993), *Trading volume and serial correlation in stock returns*. The Quarterly Journal of Economy, No. 107, pp. 205-251.

[6] Chuang, W., Lee, B.S., (2006), *An empirical evaluation of the overconfidence hypothesis*, in: *Journal on Banking & Finance*, Vol. 30, No. 9, pp. 2489-2511.

[7] Chordia, T., Swaminathan, B., (2000), *Trading volume and cross-autocorrelations in stock returns*, in: *Journal of Finance*, Vol. 55, pp. 913–935.

[8] Daniel, K., Hirshleifer, D., Subrahmanyam, A., (1998), *Investor psychology and security market under- and overreactions*, in: *Journal of Finance*, Vol. 53, No. 6, pp. 1839-1886.

[9] Daniel, K., Hirshleifer, D., Subrahmanyam, A., (2001), *Overconfidence, arbitrage, and equilibrium asset pricing*, in: *Journal of Finance*, Vol. 56, pp.921–965.

[10] Gervais, S., Odean, T., (2001), *Learning to be overconfident*. Review of Financial Studies, Vol. 14, No. 1, pp. 1-27.

[11] Glosten, L.R., Jagannathan, R., Runkle, D., (1993), *On the relation between the expected value and the volatility of the nominal excess return on stocks*, in: *Journal of Finance*, Vol. 48 No. 5, pp. 1779-1801.

[12] Granger, C.W.J., (1969), *Investigating causal relations by econometric models and cross-spectral methods*, in: *Econometrica*, Vol. 37, No. 3, pp. 424-438.

[13] Granger, C.W.J., (1988), *Some recent developments in a concept of causality*, Journal of Econometrics, Vol. 39, No. 1, pp.199-211.

[14] Hirshleifer, D., Luo, G.Y., (2001), *On the survival of overconfident traders in a competitive securities market*, in: *Journal of Financial Markets*, Vol. 4, No. 1, pp. 73-84.

[15] Hodrick, R.J., Prescott, E.C., (1997), Postwar U.S. *Business Cycles: An Empirical Investigation*., in: *Journal of Money, Credit, and Banking*, Vol. 29, No. 1 pp. 1-16.

[16] Karpoff, J.M., (1987), *The relation between price changes and trading volume: A survey*, in: *Journal of Financial and Quantitative Analysis*, Vol. 22, No. 1, pp. 109-126.

[17] Kyle, A.S., Wang, F.A., (1997), *Speculation duopoly with agreement to disagree: Can overconfidence survive the market test?,* in: *Journal of Finance*, Vol 52, pp.2073–2090.

[18] Lamoureux, C.G., Lastrapes, W.D., (1990), *Heteroskedasticity in stock return data: Volume versus GARCH effects*, in: *Journal of Finance*, Vol. 45, No. 1, pp. 221-229.

[19] Nelson, D.B., (1991), *Conditional heteroskedasticity in asset returns: A new approach*, in: *Econometrica*, Vol. 59, No. 2, pp. 347-370.

[20] Odean, T., (1998), *Volume, volatility, price, and profit when all traders are above average*, in: *Journal of Finance*, Vol. 53, No. 6, pp. 1887-1934.

[21] Ross, S., (1989), *Information and volatility: The no-arbitrage martingale approach to timing and resolution irrelevancy*, Journal of Finance, Vol. 44, pp. 1–17.

[22] Schwert, G.W., (1989), *Why does stock market volatility change over time?* in: *Journal of Finance*, Vol. 44, pp. 1115–1153.

[23] Statman, M., Thorley, S., Vorkink, K., (2003), *Investor overconfidence and trading volume*, Working Paper, Santa Clara University.

[24] Wang, F.A., (2001), *Overconfidence, investor sentiment, and evolution*, in: *Journal of Financial Intermediation*, Vol. 10, pp.138–170.

US PRICE INDICES AND THE EXCHANGE RATE: ARE RECREATIONAL PRODUCTS DIFFERENT?

Roberto **CELLINI**, Alessandro **PAOLINO**
University of Catania
Faculty of Economics, **Italy**
cellini@unict.it, michelangelo.paolino@tiscali.it

Abstract

The paper analyses the cointegration relationships and the causal links between the exchange rate of the US Dollar, on the one side, and different price indices of US products on the other side. Data are of monthly frequency and cover a period of two or three decades. We show that the exchange rate cointegrate with the Consumer Price Index and with the prices indices of several agricultural, manufactured and service goods; moreover a one-direction causal link is present, running from price to exchange rate. On the opposite, cointegrating relationships between exchange rate and price indices do not exist in the case of recreational products with "cultural" content. Tentative theoretical explanations are proposed.

Keywords: price index; exchange rate, cointegration, causality.

JEL Classification: Z10, C22, C32, F13.

1. Introduction

The relationship between exchange rate, on the one side, and price level on the other side, is an evergreen in theoretical and applied economics. Different theories predict different links between these variables. A lot of empirical investigation on the relationship between prices and exchange rates is also available, with very mixed results, depending on the considered data and methods.

Roughly speaking, two empirical approaches have been developed. The first one involves the construction and estimation of "large" models, in which several relationships connect these variables with other factors. At least in principle, this allows a detailed explanation of several features of the movements of price and exchange rate. However, the consensus on the relevant links is not unanimous. Several researchers have therefore developed an alternative data-oriented approach that favours simple time-series modelling aimed at determining the most relevant economic influences in an "atheoretical" framework; such an approach can be useful in providing empirical answers to "naive" questions concerning the links between price and exchange rate.

This second route is followed by the present paper. In particular, we take a cointegration analysis approach in order to study the pattern of time series. As it is well known, cointegration analysis allows to distinguish the long-run relationships among variables, disentangling them from the short-run components of movements. Cointegration techniques have been already used also in analysing price and exchange rates intensely, providing –also in this case– mixed results [see Edison, (1987), Rogoff, (1996), and Engle, (2000), among many others].

The novelty of this paper rests in documenting a striking evidence based on US data: a long run cointegrating relationship holds between the (nominal effective) US Dollar exchange rate on the one side, and most prices indices on the other side, including the Consumer Price Index, and all considered (nominal) prices indices referred to agricultural and industrial goods, as well as to some products belonging to the service sector. In any case, the causal link runs in one direction, from price to exchange rate.

The striking evidence is represented by the fact that such a cointegrating relationship does not exist, on the opposite, for any price index referred to recreational products. The recreational goods and services under investigation can be seen as cultural (or popular-cultural) products. We provide a tentative explanation for the absence of cointegrating relationship between recreational products and the exchange rate; the explanation is based on the public intervention in these markets, often aimed at protecting the *cultural diversity*.

The paper is organized as follows. Section 2 presents the data; Section 3 explains the method used and shows the results; Section 4 comments and concludes.

2. Data

2.1 Sources and description.

We analyze US data with a monthly frequency over the period 1980-2006.

As for the exchange rate, we consider the nominal effective exchange rate of US dollar provided by the St. Louis Federal Reserve Bank; the value expresses the amount of US dollar equivalent to one unit of foreign currency, so that a devaluation of US dollar corresponds to an increase of the considered index.[1]

As for price indices we rely on the data provided by the Bureau of Labor Statistics (Consumer Price Index by Item and Place).[2]

This paper presents the results concerning 12 different prices indices referred to specific goods' categories, along with the Consumer Price Index. The considered indices are listed in Table 1. Four goods' categories represent tradable goods (apparel; alcoholic beverage; fuels; energy); four products' categories represent services (medical care; professional services; transportation; private transportation); four categories represent "cultural" products (recreation; television; educational books; school tuition). However, the same exercise has been carried out over further goods' categories (whose data are available from the mentioned sources), with analogous outcomes (results are available upon request).

Table 1. List of variables

LNEER = log of nominal effective exchange rate
LCPI= log of consumer price index
LPDRESS= log of price of apparel
LPALCB = log of price of alcoholic beverages
LPFUEL = log of price of fuels
LPENERG= log of price of energy
LPMEDS= log of price of medical care services
LPPROS= log of price of professional services
LPTRA= log of price of transportation
LPPRTRA= log of price of private transportation
LPEDBOOK= log of price of educational books and supplies
LPRECR = log of price of recreation
LPCTV = log of price of cable television
LPELHIS = log of price of elementary and high school tuition and fees

We chose to analyse seasonally adjusted data –that are provided by the sources, along with the original data. We are aware that our choice of working with season-adjusted data could be questionable, but our primary interest is to focus on long-run links, and the adjusted data allow to avoid the problem of modelling the seasonal components, and permit to have a larger degree of freedom for the long-run analysis we are interested in.

2.2 Integration Analysis

Widespread agreement exists about the I(1) nature of price and exchange rate. Minor doubts have been cast only on consumer price index, sometimes suspected to be I(2), other times suspected to be stationary.[3]

---

[1] More precisely, the Trade Weighted Exchange Rate Index - Broad is considered (with January 1997 taking value 100).

[2] All data are free downloadable through the website www.economagic.com.

[3] The point of the supporters of the I(2) hypothesis is based on the evidence that the inflation rate is integrated of order 1. However, a large body of empirical evidence suggests that inflation is stationary in the presence of structural breaks. This evidence is documented, for instance, by several working paper of the European Central Bank published in 2004 and 2005 – see, e.g., Dossche and Everaet (2005) or Lunnemann and Matha (2004). On the opposite, the point of the supporters of the I(0) hypothesis is based on the fact that price series can appear to be stationary in the presence of appropriate deterministic structural breaks. The stationarity around a broken deterministic trend is supported, *inter alia*, by Lippi and Reichlin (1994).

As far as the data at hand is concerned, ADF tests on the series and their first difference lead to the conclusion that all the series are I(1), provided that a linear trend is present and in the absence of structural breaks – see Table 2. (The only variable for which some doubt can emerge is the series of the price of educational books, but the following analysis provides results consistent with its I(1) nature.)

Table 2. Dickey Fuller tests for unit root

	DF	ADF_1		DF	ADF_1
LNEER	-1.077	-1.586	Δ LNEER	-11.908	-10.326
LCPI	-3.004	-2.382	Δ LCPI	-11.132	-10.556
LPDRESS	-0.425	-0.412	Δ LPDRESS	-16.413	-12.085
LPALCB	-0.941	-1.151	Δ LPALCB	-13.963	-11.860
LPFUEL	-0.628	-1.310	Δ LPFUEL	-13.514	-10.234
LPENERG	-1.118	-2.106	Δ LPENERG	-12.167	-13.585
LPMEDS	-4.51	-2.58	Δ LPMEDS	-6.940	-4.450
LPPROS	-3.52	-3.24	Δ LPPROS	-10.154	-6.496
LPPRTRA	-3.010	-3.475	Δ LPPTRA	-12.086	-14.204
LPTRA	-3.010	-3.52	Δ LPTRA	-12.039	-13.986
LPEDBOOK	-3.736	-3.806	Δ LPEDBOOK	-17.379	-15.651
LPRECR	-1.686	-1.561	Δ LPRECR	-14.668	-8.819
LPCTV	-1.360	-1.852	Δ LPCTV	-10.223	-8.877
LPELHIS	-2.543	-2.850	ΔLPELHIS	-24.874	-16.226

Notes: Δ denotes the first-difference operator. ADF tests are reported. For variables in level, the trended case is considered (the 5% critical value is: -3.426) For variable in first-difference the nontrended case is considered (the 5% critical value is: -2.871)

By the way, it is interesting to report that even the real prices of the considered items appear to be non-stationary, according to the ADF tests; it means that the (log) nominal prices of items do not cointegrate with the (log of) CPI or they cointegrate under a vector different from (-1,1).[4] Under this respect, recreational goods have no specificity as compared to other goods –at least within the set of goods we are dealing with: the real price of any considered good is non stationary.

The question we ask is whether some cointegrating relationship between prices and exchange rate holds.

3. Cointegration and causality analysis

Cointegration means that there exists a stationary linear combination between two (or more) non-stationary series; the linear combination can be interpreted as the long-run link between the non-stationary series. In the case of two non-stationary time series X and Y, there is at most one stationary linear combination series. (Extensions allow for more than two series and possibly more than one cointegrating relationship; in this paper however we confine to the case of one cointegration relationship between two series.)

Consider the static equation

$$Y_t = a + bX_t + u_t \quad , \qquad t=1,2,...T . \qquad (1)$$

If the error term u_t is a stationary process, then X and Y are cointegrated [Engle and Granger, (1987)]. The appropriate test to evaluate the stationary nature of residuals is the (Augmented) Dickey

[4] The cointegration under a vector different from (-1,1) occurs for three items, namely, medical services, transportations, and private transportation; the no-cointegration in the remaining nine cases.

Fuller test.[5] The regression residuals can be interpreted as the "error" of current variable *Y* with respect to its long-run equilibrium value dictated by the cointegrating relationship. According to the representation theorem by Granger, if two integrated variables cointegrate, an error correction mechanism is operative, which means that *Y* and/or *X* have to move in order to correct the disequilibrium with respect to the long-run relationship.

The cointegration analysis offers powerful tools to look at the causality issue. The representation theorem [Engle and Granger, (1987)] states, loosely speaking, that a cointegration relationship can be represented as a model with error correction mechanisms which entails (at least) one Granger causal ordering.

Phillips and Loretan (1992)[6] suggest to look at the significance of the error-correction term directly in order to assess the existence of long run (cointegrating) relationships. This procedure allows to assess the direction of the causality, in the cointegrating relationship.

Specifically, let us consider the following system representing the dynamics of the cointegrated variables *X* and *Y*, where Δ is the first-difference operator and EC denotes the error correction term, i.e., the fitted residuals of the static long-run regression corresponding to eq. (1):

$$\Delta Y_t = \beta + \alpha EC_{t-1} + \sum_{i=1} \alpha_i \Delta Y_{t-1} + \sum_{j=1} \lambda_j \Delta X_{t-j} + \varepsilon_t \tag{2a}$$

$$\Delta X_t = \phi + \gamma EC_{t-1} + \sum_{i=1} \gamma_i \Delta X_{t-1} + \sum_{j=1} \eta_j \Delta Y_{t-j} + \varepsilon'_t \tag{2b}$$

There are different concepts of causality, with respect to the system of Eqs. (2a,b), as concerns either the level or the first difference of variables.

On the one hand, short-run Granger-causality refers to the (stationary) variables ΔY and ΔX. In particular, ΔX is said to be weakly exogenous for the parameters of the regression (2b) if γ is not significantly different from zero. If also coefficients η_j are not significantly different from zero, then there is no Granger-causal link from ΔY to ΔX. In such a case, ΔX is *strongly* exogenous and (2a) can be used for prediction purposes. Obviously the same holds, *mutatis mutandis*, for ΔY, which is weakly exogenous if $\alpha = 0$, and strongly exogenous in the case where also coefficients λ_j are not statistically different from zero.

On the other hand, long-run Granger-causality may also be studied: it refers to the links between the *levels* of *Y* and *X*. Under this perspective, Granger and Lin (1995) propose a measure of the strength of causality between (cointegrated) variables; in particular, the strength of causality of *Y* upon *X* can be approximated by:

$$M_{Y \to X} = \log\left[1 + \frac{\gamma^2 (1-\rho^2)}{(\gamma\rho - \alpha)^2}\right] \quad , \rho = corr(\varepsilon, \varepsilon') \tag{2c}$$

A particular case of such a measure index is the case in which either α or γ are not statistically significant.

Clearly, if α is not different from zero, then $M_{X \to Y} = 0$, i.e., *X* does not cause *Y*. If γ is not different from zero, $M_{Y \to X} = 0$, which means that *Y* does not cause *X*. If simultaneously $\gamma = \alpha = 0$ no error correction mechanism is operative and variables do not cointegrate.

With reference of data at hand, all the mentioned procedures (Engle-Granger, Johansen, Phillips-Loretan) give substantially identical results as concerns the cointegration relationship among variables. For the sake of brevity we report only the procedure *à la* Phillips - Loretan (1992)

---

[5] An alternative method to evaluate the cointegration among variables is the one proposed by Johansen, relying on LR Test Based on Maximal Eigenvalue of the Stochastic Matrix.

[6] See also Kremers *et al*. (1992) and Inder (1993).

For the estimation purpose, we chose to regress the following equations, where *Y* indicates the exchange rate and *X* indicates a price index, in the cases examined in this present paper:

$$\Delta Y_t = \beta + \alpha Y_{t-1} + \alpha_0 X_{t-1} + \sum_{i=1} \alpha_i \Delta Y_{t-1} + \sum_{j=1} \lambda_j \Delta X_{t-j} + \varepsilon^Y{}_t \tag{3a}$$

$$\Delta X_t = \phi + \gamma X_{t-1} + \gamma_0 Y_{t-1} + \sum_{i=1} \gamma_i \Delta X_{t-1} + \sum_{j=1} \eta_j \Delta Y_{t-j} + \varepsilon^X{}_t \tag{3b}$$

Table 4 reports the coefficients' estimates for each pair of variables, corresponding to our preferred specification, in accordance with the significance of the terms of the lag-polinomials of ΔX and ΔY. In particular, we considered the presence of 1st, 2nd, and 12th lag. However, the second and the 12th lag terms appeared to be never significant.

Of course, *Y* and *X* cointegrate if the error correction mechanism is operative. In order to have a stable adjustment process of the variables towards their long-run levels, α and/or γ must lie in the interval (-1, 0). To this end, three possibilities exist: (a) α is included in the interval (-1,0) and it is statistical significant (in this case, *Y* moves in order to adjust the error; if γ is not statistically significant, a one-direction causal link in the long run rums from *X* to *Y*); (b) γ is included in the interval (-1,0) and it is statistical significant (in this case, *X* moves in order to adjust the error; if α is not statistically significant, a one-direction causal link in the long run rums from *Y* to *X*); (c) both α and γ are significant and negative parameters included in the interval (-1,0): this means that both variables react to error. If neither α nor γ are included in the interval (-1,0), *X* and *Y* do not cointegrate since the error (that is, the deviation from the long run equilibrium relationship) is not corrected. Eventually, if no error correction mechanism is operative, integrated variables do not cointegrate.

In our analysis we consider the significance at the 15% level (if not differently specified). From Table 3 it is immediate to see what follows.

i) Parameter α is negative and significant, in nine cases: in the case of the general CPI, and in the eight cases of non-cultural products (both the four tradable items, dress, alcoholic beverages, fuel and energy, and the four services, medical services, professional services, transportations and private transportations); in eight out of these nine cases, parameter γ is zero, denoting that the causal link runs from price to exchange rate (the exception is the private transport service, where the causal link is two-way). Only one out of these nine cases provides an estimate of η not different from zero (alcoholic beverages) meaning strong exogeneity; in the remaining cases, η is statistically different from zero, meaning short-run links and leading to reject the strong exogeneity.

(ii) Parameter α is not statistically different from zero and simultaneously also parameter γ is not statistically different from zero, in each of the four cases pertaining to cultural items, educational books, recreation, cable TV, school tuition fees). This means that no error correction mechanism is operative (in other words, no cointegration links do exist) for such products. Each of these goods is of "cultural nature".

Table 3. Long-run causality and short-run dynamics

Y = LNEER X = LP*	α (ECM coef)	β	α_0	α_1	λ_1	Regression Statistics	γ (ECM coef)	ϕ	γ_0	γ_1	η	Regression statistics
LCPI	−0.0131	0.0031	0.0115	0.3480	−0.0138	R2=0.16	0.0016	0.0067	−0.0030	0.4097	0.0074	R2=0.34
	(−1.702)	(0.137)	(1.065)	(6.584)	(−0.049)	DW=1.94	(0.811)	(1.602)	(−2.091)	(7.900)	(0.756)	DW=1.91
	[0.090]	[0.891]	[0.288]	[0.000]	[0.961]	h=1.63	[0.418]	[0.110]	[0.037]	[0.000]	[0.450]	h=2.13
LPDRESS	−0.0118	−0.0425	0.0203	0.3366	−0.2302	R2=0.17	0.0026	0.0121	−0.0054	−0.0371	0.0107	R2=0.12
	(−3.471)	(−1.158)	(2.032)	(6.440)	(−1.538)	DW=1.94	(0.684)	(0.874)	(−4.162)	(−0.655)	(0.540)	DW=2.00
	[0.001]	[0.248]	[0.043]	[0.000]	[0.125]	h=1.52	[0.495]	[0.383]	[0.000]	[0.513]	[0.590]	h=ND
LPALCB	−0.0125	0.0047	0.0106	0.3472	−0.0184	R2=0.16	0.0005	0.0089	−0.0021	0.2190	0.0291	R2=0.10
	(−1.792)	(0.216)	(1.069)	(6.553)	(−0.104)	DW=1.94	(0.148)	(1.323)	(−0.956)	(3.970)	(1.760)	DW=1.95
	[0.074]	[0.829]	[0.286]	[0.000]	[0.917]	h=1.69	[0.883]	[0.187]	[0.340]	[0.000]	[0.079]	h=2.98
LPFUEL	−0.0053	0.0237	0.00036	0.3431	0.0566	R2=0.16	−0.0017	0.0165	−0.0015	0.2841	−0.0322	R2=0.09
	(−1.503)	(1.054)	(0.050)	(6.515)	(0.883)	DW=1.94	(−0.272)	(0.853)	(−0.488)	(5.1424)	(−0.710)	DW=2.01
	[0.134]	[0.293]	[0.960]	[0.000]	[0.378]	h=1.55	[0.786]	[0.394]	[0.626]	[0.000]	[0.478]	h=−.75
PLENERG	−0.0054	0.2521	0.00015	0.3412	0.0004	R2=0.15	−0.0096	0.0273	0.0044	0.3520	0.0132	R2=0.12
	(−2.120)	(1.496)	(0.033)	(6.468)	(0.011)	DW=1.94	(−1.214)	(0.956)	(1.013)	(6.426)	(0.148)	DW=1.80
	[0.035]	[0.136]	[0.974]	[0.000]	[0.991]	h=1.72	[0.226]	[0.340]	[0.312]	[0.000]	[0.882]	h=9.13
LPMEDS	-0.012	0.030	0.005	0.347	-0.147	R2=0.157	0.0001	0.013	-0.003	0.464	0.001	R2=0.59
	(-1.598)	(2.251)	(0.899)	(6.555)	(--0.360)	DW=1.94	(0.818)	(7.932)	(-3.812)	(9.378)	(0.138)	DW=2.24
	[0.111]	[0.025]	[0.369]	[0.000]	[0.719]	h=1.61	[0.414]	[0.000]	[0.002]	[0.000]	[0.890]	h=-4.55
LPPROS	-0.012	0.022	0.007	0.348	-0.011	R2=0.157	-00004	0.022	-0.004	0.052	-0.002	R2=0.528
	(-1.577)	(1.534)	(0.953)	(6.566)	(-0.024)	DW=1.94	(-0.430)	(12.68)	(-3.867)	(0.936)	(-0.275)	DW=2.03
	[0.116]	[0.126]	[0.341]	[0.000]	[0.980]	h=1.67	[0.667]	[0.000]	[0.000]	[0.350]	[0.783]	h=-1.75
LPTRA	-0.013	-0.011	0.015	0.346	0.010	R2=0.160	-0.010	0.035	0.004	0.369	0.016	R2=0.140
	(-2.134)	(-0.400)	(1.355)	(6.589)	(0.140)	DW=1.94	(-1.359)	(1.749)	(0.891)	(6.785)	(0.436)	DW=1.77
	[0.034]	[0.689]	[0.176]	[0.000]	[0.888]	h=1.58	[0.175]	[0.081]	[0.373]	[0.000]	[0.663]	h=9.11
LPRRTRA	-0.012	-0.10	0.014	0.345	0.007	R2=0.159	-0.012	0.039	0.005	0.373	0.014	R2=0.140
	(-2.09)	(-0.343)	(1.259)	(6.756)	(0.106)	DW=1.94	(-1.475)	(1.777)	(1.047)	(6.911)	(0.339)	DW=1.77
	[0.037]	[0.732]	[0.209]	[0.000]	[0.915]	h=1.59	[0.141]	[0.076]	[0.298]	[0.000]	[0.734]	h=7.90
LPDBOOK	−0.0064	0.0251	0.00094	0.3438	0.0372	R2=0.16	−0.0031	0.0223	−0.00017	−0.0372	0.0153	R2=0.05
	(−0.912)	(2.791)	(0.176)	(6.419)	(0.355)	DW=1.93	(−1.066)	(4.628)	(−0.045)	(−0.661)	(0.532)	DW=2.02
	[0.363]	[0.006]	[0.860]	[0.000]	[0.723]	h=2.03	[0.287]	[0.000]	[0.964]	[0.509]	[0.595]	h=ND
LPRECR	−0.0095	0.1066	−0.0132	0.3076	0.2191	R2=0.15	−0.0041	0.0297	−0.0020	−0.1999	0.01972	R2=0.10
	(−0.823)	(1.396)	(−0.542)	(4.039)	(0.523)	DW=1.90	(−0.909)	(2.114)	(−0.932)	(−2.598)	(1.409)	DW=2.00
	[0.411]	[0.165]	[0.589]	[0.000]	[0.602]	h=3.03	[0.365]	[0.036]	[0.353]	[0.010]	[0.161]	h=.003
LPCTV	−0.0069	0.0681	−0.0062	0.3300	−0.1605	R2=0.18	−0.0025	−0.0034	0.0043	0.1677	0.0474	R2=0.05
	(−0.774)	(2.613)	(−0.915)	(4.570)	(−1.050)	DW=1.89	(−0.757)	(−0.267)	(0.991)	(2.238)	(1.338)	DW=1.94
	[0.440]	[0.010]	[0.361]	[0.000]	[0.295]	h=2.56	[0.450]	[0.790]	[0.323]	[0.026]	[0.183]	h=3.52
LPELHIS	−0.0079	0.0261	0.0019	0.346	0.0874	R2=0.16	−0.0012	0.0301	−0.0034	−0.3943	−0.0160	R2=0.19
	(−1.134)	(2.487)	(0.429)	(6.507)	(0.800)	DW=1.94	(−0.577)	(6.040)	(−1.025)	(−7.612)	(−0.634)	DW=2.17
	[0.258]	[0.013]	[0.668]	[0.000]	[0.424]	h=1.87	[0.564]	[0.000]	[0.306]	[0.000]	[0.527]	h=−4.08

Notes: The Table reports the estimates of parameters of eqs. (3a) and (3b).t-stat. in parenthesis, p-vaule in squared brackets.

A qualitatively similar evidence is provided by the Granger-Lin index of long-run causality (see Table 4): in the cases of non cultural goods and services the strength of causality running form price to exchange rate is much higher that the strength of the opposite link; in the case of cultural products, the strength of causality is low in both directions.

Table 4. The Granger-Lin *M* statistics

	$M_{x \to y} \equiv M_{p \to e}$	$M_{y \to x} \equiv M_{e \to p}$
LCPI	8.891	0.014
LPDRESS, LPALCB, LPFUEL, LPENERG (average)	2.438	0.346
LPMEDS, LPPROS, LPTRA, LPPRTRA (average)	3.021	0.386
LPDBOOK, LPRECR, LPCTV, LPELHIS (average)	1.588	0.384

It is also interesting to note that the long-run elasticity of (nominal effective) exchange rate to consumer price index is equal to -$\alpha_0 / \alpha =$0.878 and a Wald test on the hypothesis $\alpha_0 = -\alpha$ can not reject this null ($\chi^2 = 0.1605$, p=0.689). This evidence is of course consistent with the PPP theory, provided that the foreign inflation is zero. A similar conclusion holds fore three out the eight non-cultural products under consideration.

4. Discussion, policy implications and conclusions

The links between price and exchange rate can be of different nature. On the one hand –just to mention possible links– exchange rate may affect production cost, as long as imported inputs are used. On the other hand, exchange rate has to move in order to make the law of "one price" (and the purchase power parity law) fulfilled.

A lot of empirical work is available about these themes. Four streams of recent (and relevant to our purpose) empirical literature can be listed.

a. The investigation about the property of the real exchange rate, as defined by $RER = E \cdot P^f / P^d$, where *E* denotes the (bilateral or effective, according to the different analyses) nominal exchange rate, P^f the foreign price level and P^d the domestic price. While *E*, P^f and P^d are generally accepted to be I(1), different conclusions emerge as concerns *RER*: according to a part of the available literature, *RER* are not stationary [see Rogoff (1996) just to mention a paper with a comprehensive review]; the evidence, however is challenged by different results [Abuaf and Jorion, (1990), Frankel, (1986), (1990), Glen, (1992), and so on]; Mussa (1986) provides a discussion about the pros and cons of different procedures and choices about (time and country) sample. In any case, a non stationary nature of RER can not be consistent with the idea that exchange rate move in order to assure the purchasing power parity. Indeed, PPP would imply a constant value for RER. Tests on the PPP theory basing on cointegration property among time series are presented, among other, by Edison (1987), Engel (2000) and Breitung and Candelon (2005).

b. The investigation about the quantitative importance of different reasons for the (possible) failure of the PPP: let us think of the presence of non-tradable goods, the behaviour of the relative price of tradable vs. non-tradable goods, volatile exchange rate, sticky price of goods, and so on. In this vein of literature we have to mention the recent investigation of Betts - Kehoe (2006) and Burstein *et al.* (2006) who explain the failure of PPP on the basis of movements of relative price of non-tradable vs. tradable goods in single countries, and the analyses of Giovannini (1988) or Chari *et al.* (2002) who find –on the opposite– that differences of price of traded goods can provide a quite exhaustive explanation of the failure.

c. The investigation about the (possible) failure of the law of one price in reference with a specific good; within this stream one can mention Asplund and Friberg (2001), Knetter (1989, 1993), Isard (1977), among many others. All these studies document relevant failures of the one price law, even in the case of homogenous tradable goods.

d. The investigation about the effect of the adhesion to a monetary union upon price dispersion; notably, the experience of the European Monetary Union provides an enormous experimental basis: Baye *et al.* (2002), Engle and Rogers (2004), Lutz (2002), Rogers (2002) provide evidence about the fact that a certain degree of *reduction* in the prices' dispersion (for specific goods) has been occurring across the EMU countries over the first half of the Nineties, but this process has stopped thereafter, so that the adoption a common money does not appear to enhance the law of one price across countries.

We mentioned such streams of applied economic literature, since our present analysis has something to do with these investigations, even if we have taken a much more restricted perspective. In fact, we have aimed at analyzing the direction of the causal links between exchange rate (on the one side) and price (on the other side), in a basket of goods.

Our main interest has been to assess whether recreational services (i.e., a subset of goods and services belonging to culture, in a broad sense, or "popular culture" in a more restricted sense) show any specificity. Our answer has been positive: while traded goods appear to be linked to exchange rate through a cointegration relationship, and while such a cointegrating links also holds for the considered services, such a link does not emerge in reference to each of the considered recreational (or cultural) products.

We have found a sort of "law"- which holds for all the considered non-cultural products: price and exchange rate are linked by cointegrationg relationship, and the causality runs from price to exchange rate. This law does not hold for the considered cultural products, whose price indices emerge to be not cointegrated with the exchange rate.

We are aware that some cautionary notes are necessary; just to mention the most relevant ones: firstly, we have limited our attention to linear dynamics, and non-linear adjustment behaviours are not taken into consideration. Secondly, we have limited ourselves to examine the possible existence of cointegration relations, without having provided deep analyses of the signs of the long-run parameter estimates. Thirdly, though representative, the choice of the products under consideration is partial. Nevertheless, the specificity of the long-run pattern of the price of cultural goods and service with respect to different goods is clear, and it deserve some explanations.

It is true that the international markets of cultural goods and services are probably less integrated in comparison to the international markets of agricultural and manufactured goods. In effect, even if the trade of cultural goods and services doesn't find a different treatment in international trade agreements (GATT and GATS), important restrictions are documented by several studies (see, among many others, Guevremont 2006). However, this point has little to do with the non-traded nature of these products: we have found that the cointegrating links hold also for (little tradable) services, like medical care services, or professional services.

It is also true that the public intervention in the considered cultural market is probably heavy. However, the explanation of the exception can not rest only on this fact, provided that public intervention is equally heavy in other sectors considered here, like transportations or medical services.

The cultural goods and services are specific because they have an idiosyncratic content, along with the economic and commercial value. They convey identities, value and meanings and play a special role in a community. Thus, according to some views, culture can not be left to the uncertain tastes of the invisible hand, and the government has to play a role: free trade could be a threat to cultural diversity and national identity; hence, policy has to protect and promote cultural diversity. But, under this respect, cultural policies could represent a form of protectionism.

Cultural protectionism finds its legitimacy also in the concepts of *cultural exception* and *cultural diversity*. On 20th October 2005, the UNESCO General Conference approved the *Convention on the Protectionism and Promotion of the Diversity of Cultural Expressions* that reinforces the notions enshrined in the *UNESCO Universal Declaration on Cultural Diversity* (2001) that cultural diversity is a *common heritage of humanity* and that its defence must be considered *an ethical imperative, inseparable from respect of human dignity*.

Every country in the world treats at least some aspect of its domestic cultural life as a public goods, even if the range of activities receiving government support varies widely across countries and regions. Several countries, like Canada and France, subsidize own production and exportation of cultural goods and services clearly adopting a form of cultural protectionism. In some European countries, including Italy, fix price arrangements on cultural goods (notably, books) are permitted (or were permitted during part of time-sample we have analysed), contrary to the US, and so on.

Under this respect, active cultural policy – enhancing goods' differentiation– represents an obstacle to arbitrage force in markets, thus contributing to the failure of the one-price law in the markets of goods with cultural content, and eventually to the PPP principle. This holds irrespective of the tradable nature of products involved.

5. References

[1] Abuaf, N. and Jorion, P., (1990), *Purchasing Power Parity in the Long Run*, in: *Journal of Finance*, 45(1): 157-74.

[2] Asplund, M. and Friberg, R., (2001), *The law of one price in Scandinavian duty-free stores*, in: *American Economic Review*, 91(6): 1072-1083.

[3] Baye, M. R., Gatti, R., Kattuman, P. and Morgan, J., (2002), *Online pricing and the euro changeover: Cross-country comparisons*, working paper, Center for Financial Studies, University of Frankfurt/Main.

[4] Betts, C. and Kehoe, T., (2006), *U.S. real Exchange rate fluctuation and relative price fluctuation*, in: *Journal of Monetary Economics*, 53: 1297-1326.

[5] Breitung, J. and Candelon, B. (2005), *Purchasing Power Parity during Currency Crises: A Panel Unit Root Test under Structural Breaks*, in: *Review of World Economics*, 141: 124-140.

[6] Burstein, A., Eichenbaum, M. and Rebelo, S., (2006), *The importance of nontradable goods' price in cyclical real exchange rate fluctuations*, Japan and the World Economy, 18(2): 247-53.

[7] Dossche, A. and Everaet, G., (2005), *Measuring Inflation Persistence. A Structural Time Series Approach*, ECB working paper, No. 495.

[8] Edison, H., (1987), *Purchasing Power Parity in the Long Run: A Test of the Dollar/Pound Exchange Rate (1890-1978),* in: *Journal of Money, Credit, Banking*, 19(3): 376-387.

[9] Engel, C., (2000), *Local-currency pricing and the choice of exchange-rate regime*, in: *European Economic Review*, 44: 1449-1172.

[10] Engle, R.F. and Granger, C., (1987), *Co-integration and Error-Correction: Representation, Estimation and Testing*, in: *Econometrica*, 55: 251-276.

[11] Engel, C. and Rogers, J.H., (1996), *How wide is the border?,* in: *American Economic Review*, 86: 1112-1125.

[12] Engel, C. and Rogers, J.H., (2001), *Deviations from the purchasing power parity: Causes and welfare costs*, in: *Journal of International Economics*, 55: 1243-1272.

[13] Engel, C. and Rogers, J.H., (2004), *European product market integration after the euro*, in: *Economic Policy*, 19: 347-384.

[14] Frankel, J.A., (1986), *International Capital Mobility and Crowding-out in the U.S. Economy: Imperfect Integration of Financial Markets or Goods Market?,* in Hafer R. W. (ed.), *How open is the U.S. economy?*, Lexington: Lexington Books, pp. 33-67.

[15] Frankel, J.A., (1990), *Zen and the Art of Modern Macroeconomics: A Commentary*, in Haraf W. S. – Willet T. D. (a cura di), *Monetary policy for a volatile global economy*, Washington DC, American Entreprise Institute for Public Policy Research, pp. 117-123.

[16] Giovannini, A., (1988), *Exchange rates and Traded Goods Prices*, in: *Journal of International Economics*, 24(1/2): 45-68.

[17] Glen, J.D., (1992), *Real Exchange Rates in the Short, Medium, and Long Run*, in: *Journal of International Economics*, 33(1/2): 147-166.

[18] Granger, C.W.J. and Lin, J.-L., (1995), *Causality in the Long Run*, in: *Econometric Theory*, 11: 530-36.

[19] Gregory, A.W. and Hansen, B.E., (1996), *Residual-Based Tests for Cointegration in Models with Regime Shifts*, in: *Journal of Econometrics*, 70: 99-126.

[20] Guèvremont, V. (2006), *Treatment of cultural goods and services in international trade agreements*, Mimeo, University Paris 1, Panthéon, Sorbonne.

[21] Inder, B., (1993), *Estimating Long-run Relationships in Economics*, in: *Journal of Econometrics*, 75: 53-68.

[22] Isard, P., (1977), *How Far Can We Push the Law of One Price?,* in: *American Economic Review*, 67(5), pp. 942-948.

[23] Knetter, M.M., (1989), *Price Discrimination by U.S. and German Exporters*, in: *American Economic Review*, 79(1): 198-210.

[24] Knetter, M.M., (1993), *International Comparisons of Price-to-Market Behavior*, in: *American Economic Review*, 83(3): 473-486.

[25] Kremers, Ericsson J.N., and Dolado, J., (1992), *The Power of Cointegration Tests*, in: *Oxford Bullettin of Economics and Statistics*, 54: 325-348.

[26] Lippi, M. and Reichlin, L., (1994), *Diffusion of Technical Change and the Decomposition of Output into Trend and Cycle*, in: *Review of Economic Studies*, 61: 19-30.

[27] Lunnemann, P. and Matha, T.Y., (2004), *How Persistent is Disaggregate Inflation?,* ECB working paper No. 415.

[28] Lutz, M., (2002), *Price convergence under EMU? First estimates*, *working paper*, Institute of Economics, University of St Gallen.

[29] Mussa, M., (1986), *Nominal Exchange Rate Regimes and the Behavior of the Real Exchange Rates: Evidence and Implications*, Carnegie-Rochester Ser. Public Policy, 25: 117-213.

[30] Phillips, P.C.B., and Loretan, M., (1991), *Estimating Long-Run Economic Equilibria*, in: *Review of Economic Studies*, 58: 405-35.

[31] Rogers, J.H., (2002), *Monetary union, price level convergence, and inflation: How close is Europe to the United States?,* Board of Governors of the Federal Reserve System, International Finance Discussion Paper No. 2002-2740.

[32] Rogoff, K., (1996), *The purchasing power parity puzzle*, in: *Journal of Economic Literature*, 34: 647-668.

A SYSTEMATIC ANALYSIS OF PREFERENCE CHANGE IN CO-BRANDING

Chia-Lin **LEE**, Reinhold **DECKER**
Bielefeld University, **Germany**
Department of Business Administration and Economics
clee1@wiwi.uni-bielefeld.de, rdecker@wiwi.uni-bielefeld.de

Abstract:

This paper presents current theoretical and empirical findings on consumers' preference change in co-branding. We develop a conceptual model to illustrate attitudinal changes in co-branding based on the findings of previous research. We argue that attitude change is influenced by three important effects, namely the extension effect, the mutual effect and the reciprocal effect. It is shown how the interactions of these effects can be used to systematically explain the rationale behind preference change in co-branding. Our study also takes an initial step toward the understanding of the connection between product/brand evaluation and the success of alliance formation. Finally, we provide suggestions for marketing managers and motivate the need for further research in the field of strategic marketing.

Keywords: co-branding, attitude change, preference change, consumer behaviour.

JEL Classification: M31

1. Introduction

Most firms nearly always search for potential growth opportunities in the market they serve. One effective way is co-branding. Co-branding can avoid the possible failures derived from over-leveraging the equities of existing brands [Swaminathan *et al.*, (2001), James *et al.*, (2006)] and can reduce the introduction cost of new products [Kotler and Keller, (2006)]. Examples of co-branding include the *Oral-B Rembrandt* whitening pen, the *Sony-Ericsson* mobile phone, and the *NutraSweet* sweetener in *Diet-Coke*. In the optimal case, co-branding strategies make use of the salient attributes of the allying brands and offer opportunities for both players to reach a new market. However, the existence of a co-branding alliance can also cause an endogenous competition on consumer preferences (i.e., some consumers may change their preferences from one of the partnering brands to the other).

Venkatesh *et al.* (2000) argue that the occurrence of preference change is crucial because it influences the success of forming a co-branding alliance. However, their analysis totally ignores an important issue behind preference change, namely the consumers' evaluation of co-branding, which is a major topic in co-branding research [e.g., Park *et al.*, (1996), Simonin and Ruth, (1998)]. Therefore, the aim of this paper is to analyze preference change by relating it to the overall evaluation process (i.e., to perceptions and attitudes). To our knowledge, this study is the first one to provide a systematic analysis of preference change in the context of co-branding.

The remainder of this paper is organized as follows. Section 2 provides some definitions and a brief review of the co-branding literature. Section 3 summarizes previous findings on consumers' attitude change by means of a conceptual model. Section 4 offers a systematic analysis of consumers' preference change and applies the results in a three-brand scenario. Section 5 concludes the paper with some discussions and a short outlook on future research directions.

2. Definitions and literature review

The most common definition of "co-branding" (or "composite brand extension", "brand alliance") refers to the combination of two brand names to launch a single and unique product within a short- to long-term cooperation [Park *et al.*, (1996), Boo and Mattila, (2002), Kumar, (2005)]. In particular, we exclude the terms "product bundling" and "joint sales promotion" because they involve creating two or more products with the same or different brands [Leuthesser *et al.*, (2003), Hadjicharalambous, (2006)].

There are mainly two types of co-branding: Joint venture co-branding refers to two companies financially cooperating to offer a co-branded product [Kotler and Keller, (2006)]. The respective products are often from the same product category such as *Sony-Ericsson* mobile phones. Ingredient co-branding refers to the fact that a branded ingredient is part of a product introduced or promoted by another brand [Norris, (1992)]. Maybe the most famous example is the personal computers featuring "*Intel-inside*". The present study focuses on co-branded consumer durables such as co-branded mobile phones.

Successful co-branding may appeal to the consumers because it reinforces the attribute profiles of the product [Park *et al.*, (1996)] and differentiates the product by offering a quality assurance to the consumers [Rao and Ruekert, (1994)]. However, there are at least three possible problems when a firm wants to initiate a co-branding alliance. First, co-branding may give one of the allying brands the opportunity to penetrate the other's market [Leuthesser *et al.*, (2003)]. Second, the composite brand name may also dilute the brand equities of the partnering brands [Leuthesser *et al.*, (2003)]. Finally, a free-rider problem in sharing the profit may exist when the two brands have asymmetric contributions to the partnership [Simonin and Ruth, (1998)].

Co-branding is regarded as one type of brand extension [Park *et al.*, (1996), Hadjicharalambous, (2006)] and it is also viewed as one type of new product development strategy [Park *et al.*, (1996), Bouten, (2006), Hadjicharalambous, (2006)]. Therefore, marketing managers are not only involved in the introduction process but are also interested in the diffusion pattern of the new co-branded product. Figure 1 depicts the co-branding's nature as sketched above. If there exists a co-branding alliance formed by two specific brands, say *A* and *B*, the co-branded product *AB* is therefore regarded as a new and extended product launched by these two brands.

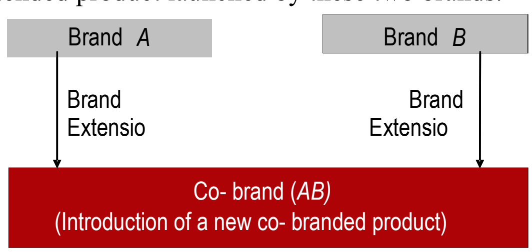


Figure 1. The nature of co-branding

3. A conceptual model of consumers' attitude change in co-branding

Consumer evaluation of co-branding is an essential topic in marketing and the corresponding cognitive process is a complex issue built on three relevant psychological theories, namely information integration [Anderson, (1981)], attitude accessibility [Fazio, (1989)], and contrast effects [Lynch *et al.*, (1991)]. In this section, we review previous research results on attitude change in co-branding and offer four statements for introducing three main effects that influence the attitude change. Here, the term "attitude change" refers to the changes of existing attitudes toward the parent (partnering) brands.

3.1 The mutual effect

The influence resulting from the "product fit" and the "brand fit" is called the "mutual effect" in this paper. The product fit [Simonin and Ruth, (1998), Bouten, (2006)] between the partnering brands has a direct impact on consumers' attitudes toward the co-brand. Previous research results show that if there exists a high product fit (e.g., because of complementing product attributes [Park *et al.*, (1996)] or a high relatedness between the product categories of the partnering brands [Simonin and Ruth, (1998)], consumers will have a favorable attitude toward the co-brand. Many studies have used the product fit to construct a theoretical model or to conduct an empirical analysis in the field of co-branding [e.g., Boo and Mattila, (2002), Bouten, (2006)].

Another important factor is the brand fit. A high fit of brand image (e.g., *Mercedes Benz* with *Louis Vuitton*) is proved to positively influence consumers' attitudes toward the co-brand [Simonin and Ruth, (1998), Bouten, (2006)]. That is, if the consumers perceive a distinct consistency between the images of the allying brands, they will have a favorable attitude toward the co-brand. This consistency can be reflected in the positioning strategy (e.g., both brands produce luxury products) and the overall performance (e.g., both brands are compatible in terms of market shares or sales volumes in their respective markets). Based on the findings of previous studies, we can formulate the following statement(s) about the mutual effect:

S_1: *A good (poor) product and brand fit results in a positive (negative) mutual effect and yields a favorable (unfavorable) attitude toward the co-brand.*

3.2 The extension effect

The prior attitude toward the parent brand is associated with the attitude toward the extended product in the brand extension context [Aaker and Keller, (1990)]. It can be measured in terms of the perceived quality [Aaker and Keller, (1990), Zeithaml, (1998)] and the prior purchase experience [Swaminathan, (2003)]. A high perceived quality or significant prior purchase experience regarding the parent brands implies a favorable attitude toward them [Aaker and Keller, (1990), Swaminathan, (2003)]. Hence, "high perceived quality" and "significant prior purchase experience" can be utilized as indicators representing favorable prior attitudes toward a brand. In addition, a significant prior experience can be used as a measure to represent a higher level of brand loyalty [Swaminathan *et al.*, (2001)].

Several scholars have also argued that the prior attitude plays an important role in the evaluation process of co-branding [Simonin and Ruth, (1998), Boo and Mattila, (2002), Lafferty and Goldsmith, (2005)]. Among these studies, Simonin and Ruth (1998) claim that the prior attitude toward one of the partnering brands is positively related to the consumer's attitude toward the co-brand and the post-exposure attitude toward that brand. Since co-branding is one type of brand extension, we term the influence resulting from the prior attitude the "extension effect". Thus, the following two statements can be written down:

S_2: *A favorable (unfavorable) prior attitude toward one of the partnering brands results in a positive (negative) extension effect and yields a relatively favorable (unfavorable) post-exposure attitude toward that brand.*

S_3: *A favorable (unfavorable) prior attitude toward one of the partnering brands results in a positive (negative) extension effect and yields a relatively favorable (unfavorable) attitude toward the co-brand.*

3.3 The reciprocal effect

The reciprocal effect first appeared in the brand extension context [Aaker and Keller, (1990), Lane and Jacobson, (1997), Swaminathan, (2003)] but has been applied to co-branding as well [Park *et al.*, (1996), Swaminathan, (1999)]. Different studies use different names to term this effect, such as the feedback effect [Park *et al.*, (1996)], the spillover effect [Simonin and Ruth, (1998)], and the post-effect [Leuthesser *et al.*, (2003)]. In this paper, the reciprocal effect is defined as an influence resulting from the attitudes toward the co-brand on each of the allying brands. Besides, the positive (negative) reciprocal effect yields a relatively favorable (unfavorable) post-exposure attitude toward each of the partners [Simonin and Ruth, (1998)]. Therefore, we conclude the following statement:

S_4: *A favorable (unfavorable) attitude toward the co-brand results in a positive (negative) reciprocal effect and yields a relatively favorable (unfavorable) post-exposure attitude toward each of the partnering brands.*

In conclusion, the process of attitude change can be described as follows: The extension effect and the mutual effect have direct impacts on consumers' attitudes toward the co-brand (cf. S_1 and S_3). The post-exposure attitude toward each of the allying brands will be affected by both the extension effect and reciprocal effect (cf. S_2 and S_4). Therefore, the possibility that a consumer will change her/his brand attitudes toward each of the partnering brands after the alliance will depend on the strength of the interactions of the considered effects. Figure 2 visualizes this process.

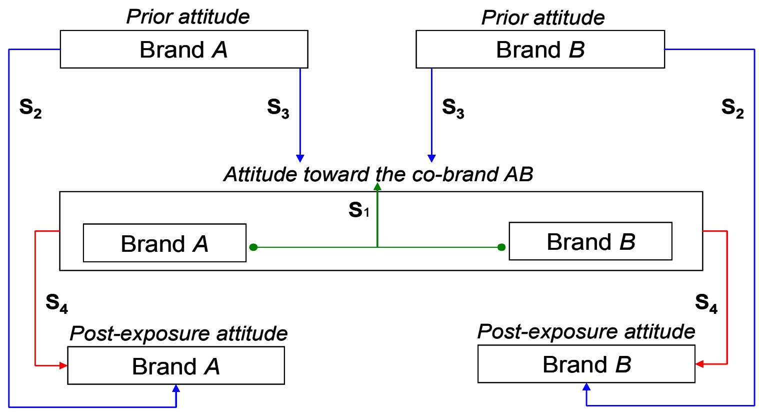


Figure 2. A conceptual model of attitude change in co-branding

4. Analysis of preference change in co-branding

Preferences are formed by the rank order of attitudes [Bass and Talarzyk, (1972)]. Hence, the attitude change can also trigger a preference change. To further discuss the latter term, we assume that several brands (termed *A*, *B* and *Y*, *Z* in the following) exist in the market of interest. Moreover, we consider two points of time as well as the intermediate period between both. At the first point of time ($t = 1$), the alliance is formed by brand *A* and *B* and releases the first co-branded product *AB*(1). Brand *A* and *B* are assumed to stop introducing their own products after the existence of the partnership. At the second point of time ($t = 2$), the alliance releases the second co-branded product *AB*(2). We further assume that a preference change will only occur after having purchased the first co-branded product [Simonin and Ruth, (1998)] in the intermediate period. So, the consumer preferences at time $t = 1$ are not affected by co-branding. The sequence of events is summarized in Figure 3.

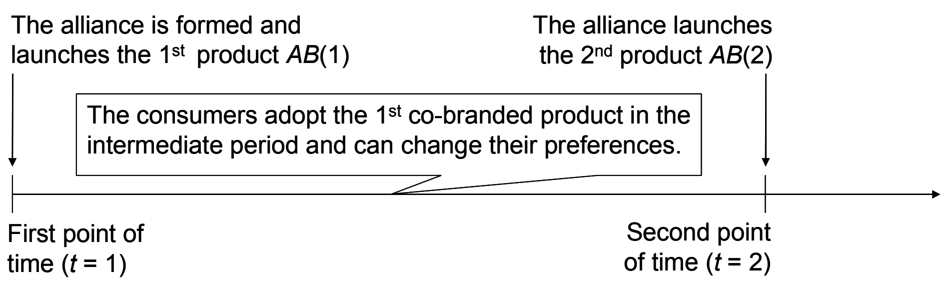


Figure 3. The sequence of events

Besides, the considered market is assumed to comprise several market segments. The consumers belonging to one segment prefer one specific brand. At the first point of time, we categorize the segments into two groups: One group is composed of those segments preferring the partnering brands (segment *A* and *B*) and the other includes the segments preferring the competing brands (segment *Y* and *Z*). Since the rationale of preference change is identical for each of the segments in the same group, we only focus on the process of preference change in segment *A* and *Z* in the next sections.

4.1 The preference change in segment A

To continue our analysis in this section, we have to assume that the consumers in segment *A* have a stable attitude toward brand *B* during the relevant time frame. That is, we only consider the reciprocal effect from the co-branded *AB* on brand *A*. This assumption is somewhat similar to the concept of comparative static analysis in economics and it will help us reduce the complexity of the following analysis. Besides, it should be noted that the co-brand *AB* does not have a reciprocal effect on consumers' attitude toward brand *Z*.

At time $t = 1$ all consumers belonging to segment *A* have a favorable prior attitude toward brand *A* and therefore prefer this brand. The preference change in segment *A* can be explained by three routes (see Figure 4).

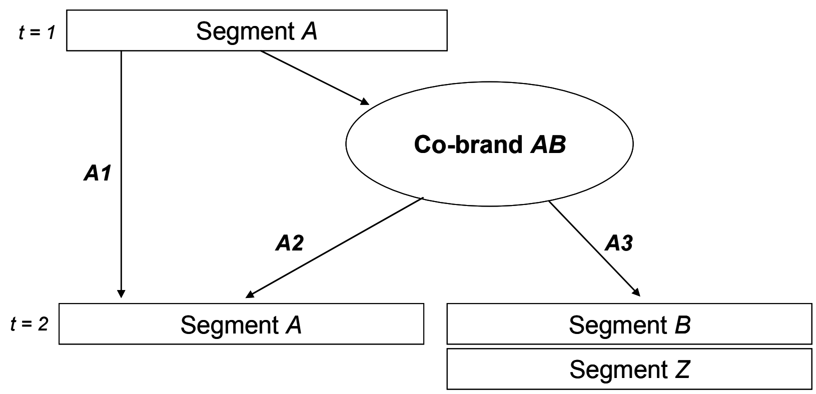


Figure 4. Routes of preference change regarding segment A

4.1.1 Route A1

We argue that a certain fraction of the consumers in segment *A* have a stable preference at time $t = 2$. This stable preference results from their extremely favorable prior attitude toward brand *A*. In other words, a significant positive extension effect (according to S_2) dominates the evaluation process. Two supporting arguments are provided below.

The first argument is related to brand familiarity, which can be defined as "the number of product-related experiences (product usage) that have been accumulated by consumers" [Alba and Hutchinson, (1987)]. Based on this definition, those consumers with an extremely favorable prior attitude toward brand *A* also have a significant prior purchase experience regarding this brand. Since brand familiarity can positively moderate the impact of prior attitude on post-exposure attitude [Simonin and Ruth, (1998)], a high level of brand familiarity will lead to stable preferences.

The second argument is related to brand loyalty. A favorable prior attitude implies brand loyalty [Dyson *et al.*, (1996)]. Those consumers who have an extremely favorable prior attitude toward brand *A* can be assumed to be completely loyal to this brand. It is commonly recognized that brand loyalty is highly resistant to change [Blackwell *et al.*, (2005)]. Therefore, the respective consumers are the most unlikely to change their brand attitudes after experiencing the co-branded product *AB* because they will ignore the potential inconsistent information and defend their well-established attitudes [Smith and Mackie, (2007)]. Therefore, an extremely favorable prior attitude toward brand *A* results in a stable preference.

4.1.2 Route A2 and A3

Although the remaining consumers of segment *A* also have a favorable prior attitude toward brand *A*, their attitudes are more amenable to change compared to those of the completely loyal consumers [Swaminathan *et al.*, (2001)]. Hence, at time $t = 2$, their preferences may stay with brand *A* or switch to another brand depending on the different levels of perceived product and brand fit.

If the respective consumers perceive a better fit from the alliance, a positive mutual effect will exist and subsequently the consumers will have a favorable attitude toward the co-brand *AB* (S_1) as well as a positive reciprocal effect on brand *A* (S_4). The resulting favorable post-exposure attitude implicates that the consumers still prefer brand *A* (route A2).

However, if the consumers perceive a poorer fit, their post-exposure attitudes toward brand *A* will be unclear. The negative mutual effect (originated from a poorer fit), together with the positive extension effect (S_3), will influence their attitude toward co-brand *AB* (S_1). The interplay may generate a favorable or an unfavorable attitude toward the co-brand and yield a positive or negative reciprocal effect (S_4) on brand *A*.

Hence, consumers may still prefer brand *A* (route A2) because the rank order of their attitudes at time 2 is the same as the one at time 1. On the other hand, it is also possible that their attitude toward brand *A* is adversely affected, and thus the rank order of their attitudes changes. In this case, the degree of favorability of brand *A* is lower than the other brands. Finally, the preference is likely to shift to any other competing brand (say brand *Z*) or to stay with brand *B* (route A3). The latter is called the "shift-in preference" [Venkatesh *et al.*, (2000)] which means that some consumers shift their preferences from one brand to its partner. Table 1 summarizes the interaction of the three effects and the preference change in segment *A*.

Table 1. Preference change in segment A

Route	Extension effect	Mutual effect	Reciprocal effect	Final segment
A1	Highly positive	-	-	Segment *A*
A2	Positive	Positive	Positive	Segment *A*
	Positive	Negative	Positive	Segment *A*
	Positive	Negative	Negative	Segment *A*
A3	Positive	Negative	Negative	Segment *B* (the partner) or segment *Z* (the competing brand)

4.2 The preference change in segment *Z*

At time $t = 1$, all consumers belonging to segment *Z* have a favorable prior attitude toward brand *Z* and therefore prefer this brand. Their preferences may also change at time $t = 2$. According to Figure 5, we can use five routes to explain the phenomenon of preference change.

4.2.1 Route Z1 and Z2

We argue that one group of consumers has a habitual buying behavior due to the well-established attitude toward brand *Z*. Hence, these consumers' preferences are stable (route Z1). Besides, some members of segment *Z* are not aware of co-brand *AB* and may shift their preferences to one of the remaining (but not explicitly considered) brands named *Y* in our example due to variety seeking (route Z2).

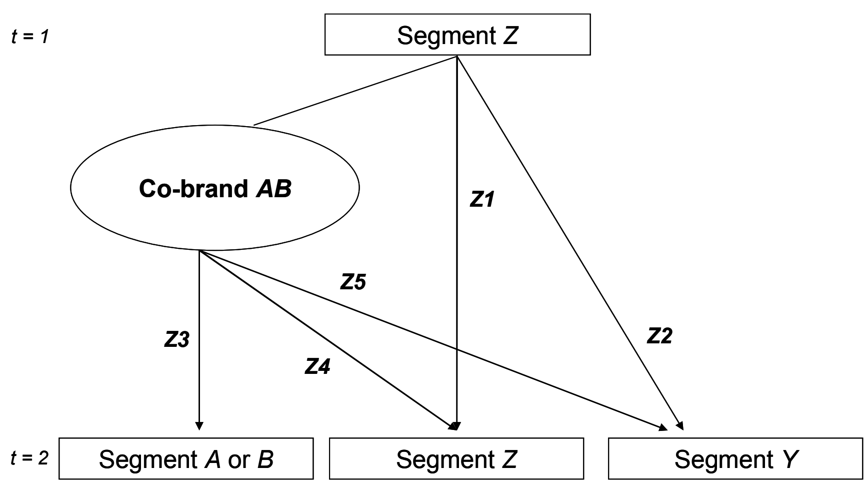


Figure 5. Routes of preference change in segment Z

4.2.2 Route Z3, Z4 and Z5

The rest of the consumers in segment *Z* are assumed to purchase the first co-branded product *AB*(1) in the intermediate period. In this case, the possibility of staying with brand *Z* depends on the interaction of the strength of the three main effects. If these consumers have a favorable attitude toward co-brand *AB*, a positive reciprocal effect on one of the allying brands *A* or *B* will exist (S_4). Besides, if these consumers have a favorable prior attitude toward brand *A*, they will have a favorable post-exposure attitude toward this brand (S_2). Accordingly, their attitude toward brand *A* will be enhanced and the rank order of the brand attitudes may change. Their preference is likely to stay with brand *A* (route Z3).

On the contrary, if the consumers have an unfavorable attitude toward co-brand *AB*, a negative reciprocal effect will exist and dilute their attitude toward brand *A* and *B*. Thus, their preferences will not stay with brand *A* or *B* at time $t = 2$. In this case, since the consumers' initial preference is brand *Z* and their attitude toward brand *A* (*B*) is diluted at time $t = 2$, they will definitely not stay with brand *A* (*B*). Consequently, depending on the rank order of their attitudes, the preferences may stay with brand *Z* (route Z4) or shift to a different competing brand *Y* (route Z5).

4.3. An application: preference change in a three-brand scenario

Let us now assume that the market of interest consists of exactly three brands: *A*, *B*, and *Z*. Brand *A* and *B* are supposed to form a co-branding partnership while brand *Z* is the competing brand. Each brand is assumed to be preferred by one segment (say segment *A* prefers brand *A*), and each consumer prefers only one brand at a certain point of time. Furthermore, $M_{A(1)}$, $M_{B(1)}$ and $M_{Z(1)}$ denote the sizes of segment *A*, *B* and *Z* at time $t = 1$. Analogously, $M_{A(2)}$, $M_{B(2)}$ and $M_{Z(2)}$ represent the counterparts at time $t = 2$.

The relationship between preference change and segment size can then be explained as follows: Co-brand *AB* is formed at time $t = 1$ and the consumers belonging to the three segments may change their preferences at time $t = 2$. If so, segment size $M_{A(2)}$ of brand *A* will be composed of three parts, namely F_{AA}, F_{BA}, and F_{ZA}. Here, F_{AA} refers to the proportion of consumers who stay with segment *A,* whereas F_{BA} and F_{ZA} denote the proportions of consumers who shift their preferences from brand *B* or *Z* to *A*. The same explanations can be applied to the notations of segments *B* and *Z*. Figure 6 concludes this evolution and provides a simplified basis for deepening behavioral studies in co-branding [e.g. in a quantitative respect as suggested by Venkatesh *et al.* (2000)].

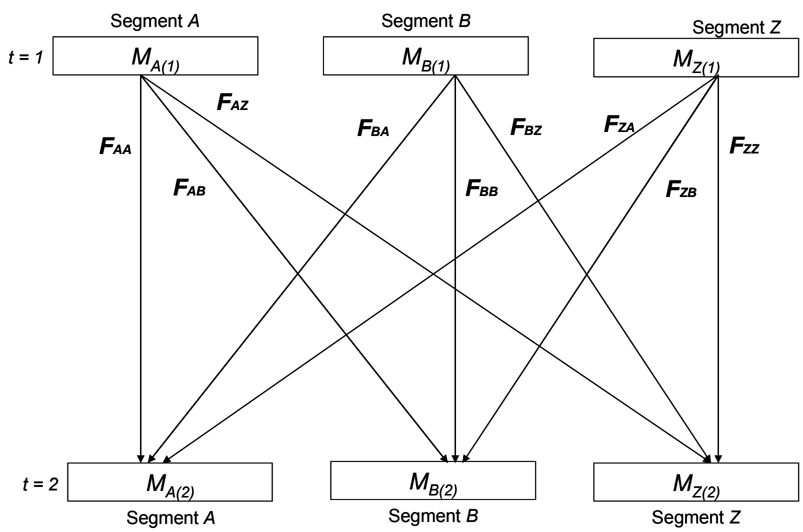


Figure 6. Preference change and evolution of segments

5. Conclusions

The incentive for a company to form a co-branding alliance is to gain an "added value" from the partnership. This added value, among others, can be referred to the opportunity for one of the partnering brands to build up its brand awareness at the other's customer base [Kippenberger, (2000), Leuthesser *et al.*, (2003)]. For instance, the co-branded credit card *Citibusiness / AAdvantage* not only enables *American Airlines* (AA) to build up its brand awareness in *Citibank*'s customer base but also provides an opportunity for *Citibank* to gain more transactions from flight tickets purchased by AA's customers who want to save the extra miles.

However, if the above example is presented in the same product category, Venkatesh *et al.* (2000) argue that the added value could cause consumers' preference change (shift-in preference) and the subsequent endogenous competition on consumer preferences. Finally, the alliance may end up because of one brand's loss in preference share. Hence, preference change plays an essential role in analyzing the success of alliance formation. This analysis provides the grounds of preference change and takes an initial step toward the understanding of the connection between product/brand evaluation and alliance success [e.g., for adapting the Venkatesh *et al.* (2000) model].

As in any research, our work is not without limitations. First, we did not include the influence of a brand's position into our analysis (e.g., *Sony-Ericsson* or *Ericsson-Sony*), although the order of the brand names can be assumed to influence consumers' attitudes toward the co-brand and the allocation of the reciprocal effect [Park *et al.*, (1996)]. Future research could address this interesting issue. Second, we did not discuss the role that brand familiarity plays in consumers' evaluation of co-branding. Simonin and Ruth (1998) conclude that the two brands with different levels of familiarity have unequal contributions to the formation of the composite concept. Hence, one can then use the level of brand familiarity as a weight to moderate the contribution to the alliance and the reciprocal effects on each partner. Third, we did not fully address the fit between the current and the co-branded products because we simply assumed that a better fit exists within our definition of co-branding. Finally, when adapting the Venkatesh *et al.* (2000) model to the above-sketched co-branding context, one could utilize our analysis and add attributes for each brand to explore the relationship between attribute evaluation and alliance success.

6. References

[1] Aaker, D.A. and Keller, K.L., (1990), *Consumer Evaluations of Brand Extensions*, in: *Journal of Marketing*, 54(1), pp. 27 – 41.

[2] Alba, J. W. and Hutchinson, W.J., (1987), *Dimensions of Consumer Expertise*, in: *Journal of Consumer Research*, 13(4), pp. 411 – 454.

[3] Anderson, N.H., (1981), *Foundation of Information Integration Theory*. New York: Academic Press.

[4] Bass, F.M., and Talarzyk, W.W., (1972), *An Attitude Model for the Study of Brand Preferences*, in: *Journal of Marketing Research*, 9(1), pp. 93 – 96.

[5] Blackwell, R.D., Miniard, P.W., and Engel, J.F., (2005), *Consumer Behavior* (10th ed.). Mason: Thomson South-Western.

[6] Boo, H. C. and Mattila, A., (2002), *A Hotel-restaurant Brand Alliance Model: Antecedents and Consequences*, in: *Journal of Foodservice Business Research*, 5(2), pp. 5 – 24.

[7] Bouten, L., (2006), *Co-branding in New Product Development*, 19th EMAC Doctoral Colloquium, University of Economics and Business, Athens.

[8] Dyson, P., Farr, A. and Hollis, N., (1996), *Understanding, Measuring, and Using Brand Equity*, in: *Journal of Advertising Research*, 36(6), pp. 9 – 21.

[9] Fazio, R., Powell, M. and Willams, C., (1989), *The Role of Attitude Accessibility in the Attitude-to-Behavior Process*, in: *Journal of Consumer Research*, 16, pp. 280 – 288.

[10] Hadjicharalambous, C., (2006), *A Typology of Brand Extensions: Positioning Cobranding As a Sub-Case of Brand Extensions*, in: *Journal of American Academy of Business*, 10(1), pp. 372 – 377.

[11] James, D.O., Lyman, M. and Foreman, S., (2006), *Does the Tail Wag the Dog? Brand Personality in Brand Alliance's Evaluation*, in: *Journal of Product and Brand Management*, 15(3), pp. 173 – 183.

[12] Kippenberger, T., (2000), *Co-branding as A New Competitive Weapon*, The Antidote, 5(6), pp. 12 – 15.

[13] Kotler, P. and Keller, K.L., (2006), *Marketing Management* (12th ed.). Upper Saddle River: Prentice Hall.

[14] Kumar, P., (2005), *The Impact of Cobranding on Customer Evaluation of Brand Counterextensions*, in: *Journal of Marketing*, 69(3), pp. 1 – 18.

[15] Lafferty, B.A. and Goldsmith, R.E., (2005), *Cause-brand Alliances: Does the Cause Help the Brand or Does the Brand Help the Cause?* in: *Journal of Business Research*, 58(4), pp. 423 – 429.

[16] Lane, V.R. and Jacobson, R., (1997), *The Reciprocal Impact of Brand Leveraging: Feedback Effects from Brand Extension Evaluations to Brand Evaluation*, in: *Marketing Letters*, 8(3), pp. 261 – 271.

[17] Leuthesser, L., Kohli, C. and Suri, R., (2003), *2+2=5? A Framework for Using Co-branding to Leverage a Brand*, in: *Brand Management*, 11(1), pp. 35 – 47.

[18] Lynch, J.G., Chakravarti, D. and Mitra, A., (1991), *Contrast Effects in Consumer Judgements: Changes in Mental Representations or in the Anchoring of Rating Scales?*, in: *Journal of Consumer Research*, 18(3), pp. 284 – 297.

[19] Norris, D.G. (1992), *Ingredient Branding: A Strategy Option with Multiple Beneficiaries*, in: *Journal of Consumer Marketing*, 9(3), pp. 19 – 31.

[20] Park, C.W., Jun, S.Y. and Shocker, A.D., (1996), *Composite Branding Alliances: An Investigation of Extension and Feedback Effects*, in: *Journal of Marketing Research*, 33(4), pp. 453 – 466.

[21] Rao, A.R. and Ruekert, R.W., (1994), *Brand Alliances as Signals of Product Quality*, in: *Sloan Management Review*, 36(1), pp. 87 – 97.

[22] Simonin, B.L. and Ruth, J.A., (1998), *Is A Company Known by the Company It Keeps? Assessing the Spillover Effects of Brand Alliances on Consumer Brand Attitudes,* in: *Journal of Marketing Research*, 35(1), pp. 30 – 42.

[23] Smith, E.R. and Mackie, D.M., (2007), *Social Psychology*. New York: Psychology Press.

[24] Swaminatham, V., (1999), *Do Co-branding Strategies Influence Brand Choice? An Empirical Analysis*, American Marketing Association – Conference Proceedings. Chicago: ABI/INFORM Global.

[25] Swaminathan, V., (2003), *Sequential Brand Extensions and Brand Choice Behavior*, in: *Journal of Business Research*, 56, pp. 431 – 442.

[26] Swaminathan, V., Richard, J.F., and Reddy, S.K., (2001), *The Impact of Brand Extension Introduction on Choice*, in: *Journal of Marketing*, 65(4), pp. 1 – 15.

[27] Venkatesh, R., Mahajan, V. and Muller, E., (2000), *Dynamic Co-Marketing Alliances: When and Why Do They Succeed or Fail?*, in: *International Journal of Research in Marketing*, 17, pp. 3 – 31.

[28] Zeithaml, V.A., (1988), *Consumer Perceptions of Price, Quality and Value: A Means-end Model and Synthesis of Evidence*, in: *Journal of Marketing*, 52, pp. 2 – 22.

Journal of Applied Economic Sciences
Volume IV/ Issue 1(7)/ Spring 2009

HERD BEHAVIOR IN MALAYSIAN CAPITAL MARKET: AN EMPIRICAL ANALYSIS

Jarita **DUASA**, Salina H. **KASSIM**
International Islamic University Malaysia, **Malaysia**
Faculty of Economics and Management Sciences,
jarita@iiu.edu.my, ksalina@iiu.edu.my

Abstract:

This study examines the existence of herd behavior among foreign investors in the Malaysian capital market. In methodology, the study analyzes the herd behavior by estimating vector error correction (VECM) model of FPI inflows as well as FPI outflows from/to major investors such as the United States, United Kingdom, Singapore and Hong Kong using quarterly data covering the period of Q1:1991 to Q3:2007. In addition, the variance decompositions and impulse response functions analyses are also adopted for further inferences. The findings provide empirical support on the existence of herd instinct among the foreign investors in the Malaysian capital market. The findings imply that the Malaysian short-term capital market can be volatile due to the herd behavior of the major portfolio investors. It is therefore imperative for Malaysia to take the necessary pre-cautionary measures to ensure that an aggregate reversal in portfolio investment flows would not result in a de-stabilising impact on the economy. This study emphasizes the need to improve market supervision and strengthen the financial structure in efforts to reduce the impact of capital reversal on the Malaysian financial market.

Keywords: foreign portfolio investment, herd behavior, VECM, impulse Response, Variance Decomposition

JEL Classification: G15, C32, C12.

1. Introduction

In the wake of increased global capital mobility, the issue of ensuring stability of foreign capital flows has captured the research interests of many. Countries which are highly opened to capital flows are getting more concerned about its stability due to the unprecedented magnitude of financial crises resulting from the highly volatile nature of foreign capital flows. Studies on the causes of capital flows cite various reasons leading to the volatile nature of foreign portfolio investment. In view of the importance of identifying factors affecting stability of portfolio investment particularly for a small open economy such as Malaysia, this study examines the causes of foreign portfolio investment flows for the case of Malaysia. In particular, this study examines the existence of herd behavior among foreign investors in the Malaysian capital market.

2. Literature review

Foreign portfolio investments are shown to be influenced by changes in economic and financial "fundamentals", [Corsetti *et al.*, (1998); Kaminsky, (1998); Krugman, (1979)] which can be categorized into the external and internal factors. The external factors include changes in macroeconomic and financial variables that are outside the host country that give impact on the volume of capital flows. Increased availability of financial capital coupled with sustained decline in the global interest rate and recessions in the industrialized economies have often been quoted as the major external factors that increase capital inflow into developing countries [Calvo *et al.*, (1993), (1996); Chuhan *et al.*, (1993); and Fernandez-Arias, (1996)]. Calvo *et al.* (1993) in particular, find evidence of lower foreign interest rates result in increased foreign capital flows into Latin America. Essentially, the external factors result in greater foreign capital inflow into the host economy as investors engage in "flight to quality" by shifting their funds to safer investment haven.

Internal factors are country-specific characteristics which have direct impact on the country's capital flow. This includes financial market-related factors such as improvement in investment infrastructure and macroeconomic-related variables relating to supply and demand shocks. Bachetta and Wincoop (1998) document positive relationship between capital inflow and favorable macroeconomic condition of a country bring about by successful macroeconomic stabilization policy and improvement in the financial market due to liberalization policies. Kim (2000) also highlighted successful resolution of debt problems among developing countries as well as advancement in transaction technologies as additional pull factors affecting the flow of capital.

Apart from the real or fundamental changes, there is a growing literature suggesting that capital flow, in particular, foreign portfolio investment are sensitive to the aggregate behavior in the financial market. In this context, investors react to financial market rumors even though the fundamentals are not deteriorating [Kaminsky and Schmuckler, (1999); Banerjee, (1992)]. The so-called "herding behavior" is based solely on "market sentiment" rather than objective assessment of market fundamentals partly because acquiring information can be costly in terms of time and money [see for example, Eichengreen and Mody, (1988); Kumar and Prasad, (2002)]. Herding behavior could result in massive outflow or inflow of funds and is purely contagion as investors react without undertaking careful assessment of the validity of the news that they are reacting to. In a related study, Baek (2006) finds that portfolio investment in Asia is caused by investors' appetite for risks, while portfolio investment in the Latin American is caused by real or fundamental factors in the domestic and global economy. Investment in Asia, therefore, is considered as more volatile in nature compared to that in Latin America since it is sensitive to market mood as opposed to true economic fundamentals.

Thus, this study attempts to analyze the herd behavior of investments by major foreign investors in Malaysia, namely the United States, United Kingdom, Singapore and Hong Kong using the **vector error correction model** (VECM). The study also undertakes the **variance decompositions** (VDC) analysis and **impulse response functions** (IRF) for further inferences.

An area of novelty of this study is that it analyzes portfolio investment behavior of the individual investing country rather than aggregated investment analysis in Malaysia. This country-by-country analysis enables detailed inferences to be made with respect to the investment behavior of the major investing countries. The rest of the paper is organized as follows: the next section provides some background information on foreign portfolio investment based on the Malaysian experience. In particular, this section highlights investment behavior of the top four major investing countries in Malaysia. Section 4 presents the empirical methods and preliminary analysis of the data. Section 5 highlights the empirical findings including the data preliminaries and the results based on the unrestricted VAR and VECM tests. Further inferences are then made based on the VDC and IRF analysis. Finally, section 6 concludes and draws several policy recommendations from the major findings of the paper.

3. Pattern of foreign portfolio investment inflows and outflows in Malaysia

During the period 1991 to 2007, foreign portfolio investment (FPI) in Malaysia has been rather volatile (see Table 1). The amount of total FPI (comprising of both inflow and outflow) ranged from RM40.6 billion in 1991 to a historical high of RM729.1 billion in 2007. Foreign portfolio investment has been very volatile in the pre-1997 period but has become more stable in the post-1997/1998 Asian crisis period. Total foreign portfolio inflow and outflow also reached record highs in 2007 at RM376.4 billion and RM352.6 billion, respectively. In terms of net portfolio investment, the lowest net portfolio investment was recorded during the Asian financial crisis in 1997 at negative RM28.4 billion. Net foreign portfolio investment has shown encouraging trend in the post-2000 period by recording positive flows since 2003, except for 2005 which recorded a negative net FPI of RM6.8 billion.

Of total FPI into Malaysia, approximately 80 percent originated from four countries, namely the US, the UK, Singapore and Hong Kong (see Table 2). On average, in the 1991-2007 period, 11.5 percent of total FPI comes from the US, 17.1 percent from the UK, 36.6 percent from Singapore and 22.6 percent from Hong Kong. An interesting observation of FPI from these countries is that the share of FPI inflows from these countries has continued to decline. In particular, in 1991, around 94.4 percent of total FPI came from these countries, while in 2007, the share has dropped significantly to only 79.3 percent. The decline was contributed by lower FPI from Singapore (from 54.5 percent in 1991 to 23.2 percent in 2007) and Hong Kong (from 24.4 percent in 1991 to 17.4 percent in 2007). The contribution of FPI from the US has increased from 5.1 percent in 1991 to 20 percent in 2007, while that from the UK has also increased from around 15 percent in the 1990s to around 22 percent in the post-2000 period. On aggregate basis, the decline in the contribution of these countries to total FPI inflow also indicate the increasing importance of FPI from other sources such as from "other countries" which details are not being specified by the Malaysian Central Bank – Bank Negara Malaysia.

Table 1. Total Foreign Portfolio in Malaysia, 1991 – 2007

	Total Inflow	Total Outflow	Total FPI	Net
1991	19,346	21,274	40,620	-1,928
1992	60,935	53,043	113,978	7,892
1993	187,779	162,128	349,907	25,651
1994	238,454	224,425	462,879	14,029
1995	106,414	101,054	207,468	5,360
1996	144,933	136,090	281,023	8,843
1997	156,162	184,517	340,679	-28,355
1998	57,028	58,286	115,314	-1,258
1999	43,598	42,532	86,130	1,066
2000	54,529	63,274	117,803	-8,745
2001	37,910	39,891	77,801	-1,981
2002	54,383	59,381	113,764	-4,998
2003	76,013	65,164	141,177	10,849
2004	135,107	100,419	235,526	34,688
2005	127,298	134,137	261,435	-6,839
2006	172,661	161,579	334,240	11,082
2007	376,444	352,612	729,056	23,832
Average	120,529	115,283	235,812	5,246

Source: Bank Negara Malaysia's *Quarterly Bulletin,* International Monetary Fund's *IMF Financial Statistics* of various issues and own calculation.

Table 2. Inflow of Foreign Portfolio Investment in Malaysia by Major Investing Countries, 1991 – 2007

	US	UK	Singapore	Hong Kong	% of Total FPI Inflow
1991	995	2,174	10,359	4,731	94.38
1992	4,361	13,471	31,596	9,853	97.29
1993	9,135	26,100	113,307	31,343	95.80
1994	35,028	36,004	114,018	37,267	93.23
1995	13,778	12,304	52,154	24,109	96.18
1996	8,870	17,654	70,198	41,699	95.51
1997	9,878	20,646	75,373	42,229	94.85
1998	5,625	6,867	22,239	17,477	91.55
1999	2,871	5,856	18,157	8,474	81.10
2000	4,749	8,160	16,072	17,155	84.61
2001	7,353	7,578	7,530	8,703	82.21
2002	7,258	12,085	11,068	13,720	81.15
2003	9,171	19,621	15,192	20,279	84.54
2004	20,131	28,943	34,990	29,900	84.35
2005	20,116	27,331	31,737	25,904	82.55
2006	30,030	36,946	31,169	28,537	73.37
2007	74,758	71,077	87,177	65,441	79.28
Average	15,536	20,754	43,667	25,107	87.17

Source: Bank Negara Malaysia's *Quarterly Bulletin,* International Monetary Fund's *IMF Financial Statistics* of various issues and own calculation.

Table 3. Outflow of Foreign Portfolio Investment in Malaysia by Major Investing Countries,

1991 – 2007

	US	UK	Singapore	Hong Kong	% of Total FPI Outflow
1991	692	2,358	11,722	5,336	94.52
1992	4,168	13,608	27,009	6,792	97.24
1993	6,436	26,532	98,997	24,584	96.56
1994	31,576	31,045	101,809	44,399	93.05
1995	7,483	10,733	52,079	26,617	95.90
1996	5,832	16,012	67,591	41,386	96.13
1997	10,219	20,562	80,316	64,055	94.92
1998	5,479	5,675	24,636	18,915	93.86
1999	3,768	6,248	18,489	8,907	87.96
2000	4,338	10,262	20,478	19,120	85.66
2001	3,670	10,249	9,156	8,466	79.07
2002	7,075	13,089	12,921	13,803	78.96
2003	7,728	19,213	14,169	14,391	85.17
2004	17,213	22,361	21,251	25,531	86.00
2005	20,060	26,113	36,295	26,458	81.21
2006	24,148	31,928	41,537	22,162	74.13
2007	74,750	62,720	79,916	62,288	79.31
Average	13,802	19,336	42,257	25,483	87.50

Source: Bank Negara Malaysia's *Quarterly Bulletin,* International Monetary Fund's *IMF Financial Statistics* of various issues and own calculation.

Similar to inflow, around 80 percent of FPI outflow resulted from the US, the UK, Singapore and Hong Kong (see Table 3). In the period of 1991-2007, around 10 percent of total FPI outflow went to the US, 17.5 percent to the UK, 37.6 percent to Singapore and 23 percent to Hong Kong. Over the years, outflows of FPI to the US showed an increasing trend, while outflows to the UK, Singapore and Hong Kong seemed to be declining. As in the case of inflow, the total outflows to these countries have also been declining in view of the increased outflow to the "other countries".

Similar declining trends of FPI inflows from Singapore and Hong Kong as well as FPI outflows to UK, Singapore and Hong Kong give preliminary indication that there could be a strong common behavior among these foreign investors towards investing in the country and pulling out the investment out from the country. This behavior could be tested by proper empirical tests which will be conducted in this study.

4. Methodology

Data of FPI inflows and outflows from/to the United States, United Kingdom, Singapore and Hong Kong as well as real GDP are quarterly, ranging from Q1:1991 to Q3:2007 and sourced from Bank Negara Malaysia's *Quarterly Bulletin* and International Monetary Fund's *IMF Financial Statistics* of various issues. The raw data obtained for all variables are in RM million and the base year for real GDP is 1987. All variables are expressed in their logarithmic transformation, denoted by italic small letters. Δ denotes the first difference operator.

To evaluate the integration properties of the variables, we employ standard augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests [Dickey and Fuller, (1979); Phillips and Perron, (1988)]. A variable is said to be integrated of order d, written $I(d)$ if it requires differencing d times to achieve stationarity. For cointegration, we employ the VAR based tests of Johansen [1988] and Johansen and Juselius [1990].

To test the herd behavior among foreign investors in Malaysian portfolio market, the vector autoregressive (VAR) model is adopted first on all FPI inflows (US, UK, Singapore and Hong Kong). In this analysis, there is a set of p=4 endogenous variables, *z = [fpiius, fpiiuk, fpiis, fpiihk]* where *fpiius, fpiiuk, fpiis* and *fpiihk* refer to the logarithm of US FPI inflow, UK FPI inflow, Singapore FPI inflow and Hong Kong FPI inflow, respectively. Following Johansen (1988,1991) and Johansen and Juselius (1990,1992), we consider a *p*-dimensional vector time series *zt* and model it as an Unrestricted Vector Autoregression (VAR) involving up to *k*-lags of z_t.

$$z_t = A_1 z_{t-1} + ... + A_k z_{t-k} + \mu + \varepsilon_t, \quad \varepsilon_t \sim niid(0, \Sigma) \tag{1}$$

where z_t is a *(px1)* matrix and each of the A_i is a *(pxp)* matrix of parameters. The Johansen approach is used with the consideration that it enables hypotheses tests concerning the matrix and the number of equilibrium relationships to be carried out. Before test of cointegration could be done, we have to choose the maximum lag length, *k*, in the Unrestricted Vector Autoregression Model (VAR). Choosing the appropriate lag length is important since a *k* too small will invalidate the tests, whereas a *k* too large may result in a loss of power [Kanioura, (2001)]. The appropriate lag is chosen by checking the residuals of VAR model with one lag after another and the selection of lag is based on the one that has the absence of serial correlation in the residuals. Being aware of the lag order, then we construct the long-run equations (Unrestricted VAR model) for the series. The analysis is carried out further by doing the Johansen cointegration test with *k-1* lag. The determination of the number of cointegrating vectors is based on the *maximal eigenvalue* and the *trace* tests.The vector error correction model (VECM) restricts the long-run behaviour of the endogenous variables to converge to their cointegrating relationships while allowing for short-run adjustment dynamics. In this case, the cointegration terms are the correction terms since a series of partial short-run adjustments correct gradually the deviation from long-run equilibrium. The VECM corresponds to a restricted VAR of order *k-1* for the first differenced series, with the inclusion of error-correction terms for the cointegrating vectors. We write a *p*-dimensional vector error correction model (VECM) as follows:

$$\Delta y_t = \sum_{i}^{k-1} \Gamma_i \Delta y_{t-i} + \Pi y_{t-1} + \mu + \varepsilon_t \quad , \quad t = 1, \ldots T \tag{2}$$

where y_t is the set of *I*(1) variables discuss above; $\varepsilon_t \sim niid(0,\Sigma)$; μ is a drift parameter, and Π is a (*p* x *p*) matrix of the form $\Pi = \alpha\beta'$ where α and β are both (*p* x *r*) matrices of full rank, with β containing the r cointegrating vectors and α carrying the corresponding loadings in each of the *r* vectors. The adjustment coefficients in matrix α refer to the coefficients of the Error Correction (ECM) terms.

Additionally, we adopt an innovation accounting by simulating variance decompositions (VDC) and impulse response functions (IRF) for further inferences. VDC and IRF serve as tools for evaluating the dynamic interactions and strength of causal relations among variables in the system. The VDC indicate the percentages of a variable's forecast error variance attributable to its own innovations and innovations in other variables. Thus, from the VDC, we can measure the relative importance of fluctuation of one country FPI inflow in accounting for fluctuation in FPI inflows from other countries. Moreover, the IRF trace the directional responses of a variable to a one standard deviation shock of another variable. This means that we can observe the direction, magnitude and persistence of FPI inflow of one country to variation in FPI inflows from other countries.

For similar objective, the above method is repeated for FPI outflows from Malaysia to the United States, United Kingdom, Singapore and Hong Kong. In this case, the VAR model is applied on FPI outflows to all these foreign countries (denoted as *fpious, fpiouk, fpios* and *fpiohk*).

5. Empirical findings

As a preliminary step, we first subject each variable to Augmented Dickey Fuller (ADF) and Phillip-Perron (P-P) unit root tests. The results of the tests are displayed on Table 4. The results generally suggest that most variables are integrated of order one as the null hypothesis that the series are not stationary is accepted at level but rejected at first difference. In other words, the variables are stationary at first difference or I(1).

Table 4. Unit Root Tests

Variable	ADF test statistic (with trend and intercept)		P-P test statistic (with trend and intercept)	
	Level	*First Difference*	*Level*	*First Difference*
rgdp	-2.44	-3.71**	-3.34*	-9.48***
fpiius	-3.40*	-10.52***	-3.33*	-10.77***
fpiiuk	-3.47*	-10.83***	-3.46*	-10.69***
fpiis	-2.71	-7.92***	-2.69	-7.87***
fpiihk	-2.97	-6.96***	-2.97	-6.83***
fpious	-3.28*	-6.55***	-3.16	-10.33***
fpiouk	-3.83**	-10.91***	-3.75**	-10.92***
fpios	-2.59	-7.69***	-2.66	-7.72***
fpiohk	-2.90	-5.88***	-2.49	-5.90***

Note: *** , ** and * denote significance at 1%, 5% and 10% level, respectively.

5.1. Analysis on FPI inflows into Malaysia

The first VAR model developed consists of 4 endogenous variables of capital inflows: *z = [fpiius, fpiiuk, fpiis, fpiihk]*. For this model, the maximum lag length, *k*, of 2 is chosen. Based on *Maximum Eigenvalue* and *Trace* tests of cointegration, there are two cointegrating vectors existed among the variables. Table 5 provides detail results of these cointegration tests.

Table 5. Johansen Cointegration Tests Results

Null Hypothesis about Rank (r)	*Max-Eigen* Statistic	5% Critical Value	*Trace* Statistic	5% Critical Value
r=0	29.18	27.58	69.94	47.86
r≤1	25.68	21.13	40.76	29.80
r≤2	12.18	14.26	15.08	15.49
r≤3	2.90	3.84	2.89	3.84

Normalising *fpiis* for cointegrating vector 1 and *fpiiuk* for cointegrating vector 2, following are the suggested vectors:

CV1 = *fpiis* + 1.55*fpiius* – 2.51*fpiihk* + 0.326
CV2 = *fpiiuk* – 0.64*fpiius* – 0.33*fpiihk* -0.455

We then proceed with an estimated error correction model using the 4 foreign portfolio inflow variables to illustrate how the cointegration results might be utilised. The vector error correction model (VECM) restricts the long-run behaviour of the endogenous variables to converge to their cointegrating relationships while allowing for short-run adjustment dynamics. Table 6 displays 4 short-run equations for Δ*fpiius,* Δ*fpiiuk,* Δ*fpiis* and Δ*fpiihk*, respectively. All coefficients of short-run equation are coefficients relating to the short run dynamics of the model's convergence to equilibrium and coefficients of lag CV (error correction term) represent the speed of adjustment. From all 4 short-run equations, equation 1 is selected to be used for further inferences as it has at least one lag error correction term ($CV1_{t-1}$) which is significant with negative sign. The negative sign of the ECM terms or cointegrating vectors is rather better results to be considered since it is the correct sign of the error correction. The significant of an error correction term shows the evidence of causality in at least one direction.

To support the selection of equation 1, we apply a number of diagnostic tests to the error correction model. We find no evidence of serial correlation, heteroskedasticity and ARCH (Autoregressive Conditional Heteroskedasticity) effect in the disturbances. The model also passes the Jarque-Bera normality test which suggesting that the errors are normally distributed.

Table 6. The Vector Error Correction Mo**del**

Equation / Ind. Variable	Dependent Variable (1) $\Delta fpiius$	(2) $\Delta fpiiuk$	(3) $\Delta fpiis$	(4) $\Delta fpiihk$
constant	0.06	0.03	-0.01	0.02
$\Delta fpiius_{t-1}$	0.27	0.32*	0.10	0.09
$\Delta fpiius_{t-2}$	0.13	0.25*	0.15	0.16
$\Delta fpiiuk_{t-1}$	-0.68**	-0.73***	-0.08	-0.39
$\Delta fpiiuk_{t-2}$	0.06	-0.05	0.23	0.05
$\Delta fpiis_{t-1}$	-0.41	0.10	-0.56**	-0.28
$\Delta fpiis_{t-2}$	-0.32	0.04	-0.42*	-0.02
$\Delta fpiihk_{t-1}$	0.34	0.18	0.49**	0.49**
$\Delta fpiihk_{t-2}$	-0.15	-0.38*	-0.15	-0.24
$CV1_{t-1}$	-0.17**	-0.11	0.04	0.12*
$CV2_{t-1}$	0.92***	0.17	0.36*	0.39
Included observation	64	64	64	64
Adjusted R^2	0.32	0.14	0.15	0.14
F-statistic	3.99***	2.03**	2.09**	2.02**
Diagnostic test:				
Far	0.63			
Farch	0.27			
JBnormal	1.41			
Fhet	0.68			

Notes: 1. *Far* is the F-statistic of Breusch-Godfrey Serial Correlation LM Test, *Farch* is the F-statistic of ARCH Test, *JBnormal* is the Jarque-Bera Statistic of Normality Test, *Fhet* is the F-statistic of White Heteroskedasticity Test. 2. *** , ** and * denote significance at 1%, 5% and 10% level, respectively.

From an estimated VAR, we compute variance decompositions and impulse-response functions, which serve as tools for evaluating the dynamic interactions and strength of causal relations among variables in the system. The results of variance decomposition and impulse response functions are displayed in Table 7 and Figure 1, respectively.

From Figure 1, the IRF can produce the time path of dependent variables in the VAR, to shocks from all the explanatory variables. It could be seen that FPI inflow from Hong Kong does react significantly to FPI inflow innovations from Singapore and UK as it respond positively for the first 5 quarters and then subsides to zero afterwards. The figure also shows that the FPI inflow from UK responds positively to a shock in FPI inflows from Hong Kong and Singapore for about 9 quarters before it subsided to zero. Shock in FPI flow from the US, however, does not give significant impact on both FPI inflows from Hong Kong and UK. Interestingly, FPI inflow from the US reacts positively and significantly to shock in all other countries inflows with the longest period of reaction on shock is 16 quarters from UK inflow. These results imply that FPI inflow from the US is highly sensitive to shock in FPI inflows from other countries but shock in the US FPI inflow itself is rather insignificant to the other countries FPI inflows.

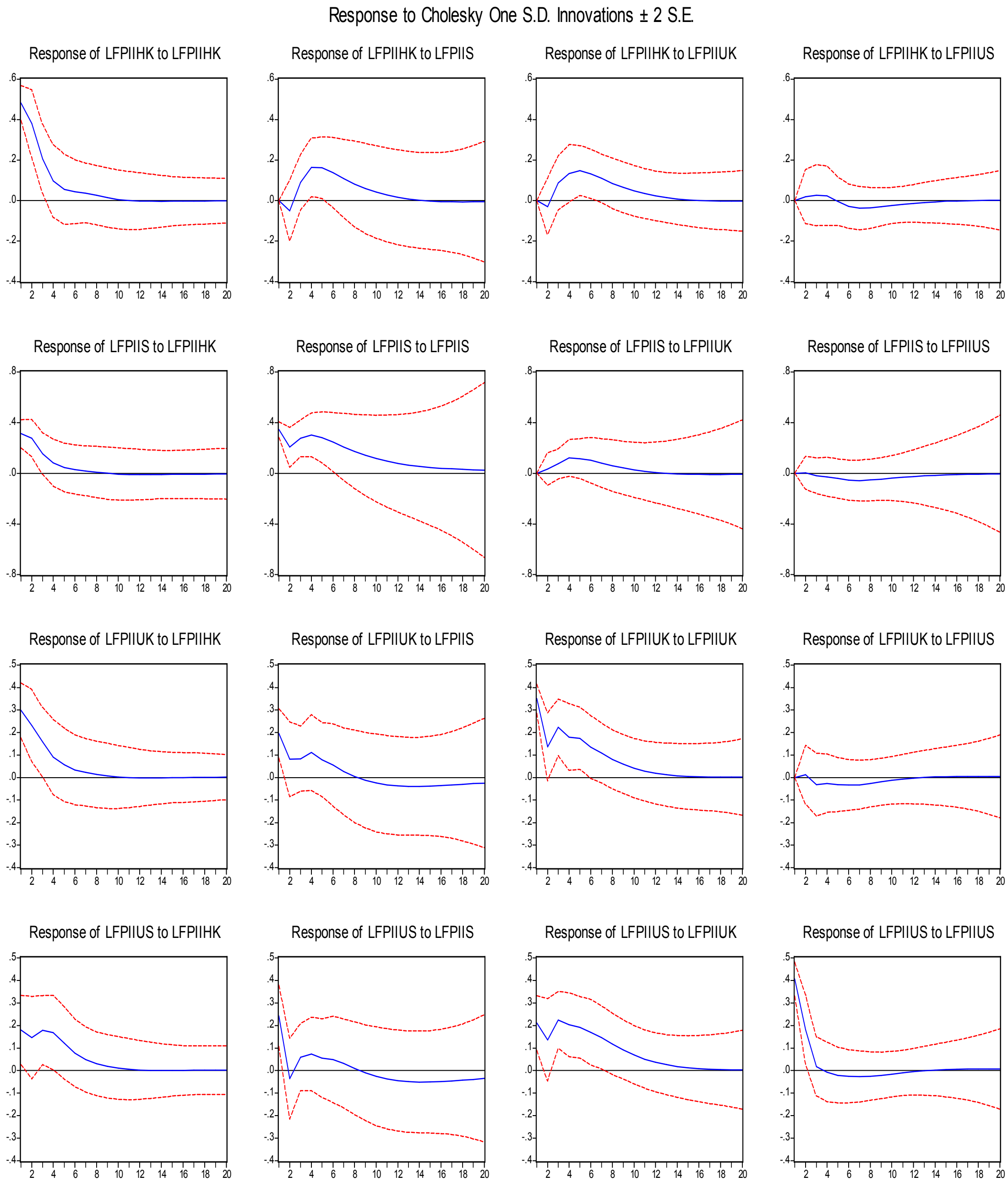


Figure 1. Impulse Response Functions, FPI inflows

As discussed earlier, the variance decomposition is an alternative method to IRF for examining the effects of shocks to the dependent variables. It determines how much of the forecast error variance for any variable in a system is explained by innovations to each explanatory variable, over a series of time horizons. Usually own series shocks explain most of the error variance, although the shock will also affect other variables in the system. From Table 7, looking along the main diagonal, the results reveal that the own shock is relatively high for FPI inflow from Hong Kong with 88 percent even at longer time horizon (20 quarters). This implies the exogeneity of FPI inflow from Hong Kong in variance decompositions, as after the first quarter after the shock, the variance appears to be less explained by innovations in other explanatory variables. On the other hand, the results shows that the percentage of variance explained by own shock for other countries FPI inflows are relatively smaller especially from Singapore and the US. In fact, the own shock's contribution is declining in long run for US inflow from 56 percent in first quarter to 8 percent in 20th quarter period which indicates that the US inflow is highly endogenous.

The VDC substantiate the significant role played by UK, Singapore and Hong Kong FPI inflows in accounting for fluctuations in US FPI inflow. At 2 quarter horizon, the fraction of US inflow forecast error variance attributable to variations in UK, Singapore and Hong Kong inflows are 24 percent, 14 percent and 12 percent, respectively. The explanatory power of all variables continuously increases at longer horizon and at 20-quarter horizon the contributions are 60 percent, 15 percent and 17 percent, respectively. Obviously, for FPI inflow from Singapore, percentage of forecast variance in it is largely explained by innovation in FPI inflow from Hong Kong, among other explanatory variables as it maintains higher percentage than the other. As for FPI from UK, its forecast error variance largely explained by variations in Hong Kong and Singapore inflows. The results again strengthen the findings earlier that the US FPI inflow has insignificant role in determining the variation of other countries FPI inflows, but FPI inflow from the US is significantly determined by FPI inflows from other countries.

Table 7. Variance Decompositions of FPI inflows

Period (Qtr)	Variance Decomposition of *fpiihk*				
	S.E.	*fpiihk*	*fpiis*	*fpiiuk*	*fpiius*
2	0.801970	99.07307	0.427826	0.498983	0.000121
4	1.066667	96.83784	1.336718	1.510587	0.314858
6	1.304335	94.60676	1.391247	2.467931	1.534066
8	1.527872	92.69070	1.714459	3.174081	2.420762
10	1.721136	91.21406	1.949895	3.737675	3.098375
12	1.900109	90.11585	2.112914	4.140608	3.630627
14	2.064232	89.28512	2.242340	4.450299	4.022243
16	2.216725	88.65103	2.341633	4.681973	4.325364
18	2.359700	88.15509	2.418791	4.864376	4.561739
20	2.494569	87.75829	2.480639	5.009976	4.751094

Period (Qtr)	Variance Decomposition of *fpiis:*				
	S.E.	*fpiihk*	*fpiis*	*fpiiuk*	*fpiius*
2	0.650984	64.30578	34.61180	0.792431	0.289991
4	0.859387	55.26631	37.54554	6.350516	0.837635
6	1.076842	52.39249	35.48135	8.787317	3.338839
8	1.274401	49.88800	35.53050	10.12215	4.459361
10	1.447493	48.15726	35.33862	11.11625	5.387863
12	1.607426	47.05021	35.11842	11.77309	6.058280
14	1.753464	46.20656	35.00983	12.24785	6.535765
16	1.888934	45.58988	34.91444	12.59610	6.899579
18	2.015627	45.11498	34.83902	12.86660	7.179398
20	2.134919	44.73902	34.78127	13.07857	7.401146

Period (Qtr)	Variance Decomposition of *fpiiuk:*				
	S.E.	*fpiihk*	*fpiis*	*fpiiuk*	*fpiius*
2	0.637880	39.82943	15.71156	44.36944	0.089572
4	0.885342	30.65441	15.66059	52.87568	0.809318

6	1.118448	29.29182	15.36570	53.02401	2.318474
8	1.327698	27.88087	15.62400	53.31246	3.182666
10	1.514197	26.94234	15.73254	53.42858	3.896542
12	1.684422	26.35741	15.80449	53.44059	4.397511
14	1.840102	25.91303	15.86282	53.45919	4.764965
16	1.984306	25.59373	15.90677	53.45724	5.042263
18	2.119027	25.34697	15.93926	53.45836	5.255413
20	2.245808	25.15318	15.96543	53.45708	5.424304

Period (Qtr)	Variance Decomposition of *fpiius:*				
	S.E.	*fpiihk*	*fpiis*	*fpiiuk*	*fpiius*
2	0.625642	11.93461	14.34689	23.93155	49.78695
4	0.828903	12.98024	13.84663	44.34451	28.82861
6	1.076470	16.77142	13.19045	52.25831	17.77983
8	1.300087	16.75827	13.76990	56.08722	13.38460
10	1.507020	16.97606	14.05466	57.71948	11.24980
12	1.696865	16.98888	14.19842	58.76756	10.04513
14	1.870058	16.98448	14.34045	59.36562	9.309452
16	2.029979	16.97555	14.43772	59.78161	8.805127
18	2.178859	16.96320	14.50759	60.08615	8.443052
20	2.318460	16.95379	14.56489	60.31331	8.168008
Cholesky Ordering: LFPIIHK LFPIIS LFPIIUK LFPIIUS					

5.2 Analysis on FPI outflows from Malaysia

In this second part of the analysis, we adopt similar VAR model using FPI outflows from Malaysia to all four foreign countries in study. Thus, the VAR model with four endogenous variables used is *z = [fpious, fpiouk, fpios, fpiohk]* where *fpious, fpiouk, fpios* and *fpiohk* refer to the logarithm of US FPI outflow, UK FPI outflow, Singapore FPI outflow and Hong Kong FPI outflow, respectively. Lag 3 is chosen as the optimal lag for the VAR model. However, we found no cointegration existed among the variables using *Maximum Eigenvalue* and *Trace* tests of cointegration as shown on Table 8.

Table 8. Johansen Cointegration Tests Results

Null Hypothesis about Rank (r)	*Max-Eigen* Statistic	5% Critical Value	*Trace* Statistic	5% Critical Value
r=0	19.54	27.58	44.71	47.86
r≤1	13.29	21.13	25.17	29.80
r≤2	10.17	14.26	11.88	15.49
r≤3	1.71	3.84	1.71	3.84

Since there is no cointegration traced, we conduct Toda and Yamamoto (1995) non-causality test to establish the direction of causation between the two variables. The main advantage of this test over Granger causality test is that it does not require pre-tests of stationarity or cointegration between the series.[1]

The Toda and Yamamoto (1995) procedure essentially suggests the determination of the *d-max*, namely, the maximal order of integration of the series in the model, and to intentionally over-fit the

[1] The unit root and cointegration tests are usually required before testing for causality. This might contribute to possible pretest biases due to the sensitivity of stationary or cointegration tests. The pretest biases might be severe as the power of unit root tests is known to be very low and tests for cointegrating rank in Johansen (1991) are not very reliable for finite samples [*see* Reimers (1992) and Toda and Yamamoto (1995)].

causality test underlying model with additional *d-max* lags – so that the VAR order is now $p = k + d$, where k is the optimal lag order. This modified version of the Granger causality test is employed to establish a causal relationship between variables in this study. The test is done by estimating a two-equation system:

$$Y_t = \alpha_1 + \sum_{i=1}^{k+d\max} \beta_i Y_{t-i} + \sum_{i=1}^{k+d\max} \delta_i X_{t-i} + \mu_t \quad (3)$$

$$X_t = \alpha_2 + \sum_{i=1}^{k+d\max} \phi_i Y_{t-i} + \sum_{i=1}^{k+d\max} \theta_i X_{t-i} + \nu_t \quad (4)$$

where d-max is the maximal order of integration of the series in the system and μ_t and ν_t are error correction terms that are assumed to be white noise. The Wald tests were then applied to the first k coefficient matrices using the standard χ^2-statistics. The null hypothesis set for equation (3) is $\delta_i = 0 \forall_i \leq k$ and for equation (4) is $\phi_i = 0 \forall_i \leq k$. From equation (3), X "Granger-causes" Y if its null hypothesis is rejected and from equation (4), Y "Granger-causes" X if its null hypothesis is rejected. Unidirectional causality will occur between two variables if either null hypothesis of equation (3) or (4) is rejected. Bidirectional causality existed if both null hypotheses are rejected and no causality existed if neither null hypothesis of equation (3) nor equation (4) is rejected.

Table 9 displays results obtained for Toda and Yamamoto non-causality tests. Obviously, there are bidirectional causality between FPI outflow to Singapore and FPI outflow to Hong Kong. In other words, capital outflow to Singapore causes capital outflow to Hong Kong, vice versa. In addition, significant Wald test coefficient of US capital outflow in UK outflow equation indicates that capital outflow to UK is caused by capital outflow to US.

Table 9. Toda-Yamamoto Non-causality test results

Equation	Wald test (χ^2)
fpiohk (in *fpiouk* equation)	3.26 (0.35)
fpios (in *fpiouk* equation)	4.24 (0.24)
fpious (in *fpiouk* equation)	**6.4 (0.09)**
fpiohk (in *fpious* equation)	0.54 (0.91)
fpios (in *fpious* equation)	2.29 (0.51)
fpiouk (in *fpious* equation)	4.90 (0.18)
fpiohk (in *fpios* equation)	**8.08 (0.04)**
fpiuk (in *fpios* equation)	5.12 (0.16)
fpious (in *fpios* equation)	5.77 (0.12)
fpios (in *fpiohk* equation)	**11.12 (0.01)**
fpious (in *fpiohk* equation)	4.82 (0.19)
fpiouk (in *fpiohk* equation)	1.29 (0.73)

Note: The figures in parentheses are the p-values.

Overall, results from the analyses of both FPI inflows and FPI outflow to/from Malaysia from/to the four main trading partners clearly show the existence of herd behavior among foreign investors in the Malaysian short-term capital market. In particular, the study found that capital inflow from the US is highly influenced by other countries' inflows especially from the UK. As for outflows, capital outflow from the UK is highly affected by capital outflow from the US and both Singapore and Hong Kong capital outflows are influenced by each other. This clearly explains the reason why the country's portfolio market was badly affected by the financial crisis in 1997 which worsened the balance position of Malaysian capital market.

5. Conclusion

By employing the VECM model to analyze the behaviour of FPIs by major investing countries, the study finds evidence supporting the existence of herd behavior among the foreign investors in the Malaysian capital market. The existence of herd behaviour among the foreign investors is not only evident in the outflows of capital but the inflows of capital as well.

Based on a country-by-country analysis, the study finds that FPI inflows from the US is highly sensitive to changes in inflows from other major investing countries, suggesting the evidence of herd behaviour among the US foreign portfolio investors in the Malaysian capital market. On the other hand, shocks in the US FPI inflow itself are insignificant to the other countries FPI inflows, suggesting that the inflows from other investing countries are not based on merely following the behaviour of the US foreign portfolio investor. This finding is further supported by the VDC analysis since a significant proportion of the fluctuations in the US FPI inflows are being explained by the innovations in the other major portfolio investing countries, namely the UK, Singapore and Hong Kong. There is also some evidence of herd behaviour among the UK foreign portfolio investors as shocks and variations in FPI from Hong Kong is significant in accounting for the variations in the UK FPI. As in the previous case, the results support that the US FPI inflow has insignificant role in determining the variation of other countries FPI inflows but FPI inflow from the US is highly determined by FPI inflows from other countries. There is also evidence of herd behaviour in the FPI outflows among the major foreign investing countries. In particular, the study finds significant bi-directional causation running from Singapore to Hong Kong FPIs and uni-directional causation running from UK to US FPIs.

In general, the study finds clear evidence of the existence of herd behavior among the major foreign investors in the Malaysian capital market. This finding implies that, during the period under review, the Malaysian short-term capital market can be volatile as it is influenced more by market "mood". Being a small yet highly open economy, Malaysia could be highly susceptible to the swings in market mood. It is therefore imperative for Malaysia to take the necessary pre-cautionary steps to ensure that an aggregate reversal in portfolio investment flows would not impose a de-stabilising impact on the economy. Learning from the experience during the financial crisis 1997/1998, this study emphasizes the need to improve market supervision and strengthen the financial structure in efforts to reduce the impact of capital reversal on the Malaysian financial market.

6. References

[1] Bacchetta, P. and Wincoop, E.V., (1998), *Capital flows to emerging markets: liberalization, overshooting and volatility*, CEPR Discussion Papers *1889*, C.E.P.R. Discussion Papers.

[2] Baek, In-Mee., (2006), *Portfolio investment flows to Asia and Latin America: pull, push or market sentiment?*, in: *Journal of Asian Economics*, 17:363-373.

[3] Banerjee, Abhijit V., (1992), *A simple model of herd behavior*, *Quarterly Journal of Economics*, Vol. CVII (3): 797-817.

[4] Bank Negara Malaysia. Monthly Statistical Bulletin. *Various issues.*

[5] Calvo, G., Leiderman, L., Reinhart, C. M., (1993), *Capital inflows and real exchange rate appreciation in Latin America.* IMF Staff Papers, 40(1): 108-151.

[6] Calvo, G., Leiderman, L., Reinhart, C. M., (1996), *Inflows of capital to developing countries in the 1990s,* Journal of Economic Perspectives, 10(2): 123-139.

[7] Chuhan, P., Classens, S., and Mamingi, N., (1993), *Equity and bond flows to Latin America and Asia: the role of global and country factors*, Journal of Development Economics, 55: 439-463.

[8] Corsetti, Giancarlo & Pesenti, Paolo and Roubini, Nouriel, (1999*), What caused the Asian currency and financial crisis?,* Japan and the World Economy*,* Elsevier, 11(3): 305-373.

[9] Dickey, D. A. and Fuller, W. F., (1979), *Distribution of the estimates for autoregressive time series with a unit root*, Journal of the American Statistical Association, 74: 427-431.

[10] Eichengreen, B. and Mody, A., (1988), *What explains changing spreads on emerging market debt: fundamentals or market sentiment?,* NBER Working Paper, No. 6408, National Bureau of Economic Research.

[11] Fernandez-Arias, E., (1996), *The new wave of private capital inflows: push or pull ?,* Journal of Development Economics, 48(2): 389-418.

[12] Johansen, Søren, (1991), *Estimation and hypothesis testing of cointegrating vectors in Gaussian vector autoregressive models*. Econometrica, 59, no. 006: 1551-1580

[13] Johansen, Søren, (1998), *Statistical analysis of cointegration vector*, Journal of Economics Dynamics and Control. 12: 231-254.

[14] Johansen, Søren and Juselius, Katarina, (1990), *Maximum likelihood estimation and inference on cointegration: with applications to the demand for money*. Oxford Bulletin of Economics and Statistics, 52, 169-210.

[15] Johansen, Søren & Juselius, Katarina, (1992), *Testing structural hypotheses in a multivariate cointegration analysis of the PPP and the UIP for UK*. Journal of Econometrics, 53: 211-244.

[16] Kaminsky, Graciela, (1998), *Currency and banking crises: the early warnings of distress*, International Finance Discussion Papers 629, Board of Governors of the Federal Reserve System, October.

[17] Kaminsky, Graciela L. and Schmuckler, Sergio L., (1999), *What triggers market jitters?* The World Bank Policy Research Working Paper No. 2094.

[18] Kanioura, Athina, (2005), *A cointegration analysis of US interest rates*, Working Paper no. 3. Department of Economics, University of Sheffield, 2001

[19] Kim, Yoonbai, (2000), *Causes of capital flows in developing countries*, Journal of International Money and Finance, Elsevier, 19: 235-253.

[20] Kumar, M. S. and Persaud, A., (2002), *Pure contagion and investors shifting risk appetite: Analytical issues and empirical evidence*, International Finance, 3(5): 401-436.

[21] Krugman, Paul, (1979), *A model of balance-of-payments crises,* Journal of Money, Credit and Banking, 11: 311-325.

[22] Phillips, P. and Perron, P., (1988), *Testing for a unit root time series regression.* Biometrika. 75(2): 335-346.

[23] Toda, H. Y. and Yamamoto, (1995), *Statistical inference in vector autoregressions with possibly integrated processes*, Journal of Econometrics. 66: 225-250.

LONG-TERM CARE: REGIONAL DISPARITIES IN BELGIUM

Güngör **KARAKAYA***
Université Libre de Bruxelles, SBS-EM, DULBEA, **Belgium**
gungor.karakaya@ulb.ac.be

Abstract:
In this paper we analyze the problem of population ageing in terms of non-medical care needs of persons who are dependent or have lost their autonomy, in order to provide the various public and private administrations active in these fields with some food for thought. The anticipated increase in dependency poses significant challenges in terms of needs evolution and financing. Using administrative data on the Belgian population to build indicators on the prevalence of dependency at home in the three regions in 2001, we find that the likelihood of a sustained increase in the Flemish prevalence rates ultimately amplifies the magnitude of the financing problems that the Flemish dependency insurance scheme has experienced since its first years of operation. Results also show that the smaller increases or the decreases (according to the scenario selected) expected in Wallonia and Brussels are likely to mitigate concern about the sustainability of any long-term care insurance in Wallonia and therefore to facilitate its eventual introduction.

Keywords: Long-term care, Old age assistance, Demographic changes, Regional inequalities, Projection

JEL Classification: I12, I18, J11, J14

1. Introduction

In Belgium, the conditions governing community support for dependency and the situation in terms of dependency insurance are not uniform among the three regions of the country[1]. Some benefits associated with dependency are covered at the federal level, such as the allocation of aid to the elderly. Attempts to introduce federal dependency insurance have been hampered by the debate around the jurisdiction for assistance to the elderly, which results in differentiated situations between regions. Flanders has its own insurance (*Vlaamse Zorgverzekering*), which began to provide benefits in 2002[2]. Affiliation to this system is mandatory for anyone living in the Flemish Region. The Brussels Region being officially bilingual, access to the Flemish dependency insurance scheme is optional for inhabitants of Brussels. No social insurance system covering specific needs for dependency-related non-medical assistance is in force in the French or German-speaking parts of the country.

Within the institutional framework of Belgium, the prospect of a differentiated evolution of dependency linked to demographic differences between regions could have consequences for the federal entities. On the one hand, ageing may potentially jeopardize the budgetary balance of the Flemish care insurance scheme, the only existing at the moment. On the other hand, an excessive increase in needs without adequate coverage would mean an unbearable increase in financial and non-financial burdens for the dependants. In order to provide an adequate response to this increase in responsibility, the possible introduction of dependency insurance where it does not exist should be considered only if it is viable in the long term. Overly pessimistic forecasts as to the future budgetary costs could deter the government from embarking upon the introduction of dependency insurance, so it is important to examine the impact of alternative scenarios regarding the evolution of dependency.

This paper attempts to assess the extent of dependency at home in all three Belgian regions using indicators based on administrative data, similar to those used in the Flemish dependency insurance scheme, and presents long-term development prospects up to 2050.

---

* This paper is a revised extension of Karakaya *et al*. (2006).

[1] For different views on dependency insurance in Belgium, see Cantillon (2004), Breda (2004), Vansteenkiste (2004), Ruz Torres (2004), Jousten (2004) and Pacolet (2004) in a special issue of *Revue Belge de Sécurité Sociale (Belgian Review of Social Security)*.

[2] Decree of 30 March 1999 (published in *the Moniteur Belge* on May 28, 1999) on the organization of Flemish care insurance ("*Vlaamse zorgverzekering*").

This study is structured as follows. Section 2 gives a brief overview of the literature regarding the evolution of dependency. Section 3 presents the problem of quantification of dependants on the basis of indicators constructed using administrative data. Then, the methodology for projecting these indicators up to 2050 and the presentation of results constitute the fourth section. Section 5 attempts to assess the budgetary cost of dependency insurance in Wallonia. Finally, section 6 includes some comments on the results and their long-term consequences.

2. Overview of the literature

Dependency insurance, when it exists, provides long-term care services to people who are limited in their ability to function independently for their daily activities at home or in an institution. These services can be financial or material assistances. The long-term care needs are most prevalent for the oldest age groups who are most at risk of long-standing chronic conditions causing physical or mental disability (impairment, activity limitation and participation restriction).

The simplest approach to projecting dependency is to make the assumption that dependency prevalence rates by age are constant over time. Given the ageing population, empirical studies using this methodology indicate a significant increase in dependency. However, it represents only one of the approaches that are found in the demographic literature, in which we can identify three main theories about the evolution of health status and/or dependency in relation to the lengthening of lifespans [Robine *et al.* 1991].

The "*expansion of morbidity*" *theory* [Gruenberg, (1977), Kramer, (1980)] attributes the drop in mortality observed to a decline in the lethality of chronic diseases rather than to a slowdown in their growth rate. A decrease in mortality or its deferral to an older age would lead to much more severe chronic disease. Therefore, life expectancy in dependency increases and the prevalence rates of dependency at every age are relatively high.

The "*compression of morbidity*" *theory* [Fries, (1980), (1989)] points out that the onset of chronic disease is occurring increasingly late in life, postponing morbidity to older ages (end of life) thereby reducing the duration of dependency. This evolution, it is said, results from medical and technological advancements that increase the healthy lifespan while reducing the number of years spent in dependency, which means a decrease in the dependency rates by age.

According to the "*dynamic equilibrium*" *theory* that combines both the expansion and compression scenarios [Manton, (1982), Manton *et al.*, (1997)], the decline in mortality is in particular attributable to a slowdown in the progression rate of chronic diseases with less severe chronic disease and disability conditions. Life expectancy in dependency therefore remains at a relatively constant level. The prevalence rate of dependency at every age decreases at a rate similar to the projected increases in life expectancy.

The recent economic literature also takes into account the prospects of a compression of morbidity. In a review of the economic literature on long-term care, Norton (2000) notes that the economic burden related to the ageing population and the increasing need for long-term care could be significantly less onerous that might be believed at first sight. The magnitude of the problem depends on the interrelationship between lifetime and expenditure.

Lakdawalla and Philipson (2002) incorporate these perspectives into a theoretical model. They assume that the number of dependants residing in institutions has risen more slowly than the population of the United States. In their model, the increase in life expectancy may have as a result, at the macroeconomic level, a reduction in the (medical and non-medical) costs of long-term care. Two factors are involved in this finding, paradoxical at first glance. On the one hand, longer lifespans in good health could increase the supply of care (provided by healthy people, even older ones, including healthy spouses) and therefore reduce the price. On the other hand, the narrowing of the gap in the life expectancy of men and women could have the effect of reducing the number of widows whose isolation eventually leads them to enter an institution.

Various empirical studies have tried to take into account the potential effects of these various scenarios in Europe. Jacobzone (2000) presents projections, up to 2020, of the dependency rate for different OECD countries. Two methodologies are adopted to count dependants. The static approach combines demographic projections with the latest known prevalence figures using constant rates of disability or institutionalization by age (pessimistic scenario). The dynamic approach is to project the trend of declining disability and institutionalization rates (optimistic scenario). Extrapolations are based on data for the year 1995 and data before 1995. Jacobzone asserts that the number of dependent elderly people correlates with the evolution of the elderly population in a state of severe disability. The increase in the number of dependent elderly people is expected to be between 43% and 61% in Canada and between 25% and 43% in France according to this scenario.

In France, Bontout *et al.* (2002) have prepared projections, up to 2040, of the number of dependants aged 60 and over, on the basis of prevalence rates by sex and age from the survey *Handicaps-Incapacités-Dépendance (HID, INSEE)* and demographic projections by the *INSEE*. Data from *HID* come from surveys conducted with couples living at home and couples living in institutions. Dependency is measured using the indicators *AGGIR (Autonomie Gérontologique Groupe Iso-Ressources)* and *EHPA (Etablissements d'Hébergement Pour Personnes Agées)*. The projections are predicated on three dependency evolution scenarios based on trends observed in the past. The pessimistic scenario assumes that increases in life expectancy without dependency evolve in parallel with general life expectancy. The dependency prevalence rates then fall at the same rate as the increases in projected life expectancy. The optimistic scenario is based on the assumption that dependency prevalence rates decrease at a rate similar to what was observed during the 90s, i.e. at a faster pace than the gains in life expectancy. In the central scenario, dependency prevalence rates decrease at a slower rate than in the optimistic scenario. Whatever the scenario, the ageing of the French population will likely increase the number of dependants aged 60 and over by 2040, to an extent which varies markedly according to the dependency scenario. Over the forty years, that increase was 35% for the optimistic scenario, 55% for the central scenario and 80% for the pessimistic scenario. In addition, the increase in dependency will be greatest among people of more than 80 years old.

The projections made by Duée and Rebillard (2004) are based on the same sources as those of Bontout *et al.* (2002). They implement three comparable scenarios to project the number of dependants aged 60 and older by 2040. Duée and Rebillard, however, use two successive waves of data from the *HID* survey, which allows them to simulate transitions in or out of dependency using the *Destinie* microsimulation model. Over the forty years, the increasing number of dependants over the age of 60 would be approximately 18% for the optimistic scenario, 43% for the central scenario and 84% for the pessimistic scenario. These rates are lower than those of Bontout *et al.* (2002) for the central and optimistic scenarios. Although the method based on dependency transitions should in theory allow the use of more precise information, the data that it draws on are less robust than when using prevalence. Ultimately, if we are to believe Le Bouler (2005), the choice of assumptions on the evolution of dependency is far more crucial than the use of either of these techniques.

For Belgium, to our knowledge, the existing projections of the number of dependants have been based on the assumption that the dependency rate by age remains constant over time. At the national level, the results obtained by the *Bureau fédéral du Plan* [Mestdagh and Lambrecht, (2003)] on the basis of the dependency rate by age from the *Enquête de Santé par Interview (Health Survey by Interview)* involve very significant long-term changes. Between 2001 and 2050, the increase in the number of people receiving home care would be 124% and that in the number of people residing in institutions 166%.

There are also separate projections for Flanders and the French-speaking part of Belgium. Breda *et al.* (2000) include in their preparatory study for the introduction of the Flemish dependency insurance scheme a projection of the number of dependants in Flanders from two demographic forecasts. They point out that recent studies predict an increase in life expectancy in good health but are confined, however, to a projection based on constant dependency rates. Their results indicate a 58% growth in the number of dependants at home and in institutions between 2000 and 2050. A similar exercise was carried out by Ethgen *et al.* (2003) with regard to the projected number of people aged over 75 in high dependency at home, and of their cost for the French and German-speaking parts of the country. In all the cases considered, the cumulative reserves curve presents a parabolic aspect, which predicts a lack of financing related to the ageing population sooner or later. This unfavourable picture is highly determined by the assumption that dependency rates by age will be constant over time. The differences in the methodology, data and definition of dependency used make it difficult to compare the dependency rates obtained from these two studies.

The regional dimension of dependency in Belgium was also considered indirectly by studies on life expectancy in good health. Van Oyen *et al.* (1996) studied the inequalities between the Belgian regions in terms of general life expectancy and life expectancy in good health during the years 1989-1990. Life expectancy in good health is evaluated on the basis of subjective measures of health status calculated using a sample of 2,640 people, randomly selected, who are asked to describe their state of health. The interviewees are considered healthy if they answer that their health is very good, good or fair. In a more recent work, Van Oyen *et al.* (2002) measure health status according to three indicators of life expectancy in good health obtained on the basis of the *Health Survey by Interview* of 1997: a subjective indicator of health status; an indicator based on the physical and functional limitations resulting from chronic disease; and an indicator of life expectancy in good mental health based on mental health and psycho-affective problems such as anxiety, psychological illness and depression.

	Flemish region				Walloon region			
Age	**GLE**	**HLE**	**DFLE**	**MLE**	**GLE**	**HLE**	**DFLE**	**MLE**
Males								
15	(58.9)	(56.5)			(56.6)	(50.2)		
25	50.42	39.47	39.11	38.55	48.07	34.19	36.97	34.22
	(49.5)	(47.3)			(47.3)	(41.1)		
35	40.92	30.55	30.08	31.87	38.77	25.48	28.15	27.46
	(40.0)	(37.9)			(37.9)	(32.0)		
45	31.57	22.19	21.01	25.10	29.71	18.15	20.00	21.77
	(30.6)	(29.3)			(28.9)	(23.4)		
55	22.78	14.48	13.26	18.03	21.29	11.69	12.40	16.09
	(21.9)	(20.5)			(20.4)	(16.1)		
65	14.93	8.55	7.07	12.20	13.90	7.62	7.14	10.15
	(14.3)	(13.3)			(13.2)	(9.2)		
75	8.70	4.31	2.80	7.19	8.18	3.80	3.24	5.78
	(8.5)	(7.2)			(8.0)	(3.5)		
85	4.42	2.27	0.57	3.95	4.22	2.65	0.98	1.18
Females								
15	(65.2)	(61.3)			(63.9)	(58.1)		
25	56.06	39.82	40.35	39.10	54.96	33.48	36.65	32.76
	(55.5)	(52.0)			(54.1)	(48.7)		
35	46.28	30.87	31.05	32.11	45.27	25.39	27.98	26.95
	(45.7)	(42.4)			(44.5)	(39.6)		
45	36.74	23.67	22.57	26.05	35.86	18.55	20.30	21.44
	(36.2)	(33.3)			(35.1)	(30.7)		
55	27.56	16.36	14.70	19.58	26.85	12.62	12.44	16.15
	(27.1)	(24.3)			(26.0)	(22.1)		
65	18.85	9.96	8.18	12.89	18.36	6.97	6.26	10.89
	(18.5)	(16.0)			(17.7)	(14.3)		
75	11.18	5.00	3.65	7.45	10.86	3.76	2.13	6.97
	(11.2)	(9.1)			(10.6)	(8.2)		
85	5.36	2.73	1.03	4.33	5.18	1.66	0.73	2.28

Table 1. General life expectancy (GLE), healthy life expectancy (HLE), disability-free life expectancy (DFLE) and mental life expectancy (MLE) by region in Belgium, 1997 (1989-1990)

Source: Van Oyen *et al.* (1996) and Van Oyen *et al.* (2002). *Note*: Results for 1989-1990 in parentheses.

As shown in Table 1, both studies obtain general life expectancies and expectations of healthy life in the Flemish Region vastly superior to those of the Walloon Region, for men as well as for women. According to Van Oyen *et al.* (1996), general life expectancy and healthy life expectancy among 15-year-old men amount to 58.9 years and 56.5 years respectively for the Flemish Region, but only 56.6 years and 50.2 years for the Walloon Region. For men of 65, they correspond to 14.3 years and 13.3 years respectively for the Flemish Region, and only 13.2 years and 9.2 years for the Walloon Region. The life expectancies of women also reveal discrepancies between regions. In Flanders, about

3.3% of men and 4.3% of women answered that they suffer bad or very bad health, while these rates are 9.4% and 7.7% respectively for Wallonia[3].

The results for 1997 in Table 1 go in the same direction [Van Oyen *et al.*, (2002)]. The gap in general life expectancy and life expectancy without disability between Flanders and Wallonia is approximately 2 years in favour of Flemings, for 25-year-old men. For healthy life expectancy and life expectancy in good mental health, this difference rises to more than 4 years. For people older than 65, the differences in life expectancy become blurred and in some cases favor Wallonia. For 25-year-old women, the differences between regions are 1 year for general life expectancy, 4 years for life expectancy without disability, and 6 years for healthy life expectancy and mental life expectancy, again to the advantage of the Flemish Region. At age 65, the differences between regions fall to 0.5 years for general life expectancy, 2 years for life expectancy without disability and mental life expectancy, and 3 years for healthy life expectancy.

3. Assessment of dependency at home in 2001

In order to assess dependency in the different regions, we use an approach based on administrative data. The program will consist in developing various indicators for people likely to be supported/assisted by dependency insurance through data/information similar to those used by the Flemish dependency insurance scheme.

3.1. Difficulties in measuring dependency

The evaluation of assistance requirements has a multidimensional aspect. Ideally, it should take into account not only physical but also psychological dependency, as well as its ability or that of its environment to accept the consequences of dependency. The complexity of measuring dependency and the lack of consensus as to its definition have prompted the development of a multitude of dependency indicators in the gerontological literature. In France, for instance, many indicators have been developed to assess dependency.

In Belgium, the available indicators of dependency are by no means so well developed. The *Enquête de Santé par Interview (Health Survey by Interview)* provides a set of indicators specifically geared to dependency: subjective health, mental health, limitations in daily life activities and social health. However, the data specifically geared to dependency are available only for a sample of individuals and could not be used to serve as a basis for a dependency insurance system.

The lack of data directly on dependency for the entire population and the concern to minimize the cost and time required for the introduction of long-term care insurance led the Flemish Government to base itself on existing administrative data to establish the severity and duration of dependency. Most of the criteria used since the introduction of the measure are based on the presence of one or more certificates issued under various federal or regional aid programs. The following criteria have been used for dependency at home:

- Katz scale revised for the *INAMI (Institut National d'Assurance Maladie-Invalidité)* in the context of home nursing care (minimum score of B);
- BEL-profile scale for family aid (minimum score of 35 on the BEL-profile scale);
- medico-social scale used to assess the degree of autonomy for the following benefits: integration, aid for the elderly and third-party assistance (minimum score of 15 on the medico-social scale);
- supplementary family allowances (for disabled children who suffer from a physical and/or mental incapacity of at least 66% and have at least 7 points as a degree of autonomy);
- supplementary family allowances (for at least 18 points on the medico-social scale) from 1 May 2003;
- request for attendance in a day or short-stay centre (a score of at least C on the rating scale justifying the request).

For dependency in an institution, from 1 July 2002, a minimum score of C on the assessment scale for residential care or a certificate of residence in a psychiatric care institution. From 1 January 2003, the procedure has been extended to include scores B and, from 1 January 2004, scores O and A.

[3] These results are not presented in Table 1.

This pragmatic solution reveals several limitations to measuring dependency.

a. Problems of validity of content, classification, interpretation and heterogeneity of the measures used were reported [Swine *et al.*, (2003)]. The indicators used, often developed from the Katz index for use in the context of a particular measure, have sometimes suffered further processing to allow their use in another context. Other indicators, such as the one used for the allocation of the "aid for the elderly" grant, combine daily life activities and activities instrumental in daily living by using items defined in a very broad way and combining ordinal variables in an additive way. As these authors point out, *[i]t is interesting to note that all of the existing financing system of care for dependants is based on an incomplete scientific approach* [Swine *et al.*, (2003), p. 11]. Until recently, no evaluation of the dependency indicators used (or likely to be) has been specifically carried out for Belgium, particularly with regard to their sensitivity to the evaluator.

b. Although these indicators were defined in a more scientifically rigorous way, it should be noted that most of them are related to medical care, while the objective here is to measure non-medical needs. However, people's need of assistance is not strictly related to functional limitations. As Davin *et al.* (2005) recently showed, they are often related to the socio-environmental context.

c. A significant interaction between dependency insurance based on these measures and the level of federal jurisdiction should also be emphasized. Thus, an easing of conditions in order to benefit from one or the other of these federal measures will directly lead to an increase in the number of people eligible for dependency insurance. This insurance intervenes in a way as a complement to other federal and regional aids.

d. The quantification of dependency is based on the presence of certificates. Some people may have a certificate proving entitlement to benefits taken into account in the construction of indicators and may no longer need it. On the other hand, persons in a position to obtain certificates giving the right to care but having not applied for it might be tempted to do so if aid were to be granted to the beneficiaries of these services.

e. Some of these measures reflect the use of services and can thus be influenced by the supply of these services. So, the number of home care (*Forfaits/packages B or C*) beneficiaries is influenced by the availability of home-assistance services.

3.2. Why rely on administrative data?

Despite the various disadvantages listed in the previous section, given the way the entitlement to benefits for both formal and informal assistance will be decided in the Flemish dependency insurance scheme, it is interesting to examine the results of the calculation of a similar dependency indicator for all three regions of the country. This approach compensates for the lack of data in assessing the number of dependants [Ruz Torres, Karakaya and Plasman, (2002), Ruz Torres, (2004)].

The great advantage of these administrative data is that they cover all of the Belgian population and not only a sample. Moreover, the use of data similar to those applied in Flanders helps us to evaluate to what extent this type of indicator can identify dependency. Finally, the use of a coherent database helps avoid double counting that might occur when calculating an index based on disparate sources.

Given the above-mentioned limitations, the regional differences reflected could be due in part not to people's health status and needs but to differences in applying or checking these criteria. It is worth recalling here the potential problems associated with the lack of scientific validation of most indicators, and the potential influence of supply. Since the effect of these limitations on the evaluation of the scale of dependency may vary from one region to another, we should be cautious when interpreting the results as reflecting dependency differentials between regions.

However, the weaknesses of our approach more or less faithfully reflect those of any system in which recognition of dependency were to be based on administrative criteria. The results for 2001 represent a quantification of people with some indicators expected to approximate the degree of dependency. In the event that such a system were introduced by federal entities other than the Flemish Region, the same problems would arise. Moreover, the use of data related to 2001, prior to the initiation of the Flemish long-term care insurance scheme, exclude the possibility that the regional differentials recorded may result from an entailment effect directly linked to the introduction of dependency insurance in Flanders.

3.3. Description of data

The construction of the dependency indicators is based on administrative data aggregated by mutual companies (*mutualité*) relating to the year 2001. They include in particular variables determining the right to the social franchise and chronic disease packages. Both measures are aimed to help people in precarious social situations and/or requiring recurrent medical care in the context of federal health care insurance.

These data are broken down by age, sex and region. The variables used are as follows:

- beneficiaries of home nursing care (*Forfaits/packages B and C*);
- beneficiaries of physiotherapy packages *E*;
- beneficiaries of supplementary family allowances;
- beneficiaries fulfilling the conditions for the granting of the integration allowance in cat. III or IV (twelve points or more);
- beneficiaries fulfilling the conditions for the granting of the "aid for the elderly" allowance in cat. II, III or IV (twelve points or more);
- beneficiaries of the allowance for third-party assistance (allowances for the handicapped);
- beneficiaries of a disability benefit (permanent or not) paid to the disabled person who can be regarded as having a dependent person (needing the assistance of a third person);
- beneficiaries of the lump-sum allowance for third-party assistance;
- beneficiaries of an allowance for the handicapped.

These variables include fairly similar concepts to criteria for entitlement to assistance from the Flemish care insurance scheme. Their definition is, however, often less restrictive. For the aid to the elderly allowance and the integration allowance, the Flemish criteria require a minimum of 15 points on a revised Katz scale, while the indicators available concern those who receive more than 12 points. Similarly, entitlement to the supplementary allowance does not take into account the more restrictive conditions on handicap laid down in the Flemish provisions. The physiotherapy package *E* is not among the criteria considered for Flemish dependency insurance.

It is also worth noting that the data used do not give an indication either on the BEL scale for home care (a scale also featuring among the criteria taken into account for Flemish long-term care) or on stays in institutions.

3.4. Evaluation of dependency at home in 2001

Different combinations of criteria deemed relevant and appearing in the available administrative data enable the construction of alternative indicators of dependency. According to the choice of more or less strict criteria, the estimated number of persons considered in a situation of dependency varies widely. Several indicators quantifying dependants have been calculated using various combinations of the above-mentioned variables. The two indicators giving the most extreme results are:

- *the most restrictive* indicator, taking into account only beneficiaries of home nursing care (packages B and/or C). These categories include in principle the most heavily dependent people not in an institution, but they are also the ones that present the greatest potential problems of comparability between regions.
- *the least restrictive* indicator, incorporating all the criteria that we have, offering as wide as possible an image of dependency at home. People are considered dependent if they satisfy at least one criterion related to the above-mentioned social franchise[4] or chronic disease packages. This indicator also includes beneficiaries of the physiotherapy package *E*, not taken into account for Flemish dependency insurance.

The estimated national prevalence rate at all ages varies from 3.43% according to the least restrictive indicator to only 0.36% according to the most restrictive indicator.

Our comments focus on the prevalence rates (ratio between people considered dependent and the total population) calculated by means of the least restrictive indicator. We present prevalence rates

[4] Such as allowances for the handicapped and supplementary family allowances.

rather than life expectancies in dependency and in good health, since these rates, combined with population growth by age, determine the number of potentially dependent people projected later in this paper.

Furthermore, prevalence rates give an (admittedly imperfect) idea of the burden of dependency compared to the number of potential carers able to look after dependent people through the financing of the system or informal aid. Informal care is often provided by relatives or friends, and in nearly half of all cases the main carer is the spouse (a woman in two-thirds of cases). In only a third of cases, aid is given by the children [Bontout *et al.* (2002)]. A more precise approach relating to informal carers is proposed in section 4.3.

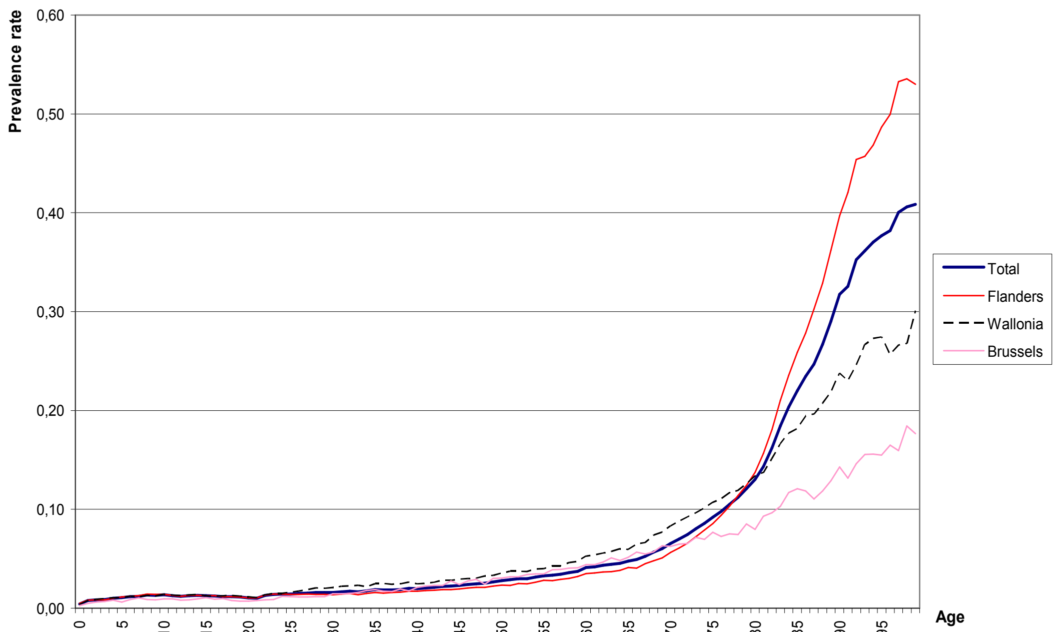


Figure 1. Prevalence rates of the least restrictive dependency indicator by region and age (in 2001).

Source: Calculations based on data provided by the Belgian "*mutualités*".

Figure 1 compares the prevalence rates relating to the three administrative regions of Belgium. For people under the age of 80, the Walloon prevalence rates are higher than those of the other two regions. Between 30 and 72 years, the Flemish rates are lower than those of the Region of Brussels-Capital. From the age of 80, the Flemish rates become far higher than those of the other two regions. These differences between regions increase with increasing age.

Given the above-mentioned limitations, caution is necessary in interpreting the data observed, notably in relation with the supply effect of assistance services. The image of dependency reflected by the indicator used seems to indicate lower risks of becoming dependent in the regions of Wallonia and Brussels than in Flanders. The fact that the Region of Brussels-Capital records the lowest prevalence rates could suggest that it represents the Belgian region least affected by dependency. This may seem surprising, given the data from the *Belgian Health Survey by Interview (Enquête de Santé par Interview)* in 2001. For most of the indicators related to dependency and calculated through this survey, the gap between Brussels and the other regions is far smaller than the one that appears in our administrative data. According to the data on applications for official recognition of invalidity/disability, people with a pending application and persons who have not applied (whereas they would be in situation to do so) are relatively numerous in Brussels. This is a possible explanation for the underestimation of Brussels prevalence rates calculated on the basis of certificates, and it illustrates a deficiency in the administrative data for measuring dependency. The effects of this limitation inherent in the indicators used, however, appear much less marked in the other two regions. Moreover, the limitations to measure dependency (as above-mentioned) and in particular the lack of socio-economic and environmental parameters to estimate dependency are more pronounced in Brussels than in Flanders and Wallonia. Indeed, psycho-affective problems such as anxiety, psychological illness and depression are more present in urban-type sites [Van Oyen *et al.*, (2002)]. As a result, the non-consideration of socio-economic and environmental aspects contributes to explain the low prevalence rates calculated for Brussels.

In our results on the basis of administrative data, life expectancy in dependency is higher in Flanders than in Wallonia or Brussels in 2001. According to Van Oyen *et al.* (1996) and Van Oyen *et al.* (2002) as shown in Table 1, both the general life expectancy and healthy life expectancy of Walloons, calculated by subjective means, are lower than those of Flemings. The difference between the two values is also much larger in Wallonia, particularly among men.

Several explanations for this apparent contradiction may be advanced. First, life expectancy in good health based on a subjective perception and life expectancy without dependency do not have the same meaning, regardless how dependency is measured. In addition, our dependency indicator is based on a mix of indicators referring for the most part to limitations in daily life activities and in the activities instrumental to daily living. But some physical limitations seem to be much more prevalent in Flanders than in the other regions (limitations concerning dressing and undressing, washing hands and face, cutting and eating food, going to the toilet, urinary continence)[5]. Indeed, Van Oyen *et al.* (2002) also obtain a greater life expectancy in dependency in Flanders from the age of 75 when using an indicator solely based on the functional limitations (see Table 1). Moreover, between 1997 and 2001, in Flanders, there was a relatively larger increase in the functional difficulties reflected by the ten items included in the two *Belgian Health Surveys by Interview*. Our results are clearly less surprising in the light of these findings.

A third point of comparison is the number of dependants at home having actually benefited from the Flemish care insurance scheme. Compared to the beneficiaries in 2003 (i.e. the second year of operation of care insurance), the number of dependants we find is much higher (59%), which is not surprising given that our indicators are much less restrictive than those actually employed in Flanders. However, the comparison of the number of dependants by age allows a reassuring finding about the quality of the indicator presented in this paper: the overestimation occurs mainly in the youngest age groups. The little restrictive nature of our dependency indicators, compared to Flanders, is therefore much more marked in the lower age groups. A first potential explanation is the non-consideration of more restrictive conditions on handicap laid down in the Flemish provisions for the entitlement to the supplementary family allowance. A second potential explanation is the physiotherapy package *E* that is not among the criteria considered for Flemish dependency insurance. Indeed, the elderly beneficiaries of the physiotherapy package *E* often benefit from other supports ("aid for the elderly" allowance, home nursing care, allowance for third-party assistance), while the youngest may well be entitled only to the physiotherapy package *E* without fulfilling the other conditions used for dependency at home. As a consequence, taking into account the condition of physiotherapy package *E* will overestimate the dependency in the lower age groups. For people aged 65 and over, the overestimation rate is only 11%, and for those aged 75 and over the results are very similar: the difference is only 1.7%[6]. For this reason, our presentation will focus on the prevalence rates among the elderly.

4. Projection by 2050 of people in dependency

This section describes the projection by 2050 of dependency rates (at home) calculated for 2001 on the basis of population forecasts. Because of the methodology used, the projections of these rates reflect only the effects of expected demographic changes. They do not take into account the possible increase in demand for formal care directly caused by the introduction of the dependency insurance scheme or the modification of its terms, nor the potential problem of adverse selection

---

[5] Caution is necessary in reading the *Belgian Health Survey by Interview (Enquête de Santé par Interview)* because of the subjective nature of data it collects.

[6] More details are given in section 5.

inherent in the coexistence of competing insurance systems [see for instance Ruz Torres, (2004)]. The projection of the number of dependants does not take into account important aspects such as the change in methods of care.

4.1. Methodology

From the dependency prevalence rates computed for 2001, thanks to the indicators presented in the previous section, life expectancy in dependency by age was calculated for 2001 using Sullivan's method. To this end, we rely on the probabilities of death and life expectancy by age and region for the period 2001-2050 [*Institut National de Statistique* and *Bureau fédéral du Plan*, (2001)].

Once life expectancy in dependency had been determined for 2001, we did it evolve according to four different scenarios. Thanks to these projections of life expectancy in dependency, it is possible to calculate future dependency prevalence rates until 2050.

4.1.1. Survival table and calculation of life expectancy with and without dependency

The method devised by Sullivan (1971) combines the prevalence of disability observed at each age in the population with survival table data to break down life expectancy according to different functional disability states[7]. Basically, the idea is to calculate life expectancy without disability or dependency after deducting the period of disability or dependency from the total lifetime.

The starting point is the survival table, which relates mortality conditions in a given year to a fictitious birth cohort (about 1,000,000 people), assuming that these people would know these specific conditions throughout their life, by deducting a theoretical life expectancy at birth or at each age, which is a reflection of the mortality conditions for that year. These probabilities of death, deriving from projections for 2001, represent the deaths of the year by age relative to the population of this age on 1 January 2001. Life expectancy calculated from the table is then split into life expectancies without and with dependency by means of the Sullivan's method.

Life expectancy without dependency is obtained by subtracting life expectancy in dependency (different in each scenario) from the total life expectancy (coming from demographic projections).

4.1.2. Assumptions about the evolution of life expectancy in dependency

Projections of life expectancy with and without dependency have been conducted based on the situation in 2001 by considering four different scenarios. The first rests on the assumption that the prevalence rates are invariable by age, while the other three consider different assumptions about the evolution of life expectancy in dependency.

Pessimistic scenario A: the prevalence rates by age calculated for 2001 remain constant up to 2050. This scenario assumes that the number of dependants varies in parallel with the total population. It does not take into account the actual downward trend in dependency rates for certain ages and consequently it potentially overestimates the number of dependants. For that reason, it may be regarded as the upper limit in the quantification of dependants. It also provides comparisons with existing projections based on the same hypothesis [Breda *et al.*, (2000); Ethgen *et al.*, (2003); Mestdagh and Lambrecht, (2003)]. This is a scenario of expanding morbidity.

Pessimistic scenario B: a situation where the lifetime spent in dependency changes in parallel with total life expectancy. Total life expectancy and life expectancy in dependency experience identical growth rates. It is a scenario of dynamic equilibrium. In the first scenario, the prevalence rates by age are held constant, while in the second scenario, life expectancy in dependency by age grows at the same rate as total life expectancy and the prevalence rates change accordingly.

Central scenario: life expectancy in dependency remains constant over time. The extra years of life are only years of life without dependency and the emergence of dependency shifts to older ages, which means a decrease in the dependency rates by age. This scenario of morbidity compression is our central scenario. Indeed, Bontout *et al.* (2002) and Duée and Rebillard (2004) cite several authors who report this scenario as the most likely.

Optimistic Scenario: life expectancy in dependency decreases over time. In other words, any increase in life expectancy would be accompanied by a decrease in life expectancy in dependency,

[7] For examples of its application, see Saito *et al.* (1991) or Bossuyt *et al.* (2000).

which would mean a considerable improvement in the health status of the population. We assume that the reduction in life expectancy in dependency between 2001 and 2050 is equivalent to half the general life expectancy growth rate over the same period. The average growth rate of the corresponding life expectancy in dependency is applied to determine the life expectancy in dependency for the intermediate years. This is a scenario of high morbidity compression.

As Bontout *et al.* (2002) note, the credibility of these scenarios depends on the timeframe considered. For instance, decisive progress to limit the development of dementia become more and more probable when we lengthen the projection period, and as a result the credibility of the most optimistic scenarios increases.

4.1.3. Projection methodology for dependency prevalence rates

Once the projections of life expectancy in dependency have been carried out, the dependency prevalence rates are calculated between 2001 and 2050 using the four above-mentioned scenarios.

In the first (pessimistic A) scenario, the prevalence rates from 2001 to 2050 are assumed constant. Therefore, the number of dependants estimated for each year of life is directly obtained by multiplying the prevalence rates by the corresponding population.

The other three scenarios apply different assumptions on the evolution of general life expectancy and life expectancy in dependency (as mentioned above). As the value of life expectancy in dependency by age varies according to the scenario used for the projection, the prevalence rates obtained by age will be different for each scenario.

Figure 2 presents the implications of these scenarios in terms of prevalence rates by age at the end of the period (2050) at national level. Compared to the pessimistic scenario A, the curves relating to the other scenarios also show a reduction in prevalence rates between 2001 and 2050. The decrease in prevalence rates is especially noticeable in the over-60 age groups. This drop is all the more pronounced as the degree of optimism of the scenarios considered increases: it is mostly from the age of 60 that the differences between scenarios count[8].

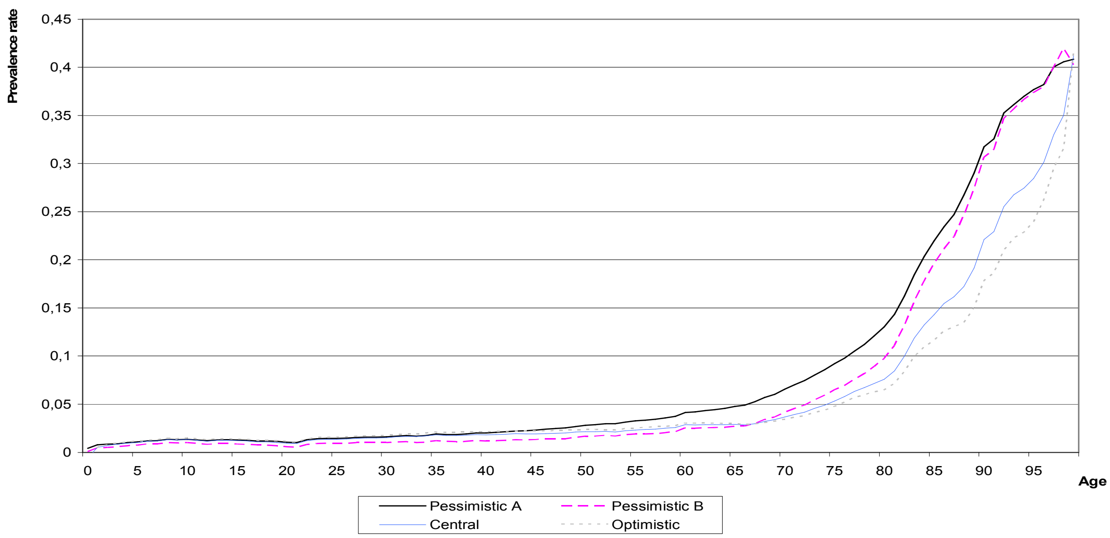


Figure 2. Prevalence rates by age according to different scenarios, for Belgium in 2050.

Source: Calculations based on data provided by the INS-BfP and the Belgian *mutualités*.

[8] The peak observed beyond the age of 95 can be explained by the low number of population for this age group.

4.2. Results of the projection of dependency at home

Projections were made for each year up to 2050 by using several alternative indicators of dependency and the four above-mentioned scenarios. Only the results obtained on the basis of the least restrictive dependency indicator are shown[9]. The presentation focuses on projections of the dependency prevalence rates for the elderly aged over 60 and for people over 75, age groups for which the number of dependants obtained is close to reality and the differences between scenarios are obvious.

4.2.1. Projection of the national prevalence rates

Figure 3 shows a decrease in the national prevalence rates for people aged over 60 until the second half of 2020 for three out of four scenarios, with the exception (of course) of pessimistic scenario A, in which the prevalence rates by age remain hypothetically constant over time. This drop is even more pronounced when one brings in a more optimistic scenario.

The total prevalence rates for people aged over 60 increase in the period from 2030 to 2050, except in the optimistic scenario. This growth is most pronounced in the most pessimistic scenarios. In addition, over the period 2001-2050, the prevalence rates decline only in the central and optimistic scenarios, while they increase in the other two scenarios. According to the scenario, the prevalence rates for Belgium are situated in a range from 67 to 120 dependants over 60 per thousand using the least restrictive indicator.

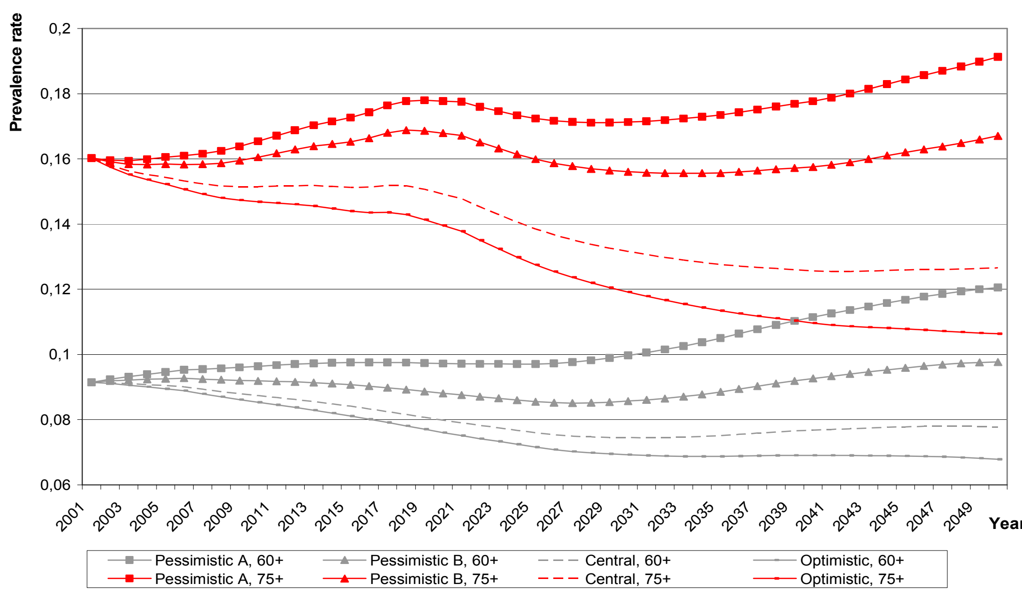


Figure 3. Projections of dependency prevalence rates at national level, for people aged over 60 and over 75 and according to different scenarios.

Source: Calculations based on data provided by the INS-BfP and the Belgian "*mutualités*".

As dependency rises with age, the prevalence rates for those aged over 75 are relatively higher than those of people over 60. During the fifty years surveyed, these prevalence rates vary from 106 to 191 dependants over the age of 75 per thousand in Belgium. The evolution of the prevalence rates for people aged over 75 is parallel to that of people over 60. However, for those over 75, the drop in the prevalence rates occurs from the end of 2010. In addition, from 2001 to 2020, we even see an increase in the rates in the case of the two most pessimistic scenarios[10].

---

[9] In general, the evolutions of the other indicators calculated are relatively similar.

[10] The proportion of dependent people calculated with the most restrictive indicator shows a similar evolution, except that the prevalence rates are, of course, much lower. During the period 2001-2050, they range

4.2.2. Projection of prevalence rates by region

Figure 4 shows the regional prevalence rates obtained on the basis of the central scenario, in which life expectancy in dependency remains unchanged over time. The prevalence rates for people aged over 75 decrease spectacularly from the late 2010 and up to the end of 2030, and then remain relatively constant (or even increase slightly) until 2050. These findings are valid for all three regions.

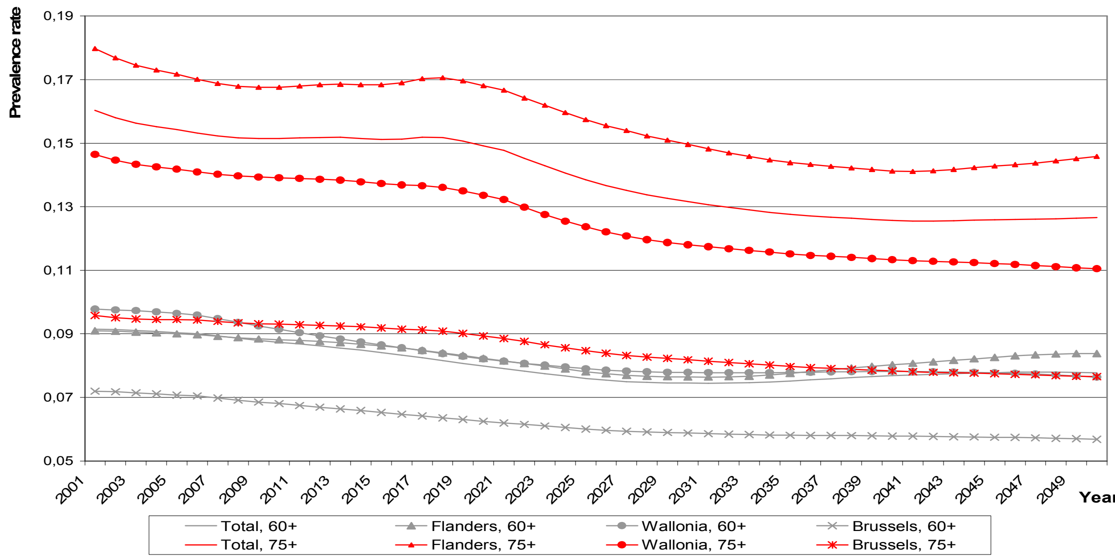


Figure 4. Projections of dependency prevalence rates by region, for people aged over 60 and over 75 using the central scenario.

Source: Calculations based on data provided by the INS-BfP and the Belgian "*mutualités*".

The prevalence rates show a more marked dependency in Flanders at the beginning of the period. The evolution of the Flemish prevalence rates also differs from that of the other regions. From the beginning of 2040, there is an increase in the prevalence rates for Flanders while they decrease for the other two regions. The prevalence rates for Wallonia and Brussels seem, however, to be stabilizing by 2050. If the upward trend of the prevalence rate observed between 2045 and 2050 extended beyond our projection horizon, the Flemish Region could in the very long term experience a substantial increase in the number of dependants as measured by the indicator used. This evolution induced by the increase in the prevalence rates will be all the more pronounced since the older population will grow as well. The prevalence rates for all ages show similar growth: a slight decrease in Brussels and Wallonia and an increase in Flanders (in the case of the central scenario).

For those aged over 75, the differences in the prevalence rates between the regions remain relatively constant throughout the period analyzed (2001-2050). Individuals over the age of 75 living in Flanders are again more likely to become dependent than people of the same age group in the other two regions. We should once again read the Figure 4 keeping in mind the limitations of administrative data, in particular as regards the low prevalence rates recorded in Brussels.

For over 60-year-olds, the evolution of the prevalence rates throughout the period considered is similar to that observed for persons aged 75 and over, with less marked gradients. The national, Flemish and Walloon rates vary in the same direction and are very close to each other. At the beginning of the period, the Walloon rates are even higher than those of Flanders and Belgium as a whole, so that by 2050 the opposite situation prevails: the Flemish rates should be higher than the Walloon and national rates. Once again, Flanders could eventually face a considerable increase in the number of dependants in the long-term.

---

from 10 to 21 dependants over 60 years old per thousand and from 20 to 38 dependents over 75 years old per thousand in Belgium (calculated using the most restrictive indicator).

The differences between the regions do not evolve in the same way according to the scenario used. There is indeed a decreasing difference in the prevalence rates between the regions when using more optimistic scenarios about the evolution of life expectancy in dependency.

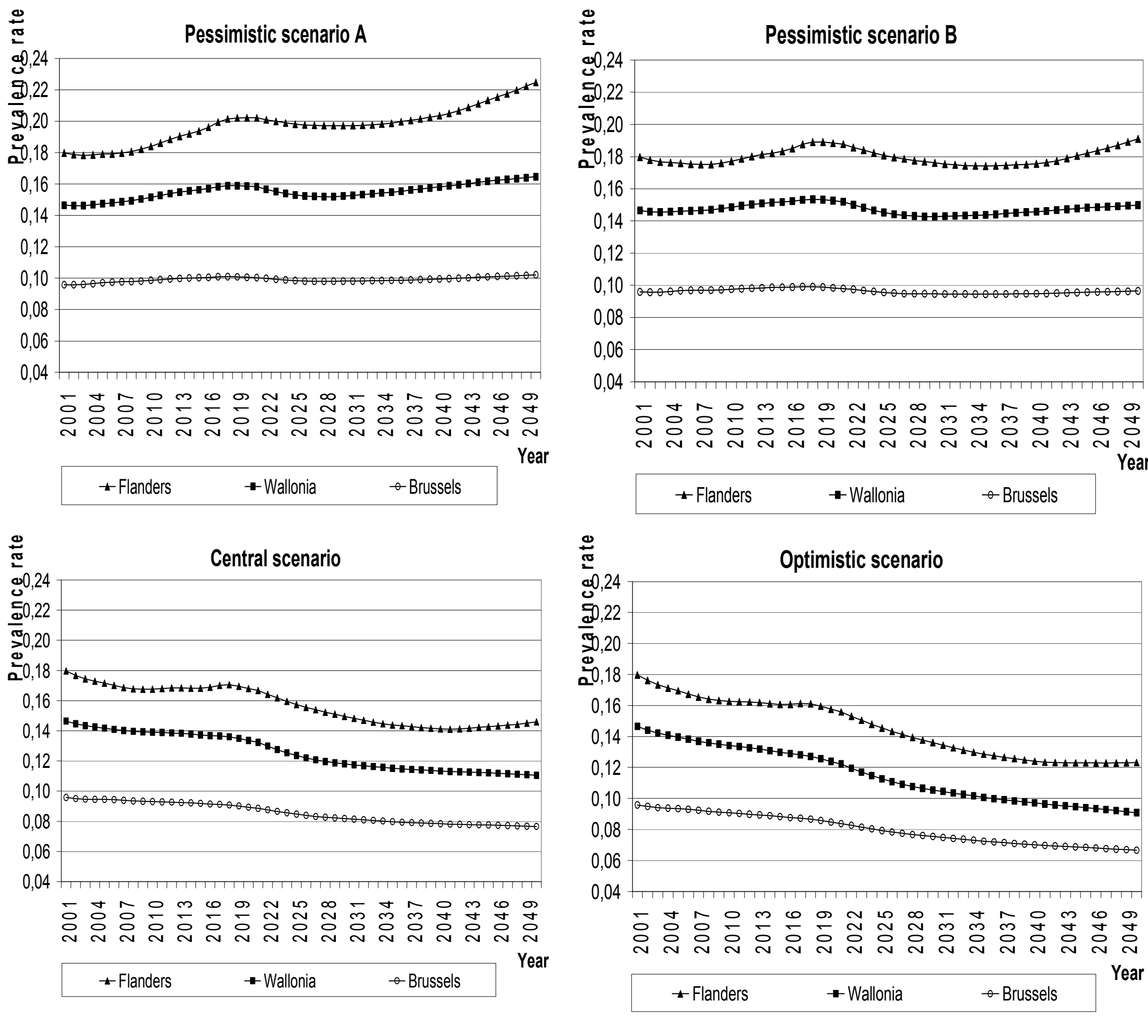


Figure 5. Projections of dependency prevalence rates by scenario and region using the least restrictive indicator and for people aged over 75

For people aged over 75 (Figure 5), the most significant differences in prevalence rates among the regions are obtained using pessimistic scenario A. They diminish over the study period as the degree of optimism of the considered scenarios increases. In other words, whereas the divergences increase significantly with pessimistic scenario A and in a less pronounced way with pessimistic scenario B throughout the period, they remain relatively constant over time under the central scenario and are even reduced in the case of the optimistic scenario. In 2001, the gaps between Wallonia and Flanders are 3.3 percentage points in favour of the Walloon Region. In 2050, they amount to 6 percentage points in the case of pessimistic scenario A, 4.1 percentage points with pessimistic scenario B, 3.5 percentage points with the central scenario and 3.2 percentage points with the optimistic scenario. In the latter case only, the differences between Flanders and Wallonia decline. As regards the inequalities between the Walloon Region and the Region of Brussels-Capital, the prevalence rate gaps, in favour of Brussels, rise from 5 percentage points in 2001 to 6.2 in 2050 for pessimistic scenario A, and to 2.4 percentage points for the optimistic scenario. Thus, the choice of scenario significantly influences the magnitude of the differences between the three regions.

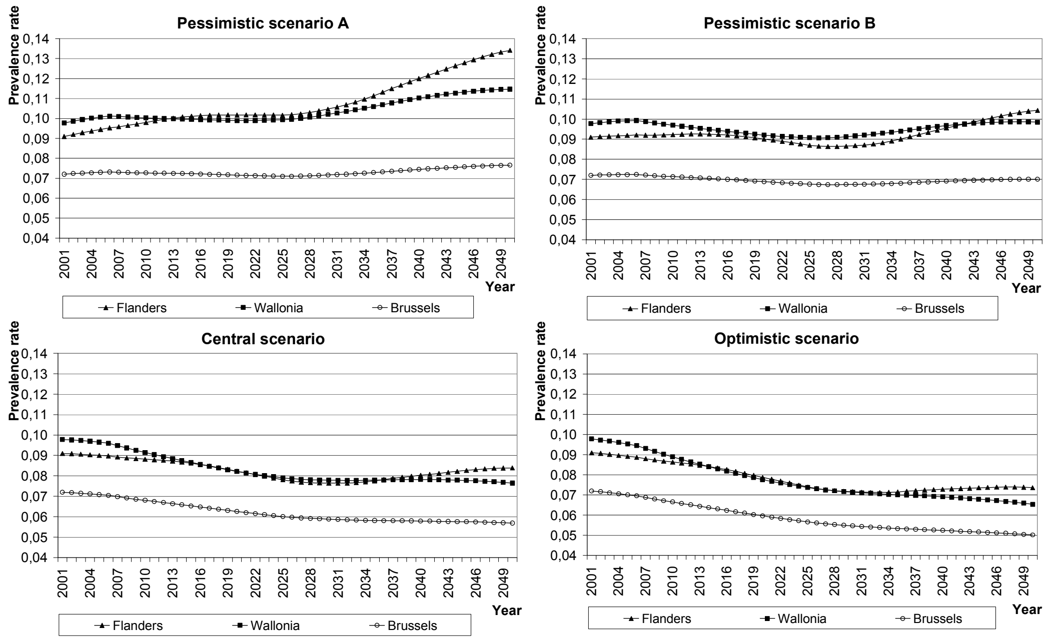


Figure 6. Projections of dependency prevalence rates by scenario and region using the least restrictive indicator and for people aged over 60

According to Figure 6, the situation regarding the gap between regions depending on the optimism degree of scenario does not seem very different for people aged 60 and older. It should be noted that for this age group, the Flemish prevalence rates are initially lower than the Walloon rates but end up exceeding those of the other regions, whichever scenario is used.

4.3. Projection of the number of dependants at home

Table 2 shows the growth rates between 2001 and 2050 in the number of dependants and in the population over the age of 60, over 75 and the total population[11].

Table 2. Evolution of the number of dependants between 2001 and 2050 (growth rates in %)

	Dependants (by scenario)				Population
	Pessimistic A	Pessimistic B	Central	Optimistic	
60+					
Belgium	105.54	67.34	33.64	16.82	57.55
Flanders	127.91	78.22	43.71	26.48	56.22
Wallonia	89.87	63.51	27.35	9.31	63.30
Brussels	54.23	41.45	15.06	1.90	45.73
75+					
Belgium	158.65	126.57	72.51	45.45	115.81
Flanders	186.37	144.38	87.40	58.88	127.73
Wallonia	139.10	118.06	61.66	33.45	112.22
Brussels	72.13	62.77	29.81	13.33	61.45
All age					
Belgium	58.67	21.48	10.64	5.25	6.18
Flanders	73.69	28.44	16.67	10.81	1.70
Wallonia	48.73	18.64	7.20	1.52	11.87
Brussels	32.86	12.31	5.29	1.81	13.98

Source: Calculations based on data provided by the INS-BfP and the Belgian "*mutualités*".

[11] As was the case for all the results presented, the number of dependants is obtained using the least restrictive indicator.

Given the significant variations in the number of dependants obtained according to the indicator used, we do not present the number of dependants obtained, but the growth rates for dependants between 2001 and 2050. Indeed, these growth rates being largely determined by demographic change, they are relatively less sensitive to the definition of the indicator used than to the number of dependants.

We saw that the least restrictive indicator overestimates the number of dependants belonging to the youngest age groups (i.e. below the age of 60)[12]. The effect of demographic changes is thereby attenuated, and we obtain relatively low growth rates for the total number of dependants in the overall population. That is why we have a growth rate for the number of dependants at home of only 58.67% using pessimistic scenario A (assuming the prevalence rates are constant over time), while Mestdagh and Lambrecht (2003) estimate the growth at 124%.

For the 60 and over and 75 and over age groups, however, our indicator gives results much closer to the number of people supported/assisted by the Flemish care insurance scheme. For these two age groups, our growth rates for the number of dependants are situated in the same order of magnitude as those of Mestdagh and Lambrecht (they estimate the growth at 170% and 130% for the 75 and over and 60 and over age groups respectively). Whatever the scenario and the age group selected, the growth rate between 2001 and 2050 in the number of Flemish dependants is higher than that for the other two regions.

The burden of dependency for informal carers can be determined in a more subtle way than with the prevalence rates previously studied, by comparing the evolution in the number of dependants with that of potential carers. As Bontout *et al.* (2002) report, the majority of informal carers consists of people aged between 50 and 79. The population of this age group increases by 23.89% in Belgium between 2001 and 2050, more slowly than the number of elderly dependants (except dependants aged 60 and over for the optimistic scenario). The growth in the number of potential carers, 19.82% for Flanders, 30.01% for Wallonia, and 29.19% for Brussels, accentuates the unfavorable findings for Flanders in terms of the dependency burden in all cases.

5. Assessing the cost of long-term care insurance in Wallonia

The purpose of this section is to calculate the budgetary cost of introducing dependency insurance in Wallonia, beginning at the same time as the *Vlaamse Zorgverzekering* (Flemish insurance), equivalent to the latter and undergoing the same adjustments as those observed in Flanders. In other words, we assume that the scope, the financing and the services supplied by the Walloon Region, and their adaptations, will be identical to the Flemish dependency insurance scheme.

The prevalence rates by region and age group, calculated using administrative data and the least restrictive indicator, are applied to the corresponding population in order to obtain the number of people fulfilling the conditions required to enjoy benefits paid out by the Flemish care insurance scheme. We should remember, however, that these conditions are quite similar, but often less restrictive than in Flanders. From the actual data related to the number of Flemish aid beneficiaries (in 2002, 2003 and 2004), we are able to adapt our results about Flanders to each age group. This "correction" by age group is then applied to the data calculated for the Walloon region (Table 3). This is done by multiplying, for each region and each age group, our estimations of the number of dependants by the ratio between the actual number of Flemish dependants and the number of Flemish dependants estimated on the basis of our prevalence rates. In addition to the evolution of the regional prevalence rates, this approach makes the implicit assumption that the propensity to resort to *MRS*, *MRPA* and *MSP* by age group is the same in Flanders and in Wallonia. It also takes into account the supply effect (insofar as it exists), but also the other factors such as changes, adjustments or amendments to the Flemish legislation observed since the birth of the *Vlaamse Zorgverzekering* until the end of 2004. As previously mentioned on several occasions, the data and criteria used in the calculation of our prevalence rates overestimate the number of dependants in Flanders for the youngest age groups, while this is not the case for others (Table 3). The column "known" shows the number of people actually receiving assistance in Flanders, while the column "before correction" shows the number of aid beneficiaries estimated using administrative data and the least restrictive

[12] More details are given in the next section.

indicator. Finally, the column "after correction" corrects our estimated results for each age group thanks to the actual number of dependants in the Flemish Region (as mentioned above). The correction is then applied to the other two regions.

Table 3. Number of dependants (central scenario)

		31/12/2002	Known	Before correction		After correction			
	Flanders	Flanders	Wallonia	Brussels	Belgium	Flanders	Wallonia	Brussels	Belgium
0-18	2,626	14,791	9,161	1,782	25,734	2,626	1,626	316	4,569
19-25	1,604	6,525	3,905	883	11,312	1,604	960	217	2,781
26-44	6,617	26,390	21,427	5,188	53,006	6,617	5,373	1,301	13,290
45-64	13,329	40,368	32,915	7,468	80,751	13,329	10,868	2,466	26,663
65-69	7,137	13,532	10,147	2,156	25,835	7,137	5,352	1,137	13,626
70-74	11,779	18,431	13,893	2,558	34,882	11,779	8,879	1,635	22,293
75-79	17,940	22,555	14,632	2,666	39,854	17,940	11,638	2,121	31,699
80-84	21,393	23,382	11,843	2,363	37,588	21,393	10,835	2,162	34,390
85-89	19,364	18,661	6,691	1,450	26,802	19,364	6,943	1,505	27,812
90-94	13,892	12,178	3,733	892	16,802	13,892	4,258	1,017	19,168
>=95	4,657	3,340	967	278	4,585	4,657	1,348	387	6,393
0-64	*24,176*	*88,074*	*67,407*	*15,322*	*170,803*	*24,176*	*18,827*	*4,300*	*47,303*
>=65	*96,162*	*112,079*	*61,906*	*12,363*	*186,348*	*96,162*	*49,254*	*9,964*	*155,380*
TOTAL	**120,338**	**200,153**	**129,314**	**27,684**	**357,151**	**120,338**	**68,081**	**14,264**	**202,683**
		31/12/2003	**Known**	**Before correction**		**After correction**			
	Flanders	Flanders	Wallonia	Brussels	Belgium	Flanders	Wallonia	Brussels	Belgium
0-18	2,673	14,680	9,112	1,785	25,577	2,673	1,659	325	4,657
19-25	1,689	6,492	3,901	882	11,276	1,689	1,015	230	2,934
26-44	6,777	26,085	21,210	5,172	52,466	6,777	5,511	1,344	13,631
45-64	13,090	40,448	33,112	7,494	81,055	13,090	10,716	2,425	26,231
65-69	7,055	13,351	9,966	2,123	25,440	7,055	5,266	1,122	13,443
70-74	11,898	18,342	13,527	2,488	34,357	11,898	8,775	1,614	22,287
75-79	18,450	22,401	14,458	2,589	39,448	18,450	11,908	2,132	32,490
80-84	25,486	26,235	13,047	2,573	41,854	25,486	12,675	2,499	40,660
85-89	19,106	16,932	6,145	1,317	24,394	19,106	6,934	1,486	27,526
90-94	15,024	12,532	3,810	907	17,248	15,024	4,567	1,087	20,678
>=95	4,989	3,543	1,019	288	4,850	4,989	1,435	406	6,831
0-64	*24,229*	*87,704*	*67,335*	*15,333*	*170,373*	*24,229*	*18,901*	*4,323*	*47,453*
>=65	*102,008*	*113,336*	*61,971*	*12,285*	*187,592*	*102,008*	*51,560*	*10,347*	*163,914*
TOTAL	**126,237**	**201,040**	**129,307**	**27,618**	**357,965**	**126,237**	**70,460**	**14,670**	**211,367**
		30/09/2004	**Known**	**Before correction**		**After correction**			
	Flanders	Flanders	Wallonia	Brussels	Belgium	Flanders	Wallonia	Brussels	Belgium
0-18	2,781	14,602	9,053	1,785	25,440	2,781	1,724	340	4,845
19-25	1,713	6,419	3,922	881	11,222	1,713	1,047	235	2,995
26-44	7,083	25,773	20,935	5,149	51,857	7,083	5,753	1,415	14,251
45-64	14,908	40,669	33,407	7,535	81,611	14,908	12,246	2,762	29,916
65-69	8,132	13,114	9,805	2,096	25,015	8,132	6,080	1,300	15,512
70-74	13,874	18,187	13,153	2,422	33,762	13,874	10,034	1,847	25,755
75-79	21,551	22,310	14,234	2,507	39,051	21,551	13,749	2,422	37,722
80-84	32,603	27,684	13,698	2,666	44,047	32,603	16,132	3,139	51,874
85-89	23,167	16,723	6,163	1,303	24,189	23,167	8,538	1,805	33,510
90-94	19,315	12,846	3,924	927	17,697	19,315	5,900	1,394	26,609
>=95	6,146	3,688	1,051	296	5,035	6,146	1,751	494	8,390
0-64	*26,485*	*87,464*	*67,317*	*15,349*	*170,130*	*26,485*	*20,770*	*4,752*	*52,007*
>=65	*124,788*	*114,552*	*62,027*	*12,217*	*188,795*	*124,788*	*62,183*	*12,400*	*199,372*
TOTAL	**151,273**	**202,016**	**129,344**	**27,566**	**358,925**	**151,273**	**82,953**	**17,152**	**251,379**

Source: Het Vlaams Zorgfonds. Calculations based on data provided by the *Vlaams Zorgfonds*, the INS-BfP and the Belgian "*mutualités*".

So on 31 December 2002, the number of Flemish dependants observed was 120,338, whereas it amounts to 200,153 in our estimations. The overestimation is mainly concentrated at the level of the lowest age groups (0-64). The correction to our estimations is used to solve the problem resulting from the administrative data and criteria used in the calculation of prevalence rates. These corrections have also focused on 2003 and 2004.

After "correction", we use the average cost of Flemish care insurance for the Walloon Region. The average budgetary cost for Flemish care insurance in period t ($Cm_{F,t}$) corresponds to the ratio between the total Flemish expenditure in t ($D_{F,t}$) and the number of Flemish dependants in t ($N_{F,t}$):

$$Cm_{F,t} = \frac{D_{F,t}}{N_{F,t}} \qquad t = 31/12/2002,\ 31/12/2003 \text{ or } 30/09/2004 \qquad (1)$$

According to the actual Flemish data, the number of files accepted, or dependants, is 120,338 and the total expenditure is € 180,368,000 on 31 December 2002 (we do not take into account transfers to the budget for the following year). Thus, the average cost amounts to approximately € 1,500 (€ 180,368,000/120,338) in 2002. The average cost is about € 1,485 (€ 187,518,000/126,237) and € 1,375 (€ 205,188,000/149,459) on 31 December 2003 and 30 September 2004 respectively. We then apply these average budgetary costs to the Walloon Region in order to determine the total cost of a Walloon long-term care insurance scheme in 2002, 2003 and 2004.

The total budgetary cost of a long-term care insurance scheme in Wallonia in time t ($Ct_{W,t}$) corresponds to the product of the average budgetary cost in t ($Cm_{F,t}$) and the number of dependants in Wallonia in t "after correction" ($\tilde{N}_{W,t}$):

$$Ct_{W,t} = Cm_{F,t} * \tilde{N}_{W,t} \qquad t = 31/12/2002,\ 31/12/2003 \text{ or } 30/09/2004 \qquad (2)$$

The net budgetary cost of a dependency insurance scheme in Wallonia in time t ($Cn_{W,t}$) is equal to its total budgetary cost in t ($Ct_{W,t}$) minus the sum of contributions paid in t by the affiliates of the Walloon care insurance scheme ($R_{W,t}$, i.e. minus the number of affiliates multiplied by the annual personal fee which is € 10 in 2002 and € 25 from 1 January 2003).

Formally, the net budgetary cost is given by:

$$Cn_{W,t} = Ct_{W,t} - R_{W,t} \qquad t = 31/12/2002,\ 31/12/2003 \text{ or } 30/09/2004 \qquad (3)$$

Table 4 shows the total and net budgetary costs of a dependency insurance scheme in Wallonia in 2002, 2003 and 2004.

Table 4. Cost of a long-term care insurance scheme in Wallonia (in euros)

	Scenario	2002	2002	2003	2004
Personal contribution		€ 25	€ 10	€ 25	€ 25
Affiliates		2,294,347	2,294,347	2,303,085	2,311,668
Total of contributions		51,438,471	22,943,470	51,634,374	52,243,801
Total cost	Pessimistic A	102,262,916	102,262,916	105,108,898	120,009,333
	Pessimistic B	102,095,612	102,095,612	104,773,229	118,785,702
Net cost	Pessimistic A	50,824,445	79,319,446	53,474,524	67,765,532
	Pessimistic B	50,657,141	79,152,142	53,138,855	66,541,901

Source: Het Vlaams Zorgfonds and calculations based on data provided by the *Vlaams Zorgfonds*, the INS-BfP and the Belgian "*mutualités*".

In other words, the establishment of dependency insurance in Wallonia at the same time as the *Vlaamse Zorgverzekering* would need to find a new budget requiring a sacrifice amounting to about € 79 million in 2002 with an annual contribution per person of € 10 (as was the case at the beginning of the *Vlaamse Zorgverzekering*), whereas if it were € 25 from the beginning, the net cost would be only € 50 million (i.e. 36% lower). With adaptations to the Walloon long-term care insurance scheme similar to those of the *Vlaamse Zorgverzekering*, the net cost or the budget to be found would reach about 53 million and 67 million in 2003 and 2004 respectively. It should be noted that the increased costs in 2004 mainly result from the extension of Flemish care insurance.

Knowing the estimated number of dependants in the Walloon Region (Table 3), it is also possible to calculate the cost of long-term care insurance in Wallonia with a system more or less generous than the *Vlaamse Zorgverzekering* (in terms of average cost).

6. Conclusion

The lack of data about dependency in Belgium has prompted the Flemish Region to use certificates/attestations from administrative data to grant the right to benefit from Flemish long-term care insurance. The main objective of this study is to assess the extent of dependency at home in the three Belgian regions using similar indicators, based on administrative data, and to obtain forecasts for its long-term evolution in the three regions of the country.

As a first step, we use administrative data on the Belgian population to construct indicators of the prevalence of dependency at home in the three regions in 2001. The results obtained indicate highly differentiated prevalence rates between the regions in 2001. *A priori*, the differences measured between the regions seem not unrealistic in view of the measurement tool selected, but they do not only measure differences in dependency. They must be interpreted in the light of the limitations of the indicator used, which is based on administrative data for the recognition of dependency. The administrative criteria used, especially taking into account the functional limitations, seem to show a better understanding of dependency in Flanders than in the other two regions. In particular, the measurement tool strongly underestimates dependency in the Region of Brussels.

In a second stage, we conduct several projections of the evolution of the prevalence rates obtained from population projections and life expectancy up to 2050. For this purpose, four scenarios are considered: constant prevalence rates by age, life expectancy in dependency increasing in parallel with general life expectancy, constant life expectancy in dependency (central scenario), and decreasing life expectancy in dependency.

With robust growth in the older population, we can expect an increase in the number of heavily dependent people, which presents significant long-term challenges in terms of needs evolution and financing. The magnitude of this effect, important in the scenario of unchanged prevalence rates by age, is significantly reduced when we consider more realistic scenarios.

The implementation of these more realistic scenarios concerning the evolution of morbidity offers more optimistic projections than with constant prevalence rates by age.

When comparing projections of the dependency indicator for the different regions, the quasi-stability of prevalence rates among the elderly in Wallonia and Brussels in the long term contrasts with the projected increase in Flanders. Since the prevalence rates presented for people over the age of 60 and those over 75 give an (imperfect) indication of the relationship between dependants and potential carers, lower rates in Brussels and Wallonia imply that the burden of long-term care might be divided among a larger number of non-dependent people.

However, the differences between the regions are reduced when we take into account more optimistic (and therefore more realistic) scenarios. The evolution of these differences between 2001 and 2050 also varies according to the scenario chosen. Despite these nuances, the different scenarios are in concord with each other: the Flemish Region should expect a more significant evolution in the burden of dependency at home (as measured by the indicator considered) than the other regions. The prevalence rates should even rise in Flanders in the final period, while remaining stable or declining in the other regions. Given that the relationship between the number of dependants and the number of people potentially supporting dependency evolves unfavorably at the end of the period, and perhaps beyond if the trend continues, the burden of long-term care may be more difficult to bear in Flanders in the very long term.

However, we must keep in mind the limitations of the indicator used to measure dependency. In addition, the projection methodology used is mainly based on demographic changes and does not take into account potential changes in behavior, particularly associated with a development in the supply of services, the introduction of Walloon dependency insurance or problems of adverse selection. The interactions with federal measures that provide the basis for the allocation of interventions are not taken into account either.

The prospects of a reduction in the Walloon and Brussels prevalence rates are likely to temper concerns about the sustainability of any long-term care insurance in Wallonia and thus to facilitate its eventual introduction. The increase expected in Flanders ultimately reinforces the magnitude of the financing problems that the Flemish dependency insurance scheme has experienced since its initial years of operation.

These findings will have different consequences according to how much responsibility the dependency insurance organization would have.

A possible second system of long-term care insurance, introduced concurrently with the Flemish insurance scheme and based on administrative criteria, would cover a population with relatively low and declining prevalence rates (resulting from a better situation in terms of dependency). We can therefore imagine that in the very long term, this system will be in a position to offer more generous benefits than the Flemish long-term care insurance scheme, all other things being equal. In Brussels, where the choice between the two systems should in any case remain possible, such a situation could exacerbate the problem of adverse selection.

The differences in dependency measured between the regions appear to decrease when we choose more optimistic evolution scenarios for life expectancy in dependency, but this does not totally compensate for the initial high divergences. It would seem necessary to find an adequate response to the problem of the quality of the dependency indicator so as to take better account of dependency in all three regions. This is particularly important for Wallonia and Brussels, and is seen as a necessary step before the introduction of long-term care insurance in Wallonia although the French and German-speaking parts of the country can not afford to finance such an insurance and they deem this long-term care insurance falls within the competence of Social Security (a federal jurisdiction) contrary to the Flanders that considers this insurance as a part of the assistance policy to the elderly (a regionalized jurisdiction). Or, if we wish to dream, before the establishment of a hypothetical federal system of dependency insurance.

Our findings have also enabled us to draw some conclusions in terms of regional policy on the dependency or loss of autonomy of the elderly.

Thus, the fact that the dependency prevalence rates measured using administrative data are higher in Flanders means that the inhabitants of this region would benefit more than others from a uniform application of long-term care insurance based on the same administrative criteria throughout the country. In the hypothetical case that dependency insurance based on identical administrative criteria were to be introduced at federal level, inequalities in support/assistance between regions could be perceived as unfair in cases where they were mainly determined not by different needs but by imperfections in the measuring instrument. According to this assessment of dependency insurance, establishing a federal system of dependency insurance would be beneficial only for Flanders, to the detriment of Wallonia and Brussels.

7. References

[1] Bontout, O., Colin, C. and Kerjosse, R., (2002), *Personnes âgées dépendantes et aidants potentiels: une projection à l'horizon 2040*, Etudes et résultats, No. 160, DREES (Direction de la recherche, des études, de l'évaluation et des statistiques, Ministère de la santé et de la protection sociale, France).

[2] Bossuyt, N. and Van Oyen, H., (2000), *Espérance de vie en bonne santé selon le statut socio-économique en Belgiqu*e, Institut Scientifique de la Santé Publique, Bruxelles.

[3] Breda, J., (2004), *Pourquoi une assurance-dépendance?*, in: *Revue Belge de Sécurité Sociale*, 1er trimestre 2004, pp. 19-34.

[4] Breda, J., Geerts, J. and Teughels, S., (2000), *Aanvullende expertisenstudie inzake de externe kostprijs van de implementatie van de Zorgverzekering*, UFSIA-Vakgroep Sociologie en Sociaal Beleid, Mei.

[5] Cantillon, B., (2004), *L'indispensable réorientation de l'assurance-dépendance flamande*, in: *Revue Belge de Sécurité Sociale*, 1er trimestre 2004, pp. 9-18.

[6] Davin, B., Paraponaris, A. and Verger, P., (2005), *Facteurs démographiques et socio-économiques associés aux besoins d'aide des personnes âgées vivant à domicile : une étude à partir de l'enquête Handicaps-Incapacités-Dépendance*, in: *Revue d'Epidémiologie et de Santé Publique*, 53, pp. 509-524.

[7] Duée, M. and Rebillard, C., (2004), *La dépendance des personnes âgées : une projection à long terme*, INSEE, Direction des études et synthèses économiques.

[8] Ethgen, O., Gillain, D., Gillet, P., Gosset, C., Jousten, A. and Reginster, J-Y., (2003), *Etude relative à la mise en œuvre des mécanismes d'aides en faveur des personnes en perte d'autonomie. ULg – Rapport final*, Université de Liège, Janvier.

[9] Fries, J.F., (1980), *Aging, natural death and the compression of morbidity*, in: *New England Journal of Medicine*, 303(3), pp. 130-135.

[10] Fries, J.F., (1989), *The compression of morbidity: near or far?*, in: *The Milbank Quarterly*, 67(2), pp. 208-232.

[11] Gisle, L., Buziarsist, J., Van der Heyden, J., Demarest, S., Miermans, P.J., Sartor, F., Van Oyen, H. and Tafforeau, J., (2002), *Enquête de santé par Interview, Belgique 2001*, Institut Scientifique de la Santé Publique - IPH/EPI REPORTS No. 2002-22.

[12] Gruenberg, E.M., (1977), *The failures of success*, Health and society, 55, pp. 3-24.

[13] Het Vlaams Zorgfonds - Begrotingscontrole 2004.

[14] Het Vlaams Zorgfonds (jaarverslag 2003).

[15] Het Vlaams Zorgfonds (jaarverslag 2001-2002).

[16] Het Vlaams Zorgfonds - Situatie per 30 juni 2002 en 31 december 2002.

[17] Het Vlaams Zorgfonds - Statistische analyse, situatie per 30 december 2002.

[18] Het Vlaams Zorgfonds - Statistisch overzicht van de Vlaamse Zorgverzekering per 31 december 2003.

[19] Institut National de Statistique and Bureau fédéral du Plan, (2001), *Démographie mathématique, Perspectives de population 2000-2050 par arrondissement*, Bruxelles.

[20] Jacobzone, S., (2000), *Santé et vieillissement. Les perspectives de prise en charge de la dépendance*, Isuma, 1(2).

[21] Jacobzone, S., Cambois, E. and Robine, J.M., (2000), *La santé des personnes âgées dans les pays de l'OCDE s'améliore-t-elle assez vite pour compenser le vieillissement de la population?*, in: *Revue économique de l'OCDE*, No. 30.

[22] Jousten, A., (2004), *Assurance dépendance en Belgique : réflexions d'un économiste*, in: *Revue Belge de Sécurité Sociale*, 1er trimestre 2004, pp. 93-112.

[23] Karakaya, G., & Ruz Torres, R., (2006), *Inégalités régionales en matière de dépendance en belgique: perspectives à long terme*, in: *Revue Belge de Sécurité Sociale*, 2e trimestre, pp. 185-216.

[24] Kramer, M., (1980), *The rising pandemic of mental disorders and associated chronic diseases and disabilities*, in: *Acta Psychiatrica Scandinavica*, 62, pp. 282-297.

[25] Lakdawalla, D. and Philipson, T., (2002), *The rise in old-age longevity and the market for long-term care*, in: *American Economic Review*, 92(1), pp. 295-306.

[26] Le Bouler, S., (2005), *Prospective des besoins d'hébergement en établissement pour les personnes âgées dépendantes. Premier volet : détermination du nombre de places en 2010, 2015 et 2025*, Rapport de mission, Commissariat général du Plan, Juillet.

[27] Manton, K.G., (1982), *Changing concepts of morbidity and mortality in the elderly population*, in: *Health and society*, 60, pp. 183-244.

[28] Manton, K.G., Corder, L. and Stallard, E., (1997), *Chronic disability trends in elderly United States populations: 1982-1994*, Proceedings of the National Academy of Science, 94(6), pp. 2593-2598.

[29] Mestdagh, J. and Lambrecht, M., (2003), *The AGIR project: Ageing, health and retirement in Europe - Use of health care and nursing care by the elderly: Data for Belgium*, Working Paper No. 11-03, Federal Planning Bureau.

[30] Norton, E.C., (2000), *Long-term care*, in A.J. Cuyler & J.P. Newhouse (Eds.), in: *Handbook of Health Economics*, Vol. 1B, pp. 955-994.

[31] Pacolet, J., (2004), *Discussion sur l'introduction d'une assurance-dépendance dans l'UE*, in: *Revue Belge de Sécurité Sociale*, 1er trimestre 2004, pp. 113-140.

[32] Robine, J-M., Bucquet, D. and Ritchie, K., (1991), *L'espérance de vie sans incapacité, un indicateur de l'évolution des conditions de santé au cours du temps : vingt ans de calcul*, in: *Cahiers québécois de démographie*, 20(2), pp. 205-235.

[33] Ruz Torres, R., (2004), *Aspects économiques d'une assurance-dépendance en Belgique francophone et germanophone*, in: *Revue Belge de Sécurité Sociale*, 1er trimestre 2004, pp. 69-91.

[34] Ruz Torres, R., Karakaya, G. and Plasman, R., (2002), *Etude relative à la mise en œuvre de mécanismes d'aides en faveur des personnes en perte d'autonomie : quantification des bénéficiaires sur base de données administratives*, mimeo, DULBEA, Université libre de Bruxelles, Décembre.

[35] Saito, Y., Crimmins, E.M. and Hayward, M.D., (1991), *Stabilité des estimations de l'espérance de vie sans perte d'autonomie calculées au moyen de deux méthodes de construction de tables de survie*, in: *Cahiers québécois de démographie*, 20(2), pp. 291-327.

[36] Sullivan, D.F., (1971), *A single index of mortality and morbidity*, in: *HSMHA Health Reports*, 86, pp. 347-354.

[37] Swine, Ch., Lucas, P., Gillet, P., Gillain, D. and Falez, F., (2003), *Etude relative à la mise en œuvre de mécanismes d'aides aux personnes en perte d'autonomie. Etude de validation de différents instruments de mesure de la dépendance*, UCL-ULB-Ulg, Janvier.

[38] Van Oyen, H., Bossuyt, N., Deboosere, P., Gadeyne, S. and Tafforeau, J., (2002), *Differences in health expectancy indicators in Belgium by region*, in: *Archives of Public Health*, 60, pp. 341-362.

[39] Van Oyen, H., Tafforeau, J. and Roelands, M., (1996), *Regional inequities in health expectancy in Belgium*, in: *Social Science and Medicine*, 43(11), pp. 1673-1678.

[40] Vansteenkiste, S., (2004), *La sécurité sociale flamande, belge et européenne – aspects juridiques de l'assurance-dépendance, en droit belge et européen*, in: *Revue Belge de Sécurité Sociale*, 1er trimestre 2004, pp. 35-68.

MODELLING REAL GDP PER CAPITA IN THE USA: COINTEGRATION TESTS

Ivan O. **KITOV**, Oleg I. **KITOV**, Svetlana A. **DOLINSKAYA**
Russian Academy of Sciences, Moscow, **RUSSIA**
Institute for the Geospheres' Dynamics
ikitov@mail.ru

Abstract

A two-component model for the evolution of real GDP per capita in the United States is presented and tested. First component of the growth rate of GDP represents the growth trend and is inversely proportional to the attained level of real GDP per capita, with the nominator being constant through time. Second component is responsible for the fluctuations around the growth trend and is defined as a half of the growth rate of the number of 9-year-olds. This nonlinear relationship between the growth rate of real GDP per capita and the number of 9-year-olds in the US is tested for cointegration. For linearization of the problem, the population time series is predicted using the relationship. Both single year of age population time series, the measured and predicted one, are shown to be nonstationary and integrated of order 1 – the original series have unit roots and their first differences have no unit root. The Engel-Granger procedure is applied to the difference of the measured and predicted time series and to the residuals of a linear regression. Both tests show the existence of a cointegrating relation. The Johansen test results in the cointegrating rank 1. Since the cointegrating relation between the measured and predicted number of 9-year-olds does exist, the VAR, VECM, and linear regression are used in estimation of the goodness of fit and root mean-square errors, (RMSE). The highest R^2=0.95 and the lowermost RMSE is obtained in the VAR representation. The VECM provides consistent, statistically reliable, and significant estimates of the slope in the cointegrating relation. Econometrically, the tests for cointegration show that the deviations of real economic growth in the US from the growth trend, as defined by constant annual increment of real per capita GDP, are driven by the change in the number of 9-year-olds.

Keywords: real GDP per capita, population estimates, cointegration, VAR, VECM, USA

JEL Classification: E32, E37, C53, O42, O51

1. Introduction

There are several macroeconomic variables, which are crucial for both theoretical consideration and practical usage. Undoubtedly, real economic growth is the most important among them. It defines the rate of economic evolution as associated with the increasing volume and quality of goods and services available for a society as a whole and for every member of the society in particular. Conventional economic concepts assume that the growth rate of real GDP reflects routine efforts of each and every economically active person, including those involved in the process of design and control of economic environment. Also, the interactions between economic agents are considered as partly controllable by economic authorities, which base their short-run actions and long-run approaches in the state of the art theories and experience. Such theories have to describe numerous aspects of the interactions between regular agents, and between the agents and the authorities as well. The literature devoted to various problems of real economic growth is extensive. A modern and almost comprehensive review of the achievements in the mainstream economics is available in the *Handbook of Economic Growth* [Agnion and Durlauf, (2005)].

There is an alternative, but simple and natural explanation using a sole cause for real economic growth [Kitov, (2006)]. Under the framework of the economic concept we have been developing since 2005, the only force driving macroeconomic evolution must be associated with some population group of specific (but constant over time) age. The intuition behind this concept is inherently related to the observation of personal income distribution (PID) in the United States. During the years of continuous and relatively accurate measurements of PID between 1960 and 2007, there was practically no change in the distributions, when they are normalized to the total population of 15 years of age and above (i.e. the working age population) and nominal per capita GDP [Kitov, (2005)]. This normalization reduces the PIDs to the portion of total income obtained by a given portion of the working age population. Some minor changes observed in the normalized PIDs are likely explained by the change in the age structure of the US society and the increase of the period when age-dependent average income grows with work experience [Kitov, (2005)]. Effectively, the PIDs demonstrate a rigid hierarchy completely reproduced by every new cohort and also by immigrants. The cohort independence is supported by the absence of any significant change with time in the normalized PID in all age groups defined by the US Census Bureau (2002) as reported in [Kitov, (2006); Kitov, (2008)].

In the economic models developed in econophysics (a branch of statistical physics) there has been a severe constrain and concern related to "conservation of energy" in actual economies [Gallegatia *et al.*, (2006)]. In reality, the gross income, as driven by the production of goods and services, is changing over time. The "frozen" hierarchy of personal incomes resolves the contradiction between the production and exchange in physical models of economy - no change in total income can affect fundamental properties of the economy as a physical system. The rigidity of the overall and age-dependent PIDs does not permit any age group of the population to improve or to lose relative income position in the economic system as a whole. Nominal changes in the absolute level of income are possible, however. In relative terms, a closed economic system has a constant structure.

In physics, there are many similar systems, where distribution of sizes is characterized by a mixture of quasi-exponential and power law distributions, as it is observed in the PIDs measured in the US [Yakovenko and Dragulescu, (2001)]. For example, in seismology the frequency distribution of seismic magnitudes, i.e. the recurrence curve introduced by Guttenberg and Richter, has these two braches – an exponential and a power law ones. Similarly to that in the Earth, any developed (there are no reliable data for developing economies or economies in transition to make any conclusion) economic system reacts to the influx of external "energy" (which is obviously not an equivalent to physical energy but is related to it) and develops the observed hierarchy of personal income distribution. The influx is provided by the existing internal economic agents and also by those who join the economy, i.e. is represented by a net sum of personal productive efforts or energy input. In a stationary case, when the number and age distribution of people is fixed and, hence, the influx is constant, there exists a nonzero economic growth trend (economic potential), which is described by a constant annual increment of real GDP per capita, as actually observed in developed countries [Kitov, (2005)]. Because the increment is constant through years, the growth rate is inversely proportional to the attained level of real GDP per capita.

In a non-stationary case, when the influx of "energy" is disturbed by the changes in the number of people joining the economy, one observes some fluctuations around the nonzero growth trend. It has been found in [Kitov, (2005)] that these fluctuations of real GDP per capita around some constant annual increase are normally distributed. Our model [Kitov, (2006)] assumed that there are no endogenous economic sources of these fluctuations, such as changes in demand and supply, inspirited or/and internally controlled by some economic agents or authorities. These fluctuations, which look like pure random innovations, are defined by the only *external* (exogenous) force. (We would like to stress again that the growth trend is of the endogenous nature.) For real economic growth, this force is the change in a single year of age population. This age is a country-specific one. In the USA and the UK, it makes nine years of age. In other European countries and Japan the age is eighteen years [Kitov, (2006)].

Therefore, one can explicitly formulate a two component model of real economic growth. Empirically, it is based on the observations of the PID in the USA and the normal distribution of annual increments of real GDP per capita in developed countries. This model is absolutely parsimonious since includes only one variable and one constant explaining the whole evolution of an economy, as expressed in monetary units. The model has described the evolution of real GDP per capita in the USA, the UK, France [Kitov, (2006)], and Japan [Kitov, (2006)].

Physics and economics both require any quantitative model to be validated by standard statistical and econometric procedures. Juselius and Franchi [Juselius, and Franchi, (2007)] have proposed the cointegrated vector auto-regression (VAR) as an adequate framework of such validation. The principal idea behind their approach consists in the estimation of statistical properties of the variables defining the models as themselves and in combinations in order to distinguish between probable and unlikely theoretical assumptions. They have also carried out an important initial analysis of conventional theoretical models of real economic growth, RBC and DSGE, and found that some principal assumptions underlying the models are not empirically supported. In a sense, we follow their procedure and also some statistical procedures developed in [Kitov, Kitov, and Dolinskaya, (2007)].

The high standard introduced in [Juselius, and Franchi, (2007)] establishes that any economic model should come from and be justified by empirical data, not from "the easiness of mathematical formulation". At least, the involved variables should meet minimal requirements established by models themselves. Such an approach has been successfully applied in hard sciences and brought a well-recognized reliability of scientific knowledge and technical inventions such as aircrafts, bridges, and so on. The reliability follows from an extensive statistical test of each and every parameter, variable, empirical relationship or fundamental law. Obviously, any physical (and economic) model is actually an approximation to a finite set of statistical links (or scatter plots) between measured variables [Ormerod, (2005)].

Our model describes the measured time series of real GDP per capita in the USA between 1960 and 2002 and allows predictions of the growth of real GDP per capita at various time horizons. The accuracy of these predictions depends on the accuracy of relevant population estimates. In this paper, we test the model (and corresponding data) in econometric sense and demonstrate the existence of a (nonlinear) cointegrating relation between real economic growth and population. The level of confidence associated with the obtained cointegrating relation is high as supported by various statistical tests. The model also involves the lowermost possible number of variables and does not contain any structural breaks. We consider a developed economy as a natural (in sense of physics) system, which evolves according to its own strict laws. Because the system is characterized by a rigid structure of personal income distribution no internal part, including economic authorities, can accelerate the evolution of the system as a whole by economic means. Of course, any part of the system can hamper or stop the evolution, as demonstrated by socialist and developing countries. The predictability and controllability (through demography) of real economic growth are important features of our model, which are wrongly denied by some (econo-) physicists [Gallegatia *et al.*, (2006), Kitov, and Kitov, (2008)].

The remainder of the paper is organized as follows. Section 2 presents a two-component model for real economic growth and the data used in the study. The model is reversed in order to obtain the number of 9-year-olds from measured economic growth, as expressed by real GDP per capita. Section 3 is devoted to the estimation of basic statistical properties of the variables, including the order of integration. Section 4 contains three different tests for cointegration between the measured number of 9-year-olds in the USA and that predicted from the measured GDP – two associated with the Engle-Granger approach and also the Johansen test. Section 5 presents a number of VAR and vector error correction (VEC) models as well as some estimates of root mean square errors (RMSE) and goodness-of-fit. Section 6 discusses principal results and concludes.

2. Model and data

There is a measured macroeconomic variable characterized by a long-term predictability for a large developed economy. This is the annual increment of real GDP per capita [Kitov, (2006)]. One can distinguish two principal sources of the intensive part of real economic growth, i.e. the evolution of real GDP per capita, G: the change in the number of 9-year-olds, and the economic growth trend associated with per capita GDP, G_t. The trend has the simplest form – no change in mean annual increment, as expressed by the following relationship:

$$dG_t(t)/dt = A \quad (1)$$

where G(t) is the absolute level of real GDP per capita at time t, A is an empirical and country-specific constant. The solution of this ordinary differential equation is as follows:

$$G_t(t) = At + B \quad (2)$$

where $B=G_t(t_0)$, t_0 is the starting time of the studied period. Then, the relative growth rate (or economic growth trend) of real GDP per capita is:

$$g_{trend}(t) = dG_t/G_tdt = A/G \quad (3)$$

which indicates that the (trend) rate is inversely proportional to the attained level of the real GDP per capita and the growth rate should asymptotically decay to zero.

One principal correction has to be applied to the per capita GDP values published by the Bureau of Economic Analysis (2006). This is the correction for the difference between the total population and the population of 15 years of age and above, as discussed by Kitov (2006). Our concept requires that only this economically active population should be considered when per capita values are calculated.

Following the general concept of the two principal sources of real economic growth [Kitov, (2006)] one can write an equation for the growth rate of real GDP per capita, $g_{pc}(t)$:

$$g_{pc}(t) = dG(t)/(dt\cdot G(t)) = 0.5dN_9(t)/(dt\cdot N_9(t)) + g_{trend}(t) \quad (4)$$

where $N_9(t)$ is the number of 9-year olds at time t. One can obtain a reversed relationship defining the evolution of the 9-year-old population as a function of real economic growth:

$$d(\ln N_9(t)) = 2(g_{pc} - A/G(t))dt \quad (5)$$

Equation (5) defines the evolution of the number of 9-year-olds as described by the growth rate of real GDP per capita. The start point of the evolution has to be characterized by some (actual) initial population. However, various population estimates (for example, post- and intercensal one) potentially require different initial values and coefficient A.

Instead of integrating (5) analytically, we use the annual readings of all the involved variables and rewrite (5) in a discrete form:

$$N_9(t+\Delta t) = N_9(t)[1 + 2\Delta t(g_{pc}(t) - A/G(t))] \quad (6)$$

where Δt is the time unit equal to one year. Equation (6) uses a simple representation of time derivative of the population estimates, where the derivative is approximated by its estimate at point t. The time series g_{pc} and N_9 are independently measured variables. In order to obtain the best prediction of the $N_9(t)$ by the trial-and-error method one has to vary coefficient A and (only slightly in the range of the uncertainty of population estimates) the initial value - $N_9(t_0)$. The best-fit parameters can be obtained by some standard technique minimising the RMS difference between predicted and measured series. In this study, only visual fit between curves is used, with the average difference minimised to zero. This approach might not provide the lowermost standard deviation.

Equation (6) can be interpreted in the following way – the deviation between the observed growth rate of GDP per capita and that defined by the long-tern trend is completely defined by the change rate of the number of 9-year olds. A reversed statement is hardly to be correct - the number of people of some specific age can not be completely or even in large part defined by contemporary real economic growth. Specifically, the causality principle prohibits the present to influence the birth rate nine years ago. Econometrically speaking, the number of 9-year olds has to be a weakly exogenous variable relative to contemporary economic growth. This property of the variables is used in the VAR models in Section 5.

In fact, Eq. (6) provides an estimate of the number of 9-year-olds using only independent measurements of real GDP per capita. Therefore, the amplitude and statistical properties of the deviation between the measured and predicted number of 9-year olds can serve for the validation of (4) and (5). In Sections 3 through 5 we use the predicted number of 9-year-olds for statistical estimates instead of the real GDP per capita readings themselves. The link between population and economic growth is effectively nonlinear and there would be difficult to study it in a linear representation. Since both involved variables are measured with some uncertainty and probably are nonstationary, the cointegrated VAR analysis should be an appropriate one.

There are numerous revisions and vintages of the population estimates. Figure 1 compares post- and intercensal population estimates of the number of 9-year olds between 1960 and 2002 [U.S. Census Bureau, (2007)]. The error of closure, i.e. the difference between the census count and the postcensal estimate at April 1, 2000, is 57233. The error of closure for the population group between 5 and 13 years of age is 1309404, however, i.e. approximately twice as large for every single year of age as that for the 9-year-olds. For the intercensal estimate, this error of closure is proportionally distributed over the 3653 days between April 1, 1990 and April 1, 2000 [U.S. Census Bureau, (2004)]. Hence, the level of the intercensal estimate is represented by the level of the postcensal one plus corresponding portion of the error of closure. The curves in Figure 1 demonstrate a growing divergence between these two estimates. There are also some non-zero corrections between adjacent years of birth in wider age groups. After April 2000, both estimates in Figure 1 are apparently postcensal with different bases in 2000. Even this minor deviation between the estimates might be of importance for statistical tests and inferences and both are analyzed in this study.

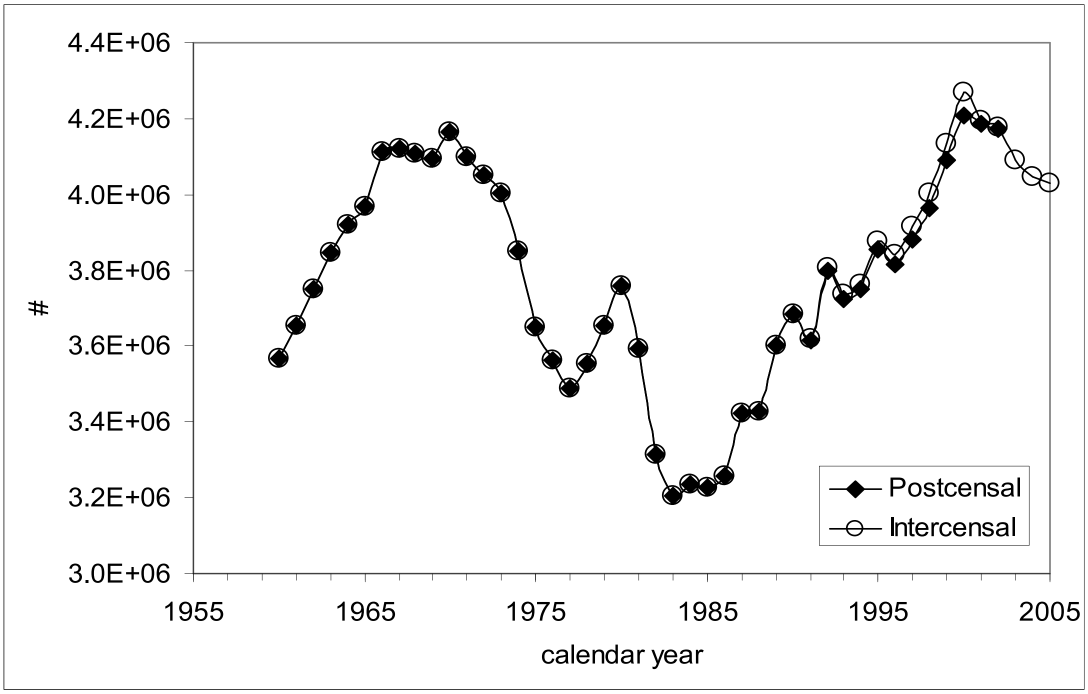


Figure 1. Comparison of the postcensal and intercensal estimates of the number of 9-year olds reported by the US Census Bureau (2007). The difference is observed only during the years between 1990 and 2002.

Real GDP per capita is estimated using total real GDP and the number of people of 15 years of age and above. This excludes from the macroeconomic consideration those who do not add to real economic growth [Kitov, (2006)]. Figure 2 depicts the growth rate of real GDP per capita in the USA between 1960 and 2002 used in the study. In average, the growth rate is 0.020 with standard deviation of 0.022. There are seven negative readings coinciding with the recession periods defined by the National Bureau of Economic Research (2007).

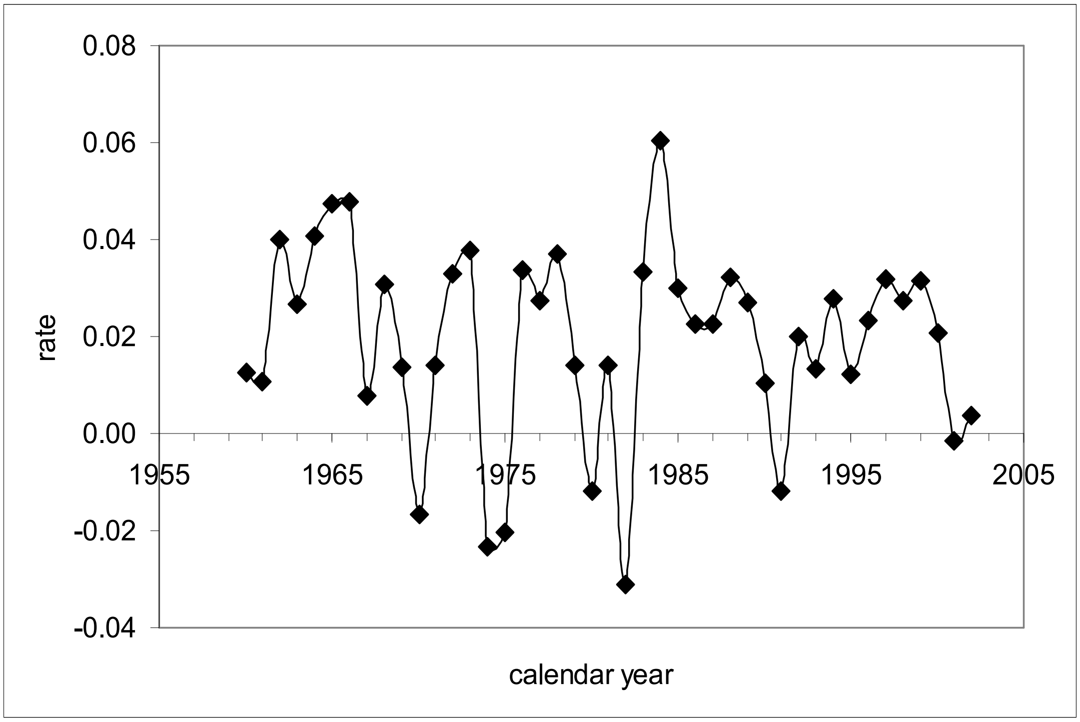


Figure 2. The growth rate of real GDP per capita in the USA between 1960 and 2002. The growth rate is corrected for the difference between total population and that above 15 years of age [Kitov, I., (2006)].

The period between 1960 and 2002 has been chosen by the following reasons. Before 1960, the single year of age population estimates are not reliable and might introduce a significant distortion in statistical estimates and inferences. After 2002, the GDP values are prone to comprehensive NIPA revisions of unknown amplitude, which historically occurred about every 5 years [Fixler, and Green, (2005)]. The most recent comprehensive revision was in 2003 and spanned the years between 1929 and 2002.

3. Unit root tests

The technique of linear regression for obtaining statistical estimates and inferences related to time series is applicable only to stationary series, as Granger and Newbold showed [Granger, and Newbold, (1967)]. Two or more nonstationary series can be regresses only in the case when there exists a cointegrating relation between them [Hendry, and Juselius, (2001)], with several precautions discussed in [Engle, and Granger, (1987)]. Therefore, the first step in any econometric analysis of time dependent data sets is currently consists is the estimation of the order of integration of involved series. Unit root tests applied to original series and their first and higher order differences are a useful tool to determine the order of integration.

Standard econometric package Stata9 provides a number of appropriate procedures implemented in an interactive form. The Augmented Dickey-Fuller (ADF) and the modified DF t-test using a generalized least-squares regression (DF-GLS) are used in this study. Potentially, the tests provide adequate results for the available short series consisting of only 41 annual readings - the real GDP per capita and the number of 9-year olds. Small samples are usually characterized by a limited reliability of statistical inferences.

There are four original time series tested for unit roots - the measured and predicted according to (6) number of 9-year-olds between 1962 and 2002. Each of the series contains two versions - a postcensal and intercensal one (for the period between 1990 and 2000, i.e. between two decennial censuses). The difference is minor, as Figure 1 demonstrates, but the intercensal series potentially contains such artificial features as autocorrelation introduced by the Census Bureaus during the revision associated with the error of closure. Statistically, the postcensal time series might be less "contaminated" than the intercensal one.

Some results of the unit root tests for the four original series are listed in Table 1. All these series are characterized by the presence of unit roots - the test values are significantly larger than the 1% critical values. In the ADF tests, trend specification is *constant* and the maximum lag order is 3. In the DF-GLS tests, the maximum lag is 4 and the same trend specification is used. Hence, one can conclude that the studied time series are nonstationary. The order of integration is not clear, however.

Table 1. Unit root tests for the measured and predicted number of 9-year-olds. Trend specification is constant

Test	Lag	Intercensal		Postcensal		1% critical
		predicted	measured	predicted	measured	
ADF	0	-1.50	-0.72	-1.51	-0.70	-3.65
	1	-2.10	-1.39	-2.10	-1.40	-3.66
DF-GLS	1	-2.34	-1.52	-2.35	-1.55	-2.63
	2	-1.82	-1.60	-1.82	-1.62	-2.63

The first differences of the measured and predicted number of the 9-year-olds (the postcensal version) between 1962 and 2002 (the reading for 1961 is also used in the difference) are presented in Figure 3. There is no visible trend in the data and one can presume a constant as trend specification. The average value is 9600 and 12625, and standard deviation is 152487 and 105287 for the measured and predicted time series, respectively.

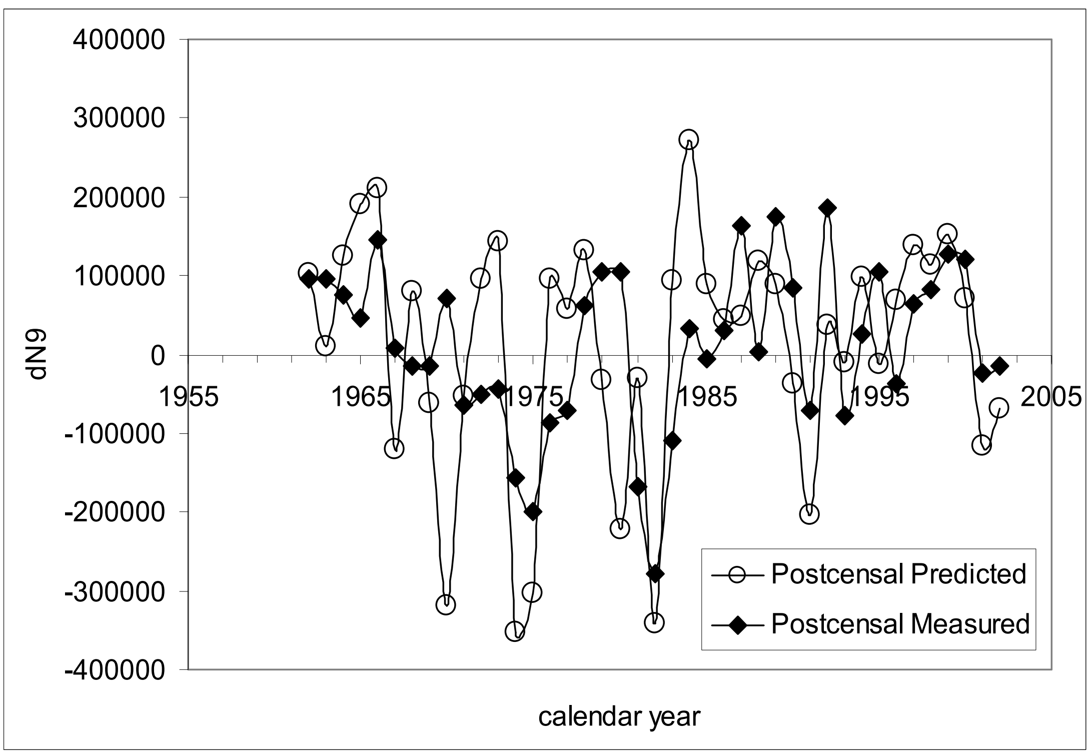


Figure 3. The first differences of the measured (postcensal) and the predicted number of 9-year-olds. There is no visible trend in the time series with average values 9600 and 12625, respectively. Standard deviation is 152487 and 105287.

Table 2 summarizes some results of the unit root tests as applied to the first differences. The predicted time series are definitely characterized by the absence of unit roots, as the ADF and DF-GLS both demonstrate for the maximum lag order 2. For lag 3, the ADF gives values just marginally below the 1% critical value. The measured time series have specific autoregressive properties intrinsically related to the methodology of population revisions and are characterized by mixed results for the unit root tests. The DF-GLS test rejects the null hypothesis of the presence of a unit root for all lags from 1 to 3. The ADF rejects the null only for lag 0. Bearing in mind the shape of the measured original curves in Figure 1, which demonstrate a quasi-sinusoidal behaviour without any significant linear trend; one can assume that their first differences are stationary. In this study, the absence of unit roots in all the first difference series is accepted.

Table 2. Unit root tests for the first differences of the measured and predicted number of 9-year-olds. Trend specification is constant. The maximum lag order is 3.

Test	Lag	Postcensal		Intercensal		1% critical
		predicted	measured	predicted	measured	
ADF	0	-4.86*	-4.22*	-4.87*	-4.27*	-3.65
	1	-4.66*	-3.37	-4.67*	-3.39	-3.66
	2	-3.86*	-2.80	-3.86*	-2.80	-3.66
	3	-3.44	-3.22	-3.44	-3.20	-3.67
DF-GLS	1	-4.64*	-3.01*	-4.54*	-3.02*	-2.63
	2	-3.67*	-2.48	-3.67*	-2.48	-2.63
	3	-3.12*	-2.84*	-3.12*	-2.84*	-2.63

The presence of unit roots in the original series and the absence of unit roots in the first differences evidences that the former series are integrated ones of order 1. This fact implies that cointegration analysis has to be carried out before any linear regression because the latter is potentially a spurious one.

4. Cointegration test

The assumption that the measured number of 9-year-olds in the USA, $N_{9m}(t)$, and that predicted from the real economic growth, $N_{9p}(t)$, are two cointegrated non-stationary time series is equivalent to the assumption that their difference, $\varepsilon(t)=N_{9m}(t) - N_{9p}(t)$, is a stationary or I(0) process. The predicted and measured series corresponding to the post- and intercensal population estimates are shown in Figures 4 and 6, and their differences in Figures 5 and 7, respectively.

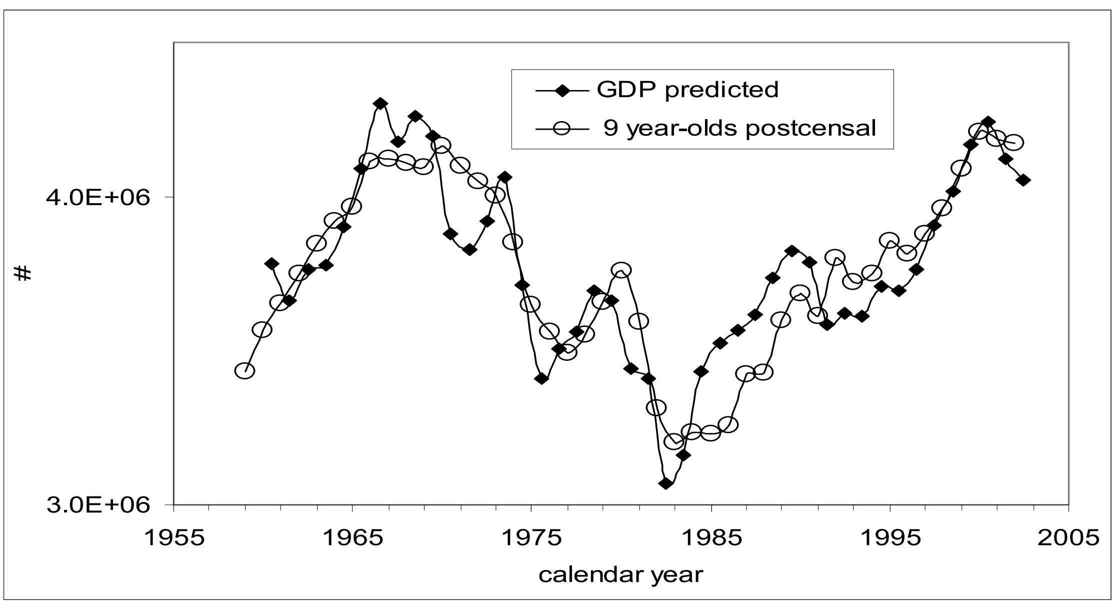


Figure 4. Comparison of the measured and predicted postcensal population estimates between 1960 and 2002.

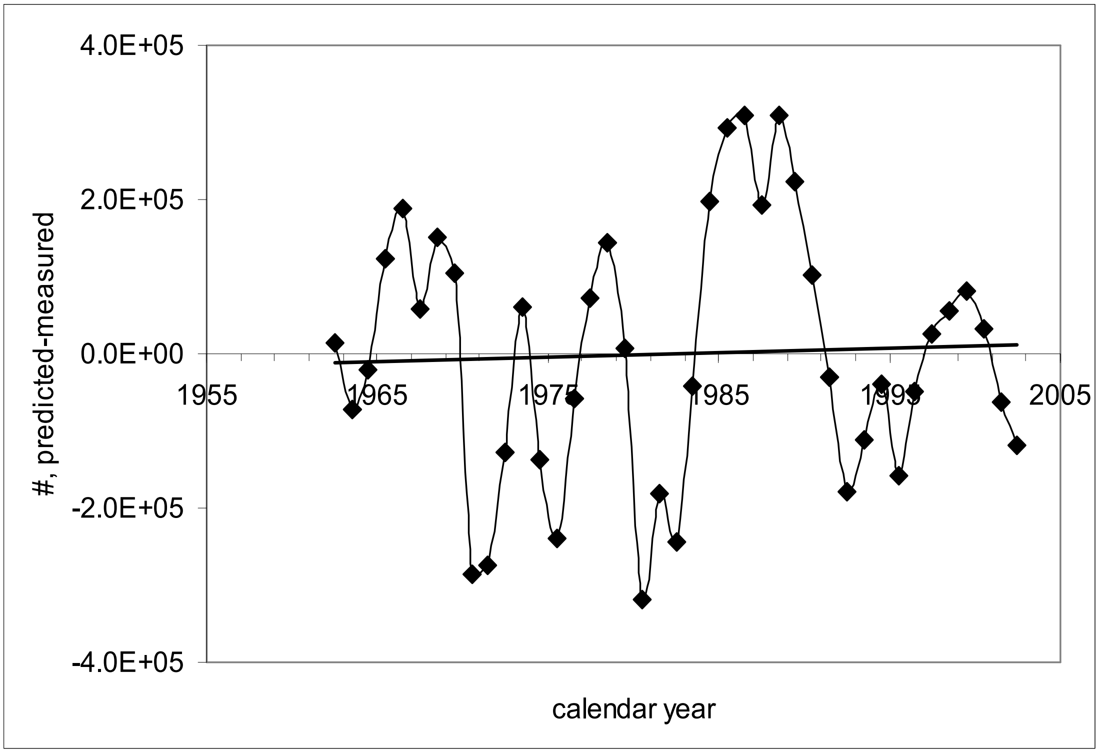


Figure 5. The difference between the measured and predicted population estimates presented in Figure 4. For the period between 1962 and 2002, the average difference is 0 and standard deviation is 164926 for coefficient A=547.1325 and the initial value for the population of 3900000 in 1959. Linear regression is represented by a bold straight line.

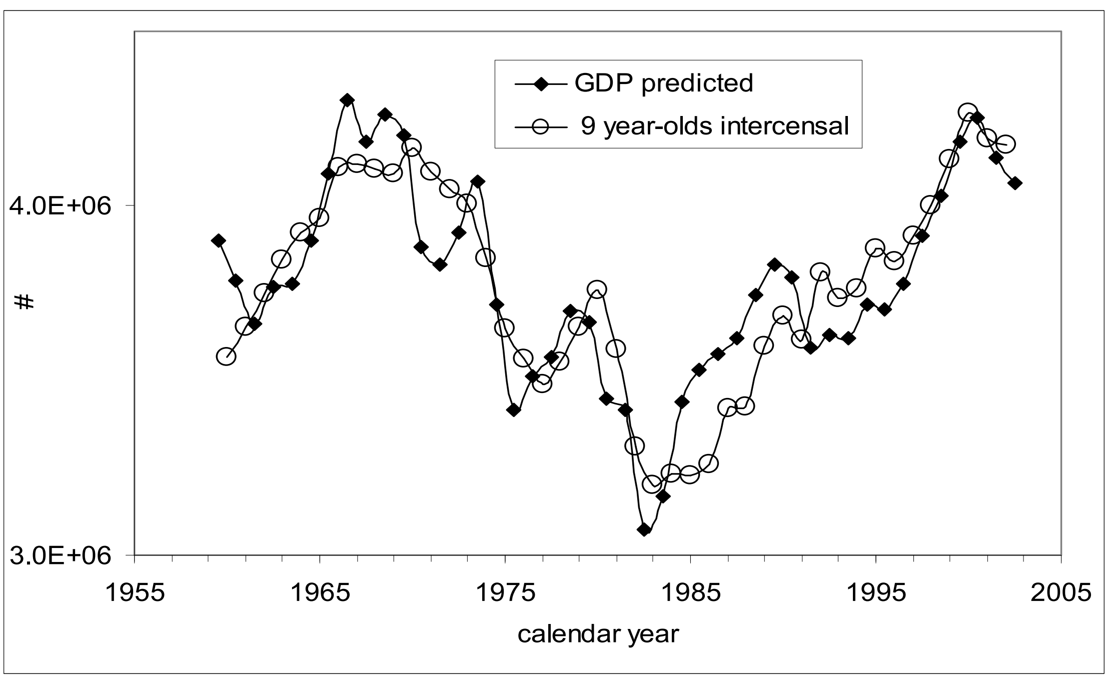


Figure 6. Comparison of the measured and predicted intercensal population estimates between 1960 and 2002.

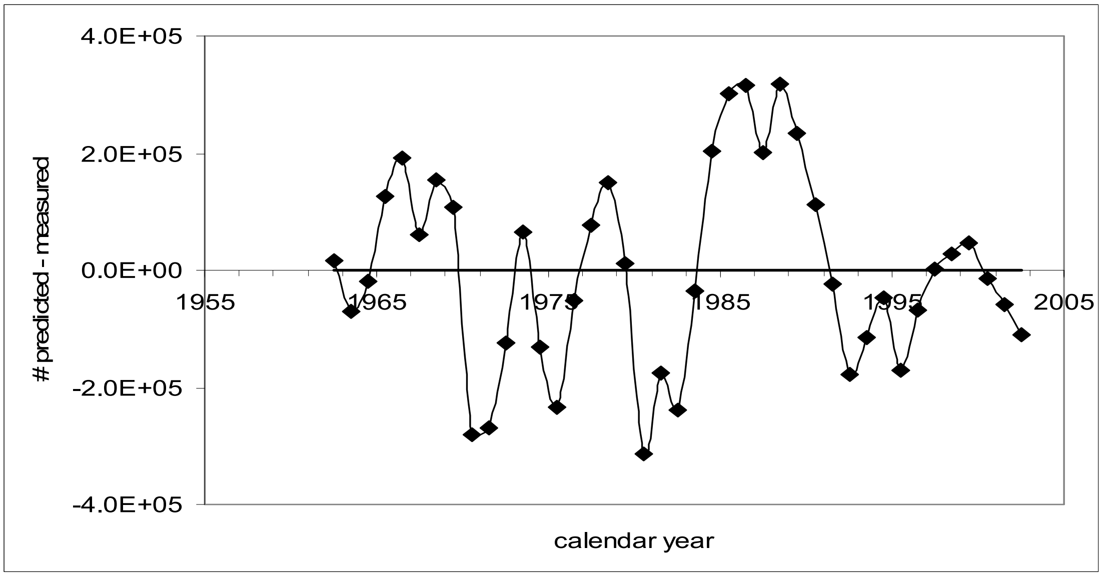


Figure 7. The difference between the measured and predicted population estimates presented in Figure 6. For the period between 1962 and 2002, the average difference is -1 and standard deviation is 165744 for coefficient A=546.079 and the initial value of population of 3900000 in 1959.
Linear regression is represented by a bold straight line.

It is natural to start with unit root tests in the difference. If ε(t) is a non-stationary variable having a unit root, the null hypothesis of the existence of a cointegrating relation can be rejected. Such a test is associated with the Engle-Granger approach [Engle, and Granger, (1987)], which requires the $N_{9m}(t)$ to be regressed on the $N_{9p}(t)$ as the first step, however. It is worth noting, that the predicted variable is obtained by a procedure similar to that of linear regression and provides the best visual fit between corresponding curves. The Engle-Granger approach is most reliable and effective when one of the two involved variables is weakly exogenous, i.e. is driven by some forces not associated with the second variable. This is the case for the GDP per capita and the number of 9-year-olds. The latter variable is hardly to be driven by the former one. The existence of an opposite causality direction is the main object of this study.

The results of the ADF and DF-GLS tests, listed in Table 3, demonstrate the absence of a unit root in the measured-predicted difference series for both the post- and intercensal population estimates. Since the predicted series are constructed in the assumption of a zero average difference, trend specification in these tests is "*none*". The maximum lag order in the tests is 3. These results give strong evidences in favor of the existence of a cointegrating relation between the measured and predicted time series. Therefore, from the econometric point of view, it is difficult to deny that the number of 9-year-olds is *the only* defining force behind the observed fluctuations of the real economic growth. These fluctuations are observed around the growth trend defined by constant annual increment, A, of the real GDP per capita.

Table 3. Unit root tests for the differences between the measured and predicted number of 9-year-olds. Trend specification is constant. The maximum lag order is 3.

Test	Lag	Time series		1% critical
		postcensal	intercensal	
ADF	0	-2.87*	-2.85*	-2.64
	1	-3.67*	-3.59*	-2.64
	2	-2.99*	-3.92*	-2.64
	3	-2.90*	-2.83*	-2.64
DF-GLS	1	-3.55*	-3.47*	-2.64
	2	-2.98*	-2.92*	-2.64

	3	-2.92*	-2.85*	-2.64

The next step is to use the Engle-Granger approach again and to study statistical properties of the residuals obtained from linear regressions of the measured and predicted single year of age populations. A pitfall of the regression analysis consists in a slight time shift between the measured and predicted series – the former variable is assigned to July 1 (averaged population) and the latter to December 31 (cumulative GDP increase) of the same year. Such a phase shift, apparently, results in a deterioration of regression results but can not be recovered since only annual population estimates are available before 1980.

Table 4 presents a summary of relevant unit root tests with the same specifications as accepted for the difference of the same series. The null hypothesis of a unit root presence is rejected for both time series and all time lags. Therefore, the residuals of the regression build an I(0) time series, and the Engle-Granger tests proves that the predicted and measured variables are cointegrated.

Table 4. Unit root tests for the residual time series of a linear regression of the measured series on the predicted one. The measured and predicted series are the numbers of 9-year-olds. Trend specification is *none* (zero average value of the residuals) and maximum lag order 3.

Test	Lag	Time series		1% critical
		postcensal	intercensal	
ADF	0	-3.03*	-3.02*	-2.64
	1	-3.88*	-3.86*	-2.64
	2	-3.15*	-3.13*	-2.64
	3	-3.05*	-3.01*	-2.64
DF-GLS	1	-3.71*	-3.69*	-2.64
	2	-3.06*	-3.04*	-2.64
	3	-2.98*	-2.95*	-2.64

The Johansen [Johansen, (1988)] approach is based on the maximum likelihood estimation procedure and tests for the number of cointegrating relations in the vector-autoregressive representation. The Johansen technique allows simultaneous testing for the existence of cointegrating relations and determining their number (rank). For two variables, only one cointegrating relation is possible. When cointegration rank is 0, any linear combination of the two variables is a non-stationary process. When the rank is 2, both variables have to be stationary. When the Johansen test results in rank 1, a cointegrating relation between the involved variables does exist.

In the Johansen approach, one has first to analyze some specific properties of the underlying VAR model for the two variables. Table 5 lists selection statistics for the pre-estimated maximum lag order in the VAR. Standard trace statistics is extended by several useful information criteria: the final prediction error, FPE; the Akaike information criterion, AIC; the Schwarz Bayesian information criterion – SBIC; and the Hannan and Quinn information criterion, HQIC. All tests and information criteria in Table 5 indicate the maximum pre-estimated lag order 1 for VARs and vector error-correction models, VECMs. Therefore, the maximum lag order 1 was used in the Johansen tests along with *constant* as the trend specification.

Table 5. Pre-estimation lag order selection statistics. All tests and information criteria indicate the maximum lag order 1 as an optimal one for VARs and VECMs.

	Lag	LR	FPE	AIC	HQIC	SBIC
postcensal	1	63.03*	5.8e+09*	25.31*	25.36*	25.44*
intercensal	1	61.63*	6.1e+09*	25.38*	25.42*	25.51*

FPE - the final prediction error, AIC - the Akaike information criterion, SBIC - the Schwarz Bayesian information criterion, HQIC - the Hannan and Quinn information criterion

The properties of the VAR error term have a critical importance for the Johansen test [Hendry, and Juselius, (2001)]. A number of diagnostic tests was carried out for the VAR residuals. The Lagrange multiplier test for the postcensal time series resulted in χ^2 of 0.34 and 0.09 for lags 1 and 2, respectively. This test accepts the null hypothesis of the absence of any autocorrelation at these lags. The Jarque-Bera test gives χ^2=7.06 (Prob>0.03) with skewness=0.96 and kurtosis=3.77, the skewness being of the highest importance for the normality test and the validity of statistical inference. Hence, the residuals are probably not normally distributed, as expected from the artificial features of the measured population time series. The VAR model stability is guaranteed by the eigenvalues of the companion matrix, which are lower than 0.63. As a whole, the VAR model accurately describes the data and satisfies principal statistical requirements applied to the residuals.

Table 6 represents some results of the Johansen tests. In both cases the cointegrating rank is 1. Hence, there exists a long-run equilibrium relation between the measured and predicted number of 9-year-olds in the USA. The predicted number is obtained solely from the readings of real GDP per capita measured and reported by the BEA (2007). We do not test for the causality direction between the variables because the only possible way of influence, if it exists, is absolutely obvious.

Table 6. Johansen test for cointegration rank for the measure and predicted time series. Trend specification is constant. Maximum lag order is 2.

Time series	Rank	Eigenvalue	SBIC	HQIC	Trace statistics	5% critical value
postcensal	1	0.397	52.48*	52.23*	2.198*	3.76
intercensal	1	0.379	52.55*	52.30*	2.117*	3.76

In this Section, three different tests have demonstrated at a high level of confidence that the measured and predicted number of 9-year-olds in the USA are cointegrated. One can use the cointegrating relation for a reliable prediction of real economic growth in the USA. This finding proves that the evolution of a developed economy is predictable in principle.

5. VAR, VECM, and linear regression

Now, it is proved by standard econometric tools that the measured and predicted single year of age population series are cointegrated. Therefore, the estimates of the goodness-of-fit, R^2, and RMSE in various statistical representations have to be valid and can provide important information on the accuracy of relevant population and economic measurements, and the relation itself.

The VAR representation provides a good estimate of R^2 and RMSE due to strong noise suppression. In practice, AR is a version of a weighted moving average, which optimizes noise suppression throughout the whole series. Two VAR models are possible, however, with the predicted time series used as an exogenous predictor and as an endogenous variable. Table 7 summarizes some results of the VAR models and demonstrates that the goodness of fit is excellent, with the highest R^2~0.95 and the lowermost RMSE near 72000 corresponding to the exogenous predicted time series for the postcensal population estimates. This version of VAR uses the maximum lag order 2, and the Table confirms that coefficient L2 is not significant in line with the previous estimates of the maximum lag. The coefficient for the predictor is significant.

Table 7. VAR models for the measured and predicted number of 9-year-olds for the postcensal and intercensal estimates. Maximum lag order is 2. Two cases for the predicted time series are considered - endogenous and exogenous one.

Measured-Predicted VAR	RMSE	R^2	Measured		Predicted		
			L1	L2	L0	L1	L2
exogenous - postcensal	71645	0.9489	0.82* [0.13]	-0.12 [0.12]	0.34* [0.06]	-	-
endogenous - postcensal	89300	0.9229	0.85* [0.19]	-0.17 [0.15]	-	0.33* [0.11]	-0.03 [0.12]
exogenous - intercensal	73954	0.9474	0.82* [0.13]	-0.11 [0.12]	0.35* [0.06]	-	-
endogenous - intercensal	92440	0.9202	0.88* [0.19]	-0.17 [0.16]	-	0.33* [0.11]	-0.05 [0.12]

The VECM representation uses information additional to that provided by the VAR models due to separation of noise and equilibrium relation. So, it potentially provides an improvement on the VAR models. Table 8 lists some results obtained in the VECM (cointegrated VAR) representation.

Coefficient β, defining the link between the measured and predicted series, is significant in both cases confirming the existence of a cointegrating relation. Coefficients α_1 and α_2 define the input of the cointegrating relation to the I(0) time series of lagged first differences of the measured and predicted series. Their estimates are significant and show a relatively large error correction effect. Coefficients of the LD terms are both insignificant as corresponded to the largest lag order 2. The values of R^2 are relatively high (0.34 and 0.32) and RMSE is ~90000 and 130000 for the postcensal and intercensal series, respectively. The RMSE values are slightly larger than those from the VAR models.

Table 8. VECM for the postcensal and intercensal estimates of the number of 9-year-olds. The maximum lag is 2. Cointegrating rank 1 for the relationship between the measured and predicted time series.

Measured-Predicted VECM	RMSE	R^2	β	α_1	α_2	Measured LD	Predicted LD
postcensal	89839	0.3446	-1.21* [0.11]	-0.24* [0.10]	0.28 [0.17]	0.11 [0.16]	0.06 [0.13]
intercensal	93007	0.3181	-1.24* [0.12]	-0.22* [0.10]	0.29 [0.16]	0.11 [0.16]	0.08 [0.13]

Finally, Table 9 is representing the results of linear regressions. These results are biased by the time shift between the series and are inferior to those obtained using VAR and VECM. The moving average technique, however, provides a slight improvement in the statistical estimates. This effect is inherently related to noise suppression in the time series.

Table 9. Results of linear regression of the measured time series on the predicted one.

Time series	Regression	Tangent	Constant	R^2	RMSFE
postcensal	M vs. P	0.85* [0.09]	569325 [326128]	0.71	160000
	M vs. MA(2)	0.94* [0.07]	221197 [274652]	0.81	130000
	M vs. MA(3)	1.09* [0.06]	-318464 [231765]	0.89	99985
intercensal	M vs. P	0.86* [0.09]	511114 [330855]	0.72	160000
	M vs. MA(2)	0.96* [0.06]	167488 [281940]	0.81	130000
	M vs. MA(3)	1.04* [0.07]	-126233 [249095]	0.86	110000

M – measured time series
P – predicted time series
MA(N) – N-year moving average

Despite a very high goodness-of-fit, approaching 0.95, in the VAR representation, the RMSE estimates are relatively large. This severely complicates the usage of Eq. (4) for the prediction of real economic growth in the USA. The RMSEs are comparable in amplitude with the uncertainty of the population estimates, especially at younger ages [West, Robinson, (1999)]. In addition, a conservative estimate of the uncertainty of growth rate of real GDP is between 0.5 and 1 percentage point, which includes also the uncertainty associated with CPI and GDP deflator. In order to distinguish between these measurement errors and true deviations in the cointegrating relations one needs a substantial improvement in population estimates.

6. Conclusion

There is an equilibrium (nonlinear) long-run relation between the number of 9-year-olds and real GDP capita in the United States. This fact implies that real economic growth, as expressed in monetary units, is practically predetermined by the age structure of the US society. An increasing number of 9-year-olds would guarantee an accelerating growth, extra to that defined by the constant annual increment of real GDP per capita.

At low frequencies, the behavior of the number of 9-year-olds in the USA is characterized by a visible period of about 30 years, between the peaks in 1970 and 2000. Such long-period oscillations in economic evolution are well-know since the 1920s, when Russian economist Nikolai Kondratiev published his original analysis. Our model gives a natural explanation of the Kondratiev waves – they are related to the natural increases and decreases in birth rate (and/or migration). For numerous reasons, the birth rate fluctuates and cycles are observed at all frequencies.

A bad news for the USA is that the ten to fifteen years since 2000 will be probably associated with a decreasing branch of the K-wave. Taking into account the effect of the decreasing background growth rate associated with the increasing real GDP per capita in Eq. (3), one can expect a significant deceleration in the US economy as expressed by a lower growth rate of real GDP per capita. However, if the total population will continue to grow at an annual rate of 1 per cent, as has been observed in the USA during the last forty years, the negative effect of the N_9 decrease will be compensated. In developed European countries, the effect of the total population growth is practically negligible and they seemingly do not grow so fast as the USA does. There is just an illusion of an elevated growth rate, which disappears when one uses per capita GDP values.

The fluctuations of the annual increment of real GDP per capita around the average level represent a random process. This stochastic component is driven only by one force and can be actually predicted to the extent one can predict the number of 9-year-olds at various time horizons. The population estimates for younger ages in previous years provide an excellent source for this prediction. The growth rate of a single year population can be predicted with a higher accuracy because the levels of adjacent cohorts change proportionally. Therefore, the number of 7-year-olds today is a very good approximation to the number of 9-year-olds in two years. Theoretically, one can use the younger populations for an exact prediction. In practice, the current methodology of population estimates does not provide adequate precision and only long-term changes have a high enough signal (true change) to noise (measurement error) ratio to resolve of the link between real economic growth and population, as Figures 4 and 6 illustrate.

The concept we have been developing links the fluctuations of real growth rate to young people (9-year-olds) likely being outside the structure of economic production. However, they bring to the economic system a nonzero and changing input, which can be interpreted as demand for goods and services. Those economic agents who are currently inside the system can not change real demand per capita due to the rigid PID. Immigrants and the population decrease associated with deaths also cannot change per capita GDP values because the PID does not demonstrate any effect of these potential sources of changes. One can presume that the hierarchy of personal incomes momentarily recovers to its origin structure, when accommodating the disturbances induced by these two sources.

The model of real economic growth tested in this study is supported by the results reported in [Juselius and Franchi, (2007)] that the principal source of economic variations is the demand for consumption and for labor but not shocks to technology or total factor productivity. (Labor productivity in developed countries is driven only by real economic growth and labor force participation rate [Kitov and Kitov, (2008)]. The latter also is an unambiguous function of real economic growth, as expressed by real GDP per capita [Kitov and Kitov, (2008)].) Newcomers entering the economy, as represented by 9-year-olds, somehow bring and introduce their long-term demand for consumption into the economic system. This demand has been changing over time according to the variations in the number of 9-year-olds and induces relevant changes in the demand for labor. A complication to conventional models is the decelerating economic trend, as defined by Eq. (3).

Expenditures in developed economies cannot be separated into two distinct parts, which are usually described as saving (investment) and consumption, the former being the driving force of shocks to technology and total factor productivity. Many theories of endogenous economic growth, however, are based on this assumption and stress the importance of investment for the rate of economic growth. Under our framework, there is no direct link between real economic growth, as expressed in monetary units (per capita), and technological content. In other words, any set of technological breakthroughs achieved during a certain period, for example one year, has the same money valuation. What important for the monetary size is only changes in quantitative characteristics of population – the age structure. We also do not share the opinion or assumption that investments are made for the sake of economic growth *per ce*. One hardly can imagine that an owner, shear holder or manager who really wants an overall economic growth and decides what input s/he can bring to the process. Investment decisions are rather made for a sole purpose, which is psychologically and economically justified, one wishes by all means to elevate the current position in relevant PID.

Technological innovations (not only purely technological, but also cultural in a broader sense) have been stimulating the growth in the diversity of goods and services. At the same time, the innovations were helpful in creating new tools for deposing some people from their top positions in the PID. The rigidity of the PID does not allow joining the top positions – only deposing is possible (when working age population does not change). However, not all technologically excellent discoveries guarantee income increase.

Therefore, the main purpose to invest is to progress in the income pyramid to higher steps. This is a routine, strong and long-run interest and demand. Sometimes it uses not the best sides of human psychology and reflexes. But, in general, it makes what it should make – brings random and deterministic innovations in technologies. Juselius and Franchi [Juselius and Franchi, (2007)] justified our concept by empirical analysis. No technological innovations induce fluctuations in economic growth. (We do not consider here technical policy aimed at the selection of sound innovations, which can definitely bring a better result for the society as a whole. For example, investments in military technologies brought a large-scale profit to many areas of civil techniques.) The authors of [Juselius, and Franchi, (2007)] deny the possibility of technology, whatever it is, to drive monetary side of social life.

The Great Moderation is easily explained in our framework. Amplitude of the fluctuations of the defining age population around the constant level has been decaying since the 1980s, as Figure 5 and 7 demonstrate. The reasons behind the smoothing of the population changes are beyond the scope of this study but deserve a special attention. The economic growth trend, as a part of the growth rate of real GDP, has been also decreasing with increasing per capita GDP level as denominator. Inflation in the USA and other developed countries is driven by the change in the level of labor force [Kitov, (2006), (2007), Kitov, Kitov, and Dolinskaya, (2007)], which in turn, is defined by real GDP per capita and total population. Therefore, the observed decrease in the volatility of the GDP growth rate leads to lower fluctuations in inflation. The Great Moderation is not going to leave the scene in the future.

7. References

[1] Bureau of Economic Analysis, (2007), *National Income and Product Accounts*. *Table 1.1.6. Real Gross Domestic Product, Chained Dollars*, retrieved March 5, 2007 from http://bea.gov/be/dn/nipaweb/

[2] Chiarella, C., Gao, S., (2002), *Type I spurious regression in econometrics*, Working Paper No 114, School of Finance and Economics, University of Technology Sydney

[3] Engle, R., Granger, C., (1987), *Cointegration and error correction: representation, estimation, and testing,* in: *Journal of Econometrics*, 55, 251-276

[4] Gallegatia, M., *et al.*, (2006), *Worrying trends in econophysics*, in: *Physica A*., 370, pp. 1–6

[5] Granger, C., Newbold, P., (1967), *Spurious regression in econometrics*, in: *Journal of Econometrics*, 2, 111-120

[6] Fixler, D.J., Green, B.T., (2005), *Reliability of the NIPA Estimates of U.S. Economic Activity*, retrieved February 26, 2007 from http://www.bea.gov/scb/pdf/2005/02February/0205_ NIPAa.pdf

[7] Agnion, P., Durlauf, S., (2005), *Handbook of Economic Growth*, Elsevier B.V.

[8] Hendry, D., Juselius, K., (2001), *Explaining Cointegration Analysis: Part II,* in: *Energy Journal*, 22, 75-120

[9] Johansen, S., (1988), *Statistical analysis of cointegrating vectors*, in: *Journal of Economic Dynamics and Control*, 12, 231-254

[10] Juselius, K., Franchi, M., (2007), *Taking a DSGE Model to the Data Meaningfully*, Economics Discussion Papers, No 2007-6. Retrieved March 5, 2007 from http://www.ejournal.org/economics/discussionpapers/2007-6

[11] Kitov, I., (2005), *A model for microeconomic and macroeconomic development*, Working Paper 05, ECINEQ, Society for the Study of Economic Inequality.

[12] Kitov, I., (2005), *Modeling the overall personal income distribution in the USA from 1994 to 2002*, Working Paper 07, ECINEQ, Society for the Study of Economic Inequality

[13] Kitov, I., (2005), *Modeling the average income dependence on work experience in the USA from 1967 to 2002,* Working Paper 11, ECINEQ, Society for the Study of Economic Inequality

[14] Kitov, I., (2006), *Modeling the age-dependent personal income distribution in the USA*, Working Papers 17, ECINEQ, Society for the Study of Economic Inequality.

[15] Kitov, I., (2006), *GDP growth rate and population*, Working Paper 42, ECINEQ, Society for the Study of Economic Inequality.

[16] Kitov, I., (2006), *Real GDP per capita in developed countries*, MPRA Paper 2738, University Library of Munich, Germany, http://ideas.repec.org/p/pra/mprapa/2738.html

[17] Kitov, I., (2006), *The Japanese economy*, MPRA Paper 2737, University Library of Munich, Germany, http://ideas.repec.org/p/pra/mprapa/2737.html

[18] Kitov, I., (2006), *Inflation, Unemployment, Labor Force Change in the USA*, Working Paper 28, ECINEQ, Society for the Study of Economic Inequality.

[19] Kitov, I., (2006), *Exact prediction of inflation in the USA*, MPRA Paper 2735, University Library of Munich, Germany, http://ideas.repec.org/p/pra/mprapa/2735.html

[20] Kitov, I., (2007), *Inflation, Unemployment, Labor Force Change in European countries*, in T. Nagakawa (Ed.), *Business Fluctuations and Cycles*, pp. 67-112, Hauppauge NY: Nova Science Publishers.

[21] Kitov, I., Kitov, O., Dolinskaya, S., (2007), *Inflation as a function of labor force change rate: cointegration test for the USA*, MPRA Paper 2734, University Library of Munich, Germany, http://ideas.repec.org/p/pra/mprapa/2734.html

[22] Kitov, I., Kitov, O., Dolinskaya, S., (2007), *Relationship between inflation, unemployment and labor force change rate in France: cointegration test*, MPRA Paper 2736, University Library of Munich, Germany, http://ideas.repec.org/p/pra/mprapa/2736.html

[23] Kitov, I., (2008), *Modeling the evolution of age-dependent Gini coefficient for personal incomes in the U.S. between 1967 and 2005*, MPRA Paper 10107, University Library of Munich, Germany, http://mpra.ub.uni-muenchen.de/10107/

[24] Kitov, I., Kitov, O., (2008), *The driving force of labor productivity*, MPRA Paper 9069, University Library of Munich, Germany, http://ideas.repec.org/p/pra/mprapa/9069.html

[25] Kitov, I., Kitov, O., (2008), *The Driving Force of Labor Force Participation in Developed Countries*, in*: Journal of Applied Economic Sciences*, *Spiru Haret* University, Faculty of Financial Management and Accounting Craiova, vol. II_3(5)_Fall2008, pages 203-222.

[26] National Bureau of Economic Research, (2007), *Business Cycle Expansions and Contractions*, retrieved March 5, 2007 from http://www.nber.org/cycles/cyclesmain.html#announcements

[27] Ormerod, P., (2005), *Complexity and the limit of knowledge*, Futures, v. 37

[28] U.S. Census Bureau, (2002), *Technical Paper 63RV: Current Population Survey - Design and Methodology*, issued March 2002, retrieved February 26, 2007 from http://www.census.gov/prod/2002pubs/tp63rv.pdf

[29] U.S. Census Bureau, (2004), *National intercensal estimates (1990-2000)*, retrieved February 26, 2007 from http://www.census.gov/popest/archives/methodology/ intercensal_nat_meth.html

[30] U.S. Census Bureau, (2006), *Methodology. United States Population Estimates by Age, Sex, Race, and Hispanic Origin: July 1, 2005*, retrieved February 26, 2007 from http://www.census.gov/popest/topics/methodology/2006_nat_meth.html

[31] U.S. Census Bureau, (2007), *Population Estimates*, retrieved February 26, 2007 from http://www.census.gov/popest/estimates.php

[32] West, K.K., Robinson, J.G., (1999), *What Do We Know About the Undercount of Children?*, US Census Bureau working paper 039, issued August 1999, retrieved March 5, 2007 from http://www.census.gov/population/www/documentation/twps0039/twps0039.html

[33] Yakovenko, V., Dragulescu, A. (2001), *Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States,* in: *Physica* A, v. 299, pp. 213-221.

Journal of Applied Economic Sciences
Volume IV/ Issue 1(7)/ Spring 2009

DETERRENCE OF A CRIMINAL TEAM: HOW TO RELY ON ITS MEMBERS' SHORT COMINGS ?

Eric **LANGLAIS**[1 2]
EconomiX-CNRS-Paris X and CEREFIGE-Nancy University
eric.langlais@univ-nancy2.fr

Abstract:

*In this paper, we assume that a criminal organization is an agency where the Principal and the Agent have different sensibilities towards the risk of arrestation and punishment, and at the same time have different skills with respect to general organization tasks, crime realization or detection avoidance activities (*i.e. *allowing to reduce the probability of detection). In this set up, we first compare two regimes of exclusive sanctions (either the sanctions are borne by the Principal/beneficiary of the crime,* or *they are borne by the Agent/perpetrator of the crime), and we analyze the comparative efficiency of the various instruments which are at the disposal of public authorities to prevent corporation in criminal activities (frequency of control and level of monetary penalties). Finally, we study a case with joint liability.*

Keywords: Criminal teams, corporate criminality, state dependent risk aversion, deterrence, monetary penalties *versus* detection.

JEL Classification: K13, K4.

1. Introduction

A criminal organization may be seen as an agency where the Principal and the Agent have different sensibilities towards the risk of being caught and punished, and at the same time have different skills with respect to general organization tasks, crime realization or detection avoidance activities (*i.e.* allowing to reduce the probability of detection). We introduce a basic framework of criminal groups activity along this line in order to compare two regimes of exclusive sanctions (either the sanctions are borne by the Principal/beneficiary of the crime, *or* they are borne by the Agent/perpetrator of the crime) depending on their social benefits 1/ in terms of their deterrence effects of crime, 2/ in terms of the probability of detection. Then we analyze the comparative efficiency of the various instruments which are at the disposal of public authorities to prevent corporation in criminal activities (frequency of control and level of monetary penalties).

Concerning criminality team, it is often claimed in public opinion that sanctions which are inflicted to the sleeping partner in a crime, or to the beneficiary of a fraudulent act should be more heavy than the punishments applied to the perpetrator of the offences who may be sometimes entrapped in the crime. On the other hand, that public deterrence of criminal activity focuses at first on perpetrators may be explained by the existence of large costs of detection of Principals, as compared to the small monitoring costs of Agents. Our results suggest that, as far as there is no strict constraints on the resources allotted to the control and repression of criminal activities (public authorities may levy more and more resources to develop this activity), then the regime with exclusive sanction on the perpetrator always allows to obtain larger deterrence effects on Principals than the regime of exclusive sanction on the Principal, and that these advantages increase with the probability of public control and detection. On the other hand, the regime of Agent's liability gives him more incentives to cheat, thus leading to a smaller probability of detection of criminal activities.

We also find that under mild conditions about the choice of fines, the regime of joint liability/mixed sanctions allows to obtain larger effects in terms of crime deterrence than the regime of exclusive sanctions upon the Principal; nevertheless, this also yields more difficulties in detecting

[1] *Contact:* University Nancy 2 - UFR AES, 4 Rue de la Ravinelle, CO 7026, 54035 Nancy cedex, France.

[2] I am thankful to Paul Calcott, Giuseppe Dari-Mattiacci and Francesco Parisi for valuable comments and suggestions on different previous versions of this work, and the participants to the 2nd French-German Talks in L&E organized by the Center for the Study of Law and Economics – Universität des Saarlandes (December 2005). The usual disclaimer applies.

the crime, since the Agent have more incentives to invest in activities avoiding public detection. In contrast, it yields (under reasonable conditions regarding the choice of fines) less crime deterrence but more detection than a regime of unilateral sanction upon the Agent.

Section 2 gives some related works on the topics. Section 3 describes our framework and the main behavioral assumptions of the model. Then, in section 4, we characterize the optimal contract (effort,monetary transfer) between the Principal and the Agent, in each regime of sanction (exclusive sanction upon the Principal *versus* exclusive sanctions upon the Agent). We show that exclusive sanctions upon the Agent allow to obtain more compliance from the Principal, but induce higher efforts in cheating from the Agent. Section 5 displays the complete comparative statics of the model. Our results suggest that, as far as there is no strict constraints on the resources allotted to the control and repression of criminal activities (public authorities may levy more and more resources to develop this activity), the advantages in favour of a regime of exclusive sanctions on Agents increase with the probability of public control and detection. Section 6 studies the case for joint liability. We show that it implies more deterrence than the regime of exclusive sanction upon the leader, but at a cost since "~joint liability~" also reduces the effectiveness of public monitoring as compared to the second regime, once again. It means that it is more efficient to punish both partners in a criminal teams than the "stronger partner" (leader) alone, since the probability of crime is smaller when both bear a sanction - albeit, the frequency of detection of the team is smaller. In contrast, there is unfortunately no reason to believe that it is also more efficient to punish both partners than the "~weaker~" alone; maybe this is not the case: more conditions over basic assumptions regarding the technology of avoidance and individual preferences are required (not found here). Section 7 gives a brief conclusion.

2. Related literature

At least three kinds of literature are worth mentioning, since this work may be connected to each of them: corporate crime, gatekeepers' liability, and criminal organizations.

Corporate crime. At a basic level, the issue of corporate crime has been studied first with reference to employer-employee relationships. Arlen (1994) puts the distinction that corporate crime is not crime committed by corporations, but corresponds to illegal activities undertaken by individuals belonging to a corporation, but pursuing their own selfish interest - even though the offence incidentally benefits the corporation. Thus, the rational for corporate liability is that public monitoring of corporate crimes is most of the time difficult and costly to implement, enabling not enough (inefficient) deterrence, while in contrast, corporations have an advantage in detecting wrongdoings by their employees. As a matter of fact, vicarious criminal liability (shifting liability from employees to the corporation) increases corporation's enforcement expenditures, thus increases the probability of detection and then reduces the number of malicious employees who commit the offence. Nevertheless, Arlen's analysis highlights the potentially perverse incentives of strict vicarious liability, coming from opposite forces created by the very increase in enforcement expenditures: as the probability of corporation's employees detection increases, so is the probability of public detection, allowing the government to increase the corporation's expected liability for these crimes. As a result, strict vicarious liability may finally lead a corporation to spend less in enforcement than it would absent of vicarious liability: this occurs soon as the benefits of the reduction in the number of offences committed by their employees fall well down the expected costs associated to enforcement and liability. Chu and Qian (1995) have argued that one solution to overcome the problem is to introduce vicarious liability under a negligence rule. Specifically, courts may provide enough incentives to induce from Principal honest reports either by lowering the due care level or by lowering the level of delegated liability.

Shavell (1997) observes that corporations have limited ability to penalize their wrongdoing employees. Although standard economic analysis of compensation concludes that optimal damages must be set to the value of harm, in practice corporations impose limited penalties on their employees for causing harm to third parties, the major sanction imposed being to be dismissed from their job. One way firms may remedy this problem is by paying supernormal (above-market) wages: this is because when the employee have more to lose when he commits an error or an offense and is dismissed. Thus supernormal wages induce more care and prevention of accidents. Nevertheless, private incentives of firm to the use of above-market wages deviate from social ones. On the one hand, this reflects that supernormal wages correspond to additional costs for a firm, whereas it is not for society. On the other, market price of firm product may exceed social costs, as they reflect the increase in the private costs following the adoption of supernormal wage. Thus the undesirable decrease in consumers' purchases may render the setting of wage at a lower level than harm socially advantageous.

Gatekeepers' liability: auditors and lawyers. Recent corporate scandals (Enron, Worldcom) since the beginning of 2000s, and the even more recent subprime crisis which has been the departure of a major financial crisis at an international level, have made apparent the essential role of professional service providers, such as auditors, corporate lawyers, and securities analysts, in detecting and revealing corporate misconduct on the part of their clients and at the same time, the failure of the internal control inside large diversified financial organizations. Kraakman (1986) has revised the classical gatekeeping theory, clarifying that a gatekeeping strategy requires gatekeepers *who can and will prevent misconduct reliably, regardless of the preferences and market alternatives of wrongdoers*. After the historic debacle of Enron, there has been a renewal of the debate, beginning with Coffee (2002) who has argued that *Enron is a maddeningly idiosyncratic example of pathological corporate governance, which by itself cannot provide evidence of systematic governance failure* but properly understood it can explain why and when reliance may not be justified on "reputational intermediaries," such as auditors, securities analysts, attorneys. His proposal is to convert gatekeepers into insurers, but cap their insurance obligations based on a multiple of the highest annual revenues the gatekeepers recently had received from their wrongdoing clients. Partnoy (2004) makes the point that the problem of gatekeeper liability is a shift in scholarship view which had more focused on reputation than on regulation or civil liability. Many scholars have argued that liability should not be imposed on gatekeepers and that reputation-related incentives alone would be sufficient. Partnoy compares various proposals and concludes that a contractual system based on a percentage of the issuer's liability would be preferable to a regulatory system with caps based on a multiple of gatekeeper revenues. Schäfer (2004) argues that a wrong audit that causes damages to shareholders should be strictly regarded as a tort case. We also argue that a rule of gross negligence or of gross violation of professional standards in tort law can avoid the problems of underdeterrence as well as of overdeterrence in the compensation of pure financial loss in tort. However, we also argue that a wrong audit should lead to contractual liability, if it was made to prepare the sale of a company or parts of it from inside investors to outside investors or to prepare an initial public offering. Ganuza and Gomez (2007) consider a framework where they analyze the imposition of duties of care and reporting on gatekeepers, conditional on their having observed an underlying wrongdoing or misconduct of their clients. They make the assumption that the gatekeeper observes the state of the world affecting misconduct, and that the public authority (Courts or a regulator) are unable to verify whether misconduct had or not been observed by the lawyer or auditor. However it is costlessly verifiable ex post. The main results are that standards of professional behavior by auditors or lawyers may well be sufficient as incentives, and that the implications of the model tend to imply that the distinction between voluntary violation of duties and mere negligence is not very useful contrary to what is specified in existing Laws.

Criminal organization and corruption. A specific case of a crime in a team is corruption. Marjit and Shi (1998) and Jacquemet (2006) provide two different surveys of the strategic approach of corruption. One of the main issue is not how to punish, but in contrast how to reward a corrupted official in order to better control crime. Chappe and At (2005, 2008) develop two dynamic models of crime with and without information acquisition and study the conditions under which it is optimal for a criminal to delay commission of a crime rather than committing it immediately. They address the issue of the optimal fine and level of deterrence. However, they do not consider the question of how liability may be allocated among the gang members. Garoupa (2000) provides a comparison of criminal organizations such as mafias with governments, and analyses the optimal contract between a gang member and the gang authority. The mafia extorts a rent to criminals when they commit a crime, which runs as a barrier to entry on the criminal market that enables public authorities to save on enforcement costs. Nevertheless, the models do not take into account the interactions within the gang (the mafia is a moral person) and thus does not consider the issue of liability of the Principal.

A main reason explaining why the allocation of liability is neutral in these works, comes from the fact that this a pure transfer between both risk neutral parties, according to Shavell's argument.

Privileggi and ali (2001) have focused on what occurs when one of the party is risk averse. They assume that the (risk neutral) Principal may delegate to the (risk averse) Agent the realization of the wrongdoings, thus leading to an agency problem. The Agent exerts an effort that negatively affects the probability of detection, but bears a cost in utility terms. The Principal has the opportunity to perfectly observe the decision of the Agent. In this framework, Privileggi and ali show that shifting the liability upon the Agent, all else equal (holding constant both the monetary sanctions and the probability of public detection), allows to obtain larger deterrence effects on the Principal than when sanctions only affect the Principal. Hence, a regime of exclusive sanctions (liability) on the Agent favors a better public monitoring of illegal activity. Nevertheless, they also show that the regime of exclusive sanction on the Agent may induce him to exert more effort in cheating, which leads to countervailing incentive effects on individuals who still find illegal behavior profitable: given that the probability of detection may be lower than in the other regime, public authorities may be faced with greater difficulties in repressing illegal activities. In the case where the Agent displays constant absolute risk aversion, Privileggi and ali show that there exists a kind of complementarity between the level of monetary sanctions and the level of probability of detection: a calibration of their model exhibits that for any level of the CARA index, there always exists a combination consisting in a probability of public detection and a level of sanctions which allows to implement a small effort of cheating. Specifically, strong public enforcement policies based on large probability of detection and at the same time high level of monetary sanctions render the shift of the responsibility upon the Agent socially beneficial - that is, it allows more Principal to renounce illegal behavior, and facilitates public detection since Agents undertake less efforts in cheating.

Our papers takes a different view regarding two main behavioral assumptions.

One the one hand, a the main feature of the analysis of Privileggi and ali (2001) makes their framework close to the basic agency model of the employer-employee - the Agent's cost of the effort is expressed in utility terms. In contrast, our paper introduces an alternative specification for the technology of effort. We assume here that the illegal activity is costly, *i.e.* individual who engage in it has to spend money coming from the use of productive factors, which may be scarce and specific. The rational for that is that when such offences are developed at a large scope, they become strictly speaking a parallel economy requiring at a basic level the coordination and cooperation between several individuals. Each may be endowed with specific human and non human assets which are worth for this activity: hence, corporate criminality is an industry. Remark that when the assumption about individuals' risk neutrality is relaxed, the issue of the nature of the cost of effort matters, although it is irrelevant in the case where individuals are risk neutral. Is it a monetary cost, having the characteristics of the implicit technology of production to which the effort is associated; or is it only the disutility of effort - which may have monetary equivalent, but may be complementary to wealth? The case for the monetary cost of effort, which seems actually to be the widespread interpretation in the literature on Law & Economics, is attractive but not simply for its realism. As it will be seen, it has finally the main advantage to enable a complete characterization of the various regimes of sanction, and specifically non ambiguous effects when we proceed to a complete comparative statics analysis.

On the other hand, in this paper we allow a richer representation of Agent's preferences; we introduce the state-dependent approach more usually used in the literature on safety and/or value of life analysis. The rational for such an assumption is that, although we do not address directly this issue here, these state-dependant preferences allow to take into account the fact that public authorities may use non monetary sanctions where criminal activities are detected and criminal individual are arrested, such as: imprisonment, full or partial loss of civic rights and so on. The basic intuition says that individual even when engaged in illegal activities have an intrinsic preference for freedom (not being detected), and to the extent that going on with their wrongdoing is worth to them, all else equal, such that they are better off when they have the opportunity to cheat than when cannot. This is specifically what the state-dependent representation allows with a parsimonious model of individual preferences.

3. The model

Consider the case where the criminal activity or illegal behavior allows the Principal to obtain a payment equal to $B > 0$. The Principal is not the perpetrator of this malicious act, but he delegates[3] it to an Agent, who bears a monetary cost given by $C(x) = x$. On the other hand, the Principal rewards Agent's efforts with a payment equal to $w > 0$ in case of success, *i.e.* when the offences act has not been detected by the authority. Let the aggregate technology of monitoring be characterized by $p(x) = p_0 - \hat{p}(x)$, where p_0 denotes public monitoring such as the frequency of control by the public authority, and $\hat{p}(x)$ corresponds to the private efforts of concealing the illegal activity undertaken by the Agent, where x is his effort or expenditures in wrongdoing.

Assumption 1:

1.1: $\hat{p}^{'}(x) = -p'(x) > 0, \quad \hat{p}^{''}(x) = -p''(x) < 0$;

1.2: $\lim_{x\to\infty} \hat{p}(x) = p_0$

Assumption 1.1 says that even if the Agent has the opportunity to run counter to public efforts of detection, the marginal return of effort in cheating become smaller and smaller - there nevertheless exist decreasing returns to scale in this activity. Assumption 1.2 is introduced essentially for technical reasons. It may be argued that as a matter of fact, the Agent is not allowed to increase without any upper limit his effort (x may be only define on a subset $[0, x_{\max}]$, with $x_{\max} < \infty$: hence, once the maximum possible value is reached, public authorities may obtain any deterrence effect without the threat of induced effects on the probability of detection). Nevertheless, from a social point of view the issue is how much does it cost to reach high values for x ? The final section partly investigates this point. We do not introduce explicitly the social welfare objective, but we address the problem of the impact of the various instruments at the disposal of public authority, depending on whether or not public authorities have limited resources in the monitoring of illegal activities.

Let us focus on assumptions about individuals' preferences. We assume that the Principal is a risk neutral individual. In contrast, the Agent is supposed to be a risk averse one, with a state-dependent representation of his preferences, where u_1 denotes his utility index when he is not detected, while u_0 denotes his utility index when he is detected. Both utility functions are supposed to be strictly increasing and strictly concave, and unique up to an affine transformation.

Although this is not the central topic of this paper, this representation enables to take into account various effects which are associated to the detection of illegal activities: the Agent is suffering a psychological penalty when he is caught, or more generally public authorities may apply non monetary sanctions in case where the fraudulent acts have been observed, such as imprisonment, full or partial loss of civic rights and so on. Hence all else equal, and to be short, we consider a case where the Agent is better off when he is not detected as compared to the case where he is. In order to represent this situation and be exhaustive about this issue, we require additional assumptions to hold[4]:

Assumption 2: at each level of Agent's wealth:

2.1: u_1 ¤ u_0,

2.2: $u_1^{'} \geq u_0^{'}$,

2.3: $-\frac{u_1^{''}}{u_1^{'}} \geq -\frac{u_2^{''}}{u_2^{'}}$.

[3] Monitoring the Agent's efforts may be a costly activity for the Principal - nevertheless, assuming it entails only fixed costs, under perfect information these may be seen as negligible (set to zero) and are ignored in the following analysis.

[4] They have been extensively discussed and justified in the literature on self-protection expenditures and/or willingness to pay for safety, health and life: see Dehez and Drèze (1987), Jones-Lee (1974) for example.

While these would appear as very strong conditions at first glance, they have a great and intuitive appeal. The first one (2.1) implies generally that being caught is never beneficial for the Agent: whatever his wealth (accumulated through illegal activities), he is better off when he escapes from public detection. Relaxing such a condition would imply that the Agent has an incentive to give himself up[5], which may be seem very strange on a priori grounds (up to pathological behaviors); more over, such an assumption would introduce a bias in the following analysis in favor of Agent's liability. The second restriction (2.2) means that each additional Euro has more value in the state where the Agent is not detected than in the state where he is: the marginal contribution of additional units of wealth to his welfare is larger when he is not detected than when he is. Such a condition is required as far as we consider that the effort corresponds to a normal good[6].

Finally, (2.3) allows the Agent to be more sensible to risk when he is not detected than when he is: according to (2.3), u_1 displays more risk aversion than u_2 at all level of Agent's wealth. While this last assumption proved to be more meaningful if we were to consider that public authorities may use the non monetary sanctions with different intensities but which cannot be precisely known by the Agent, it is sufficient to obtain that the Willingness to Pay for Safety is also a normal good; on the other hand, it is introduced here for practical motivations which will be clearer later on[7].

4. The consequences of exclusive sanctions

4.1. Exclusive sanctions upon the Principal

The expected benefit obtained by the Principal when the illegal activity is performed is equal to:

$$v_P \equiv \max_{w,x}\left(B - w - p(x)S\right),\ s.t.:$$
$$(1.1): u_1(w - x) \geq k$$
(1)

Assuming that the Principal is a monopolist, we focus solely on the case where the participation constraint of the Agent binds. In a such a situation, the Principal demands an effort which is such that:

$$\frac{1}{\hat{p}'(x_P)} = S \qquad (2)$$

and pays a reward to the Agent equal to:

$$w_P = u_1^{-1}[k] + x_P \qquad (3)$$

By analogy with the literature on safety and/or the value of life, let us consider that condition (2) says that for the equilibrium value of the expenditures allotted to cheating and concealing the illegal activity (in terms of the level of Agent's effort) then, the Social Marginal Cost of Safety (LHS of (2)) which is nothing but the ratio between the Marginal Cost of Cheating borne by the Agent to the marginal productivity of Cheating for the Principal, is just equal to the Willingness to Pay for Safety of the risk neutral Principal (RHS in (2)). Condition (3) means that the payment to the Agent

[5] Remark that a less stringent condition may be obtained assuming only that $u_1(w_1) \geq u_0(w_0)$ at least whenever $w_1 \geq w_0$, *i.e.* soon as the Agent's wealth when not detected is higher than when he is, which is always satisfied in the following analysis.

[6] That is, the effort increases when the Agent becomes richer. While we do not explicitly address the issue here, the reader may refer to Dehez and Drèze (1987, proposition 3.2) for an explicit analysis in the case of individual safety expenditures.

[7] Just remind that assumption 3 is usual in the literature when comparative statics analysis is performed.

has to cover the total cost of his expenditures in safety, up to some fixed cost (the constant $u_1^{-1}[k]$) which may be understood as representing the external opportunities of the Agent.

4.2. Exclusive sanctions upon the Agent

Consider now that liability is delegated to the Agent, but that his choice corresponding to the level of effort can be perfectly monitored by the Principal. Now, the expected benefit obtained by the Principal when the illegal activity is performed is equal to:

$$\begin{aligned} & v_A \equiv \max_{x,w}(B-w),\ s.t.: \\ & (4.1): U(x,w) \geq k \end{aligned} \tag{4}$$

with $U(x,w) = p(x)u_0(w-x-S) + (1-p(x))u_1(w-x) = Eu.$ In a such a situation, the Agent affords an effort such that[8]:

$$\frac{1}{\hat{p}'(x_A)} = \frac{u_1 - u_0}{Eu'} \tag{5}$$

where $Eu' = p(x_A)u_0'(w-x_A-S) + (1-p(x_A))u_1'(w-x_A)$, and the Principal pays the associated reward which is:

$$w_A = u_1^{-1}[k] - ec_A \tag{6}$$

where: $ec_A < 0$ denotes the certainty-equivalent[9] (expressed at w_A) of the risky prospect denoted as: $[(p(x_A), -x_A - S); (1-p(x_A), -x_A)]$. Interestingly enough, condition (5) says now that the equilibrium value of the expenditures of safety (level of effort) chosen by the Agent when he is liable, is such that the Marginal Cost of Safety (LHS of (5)), is just equal to the Willingness to Pay for Safety of the risk averse Agent (RHS in (5)). Condition (6) means that the payment to the Agent has to cover the cost adjusted of the price of the risk associated to his expenditure in cheating - and still up to some fixed cost (the constant $u_1^{-1}[k]$).

4.3. Comparative analysis of deterrence effects

One of the main attractive features of the present model is that it allows a complete non ambiguous comparative statics analysis of the optimal contract. To see this, let us first consider the main result of the paper which is the following:

Proposition 1. *All else equal:*

i) $v_A \leq v_P$.

ii) $w_A \geq w_P$.

iii) $x_A \geq x_P$.

Proof i) By definition of v_A, v_P and ec_A, and finally using (6) and (3), we have:

---

[8] See appendix 1 for SOC.

[9] See appendix 2.

$$
\begin{aligned}
v_A &\equiv b - w_A \\
&= b - \left[u_1^{-1}[k] + x_A + ec_A\right] \\
&\leq b - u_1^{-1}[k] - x_A - p(x_A)S \\
&\leq \max_x \left(b - u_1^{-1}[k] - x - p(x)S\right) \\
&\equiv v_P
\end{aligned}
$$

Hence the result. ii) is a straightforward consequence of i), starting with $v_A \equiv B - w_A \leq v_P \equiv B - w_P - p(x_P)S$ implies that $w_P + p(x_P)S \leq w_A$; hence the result, given that $p(x_P)S > 0$.

iii) Consider now conditions (2) and (5); if the Agent were risk neutral, his WTP for safety would also be equal to S; as a result, it is not so obvious on a priori ground whether $x_A > or < x_P$. However, remark that by concavity of function u_1, it comes that:

$$u_1(w - x_A) \geq S.u_1'(w - x_A) + u_1(w - x_A - S)$$

Thus substracting each side with $u_0(w - x_A - S)$, then dividing by Eu' and rearranging yields:

$$
\begin{aligned}
&\frac{u_1(w - x_A) - u_0(w - x_A - S)}{Eu'} \\
&\geq S \times \frac{u_1'(w - x_A)}{Eu'} + \frac{u_1(w - x_A - S) - u_0(w - x_A - S)}{Eu'} \\
&\geq S \times \frac{u_1'(w - x_A)}{Eu'}
\end{aligned}
$$

given that $u_1(w - x_A - S) \geq u_0(w - x_A - S)$. Once more, by concavity of u_1, it is easy to check that for any $p(x) > 0$, then $\frac{u_1'(w - x_A)}{Eu'} \geq 1 \Leftrightarrow u_1'(w - x_A) \geq u_0'(w - x_A - S)$; hence:

$$\frac{u_1(w - x_A) - u_0(w - x_A - S)}{Eu'} \geq S$$

saying that any risk averse (with a state-dependent representation of preferences) decision maker will have a WTP for safety larger than a risk neutral decision maker. To conclude, just remark finally that $\left(\frac{1}{\hat{p}'(x)}\right)$ (LHS in (5)) is a increasing function of x : hence: $x_A \geq x_P$.

As it is easy to check, the way liability is allotted does not matter in the case where both the Principal and the Agent are risk neutral, since the level of effort and the reward paid to the Agent are equal to x_P and w_P respectively. In contrast, the way sanctions are allotted matters under risk aversion as far as efficiency in liability setting is concerned. Part i) of proposition 1 says that shifting liability from Principal to Agent entails more deterrence effects on sleeping partners in criminal and illegal activities: all else equal, the participation constraint of the Principal binds in the Agent's liability system before it binds in the Principal's liability one. But a straightforward consequence of part iii) of proposition 1 is the following:

Corollary 2. *All else equal, the probability of detection is smaller in the Agent's liability regime than in the Principal's liability regime.*

The driving force behind this last result is the size of the willingness to pay of the team's member who is liable. In each pure regime of sanction and punishment, the Agent's expenditures in cheating and concealing the illegal activity basically reflects the willingness to pay of the party in the crime who will bear the sanction in case of detection. In fact, in a state-independent context, a risk neutral individual may have a willingness to pay higher than a risk averse one[10]: this explains the ambiguous findings by Privileggi, Marchese and Cassone (2001). In contrast, in a state-dependent world, a risk neutral individual always has a willingness to pay smaller than a risk averse one. As a result, Agent's efforts in avoiding detection and concealing the illegal activity are larger when he is liable than when the Principal is. Hence shifting liability from Principal to Agent may be socially worth according to deterrence effects on Principals. But on the second hand, it becomes less easy to detect illegal activities, since shifting the burden of the liability on the perpetrator of wrongdoings gives him more incentive to cheat. These countervailing effects are well known since Arlen (1994) and Shavell (1997). This also confirms the intuitions developed by Sanchirico (2005).

5. Comparative statics: probability *versus* penalty

In a second step, we compare the impact of the instruments which are available to public authorities namely the level of monetary sanctions and the frequency of control. An interesting point which deserves to be highlighted is that when the Principal is legally liable, thus the main instruments at the disposal of authorities to repress crimes and dishonest behaviors, namely (p_0, S) have far different effects depending on the regime of liability.

Proposition 3. *All else equal, in the regime of Principal's liability:*
i) The analysis of the comparative statics gives:

	w_P	x_P	v_P
p_0	*independent*	*independent*	$-$
S	$+$	$+$	$-$

ii) Increasing the frequency of controls have more deterrence effects on Principals than raising the level of monetary sanctions.

Proof i) To begin with, the frequency of control of illegal activities - *i.e.* the choice of p_0 by the authority - has no effect neither on the effort undertaken by the Agent, nor on the payment he obtains from the Principal. Further more, it has no effect of the Agent's utility level, as far as his participation constraint always binds. An increase in p_0 simply reduces the expected outcome of the Principal in this case:

$$\frac{\partial v_P}{\partial p_0} = -S < 0$$

In contrast, an increase in the penalty S paid by the Principal when he is detected induces effects, on the one hand, on the activity of the Agent and on the payment he receives, since the Agent increases his effort and receives a higher payment (the Principal gives him more incentives to invest in effort):

$$\frac{\partial x_P}{\partial S} = \frac{1/S}{\frac{p''(x_P)}{-p'(x_P)}} > 0$$

$$\frac{\partial w_P}{\partial S} = \frac{\hat{p}'(x_P)}{\frac{p''(x_P)}{-p'(x_P)}} > 0$$

[10] See Langlais (2005) fro a general analysis.

and on the expected utility level of the Principal on the second:

$$\frac{\partial v_P}{\partial S} = -p(x_P) = -(p_0 - \hat{p}(x_P)) < 0$$

the Agent being not affected by the increase in S.

ii) Soon as $S > 1$, $S > p(x_P)$, hence the result.

To conclude for the moment about this regime, the monetary sanction proved to be a less efficient instrument to deter crime as compared to the frequency of control, since the (direct) impact of the latter instrument on Principal's satisfaction level is larger (in absolute value): hence, the reservation utility of Principal may be more easily reached, all else equal, without pervasive effects on the probability of detection - since the increase in p_0 has a one to one effect on the total probability of criminals' detection $p(x) = p_0 - \hat{p}(x)$, without any induced (additional) effects on $\hat{p}(x)$ which may come from the protective measures undertaken by the Agent. To the contrary, given that any increase in the monetary sanction S leads the Agent to produce more effort, so that the total probability of detection decreases, this second instrument has effects which are more uncertain: the higher the sanction, the smaller the probability of detection and the decrease in the Principal's utility.

When the Agent is liable, the comparative statics analysis provides richer results.

Proposition 4. *All else equal, in the regime of Agent's liability:*

i) The analysis of the comparative statics gives:

	w_A	x_A	v_A
p_0	$+$	$+$	$-$
S	$+$	$+$	$-$

ii) When x_A becomes large enough, raising the level of monetary sanctions entails more deterrence effects on Principal than increasing the frequency of controls.

Proof i) An increase in the probability of control first affects the utility level of the Principal:

$$\frac{\partial v_A}{\partial p_0} = -\frac{u_1 - u_0}{Eu'} < 0$$

with additional (induced) effects on the terms of contract between the Principal and the Agent:

$$\frac{\partial w_A}{\partial p_0} = \frac{u_1 - u_0}{Eu'} > 0$$

$$\frac{\partial x_A}{\partial p_0} = \frac{2\left(\frac{1}{\hat{p}'(x_A)}\right)\left[\left(\frac{u_1' - u_0'}{Eu'}\right) + \frac{1}{2}\left(\frac{1}{\hat{p}'(x_A)}\right)\left(-\frac{Eu''}{Eu'}\right)\right]}{\Omega} > 0$$

where $\Omega > 0$ by the SOC (see appendix 1). An increase in the sanctions have the following impact:

$$\frac{\partial v_A}{\partial S} = -p(x_A)\left(\frac{u_0'}{Eu'}\right) < 0$$

and the induced effects on the contract are:

$$\frac{\partial w_A}{\partial S} = p(x_A)\left(\frac{u_0'}{Eu'}\right) > 0$$

$$\frac{\partial x_A}{\partial S} = \frac{\left(\frac{u_0'}{Eu'}\right)\left[1 + p(x)\left(\frac{1}{\hat{p}'(x_A)}\right)\left(\left(-\frac{Eu''}{Eu'}\right) - \left(-\frac{u_0''}{u_0'}\right)\right) + p(x)\left(\frac{u_1' - u_0'}{Eu'}\right)\right]}{\Omega} > 0$$

Remark that the numerator in $\frac{\partial x_A}{\partial S}$ should be of any sign, in the absence of assumption 2.3 since, for any positive probability of being caught:

$$\left(-\frac{Eu''}{Eu'}\right) \leq or \geq \left(-\frac{u_0''}{u_0'}\right) \Leftrightarrow \left(-\frac{u_1''}{u_1'}\right) \leq or \geq \left(-\frac{u_0''}{u_0'}\right)$$

Hence, assumption 2.3 is sufficient to obtain the (intuitive) positive sign.

ii) It is easy to see that:

$$-\frac{\partial v_A}{\partial p_0} \leq or \geq -\frac{\partial v_A}{\partial S} \Leftrightarrow p_0 \leq or \geq p(x_A) + \frac{u_1 - u_0}{Eu'}$$

suggesting that as the Agent's effort attains large values, $p(x_A) \to p_0$, and thus the monetary sanctions may become more efficient than the probability of detection to deter Principals to undertake the illegal activity since $\frac{u_1 - u_0}{Eu'} > 0$.

As compared to the previous case, both instruments have now pervasive effects, since they give more incentives to the Agent in cheating: the higher the intensity of public intervention, the harder the detection of wrongdoing. Moreover, comparing the sensitivity of the Principal to each of the instruments between regimes, we obtain:

Corollary 5. *All else equal:*

i) The higher the frequency of controls, the larger the advantages of a Agent's liability with respect to a Principal's liability, in terms of Principal's deterrence.

ii) The larger the level of monetary sanction, the smaller the advantages of a Agent's liability with respect to a Principal's liability, in terms of Principal's deterrence.

Proof i) The proof is direct given that:

$$-\frac{\partial v_A}{\partial p_0} = \frac{u_1 - u_0}{Eu'} \geq -\frac{\partial v_P}{\partial p_0} = S \;:(A)$$

$$-\frac{\partial v_A}{\partial S} = (p_0 - \hat{p}(x_P))\left(\frac{u_0'}{Eu'}\right) \leq -\frac{\partial v_P}{\partial S} = p_0 - \hat{p}(x_P) : (B)$$

ii) Using condition (B), it is direct that as $x_P \to \infty$, then $-\frac{\partial v_A}{\partial S} = 0 = -\frac{\partial v_P}{\partial S}$, and thus $v_P \to v_A$.

This last result means that the probability of detection is more efficient in the Agent's liability system than in the Principal's one, while to the converse the monetary sanctions have more effects in the Principal's regime than in the Agent's one such that for large value of the sanctions both regimes tend to reach the same results on the deterrence of Principal, roughly speaking.

6. Joint liability and sanctions

For practical purposes, a main criticism against the previous analysis is that generally speaking, the penal code imposes that all the members in a criminal teams will be punished in case of arrest. We now analyze in our framework the consequences of joint liability in terms of deterrence.

Assume now that for a given level of public expenditures in deterrence of criminal activities, the probability that the Agent be caught is $p(x)q$ where $q \in (0,1)$, while the probability that the Principal be caught is $(1-\theta)p(x)$ where $\theta \in (0,1)$. Let us denote f_P and f_A the penalty inflicted respectively to the Principal and the Agent in case of arrest. The optimal contract $(\hat{x},\hat{w})$ is the solution to:

$$\max_{w,x}\{(b-w-(1-\theta)p(x)f_P)\ s.t.:\ U(x,w) \geq k\}$$

where: $U(x,w) = p(x)qu_0(w-f_A-x)+(1-p(x)q)u_1(w-x)$ corresponding to an effort which satisfies:

$$\frac{1}{-p'(\hat{x})} = q \times \frac{u_1(\hat{w}-\hat{x})-u_0(\hat{w}-f_A-\hat{x})}{Eu'} + (1-\theta)\times f_P \tag{7}$$

with $Eu' = p(\hat{x})qu_0'(w-f_A-\hat{x})+(1-p(\hat{x})q)u_1'(w-\hat{x})$, and a monetary transfer given by:

$$\hat{w} = u_1^{-1}[k]+\hat{x}+p(\hat{x})qf_A+\pi_1(\hat{w}-\hat{x}-p(\hat{x})qf_A) \tag{8}$$

The structure of the RHS in (7), which is defined roughly speaking as a weighted sum of two individual WTP, reflects that each one of the Principal and the Agent have now to bear a specific kind of risk: on the one hand, the risk of being detected; on the second, the risk of being arrested and punished. These risks are reallocated through the contract; as a consequence, the contractual effort is tailored to both the willingness to pay of the Agent (first term in the RHS of (7)) and the willingness to pay of the Principal (second term in the RHS of (7))[11].

In the next proposition, we compare the regime of joint liability and sanction with the case of exclusive sanction upon the Principal.

Proposition 6. *Assume that* $(1-\theta)f_P \leq f \leq qf_A+(1-\theta)f_P$ *; then:*

i) $\hat{x} \geq x_P$.

ii) $\hat{v} \leq v_P$.

iii) $\hat{w} \geq w_P$.

Proof i) Let us compare the RHS in (7) and (2), for a given value of (w,x) ; by assumption of concavity, the first term in the RHS of (7) satisfies: $\frac{u_1-u_0}{Eu'} \geq f_A$ (see the proof of part iii) in proposition 1). Hence, under the assumption $f \leq qf_A+(1-\theta)f_P$, we have:

[11] Thus, condition (7) mimics the Bowen-Lindahl-Samuelson condition which characterizes the optimal production of public goods; detection avoidance may be seen as a *club good*, whose optimal level has to be tailored to the willingness to pay of the members of the club up to the specific investments in prosecution and punishment of enforcers.

$$q\frac{u_1-u_0}{Eu'}+(1-\theta)f_P\geq qf_A+(1-\theta)f_P\geq f$$

According to (7) and (2), this implies that $\frac{1}{-p'(\hat{x})}\geq\frac{1}{-p'(x_P)}$; hence the result follows.

ii) According to (8), we obtain:

$$\begin{aligned}\hat{v}&\equiv b-\hat{w}-(1-\theta)p(\hat{x})f_P\\&\leq b-u_1^{-1}[k]-\hat{x}-p(\hat{x})\left(qf_A+(1-\theta)f_P\right)\\&\leq\max_x\left(b-u_1^{-1}[k]-x-p(x)\left(qf_A+(1-\theta)f_P\right)\right)\end{aligned}$$

which is nothing else but the definition of v_P where $qf_A+(1-\theta)f_P$ has been substituted for f; but as it is easy to verify (applying the envelop theorem), v_P is a decreasing function of f. Thus, it is straightforward to see that if: $f\leq qf_A+(1-\theta)f_P$ then we also have:

$$\begin{aligned}\hat{v}&\leq\max_x\left(b-u_1^{-1}[k]-x-p(x)\left(qf_A+(1-\theta)f_P\right)\right)\\&\leq v_P=\max_x\left(b-u_1^{-1}[k]-x-p(x)f\right)\end{aligned}$$

Finally, part iii) is a straightforward consequence of part i) and ii): if $(1-\theta)f_P\leq f$, we have according to i):

$$\hat{w}-w_P\geq p(x_P)f-(1-\theta)p(\hat{x})f_P\geq p(x_P)\left(f-(1-\theta)f_P\right)\geq 0$$

given that $p(\hat{x})\leq p(x_P)$.

Proposition 6 implies that for mild conditions about the choice of fines, the regime of joint liability/mixed sanctions allows to obtain larger effects in terms of crime deterrence than the regime of exclusive sanctions upon the Principal; nevertheless, this also yields more difficulties in detecting the crime, since the Agent have more incentives to invest in activities avoiding public detection.

A direct comparison between (5) and (7) is quite intractable. Thus, in order to compare the regime of joint liability and sanction and the case of exclusive sanction upon the Agent, let us proceed in a different way. Intuitively speaking, there is no discontinuity between the regime of mixed sanction/liability and the regime of exclusive sanction upon the Agent: it is easily seen that starting from the former and raising continuously the probabilities θ and q up to their maximal value 1, we reach the second regime. Hence, we first study the comparative statics in the regime of joint sanction; then, we use these results in order identify the sufficient conditions which are required to compare both regimes. In appendix 3, it is shown that:

Proposition 7. *All else equal:*

i) $\hat{x}$ *is an increasing function of* q *and a decreasing function of* θ*; moreover, if* $f_A\geq f_P$ *then* $\hat{x}$ *is more sensible to* q *than t* θ*, i.e.* $\frac{\partial\hat{x}}{\partial q}>-\frac{\partial\hat{x}}{\partial\theta}$.

ii) $\hat{v}$ *is a decreasing function of* q *and an increasing function of* θ*; moreover, if* $f_A\geq f_P$ *then* $\hat{v}$ *is more sensible to* q *than to* θ*, i.e.* $-\frac{\partial\hat{v}}{\partial q}>\frac{\partial\hat{v}}{\partial\theta}$.

The previous propositions implies that the level of avoidance activity which is the solution to (7) when the fine and the probabilities of detection and punishment are set to the level (f_A,q,θ) is larger than when those variables are equal to $(f_A,1,1)$; *i.e.* $\hat{x}(f_A,q,\theta)\leq\hat{x}(f_A,1,1)$ - and in contrast the corresponding expected profit for the Principal is smaller: *i.e.* $\hat{v}(f_A, q, \theta) \leq \hat{v}(f_A, 1, 1)$. Given that $\hat{x}(f_A, 1, 1)$ is by definition the solution to (5) where f has been substituted by f_A, and that x_A is an increasing function of f, a straightforward consequence is:

Corollary 8. *Assume that* $f_P \leq f_A \leq f$ *; then:*

i) $\hat{x} \leq x_A$.

ii) $\hat{v} \geq v_A$.

The regime of mixed sanction entails (under reasonable conditions regarding the choice of fines) less crime deterrence but more detection than a regime of unilateral sanction upon the Agent.

Remark that in order to compare our different regimes, we have obtained sufficient conditions which may deserve some comments.

First, let us consider the sufficient condition of proposition 6 (which implies the various values of the fines): $(1-\theta) f_P \leq f \leq q f_A + (1-\theta) f_P$; remark that the LHS inequality is satisfied (but not uniquely) soon as: $f_P \leq f$; if this inequality holds, then the RHS inequality may be written as: $f - f_P \leq q f_A - \theta f_P$, and it must be finally that $q f_A \geq \theta f_P$ holds.

On the other hand, the sufficient conditions of proposition 4 and corollary 5 together require that: $f_P \leq f_A \leq f$. Once more, these conditions are only sufficient to guarantee the results; nevertheless they suggest that:

- as the enforcer focuses its monitoring efforts on one member of the criminal team (either the leader, or the active agent) rather than on both members, thus raising the corresponding specific probability of detection up to its potential maximal level, then it is recommended to also use larger fines (f) than the level which is individually applied in case of joint liability and sanction;
- at the same time, the total expected amount of fines raised by the enforcer in case of joint liability must be at least as large as the level applied in case of exclusive sanction;
- finally, it is suggested that in case of joint liability and sanction, the individual sanction/fine both in level and expected terms, which are applied to the active partner must be at least as large as those raised on the sleeping partner/leader of the team.

With respect to the state-dependent characteristics of Agent's preferences, which enable to encompass cases where enforcers use non monetary sanctions: our results suggest that they are consistent the usual view on the use of fine and imprisonment; as the probability of convincing the Agent increases so is the level of the fine put on him, although prison sentencing or any other non monetary sanction be used at the same time.

7. Conclusion

The paper has addressed the issue of the choice of a regime of sanction in order to control criminal teams activities. It has followed a basic positive view according to previous works such as Privileggi, Marchese and Cassone (2001). But in contrast, the results are qualified here (almost all the time) in a general and non ambiguous way, allowing us to evaluate how the nature of the activity of detection avoidance together with the existence of a cooperative/non cooperative behavior between partners in a criminal team, facilitates the action of enforcers. We also have introduced (at least implicitly) more policy instruments than in previous studies: in addition to the monitoring activity of enforcers and the use of monetary sanctions in case of arrest, we have assumed that authorities have the opportunity to apply non monetary sanctions (prison sentence, loss of civil rights, incapacitation and so on); this enables us to consider that (at least some) criminals have state-dependent preferences with different utility indexes depending on whether a criminal is or not arrested and punished.

The first main result of the paper is about the effects of liability shifting between the members of a criminal organization (we compare pure regimes of sanction in this set up) when they have specialized tasks in the team and a different sensibility to the risk of arrest and punishment. Specifically, we assume here that the perpetrator of the crime is a state-dependent risk averse individual, and that he bears a (linear) monetary cost due to the activity of detection avoidance, whereas the leader is risk neutral and bears only fixed costs (due to general management of the team and the monitoring of the members). In this set up, we have shown as in Privileggi, Marchese and Cassone (2001)'s paper that exclusive sanctions upon the Agent/perpetrator of the crime induces more deterrence of the Principal (leader of the team) than a regime of exclusive sanctions upon the Principal: it is more efficient to punish the "~weaker~" individual alone (more sensible to the risk of sanctions), in the sense that the frequency of crime is smaller. But the counterpart of this result [see also Sanchirico (2006)] is that it becomes more difficult for public authorities to detect illegal activities, in the sense that the frequency of detection and punishment becomes smaller in the former regime. We have shown that both results occur whatever the value of the fine, and the intensity of the public monitoring (likelihood of controls), which is in contrast with Privileggi and ali (2001)'s analysis for state-independent risk averse Agent having a disutility for efforts in the detection avoidance activity.

The second main result of the present paper is about the effects of a regime of "joint liability and sanction": we have shown that it implies more deterrence than the regime of exclusive sanction upon the leader, but with a cost since "~joint liability~" also reduces the effectiveness of public monitoring as compared to the second regime, once again. It means that it is more efficient to punish both partners in a criminal teams than the "stronger partner" (leader) alone, since the probability of crime is smaller when both bear a sanction - albeit, the frequency of detection of the team is smaller. In contrast, there is unfortunately no reason to believe that it is also more efficient to punish both partners than the "~weaker~" alone; maybe this is not the case: more conditions over basic assumptions regarding the technology of avoidance and individual preferences are required (not found here).

In this paper, we have followed a descriptive view on the issue of criminal teams. This point of view has to be motivated, given that there exists a large body of literature focusing rather on the optimal enforcement of the penal code [see Garoupa, (1997), Polinsky and Shavell, (2000)]. On the one hand, it is well known that introducing risk aversion leads to puzzling results regarding the design of the optimal law enforcement policy [see Polinsky and Shavell, (1979), Neilson, (1998)]: under risk aversion, maximal fines and small probability of control may be or not optimal, depending on whether criminals are more or less sensible to the frequency than to the severity of the sanction [see also Neilson and Winter, (1997)]. On the other hand, it must be reminded that there exist other goals of criminal law, notably incapacitation, rehabilitation, and retribution [see Shavell, (1987)]. Our paper may be understood as suggesting that there exist cases where these different goals (such as the incapacitation of all the gang members) may compete with the deterrence objective.

8. References

[1] Arlen, J., (1994), *The potentially perverse effects of corporate criminal liability*, in: *The Journal of Legal Studies,* 23, pp. 833 – 867.

[2] Chappe, N. and At C., (2005), *Crime timing*, in: Economics Bulletin, 11, pp. 1 – 7.

[3] Chappe, N. and At C., (2008), *Timing of crime, learning and sanction,* in: *Review of Law and Economics,* 4.1, Article 3.

[4] Chu, C. and Qian, Y., (1995), *Vicarious liability under a negligence rule*, in: *International Review of Law and Economics,* 15, pp. 305 – 322.

[5] Coffee, J., (2002), *Understanding Enron: It's About the Gatekeepers, Stupid*, Columbia Law & Economics Working Paper No. 207.

[6] Dehez, P. and Drèze, J., (1987), *State-dependent utility, the demand for insurance and the value of safety,* in Drèze: *Essays on economic decisions under uncertainty,* Cambridge University Press.

[7] Ganuza, J. and Gomez, F., (2007), *Should we trust gatekeepers?* Auditors' and lawyers' liability for clients' misconduct, mimeo, in: *International Review of Law and Economics*, 27, pp. 96 – 109.

[8] Garoupa, N., (1997), *The theory of optimal law enforcement*, in: *Journal of Economic Surveys*, 11, pp. 267 – 295.

[9] Garoupa, N., (2000), *The economics of organized crime and optimal law enforcement*, in: *Economic Inquiry,* 38, pp. 278 – 288.

[10] Garoupa, N., (2001), *Optimal magnitude and probability of fines*, in: *European Economic Review*, 45, pp. 1765 – 1771.

[11] Jacquemet, N., (2006), *Microéconomie de la corruption*, in: *Revue Française d'Economie*, 20, pp. 117 – 159.

[12] Jones-Lee, M., (1974), *The value of changes in the probability of death or injury*, in: *Journal of Political Economy,* 82, pp. 835 – 849.

[13] Kraakman, R., (1986), *Gatekeepers: the anatomy of third-party enforcement strategy*, in: *Journal of Law, Economics and Organization,* 2, pp. 53 – 104.

[14] Langlais, E., (2005), W*illingness to pay for risk reduction and risk aversion without the expected utility assumption*, in: *Theory and Decision*, 59, pp. 43 – 50.

[15] Marjit, S. and Shi, H., (1998), *On Controlling Crime with Corrupt Officials*, in: *Journal of Economic Behavior and Organization*, 34, pp. 163 – 172.

[16] Neilson, W. and Winter, H., (1997), *On criminals' risk attitude*, in: *Economics Letters*, 55, pp. 97 – 102.

[17] Neilson, W., (1998), *Optimal punishment schemes with state-dependent preferences*, in: *Economic Inquiry,* 36: 266-271.

[18] Partnoy, F., (2003), *Strict Liability for Gatekeepers: A Reply to Professor Coffee*, in: University San Diego Law and Economics Research Paper No. 3*,* University of San Diego - School of Law.

[19] Polinsky, M. and Shavell, S., (1979), *The optimal trade-off between the probability and magnitude of fines,* in: *American Economic Review*, 69, pp. 880 – 891.

[20] Polinsky, M. and Shavell, S., (2000), *The theory of public enforcement of law*, in: *Journal of Economic Literature*, 38, pp. 45 – 76.

[21] Privileggi, F., Marchese, C. and Cassone, A., (2001), *Agent's liability versus Principal liability when attitudes toward risk differ*, in: *International Review of Law and Economics*, 21, pp. 181 – 195.

[22] Sanchirico C. (2006), *Detection avoidance*, in: *New York University Law Review* 81, pp. 1331 – 1399.

[23] Shavell, S., (1985), *Criminal Law and the Optimal Use of Nonmonetary Sanctions as a Deterrent*, in: *Columbia Law Review*, 85, pp. 1232 – 1262.

[24] Shavell, S., (1997), *The optimal level of corporate liability given the limited ability of corporations to penalize their employees*, in: *International Review of Law and Economics*, 17, pp. 203 – 213.

[25] Schäfer, H-B, (2004), *Efficient Third Party Liability of Auditors in Tort Law and Contract Law*, in: *Supreme Court Economic Review*, 12, pp. 181 – 208.

APPENDIX 1

Checking explicitly second order conditions of maximization may be useful there after, let us totally differentiate the FOC (5) written as $-Eu' + p'[u_{0-}u_1] = 0$, the SOC require that:

$$Eu'' - Eu' + p''(u_0 - u_1) + 2(-p')(u_0' - u_1') \leq 0$$

which may also be written after some straightforward but tedious manipulations as $\Omega \times Eu' \geq 0$, where:

$$\Omega = \left(\frac{1}{-p'(x_A)}\right)\left(\frac{p''(x_A)}{-p'(x_A)}\right) + 2\left[\left(\frac{u_1' - u_0'}{Eu'}\right) + \frac{1}{2}\left(\frac{1}{-p'(x_A)}\right)\left(-\frac{Eu''}{Eu'}\right)\right] \geq 0$$

where $-p'(x_A) = \hat{p}'(x_A)$ and $p''(x_A) = -\hat{p}''(x_A)$. Remark that Ω is used when we turn to the analysis of comparative statics. Remark also that assumption 2.2 is a sufficient condition for SOC to hold.

APPENDIX 2

In our set up, state 0 corresponds to the case where the Agent is detected and caught by the public authorities, while state 1 is the state where he is not caught. Consider any gamble defined as: $(p(x), w_0; (1-p(x)), w_1)$. Thus, the certainty equivalent at wealth w may be defined as the amount of wealth which is accepted by the Agent to be not detected by the authority and to have the same level of satisfaction as the gamble itself:

$$\begin{aligned} & p(x)u_0(w_0) + (1-p(x))u_1(w_1) \\ &= u_1(w + ec) \\ &= u_1(p(x)w_0 + (1-p(x))w_1 - \pi_1(w; p(x)w_0 + (1-p(x))w_1)) \end{aligned}$$

with $\pi_1(.) > 0$ denoting the Arrow-Pratt absolute risk premium associated to u_1 . Hence, defining $w_0 = w - x - S$, and $w_1 = w - x$, we have: $ec = -x - p(x)S - \pi_1(w; -x - p(x)S)$

APPENDIX 3

Let us define the function: $\lambda(f_A, q) = q\frac{u_1(w-x) - u_0(w - f_A - x)}{p(x)qu_0'(w - f_A - x) + (1-p(x)q)u_1'(w-x)}$ which takes positive values on its domain $\Re \times [0,1]$.

- Differentiating the system (7)-(8) in q and rearranging yields:

$$D\frac{\partial x}{\partial q} = \Omega\frac{\partial w}{\partial q} + \frac{\lambda(f_A, q)}{q}\left(1 + p(x)q\frac{u_1' - u_0'}{p(x)qu_0' + (1-p(x)q)u_1'}\right)$$

$$(\lambda(f_A, q)(-p') - 1)\frac{\partial x}{\partial q} = -\frac{\partial w}{\partial q} + p(x)\frac{\lambda(f_A, q)}{q}$$

where:

$$D = \left(\frac{p''}{(p')^2} + \Omega + +\lambda(f_A, q)(-p')q\frac{u_1' - u_0'}{p(x)qu_0' + (1-p(x)q)u_1'}\right) > 0$$

and:

$$\Omega = q\frac{u_1' - u_0'}{p(x)qu_0' + (1-p(x)q)u_1'} + \lambda(f_A, q)\left(-\frac{Eu''}{Eu'}\right) > 0,$$

under assumption 2. Solving for $\frac{\partial x}{\partial q}$ leads to:

$$\frac{\partial x}{\partial q} = \frac{\lambda(f_A, q)}{q} \times \frac{1 + p(x)\left(\Omega + q\frac{u_1' - u_0'}{p(x)qu_0' + (1-p(x)q)u_1'}\right)}{D + \Omega\left(\lambda(f_A, q)(-p') - 1\right)}$$

The denominator is positive according to the SOC; the numerator is also positive: this implies that $\frac{\partial x}{\partial q} > 0$. As a result, we also have the impact on the Principal's expected return is: $\frac{\partial v}{\partial q} = -\frac{\lambda(f_A, q)}{q} p(x) < 0.$

▪ Differentiating the system (7)-(8) in θ and rearranging yields:

$$D\frac{\partial x}{\partial \theta} = \Omega\frac{\partial w}{\partial \theta} - f_P$$

$$\left(\lambda(f_A, q)(-p') - 1\right)\frac{\partial x}{\partial \theta} = -\frac{\partial w}{\partial \theta}$$

Solving for $\frac{\partial x}{\partial \theta}$ leads to:

$$\frac{\partial x}{\partial \theta} = \frac{-f_P}{D + \Omega\left(\lambda(f_A, q)(-p') - 1\right)}$$

hence implying that $\frac{\partial x}{\partial \theta} < 0$. As a result, we also have: $\frac{\partial v}{\partial \theta} = p(x) f_P > 0.$

▪ Finally, remark that if $f_A \geq f_P$ then $\frac{\lambda(f_A, q)}{q} \geq f_A \geq f_P \Rightarrow \frac{\partial x}{\partial q} > -\frac{\partial x}{\partial \theta}$: when both θ and q increase, then x increases; moreover, if $f_A \geq f_P$ then $\frac{\lambda(f_A, q)}{q} \geq f_A \geq f_P \Rightarrow -\frac{\partial v}{\partial q} > \frac{\partial v}{\partial \theta}$: when both θ and q increase, then v decreases.

Journal of Applied Economic Sciences
Volume IV/ Issue 1(7)/ Spring 2009

REPRESENTATION-CONSTRAINED CANONICAL CORRELATION ANALYSIS: A HYBRIDIZATION OF CANONICAL CORRELATION AND PRINCIPAL COMPONENT ANALYSES

Sudhanshu K **MISHRA**
North-Eastern Hill University, Shillong, Meghalaya, **India**
Department of Economics
mishrasknehu@yahoo.com

Abstract:

The classical canonical correlation analysis is extremely greedy to maximize the squared correlation between two sets of variables. As a result, if one of the variables in the dataset-1 is very highly correlated with another variable in the dataset-2, the canonical correlation will be very high irrespective of the correlation among the rest of the variables in the two datasets. We intend here to propose an alternative measure of association between two sets of variables that will not permit the greed of a select few variables in the datasets to prevail upon the fellow variables so much as to deprive the latter of contributing to their representative variables or canonical variates.

Our proposed Representation-Constrained Canonical correlation (RCCCA) Analysis has the Classical Canonical Correlation Analysis (CCCA) at its one end (λ=0) and the Classical Principal Component Analysis (CPCA) at the other (as λ tends to be very large). In between it gives us a compromise solution. By a proper choice of λ, one can avoid hijacking of the representation issue of two datasets by a lone couple of highly correlated variables across those datasets. This advantage of the RCCCA over the CCCA deserves a serious attention by the researchers using statistical tools for data analysis.

Keywords: Representation, constrained, canonical, correlation, principal components, variates, global optimization, particle swarm, ordinal variables, computer program, FORTRAN

JEL Classification: C13, C43, C45, C61, C63, C87

1. Introduction

We begin this paper with reference to a dataset that, when subjected to the classical canonical correlation analysis, gives us the leading (first or largest) canonical correlation which is misleading. It is misleading in the sense that, in this example, the canonical correlation (which is the coefficient of correlation between the two canonical variates, each being a linear weighted combination of the variables in the associated dataset) is, indeed, not a measure of the true association of the variables in the two datasets, but, instead, the datasets have been hijacked by a lone couple of variables across the two datasets.

Table 1.1. Simulated Dataset-1 for Canonical correlation

Sl	X_1 or Dataset-1				X_2 or Dataset-2				
No.	X_{11}	X_{12}	X_{13}	X_{14}	X_{21}	X_{22}	X_{23}	X_{24}	X_{25}
1	0.7	2.6	0.1	1.7	0.2	0.8	1.6	0.5	1.6
2	1.5	1.7	1.2	1.5	1.6	2.4	2.3	1.4	3.2
3	2.3	0.3	2.7	1.2	2.5	2.9	0.6	1.3	4.8
4	0.6	2.0	0.9	2.8	2.8	2.5	1.1	1.8	1.4
5	0.1	0.9	1.6	1.8	2.2	2.7	2.1	0.2	0.4
6	1.9	1.1	1.7	2.6	1.5	2.2	2.2	2.0	4.0
7	1.0	2.7	2.4	2.7	1.0	0.2	2.0	0.4	2.2
8	1.8	2.9	1.4	0.9	1.7	1.0	1.8	1.2	3.8
9	2.8	0.1	1.8	0.4	2.3	0.6	1.7	0.6	5.8
10	1.4	0.6	2.8	1.4	2.6	1.8	0.8	1.7	3.0
11	1.2	2.5	2.9	0.8	2.1	0.7	1.4	2.3	2.6
12	1.1	1.3	0.2	2.5	0.7	1.5	1.0	2.2	2.4

Sl No.	X_1 or Dataset-1				X_2 or Dataset-2				
	X_{11}	X_{12}	X_{13}	X_{14}	X_{21}	X_{22}	X_{23}	X_{24}	X_{25}
13	3.0	1.9	1.1	1.6	0.1	0.1	2.7	3.0	6.2
14	2.0	0.8	0.6	1.3	1.9	0.5	0.4	0.8	4.2
15	1.6	2.2	2.6	1.9	1.4	1.3	1.3	2.5	3.4
16	2.9	0.7	1.9	2.9	2.4	1.2	2.5	2.1	6.0
17	1.3	1.4	2.0	0.2	1.8	2.8	0.3	2.6	2.8
18	0.8	0.2	2.3	2.0	2.9	1.4	3.0	0.7	1.8
19	1.7	0.5	1.3	0.1	2.0	0.9	2.9	1.5	3.6
20	2.1	2.4	0.7	0.5	0.9	2.3	0.7	0.3	4.4
21	2.5	1.0	3.0	2.2	1.2	2.6	2.6	1.0	5.2
22	2.2	2.8	2.5	0.7	3.0	3.0	0.2	1.9	4.6
23	0.5	0.4	0.8	1.0	0.8	0.4	0.1	1.1	1.2
24	2.7	2.1	1.5	2.3	1.1	1.1	0.9	2.7	5.6
25	2.4	1.8	0.5	0.3	2.7	1.6	2.8	0.1	5.0
26	0.2	1.6	0.3	1.1	0.6	0.3	2.4	2.8	0.6
27	0.9	2.3	0.4	0.6	1.3	1.7	1.5	2.4	2.0
28	2.6	3.0	2.2	3.0	0.5	1.9	1.9	1.6	5.4
29	0.4	1.2	1.0	2.4	0.4	2.0	0.5	2.9	1.0
30	0.3	1.5	2.1	2.1	0.3	2.1	1.2	0.9	0.8

In Table 1.1 the dataset X is presented which is a pooled set of two datasets, X_1 and X_2, such that $X=[X_1|X_2]$. The first dataset has m_1 (=4) variables and the second dataset has m_2 (=5) variables, each in n (=30) observations. These seemingly normal datasets, when subjected to the classical canonical correlation analysis, yield canonical correlation between the composite variables, z_1 and z_2 (the canonical variates), $r(z_1,z_2)=1.0$: $z_1=\sum_{j=1}^{4} w_j x_{1j}; x_{ij}\in X_1$; $z_2=\sum_{j=1}^{5} w_j x_{2j}; x_{2j}\in X_2$. The weight vectors are: w_1=(1, 0, 0, 0, 0) and w_2=(0, 0, 0, 0, 1). This anomalous situation has arisen due to the fact that x_{25} is perfectly linearly dependent on x_{11} and the canonical correlation, $r(z_1,z_2)$, is in fact $r(x_{11},x_{25})$. Other variables have no contribution to z_1 or z_2. It follows, therefore, that z_1 and z_2 do not represent other variables in X_1 and X_2. Nor is the canonical correlation, $r(z_1,z_2)$, a correlation between the two sets, X_1 and X_2, in any relevant or significant sense. Thus, the leading canonical correlation may deceive us if we are only a little less careful to look into the correlation matrix encompassing all variables.

Such examples may be multiplied *ad infinitum*. If one is cautious, the anomalous cases can be detected. However, such cases, if not detected, make scientific analysis and interpretation of empirical results rather hazardous. One may easily be misled to a conclusion that such two datasets are highly correlated while the truth may be quite far from it.

2. Objectives of the Present Work

We intend here to propose an alternative measure of association between two sets of variables that will not permit the greed of a select few variables in the datasets to prevail upon the fellow variables so much as to deprive the latter of contributing their say and share to the representative variables (ς_1 and ς_2), which they make by their participation in the linear combination. We may not call $\varsigma_1=\sum_{j=1}^{m_1}\omega_{1j}x_{1j}$ and $\varsigma_2=\sum_{j=1}^{m_2}\omega_{2j}x_{2j}$ the canonical variables (defined before as $z_1=\sum_{j=1}^{4} w_j x_{1j}$; $z_2=\sum_{j=1}^{5} w_j x_{2j}$ obtained from the classical canonical correlation analysis).

In the classical canonical correlation analysis the objective is to maximize $r^2(z_1,z_2): z_1=\sum_{j=1}^{m_1} w_{1j}x_{1j}; z_2=\sum_{j=1}^{m_2} w_{2j}x_{2j}$ irrespective of $r(z_1,x_{1j}): x_{1j}\in X_1$ and $r(z_2,x_{2j}): x_{2j}\in X_2$, and, therefore, $r^2(z_1,z_2)$ is subject to an unconstrained maximization. However, in the method that we are proposing here, the objective will be to maximize $r^2(\varsigma_1,\varsigma_2)$: $\varsigma_1=\sum_{j=1}^{m_1}\omega_{1j}x_{1j}$ and $\varsigma_2=\sum_{j=1}^{m_2}\omega_{2j}x_{2j}$ with certain constraints in terms of $r(\varsigma_1,x_{1j}): x_{1j}\in X_1$ and $r(\varsigma_2,x_{2j}): x_{2j}\in X_2$. These constraints would ensure the representativeness of ς_1 to X_1 and that of ς_2 to X_2. Hence, the proposed method may be called the *Representation-Constrained Canonical Correlation Analysis*.

3. The Nature and Implications of the Proposed Constraints

There are a number of ways in which the canonical variates can be constrained insofar as their association and concordance with their fellow variables in their respective native datasets are concerned. In other words, their representativeness to their native datasets can be defined variously. We discuss here some of the alternatives in terms of correlation as a measure of representativeness.

(i) Mean absolute correlation principle: A (constrained) canonical variate $\varsigma_a = \sum_{j=1}^{m_a} \omega_{aj} x_{aj}; x_{aj} \in X_a$ is a better representative of X_a if the mean absolute correlation, $\sum_{j=1}^{m_a} | r(\varsigma_a, x_{aj}) |$, is larger. This approach is equalitarian in effect.

(ii) Mean squared correlation principle: A (constrained) canonical variate $\varsigma_a = \sum_{j=1}^{m_a} \omega_{aj} x_{aj}; x_{aj} \in X_a$ is a better representative of X_a if the mean squared correlation, $\sum_{j=1}^{m_a} r^2(\varsigma_a, x_{aj})$, is larger. This approach is elitist in effect, favouring dominant members.

(iii) Minimal absolute correlation principle: A (constrained) canonical variate $\varsigma_a = \sum_{j=1}^{m_a} \omega_{aj} x_{aj}; x_{aj} \in X_a$ is a better representative of X_a if the minimal absolute correlation, $\min_j[| r(\varsigma_a, x_{aj}) |]$, is larger. A larger $\min_j[| r(\varsigma_a, x_{aj}) |]$ implies that the minimal squared correlation, $\min_j[r^2(\varsigma_a, x_{aj})]$, is larger. This approach is in favour of the weak.

These three approaches lead to three alternative objective functions:

(i). Maximize $r^2(\varsigma_1,\varsigma_2) + \lambda[\sum_{j=1}^{m_1} | r(\varsigma_1, x_{1j}) | / m_1 + \sum_{j=1}^{m_2} | r(\varsigma_2, x_{2j}) | / m_2] : \varsigma_1 = \sum_{j=1}^{m_1} \omega_{1j} x_{1j};\ \varsigma_2 = \sum_{j=1}^{m_2} \omega_{2j} x_{2j}.$

(ii). Maximize $r^2(\varsigma_1,\varsigma_2) + \lambda[\sum_{j=1}^{m_1} r^2(\varsigma_1, x_{1j}) / m_1 + \sum_{j=1}^{m_2} r^2(\varsigma_2, x_{2j}) / m_2] : \varsigma_1 = \sum_{j=1}^{m_1} \omega_{1j} x_{1j};\ \varsigma_2 = \sum_{j=1}^{m_2} \omega_{2j} x_{2j}.$

(iii). Maximize $r^2(\varsigma_1,\varsigma_2) + \lambda\left[\min_j[| r(\varsigma_1, x_{1j}) |] + \min_j[| r(\varsigma_2, x_{2j}) |]\right] : \varsigma_1 = \sum_{j=1}^{m_1} \omega_{1j} x_{1j};\ \varsigma_2 = \sum_{j=1}^{m_2} \omega_{2j} x_{2j}.$

In these objective functions, the value of λ may be chosen subjectively. If $\lambda = 0$, the objective function would degenerate to the classical canonical correlation analysis, but λ has no upper bound. Also note that if the first term is $| r(\varsigma_1,\varsigma_2) |$ rather than $r^2(\varsigma_1,\varsigma_2)$ and $\lambda \neq 0$, its implied weight vis-à-vis the second term increases since $| r(\varsigma_1,\varsigma_2) | > r^2(\varsigma_1,\varsigma_2)$ for $| r | < 1$.

4. The Method of Optimization

The classical canonical correlation analysis [Hotelling, (1936)] sets up the objective function to maximize $r^2(\varsigma_1,\varsigma_2) : \varsigma_1 = \sum_{j=1}^{m_1} \omega_{1j} x_{1j};\ \varsigma_2 = \sum_{j=1}^{m_2} \omega_{2j} x_{2j}$ and using the calculus methods of maximization resolves the problem to finding out the largest eigenvalue and the associated eigenvector of the matrix, $[X_1'X_1]^{-1} X_1'X_2 [X_2'X_2]^{-1} X_2'X_1$. The largest eighen-value turns out to be the leading $r^2(z_1, z_2) : z_1 = \sum_{j=1}^{m_1} w_{1j} x_{1j};\ z_2 = \sum_{j=1}^{m_2} w_{2j} x_{2j}$, and the standardized eigenvector is used to obtain w_1 and w_2. However, a general calculus-based method cannot be applied to maximize the (arbitrary) objective function set up for the constrained canonical correlation analysis. At any rate, the first and the third objective functions are not amenable to maximization by the calculus-based methods.

We choose, therefore, to use a relatively new and more versatile method of (global) optimization, namely, the Particle Swarm Optimization (PSO) proposed by Eberhart and Kennedy (1995). A lucid description of its foundations is available in Fleischer (2005). The PSO is a biologically inspired population-based stochastic search method modeled on the ornithological observations, simulating the behavior of members of the flocks of birds in searching food and communicating among themselves. It is in conformity with the principles of decentralized decision making [Hayek, (1948); (1952)] leading to self-organization and macroscopic order. The effectiveness of PSO has been very encouraging in solving extremely difficult and varied types of nonlinear optimization problems [Mishra, (2006)]. We have used a particular variant of the PSO called the Repulsive Particle Swarm Optimization [Urfalioglu, (2004)].

5. Findings and Discussion

We have subjected the data in Table 1.1 to the representation-constrained canonical correlation analysis with the three alternative objective functions elaborated in section-III. The first term, measuring the degree of association between the two datasets, X_1 and X_2, is in the squared form, that is $r^2(\varsigma_1,\varsigma_2)$, although we have reported its positive square root ($=|r(\varsigma_1,\varsigma_2)|$) in Table 1.2. The three objective functions have been optimized for the different values of λ, varying from zero to 50 with an increment of 0.5. For the first objective function, the values of $|r(\varsigma_1,\varsigma_2)|$, mean absolute $r(\varsigma_1,x_1)$ and mean absolute $r(\varsigma_2,x_2)$ at different values of λ have been plotted in Figure 1.1. Similarly, for the second objective function, the values of $|r(\varsigma_1,\varsigma_2)|$, mean squared $r(\varsigma_1,x_1)$ and mean squared $r(\varsigma_2,x_2)$ at different values of λ have been plotted in Figure1.2., Figure 1.3 presents $|r(\varsigma_1,\varsigma_2)|$, minimum absolute $r(\varsigma_1,x_1)$ and minimum absolute $r(\varsigma_2,x_2)$ relating to the 3rd objective maximized at different values of λ.

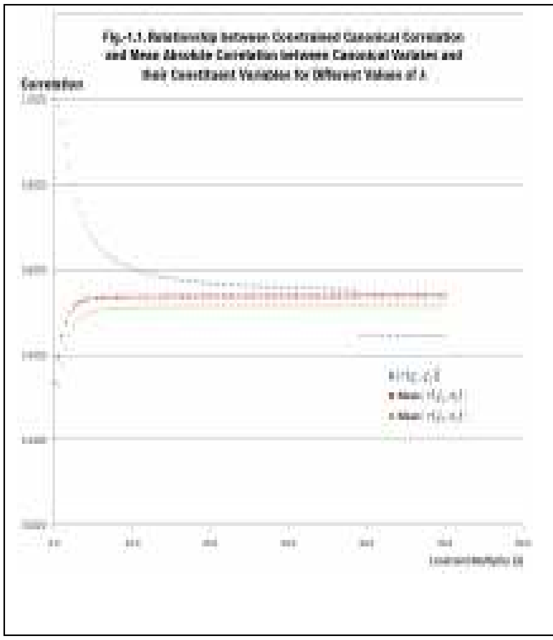
Table 1.2. Relationship between Constrained Canonical Correlation and Representation Correlation between Canonical Variates and their Constituent Variables for Different Values of λ

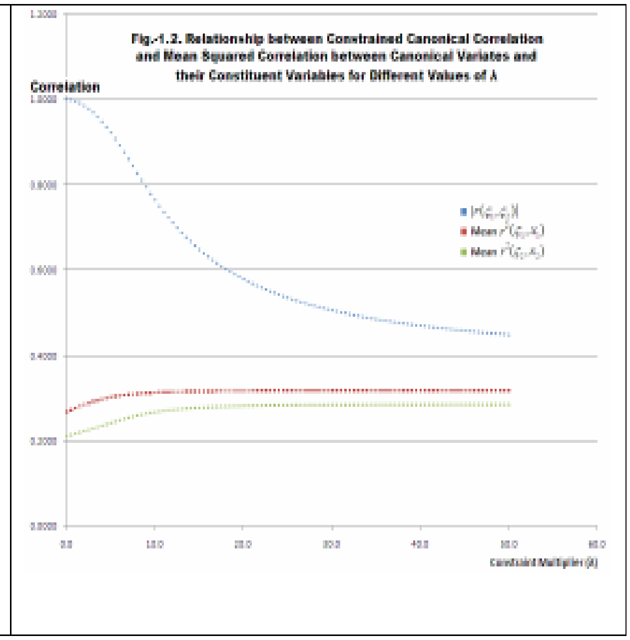
Sl No	λ	Canonical $\lvert r(\varsigma_1,\varsigma_2)\rvert$	Mean Absolute $r(\varsigma_1,x_1)$	Mean Absolute $r(\varsigma_2,x_2)$	Canonical $\lvert r(\varsigma_1,\varsigma_2)\rvert$	Mean Squared $r(\varsigma_1,x_1)$	Mean Squared $r(\varsigma_2,x_2)$	Canonical $\lvert r(\varsigma_1,\varsigma_2)\rvert$	Minimum Absolute $r(\varsigma_1,x_1)$	Minimum Absolute $r(\varsigma_2,x_2)$
1	0.0	1.0000	0.3342	0.2814	1.0000	0.2668	0.2121	1.0000	0.0234	0.0246
2	0.5	0.9831	0.3942	0.3254	0.9990	0.2717	0.2152	0.9755	0.1615	0.1361
3	1.0	0.9440	0.4434	0.3785	0.9961	0.2763	0.2183	0.8223	0.3921	0.2671
4	1.5	0.8942	0.4772	0.4188	0.9916	0.2805	0.2214	0.5072	0.5302	0.4618
5	2.0	0.8432	0.4992	0.4479	0.9855	0.2843	0.2244	0.4662	0.5319	0.4853
6	2.5	0.7975	0.5128	0.4679	0.9780	0.2878	0.2275	0.4556	0.5337	0.4889
7	3.0	0.7597	0.5210	0.4813	0.9691	0.2909	0.2306	0.4473	0.5338	0.4917
8	3.5	0.7298	0.5259	0.4902	0.9590	0.2938	0.2338	0.4296	0.5337	0.4968
9	4.0	0.7060	0.5290	0.4962	0.9477	0.2964	0.2369	0.4349	0.5334	0.4954
10	4.5	0.6870	0.5310	0.5005	0.9352	0.2987	0.2401	0.4230	0.5335	0.4978
11	5.0	0.6715	0.5323	0.5036	0.9217	0.3008	0.2433	0.4342	0.5337	0.4955
12	5.5	0.6590	0.5333	0.5058	0.9073	0.3027	0.2464	0.4359	0.5338	0.4950
13	6.0	0.6483	0.5339	0.5076	0.8921	0.3044	0.2495	0.4404	0.5338	0.4940
14	6.5	0.6394	0.5345	0.5089	0.8762	0.3059	0.2525	0.3743	0.5389	0.4963
15	7.0	0.6318	0.5348	0.5100	0.8599	0.3072	0.2554	0.4170	0.5337	0.4994
16	7.5	0.6251	0.5351	0.5108	0.8434	0.3083	0.2581	0.4175	0.5338	0.4992
17	8.0	0.6193	0.5354	0.5115	0.8270	0.3094	0.2607	0.4278	0.5338	0.4970
18	8.5	0.6142	0.5356	0.5121	0.8106	0.3102	0.2630	0.4167	0.5335	0.4990
19	9.0	0.6098	0.5357	0.5126	0.7945	0.3110	0.2652	0.4293	0.5337	0.4967
20	9.5	0.6056	0.5358	0.5130	0.7789	0.3117	0.2672	0.4206	0.5339	0.4986
21	10.0	0.6019	0.5360	0.5133	0.7641	0.3122	0.2690	0.3746	0.5389	0.4962
22	10.5	0.5988	0.5360	0.5136	0.7495	0.3127	0.2706	0.2904	0.5023	0.4748
23	11.0	0.5958	0.5361	0.5139	0.7359	0.3132	0.2721	0.4167	0.5338	0.4990
24	11.5	0.5931	0.5362	0.5141	0.7227	0.3136	0.2734	0.4201	0.4789	0.4281
25	12.0	0.5906	0.5362	0.5143	0.7103	0.3139	0.2746	0.4206	0.5338	0.4987
26	12.5	0.5884	0.5363	0.5144	0.6985	0.3142	0.2756	0.5150	0.4781	0.3664
27	13.0	0.5861	0.5363	0.5146	0.6872	0.3145	0.2766	0.4167	0.5337	0.4993
28	13.5	0.5842	0.5364	0.5147	0.6764	0.3148	0.2774	0.3745	0.5389	0.4964
29	14.0	0.5826	0.5364	0.5148	0.6665	0.3150	0.2782	0.3742	0.5390	0.4963
30	14.5	0.5807	0.5364	0.5150	0.6570	0.3152	0.2789	0.4022	0.4648	0.4532
31	15.0	0.5791	0.5365	0.5151	0.6478	0.3154	0.2795	0.4170	0.5338	0.4991
32	15.5	0.5778	0.5365	0.5151	0.6390	0.3155	0.2801	0.4179	0.5003	0.4860
33	16.0	0.5765	0.5365	0.5152	0.6310	0.3157	0.2806	0.2791	0.5387	0.4990
34	16.5	0.5751	0.5365	0.5153	0.6231	0.3158	0.2810	0.3992	0.4764	0.4347
35	17.0	0.5739	0.5365	0.5154	0.6158	0.3159	0.2815	0.3742	0.5388	0.4964
36	17.5	0.5728	0.5366	0.5154	0.6088	0.3160	0.2819	0.0285	0.4457	0.4501
37	18.0	0.5715	0.5366	0.5155	0.6021	0.3161	0.2822	0.2794	0.5389	0.4992
38	18.5	0.5706	0.5366	0.5155	0.5960	0.3162	0.2825	0.3811	0.4744	0.4599
39	19.0	0.5697	0.5366	0.5156	0.5898	0.3163	0.2828	0.3741	0.5389	0.4963
40	19.5	0.5688	0.5366	0.5156	0.5840	0.3164	0.2831	0.3743	0.5389	0.4962
41	20.0	0.5680	0.5366	0.5157	0.5783	0.3165	0.2834	0.3345	0.4838	0.3983
42	20.5	0.5671	0.5366	0.5157	0.5732	0.3166	0.2836	0.2795	0.5389	0.4994
43	21.0	0.5663	0.5366	0.5157	0.5682	0.3166	0.2838	0.4194	0.4718	0.4439
44	21.5	0.5655	0.5366	0.5158	0.5632	0.3167	0.2840	0.3746	0.5389	0.4963

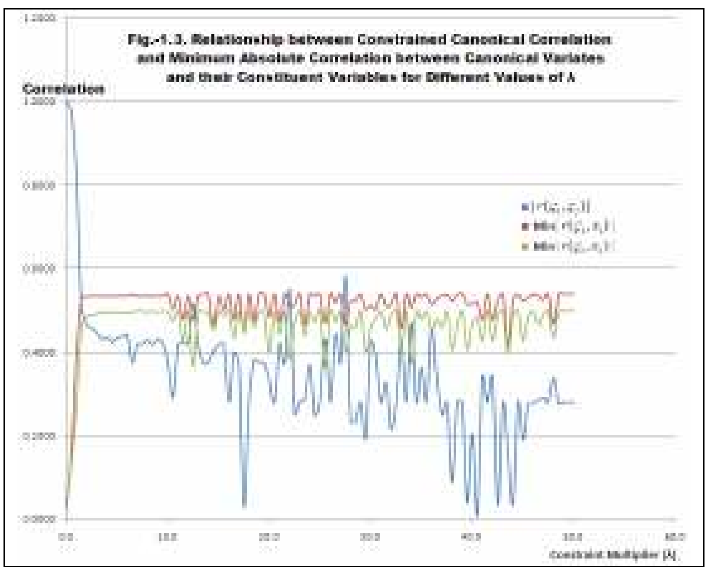
45	22.0	0.5650	0.5367	0.5158	0.5587	0.3167	0.2842	0.5496	0.5103	0.3823
46	22.5	0.5643	0.5367	0.5158	0.5542	0.3168	0.2843	0.2539	0.5138	0.4743
47	23.0	0.5635	0.5367	0.5158	0.5499	0.3168	0.2845	0.2795	0.5390	0.4993
48	23.5	0.5630	0.5367	0.5159	0.5459	0.3169	0.2846	0.2865	0.4643	0.4394
49	24.0	0.5623	0.5367	0.5159	0.5419	0.3169	0.2848	0.3688	0.5389	0.4944
50	24.5	0.5618	0.5367	0.5159	0.5383	0.3170	0.2849	0.2490	0.5347	0.4720
51	25.0	0.5612	0.5367	0.5159	0.5347	0.3170	0.2850	0.2792	0.5387	0.4994
52	25.5	0.5607	0.5367	0.5159	0.5312	0.3170	0.2851	0.4305	0.4684	0.3653
53	26.0	0.5603	0.5367	0.5160	0.5280	0.3171	0.2852	0.2793	0.5387	0.4993
54	26.5	0.5597	0.5367	0.5160	0.5249	0.3171	0.2853	0.4418	0.5176	0.4731
55	27.0	0.5592	0.5367	0.5160	0.5219	0.3171	0.2854	0.3741	0.5388	0.4963
56	27.5	0.5589	0.5367	0.5160	0.5186	0.3171	0.2855	0.5795	0.4661	0.4031
57	28.0	0.5584	0.5367	0.5160	0.5160	0.3172	0.2856	0.2335	0.5213	0.4604
58	28.5	0.5581	0.5367	0.5160	0.5131	0.3172	0.2857	0.2335	0.5213	0.4604
59	29.0	0.5575	0.5367	0.5161	0.5103	0.3172	0.2858	0.2790	0.5388	0.4993
60	29.5	0.5572	0.5367	0.5161	0.5080	0.3172	0.2858	0.1922	0.5023	0.4015
61	30.0	0.5568	0.5367	0.5161	0.5054	0.3173	0.2859	0.4223	0.5119	0.4564
62	30.5	0.5564	0.5367	0.5161	0.5030	0.3173	0.2859	0.3929	0.5016	0.4801
63	31.0	0.5561	0.5367	0.5161	0.5008	0.3173	0.2860	0.2795	0.5390	0.4993
64	31.5	0.5558	0.5367	0.5161	0.4987	0.3173	0.2861	0.3260	0.5081	0.4567
65	32.0	0.5555	0.5367	0.5161	0.4964	0.3173	0.2861	0.2140	0.5156	0.4897
66	32.5	0.5549	0.5367	0.5161	0.4942	0.3173	0.2862	0.2793	0.5389	0.4992
67	33.0	0.5547	0.5367	0.5161	0.4921	0.3174	0.2862	0.4277	0.4566	0.4137
68	33.5	0.5545	0.5367	0.5161	0.4902	0.3174	0.2863	0.2794	0.5389	0.4993
69	34.0	0.5542	0.5367	0.5161	0.4883	0.3174	0.2863	0.4708	0.5056	0.3723
70	34.5	0.5539	0.5367	0.5162	0.4865	0.3174	0.2863	0.2787	0.5388	0.4988
71	35.0	0.5539	0.5367	0.5162	0.4846	0.3174	0.2864	0.3639	0.5312	0.4787
72	35.5	0.5534	0.5367	0.5162	0.4830	0.3174	0.2864	0.2793	0.5389	0.4992
73	36.0	0.5532	0.5367	0.5162	0.4814	0.3174	0.2864	0.4560	0.5133	0.4533
74	36.5	0.5528	0.5367	0.5162	0.4796	0.3174	0.2865	0.3375	0.5282	0.4788
75	37.0	0.5524	0.5368	0.5162	0.4780	0.3174	0.2865	0.2504	0.5345	0.4600
76	37.5	0.5524	0.5368	0.5162	0.4765	0.3175	0.2865	0.2784	0.5380	0.4988
77	38.0	0.5521	0.5368	0.5162	0.4749	0.3175	0.2866	0.0886	0.5222	0.4078
78	38.5	0.5520	0.5368	0.5162	0.4733	0.3175	0.2866	0.2791	0.5372	0.4631
79	39.0	0.4469	0.5394	0.5163	0.4721	0.3175	0.2866	0.2795	0.5389	0.4992
80	39.5	0.4468	0.5394	0.5163	0.4707	0.3175	0.2866	0.0385	0.5148	0.4071
81	40.0	0.4467	0.5394	0.5163	0.4693	0.3175	0.2867	0.2028	0.5160	0.4721
82	40.5	0.4463	0.5394	0.5163	0.4681	0.3175	0.2867	0.0080	0.5182	0.4812
83	41.0	0.4463	0.5394	0.5163	0.4666	0.3175	0.2867	0.3389	0.4771	0.4282
84	41.5	0.4461	0.5394	0.5163	0.4653	0.3175	0.2867	0.2795	0.5389	0.4994
85	42.0	0.4460	0.5394	0.5163	0.4644	0.3175	0.2868	0.3389	0.4771	0.4282
86	42.5	0.4458	0.5394	0.5163	0.4631	0.3175	0.2868	0.0338	0.5248	0.4897
87	43.0	0.4456	0.5394	0.5163	0.4617	0.3175	0.2868	0.2793	0.5389	0.4993
88	43.5	0.4454	0.5394	0.5163	0.4606	0.3175	0.2868	0.1597	0.4139	0.3977
89	44.0	0.4453	0.5394	0.5163	0.4593	0.3176	0.2868	0.0338	0.5248	0.4897
90	44.5	0.4452	0.5394	0.5163	0.4586	0.3176	0.2869	0.2794	0.5389	0.4994
91	45.0	0.4451	0.5394	0.5163	0.4576	0.3176	0.2869	0.1880	0.5229	0.4274
92	45.5	0.4450	0.5394	0.5163	0.4564	0.3176	0.2869	0.2733	0.5300	0.4848
93	46.0	0.4448	0.5394	0.5163	0.4555	0.3176	0.2869	0.2786	0.5389	0.4991
94	46.5	0.4447	0.5394	0.5163	0.4547	0.3176	0.2869	0.2822	0.5354	0.4665
95	47.0	0.4445	0.5394	0.5163	0.4535	0.3176	0.2869	0.2898	0.5252	0.4905
96	47.5	0.4444	0.5394	0.5163	0.4527	0.3176	0.2869	0.2796	0.5389	0.4993
97	48.0	0.4444	0.5394	0.5163	0.4510	0.3176	0.2870	0.3372	0.4676	0.4344
98	48.5	0.4442	0.5394	0.5163	0.4509	0.3176	0.2870	0.2768	0.5389	0.4985
99	49.0	0.4440	0.5394	0.5163	0.4500	0.3176	0.2870	0.2792	0.5388	0.4993
100	49.5	0.4439	0.5394	0.5163	0.4491	0.3176	0.2870	0.2790	0.5389	0.4993
101	50.0	0.4438	0.5394	0.5163	0.4480	0.3176	0.2870	0.2784	0.5390	0.4989

From Figure 1.1 and Figure 1.2 it is clear that for increasing values of $\lambda,$ the value of $|r(\varsigma_1,\varsigma_2)|$ decreases monotonically, while the values of mean absolute (or squared) $r(\varsigma_1,x_1)$ and mean absolute (or squared) $r(\varsigma_2,x_2)$ increase monotonically. All of them exhibit asymptotic tendencies. However, for the third objective function the monotonicity of all the correlation functions is lost (shown in Figure 1.3). Of course, the trends in minimum absolute $r(\varsigma_1,x_1)$ and minimum absolute $r(\varsigma_2,x_2)$ are clearly observable. These observations may be useful to the choice of λ. For the case that we are presently dealing with, the value of λ need not exceed 10 to assure a fairly satisfactory representation of the two datasets by the corresponding canonical variates.

In particular, optimization of the second objective function has shown that the values of mean squared $r(\varsigma_1, x_1)$ and mean squared $r(\varsigma_2, x_2)$ exhibit asymptotic tendencies. For λ=50, the mean squared $r(\varsigma_1, x_1)$ is 0.3176 while the mean squared $r(\varsigma_2, x_2)$ is 0.2870.







Now, let us digress for a while to compute the first principal components of X_1 and X_2 (from the data given in Table 1.1). We find that for X_1 the sum of squared correlation (component loadings) of the component score (ξ_1) with its constituent variables is 0.317757. In other words, the first eigenvalue of the inter-correlation matrix R_1 obtained from X_1 is 1.271029, which divided by 4 (order of R_1) gives 0.317757. This is, in a way, a measure of representation of X_1 by its first principal component. Similarly, for X_2 the sum of squared correlation of the component score (ξ_2) with its constituent variables is 0.287521.

We resume our discussion for comparing these results (obtained from the Principal Component Analysis) with the results of our proposed representation-constrained canonical correlation analysis.

We observe that the asymptotic tendencies of mean squared $r(\varsigma_1, x_1)$ and mean squared $r(\varsigma_2, x_2)$ clearly point to the explanatory powers of the first principal components of X_1 and X_2 respectively.

However, if we compute the coefficient of correlation between the two component scores ($r(\xi_1, \xi_2)$=0.390767) and compare it with the constrained canonical correlation ($r(\varsigma_1, \varsigma_2)$=0.4480 for λ=50) we find that the latter is larger. Then, is the constrained canonical correlation analysis a hybrid of the classical canonical correlation and principal component analyses which has better properties of representation of data than its parents?

We conduct another experiment with the dataset presented in Table 2.1. We find that ξ_1 for X_1 has the representation power 0.333261 (eigenvalue=1.333042) while ξ_2 for X_2 has the representation power 0.382825 (eigenvalue=1.914123). The $r(\xi_1, \xi_2)$ is 0.466513. On the other hand, results of the constrained canonical correlation (for λ=49) are: mean squared $r(\varsigma_1, x_1)$= 0.33317; mean squared $r(\varsigma_2, x_2)$=0.38270 and the representation-constrained canonical correlation, $r(\varsigma_1, \varsigma_2)$= 0.48761. These findings are corroborative to our earlier results with regard to the dataset in Table 1.1.

Table 2.1. Simulated Dataset-2 for Canonical correlation

Sl	X_1 or Dataset-1				X_2 or Dataset-2					Sl	X_1 or Dataset-1				X_2 or Dataset-2				
No.	X_{11}	X_{12}	X_{13}	X_{14}	X_{21}	X_{22}	X_{23}	X_{24}	X_{25}	No.	X_{11}	X_{12}	X_{13}	X_{14}	X_{21}	X_{22}	X_{23}	X_{24}	X_{25}
1	2.7	1.9	2.4	1.2	2.6	2.3	1.5	0.1	6.6	16	1.4	1.4	0.2	0.4	1.1	2.2	2.2	2.6	2.4
2	1.7	0.1	0.8	1.8	0.4	0.2	2.3	0.2	0.1	17	1.5	0.4	2.2	1.9	1.9	0.6	2.1	1.9	5.5
3	0.2	2.8	2.6	0.9	1.3	2.0	2.0	0.3	7.3	18	0.6	1.3	2.5	2.8	2.8	2.5	2.7	2.2	4.9
4	0.4	0.3	1.1	0.2	1.5	1.3	1.1	1.8	3.3	19	2.5	1.1	0.1	1.1	2.5	1.0	1.0	2.9	3.8
5	0.9	1.8	1.6	1.4	0.8	1.2	2.4	2.4	5.7	20	1.0	2.3	1.8	1.5	2.9	1.8	1.6	2.0	5.8
6	0.5	0.9	2.7	0.7	1.4	1.6	1.2	3.0	6.2	21	0.8	1.7	1.0	1.6	1.6	2.4	0.6	1.4	4.5
7	2.0	1.0	2.9	1.7	0.3	0.1	0.4	1.1	5.4	22	0.3	1.2	2.1	0.3	2.0	1.9	0.7	0.9	4.5
8	0.1	1.6	0.5	2.7	0.7	2.1	1.3	1.7	3.1	23	1.3	0.7	1.3	2.4	2.2	0.7	0.8	1.0	3.4
9	1.2	0.6	2.8	1.0	0.1	0.9	0.1	0.8	3.7	24	2.6	1.5	2.3	0.6	1.7	2.9	2.9	2.5	7.3
10	2.9	2.1	0.4	0.8	0.5	0.3	1.7	0.4	4.7	25	3.0	2.6	1.2	3.0	2.7	2.6	2.8	1.5	7.2
11	0.7	0.5	0.6	1.3	2.1	0.5	0.3	0.7	0.7	26	1.1	2.2	0.7	2.5	2.4	0.8	2.6	1.2	3.8
12	2.8	2.5	1.5	2.9	2.3	2.8	3.0	1.6	6.5	27	1.8	2.0	1.9	2.2	1.8	1.7	1.8	0.6	6.1
13	2.2	0.2	1.7	2.3	3.0	1.1	0.5	2.7	3.9	28	1.9	2.7	3.0	2.0	1.0	1.4	1.4	0.5	9.5
14	2.1	0.8	0.9	2.6	0.9	2.7	2.5	2.1	3.6	29	1.6	2.4	0.3	0.5	0.2	0.4	0.2	2.3	6.5
15	2.3	3.0	1.4	0.1	0.6	3.0	0.9	2.8	8.4	30	2.4	2.9	2.0	2.1	1.2	1.5	1.9	1.3	7.6

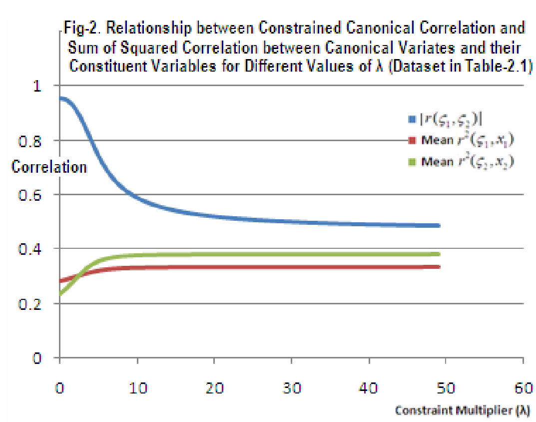


Table 2.2. Relationship between Constrained Canonical Correlation and Representation Correlation between Canonical Variates and their Constituent Variables for Different Values of λ (Dataset in Table-2.1)

Sl No.	λ	Canonical $\lvert r(\varsigma_1,\varsigma_2)\rvert$	Mean Squared $r(\varsigma_1,x_1)$	Mean Squared $r(\varsigma_2,x_2)$	Sl No.	λ	Canonical $\lvert r(\varsigma_1,\varsigma_2)\rvert$	Mean Squared $r(\varsigma_1,x_1)$	Mean Squared $r(\varsigma_2,x_2)$
1	0.0	0.95772	0.28514	0.23519	26	25.0	0.50983	0.33290	0.38230
2	1.0	0.94904	0.29212	0.26011	27	26.0	0.50804	0.33293	0.38234
3	2.0	0.91881	0.29987	0.28983	28	27.0	0.50638	0.33296	0.38238
4	3.0	0.86701	0.30782	0.31901	29	28.0	0.50475	0.33298	0.38241
5	4.0	0.80425	0.31506	0.34197	30	29.0	0.50340	0.33300	0.38244
6	5.0	0.74448	0.32066	0.35709	31	30.0	0.50203	0.33302	0.38247

Sl No.	λ	Canonical $\lvert r(\varsigma_1,\varsigma_2)\rvert$	Mean Squared $r(\varsigma_1,x_1)$	Mean Squared $r(\varsigma_2,x_2)$	Sl No.	λ	Canonical $\lvert r(\varsigma_1,\varsigma_2)\rvert$	Mean Squared $r(\varsigma_1,x_1)$	Mean Squared $r(\varsigma_2,x_2)$
7	6.0	0.69541	0.32452	0.36618	32	31.0	0.50086	0.33304	0.38249
8	7.0	0.65777	0.32703	0.37155	33	32.0	0.49966	0.33305	0.38252
9	8.0	0.62930	0.32867	0.37482	34	33.0	0.49858	0.33306	0.38254
10	9.0	0.60764	0.32976	0.37690	35	34.0	0.49755	0.33308	0.38256
11	10.0	0.59071	0.33052	0.37828	36	35.0	0.49659	0.33309	0.38257
12	11.0	0.57730	0.33106	0.37924	37	36.0	0.49571	0.33310	0.38259
13	12.0	0.56634	0.33146	0.37993	38	37.0	0.49492	0.33311	0.38260
14	13.0	0.55733	0.33176	0.38044	39	38.0	0.49409	0.33311	0.38261
15	14.0	0.54983	0.33199	0.38083	40	39.0	0.49333	0.33312	0.38262
16	15.0	0.54338	0.33217	0.38113	41	40.0	0.49265	0.33313	0.38263
17	16.0	0.53786	0.33232	0.38137	42	41.0	0.49193	0.33314	0.38264
18	17.0	0.53310	0.33244	0.38156	43	42.0	0.49132	0.33314	0.38265
19	18.0	0.52896	0.33253	0.38171	44	43.0	0.49074	0.33315	0.38266
20	19.0	0.52523	0.33262	0.38185	45	44.0	0.49018	0.33315	0.38267
21	20.0	0.52196	0.33268	0.38195	46	45.0	0.48958	0.33316	0.38268
22	21.0	0.51904	0.33274	0.38205	47	46.0	0.48912	0.33316	0.38268
23	22.0	0.51638	0.33279	0.38212	48	47.0	0.48852	0.33317	0.38269
24	23.0	0.51395	0.33283	0.38219	49	48.0	0.48812	0.33317	0.38270
25	24.0	0.51184	0.33287	0.38225	50	49.0	0.48761	0.33317	0.38270

We conduct yet another experiment with the dataset presented in Table 3.1. We find that ξ_1 for X_1 has the representation power 0.661265 (eigenvalue=2.645058) while ξ_2 for X_2 has the representation power 0.752979 (eigenvalue=3.764895). The $r(\xi_1,\xi_2)$ is 0.922764. Against these, results of the constrained canonical correlation (for λ=49) are: mean squared $r(\varsigma_1,x_1)$= 0.661261; mean squared $r(\varsigma_2,x_2)$=0.752966 and the constrained canonical correlation, $r(\varsigma_1,\varsigma_2)$= 0.923647. These results are once again corroborative to our earlier findings.

Table 3.1. Simulated Dataset-3 for Canonical correlation

Sl No.	X_1 or Dataset-1				X_2 or Dataset-2					Sl No.	X_1 or Dataset-1				X_2 or Dataset-2				
	X_{11}	X_{12}	X_{13}	X_{14}	X_{21}	X_{22}	X_{23}	X_{24}	X_{25}		X_{11}	X_{12}	X_{13}	X_{14}	X_{21}	X_{22}	X_{23}	X_{24}	X_{25}
1	0.7	1.1	1.6	1.3	1.1	0.1	2.7	1.5	1.8	16	2.1	2.7	2.7	1.8	1.8	0.9	6.4	1.6	2.3
2	1.3	1.2	0.8	1.0	0.3	1.2	1.1	0.4	0.3	17	2.2	2.9	2.6	2.3	3.0	3.0	6.9	1.9	2.8
3	1.7	2.5	1.8	1.1	1.9	2.3	5.5	2.6	2.1	18	2.6	1.9	2.8	2.2	2.5	2.8	5.0	2.4	3.0
4	2.9	2.4	1.9	2.8	2.7	2.2	4.9	2.3	1.9	19	1.6	1.3	2.4	3.0	1.7	2.1	4.3	2.0	2.9
5	0.4	1.0	0.2	0.9	0.5	2.5	1.7	1.0	0.1	20	1.9	0.9	2.9	1.9	1.5	2.0	3.6	2.1	1.4
6	0.6	0.4	2.2	0.4	1.0	0.8	2.4	1.7	0.4	21	1.5	0.2	0.4	0.7	1.3	1.6	0.8	0.2	0.9
7	0.8	0.1	0.7	0.8	0.1	0.2	0.4	0.1	0.5	22	1.8	0.5	1.1	0.5	1.2	1.4	2.1	0.8	1.0
8	2.3	2.8	3.0	2.6	2.6	2.9	6.7	2.5	2.7	23	1.4	0.7	0.5	1.6	0.4	1.9	2.6	2.2	1.5
9	1.2	2.0	0.9	1.7	2.4	0.7	3.2	1.8	2.0	24	1.0	1.6	0.3	0.1	0.7	1.1	1.2	0.3	0.7
10	2.4	2.1	2.5	2.5	2.1	1.5	3.9	2.7	1.7	25	0.5	1.8	1.4	2.7	0.2	1.8	3.3	1.3	1.3
11	0.2	0.6	0.1	1.5	1.4	0.4	0.6	0.5	0.2	26	3.0	2.6	2.3	2.4	2.0	1.7	5.1	3.0	2.2
12	2.0	1.5	0.6	0.3	0.8	0.6	1.5	0.6	1.6	27	2.5	1.4	1.3	2.1	2.3	2.6	3.8	2.9	2.4
13	0.9	0.3	1.7	2.0	1.6	0.5	2.5	0.9	0.8	28	0.3	2.2	1.2	0.2	0.9	1.3	2.3	1.1	0.6
14	2.7	3.0	2.1	2.9	2.8	2.7	7.0	2.8	2.6	29	2.8	2.3	2.0	1.4	2.9	2.4	5.8	1.4	2.5
15	1.1	0.8	1.0	1.2	0.6	0.3	2.0	1.2	1.2	30	0.1	1.7	1.5	0.6	2.2	1.0	4.2	0.7	1.1

Fig-3. Relationship between Constrained Canonical Correlation and Sum of Squared Correlation between Canonical Variates and their Constituent Variables for Different Values of λ (Dataset in Table-3.1)

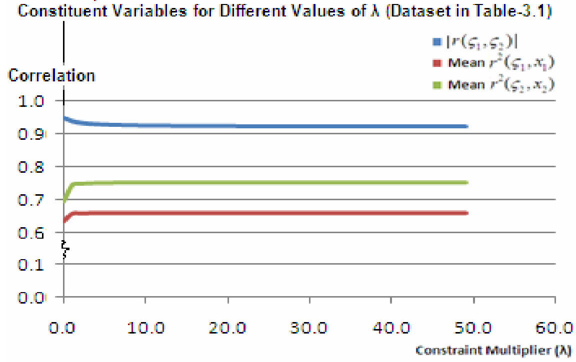


Table 3.2. Relationship between Constrained Canonical Correlation and Representation Correlation between Canonical Variates and their Constituent Variables for Different Values of λ (Dataset in Table-3.1)

Sl No.	λ	Canonical $\lvert r(\varsigma_1,\varsigma_2)\rvert$	Mean Squared $r(\varsigma_1,x_1)$	Mean Squared $r(\varsigma_2,x_2)$	Sl No.	λ	Canonical $\lvert r(\varsigma_1,\varsigma_2)\rvert$	Mean Squared $r(\varsigma_1,x_1)$	Mean Squared $r(\varsigma_2,x_2)$
1	0.0	0.94813	0.63711	0.69854	26	25.0	0.92443	0.66125	0.75293
2	1.0	0.93901	0.65895	0.74485	27	26.0	0.92437	0.66125	0.75294
3	2.0	0.93449	0.66036	0.74955	28	27.0	0.92431	0.66125	0.75294
4	3.0	0.93199	0.66076	0.75106	29	28.0	0.92426	0.66125	0.75294
5	4.0	0.93039	0.66094	0.75175	30	29.0	0.92421	0.66125	0.75294
6	5.0	0.92927	0.66104	0.75212	31	30.0	0.92417	0.66126	0.75295
7	6.0	0.92844	0.66110	0.75235	32	31.0	0.92412	0.66126	0.75295
8	7.0	0.92780	0.66114	0.75250	33	32.0	0.92409	0.66126	0.75295
9	8.0	0.92729	0.66116	0.75260	34	33.0	0.92405	0.66126	0.75295
10	9.0	0.92687	0.66118	0.75267	35	34.0	0.92401	0.66126	0.75295
11	10.0	0.92653	0.66119	0.75272	36	35.0	0.92398	0.66126	0.75295
12	11.0	0.92624	0.66121	0.75276	37	36.0	0.92394	0.66126	0.75296
13	12.0	0.92598	0.66121	0.75279	38	37.0	0.92392	0.66126	0.75296
14	13.0	0.92577	0.66122	0.75282	39	38.0	0.92389	0.66126	0.75296
15	14.0	0.92558	0.66123	0.75284	40	39.0	0.92386	0.66126	0.75296
16	15.0	0.92541	0.66123	0.75286	41	40.0	0.92383	0.66126	0.75296
17	16.0	0.92526	0.66123	0.75287	42	41.0	0.92381	0.66126	0.75296
18	17.0	0.92513	0.66124	0.75288	43	42.0	0.92378	0.66126	0.75296
19	18.0	0.92501	0.66124	0.75289	44	43.0	0.92376	0.66126	0.75296
20	19.0	0.92490	0.66124	0.75290	45	44.0	0.92374	0.66126	0.75296
21	20.0	0.92481	0.66124	0.75291	46	45.0	0.92372	0.66126	0.75296
22	21.0	0.92472	0.66125	0.75291	47	46.0	0.92370	0.66126	0.75296
23	22.0	0.92464	0.66125	0.75292	48	47.0	0.92368	0.66126	0.75297
24	23.0	0.92455	0.66125	0.75292	49	48.0	0.92366	0.66126	0.75297
25	24.0	0.92450	0.66125	0.75293	50	49.0	0.92365	0.66126	0.75297

6. A Computer Program for RCCCA

We developed a computer program in FORTRAN that we have developed and used for solving the problems in this paper (codes available at www.webng.com/economics/rccca.txt, which may also be obtained from the author on request). Its main program (RCCCA) is assisted by 13 subroutines. The user needs setting the parameters in the main program as well as in the subroutines CORD and DORANK. Parameter setting in RPS may seldom be required. This program can be used for obtaining Ordinal Canonical Correlation [Mishra, (2009)] also. Different schemes of rank-ordering may be used [Wikipedia, (2008)].

7. Concluding Remarks

Our proposed Representation-Constrained Canonical correlation (RCCCA) Analysis has the classical canonical correlation analysis (CCCA) at its one end (λ=0) and the Classical Principal Component Analysis (CPCA) at the other (as λ tends to be very large). In between it gives us a compromise solution. By a proper choice of λ, one can avoid hijacking of the representation issue of two datasets by a lone couple of highly correlated variables across those datasets. This advantage of the RCCCA over the CCCA deserves a serious attention by the researchers using statistical tools for data analysis. Our method also addresses the problem raised by Sugiyama (2007).

8. References

[1] Eberhart, R.C. and Kennedy, J., (1995), *A New Optimizer using Particle Swarm Theory*, Proceedings Sixth Symposium on Micro Machine and Human Science*:* 39–43. IEEE Service Center, Piscataway, NJ.

[2] Fleischer, M., (2005), *Foundations of Swarm Intelligence: From Principles to Practice*, Swarming Network Enabled C4ISR, arXiv:nlin.AO/0502003 v1.

[3] Hayek, F. A., (1948), *Individualism and Economic Order*, The University of Chicago Press, Chicago.

[4] Hayek, F. A., (1952), *The Sensory Order: An Inquiry into the Foundations of Theoretical Psychology*, University of Chicago Press, Chicago.

[5] Hotelling, H., (1936), *Relations between Two Sets of Variates*, in: *Biometrica*, 28: 321-377.

[6] Mishra, S.K., (2006), *Global Optimization by Differential Evolution and Particle Swarm Methods: Evaluation on Some Benchmark Functions*, available at SSRN: http://ssrn.com/abstract=933827

[7] Mishra, S.K., (2009), *A Note on the Ordinal Canonical Correlation Analysis of Two Sets of Ranking Scores*. Available at *SSRN*: http://ssrn.com/abstract=1328319

[8] Sugiyama, T., (2007), *Tests for Validity on Principal Canonical Correlation Analysis*, Atlas Conference, Interdisciplinary Mathematical & Statistical Techniques (Shanghai 2007), May 20-23, 2007, University of Science and Technology of China, Hefei, Anhui, P.R. China.

[9] Urfalioglu, O., (2004), *Robust Estimation of Camera Rotation, Translation and Focal Length at High Outlier Rates*, Proceedings of the 1st Canadian Conference on Computer and Robot Vision, IEEE Computer Society Washington, DC, USA: 464- 471.

[10] Wikipedia, (2008), *Ranking*, available at Wikipedia: http://en.wikipedia.org/wiki/Rank_order

CAN CREDIT DEFAULT SWAPS PREDICT FINANCIAL CRISES? EMPIRICAL STUDY ON EMERGING MARKETS

Hekuran **NEZIRI**
American University, **Bulgaria**
hekuranneziri@gmail.com

Abstract:
We explore the informational value of credit default swaps and the extent to which they may be linked to financial crises. After developing a theoretical framework to model the relationship between credit default swap market and equity and currency markets, we apply an empirical study which uses logistic regressions and a panel data sample of emerging markets to assess the ability of these financial instruments to predict crises. Regarding them as reflections of future expectations of investors on the outcomes of currency and equity markets, we find credit default swaps to be a significant indicator explaining the periods proceeding financial crises, at least in equity markets. The inclusion of credit default swaps as a factor in models that predict crises and their ability to improve predictions in equity market is a major contribution of this study to the existing literature.

Keywords: credit default swaps, stock market crises, currency crises, emerging market debt

JEL Classification: F3

1. Introduction

Credit default swaps have been all around the news on the current financial crisis. Economists, investors, politicians and almost everyone seemed to agree that these financial instruments contain important information that can be used to gauge the financial situation during the current financial crisis. An article in Wall Street Journal on Oct 31, 2008 reports:

Investors raise their bets on defaults in EU Countries implying that euro-zone economy heads into recession as costly bank-bailout plans could drive some European governments to default on their debt… Soaring [credit default] swap prices have preceded real calamity. Investor fears about the health of Wall Street firms like Bear Stern Cos. And Lehman Brothers Holding Inc. appeared in swap prices early on and contributed to capital flights that left the firms seeking government help. Credit default swaps have proven themselves a reliable indicator of trouble ahead.

However, can we say that an increase in credit default swap prices implies trouble in the near future for the reference entity? This question prompted us to explore the informational value of credit default swaps and the extent to which they may be linked to financial crises. Specifically, we investigate the ability of fluctuations in these premiums to predict the occurrence of financial crises.

The analysis explores the ability of credit default swaps to predict stock market and currency market crises. The hypothesis is that CDS premiums reflect future expectations of investors on outcomes in currency and equity markets. The premium allows for a clear view of investor perception of risk. Logistic models with panel data from emerging markets are used to assess the crises predictive power of premiums written on sovereign obligations of emerging markets. We check the effect on both stock and currency markets.

To our knowledge, this is the first paper to use credit default swap premiums in predicting financial crises. However, other factors which were created to reflect investor sentiment have been previously used. In principle financial crises should be preceded by periods of increased risk aversion among investors. Curdert and Gex (2007) test whether some main risk aversion indexes are able to predict crises. They find that "risk appetite" tends to decrease prior to financial crises. Also, they do not fail to mention that the opposite case is possible. Crises may follow after periods of strong "risk appetite" during which investors are excessively optimistic and hence create "speculative bubbles" on the prices of risky assets. The recent mortgage crisis is an example of increased investor "risk appetite" prior to crises. According to their findings, risk aversion indicators had more predictive ability in stock market crises then currency crises.

The study is organized as follows. Section 2 briefly discusses the market for credit default swaps and establishes a theoretical basis for relationship between CDS premiums, equity prices and currency trends. Section 3 summarizes the literature on financial crises prediction. Section 4 presents the model specification and the definition of crises. Section 5 describes the panel data. Section 6 presents the results and Section 7 offers concluding remarks.

2.Theoritical Framework

The credit default swap (CDS) market is the largest market of credit derivatives. In a CDS transaction, the protection buyer pays a series of fixed periodic payments (CDS premium) to the protection seller in exchange for a contingent payment in case of a credit event, such as bankruptcy, credit downgrade or a failure to make the scheduled payments. Duffie (1999) explains that the CDS premium is equivalent to swapping the payments from a risky security for the payments of a risk-free security in exchange of a contingent payment in case the risky security defaults. Hence, the premium reflects the credit risk of the underlying asset and is normally quoted in basis points over a reference rate, supposed to be a risk-free rate.

CDSs are actively traded on corporate bonds and sovereign debt. This paper focuses on emerging markets where most contracts reference sovereign obligations. Sovereign CDSs are considered to be the most liquid credit derivative instruments in emerging markets. The usual contract is written on notional amounts of $10 million with a five-year maturity. For instance, a 5 year CDS rate of 200 for a Bulgarian international bond means that it costs $200 per annum to insure a $10,000 face value Bulgarian international bond.

Players in the emerging market CDSs use the contracts for a number of reasons. CDSs allow speculation on the future creditworthiness of countries. Also, they allow exploitation of arbitrage opportunities that may arise from the spread between CDS and the referenced bond. In addition, CDSs are used to manage the exposure to sovereign bonds. Participants in the market utilizing these opportunities range from hedge funds, mutual funds, banks to pension funds. Because of the players involved and the theory that follows, we assume that CDS premium reflects future expectations of the investors and the premium can be used to predict outcomes in currency and equity markets in emerging countries.

First, we discuss our theoretical basis for the effect of the changes in CDS premium on the probability of a currency crisis in the emerging markets. Let N be the notional amount of a contract, **s** be the CDS rate (premium) and p be the default probability of bond payments. Seeker of protection against default will buy the contract if the present value of premium payments will be equal or less than the present value of the expected loss from default:

$$\frac{Ns(1-p)}{1+r} \leq \frac{Np(1-R)}{1+r} \tag{1}$$

where R is the recovery rate in case of default and r is the interest rate. The above equation produces an expected default probability, set by investors, that is proportional to the premium paid:

$$p \leq \frac{s}{s+1-R} \tag{2}$$

where $1 - R$ falls into [0,1]. According to Eq. (2), an increase in the CDS premium (s) at a fixed R, indicates an increased default probability on bond payments of the reference entity (p).

In cases where the reference entity is sovereign debt, the increased default concern translates into a tendency of the currency to depreciate, as concluded by Cochrane (2004). The argument comes from fiscal theory and the theory of optimal distorting taxation. Chochrane uses the analogy of money as a stock in fiscal theory to establish a relation between currency devaluations and fiscal balances. From the theory of optimal distorting taxation, Chochrane argues that a currency crash represents a choice by the government to devalue outstanding nominal debt rather than increase distortionary taxes. Hence, we expect an increase in the probability of currency crisis in a country when there is an increase in the CDS premium.

Second, the theory for the effect of the changes in CDS premium on the probability of a stock market crisis is based on Merton's (1974) model and its extension to sovereign debt by Chan-Lau & Kim (2004). First, Duffie (1999) notes that the yield from a holding a risky bond and paying a CDS premium, is equivalent to the yield from holding a risk-free bond. However, Chan-Lau & Kim (2004) observe that in practice the risk-free yield and the yield of a risky asset with CDS is not the same, because there is always a difference in bond spreads and CDS spreads evident in CDS-bond differential. The CDS-bond differential partly exists because of the different liquidity in the markets and the cheapest-to-deliver option premium in bonds. However, in the long-run, there should be a co-integration between bond spreads and CDS spreads.

Merton's model links bond and equity prices by taking a balance sheet approach. It argues that if the value of a firm's assets falls below the face value of its debt, the firm defaults. Also, there is a positive correlation between bond and equity prices, and hence equity prices and bond spreads move in opposite direction. Chan-Lau and Kim (2004) extend the model to include sovereign obligations as equivalents of the firms' debt.

From the CDS-bond differential and the relationship between equity prices and bond spreads, we infer that CDS spreads and equity prices move in opposite directions. Furthermore, we use CDS as a source of information for investor expectations in country's equity market, from our assumption that CDSs reflect future market expectations. So, we expect a positive relationship between changes in CDS premium and the probability of a crisis in stock markets.

3.Literature review

The idea underlying the empirical research of crises prediction is to identify some factors that show specific patters prior to periods of crises. The goal is to build a system that assesses the probability of crises at a specific time horizon, taking into consideration all information available at the time of prediction. There are three methodological approaches used in literature on currency and stock market crises. The first approach does not concentrate on the factors that caused the crisis but rather wants to analyze the effects of crisis on some specific sector of the economy. An example is Sachs, Tornell and Velasco (1996), who examine the implications of 1995 crises and try to answer the question of why some emerging markets were hit by the crises while others not.

Kaminsky and Reinhart (1996) devise a methodology called a "signal approach" to identify the periods where a crisis will occur. Numerous papers follow this method which looks for any pattern in individual variables prior to crisis. When a pattern is found (e.g. a deviation from mean up to a certain threshold) a signal is issued by the variable tested. The threshold is chosen so as to minimize the false signals. The advantage of this approach is that it produces easily understandable results for policy purposes. However, it ignores the interaction between independent variables and standard statistical tests cannot be applied.

This paper follows the third approach which eliminates some of the disadvantages of the signal approach by using a limited dependent variable. The method uses a logistic function to evaluate the overall effect of the explanatory variables and predict an outcome, i.e. the probability of the crises, constrained by zero and one. Kumar et al (2003) use a logistic model to study currency crises in 32 developing countries for a period of 15 years.

Factors suggested by Kaminsky, Lizondo and Reinhart (1998) and used in this paper to predict currency crises are: terms of trade, real interest rate, current account deficit, unemployment rate, GDP growth, changes in consumer prices, and returns in stock market indices. For the stock market, factors used in this paper and suggested by Boucher (2004) and also used by Curdert and Gex(2007) are price earnings ratio, stock returns and real interest rates.

The theory of financial crises suggests that economic fundamentals are the main cause of the financial crises. By economic fundamentals, we mean macroeconomic factors such as GDP growth, unemployment, inflation etc. The exact timing of the crises was first linearly determined and the crises had been predictable with economic fundamentals. Kaminsky and Lizondo (1996) specify models that predicted crises by using economic fundamentals in a non-linear fashion. However, economic fundamentals were not the only leading indicators of the crises. Investor behavior and other political and geographic factors were also taken into consideration. This paper argues for the inclusion of CDS premiums alongside economic fundamental factors to improve the predictability of the crises.

4. Definition of Crises and Model Specification

To test whether sovereign credit default swaps are a leading indicator of financial crises, we construct three models in stock markets and currency markets respectively. The first model is referred to as the "base" model and includes variables usually used in the literature to predict crises. The second model adds to the base model changes in CDS premiums, and will be indicative of the ability of the sovereign CDS to improve the forecasts of the existing models. It does so by checking the significance of the factor in presence of other factors currently used. The third model has changes in CDS premiums as the only independent variable. The dependent variable used is a qualitative variable while the independent variables are exogenous quantitative variables. Therefore, non-linear models are used to link the crisis prediction indicators as the dependent variables to the changes in CDS premiums and other quantitative variables as the independent variables.

Periods of crises are identified by constructing an indicator ($Crisis_{i,t}$), which is used as the dependent variable in the model. The indicator takes the value of 1 if there was a crisis within past 6 months and a value of 0 otherwise. In the regression models that follow, we estimate the probability that the crisis indicator is equal to 1 in a six-month horizon in both currency market and stock market of the emerging countries under consideration. Next, we define crises in currency and stock markets and then explain the regression methodology.

4.1. Currency Market Crisis Definition

Following Sachs, Tornell and Velasco 1996, Kaminski, Lizondo and Reinhart (1998) and Corsetti, Pesenti and Roubini (1999), our models specify a currency crisis when there is a simultaneous increase in currency depreciation and foreign exchange reserves losses. As a convention in literature, a currency pressure index is constructed by the following formula:

$$cpindex_{i,t} = \Delta E_{i,t} - \left(\sigma_{E_{i,t}} / \sigma_{R_{i,t}}\right)\Delta R_{i,t} \qquad (3)$$

where $\Delta E_{i,t}$ measures the devaluation of the nominal exchange rate in terms of dollars, $\Delta R_{i,t}$ measures the change in the country's foreign reserves, and $\left(\sigma_{E_{i,t}} / \sigma_{R_{i,t}}\right)$ is the ratio of standard deviations. The index has an advantage of being able to analyze speculative attacks on currencies under both fixed and flexible exchange rate systems. An increase in the reserves reflects foreign currency inflows and lowers the pressure on depreciation of the local currency because of the negative sign in the equation. So, $cpindex_{i,t}$ measures the depreciation pressure of a currency.

We define a currency crisis when the pressure of a currency goes beyond a certain threshold. In empirical studies, the threshold used falls between one and three standard deviations above the mean of the index. This paper uses the following formula to identify the crisis periods:

$$crisis_{i,t} = \begin{cases} 1, \text{ if } cpindex_{i,t} \geq \mu_{cpindex_{i,t}} + 2.0\sigma_{cpindex_{i,t}} \\ 0, \; otherwise \end{cases} \qquad (4)$$

where $crisis_{i,t}$ is the crisis indicator of country i at time t, $\mu_{cpindex_{i,t}}$ is the sample mean of the pressure index and $\sigma_{cpindex_{i,t}}$ is the standard deviation of the pressure index. With a threshold of two standard deviations above the mean, a total number of 24 crises are identified. Table 1 below shows frequencies of crises and tranquil periods in the respective countries.

Table 1. Number of observations, frequency of crises and tranquil periods in both stock markets and currency markets categorized by country

		Currency Market			Stock Market		
Country	Obs.	Crises	Tranquil	% Crises	Crises	Tranquil	% Crises
Argentina	39	3	36	8.33%	0	39	0.00%
Brazil	83	0	83	0.00%	2	81	2.47%
Bulgaria	95	0	95	0.00%	8	87	9.20%
Chile	68	0	68	0.00%	0	68	0.00%
China	68	3	65	4.62%	6	62	9.68%
Colombia	68	2	66	3.03%	0	68	0.00%
Croatia	95	1	94	1.06%	2	93	2.15%
Hungary	78	0	78	0.00%	0	78	0.00%
India	43	0	43	0.00%	1	42	2.38%
Indonesia	43	1	42	2.38%	0	43	0.00%
Israel	49	0	49	0.00%	0	49	0.00%
Malaysia	82	2	80	2.50%	0	82	0.00%
Mexico	83	2	81	2.47%	0	83	0.00%
Peru	59	1	58	1.72%	2	57	3.51%
Philippines	77	1	76	1.32%	2	75	2.67%
Poland	95	3	92	3.26%	4	91	4.40%
Russia	95	2	93	2.15%	1	94	1.06%
South Africa	95	3	92	3.26%	0	95	0.00%
South Korea	95	0	95	0.00%	1	94	1.06%
Thailand	77	0	77	0.00%	0	77	0.00%
Turkey	77	0	77	0.00%	9	68	13.24%
Total	1564	24	1540	1.56%	38	1526	2.49%

4.2. Stock Market Crisis Definition

To identify the crises periods in stock markets, we follow Mishkin and White (2002) definition of stock market crisis as falls in price of an index below some threshold during a specified period of time. We take the threshold to be 25 percent and the period to be 6 months. Country's main stock market indexes have been used as indicators of the stock market situation. Table 2 in Appendix shows the stock market indices used for respective countries. We use the following formula to identify crises in the emerging stock markets:

$$crisis_{i,t} = \begin{cases} 1, \text{ if } \left(P_{i,t} / P_{i.t-6}\right) - 1 <= -0.25 \\ 0, \; otherwise \end{cases} \tag{5}$$

where $crisis_{i,t}$ is the stock market crisis indicator of an emerging country denoted by i at time t, $P_{i,t}$ is the price of the index at time t while $P_{i.t-6}$ is the price of the index six months ago. The formula considers a crisis to be a drop in the index of more than a quarter of its value six month ago. Using the above definition, we identify a total number of 38 crises in our sample data. Table 1 above shows frequencies of crises and tranquil periods in respective countries.

4.3. Model Specification

Three models are tested to see whether changes in sovereign credit default swaps improve forecasts in emerging stock markets and currency crises respectively. The base model includes variables usually used in literature of predicting crises in emerging markets. The second model adds changes in sovereign CDS premiums to the base model, while the third tests the significance of a model which includes only changes in CDS premiums as explanatory variable. Variables in all three models are lagged by one month to check the predictability a crisis a month in advance. In all the regression models for both markets, we use a logistic function of the type:

$$f(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{1+e^x}\ . \tag{6}$$

The first model uses the following logistic regression model to estimate the probability of a crisis in stock and currency markets of emerging countries within a one-month horizon by regressing the indicator on variables commonly used in literature:

$$\Pr(crisis_{i,t} = 1) = f\left(\alpha + \sum_{k=1}^{n} \beta_k X_{i,t-1}^{k}\right) \tag{7}$$

where α and β are the coefficients while $X_{i,t-1}^{k}$ are the variables used in past studies on the field. For currency crises model, the variables used are: real interest rate, terms of trade, current account, unemployment rate, GDP growth, inflation and one month stock returns. However, there is not as much literature that addresses stock market crises as there is for currency crises. We use the same factors as in Curdert and Gex (2007) for stock market crises model, namely: price earnings ratio of the indices, one month stock returns and real interest rates.

The second model extends the first model by adding changes in sovereign CDS premiums as a factor predicting crises in emerging markets. The following equation is estimated:

$$\Pr(crisis_{i,t} = 1) = f(\alpha + \sum_{k=1}^{n} \beta_k X_{i,t-1}^{k} + \beta_{n+1} \Delta CDS_{i,t-1}) \tag{8}$$

where $\Delta CDS_{i,t-1} = (CDS_{i,t-1} / CDS_{i,t-2}) - 1$ is the lagged one-month change in sovereign CDS premiums on the international bonds of the emerging countries considered in this paper. Here, fluctuations in CDS are regressed on the presence of other factors used in the first model in order to check its ability to improve the forecasts.

The third model uses the following logistic regression equation to estimate crises probabilities by using changes in sovereign CDS premiums as the only factor:

$$\Pr(crisis_{i,t} = 1) = f(\alpha + \beta \Delta CDS_{i,t-1}) \tag{9}$$

where $\Delta CDS_{i,t-1} = (CDS_{i,t-1} / CDS_{i,t-2}) - 1$ is the lagged one-month change in sovereign CDS premiums as in the second model.

5. Data

The dataset consists of 21 emerging market countries: Argentina, Brazil, Bulgaria, Chile, China, Colombia, Croatia, Hungary, India, Indonesia, Israel, Malaysia, Mexico, Peru, Philippines, Poland, Russia, South Africa, South Korea, Thailand and Turkey. We use panel data with monthly frequencies starting from the date when CDS quotes were available for the respective reference country until August 2008. We used Bloomberg terminal to retrieve stock market index levels, P/E ratios of indexes and 5 year sovereign credit default swap premiums whereas IMF's International Financial Statistics database was used to retrieve other variables.

The choice of independent variables in both markets was based on earlier studies, and the variables were earlier found to be related to currency and stock market crises respectively. Some variables needed to be created from the retrieved series. In stock markets, six month stock return series were created by using the index price series obtained from Bloomberg. In currency markets, following Kaminsky et al (1998), we created terms of trade as the ratio between exports and imports. Table 2 below summarizes the independent variables and their sources.

Table 2. Data sources and frequencies as they were retrieved

Variables for Stock Market Crises Index	Source	Frequency and Period
Country's Main Stock Market Index	Bloomberg	Monthly, 2000:10 - 2008:8

Independent Variables in Stock Market Model	Source	Frequency and Period
5 YR Credit Default Swaps in USD	Bloomberg	Monthly, 2000:10 - 2008:8
Price/Earnings Ratio of Index	Bloomberg	Monthly, 2000:10 - 2008:8
Stock Returns	Bloomberg	Monthly, 2000:10 - 2008:8
Real Interest Rates	IMF IFS line 60 and 64.X	Monthly, 2000:10 - 2008:8

Variables for Currency Market Crises Index	Source	Frequency and Period
Exchange Rate National Currency per USD	IMF IFS line AE	Monthly, 2000:10 - 2008:8
Total Reserves Minus Gold	IMF IFS line 1LD	Monthly, 2000:10 - 2008:8

Independent Variables in Currency Market Model	Source	Frequency and Period
5 YR Credit Default Swaps in USD	Bloomberg	Monthly, 2000:10 - 2008:8
Gross Domestic Product	IMF IFS line 99B	Annual, 2000 - 2008
Current Account	IMF IFS line 78ALD	Annual, 2000 - 2008
Unemployment Rate	IMF IFS line 67R	Annual, 2000 - 2008
Terms of Trade	Kaminsky et al (1998)	Quarterly, 2000:9 - 2008:6
Country's Main Stock Market Returns	Bloomberg	Monthly, 2000:10 - 2008:8
Real Interest Rates	IMF IFS line 60 and 64.X	Monthly, 2000:10 - 2008:8
Change in Consumer Prices	IMF IFS line 64.X	Monthly, 2000:10 - 2008:8

We used linear interpolation to convert the annually and quarterly data observations into monthly frequencies. The variables converted were: GDP growth, current account and unemployment rate. However, in cases where data with lower frequency was available for some countries on these series, we retrieved the lower frequency data. For the monthly series, the last day of the month is used.

By correlation coefficients, we checked whether CDS is showing information that is already contained in another variable. Table 3 below shows correlation matrices for both markets. The correlation coefficient that measures the relationship of CDS to other variables falls in the range of -0.063 (GDP growth) and 0.548 (Inflation). Any coefficient significantly different from 1 and -1 shows that we cannot fit a linear relationship between those variables. Therefore, we argue that information contained in CDS is distinct from the one contained in other variables and hence, it can improve results.

Table 3. Correlation matrices of variables used in stock market and currency market models

Stock Market

	Return	CDS	Real Int Rate	P/E Ratio
One Month Return	1			
CDS	-0.01	1		
Real Interest Rate	0.0279	0.3433	1	
Price/Earnings Ratio	0.0447	0.0825	-0.0821	1

Currency Market

	CDS	Terms	Curr Account	Unemploy.	Inflation	Real Int Rate	GDP Growth	Return
CDS	1							
Terms of Trade	0.0442	1						
Current Account	-0.0487	0.0426	1					
Unemployment Rate	0.1371	0.3065	-0.1246	1				
Inflation	0.548	0.1443	0.0569	-0.0367	1			
Real Interest Rate	0.1129	0.0884	-0.1359	0.0699	-0.0454	1		
GDP Growth	-0.0639	0.0444	0.1508	-0.1015	0.023	0.0666	1	
One Month Return	0.0049	0.0148	0.1609	-0.0249	0.0377	-0.0614	0.1201	1

Panel data unit root test as suggested by Im *et al.* (2003) is applied to check whether we have stationary variables. The test rejected the null hypothesis that all series are non-stationary against an alternative that all series are stationary at a 10 percent level of significance. Hence, the variables used in the regressions are trend stationary.

6. Empirical Results

Models specified in the methodology were regressed to estimate the variables that affect the probability of stock market crises and currency market crises in emerging markets. We used panel data with one month lagged independent variables. Because of the short time period of the data we possessed, we did only in-sample predictions. Table 4 below shows the coefficients and marginal effects or slope coefficients of the factors supposed to affect the probability of the stock market crises whereas Table 5 below presents the same regression results for the factors assumed to affect the probability of the currency crises. First, we discuss the results in both markets. Then, we explore the predictive ability of our models and finally, we compare the predictive power of models in stock and currency markets.

In the stock market models, all three models proved to be significant. In the first model, the regression on the currently used factors in literature produced a pseudo squared-R of 0.059 and the model proved significant at a one percent confidence level. Unexpectedly, P/E ratios and real interest rates are insignificant in predicting stock market crises. In the second model, we improve the pseudo squared-R to 0.077 after adding the one month changes in the CDS premiums in the base model. One month changes in CDS premiums are statistically significant at one percent and improve forecasts of stock market crashes a month in advance in emerging markets. The factor has the expected positive sign, which reflects the assumption that ceteris paribus, an increase in premiums signals a higher probability of default in emerging stock markets. Also, changes in CDS premiums are significant also in the third model, when they are the only factor predicting crises. So, an increase in the default probability of bond payments by a country, which is derived from the credit default swap premium, can be interpreted as a factor signaling an increase in the probability of a stock market crisis in the same country.

Table 4. Logit regression models predicting stock market crises[17]

Dependent variable: Crisis Indicator	Model (1)		Model (2)		Model (3)	
Independent variables (t-1):	Coefficients	Marginal Effects	Coefficients	Marginal Effects	Coefficients	Marginal Effects
Constant	-3.505637*** [0.3481558]		-3.950127*** [0.2009473]		-3.973741*** [0.1925427]	
Sovereign Credit Defalut Swaps			.0008721*** [0.0003182]	0.00000119*** [0.00002]	0.0010651*** [0.0002776]	.0000233*** [.00001]
P/E Ratio of Stock Markets	-0.0005738 [0.0121172]	-6.79e-06 [0.00014]	-.000612 [0.0121767]	-0.00000725 [0.00014]		
Real Interest Rate	-0.170696 [0.0560089]	-0.0020196 [0.00059]	-.0334392 [0.0307316]	-0.0020251 [0.0006]		
One Month Stock Returns	-8.266678*** [3.192179]	-0.0978076*** [0.04138]	-8.640637*** [2.129019]	-0.0978626*** [0.04142]		
Observations	1521		1521		1543	
Log Likelihood	-163.73714		-160.6004		-169.41479	
Pseudo squared-R	0.0592		0.0773		0.0296	
Chi-square	20.62		26.89		10.34	
P-Value	0.0000		0.0000		0.0013	

[17]Standard errors are in brackets, * significant at 10%, ** significant at 5%, *** significant at 1%, marginal effects give the estimated slope coefficient, coefficients are estimated for the original logistic model

We do not have the same impressive results in currency markets. The regression on the factors currently used in literature produced a pseudo squared-R of 0.14 with GDP growth and inflation failing to be statistically significant as leading indicators of currency crises. Ceteris paribus, increases in stock returns or real interest rates contribute to lower probability of currency crisis while increases in terms of trade, current account or unemployment rate signal a higher probability of currency crises. Changes in CDS premiums are statistically significant at 5 percent confidence level in the third model and they have the expected positive sign, where an increase in the premiums corresponds to a higher probability of currency crises. However, a test on the slope coefficient of changes in CDS premiums in the second model cannot reject the null hypothesis that the coefficient is significantly different from zero. So, changes in CDS premiums are insignificant in presence of currently used factors in literature. Hence, the inclusion of the factor in the base model does not improve forecasts of currency crises in emerging markets.

However, one cannot draw final results from conclusions by looking at the coefficients and marginal effects of the factors only. Statistics on predictive ability of our models should also be considered. To obtain meaningful statistics about predictions from our data-sample, we set a probability threshold above which, it is decided that the model predicts a crisis. Other studies choose a threshold that maximizes the share of correctly classified observations. We follow Peltonen (2006) and use four different probability thresholds. Table 6 and Table 7 show statistics on predictive ability of our models for different probability thresholds.

Table 5. Logit regression models predicting currency market crises[18]

Dependent variable: Crisis Indicator Independent variables (t-1):	Model (1) Coefficients	 Marginal Effects	Model (2) Coefficients	 Marginal Effects	Model (3) Coefficients	 Marginal Effects
Constant	-6.738983*** [1.077809]		-6.487072*** [1.086092]		-4.23099*** [0.2203864]	
Sovereign Credit Defalut Swaps			0.8368797 [1.062977]	0.0046655 [0.00602]	1.63882** [0.6597416]	.0236229*** [0.00923]
Terms of Trade	0.8058111** [0.3655891]	0.0046468 ** [0.00215]	0.5876072** [0.3637605]	0.0024199** [0.00172]		
Current Account	0.0000107*** [4.79e-06]	6.20e-08*** [0.00000]	8.91e-06*** [5.11e-06]	3.67e-08*** [0.00000]		
Unemployment Rate	0.0830174** [0.0354746]	0.0004787** [0.0002]	0.0954109** [0.0356845]	0.0003929** [0.00018]		
Change in Consumer Prices	0.0759931 [0.0569267]	0.0004382 [0.00036]	0.1838124 [0.0814281]	0.000757 [0.00045]		
Real Interest Rate	-0.01649704 ** [0.0788175]	-0.0009513** [0.00045]	-0.1489212** [0.0946067]	-0.0006133** [0.00039]		
GDP Growth	3.375028 [4.032087]	0.0194626 [0.0232]	3.520752 [4.609704]	0.0144992 [0.01952]		
One Month Stock Returns	-8.855517* [3.613523]	-0.0510666* [0.02639]	-9.64336* [3.751803]	-0.0397134* [0.02234]		
Observations	1209		1080		1522	
Log Likelihood	-90.690727		-69.647606		-120.9699	
Pseudo squared-R	0.1439		0.1636		0.0197	
Chi-square	30.48		27.25		4.87	
P-Value	0.0000		0.0006		0.0274	

The statistics on the predictive ability of stock market crises show that the second model outperforms the base model by the share of correctly predicted crises at 0.1 and 0.15 probability

[18]Standard errors are in brackets, * significant at 10%, ** significant at 5%, *** significant at 1%, marginal effects give the estimated slope coefficient, coefficients are estimated for the original logistic model

threshold. The second model correctly predicts 27.27% and 33.33% while the base model predicts 21.05% and 16.67% of the crises at 0.1 and 0.15 thresholds respectively. At 0.25 probability threshold, the base model predicts one crisis and makes no false alarm while the second and third models correctly predict two crises out of five. Adding CDS premium changes to the base model improves the predictive ability and at best, it predicts 40% of the crises.

Changes in CDS premiums are statistically insignificant in the second model, and hence the predictive ability of the base model and the second model is the same. However, the CDS premiums are significant in the third model but underperform the base model in the predictive ability. The model does not predict any crises at a probability threshold of more than 0.1. At a threshold of 0.05, the third model correctly predicts only 4.76% of the crises while the base model predicts 16.13%. At best, the base model correctly predicts 42.46% of the crises at 0.1 and 0.15 probability thresholds.

Table 6. Forecasts of crises probabilities in currency market models[19]

Currency Market		Number of crises predicted	Crises Predicted Correctly Pr(D| +)	False Alarms Pr(~D| +)	Sensitivity Pr(+| D)	Specificity Pr(-| ~D)	Share of correctly classified obs.
Model (1)							
Threshhold							
	0.05	62	16.13%	83.87%	47.62%	95.62%	94.79%
	0.1	7	42.86%	57.14%	14.29%	99.66%	98.18%
	0.15	7	42.86%	57.14%	14.29%	99.66%	98.18%
	0.25	5	40.00%	60.00%	9.52%	99.75%	98.18%
Model (2)							
Threshhold							
	0.05	62	16.13%	83.87%	47.62%	95.62%	94.79%
	0.1	7	42.86%	57.14%	14.29%	99.66%	98.18%
	0.15	7	42.86%	57.14%	14.29%	99.66%	98.18%
	0.25	5	40.00%	60.00%	9.52%	99.75%	98.18%
Model (3)							
Threshhold							
	0.05	21	4.76%	95.24%	4.17%	98.66%	97.17%
	0.1	1	0.00%	100.00%	0.00%	99.93%	98.36%
	0.15	0	0.00%	0.00%	0.00%	100.00%	98.42%
	0.25	0	0.00%	0.00%	0.00%	100.00%	98.42%

Table 7. Forecasts of crises probabilities in stock market models*

Stock Market		Number of crises predicted	Crises Predicted Correctly Pr(D| +)	False Alarms Pr(~D| +)	Sensitivity Pr(+| D)	Specificity Pr(-| ~D)	Share of correctly classified obs.
Model (1)							
Threshhold							
	0.05	112	12.50%	87.50%	37.84%	93.40%	92.04%
	0.1	19	21.05%	78.95%	10.81%	98.99%	96.84%
	0.15	6	16.67%	83.33%	2.70%	99.66%	97.30%
	0.25	1	100.00%	0.00%	2.70%	100.00%	97.63%
Model (2)							
Threshhold							
	0.05	93	11.83%	88.17%	29.73%	94.47%	92.90%
	0.1	22	27.27%	72.73%	16.22%	98.92%	96.91%
	0.15	9	33.33%	66.67%	8.11%	99.60%	97.37%
	0.25	5	40.00%	60.00%	5.41%	99.80%	97.50%
Model (3)							
Threshhold							
	0.05	23	8.70%	91.30%	5.41%	98.61%	96.37%
	0.1	8	25.00%	75.00%	5.41%	99.60%	97.34%
	0.15	7	28.57%	71.43%	5.41%	99.67%	97.41%
	0.25	5	40.00%	60.00%	5.41%	99.80%	97.54%

[19]* Crises predicted correctly were classified when there was a prediction that a crisis will happen and the crisis in fact happened. False alarms were considered when a crisis was predicted but it did not take place in reality. Sensitivity measures the percent probability of a crisis to have been predicted when a crisis happens. Specificity measures the probability that no crisis is predicted when no crisis is taking place. Share of correctly classified observations measures the percentage of correctly predicted situations in the market out of all the data points used. D stands for "crises happening" while ~D is the opposite. + stands for "predicted crisis" while – stands for "no crisis predicted".

Looking at country-specific forecasts, the most successful prediction of a stock market crisis was for the Brazilian Bovespa index. Also, the model produced successful predictions for the Turkish stock market with the exception of predicted values for 2003. The worst results from the stock market model are for the Russian stock market index where we have a predicted crisis probability of more than 20 percent and yet there is no crisis as defined in this paper. Figure 1 below shows four graphs of interest for forecasts for stock market crises by country.

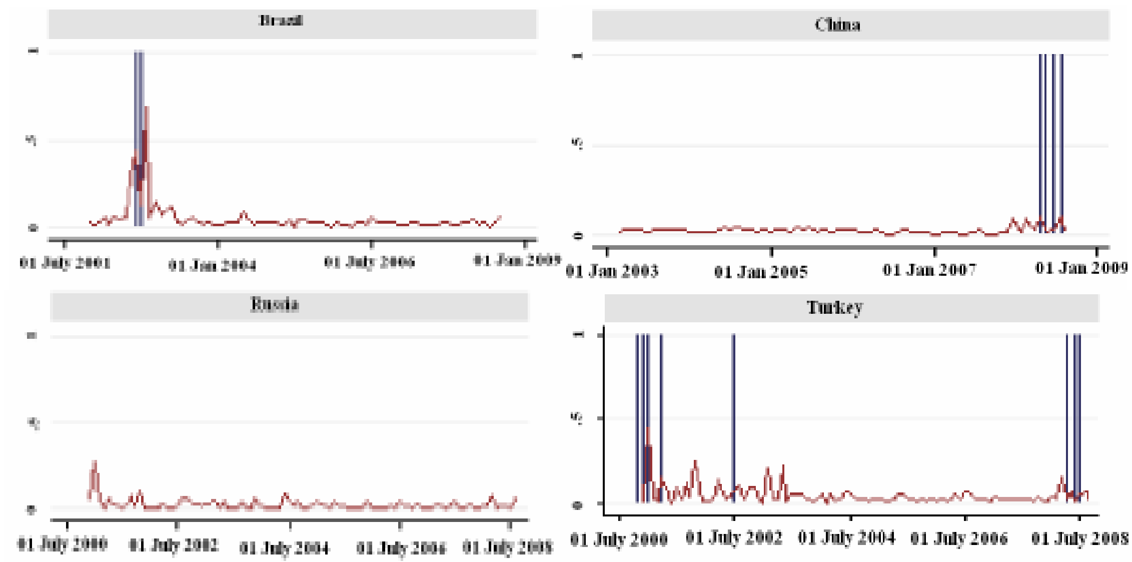


Figure 1. Graphs' of model (2) forecasts for stock market crises for four countries. The blue bar shows the occurrence of a crisis as defined in this paper and the red line shows the estimated probability of a crisis by model(2)

On the other hand, the most successful prediction of a currency market crisis was for the South African Rand in 2003, although we don't have a predicted crisis probability of more than 10 percent for this currency. The worst result from the model that predicts currency crises was for the Chinese Renminbi. The model predicts a currency crash with more than 50 percent probability during the second quarter of 2008 although there is no crisis. Figure 2 below shows the graphs of forecasts for currency market crises by country.

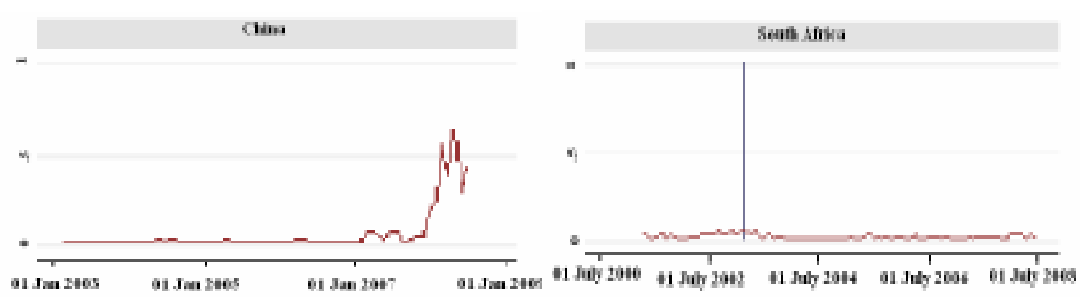


Figure 2. Graphs' of model (2) forecasts for currency market crises for two currencies. The blue bar shows the occurrence of a crisis as defined in this paper and the red line shows the estimated probability of a crisis by model(2)

To sum up, changes in sovereign credit default swap premiums have statistically significant impact on the probability of stock market crises in emerging markets, while the same is not true for the currency crises. Credit default swaps alone were able to predict 40% of the stock market crises and 4.76% of the currency market crises.

7. Conclusion

The purpose of this study was to examine whether changes in sovereign credit default swap premiums were able to predict stock market or currency market crises in emerging markets. The logistic regression results show that the one month change in a country CDS premium tends to increase one month ahead of a crisis in the stock market, while it does not have a significant implication for the currency market. Also, the predictive power was satisfactory for stock market crises. By contrast, these financial instruments were found to be insignificant in presence of other factors in explaining currency crises. The findings here confirm the results from Cudert and Cox (2007) that factors that measure investor sentiment have more predictive power in stock market crises than in currency crises. The inclusion of such a financial instrument as a factor in models that predict financial crises and its ability to improve predictions in stock market crises is the major contribution of this study in the existing literature. We established that fluctuations in CDS premiums signal trouble in stock markets.

Recently, because of the critiques directed to the industry for the low level of transparency in the CDS market, The Depository Trust & Clearing Corp. has decided to publish quotes of CDS online and free for public. It will be interesting to see whether this move will have an effect on the crises predictive ability of these instruments in the future. This issue is beyond the scope of this study; however, the informational value of these financial instruments is without doubt immense.

8. References:

[1] Boucher, Cristophe, (2005), *Stock Prices, Inflation and Stock Return Predictability*, Social Science Research Network.

[2] Chan-Lau, Jorge A.; Yoon Sook, Kim, (2004), *Equity Prices, Credit Default Swaps, and Bond Spreads in Emerging Markets,* IMF Working Paper WP/04/27.

[3] Cochrane, John, (2005), *Money as stock,* in: *Journal of Monetary Economics*, 52, pp. 501 – 528.

[4] Coudert, Virgine; Gex, Mathieu, (2006), *Can risk aversion indicators anticipate financial crises, Banque de France,* in: *Financial Stability Review,* 9.

[5] Duffie, Darrell., (1999), *Credit Swap Valuation*, in: Financial Analysts Journal, 55, pp. 73 – 87.

[6] Im, Kyung So; Pesaran, Hashem M., (2003), *On the Panel Unit Root Tests Using Nonlinear Instrumental Variables,* Cambridge Working Papers in Economics, 0347.

[7] Kaminsky, Graciela; Reinhart, Carmen M., (1996), *The Twin Crises: The Causes of Banking and Balance of Payments Problems,* in: *The American economic Review*, 89, pp. 473 – 500.

[8] Kaminsky, Graciela; Lizondo, Saul; Reinhart, Carmen M., (1998), *Leading Indicators of Currency Crises,* IMF Working Paper 97/79.

[9] Kumar, Mohan; Moorthy, Uma; Perraudin, William, (2003), *Predicting emerging market currency crashes*, in: *Journal of Empirical Finance*, 10, pp. 427 – 454.

[10] Mishkin, Frederic S.; White, Eugene N., (2002), *U.S. Stock Market Crashes and their Aftermath: Implications for Monetary Policy*, NBER Working Paper 8992.

[11] Packer, Frank; Chamaree, Suthiphongchai, (2003), *Sovereign credit default swaps*, in: *BIS Quarterly Review*, pp. 79 – 88.

[12] Peltonen, Tuomas A., (2006), *Are Emerging Market Currency Crises Predictable,* European Central Bank, Working Paper Series, 571.

[13] Sachs, Jeffrey; Tornell, Aaron; Velasco, Andres, (1996), *Financial Crises in Emerging Markets: The Lessons from 1995*, Brooking Papers on Economic activity 1.

AN OVERVIEW OF NEEDS THEORIES BEHIND CONSUMERISM

David **WARD**, Marta **LASEN**
European School of Economics
daward@tin.it,marty.la@tiscali.it

Abstract:

*Wynn and Coolidge [2004] have hypothesized that one of the key reasons why the Homo Sapiens progressed to being modern man while the Neanderthal man didn't, is that the former developed through innovation (from artefacts to advanced hunting methods) while the latter has left no trace of such evolution. Almost as if the Neanderthal man did not see the need to progress and accepted circumstances as fact. If this is true then the Homo Sapiens have not only developed psychological and objective needs but have progressively updated them as well. Maslow put it beautifully by saying "You will either step forward into growth, or you will step backward into safety". This paper is the first part of a two part series. Here we provide an overview of needs theories and discuss them in the context of consumerism, consumption and opportunities for enterprises. In part two, needs and opportunities are linked to markets, benefits and strategies through a specific 3D model based on Maslow's pyramid. To pave the way for this approach we also promote a model (PIE-Persons, Institutions and Enterprises) with the intent to help enterprises view consumers, institutions and their organisation as one interweaved entity. Needs theories are known to be crucial behind much of the understanding of human behaviour and in particular in the workplace and by the consumer. This paper examines the development of hierarchical needs theory from Maslow to Gough with the intent to better identify consumer needs, provide examples of current and past business opportunities and macroscopically show the progression from red to blue ocean strategies. The authors provide an overview of needs theories seeded through motivational theory also with the aim to uncover the differences in **having** (sometimes known as deficit needs) and **being** needs (sometimes known as growth needs) and then subsequently link them to enterprise strategies, improved consumer understanding and better market exploitation.*

Keywords: Maslow, Herzberg, needs, motivation, having, being, uniformity, diversity.

JEL Classification: A1, M2 and M3,

1. Introduction

The psychological definition of a need is that it is a trait that impels an individual to pursue a goal through an action that also gives purpose, meaning and direction for the behaviour of the individual.

As a human being we may explicitly solicit our needs in two ways, Psychologically or Objectively. In the first case we perceive the need as being an unsatisfied objective need and many scholars such as Maslow (1954) and Alderfer (1972) have investigated the hierarchical structure of psychological needs. In the second case an objective need is seen as specific manifestation of a goal to be attained and has been addressed also by academics such as by Gough (1994) and Doyal (1991).

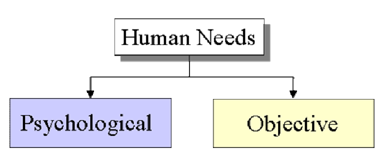


Figure 1. Basic Types of Human Needs

Recently Wynn and Coolidge (2004, 2008) have hypothesized that one of the case reasons why Homo Sapiens progressed to being man while the Neanderthal man didn't, is that the former developed through invention (everything from artefacts to sophisticated hunting methods) while the latter has left no trace of such advancement. Almost as if the Neanderthal man did not see the need to progress and accepted circumstances as fact. If this is true then Homo Sapiens have not only developed both types of basic human needs but have progressively updated them as well. Needs theories are known to be key behind much of the understanding of human behaviour and in particular workplace and consumer behaviours. The prime objective of this paper is to provide an overview of the needs theories so as to link consumer needs to enterprise strategies. We investigate the individual in terms of consumer, consumerism and consumption.

John Locke (1632-1704) explored the concept of property both in terms of human interests and aspirations as well as material goods [cited by Horwitz *et al.*, (1990)]. Locke also defined the *self* (what we call the individual today) which he considered to be a *conscious thinking thing* composed of the spiritual or material in a *simple or compounded* manner.

The apex of needs theories development occurred between the 1930s and mid. 70s, indeed it is here that we see groundbreaking work from Allport (1933), Maslow (1943, 1954), Lowry and Maslow (1979), Alderfer (1972), James (1890), Herzberg (1959) etc. who investigated the motivation of individuals in organizations and it is for this reason that we speak of motivational theory[1]. Maslow in particular developed his 5 stage (or levels) hierarchy needs model which has become so popular that even after his death in 1971 he is attributed to the 7 stage and 8 stage versions of the same model but developed by others [Lowry, (1971)].

The outcome of this 40 year period led to many new or modified needs models that have been applied also in other fields including marketing, consumer behaviour [Kotler and Keller, (2006)], human resources [Cullen, (2001)], management [Kiel, (1999), Huitt, (2001)], information [Norwood, (1999)], teaching and learning [Eccles, (2002)], athletes and sports [Fortier *et al.*, (2007)] etc. Since then interest has slowly switched to either investigating the exploitation of needs theories or explaining how individuals define needs and subsequently transform them into goals, behaviour and decisions.

Over the last century, and especially in the last 4 to 6 decades, we have also seen the evolution of the consumer and the marketplace, going from consumers of the massified product to one of hyper-personalised products and services. We have also witnessed the stratification of product offerings from horizontally to vertically differentiated products culminating in what is currently termed *inaccessible luxury* [Ward and Chiari, (2008)]. In this context it is not surprising that academics, enterprises and authorities have all focused their attention on understanding, leveraging and regulating such knowledge [Gough, (1994)]. This paper sets out to provide an overview of needs theories in the context of consumerism and paves way for subsequent enterprise strategies.

2.The Evolution of Needs, Motivation and Maslow's Model

As a human being we have always searched for the satisfaction of our basic needs, namely survival needs including food, water and shelter [Wynn and Coolidge, (2004), (2008)]. These needs are inborn and warranted by our body to sustain life rather than just improve it. One could address these needs at a primitive level since they do not actually form a construct dependent on our peers or external social factors i.e. they are both cognitive and physiological in nature. Indeed recent researches into mathematics and cognitive psychology [Lakoff and Núñez, (2002)] have shown that newly borns are capable of 'counting' up to 3 objects even without knowledge of basic arithmetic. In a similar way our body is capable of recognizing needs for warmth and nutrition.

In his discussion on consciousness, William James anticipated the model later proposed by Maslow, describing the different constituents of the empirical self on *a hierarchical scale, with the bodily 'me' at the bottom, the spiritual 'me' at the top, and the extra-corporeal material selves and the various social selves between… according to their worth* (1890)]. While the body represents the heart of the material self, clothes, family, home and possessions come immediately after as parts of human beings' life. James associates instead the social self with the recognition gained by other members of society and fellows, so that *a man has as many social selves as there are individuals who recognize him and carry an image of him in their mind* [James, (1890)]. The spiritual self finally represents the central source of effort and interest: it actually comprises individuals' subjective and inner being together with their psychic faculties, intended as sensory perceptions and emotions as well

[1] There are two types of motivational theories, namely Extrinsic and Intrinsic. In the former case motivation is stimulated by external factors such as rewards and recognition while in the latter motivation comes from within the individual, such as curiosity, challenge, self-determination.

as motor ideas and concepts [Johnson and Henley, (1990)]. Examining the different selves, James also described the existence of instinctive impulses acting as functional means to attain personal goals that fall within bodily, social or spiritual self-seeking and progress endeavours. Due to the conflictive relationship present among the different selves, it is however, evident the inherent impossibility to concurrently augment all the aspects: based on the aforesaid hierarchical view, the preservation and growth of the spiritual self clearly acquire the highest priority, after a thoughtful evaluation of *the immediate and actual, and the remote and potential* [1890].

Table 1. James' Empirical Life of Self

	Material Self	Social Self	Spiritual Self
Self-Seeking	bodily appetites and instincts, love of adornment, acquisitiveness, love of home, constructiveness etc.	desire to please, envy, be noticed, admired, love, sociability, emulation, pursuit of honour, ambition etc.	intellectual, moral and religious aspiration, conscientiousness etc.
Self-Estimation	personal vanity, modesty, pride of wealth, fear of poverty etc.	social and family pride, vainglory, snobbery, humility, shame, etc.	sense of moral or mental superiority, purity, sense of inferiority or of guilt

In 1937 Allport published a paper on dynamic psychology and discusses the functional autonomy of motives. He states that *adult motives as infinitely varied, and as self-sustaining, contemporary systems, growing out of antecedent systems, but functionally independent of them. Just as a child gradually repudiates his dependence on his parents, develops a will of his own, becomes self-active and self-determining, and outlives his parents, so it is with motives. Each motive has a definite point of origin which may possibly lie in instincts, or, more likely, in the organic tensions of infancy. Chronologically speaking, all adult purposes can be traced back to these seed-forms in infancy, but as the individual matures the tie is broken. Whatever bond remains, is historical, not functional*. Allport's analysis is remarkable because it unfolds another aspect behind needs and that is the development of the individual and the context within which such development occurs.

During the same period Abraham Maslow was studying rhesus monkeys and motivational research at the University of Wisconsin for which he was awarded a PhD in psychology in 1934. His work was so grounding breaking that it is one of the very few universally accepted works on motivation and needs that carries the authors name and for which a dedicated branch is provided (see next figure) in motivation theory.

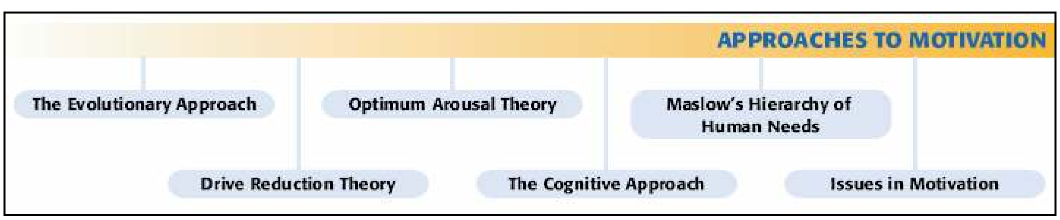


Figure 2. Approaches to Motivation

Source: http://highered.mcgraw-hill.com/sites/dl/free/0072937769/128305/Santrock_PU7e_ch11.pdf, pp. 425

He later he moved to New York's Brooklyn College where he developed his epic model now known as Maslow's five stage needs model and shown next.

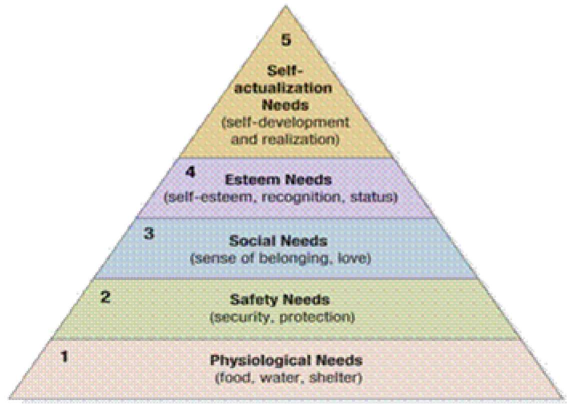


Figure 3 . Maslow's hierarchy of 5 needs

Source: Kotler and Keller, (2006). *Marketing management (12th edition),* New Jersey, Pearson Prentice Hall, pp. 185)

Maslow considered survival needs as being the first and foremost of all needs since without their complete satisfaction an individual would not progress to other types of needs. He classified these needs as Physiological Needs. Other academics dispute this concept of the development of needs since it has been demonstrated that individuals can simultaneously desire and search for more than one need at a time and may well prioritise them differently and within each stage. In fact the first stage was later restated as Biological and Physiological Needs [Gorman, (2004)] that include air, water, food, shelter, warmth, sex and sleep.

According to Maslow's original thoughts a person will strive for further and higher positioned or ranked needs only when the lower level needs (also known as "deficiency needs") have been consistently and permanently met. In doing do Maslow was stating that this striving is essentially the engine behind a individuals' personal growth and is both a measure of motivation as well as dissatisfaction/satisfaction. Maslow devised his 5 level Hierarchy of Needs motivational model while investigating motivation in workforces in organizations [Maslow, (1954)]. His pioneering work was a real eye-opener because organizations were suddenly confronted with the realisation that workforces required managing not controlling and this was only achievable if the needs of individuals were fully understood and developed. The outcome of his work is described in five stages or levels housed in a stratified pyramid, commonly known as Maslow's pyramid of needs (see figure 3). Each level or stage is summarised as follows, starting from the most basic stage (1) and ending with a stage (No.5) known as self-actualisation:

1. Physiological needs such as food, drink and shelter.
2. Safety needs such as security and protection.
3. Social needs such as belongingness and love.
4. Esteem needs such as status, self-esteem and recognition.
5. Self-actualisation needs such as self-development and realization.

We consider the 5 needs model to be split into two, unequal parts, namely the first 4 stages or levels concern the 'having' or materialistic while the last stage is termed "being' since it is non-materialistic and holistic in nature, something we consider here as 4:1 (4 having and 1 being).

Maslow thought of his model as a naturally progressive method of prioritising human needs later it became clear other fields of research could use this approach to better address consumer needs. The affection for Maslow by marketers is quickly explained since it offers a deterministic approach in an otherwise non-deterministic marketplace. In other words consumer needs are analysed, categorised and prioritised in a prescriptive[2] manner [Mintzberg *et al.*, (1998)] since the same individuals are also

[2] By prescriptive one intends that the necessary initiatives taken by enterprises and leaders to satisfy consumers hence it follows a fact-based management approach.

consumers. Indeed the marketplace, just like the workplace, is simply an environment in which the individual is called upon to play his or her role in sustaining and/or expressing well-being and wealth. Hence in the case of the marketplace the individual is the consumer while in the workplace he or she is the worker. In the context of consumerism the scope of the workplace is to provide the worker with the necessary economic status to play his or her role as a consumer. If the worker does realise his or her needs then this is directly reflected for the same person as a consumer, something that we could consider as a "knock-on" effect, or cause and effect.

A recent European report indicates (www.fao.org), that with the increase of personal incomes, individuals tend to spend a higher proportion of their income on activities that are higher placed in Maslow's hierarchy of needs, and less on satisfying their basic needs. That is why, for example, in most developed countries, only a small proportion of personal income is currently spent on food, while in developing countries this can account for a significant proportion, or even the majority, of personal expenditure. This link between personal incomes and consumerism not only rebalances and reorders consumer needs but also shows that the needs in each stage are also stratified further. In fact in the following figure (consumption versus pro capita GDP) we can observe, for example, that food and shelter are distinct just as knowledge and awareness. Also the same graph shows that the lower the income, the bigger the percentage spent to satisfy primal needs.

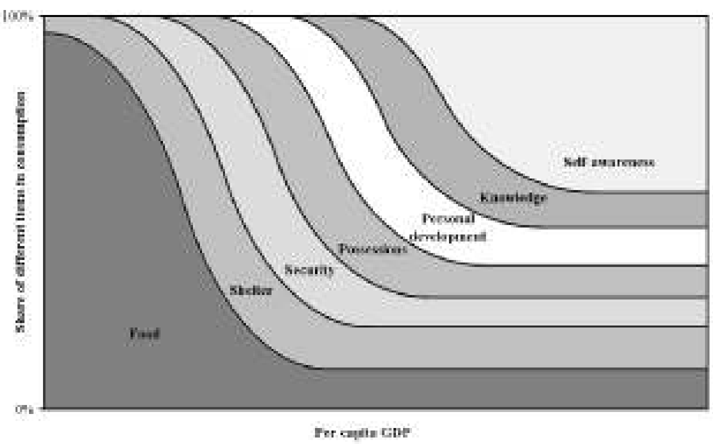


Figure 4. Relationship between needs and income

Source: www.fao.org

In the 70s Maslow is attributed to extending his needs model [Lowry, (1979)] and is often referred to as the *7 needs* model since it includes two new categorised needs namely *cognitive* and *aesthetical* needs. A further distinction in the new model is that needs are more well defined, for example, the first level identifies biological and physiological needs while safety needs are not just about personal safety but also include law and order. The emphasis on the specification of needs shows an increase in societal awareness about human needs and in particular consumer and people needs. It could also be a clear indication of the fragmentation[3] of the consumer and the rise of a consumer that is still focused in the 'having' that pertain to the first 4 levels of Maslow's pyramid) but is leaning progressively more towards the 'being' i.e. towards self-actualisation. However, doubts that the 5 needs model was incomplete [Simons *et al.*, (1987)] eventually led to understanding that between esteem needs and self-actualization there were other needs. In this paper we voice this doubt even more since the move from 'having' to 'being' needs require some superimposition and a mix of more interior needs. Another claim we make is that consumers not only have the needs claimed by for example, by Maslow but also the need to realise their needs as fast as possible (think of fame in the context of status) and also the need to dedicate as less effort as possible to satisfy or gurantee the

[3] By fragmentation we imply the birth and development of niche consumer categories or new segmentation.

need. We believe that there are other aspects that prioritise needs but this will be discussed in part two.

As said in the 7 needs model two new stages are envisaged namely Cognitive needs and Aesthetic needs. Both these needs act as a bridge between the having and being parts of Maslow's pyramid. In the modern mindset cognitive needs are expressed through a desire to understand and find meaning, thus knowledge (or lack of it) appears to be the main driver behind this need. Similarly aesthetic needs are spurred by the search for beauty, balance, form etc. and thus imply a desire to appreciate detail and self-expression. In this sense the two new needs overlap and superimpose with self-actualization. Lowry describes the needs of the 7 needs model as follows:

1. Biological and Physiological needs – air, food, drink, shelter, warmth, sex, sleep, etc.
2. Safety needs – protection from elements, security, order, law, limits, stability, etc.
3. Belongingness and Love needs – work group, family, affection, relationships, etc.
4. Esteem needs – self-esteem, achievement, mastery, independence, status, dominance, prestige, managerial responsibility, etc.
5. Cognitive needs – need for knowledge, meaning, etc.
6. Aesthetic needs – appreciation and search for beauty, balance, form, etc.
7. Self-Actualization needs – realising personal potential, self-fulfilment, seeking personal growth and peak experiences.

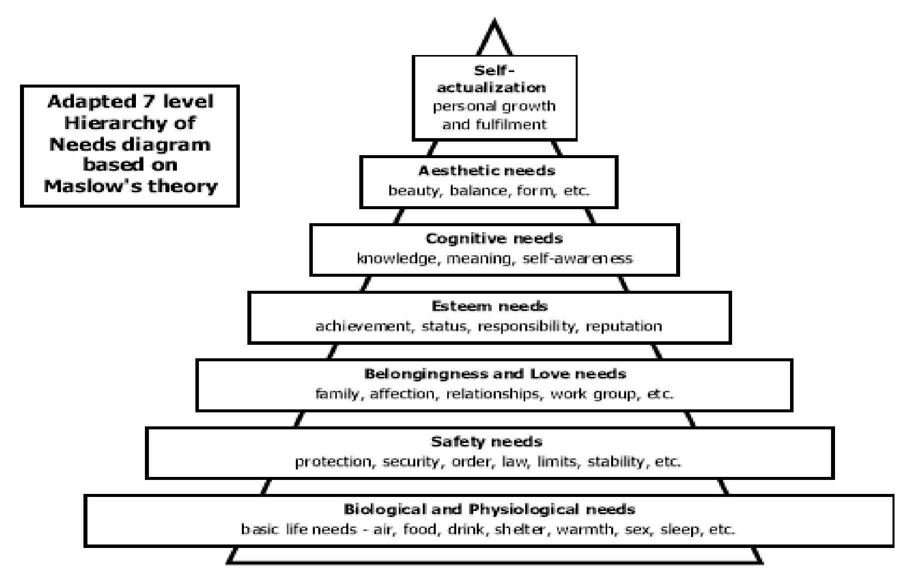


Figure 5. Maslow's hierarchy of 7 needs

Source: http://www.businessballs.com/maslowhierarchyofneeds7.pdf

It is worth noting that Maslow recognised that his initial idea that there was only one growth need, namely self-actualization, was weak and required further differentiation, he voiced this but never actually included these two new lower-level growth needs [Maslow & Lowery, (1998)]. By doing so he introduced a ladder approach to self-actualization and accepted that self-actualised individuals possess at least four characteristics: 1) Are problem or goal focused; 2) Have boundless optimism about life; 3) Believe strongly in personal growth; and 4) Have an expected attitude towards peak performance and experience such as top athletes. Indeed it could be concluded that self-actualisation manifests itself in further needs. We consider the 7 needs model to be split again into two, unequal parts, namely the first 4 stages or levels concern the 'having' or materialistic while the three upper stages reinforce the concept of 'being' i.e. they are non-materialistic and holistic in nature. In this continuum we consider the 7 needs model to be 4:3 (4 having and 3 being).

In the 8 needs model [Maslow, (1971)] a new level follows that of self-actualization in which people push self-actualization for others. Maslow imagined this need as a **transcendence** need i.e. a stage that represents the realisation of the being for the goodness of others as well as oneself e.g. scout leaders, preachers etc. In this new need people progress further because they help others to self-actualise via the realisation of their potential. By combining self-actualization and transcendence people put their knowledge to good use, by developing and disseminating their wisdom[4].

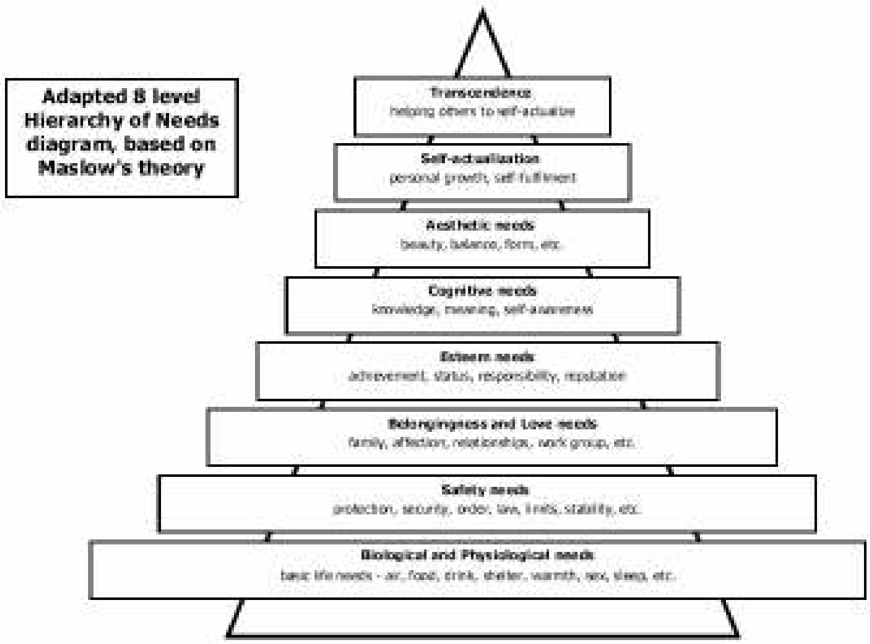


Figure 6. Maslow's hierarchy of 8 needs

Source: http://www.businessballs.com/maslowhierarchyofneeds8.pdf

According to Austrian neurologist Viktor Frankl (1959) self-actualization represents only a by-product of self-transcendence, as the true meaning of life is to be found in the world rather than within human beings. Based on this interpretation, rather than identify self-actualization and self-transcendence needs as distinct and hierarchically organized, Michael Daniels (2001) suggested that self-actualization is the same as or, at least implies, self-transcendence. Indeed if we take teaching as an example it is difficult to separate the motive to teach as self-actualization or self-transcendence or both. However, it would seem logical and sequential to think that before teaching others how to teach one needs to learn how to teach.

We consider the 8 needs model to be a 4:4 model (4 having and 4 being) and thus suggest that as human needs have progressed we see a shift, and much greater emphasis, on the 'being', which we consider to be a distinctive feature of the post-industrial and modern consumer.

So far we have seen a search to provide the most detailed, yet summarized account, of human needs. Revich (2005) takes a different approach and, starting from Maslow's 5 needs pyramid, reverts to a more simpler and universal model. This model, known as the Three Fundamental Needs (or TFN) model divides the pyramid into three areas. Revich argues that needs are time dependent (i.e. they depend on circumstance and the stage of life of the individual) and therefore their relationship and importance will change. What is interesting about this model for consumerism is that hierarchy is

[4] Daniels (2001) suggests that Maslow's ultimate conclusion that the highest levels of self-actualization are transcendent in their nature may be one of his most important contributions to the study of human behaviour and motivation.

temporary and that consumers will expand or shrink each one of the three circles of needs to match the moment, i.e. a sort of temporary evolution.

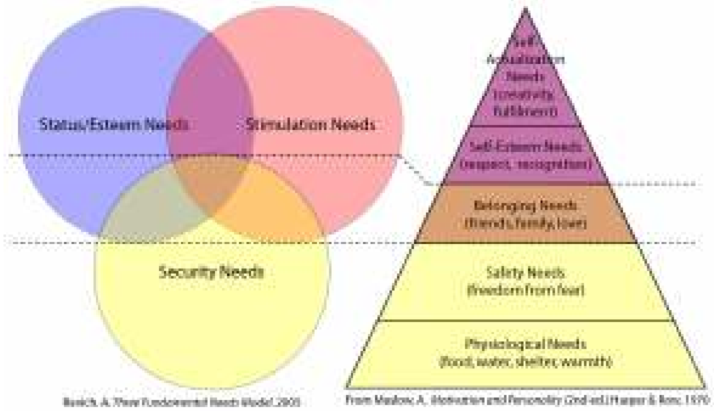


Figure 7. Revich's Three Fundamental Needs model

Source: www.workitout.ca

Maslow has spurred a whole collection of competing, dedicated or complimentary models. William Glasser (1998), for example, promotes a five needs model based on: 1. Survival, 2.Love and belonging, 3.Power and Recognition, 4. Freedom and 5. Fun. His theory concerns that of control or choice, in fact his model is also known as "Control Theory" or "Choice Theory", and is perhaps more in-line with modern society and consumerism than Maslow's model. Glasser's work has been widely acclaimed by teachers and in the classroom. He stresses that we are *genetically programmed* to satisfy our psychological needs (pp. 28).

Dean Spitzer (1996) considers motivation in terms of desires and promotes an Eight Desires model. In it he describes human desires in a non-hierarchical fashion that include; Power, Activity, Recognition, Affiliation, Competence, Ownership, Meaning and Achievement, thus making a direct connection with the work of McClelland and Maslow.

Interestingly needs theories consider the human being as an individual that has varying needs and priorities throughout life yet none of them actually focus on the differences between adults and children. Susan Price (1994) has identified the Eight Needs of children starting from physical protection such as safety and the need to be protected against physical harm. She goes on to discuss Physiological, Psychological, Social, Emotional, Intellectual, Educational and Spiritual needs. What is especially interesting is that here we see the importance of context, the environment and development of the individual and the emphasis on "special needs".

So far we have discussed needs with minimal linkage to structured consumer needs or requirements. In this context it is worth mentioning the work of Professor Noriaki Kano of Tokyo Rika University who elaborated a model to assist companies in the analysis and understanding of consumer needs. Like Herzberg's hygiene factors to be discussed later, Kano's basic needs represent the fundamental features buyers normally expect from a certain product or service. As they are taken for granted, they appear as unspoken but if left unmet, they become sources of profound customer dissatisfaction. On the contrary, consumers tend to explicitly express and seek performance needs, which are generally addressed by standard attributes that either increase or reduce satisfaction depending on their actual functioning. While focusing on these two categories allow organizations to enter and remain in the market respectively, according to Kano real excellence can be attained only through customer delight fostered by the fulfillment of excitement needs (cited by Stroud). This implies that the emphasis is on discovering, satisfying and anticipating latent consumer needs rather than declared needs, examples such as the 'self-parking' car, the remote control, Blackberry, iPod etc. are just a few products that follow this rule. Furthermore we consider Kano's model to promote blue ocean strategy rather than red ocean strategy [Kim and Mauborgne, (2005)]. As Kano's model suggests this is achieved by addressing three specific requirements, ranging from dissatisfiers to delighters:

- **Satisfying basic needs:** Allows a company to enter the market and thus overcome incumbency by competing with enterprises already there. This is typical of red ocean strategy.
- **Satisfying performance needs:** Allows a company to remain in the market and continue to compete with existing competing enterprises.
- **Satisfying excitement needs:** Allows a company to excel, to be world class, best-in-class and thus dominate the marketplace. This is especially true for blue ocean strategy.

Dissatisfiers or Basic Needs – Expected features or characteristics of a product or service that are not declared but expected e.g. hotel room cleanliness, reliability, empathy in case of complaint etc. If this basic need is not met consumers will be extremely dissatisfied. When consumers are dissatisfied they will voice that negativity strongly while on the contrary consumers say little at all.

Satisfiers or Performance Needs – Typically this is where the product or service offering is correctly positioned and delivered e.g. quality versus price, user friendliness, speed, professionalism. Again if this need is satisfied not necessarily is it expressed explicitly although it does provide a longer sensation of satisfaction and is a key motivator for further consumer purchases.

Delighters or Excitement Needs – Unexpected features or characteristics of product or service offering that distance the enterprise from the competition but also the consumer from other consumers. These needs, if met, are very strongly voiced when communicated to thirds and extremely appreciated by consumers. However, they are typically latent [Nishino *et al*, (2008)] and remain unspoken. They can be "trivial", such as anticipating a customers preferred dish, wine, table etc. at a restaurant or unsolicited personalization, or whopping such as replacing broken products free of charge even through it was the consumers fault.

Although these needs categories appear unquantifiable Nishino *et al.* proposes an analytical model based on the Japanese Kansei approach[5] in which consumer needs are collected, analysed and connected to correct product development. Kano's model, which is depicted next, shows that satisfaction/dissatisfaction and performance can be quantified and hence scaled:

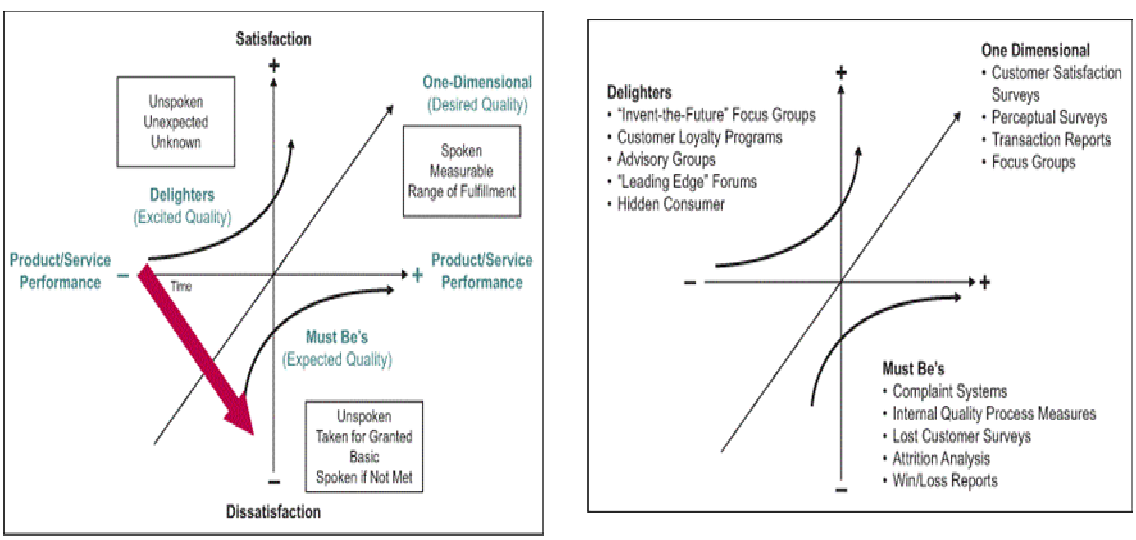


Figure 8. Kano's model of Satisfaction-Dissatisfaction versus product/service performance

This graph summarises the viewpoint of the customer in terms of the quality and performance of product or service. It visualises spoken and unspoken information from the customer.

[5] Kansei is a universally accepted and structured way of satisfying consumer needs during product development.

In a world dominated by hi-tech and informatics it is not surprising that mans' needs have also brought about a need for information. Furthermore, in the 7 needs model the cognitive need is also a need for knowledge, to know and to be informed. Norwood (1999) proposes an information needs pyramid similar to Maslow's hierarchy needs model and is proposed to describe the kinds of information that individuals seek at different levels. At the most basic and lowest level individuals needs are addressed by information that copes with situations, otherwise known as **coping information**. This information is primarily about satisfying an immediate need and thus has a very short-term time frame of application, virtually call-on-demand.

The next level of information addresses safety, defined as **helping information** by Norwood. This information satisfies a need for safety and security. Although this information can also be judged short-term it requires continuous refreshing and confirmation so as to provide reassurance. An interesting and growing requirement for lower levels of information concerns that of e-Government and e-Governance. In both contexts the objective is to bring the management and/or market-decision mechanisms into the realm of traditional public administration and the public domain, for example by getting public services online and sharing information that involve and condition the public.

Individuals will also search for *belongingness* just as Maslow suggested with love needs and social acceptance. Norwood classifies this information as **enlightening information** and represents an individuals need to seek the development of relationships (think of social networking). At the fourth level esteem needs are represented by **empowering information** and concerns self-worth and awareness. Finally, people in the growth levels of cognitive, aesthetic, and self-actualization, as seen in the 7 needs model, search for **edifying information**. Since Norwood stops here it is quite plausible that transcendence needs seen in the 8 needs model equally require information needs. In this paper we address this type of information as spiritual, mystical or **transcendence information**. This information is not just about the transcendence of the individual but also about how an individual can help others to transcend. Indeed e-government is a the response of the community to involve people in a more open and accessible way. Norwood's 5 level model is shown next and is represented as a pyramid to coincide with Maslow's original needs model.

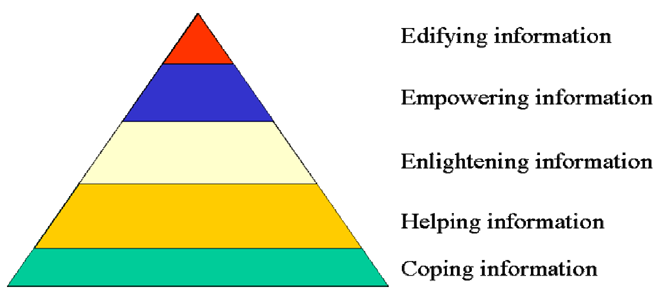


Figure 9. Norwood's information needs model

The second view of needs is represented by the work by Ian Gough, a political economist, who provides an individual's needs in the context of social assistance and social welfare. Similarly Len Doyal, a medical ethics academic, together with Gough provide an **objective view** of the human need in their publication *The Theory of Human Need* (1991). Their perspective ties needs to the participation of the individual in society, especially in terms of physical health and personal autonomy.

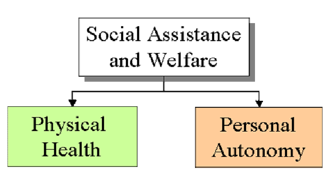


Figure 10. Gough and Doyals' needs theory model

Physical health is a rather elaborate concept since it involves nutrition, care, the environment etc. while personal autonomy is about the capacity of an individual in making informed and contextualised decisions about what should be done (to satisfy the need) and how to go about implementing such actions. Furthermore since we are speaking about individuals needs this is strictly a bottom-up approach and not vice versa.

Doyal and Gough suggest eleven broad categories of "intermediate needs" that define how the need for physical health and personal autonomy are fulfilled:

1. Adequate nutritional food and water.
2. Adequate protective housing.
3. A safe environment for working.
4. A safe physical environment.
5. Appropriate health care.
6. Security in childhood.
7. Significant primary relationships with others.
8. Physical security.
9. Economic security.
10. Safe birth control and child-bearing.
11. Appropriate basic and cross-cultural education.

Personal circumstances however, are key to the satisfaction of such needs, for example, a person with adequate physical capacity and education will find these needs much easier to attain and sustain. Consequently old age people, the handicapped, those with learning difficulties etc. will have less chance because of a deficiency in capabilities. In fact academics agree that Doyal and Gough's theory should be associated to the capability approach developed by Amartya Sen and Martha Nussbaum (1993) that discusses "substantial freedoms" including the ability to live to an old age, engage in economic transactions, participate in political activities etc. A further consideration is that the capability of an individual and relative needs will determine just how far he or she is included in society[6]. Hence Doyal and Gough also provide insight into the degree of inclusion or exclusion in society for individuals and, with the help of the capability approach, indirectly explain poverty measures from their needs theory. Ward and Farmaki (2006) link needs theory to capabilities by suggesting that socioeconomic classification of the individual in society inevitably leads to both social stratification and hierarchy based on three social exclusion groups namely Macro-Economic, Personal-Cultural and Social-Cultural. Culture therefore plays a key role in motivation [Mühlbacher *et al.*, (2006) and Hofstede, (2001)]. The very fact that differences occur affords an opportunity to differentiate product and service offerings that in turn call for different strategies. Porter suggests three strategies, namely Cost Leadership, Product Differentiation and Focus. Although this is a rather 'old' school of thought it does emphasis the need to see a business opportunity (see table 2) from more than one angle, moreover, it helps us to view product demand from a "willing to pay" price. Grant (2002) defines this price as *The price at which a product can sell in the market is the aggregate of the values derived from each of these individual attribute's*, which provides us with a tool known as Hedonic price analysis[7].

Another discussion point about needs theories is that concerning hierarchy. Needs theories have always been structured but their ordering has left academics in two clusters of thought i.e. those that favour a hierarchical sequential approach e.g. Maslow, and those who prefer a more hierarchical circumstantial approach e.g. Alderfer. Academics have not really truly discussed in detail eventual differences between these two schools of thought but James (1962) and Mathes (1981) view hierarchy as a convenient tactic. James hypothesized three levels titled *material* (physiological, safety), *social* (belongingness, esteem) and *spiritual* while Mathes equally proposed three levels of needs

[6] For example, poverty is understood as capability-deprivation due to ignorance, government oppression, lack of economic resources etc.

[7] Hedonic price analysis assesses price differences for competing products, relates these differences to several combinations of product attributes, and computes the implicit market price for each attribute [Grant, (2002)]

denominated physiological, belongingness, and self-actualization. Interestingly Mathes considered security and self-esteem as unwarranted. What appears to be common ground is that needs theories start from physiological needs and only after do we see discrepancies. Such discrepancies are certainly socio-cultural in nature [Ward and Farmaki, (2006)] but as globalization progresses higher level needs are being aligned. For enterprises this means [Morace, (2008)] global consumers and consumerism is not fiction but fact.

Alderfer (1972)] developed a comparable hierarchy with his ERG (existence, relatedness, and growth) theory. His approach modified Maslow's theory based on the work of Gordon Allport (1960, 1961) who incorporated concepts from systems theory into his work on personality. Allport believed that an individual's philosophy is founded upon their values, or basic convictions, that what the person holds about what is and is not of real importance in (their) life. From this assumption he outlined six major value-types and similar to Maslow's needs pyramid, as follows:

1. The Theoretical person is primarily concerned with the discovery of truth, to which they seek in a cognitive way.

2. The Economic individual places highest value on what is the most useful.

3. The Aesthetic person places high value of form and harmony. They believe life to be a series of events that are to be enjoyed for its own sake.

4. The Social type seeks out the love of people.

5. The Political person's dominant drive is power.

6. The Religious individual places highest value on unity. They seek to understand and experience the world as a unified whole.

The social theory of Pierre Bourdieu (1998) is offered as an alternative to the Maslow approach, providing the basis for a social critique of consumerism and an alternative evolutionary theory of consumption. In this approach, the structure of the social hierarchy both constrains the consumption of lower social strata and leads to subtle, less conspicuous consumption patterns at the top of the social hierarchy: a scenario that could provide a social foundation to the Engel curve. Engel's curve and subsequent law was formulated nearly 150 years ago and states that with rising incomes, the share of expenditures for food products (and, by extension, other things as well) declines. This law was brought forward at a time when agriculture was slowly being taken over by industry, hence the curiosity in investigating income effects on food. Engel showed that as a country develops economically, the relative importance of agriculture declines. He suggested that the primary reason for this behaviour was that as incomes increase the proportion of income spent on food declines and money is spent on higher needs such as esteem, self-actualization. Engel also found, based on surveys of family budgets and expenditure patterns, that the income elasticity of demand for food was relatively low[8]. We note that Engel's Law does NOT advocate that the consumption of food products remains unchanged as income increases!, rather it suggests that consumers increase their expenditures for foodstuffs (in % terms) less than their increases in income.

Another view of motivation theories is to consider them as part of a process in which we furnish a certain effort to achieve a certain outcome under a degree of expectancy. Vroom's expectancy theory sees motivation (in the workplace) not in terms of needs (unlike Maslow and Herzberg) rather he focuses on outcomes.

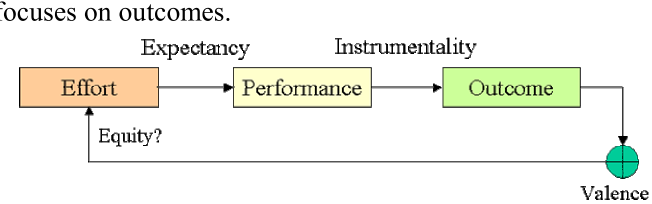


Figure 11. Expectancy theory by Vroom

Vroom, hypothesises that in order for a person to be motivated that effort, performance and motivation must be linked. Along this process the need will be subjected to three variables, namely,

[8] The resulting shift in expenditures affects demand patterns and employment structures.

Valence, Expectancy and Instrumentality. Expectancy is the belief that increased effort will lead to increased performance i.e. if I work harder then this will be better, while instrumentality is the belief that if you perform well that a valued outcome will be received. Valence is the importance that the individual places upon the expected outcome. Although Vroom, like Maslow and Herzberg, focuses on the workplace his model lends itself to additional interpretation of human needs, for example, people may place more emphasis on status and at the expense of belongingness and security even it is only temporary, think of young people participating in the casting of programs like The Big Brother. Indeed Vroom's hypothesis predicts that the effort of the individual will depend on the value the person places on the outcome. In terms of consumers this means that time and the amount of effort needed to satisfy the need will lead to different degrees of satisfaction or dissatisfaction.

The Adam's Equity Theory (1963) is another work motivation theory that asserts employees become de-motivated when the balance between effort, or input, does not match the output. In a similar fashion a consumer will search for a correct and fair balance between an input and outcome and consider their relationship fundamental in deciding if a need has been satisfied. CRM leverages and manages precisely this relationship between input and output.

As we have seen objectives and therefore goals can be key motivators, Latham and Locke (2002) speak of *goals and subconscious motivation* and summarise 35 years of goal-setting theory and motivation in the workplace. But they compare goal-setting theory (that concerns the workplace) to social-cognitive theory (the individual), a sort of bridge between the organization and the person. In their analysis we feel that they raise indirectly (and perhaps accidentally) a question of multiple goal setting (for one or more individuals) and how these breed in the context of team-playing. Could this indicate a flaw in needs theories i.e. individual needs do not stand-alone but survive and thrive also thanks to others?

3.From Needs to Opportunities

One of the most celebrated phrases of Philip Kotler is *It is no longer enough to satisfy customers but you must delight them* and he was one of the first marketing gurus to recognise that firms need to meet customer needs, retain customer loyalty and innovate to keep in step with changing needs.

Kotler (2003) argues that a person's buying choices are influenced by four psychological factors; motivation, perception, learning, and beliefs and attitudes (pp.195 – 198). he goes on to highlight that in terms of motivation celebrities such as Freud, Maslow and Herzberg carry different implications for consumer analysis and subsequent [marketing] strategy. For example, Freud claimed that people's behaviour were largely unconscious and a technique known as *laddering* can be used to trace a person's motivation. The link between motivation and needs theories is that a **motive** is a **need** that is sufficiently pressing to drive the person to act [Kotler, pp. 195, (2003)]. Hence in the buying process we have different levels of attention and drive to consume. In a similar manner we may consider the evolution of the consumer's attitude to the product during its life cycle (PLC). Wasson [cited by Kotler, pp.340, (2003)] indeed provides a link between PLC, marketing objectives and strategies, but not strictly between consumer needs and enterprise strategies.

However, the discovery (or rediscovery) of needs and development of needs theories inevitably leads to several opportunities for their application and exploitation. This implies that once a company detects a consumer need a suitable strategy is required in order to satisfy the consumer and exploit or build the business opportunity. If the opportunity turns out to be a question of product range extension, additional features, design update etc. then we consider this to be a (declared) need that is best tackled with read ocean strategy, which is typical of the prescriptive school of thought for strategy [Mintzberg *et al.* (1998)] and most likely sit in the lower levels of Maslow's pyramid. On the other hand if the opportunity satisfies a latent need then we may well have the opportunity to charter into a blue ocean i.e. use a descriptive school of thought of strategy. We suspect the blue ocean type opportunities to sit at the top of pyramid i.e. towards or in the domain of the *being*. As a counter argument though we must not forgot that the majority of the human race have still to satisfy their basic needs, so economically speaking, and as suggested recently by Prahalad (2005), the gold may lie at the bottom of pyramid[9] and not vice versa. Another useful consideration of how difficult or non-obvious it is to relate needs to opportunities can be smoking in juveniles. The need here could be interpreted in several ways e.g. belongingness through emulation (to feel like an adult) or challenge your parents and/or the establishment.

In order to assess opportunities we promote the idea that one may view these opportunities from three different perspectives; Person, Institution and Enterprise (PIE), defined as follows:

1. For the individual (person) and/or community.

- The realization of what is missing and what requires developing/satisfying
- The prioritisation of needs i.e. understanding what is more or less important and when they should be attained.
- The necessary support to favour such awareness.

2. For the institutions

- The realization of what is missing and what needs to be developed so as to satisfy both individuals and communities.
- The prioritisation of needs i.e. understanding what is more or less important and when they should be attained for the community and the good of each individual.
- Provides the necessary infrastructure, financial, socio-cultural and legislative support.

3. For enterprises

- Uncover and pinpoint both individual and community needs so as to reflect and respect local customs in a responsible and sustainable manner.
- Provide the products and services that consumers truly need and search for (hence with more emphasis on latent rather than declared needs) while respecting the environment.
- Work together with the consumers and institutions to ethically sustain the demand for goods and services.
- Provide support to individuals, communities and institutions e.g. CSR, Sponsorship etc.

The model is depicted below:

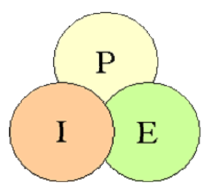


Figure 12. PIE Needs-Opportunities model

The scope of the model is to link needs to opportunities so as to develop adequate strategies for their definition and attainment. Later (in part B) we will see how we can complete the process by developing a 3D square pyramid for which, for example, each of the four sides represent needs, markets, strategies and benefits. Seen purely from an enterprise point of view this implies ensuring that internal and external contexts match, which in turn entails exploiting many of the typical strategic management tools e.g. ARC and PIE [Saloner *et al.*, (2001)], Ansoff matrix [Di Michael, (2003)], Porters 5 forces model [Porter, (1985)] etc.

To exemplify how needs can generate opportunities for enterprises the following table lists a series of consumer products and/or services (some of which have generated new industries) and strategies [Porter, (1980)]. The table is based on Maslow's derived 8 needs model and Porter's concept of strategies for enterprises and industries (1985). In the far right column we see a view of the spectrum of strategy, which goes from Red Ocean to blue ocean strategies as we progress through Maslow's pyramid of needs. Conceptually what we promote is that at the bottom of the pyramid consumer needs are declared and enterprises and industries have already tried to satisfy them (and continue to do so with success). What changes are how these needs are satisfied and the fact that enterprises need to fight-off increasingly more fierce competition. As we progress towards the top of

[9] Over 4 billion people are living on less than 2 US£ per day (www.12manage.com)

the pyramid we move into blue ocean strategy where we invent the future and anticipate the needs of the consumer and satisfy (more) the latent needs.

Table 2. Needs and Opportunities

Needs	Opportunity	Strategy
Biological and Physiological needs such as air, food, drink, shelter, warmth, sex, sleep, etc.	Food and drink dispensers, comfortable housing, central heating, beds, contraceptives, air purifiers, preventive medicine, fitness and fitness centres, co-branding etc.	Cost Leadership Product differentiation Focus
Safety needs such protection from elements, security, order, law, limits, stability, etc.	House alarms, personal anti-attack alarms, private policing, fixed financial income (e.g. fixed salary, pension etc.), life insurance, windows and doors, body guards, web-banks, savings accounts etc.	Cost Leadership Product differentiation Focus
Belongingness and Love needs such as work groups, family, affection, relationships etc.	Social networks, clubs, institutions, family, marriage and/or engagement ceremonies such as weddings and parties, religion, dating and match-making services, chat-lines etc.	Product differentiation Focus
Esteem needs such as self-esteem, achievement, mastery, independence, status, dominance, prestige, managerial responsibility etc.	Career counselling, personal development and training, cosmetics, fashion, fast cars, luxury products, home improvements, furniture, fashion clothes, drinks, lifestyle products and services	Product differentiation Focus
Cognitive needs such as knowledge, meaning, etc.	Wikipedia, further education, cookery classes, language and cultural classes, self-awareness programs, interest towards the arts and humanities, social forums	Product differentiation Focus
Aesthetic needs such as appreciation and search for beauty, balance, form, etc.	Fitness, yoga, facial and body surgery, beauty farms, tanning centres, personal trainers, indoor fitness equipment, personalised diets, cosmetics	Product differentiation
Self-Actualization needs such as realising personal potential, self-fulfilment, seeking personal growth and peak experiences.	Teaching, personal development, vocational activities, personal trainers, thrill-seeking such as bungee jumping and sky-diving, career changing	Product differentiation
Transcendence needs such as helping others to self-actualise	Teaching, family consultancy, good Samaritans, home-helps, humanitarian assistance, free on-line medical advice	Product differentiation

4. Conclusions

The 5 needs model is still very much dominant in needs theories, thus testifying this epic piece of work by Maslow, however, there are other aspects to needs and what motivates an individual. For example, Information needs by Norwood provides us with a view of society and individuals in an era in which information is a vital competitive advantage. Moreover, we emphasise that modern society is pushing individuals to seek the accomplishment of needs in the shortest possible time and with the least effort. TV Programs such as The Big Brother and American Idol testify that individuals seek not just status but also fame and fortune with limited knowledge and seeking shortcuts at all costs.

Since no universal needs model appears to be on the horizon it makes sense for enterprises to at least realise strategies that monitor and tag changes in such needs. Perhaps more important is that as enterprises acquire more sensitivity towards consumer needs and how individuals relate to them, so it becomes more likely that differences between, shareholders, stakeholders and consumers will shrink or be re-ordered. This is fortified by the social theory of Pierre Bourdieu that links consumption to macro-economics by providing a social perspective of Engel's curve.

There are conflicting views, and even concepts, in needs theories such as hierarchy, evolution of the individual and the prioritisation of needs and although this paper attempts to provide the widest possible view, the authors consider this paper to be a starting point for further development. The PIE model should help enterprises view consumers, institutions and their organisation as one interweaved entity.

Moreover, the authors feel that the major contribution of this paper is not only an overview of needs theories but especially how enterprises can link consumer needs to benefits, market identification and ultimately, strategy as well. This will be the subject of part B of this two part series of papers.

5. References

[1] Alderfer, C., (1972), *Existence, relatedness, & growth; Human Needs in Organizational Settings,* New York, Free Press.

[2] Allport, G.W., (1937), *The Functional Autonomy of Motives*, first published in the American J. of Psychology, 50, pp.141-156. available from http://psychclassics.yorku.ca/Allport/autonomy.htm

[3] Allport, G., (1960), *Personality and social encounter: Selected essays,* New York, Beacon Press.

[4] Allport, G., (1961), *Pattern and growth in personality,* New York: Holt, Rinehart and Winston.

[5] Di Michael, J. B., (2003), *The Marketing Book*, UK, Butterworth-Heinemann, pp.518-523.

[6] Bourdieu, P., (1998), *Practical Reason: On the Theory of Action*, Stanford University Press.

[7] Cullen, J. B., (2001), *Multinational Management: A Strategic Approach*, 2nd edition, South Western, pp.502-531.

[8] Daniels, M., (2001), *Maslow's concept of self-actualization,* available from http://www.mdani.demon.co.uk/archive/MDMaslow.htm.

[9] Doyal, L. and Gough, I., (1991), *A Theory of Human Need*, London, Macmillan, pp.1-365

[10] Eccles, J. E. and Wigfield, A., (2002), *Motivational Beliefs, Values and Goals*, Annual Review of Psychology, vol. 53, pp.109-132

[11] Fortier, M. S., Sweet, S. N., O'Sullivan, T. L., and Williams, G. C., (2007), *A self-determination process model of physical activity adoption in the context of a randomized controlled trial*. Psychology of Sport and Exercise, vol.8, pp. 741–757.

[12] Frankl, V. E., (1959), *Man's Search for Meaning*, 1st edition, Beacon, Boston Press.

[13] Glasser, W., (1998), *Choice Theory. A New Psychology of Personal Freedom*, New York, Harper Collins.

[14] Gough, I., (1994), *Economic Institutions and the Satisfaction of Human Needs*. Journal of Economic Issues. Volume 28, pp.25-66

[15] Gorman, P., (2004), *Motivation and Emotion*, Sussex, England, Routledge, pp.57-71.

[16] Grant, R.M., (2002), *Contemporary Strategy Analysis*, Oxford, Blackwell, pp.285-288.

[17] Herzberg, F., Mausner, B. and Snyderman, B. B., (1959), *The Motivation to Work*, New York, Wiley & Sons.

[18] Hofstede, G., (2001), *Culture's Consequences: Comparing values, behaviours, institutions and organisations across nations*, London, Sage.

[19] Horwitz, R. *et al.* (1990), *John Locke, Questions Concerning the Law of Nature*, Ithaca: Cornell University Press.

[20] Huitt, W., (2001), *Motivation to Learn: an Overview*, Educational Psychology Interactive, Valdosta, GA: Valdosta State University, available from: http://chiron.valdosta.edu/whuitt/col/motivation/motivate.html.

[21] James, W., (1890), *The Principles of Psychology*, New York, Henry Holt and Company, vol. 1, ch.10.

[22] Johnson, G. M. and Henley, T. B., (1990), *Reflections on The Principles of Psychology,* William James After a Century, Lawrence Erlbaum Associates

[23] Kano, N., (1993), *Kano's Method for Understanding Customer-defined Quality,* Center for Quality of Management Journal, Japan, vol.2.

[24] Kiel, J.M., (1999), *Reshaping Maslow's Hierarchy of Needs to Reflect Today's Educational and Managerial Philosophies*, Journal of Instructional Psychology.

[25] Kim, W.C. and Mauborgne, R., (2005), *Blue Ocean Strategy: How to Create Uncontested Market Space and Make Competition Irrelevant*, Harvard Business Press.

[26] Kotler, P., (2003), *Marketing Management*, 11th international edition, New Jersey, Prentice/Hall, pp.195-198, 340.

[26] Kotler, P. and Keller, K., (2006), *Marketing Management*, 12th edition, New Jersey, Pearson Prentice Hall, pp. 140-141, 144, 185.

[27] Lakoff, G. and Núñez, R.E., (2002), *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being,* Basic books.

[28] Latham and Locke, (2002), *Building a Practically Useful Theory of Goal Setting and Task Motivation, a 35year Odyssey*, American Psychologist, pp.705-717.

[29] Lowry, R. and Maslow, B. G., (1979), *The Journals of A. H. Maslow,* Thomson Brooks/Cole,

[30] Maslow, A.H., (1943), *A Theory of Human Motivation*, originally published in Psychological Review, vol.50, pp.370-396.

[31] Maslow, A.H., (1954), *Motivation and Personality*, Maslow Publications (available on-line).

[32] Maslow, A.H. and Lowery, R., (1998), *Toward a psychology of being*, 3rd edition, New York, Wiley & Sons.

[33] Mathes, E.W., (1981), *Maslow's Hierarchy of Needs as a Guide for Living*, Journal of Humanistic Psychology, vol. 21, no. 4, Fall, pp.69-72.

[34] McCarthy, J.E., Cannon, J. P., and Perreault, W.D., (2006), *Basic Marketing*, 16th edition, Irwin publishers.

[35] McClelland, D.C. and Burnham, D., (1976), *Power is the Great Motivator*, Harvard Business Review Classics.

[36] Mintzberg, H., Ahlstrand, B. and Lampel, J., (1998), *Strategy Safari*, Harlow, UK, FT Prentice Hall.

[37] Morace, F., (2008), *Consum-authors*, Scheiwiller, Milan. Italy (Bilingual text: Italian/English).

[38] Mühlbacher, H., Leihs, H. and Dahringer, L., (2006), *International Marketing a Global Perspective*, 3rd edition, London, Thomson, pp.178-188.

[39] Nussbaum, M.C. and Sen, A., (1993), *The Quality of Life*, Oxford: Clarendon Press.

[40] Nishino, T., Satsuta, R., Uematsu, S., and Nagamachi, M., (2008), *Identification of Customers-Latent Kansei Needs and Product Design by Rough set based approach*, Hiroshima International University.

[41] Norwood, G., (1999), *Maslow's hierarchy of needs. The Truth Vectors (Part I),* available from: http://www.deepermind.com/20maslow.htm.

[42] Prahalad, C.K., (2005), *The Fortune at the bottom of the Pyramid*, New Jersey, Pearson education.

[43] Price, S., (1994), *The special needs of children*, Journal of Advanced Nursing, 20, pp.227-232.

[44] Porter, M.E., (1980), *Competitive Advantage: Techniques for Analyzing Industries and Competitors*, New York, The Free Press, pp.2-32.

[45] Porter, M.E., (1985), *Competitive Advantage: Creating and Sustaining Superior Performance*, New York, The Free Press, pp.33-61.

[46] Revich, A., (2005), *The Three Fundamental Needs model*, available from: www.workitout.ca

[47] Saloner, G., Shepard, A. and Podolny, J., (2001), *Strategic Management*, New York, Wiley & Sons.

[48] Simons, J.A., Irwin D.B. and Drinnien, B.A., (1987), *Psychology - The Search for Understanding*, New York, West Publishing Company.

[49] Spitzer, D.R., (1996), *Power Rewards: Rewards That Really Motivate*, Management Review, pp.45-50.

[50] Stephens, D.C., (2000), *The Maslow Business Reader*, New York, Wiley & Sons.

[51] Stroud, D.J., (2008), *The Kano Analysis: Customer Needs Are Ever Changing*, Six Sigma Financial Services, available from: http://finance.isixsigma.com/library/content/c071017a.asp#author.

[52] Ward, D. and Rivani, E., (2005), *An Overview of Strategy Development Models and the Ward-Rivani Model,* Economics Working Papers, June, 2005. available from: ftp://wueconb.wustl.edu/econ-wp/get/papers/0506/0506002.doc.

[53] Ward, D. and Farmaki, E., (2006), *Loosening the Social Exclusion Knot and Exploring the Chemistry of Social Inclusion*, Social Science Research Network, http://hq.ssrn.com (working paper publication No. 630340).

[54] Ward, D. and Secondi, D., (2005), *Virtual branding Turning a Stone into a Jewel*, Economic working papers, pp.7-8, Economics Working Papers, available from: ftp://wueconb.wustl.edu/econ-wp/get/papers/0507/0507001.doc.

[55] Ward, D. and Chiari, C., (2008), *Keeping Luxury Inaccessible*, available from: http://econpapers.repec.org/paper/pramprapa/11373.htm

[56] Wynn, T. and Coolidge, F., (2004), *The Expert Neanderthal Mind*, Journal of Human Evolution, Vol. 46, pp.467-487.

[57] Wynn, T. and Coolidge, F., (2008), *Un incontro di menti nell'età della pietra (A meeting of minds during the stone age)*, Le Scienze (Italian edition of the Scientific American magazine), pp.74-83.

Additional Suggested Readings

[58] 12manage website: *Benchmarking, Extended marketing mix (7Ps),* QFD, *SWOT analysis,* www.12manage.com, last accessed 09.01.2009.

[59] Alderfer, C.P., (1969), *An Empirical Test of a New Theory of Human Needs*, in: *Organizational Behavior and Human Performance*, vol. 4, pp. 142-175.

[60] Deci, E.L. and Ryan, R.M., (2000), *The 'What' and 'Why' of Goal Pursuits: Human Needs and the Self-Determination of Behavior*, in: *Psychological Inquiry*, vol. 11, n. 4, pp. 227-268.

[61] Di Deana McDonagh, Hekkert, P., van Erp, J. and Gyi, D., (2002), *Design and Emotion: The Experience of Everyday Things*, CRC Press.

[62] Ellis, A. and Abrams, M., (2008), *Personality Theories: Critical Perspectives*, SAGE, pp.295-298.

[63] Ewen, R.B., (1992), *An Introduction to Theories of Personality*, 4th edition, Psychology Press.

[64] Eysenck, M.W., (2004), *Psychology: An International Perspective*, chapter 3, Psychology Press.

[65] Herzberg, F., (1966), *Work and the Nature of Man*, Cleveland, World Publishing Company.

[66] Heyligen, F., (1992), *A Cognitive-Systemic Reconstruction of Maslow's Theory of Self-Actualization*, in: *Behavioural Science*, vol. 37, pp. 39-57.

[67] Lawrence, P. & Nohria, N., (2001), *Driven: How Human Nature Shapes Organizations",* Harvard Business School Working Knowledge, available from: http://hbswk.hbs.edu/item/2543.html.

[68] MindTools website: *Alderfer's ERG Theory: Understanding the Priorities in People's Needs*, available from: https://www.mindtools.com/community/pages/article/newTMM_78.htm.

[69] Miner, J.B., (2005), *Organizational Behaviour 1: Essential Theories of Motivation and Leadership*, M. E. Sharpe, Armonk, USA.

[70] Motivation theory, available from: http://highered.mcgraw-hill.com/sites/dl/free/0072937769/128305/Santrock_PU7e_ch11.pdf

[71] Murray, H.A. and McAdams, D.P., (1938), *Explorations in Personality*, Oxford University Press.

[72] Price, S., (2008), *The special needs of children*, Journal of Advanced Nursing, Wiley & Sons, vol.20, issue 2, pp.227-232

[73] Silverthorne, S., (2001), *Five Questions for Paul Lawrence and Nitin Nohria*, Harvard Business School Working Knowledge, available from: http://hbswk.hbs.edu/item/3567.html

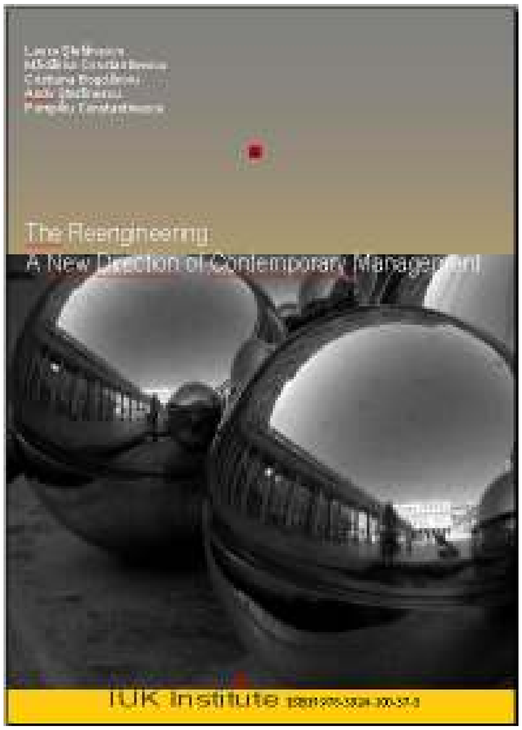
[74] A motivation to learn, a summary of needs theories concerning knowledge and learning: available at http://chiron.valdosta.edu/whuitt/col/motivation/motivate.html

[75] Wahba, A., & Bridgewell, L., (1976), *Maslow reconsidered: A review of research on the need hierarchy theory*, in: *Organizational Behavior and Human Performance*, 15, pp. 212-240.

THE REENGINEERING – A NEW DIRECTION OF CONTEMPORARY MANAGEMENT

Book Review

Jean-Paul **GAERTNER**,
l'Institut Européen d'Etudes Commerciales Supérieures*,* **France**



R*eengineering* has captured the imagination of managers and shareholders alike, sending corporations on journeys of radical business redesign that have already begun to transfigure global industry. Yet aside from earning them improvements in their business performance, the shift into more-process-centered organizations is causing fundamental changes in the corporate world, changes that business leaders are only now beginning to understand. What will the revolutions final legacy be?

The work, materialized in this volume is the result of an ample scientific approach which tries to synthesize the main tendencies and currents expressed nowadays in economic science and offer a general perspective over the actual stage of development of this science. Covering extremely complex problems, but also exciting ones, the content of the work represents a theme of proceedings and controversies over the economic contemporary thinking.

Authors begin their book rather defensively by insisting that reengineering is not merely a forgotten fad of the 1990s. And they may be right, particularly given their insistence that companies must be totally, absolutely willing to discard the old and replace it with the new. The authors make dramatic claims for the potential of reengineering, and highlight interesting victories. The book presents reengineering as a simple, straightforward way to view business processes, figure out how to make them more rational and economical, and then implement necessary changes. The authors made a splash by labeling this approach as reengineering in the 1990s. The term became a euphemism for firing people in droves and then fell into discredit. This update may be intended to rescue the concept from its bad image, but it doesn't quite succeed.

No business concept was more important to America's economic revival in the 1990s than reengineering – Introduced to the world in Michael Hammer and James Champy's Reengineering the Corporation. Already a classic, this international bestseller pioneered the most important topic in business circles today: reengineering – the radical redesign of a company's processes, organization, and culture to achieve a quantum leap in performance.

In *The Reengineering – A New Direction of Contemporary Management*, Laura Ştefănescu, Mădălina Constantinescu, Cristiana Bogdănoiu, Andy Ştefănescu and Pompiliu Constantinescu explain how some of corporations are reengineering to save hundreds of millions of dollars each year, to achieve unprecedented levels of customer satisfaction, and to speed up and make more flexible all aspects of their operations.

Now, more changes and challenges are coming to the corporation in the throes of the Internet age at the start of the 21st century. The authors have updated and revised their milestone work for the

New Economy they helped to create – promising to help corporations save hundreds of millions of dollars more, raise their customer satisfaction still higher, and grow ever more successful and adaptive in the years to come.

In *The Reengineering – A New Direction of Contemporary Management*, authors talk about the bases of organizational culture, focusing their research on the analysis, models and concepts of management, improving the self knowledge regarding the acquisition of working methods, critical methods, philosophy and action development of management. Likewise, the development of knowledge in the field of management lies behind every method or model applied by people, by creators, whose approaches they tried to emphasize in their own vision on the world and organization.

The authors also are approaching the effect that leadership and change management strategy place on acceptance of cultural change by individuals during a merger. Recent empiric qualitative and quantitative findings reveals that in many cases the change that occurs have a profound impact on leaders being often the key-element that blocks the successful cultural re-engineering. These findings allowed them to analyse how the successful change which is depending directly of the leaders' competence and training in the process of transforming the organizations ensure that individuals within the organization accept the changes prompted by the merger. Leader's communication and transparency in the change process impacts strongly on the organization's image. To gain the followers trust and confidence leaders must be effective and they have to use their knowledge and competence to ensure the success of change process.

An important vector of business globalization, of resources change and of firm business capabilities in competitive advantages is represented by the usage of information technologies for reengineering. Analyzing the impact of information technology on the working processes within an organization, the authors conclude the following: The most important management task is to eliminate the activities which do not produce added value and not to use the technology to automatise them; The ineffective activities within an organization must be eliminated and not accelerated through automatization; Organizations must think over the working processes in order to maximize the value given to the clients, at the same time minimising the use of necessary resources for product or service provision; Information technology can be used as a catalyst to optimise the working processes, as it allows new working methods and cooperation.

All these conclusions underline the fact that the implementation of information technologies must not be the main objective of technology usage, because often the working method of the organizations is not effective, and informatics' modelling of this inefficiency can only make this error grows longer. Also, the usage of information technology is a change itself, which can create synergies and opportunities to ease working flows. For this reason, the implementation of information systems should keep these realities in mind, and the implementation of the instruments provided by the information technologies should not be made unless the working flows have been optimised. This approach assumes the creation of a Business Reengineering Project, as a part of the implementation of any information system.

JOURNAL of Applied Economic Sciences

Call for Papers: Special issue on Management Research and Practice

Guest Editor:

Prof. Dr. Shankar Gargh
M. Sc. (Gold-Medalist), Ph.D., M.B.A. (U.S.A.),
LL.B., A. Inst. Pet. (London), FICCE, FISB, FICDM, FISM.

Editor-in-Chief of various Journals
Management Consultant
Consultant in Waterproofing Chemicals & Anti-Corrosive Coatings & Polymers
Environment Consultant
UN Accredited Press Representative

Significance of the Topic

Management Research and Practice issue will be dedicated to communicating innovative, new research that advances theory and practice in management and administration. The special issue is committed to expanding the scope of management theory and enriching practice by:

- encouraging multi theoretical approaches that span multiple social science disciplines
- welcoming papers using a diversity of innovative research methods
- creating focused volume that explore in depth promising new research directions, consolidate research streams, and address significant current theoretical and practical problems.

Schedule

Deadline for Submission of Papers:	15th May 2009
Expected Publication Date:	June 2009

Contact information

Dr. Shankar Lal Gargh
E-mail: management@managein.org

Call for papers: Journal of Applied Research in Finance

Editor in Chief: Mădălina **Constantinescu**
Managing Editor: Laura **Ungureanu**
Co-Editor: Laura **Ştefănescu**
Redactor: Cristiana **Bogdănoiu**

Published two times a year, the journal is the official publication of *The European Centre of Managerial and Business Studies* www.cesmaa.uv.ro, 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 ***J****ournal of* ***A****pplied* ***E****conomic* ***S****ciences* www.jaes.reprograph.ro (JAES). Two issues are published per volume. 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.

The ***J****ournal of* ***A****pplied* ***R****esearch in* ***F****inance* 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 ***J****ournal of* ***A****pplied* ***R****esearch in* ***F****inance* are welcome. The paper must be an original unpublished work written in English, not under consideration by other journals.
Invited manuscripts will be due till May 31, 2009, and shall go through the usual, albeit somewhat expedited, refereeing process.

Schedule:

Deadline for Submission of Papers:	30th May 2009
Expected Publication Date:	June (e-version) – July (hard-copy) 2009

E-mail: jarf_secretary@yahoo.com
Cc: jaes_secretary@yahoo.com

International Journal of Business Environment (IJBE)

ISSN (Online): 1740-0597 - ISSN (Print): 1740-0589

Call for Papers:
Special issue on "Globalization and Emerging Markets"

Guest Editors:
Dr.Rajesh K. **Pillania**, Management Development Institute,Gorgeon, India
Dr.Laura **Ştefănescu**, Spiru Haret University, Dolj, Romania

Need and Significance of the Topic

With a high growth rate and still significant untapped potential, emerging markets are the growth engines of the world economy. The term 'emerging market' was originally coined by IFC to describe a fairly narrow list of middle-to-higher income economies among the developing countries, with stock markets in which foreigners could buy securities. The term's meaning has since been expanded to include more or less all developing countries. How to face the changing realities of globalization requires substantial strategic thought, guidance and implementation. Contributions are invited on various aspects of globalization and emerging markets.

A Suggested List of Topical Areas: A list of suggested topics includes but is not limited to the following:

- Rise of Emerging Markets
- Liberalization in Emerging Markets
- Globalization of Emerging Markets Firms
- Conquering Emerging Markets
- Business Environment in Emerging Markets
- Corporate Governance Issues
- Issues of Corruption and Ethics
- Future of Globalization of Emerging Markets

The International Journal of Business Environment (IJBE) fosters research on how firms behave under different types of environment and aims to examine the external influences on business organizations. This journal examines a wide variety of business decisions, processes and activities within the business environment. *IJBE* publishes original and practical contributions. These can be quantitative or qualitative, surveys, reviews or cases studies, or research notes up-to-date with current research in the field.

Journal Webpage http://www.inderscience.com/browse/index.php?journalCODE=ijbe

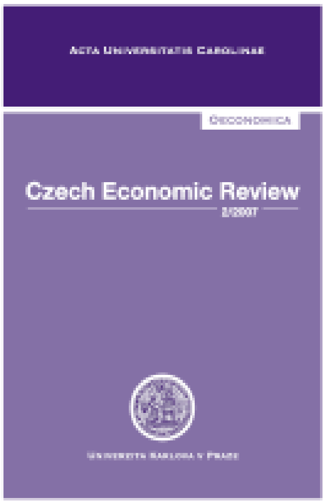
Schedule

Deadline for Submission of Papers	June 30, 2009
Expected Publication Date	October-December 2009

Contact information

Dr. Rajesh K Pillania, E-Mail: r_pillania@yahoo.com

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
Web:	http://auco.cuni.cz/
Post address:	UK FSV IES, AUCO Czech Economic Review, Opletalova 26, 110 00 Prague 1, Czech Republic

European Research Centre of Managerial Studies in Business Administration and Spiru Haret University announce:

The 3rd International Conference on
„Further Challenges for a Competitive Management"
Craiova, Romania, 23- 24 October 2009

„Further Challenges for a Competitive Management" *is* an international conference devoted to research in all branches of economics and encouraging the application of economic analysis to specific problems in both the public and private sectors. It particularly fosters quantitative studies, the results of which are of use in the practical field, and thus help to bring economic theory nearer to reality.

Theoretical and empirical papers are welcome from all parts of the international research community. Contributions which make use of the methods of mathematics, statistics and operations research will be welcomed, provided the conclusions are factual and properly explained.

Call for Papers

We are kindly invited to participate in **The 3rd International Conference on „Further Challenges for a Competitive Management"** held in **Craiova, Romania** from 23 to 24 October 2009.

This International Conference aims to promote new thinking on how institutions and institutional change can be analysed and measured and how their impact and aggregate economic performance can be evaluated. *The Program Committee* seeks contributions, which **topics** include, but are not limited to:

- Business Administration
- Risk Management
- Mathematics Models of Economical Processes
- Accounting
- General Financial Markets
- Mathematical Modelling
- Knowledge Management
- Management Information System
- ICT , e-Business, Business Intelligence
- Intelligent and Computer Systems in: economy, banking, insurance, and elsewhere
- Management of Technological Innovation and R&D
- Business Communications
- Technological Change
- Law and Economics
- Environmental Economic

We invite to submission original research contributions which describing new results, original ideas and applications. Paper have to submitted electronically at this e-mail address office_conference@yahoo.com in MS Word and PDF format (see Instructions for Authors)

The best papers will be published in ***Journal of Applied Economic Sciences*** (www.jaes.reprograph.ro) which is indexed in RePEc, CEEOL and EBSCO databases as well as in our latest ***Journal of Applied Research in Finance*.**

All the other ***papers submitted and presented*** at the works of the Conference will be published in a **Special Issue** of Journal of Applied Economic Sciences - **Further Challenges for a Competitive Management** after peer reviewing.

IMPORTANT DATES:

June 1st, 2009 – Abstract submission deadline
June 25th, 2009 - Notification of acceptance/rejection
July 10th, 2009 – Deadline for payments
July 15, 2009 - Final paper submission deadline
October 23-24th, 2009 - Conference

Instructions for Authors

The abstract should be send till June 1st, 2009, via e-mail, at the corresponding address of the conference (office_conference@yahoo.com). The abstracts of the accepted paper will be published in a volume which will be distributed among participants at the conference. The abstract should be written with Times New Roman, size 11 pt., normal, line spacing: single and it will be of approximately 500 – 700 words, no more than 1 page.

Authors should submit **final paper** in English up to 8 A4 pages, using the paper format indicated in MS Word Sample. Authors must also indicate the Conference area to which the paper is submitted. The paper must be carefully checked for correct grammar and spelling.

Paper should be submitted electronically by e-mail office_conference@yahoo.com Author name should be used as file name (e.g. popescu.doc, smith.doc, etc).

Conference Fee

Conference Fee covers conference materials, refreshments during coffee-breaks.

- **80 Euro** for foreign participants
- **80 Lei** for Romanians participants
- **50 Lei** for CESMAA members

Payment will be to perform immediately after notification of acceptance of your paper until 2009, July 10. All fees derived by the bank transfer will be supported by the sender.
Bank: Raiffeisen Bank – N.Titulescu Agency

IBAN ACCOUNT for EURO: RO87RZBR0000060008756957
IBAN ACCOUNT for LEI: RO98RZBR0000060008756953
Owner: Felicia MANEA

Please mention on the payment form that the fee is for the Conference and your name (*e.g. for Conference_2009_Smith*).
It is essential to send a copy of Payment Form (Bank Transfer):

- copy by fax: 40 251 598265
- or a scanned copy by email to: office_conference@yahoo.com