JOURNAL of Applied Economic Sciences

Quarterly Journal

JAES

Journal of Applied Economic Sciences

Volume IV
Issue 2(8)

Summer 2009

I.S.S.N. 1843-6110

Journal of Applied Economic Sciences

Editorial Board

Publisher:

European Research Centre of Managerial Studies in Business Administration

Email: jaes_secretary@yahoo.com

Volume IV/ Issue 2(8)/ Summer 2009

Journal of Applied Economic Sciences

ISSN 1843-6110

Table of Contents

MECHANISMS OF GOVERNANCE OF SUSTAINBLE DEVELOPMENT

Hrabrin **Bachev**
Institute of Agricultural Economics, Sofia, **Bulgaria**
hbachev@yahoo.com

Abstract:

In this paper we incorporate the interdisciplinary New Institutional and Transaction Costs Economics (combining Economics, Organization, Law, Sociology, Behavioural and Political Sciences), and suggest a framework for analyzing the mechanisms of governance of sustainable development. Our new approach takes into account the role of specific institutional environment; and the behavioural characteristics of individual agents (personal preferences, bounded rationality, tendency for opportunism, trust, risk aversion); and the transaction costs associated with the various forms of governance; and the critical factors of economic activity and exchanges (such as appropriability, frequency, uncertainty, and asset specificity of transactions); and the comparative efficiency of market, private, public and hybrid modes; and the potential of production structures for adaptation; and the comparative efficiency of alternative modes for public intervention. Agricultural sector is used to illustrate that new approach and support with examples.

Keywords: institutions, market, private, public and hybrid modes; agrarian sustainability

JEL Classification: D02, D23, O13, O17, Q01, Q12, Q18, Q13, Q56, L14, L22

1. Introduction

Governance of sustainable development is among most topical issues in academic, business, and policies debates [EC (2005); Raman, (2006); VanLoon *et al.*(2005)]. It is recognized that achievement of economic, social, environmental, intra and inter-generational goals of sustainable development requires effective social order and coordinated actions at individual, organizational, community, regional, national, and transnational levels. It is known that effective forms of governance are rarely universal and there is a big variation among different countries, regions, subsectors. Experience shows that different societies achieve to a different extend economic, social, environmental etc. goals of sustainable development. That is a result of specific governing structures which affect in dissimilar ways individuals behaviour, give unlike benefits, command different costs, and lead to diverse actual performances. In this paper we incorporate the interdisciplinary *New Institutional and Transaction Costs Economics* [Coase, (1960); Furuboth and Richter, (1998); North, (1990); Williamson, (1996)], and suggest a framework for analysis of mechanisms of governance of sustainable development. Our new approach takes into account the role of specific institutional environment; and behavioural characteristics of individual agents; and transaction costs associated with various forms of governance; and critical factors of economic activity and exchanges; and comparative efficiency of market, private, public and hybrid modes; and potential of production structures for adaptation; and comparative efficiency of alternative modes for public intervention. Agricultural sector is used to illustrate the approach and support with examples.

2. "Institutions matter"

Institutions are the "rules of the game", and they determine individuals' rights in society and way property rights are enforced [Furuboth and Richter, (1998); North, (1990)]. Spectrum of rights could embrace material assets, natural resources, intangibles, certain activities, labour safety, clean environment, food security, intra- and inter-generational justice. A part of property rights are constituted by formal laws, regulations, standards, court decisions etc. In addition, there are important *informal rules* determined by tradition, culture, religion, ideology, ethical and moral norms. Enforcement of various rights is done by state or other mechanisms such as community pressure, trust, reputation, private modes, self-enforcement etc. Institutional analysis is not interested in de-jure rights but de-facto rights individuals and groups possess. For instance, the "universal principles" of sustainable development have been declared [Rio Earth Summit, (1992)] and accepted by most countries. However, extend of adaptation and respecting of related rights and their practical enforcement vary significantly among countries.

Specific institutional environment affects human behaviour and directs (governs) individuals' activities "in a predictable way" [North, (1990)]. It creates dissimilar incentives and restrictions for intensifying exchange, increasing productivity, inducing private and collective initiatives, developing new rights, decreasing divergence between social groups and regions, responding to ecological and other challenges. For example, (socially) acceptable norms for labour use (employment of children, safety standards, wages), plant and livestock (animal welfare, preservation of biodiversity, usage of GM crops), and environmental resources (water use rights; permissions for pollution), could differ even between regions of the same country. Namely institutional structure eventually determines potential for and particular type of development in different communities, regions, and countries.

Institutional "development" is initiated by public authority, international actions (agreements, assistance, pressure), and private and collective actions of individuals. It is associated with modernization and/or redistribution of existing rights; and evolution of new rights and emergence of novel (private, public, and hybrid) institutions for their enforcement. For instance, sustainability initially evolved "movements" and "new ideology" in developed countries. Afterward this "new concept" extended and instituted in the body of formal laws, regulations and public support programs. Numerous initiatives of producers and consumers are wide-spreading (codes of ethical behaviour, organic farming, and system of fair-trade) being important part of institutional modernization in the area.

Diverse institutional environment contributes to different extend to achieving economic, social, environmental etc. goals of sustainable development. If for instance, private rights are not well defined, enforced, or are restricted, that would limit intensification of exchange and overall economic development. Indeed rights on major agrarian resources were not well defined during post-communist transition in Bulgaria and that led to domination of low productive, unsustainable and "gray" structures; ineffective use of large national resources; and serious economic, social and environmental problems [Bachev, (2006)]. The classical examples for importance of institutional structure are associated with the "tragedy of commons" (Hardin) and the negative externalities [Pigou, (1920)].

Thus "institutions matter" and analysis of sustainability is to be done in the specific institutional rather than in an unrealistic ("normative", desirable) context. Nevertheless, institutional aspect is commonly missing in most of suggested frameworks for analyzing and assessing sustainability. Accordingly, non-feasible norms rather than the real-life arrangements are used as criteria – e.g. the farming model in other (developed) countries, the assumption for perfectly defined and enforced property rights, the effectively working public (local, state, inter-governmental) organizations etc. Therefore, an analysis of the structure and the evolution of the real or other feasible institutional arrangements for carrying out the agrarian activities have to be included in the model [Bachev, (2004)].

3. The modes of governance

The New Institutional Economics gives a new insight on efficiency of divers market, private, public and mix modes of governance, and their potential to deal with agrarian sustainability [Bachev, (2004), (2007)]. This new approach requires embracing all modes of governance affecting individuals' behaviour which includes: *institutional environment* ("rules of the game") – that is distribution of rights and obligations between individuals, groups, communities and generations, and system(s) of enforcement of these rights and rules. In modern society a great deal of individuals activities and relations are regulated by some (general) formal and informal rules. However, there is no perfect system of preset outside rules that can govern effectively the entire activities of individuals in all possible (and quite specific) circumstances of their life and relations.

– *market modes* – various decentralized initiatives governed by free market price movements and market competition (spotlight exchanges, classical contracts, production and trade of organic products and origins, system of fair-trade). Importance of the "invisible hand" of market for effective coordination and stimulation of individuals' activities is a fundament of modern economy (and policies for development and globalization). However, there has been many "market failures" compromising sustainable development and leading to social crisis, economic crisis, ecological crisis, energy crisis.

– *private modes* ("private or collective order") – diverse private initiatives, and specially designed contractual and organizational arrangements governing bilateral or multilateral relationships

of private agents (voluntary individual or collective actions, codes of behaviour, environmental contracts, eco-cooperatives). There has been emerging a great number of private and collective forms managed by "visible hand of the manager", collective decision-making, private negotiations etc. governing successfully various aspects (and challenges) of sustainable development. Nevertheless, there exist abundant examples of "private sector failures" (lack of potential to coordinate and stimulate sustainability) demonstrating incapability to deal effectively with problems of development.

– *public modes* ("public order") – various forms of a third-party public (Government, community, international) intervention in market and private sectors such as public guidance, regulation, taxation, assistance, funding, provision. The role of public (local, national, transnational) governance has been increasing along with intensification of activity and exchange, and growing interdependence of social, economic and environmental activities (and related problems and risks). In many cases, effective organization of certain activity through market mechanism (price competition) and/or private negotiation would take a long period of time, be very costly, could not reach a socially desirable scale, or be impossible. Thus a centralized public intervention could achieve the willing state of the system faster, cheaper or more efficiently. Nonetheless, there has been a great number of bad public involvements (inaction, wrong intervention, over-regulation) leading to significant problems of sustainable development around the globe.

– *hybrid forms* – some mixture combining features of market and/or private and/or public governance - the state certifies organic producers and enforces organic standards, and thus intensifies development of organic markets and environmental sustainability.

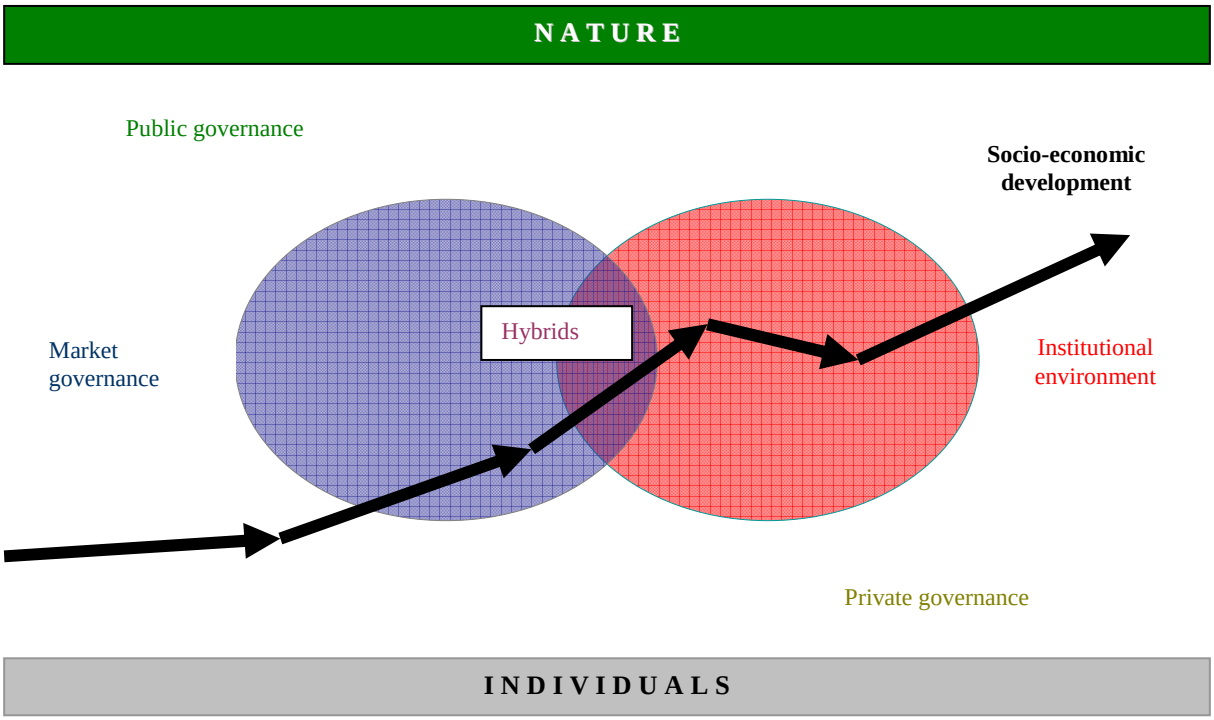


Figure 1. Mechanisms of governance of sustainability

In one person world there is no need for (any) governance since the sustainable relations between that person and the nature are achieved through a simple (production and/or consumption) management ("self-governance"). However, in the real world of limited resources, complex social interactions between many individuals (division, specialization and cooperation of labour, intensive exchanges) and conflicting interests, there is a need for a special governing mechanism to direct, coordinate, stimulate, induce and enforce individuals efforts to accomplish a sustainable development. The achievement of the state of an overall efficiency (the maximum social welfare, sustainability) is driven by various social arrangements – preset formal and informal rules (institutional environment), competition, contracting, cooperation, profit-making or non-for profit activity, collective actions, pure private order, public order, voluntary initiatives, mixed modes etc. Depending on the efficiency of system of governance which is put in place, the outcome of development is quite different (Figure 1).

Therefore, all systems for assessment of sustainability must include not only the outcome(s) of the process, which is the "current" level (the state) of sustainability. Evaluation is to embrace the system of governance put in place, which is the social mechanism responsible for the outcome. Otherwise, mere analysis of the state or trend indicators would give no adequate picture for the ability of the system to improve, sustain, or adapt to a new sustainable level. Thus the problem for assessing the efficiency of individual governing mechanisms and for selecting the most efficient one(s) is very important.

4. The costs of governance

Transaction costs are the costs associated with protection and exchange of individuals' rights [Furuboth and Richter, (1998)]. In addition to the production costs, economic agents make significant costs for coordination of their relations with other individuals: for finding best prices and partners for land, inputs and labour supply, financing, and marketing of outputs and services; negotiating conditions of exchange; completing and "writing down" contract or setting up partnership organization (coalition); coordination through collective decision-making or direct managerial orders; enforcing negotiated terms through monitoring, controlling, measuring and safeguarding; disputing through a court system or another way; adjusting or termination along with changing conditions of exchange.

Institutional environment and its development also impose significant transaction costs to individuals – for studying out and complying with various institutional restrictions (community or state norms, regulations, standards), formal registrations, efforts to deal with bureaucracy etc. A good example are current problems of many Bulgarian farms to meet EU requirements ("institutionally determined" costs) related to new product quality, food safety, eco, animal welfare etc. standards.

Transaction costs have two *behavioural* origins: individual's *bounded rationality* and *tendency for opportunism* [Williamson, (1996)]. Economic agents do not possess full information about the system (price ranges, trade opportunities, adverse effects on others, trends in development) since collection and processing of such information would be either very expensive or impossible (future events, partners intention for cheating, time and space discrepancy between individual action and adverse impacts on others). In order to optimize decision-making (to reach state of efficiency) they have to spent costs for "increasing their imperfect rationality" - data collection, analysis, forecasting, training.

Individuals are also given to opportunism and if there is opportunity for some of transacting sides to get non-punishably an extra rent from exchange (performing unwanted exchange) he will likely "steal" others rights. Two major forms of opportunism can be distinguished: pre-contractual ("adverse selection") - when some of partners use information asymmetry to negotiate better contract terms; and post-contractual ("moral hazard") - when some counterpart takes advantage of impossibility for full observation on his activities (by partner or third party) or when he take "legal advantages" of unpredicted changes in transacting conditions (costs, prices, environment). Third form of opportunism occurs in development of large organizations (known as "free-riding"). Since individual benefits are often not proportional to individual efforts, everybody tends to expect others to invest costs for organizational development and later on benefit ("free riding") from new organization [Olson]. It is very costly or impossible to distinguish opportunistic from non-opportunistic behaviour (because of bounded rationality). Therefore, agrarian agents have to protect their transactions and rights from hazard of opportunism through: ex-ante efforts to protect "absolute" (given by institutions) rights, and find reliable counterpart and design efficient mode for partners credible commitments to contracted (voluntary transferred) rights; and ex-post investments for overcoming (through monitoring, controlling, stimulating cooperation) of possible opportunism during contract execution stage.

If transaction costs were zero then mode of the governance would not be of economic importance. In such a world individuals would manage their relations with *equal efficiency* though free market, or through private organizations of different types, or in single nationwide company. All information for effective potential of transactions (exploration of technological opportunities, satisfying various demands, respecting assigned and transferred rights) would be costless available. And individuals would costless protect their (absolute, contracted) rights, and trade owned resources in mutual benefit until exhausting possibilities for increasing productivity, maximizing consumption,

and sustainable development[1]. However, very often high costs make it difficult or block otherwise efficient (mutually beneficial) transactions. Textbook cases of "market failure" are connected with negative (positive) externalities. Since free-market prices do not reflect effect on third party's welfare they cannot govern effectively relations between individuals. Maximization of social output (welfare) is not achieved, and inefficient allocation of resources and activities, and unsustainable development arrives. Hence farmers will over-produce "public bads" (noise, air, and water pollution) and under-produce "public goods" (rural amenities, ecological and cultural services; habitat for wildlife, biodiversity). That necessitates a "Government intervention" to eliminate the differences between the social and the private prices (an "internalization of externalities" through taxes, norms etc.).

The problem of "social costs" does not exist in the world of zero transaction costs and well-defined private rights [Coase, (1960)]. Situation of maximum efficiency is always achieved independent of the initial allocation of rights. However, when transaction costs are significant, then costless protection, negotiation and exchange of rights is impossible. The initial allocation of property rights between individuals is critical for the overall efficiency and sustainability. Moreover, if rights on important resources are not well-defined (e.g. rights on clean air and water) that creates big difficulties in effective allocation (e.g. unsolvable costly disputes between polluting farmers and neighbourhood). Consequently, some essential activities (and transactions) are not carried out at socially effective scale, and the existing governing structures less contribute to sustainable development [Bachev, (2007)].

Thus type of governance becomes crucial since various modes give unequal possibilities for participants to coordinate activities, stimulate acceptable behaviour of others, and protect contracted and absolute rights from unwanted expropriation. In a world of positive transaction costs rational agents will seek, chose, and develop such modes for governing of their activities and relations which maximize their benefits and minimize their total (production and transacting) costs. In the long run only efficient modes for governing of different activities will prevail (sustain) in agriculture [Bachev, (2004)]. However, sustainability of agrarian structures is necessary but not a sufficient condition for the sustainable development [Bachev and Peeters, (2005)]. The overall goals of sustainable development cannot be automatically achieved through totally decentralized actions (free market competition, private initiatives). There is a need for special (designed and installed) governance which include a significant public (community, national, transnational, global) intervention in the agrarian sector.

There is not singe (universal) mode for effective organization of all type agrarian activity in any possible natural, institutional, and economic surroundings [Bachev, (2004)]. Individual governing forms have distinct features (different advantages and disadvantages) to protect rights and coordinate and stimulate socially desirable activities. Besides, agents have specific personal characteristics – different awareness, entrepreneurships, preferences, risk aversion, tendency for opportunisms. Furthermore, efficiency of governing mode will depends on specific attributes of each activity and transaction. Therefore, individual transaction and transaction costs is to be put in the centre of analysis, and the comparative efficiency of feasible modes for governing of socially desirable activities assessed.

5. The principle governance matrix

Generally, every agrarian activity and transaction could be governed through a great variety of alterative forms. For instance, a supply of environmental preservation service could be governed as: voluntary activity of farmer; though private contracts of the farmer with interested (affected) agents; though interlinked contract between the farmer and a supplier (processor); though cooperation (collective action) with other farmers and stakeholders; though (free) market or assisted by a third-party (certifying and controlling agent) trade with special (eco, protected origins, fair-trade) products; though a public contract specifying farmer's obligations and compensation; though public order (regulation, taxation, quota for use of recourses); within hierarchical public agency or by hybrid form.

Different governance modes are alternative but not equal modes for organization of activities. Free market has big coordination and incentive advantages ("invisible hand", "power of competition"), and provides "unlimited" opportunities to benefit from the specialization and the exchange. However,

[1] Currently, there is a principle agreement ("social contract") for a global sustainable development.

market governance could be associated with a high uncertainty, risk, and costs due to the price instability, the great possibility for facing an opportunistic behaviour, the "missing market" situation etc. The special contract form ("private ordering") permits a better coordination, intensification, and safeguard of transactions. However, it may require large costs for the specification of contract provisions, for adjustments with constant changes in the conditions, for enforcement and disputing of negotiated terms etc. The internal (ownership) organization allows a greater flexibility and control on transactions (direct coordination, adaptation, enforcement, and dispute resolution by a fiat). However, the extension of the internal mode beyond the family and small-partnership boundaries (allowing achieving the minimum technological or agronomic requirements; exploration of technological economies of scale and scope) may command significant costs for development (initiation and design, formal registration, restructuring), and for current management (for collective decision making, control on the coalition members opportunism, supervision and motivation of hired labour etc.).

In order to select the best (most efficient) form for governing of a particular activity we have to assess the comparative advantages and disadvantages of practically possible forms for governance of that activity. In some cases advantages of certain mode of governance are not difficult to verify - e.g. when it gives bigger benefits (achieves socially desirable/effective scale) or commands minimum total costs. In such cases choice of most effective form of governance is easy since we can compare directly costs and benefits of alternatives. For instance, in most countries much of the agrarian activity is commonly governed in some sort of family farm, the supply of inputs or exchange of farm output are governed my market modes etc. However, in many instances, direct assessment (comparison) of costs and benefits of alternative governing arrangements are difficult or impossible to make. That is particularly true for some elements of transaction costs related to drivers' governance structures. In the later group we can include costs for finding best partners, negotiation, controlling and enforcement of contractual terms, organizational development, interlinked transacting, unrealized (failed) deals etc.

The discrete structural analysis is suggested to evaluate comparative efficiency of alternative governing forms [Williamson, (1996)]. Here the assessment of absolute levels of transaction costs of alternative governing structures is not necessary. This approach aims to evaluate relative levels of transacting costs between alternative modes of governance, and selecting that one which most economizes on transacting costs. Following that framework *first* we have to identify "critical dimensions" of transactions responsible for variation of transaction costs. "Frequency", "uncertainty", and "asset specificity" are identified as critical factors of transaction costs by Williamson [Williamson, (1996)] while "appropriability" added by Bachev and Labonne [Bachev and Labonne, (2000)].

When recurrence of transactions between the same partners is high, then both (all) sides are interested in sustaining and minimizing costs of their relations (avoiding opportunism, building reputation, setting up adjustment mechanisms). Besides, costs for development of a special private mode for facilitating bilateral (multilateral) exchange could be effectively recovered by frequent exchange. When uncertainty, which surrounds transactions increases, then costs for carrying out and secure transactions go up (for overcoming information deficiency, safeguarding against risk). Certain risks could be diminished by production management or through special market mode (purchase of insurance). However, governance of most transacting risk would require a special private forms – e.g. trade with origins; providing guarantees; using share-rent or output-based compensation; employing economic hostages; risk-pooling, inputs-supply or marketing cooperative; complete integration.

Transaction costs get very high when specific assets for relations with a particular partner are to be deployed. Relation specific investments are "locked" in transactions with a particular buyer or seller, and cannot be recovered through "faceless" market trade. Therefore, dependant investment (assets) have to be safeguarded by a special form such as long-term contract, interlinks, hostage taking, joint investment, or ownership integration. Transacting is particularly difficult when appropriability of rights on products, services or resources is low. "Natural" low appropriability has most of agrarian intellectual products (market information, meteorological forecasts, new varieties and technologies, software). Besides, all products and activities with significant (positive or negative) externalities are to be included in this group. If appropriability is low possibility for unwanted (market or private) exchange is great, and costs for protection of private rights (safeguard, detection of cheating, disputing) extremely high. Agents would either over produce (negative externalities) or under organize such activity (positive externalities) unless they are governed by efficient private or hybrid mode (cooperation, strategic alliances, long-term contract, trade secrets, or public order).

Secondly, we have to "align transactions (differing in attributes) with governance structures (differing in costs and competence) in discriminating (mainly in transaction cost economizing) way" [Williamson, (1996)]. According to the combination of specific characteristics of each transaction, there will be different most effective form for governing of activity (Table 1). Agrarian transactions with a good appropriability, high certainty, and universal character of investments (the partner can be changed anytime without significant additional costs) could be effectively carried across the free market through spotlight or classical contracts. Here the organization of transactions with a special form or within the farm (firm) would only bring extra costs without producing any transacting benefits.

Table 1. Principle modes for effective governance

Generic modes	***Critical dimensions of transactions***								
	Appropriability								
	High								*Low*
	Assets Specificity								
	Low				*High*				
	Uncertainty								
	Low		*High*		*Low*		*High*		
	Frequency								
	High	*Low*	*High*	*Low*	*High*	*Low*	*High*	*Low*	
Free market	Y	Y							
Special contract form			Y			Y			
Internal organization					Y		Y		
Third-party involvement				🚑				🚑	
Public intervention									🚑

Y - *the most effective mode*; 🚑 - a *necessity for a third party involvement*

Recurrent transactions with low assets specificity, and high uncertainty and appropriability, could be effectively governed through special contract. Relational contract is applied when detailed terms of transacting are not known at outset (high uncertainty), and a framework (mutual expectations) rather than specification of obligations is practiced. Partners (self) restrict from opportunism and are motivated to settle emerging difficulties and continue relations (situation of frequent bilateral trade). Besides, no significant risk is involved since investments could be easily (costless) redeployed to another use or users (no assets dependency). A special contract form is also efficient for rare transactions with low uncertainty, high specificity and appropriability. Dependent investment could be successfully safeguarded through contract provisions since it is easy to define and enforce relevant obligations of partners in all possible contingencies (no uncertainty surrounds transactions). Here the occasional character of the transactions does not justify the internalization within the farm (firm).

Transactions with high frequency, big uncertainty, great assets specificity (dependency), and high appropriability, have to be organized within the farm/firm (internal ownership mode). For instance, managerial and technological knowledge is quite specific to a farm, and its supply has to be governed through a permanent labour contract and coupled with ownership rights [Bachev, (2004)]. Capital investments in land are to be made on owned (long-leased) rather than seasonally rented land (high site and product specificity). All "critical" to farm material assets will be internally organized - production of forage for animals; important machineries; water supply for irrigated farming etc. While universal capital could be effectively financed by market form (e.g. bank credit), highly specific investments can be only made through an internal funding (own funds, equity sell, joint venture).

According to personality of resource owners and costs of their coalition, different type of farm (agro-firm) will be efficient - one-person, family, partnership, cooperative, corporative [Bachev (2004)]. If specific and specialized capital cannot be effectively organized within the farm (economy of scale/ scope explored, funding made), then effective governance outside farm-gates is to be used - group farming, joint ownership, interlinks, cooperative, lobbying for public intervention. When strong assets (capacity, time of delivery, site, branding) inter-dependency with upstream or downstream partner exists, then it is not difficult to govern transactions through contract modes (mutual interests for cooperation and restriction of opportunism). For instance, in Germany and France effective cooperative agreements between farmers and drinking water companies are widely used (symmetrical

dependency) and led to production methods protecting water from pollution [Hagedorn, (2002)]. However, very often farmers face unilateral dependency and need effective (ownership) organization to protect their interests. Transacting costs for initiation and maintaining such "collective organization" is usually great (big number of coalition, different interests of members, opportunism of "free-riding" type) and it is either unsustainable or does not evolve at all. That creates serious problems for efficiency (and sustainability) of individual farms - missing markets, monopoly or quasi-monopoly situation, impossibility to "induce" public intervention etc.

Third, we have to identify the situations of market and private sector failures – that is the critical points for sustainable development. Serious transacting problems arise when condition of assets specificity is combined with high uncertainty, low frequency, and good appropriability (Table 2). Elaboration of special governing structure for private transacting is not justified, specific investments are not made, and activity (restriction of activity) fails to occur at effective scale ("market failure" and "contract failure"). Similar difficulties are also encountered for rare transacting associated with high uncertainty and appropriability. In these cases, a third part (private agent, NGO, public authority) involvement in transactions is necessary (assistance, arbitration, regulation) to make them more efficient or possible. For instance, when State establishes (enforces) quality and safety standards for farm inputs (chemicals, machinery) and produces, or certify providers of services, or regulate employment relations, or guarantee minimum price for farmers, all that considerably facilitates and intensifies (market and private) transactions and increases farm sustainability. Emergence and unprecedented development of organic farming and system of fair-trade are also good examples in that respect. There is increasing consumer's demand (price premium) for the organic, semi-organic and fair-trade products in developed countries. Nevertheless their supply could not be met unless effective trilateral governance (including an independent certification and control) has been put in place.

When appropriability associated with a transaction is low, there is no pure market mode to protect and carry out activity effectively. Nevertheless, respecting others rights (unwanted exchange avoided) or "granting" additional rights to others (needed transactions carried) could be governed by "good will" or charity actions, NGOs, government or international organizations. For instance, a great number of voluntary environmental initiatives (agreements) have emerged driven by competition in food industries, farmers' preferences for eco-production, and responds to the public pressure for sound environmental management[2]. However, environmental standards are usually "process-based", and "environmental audit" is not conducted by independent party, which does not guarantee "performance outcome". Therefore, most of these initiatives are seeing as a tool for external image manipulation. Recent huge food safety, animal safety, and eco-scandals have demonstrated that such private schemes could often fail (high bounded rationality and possibility for opportunism).

In any case, voluntary initiatives could hardly satisfy the entire social demand especially if they require significant costs. Some private modes could be employed if high frequency (pay-back on investment is possible) and mutual assets dependency (thus incentive to cooperate) exists[3]. In these instances, unwritten accords, interlinking, bilateral or collective agreements, close-membership cooperatives, codes of professional behaviour, alliances, internal organization etc. are used. However, emerging of special large-members organizations for dealing with low appropriability (satisfying entire "social" demand) would be very slow and expensive, and they unlikely are sustainable in long run ("free riding" problem). Therefore, there is strong need for third-party public (state, local authority, international assistance) intervention to make such activity possible or more effective [Bachev (2004)]. For example, supply of environmental goods by farmers could hardly be governed through private contracts with individual consumers because of low appropriability, high uncertainty, and rare character of transacting (high costs for negotiating, contracting, charging all potential consumers, disputing). At the same time, supply of additional environmental protection and improvement service is very costly (in terms of production and organization costs) and would unlikely be carried out on a voluntary basis. Besides, financial compensation (price-premium) of farmers by willing consumers through a pure market mode is also ineffective due to high information asymmetry,

[2] Unprecedented development of the "codes of behaviors", eco-labeling and branding, environmental cooperatives, and "green alliances", all they are good examples in that respect.

[3] For instance, inter-dependency between a bee keeper and a neighboring orchard farm (symmetric dependency between needs of flower and needs for pollination).

massive enforcement costs etc. A third-party mode with a direct public involvement would make that transaction effective: on behalf of consumers state agency negotiates with farmers contract for "environment conservation and improvement service", coordinates activities of various agents (including direct production management), provides public payments for compensation of farmers, and controls implementation of negotiated terms.

6. Farm as a governing structure

A significant amount of agrarian activities is organized by different type farms and farming organizations. Sustainability of a farm is to characterize farm's *ability to maintain* (continue) *over time*. Since no economic organization would exist in long-term if it were not efficient (otherwise it would be replaced by more efficient arrangement), the problem of assessment of sustainability of farms is directly related to estimation of factors and level of farm efficiency. In traditional (Neoclassical) framework, farm is presented as a "production structure" and analyses of efficiency restricted to production costs ("factors productivity", "optimization of technological factors according to marginal rule"). This approach fails to explain why for a long period of time there exist so many farms with different levels of "efficiency" (productivity). In Bulgaria for instance, level of profitability and productivity in cooperatives has been 5 times lower than in private farms. Besides, there have been one million highly sustainable subsistent and non-profit farms in the country [Bachev, (2006)].

In addition to the production costs, modern farming is also associated with significant transaction costs. Therefore, "rational" agrarian agents will seek chose and/or develop most effective (less expensive) mode for organization of their transactions that minimize their bounded rationality, and safeguard their investments and rights from hazard of opportunism. When transaction costs are high, they could block otherwise effective transactions, and restrict farm size far bellow technologically optimal level. Very often high costs for market trading (finding a credit, marketing of output) and/or internal governance (deficiency of low transacting cost labour) limit farm size to miniature subsistent farming or family borders [Bachev, (2004)]. In other instances, existing effective potential to economize on market transacting costs could cause vast extension of farm size through backward, lateral or forward integration. For example, high costs for market and contract trading after 1990 has turned subsistent farming into the most effective (only possible) forms for organization of available agrarian assets (farmland, livestock) of more than a million Bulgarians [Bachev, (2006)]. On the other hand, enormous costs of market trading have caused domination of integrated and interlinked modes, and concentration of commercial farming in few thousands large agro-firms and cooperatives.

Thus in the world of positive transaction costs, farms and other agrarian organizations have a significant economic role to play. They are not only production but also a major governing structure – a form for organization of transactions and for minimization of transacting costs. Therefore, sustainability of different farms cannot be correctly understood and estimated without analyzing their comparative production and governance potential [Bachev and Peeters, (2005)].

Generally, every farm related transaction could be governed through a great variety of alterative market, contract, integral etc. forms. Each of these governing modes gives individuals dissimilar opportunities to coordinate, stimulate, and control transactions, safeguard investments from opportunistic expropriation, and profit from specialization, cooperation and exchange. One-person farm (firm) has zero internal transaction costs, but limited possibility for investment in specialized (specific) human and material capital. "Internal" opportunities for increasing productivity (through investments, exploring economy of scale and size) increases along with extension of coalition members (group farm, partnership). However, it is associated with enlargement of costs for making coalition (finding complementary and reliable partners) and internal costs for managing (coordination, reducing bounded rationality, controlling opportunism). Separation of ownership from management (cooperative, corporation) gives enormous opportunities for productivity growth but it is connected with huge transacting costs (for decreasing information asymmetry between management and shareholders, decision-making, adaptation, controlling opportunism of hired labour and between partners). Special contract form combines the potential for greater "control" on transactions with possibility to explore advantages of further specialization. Nevertheless, it could be connected with large costs for preparing and enforcement of contracts for complex occasional transactions with high unilateral dependency. Free market has big coordination and incentive advantages ("invisible hand", "power of competition"), and provides "unlimited" opportunities to benefit from specialization and

exchange. However, market governance could be associated with high uncertainty, risk, and costs due to price instability, great possibility for facing opportunistic behaviour, "missing market" situation etc.

Protection of rights and economic exchanges let more profitable use of resources but also require additional costs. Farmers and other economic agents (resource owners, consumers) will tend to govern their activity and relations though the most effective forms – that which maximize their benefits and minimize their costs. Therefore, most effective form and size of farm will be determined through optimization of total (production *and* transacting) costs, and trade-offs between the gain in productivity/benefits and the gain in transacting costs. Hence farm will be efficient (sustainable) if it manages all transactions in most economical for the owner(s) way – that is situation when there is no activity which could be carried out with net benefit. If farm does not govern activity or transactions effectively, it will be unsustainable since it experiences high costs and difficulties using institutions (possibilities, restrictions) and carrying activity and transactions comparing to other feasible organization. In that case, there will be strong incentives for exploring existing potential (adapting to sustainable state) through reduction or enlargement of farm size, or via reorganization or liquidation of farm. Thus either alternative farm or non-farm application of resources; or farm expansion through employment of additional resources; or trade instead of internal use of owned land and labour; or taking over by (merger with) another farm or organization[4], will take place.

Furthermore, transacting modes and acceptable net benefits will vary according to individual's preferences, entrepreneurship ability, risk aversion, opportunity costs of owned resources etc. Depending on personality of resource owners and (transacting) costs and benefits of their coalition, different type of farm will be preferred - one-person farm (firm), family farm (firm), group farm or partnership (firm), cooperative farm, and corporative farms [Bachev, (2004)]. Expected benefits for farmers could range from monetary or non-monetary income; profit; indirect revenue; pleasure of self-employment or family enterprise; enjoyment of agricultural activities; desire for involvement in environment, biodiversity, or cultural heritage preservation; increased leisure and free time; to other non-economic benefits. Moreover, in specific institutional environment (legal framework, support policies, tradition, access to new technology, and level of transacting costs) various types of farm will have quite different effective horizontal and vertical boundaries. For instance, in transitional conditions of high market and institutional uncertainty, and inefficient property rights and contract enforcement system, most agrarian investments happened to be in regime of high specificity. As result (over)integrated modes such as low productive subsistent household and group farming, or large production cooperatives and agro-companies, dominates in Bulgaria and East Europe [Bachev, (2006)]. Alternatively, in more matured economies, where markets are developed and institutions stable, the agrarian assets are with more universal character. Therefore, farm borders are greatly determined by the family borders, and more market and mixed (contract rather than entirely integrated) forms prevail.

In order to assess farm's efficiency and sustainability we have to put individual transaction in the centre of analysis, and assess the level of associated costs and benefits. Major types farm transactions are associated with: the know-how supply, innovation supply, land supply, labour supply, inputs supply, service supply, financing, insurance supply, and marketing of services and products. The analysis is to embrace comparative efficiency of organization (governance) of every major transaction of farm. If significant costs (difficulties) of some type transacting in relation to the feasible alternatives is in place, then farm is to be considered as non-sustainable. Given the fact that alternative form often diminish one type while increasing other kind of transacting costs, and widespread application of complex modes (interlinking credit supply with inputs supply and/or marketing), the overall (internal and external) governance costs of the farm has to be taken into account.

Next, farm's *potential* (incentives, ability) *for adaptation* to the evolving market, institutional and natural environment through effective changes in the governing forms (saving on transacting costs) and the production structure (exploring technological possibilities for growth in productivity) is to be estimated. Thus if a farm does not have a potential to stay at or adapt to new more sustainable level(s) it would be either liquidated or transformed into another type of farm. For instance, if a farm

[4] In developed countries, the sustainable development has been associated with *disappearance* of traditional farming organization in major sectors (poultry, beef, and pig) which is *taken over* by or *integrated* into related industries (Sporleder).

faces enormous difficulties meeting institutional opportunities and restrictions (e.g. new quality and environmental standards, production quotas); or it has serious problems supplying managerial capital (as it is in a one-person farm when an aged farmer has no successor), or supply of needed farmland (a big demand for non-agricultural use of land), or funding activities (insufficient own finance, impossibility to sell equity or buy credit), or marketing output (a changing demand for certain products, strong competition with the imported products), then it would not be sustainable despite the high historical or current efficiency. Currently there are numerous unsustainable farms in most EU countries, which can hardly adjust to the fundamental changes in CAP and associated enhanced competition and new food safety, environmental, animal welfare etc. standards.

Our new approach makes it clear that sustainable development does not mean sustainable farms and agrarian structures [Bachev and Peeters, (2005)]. The farms and other modes of governance evolve (modernize, adapt, transfer, disappear) according to the changes in the social and natural environment. The development of the governance must be judged depending on the contribution of dominating and newly emerging forms of governance to achieving various (social, economic, environmental etc.) goals of sustainable development. Our approach also proves inadequacy of widely used indicators for productivity of "production costs and resources" for the assessment of the efficiency (viability, sustainability) of different farming organizations. Actually it is to be expected a significant differences in the rate of profitability on investments in an agro-firm (a "profit making organization") from the "pay-back" of expenditures and resources in a cooperative ("member oriented organization"), a public farm (a "non-for profit organization") or in a self-consistent farm (giving opportunity for productive use of otherwise "non-tradable" resources such as family labour, land etc.) [Bachev, (2004)].

7. The effective modes for public intervention

There is a big variety of possible forms for public intervention in the market and private activities. The comparative analysis is to extend to the public modes and include: *firstly,* the correspondence of public involvement to the real needs of development – identified needs for a third-party intervention from Table 2. *Secondly,* an assessment of comparative advantages of alternative modes for public involvements comprising all costs – direct (tax payer, assistance agency) expenses, and transacting costs of bureaucracy (for coordination, stimulation, mismanagement), and costs for individuals' participation in public modes (for information, paper works, payments of fees, bribes), and costs for community control over and for reorganization of bureaucracy (modernization and liquidation of public modes), and the (opportunity) costs of public inaction. And *third,* estimation of comparative efficiency of selected form and other practically possible (feasible) modes of governance of socially desirable activity such as partnership with private sector; property rights modernization etc. Accordingly, public intervention is to be initiated only if there is overall *net benefit* - when effects are greater than additional (individual and social) costs for the third-party involvement [Bachev, (2007)].

Depending on the uncertainty, frequency, and necessity for the specific investment of public involvement, there will be different the most effective forms. Table 2 presents an example with the public modes for effective interventions in the "environmental transactions"[5]. Principally, interventions with low uncertainty and assets specificity would require smaller Government organization (more regulatory modes; improvement of general laws and contract enforcement). When uncertainty and assets specificity of transactions increases special contract mode would be necessary-employment of public contracts for provision of private services, public funding of private activities, temporary labour contract for carrying out special public programs, leasing out public assets for private management etc. When transactions are characterized with high assets specificity, uncertainty and frequency then internal mode and bigger public organization would be needed – permanent public employment contracts, in-house integration of crucial assets in specialized public agency (company).

[5] Eco-transactions are associated with respecting eco-rights and improving eco-performance of individual agents

Table 2. Effective modes for public intervention in environmental transactions

Level of Uncertainty, Frequency, and Assets specificity				
Low ←----------→ *High*				
New property rights	**Regulations**	**Taxes**	**Assistance and support**	**Public provision**
Private rights on natural, biological, and environmental resources; Private rights for (non) profit management of natural resources; Tradable quotas (permits) for polluting; Private rights on intellectual property, origins, protecting ecosystem services; Rights to issue eco-bonds, shares; Private liability for polluting	Quotas for emissions, and use of products and resources; Regulations for trading of protection of ecosystem services; Regulations for introduction of foreign species, GM crops; Bans for use of certain inputs and technologies; Norms for nutrition and pest management; Regulations for water protection against pollution by nitrates; Regulations for biodiversity and landscape management; Licensing for water use; Quality and food safely standards; Standards for good farming practices; Mandatory (environmental) training; Certifications and licensing; Compulsory environmental labelling; Designating environmental vulnerable and reserve zone; Set aside measures; Inspections, fines and, ceasing activities	Tax rebates, exception, and breaks; Environmental taxation on emissions or products (pesticides, fertilizers); Levies on manure surplus; Tax or levies schemes on farming or export for funding innovations and extension; Waste tax	Recommendation and information; Demonstration; Direct payments and grants for environmental actions of farms, and farmers and community organizations; Preferential credit programs; Environmental contracts; Government purchases (water and other limited resources); Price and farm support for eco-production; Funding of environment and management training programs; Assistance in farm and eco-associations	Research and development; Extension and advise; Agro-market and know-how information; Agro-meteorological forecasts; Sanitary and veterinary control; Specialized (quasi) public agency or company; Eco-monitoring; Eco-foresight; Risk assessment; Pertaining "precaution principle"

In the beginning, existing and emerging problems (difficulties, costs, risks, failures) in organization of market and private transactions have to be specified. The appropriate government involvement would be to create environment for: decreasing uncertainty surrounding market and private transactions, increasing intensity of exchange, protecting private rights and investments, and making private investments less dependent. For instance, State establishes and enforces quality and safety standards for farm inputs and produces, certifies service providers, regulates employment relations, transfers water management rights to farms associations, sets up minimum farm-gate prices etc. All that facilitates will intensify (market and private) transactions and increases sustainability.

Next, practically possible modes for increasing appropriability of transactions have to be considered. The low appropriability is often caused by unspecified or badly specified private rights [Bachev, (2004)]. In some cases, the most effective government intervention would be to introduce

and enforce *new private property rights* – e.g. rights on natural, biological, and environmental resources; on issuing environmental bonds and shares; marketing and stock trading of ecosystem services protection; tradable quotas for polluting; private rights on intellectual agrarian property and origins etc. That would be efficient when the privatization of resources or the introduction (and enforcement) of new rights is not associated with significant costs (uncertainty, recurrence, and level of specific investment are low). That Government intervention effectively transfers the organization of transactions into the market and private governance, liberalizes market competition and induces private incentives (and investments) in certain activities (the relevant part in Table 2). For instance, tradable permits (quotas) are used to control the overall use of certain resources or level of a particular type of pollution. They give flexibility allowing farmers to trade permits and meet their own requirements according to their adjustment costs and specific conditions of production. That form is efficient when a particular target must be met and the progressive reduction is dictated through permits while trading allows the compliance to be achieved at least costs (through private governance). The later let also a market for environmental quality to develop.

In other instances, it would be efficient to put in place *regulations* for trade and utilization of resources and products – standards for labour (safety, social security), product quality, environmental performance, animal welfare; norms for using natural resources, GM crops, and (water, soil, air, comfort) contamination; ban on application of certain chemicals or technologies; foreign trade regimes; mandatory training and licensing of farm operators. Large body of environmental regulations in developed countries aim changing the farmers' behaviour and restricting the negative externalities. It makes producers responsible for the environmental effects of their products or the management of products uses (e.g. waste). This mode is effective when a general improvement of the performance is desired but it is not possible to dictate what changes (in activities, technologies) is appropriate for a wide range of operators and environmental conditions (high uncertainty and information asymmetry). When the level of hazard is high, the outcome is certain and the control is easy, and no flexibility exists (for timing or the nature of socially required result), then the bans or strict limits are the best solution. However, the regulations impose uniform standards for all regardless of the costs for compliance (adjustment) and give no incentives to over-perform beyond a certain level.

Sometimes, using the incentives and restrictions of the *tax system* would be the most effective form for intervention. Different sorts of tax preferences (exception, breaks, credits) are widely used to create favourable conditions for development of certain (sub)sectors and regions, forms of agrarian organization, segment of population, or types of activities. Environmental taxation on emissions or products (inputs or outputs) is also applied to reduce the use of harmful substances. For instance, taxes on pesticides and fertilizer are used in Scandinavian countries and Austria to decrease their application and environmental damaging impact. In Holland, levies on manure surplus were used based on levies for nitrogen and phosphorus surpluses above a levy free surplus. That system creates strong incentives to minimize leakages (and not just usage), and reduce flexibility to substitute taxable for non-taxable inputs. However, it is associated with significant administrative and private costs [ECOTEC, (2001)]. Environmental taxes impose the same conditions for all farmers using a particular input and give signals to take into account the "environmental costs" inflicted on the rest of society. Taxing is effective when there is close link between activity and the environmental impact, and when there is no immediate need to control pollution or to meet targets for reduction. Tax revenue is also perceived to be important to maintain budget of special (e.g. environmental) programs. However, appropriate level of charge is required to stimulate a desirable change in farmers' behaviour. Furthermore, nitrogen emission can vary according to conditions when nitrates are applied and attempting to reflect this in tax may result in complexity and high administrating costs. Besides, distribution impact of such taxes must be socially acceptable, and implications for international competitiveness also taken into account.

In some cases, a public *assistance and support* to private organizations is the best mode for intervention. Large agrarian and rural support and development programs are widely used in all industrialized countries. They let "proportional" development of agriculture, improvement of farmers' welfare ("income parity"), and in some instances undesired effects such as over-intensification, environmental degradation, and market distortions. Public financial support for environmental actions is most commonly used instrument for improving environment performance of farmers in EU and other developed countries. It is easy to find a justification for public payments as compensation for the provision of "environmental service" by farmers. All studies shows that value placed upon landscape

exceed greatly the costs of running schemes. However, share of farms covered by various agri-environmental support schemes is not significant [EC, (2005)]. That is a result of voluntary (self-selection) character of this mode which does not attract farmers with highest environment enhancement costs (intensive and damaging producers). In some cases, low-rate of farmers' compliance with environmental contracts is a serious problem [Dupraz *et al.*, (2004)]. The later cannot be solved by augmented administrative control (enormous enforcement costs) or introducing bigger penalty (politically and juridical intolerable). A disadvantage of "payment system" is that once introduced it is practically difficult ("politically unacceptable") to be stopped when goals are achieved or there are funding difficulties. Moreover, withdraw of subsidies may lead to further environmental harm since it would induce adverse actions such as intensification and return to conventional farming. Main critics of subsidies are associated with their "distortion effect", negative impact on "entry-exit decisions" from polluting industry, unfair advantages to certain sectors in the country or industries in other countries, not considering total costs (transportation and environmental costs, and "displacement effect" in other countries). It is estimated that agro-environmental payments are efficient in maintaining current level of environmental capital but less successful in enhancing environmental quality [EC, (2005)].

Often providing *public information, recommendations, training and education* to farmers, rural agents, and consumers are most efficient form. In some cases, pure *public organization* (in-house production, public provision) will be most effective as in the case of agrarian research and education, agro-market information, agro-meteorological forecasts, border sanitary and veterinary control etc.

Usually, specific modes are effective if they are applied alone with other modes of public intervention. Necessity of combined intervention (*governance mix*) is caused by: complementarities (joint effect) of individual forms; restricted potential of some less expensive forms to achieve a certain (but not entire) level of socially preferred outcome; possibility to get extra benefits ("cross-compliance" requirement for public support); particularity of problems to be tackled; specific critical dimensions of governed activity; uncertainty (little knowledge, experience) associated with likely impact of new forms; practical capability of Government to organize (potential to control, implement) and fund different modes; and dominating (right, left) policy doctrine [Bachev, (2007)].

Besides, level of effective public intervention depends on the kind of problem. There are public involvements which are to be executed at local (ecosystem, community) level, while others require regional or nationwide governance. And finally, there are activities, which are to be coordinated at international (regional, European, worldwide) level due to strong necessity for trans-border actions (needs for cooperation in natural resources management, for exploration of economies of scale/scale, for governing of spill-over)[6] or consistent (national, local) government failures. Very frequently the effective governance of many problems (risks) requires multilevel governance with a system of combined actions at various levels involving diverse range of actors and geographical scales.

Public (regulatory, inspecting, provision) modes must have built special mechanisms for increasing competency (decrease bounded rationality, powerlessness) of bureaucrats, beneficiaries, interests groups and public at large as well as restricting possible opportunism (opportunity for cheating, interlinking, abuse of power, corruption) of public officers and other stakeholders. That could be made by training, introducing new assessment and communication technologies, increasing transparency (e.g. independent assessment and audit), and involving experts, beneficiaries, and interests groups in the management of public modes at all levels [Bachev, (2007)]. Furthermore, applying "market like" mechanisms (competition, auctions) in public projects design, selection and implementation would significantly increase the incentives and decrease the overall costs.

Principally, *pure* public organization should be used as a last resort when all other modes do not work effectively [Williamson, (1996)]. "In-house" public organization has higher (direct and indirect) costs for setting up, running, controlling, reorganization, and liquidation. What is more, unlike the market and private forms there is not an automatic mechanism (such as competition) for sorting out less effective modes[7]. Here public "decision making" is required which is associated with high costs and time, and often influenced by strong private interests (power of lobbying groups, policy makers and associates, employed bureaucrats) rather than efficiency. Along with development of general

[6] Recent epidemic of avian infection is a good example in that respect.

[7] It is not rare to see highly inefficient but still "sustainable"public organizations around the world.

institutional environment ("The Rule of Law") and the measurement, communication etc. technologies, efficiency of pro-market modes (regulation, information, recommendation) and contract forms would get bigger advantages over internal less flexible public arrangements [Bachev, (2007)].

Usually *hybrid modes* (public-private) are much more efficient than pure public forms given coordination, incentives, and control advantages. In majority of cases, involvement of farmers and other beneficiaries increase efficiency - decrease asymmetry of information, restrict opportunisms, increase incentives for private costs-sharing, and reduce management costs [Bachev, (2007)]. For instance, a hybrid mode would be appropriate for carrying out the supply of non-food services by farmers such as preservation and improvement of biodiversity, landscape, and cultural heritages. That is determined by farmers information superiority, strong interlinks with traditional food production (economy of scope), high assets specificity to farm (farmers competence, high cite-specificity of investments to farm, land, ecosystem), and spatial interdependency (need for cooperation of farmers at regional or wider scale), and not less important – farm's origin of negative externalities. Furthermore, enforcement of most labour, animal welfare, biodiversity etc. standards is often very difficult or impossible. In all these cases, stimulating and supporting (assisting, training, funding) private voluntary actions are much more effective then mandatory public modes in terms of incentive, coordination, enforcement, and disputing costs [Bachev, (2004)].

Anyway, if there is strong need for third-party public involvement but effective government intervention is not introduced in due time, agrarian "development" would be substantially deformed. Thus Government failure is also possible and often prevails. In Bulgaria for instance, there have been great number of bad examples for Government under- and over-interventions in agrarian sector during post-communist transition now [Bachev, (2006)]. Consequently, primitive and uncompetitive small-scale farming; over-integrated and personalized exchanges; ineffective agrarian bureaucracy; blocking out of all class of transactions (supply of innovation and extension, long-term credit, infrastructure and environmental goods); and development of a large informal sector, all they come out as result.

8. Conclusion

Deepening specialization and exchanges between agents opens enormous opportunities for economic growth. However, it is also associated with significant transaction costs which might disturb sustainable development. In traditional (*Neoclassical Economics*) framework with no transacting costs there is only one mechanism for governance. "Free market prices" (and market competition) effectively coordinate and stimulate entire activity of resource owners, entrepreneurs, and consumers. Accordingly all farms constantly "adapt" to price movements and social demand being equally efficient and sustainable. Rare cases of market "failures" are also recognized but perfect "government intervention" is seen as remedy. All that leads to interrupted global sustainable development.

In the real economy, there are additional important factors affecting individual choices and sustainability (institutions and transacting costs), and a great variety of effective governing mechanisms. Institutional environment is a crucial factor, which eventually determines the type of development. Individual agents tend to govern available resources in most economical way adapting to institutional environment and minimizing total (production and transaction) costs. Depending on personal characteristics of agents and critical attributes of each activity, there will be spectrum of effective structure for organization of resources, activities and exchanges – some will be governed by "invisible market hand", other by special contract forms, and some by "visible manager hands" or within complex hierarchies, other supported by third-party. At any given period of time, farms and agrarian organizations of various type and size would persist in agriculture - subsistent, family, cooperative, corporative etc. Furthermore, sustainable development does include a fundamental modernization of farming structures – size adjustment, transformation, coalition, and disappearance of farms.

Our new framework helps us better understand factors for sustainable development and "Government's role" as well. Analyses of transaction costs identify immense range of "market failures" associated with unspecified or badly specified property rights; inefficient system for enforcement of absolute and contracted rights; high uncertainty and dependency of activity, low appropriability of rights. Economic agents deal with market deficiency developing different non-market forms for effective governance (contracts, internal modes, collective actions). Nonetheless, private sector also "fails" to safeguard individual rights and carry out certain activities at effective

scale. That is particularly true for human and eco-rights, technological and infrastructural development, environmental conservation activity etc. Thus there is a strong need for a third-party public involvement in market and private transactions though institutional modernization, assistance, regulation, hybrid or public organization. However, diverse forms of public interventions are with unequal efficiency and most efficient one is to be selected taking into account overall transaction costs and contribution to sustainable development. What is more, most public interventions increasingly require concerted actions (multilateral and multilevel governance) at local, regional, national, and transnational scale. Yet "government failure" is also possible and inappropriate involvements, under or over-regulations, mismanagement etc. are widespread. Sustainability is significantly compromised when market and private sector fails, and no effective public intervention takes place.

5. References

[1] Bachev, H., (2004), *Efficiency of Agrarian Organizations*, in: *Farm Management and Rural Planning*, No 5, Kyushu University, Fukuoka, pp.135-150.

[2] Bachev, H., (2006), *Governing of Bulgarian Farms – Modes, Efficiency, Impact of EU Accession*, in: *Agriculture in the Face of Changing Markets, Institutions and Policies: Challenges and Strategies*, J.Curtiss, A.Balmann, K.Dautzenberg, K.Happe (editors), IAMO, Halle, pp.133-149.

[3] Bachev, H., (2007), *Governing of Agrarian Sustainability*, in: *ICFAI Journal of Environmental Law*, Vol. VI, No 2, ICFAI University, Hyderabad, pp.7-25.

[4] Bachev, H. and Labonne, M., (2000), *About Agrarian Innovations*, INRA, Montpellier.

[5] Bachev, H. and Peeters, A., (2005), *Framework for Assessing Sustainability of Farms*, in: *Farm Management and Rural Planning*, No 6, Kyushu University, Fukuoka, pp. 221-239.

[6] Coase, R., (1960), *The Problem of Social Costs*, in: *Journal of Law and Economics 3*, pp.1-44.

[7] Dupraz, P., Latouch, K., and Bonnieux F., (2004), *Economic Implications of Scale and Threshold Effects in Agri-environmental Processes*, paper presented at the 90 EAAE Seminars, Rennes.

[8] EC, (2005), *Agri-environment Measures, Overview on General Principles, Types of Measures, and Application. Evaluation of Measures applied to Agriculture Studies.* European Commission, Directorate General for Agriculture and Rural Development.

[9] ECOTEC, (2001), *Study on the Economic and Environmental Implications of the Use of Environmental Taxes and Charges in the EU and its Member Sates*, ECOTEC Research and Consulting, Brussels.

[10] Furuboth, E., and Richter, R., (1998), *Institutions and Economic Theory: The Contribution of the New Institutional Economics*, The University of Michigan Press, Ann Arbor.

[11] Hagedorn, K., (editor) (2002), *Environmental Cooperation and Institutional Change*, Edward Edgar, Cheltenham.

[12] Hardin, G., (1968), *The Tragedy of the Commons*, in: *Science*, Vol. 162, no.3859, pp.1243 - 1248.

[13] North, D., (1990), *Institutions, Institutional Change and Economic Performance*. Cambridge University Press, Cambridge.

[14] Olson, M., (1969), *The Logic of Collective Actions: Public Goods and the Theory of Groups*. Harvard University Press, Cambridge.

[15] Pigou, A., (1920), *Economics of Welfare*. Macmillan and Co, London.

[16] Raman, S., (2006), *Agricultural Sustainability. Principles, Processes and Prospect*, The Haworth Press Inc., New York.

[17] Sporleder, T., (1992), *Managerial Economics of Vertically Coordinated Agricultural Firms*, in: *American Journal of Agricultural Economics*, Vol. 74, No 5, pp.1226-1231.

[18] VanLoon, G., Patil, S., and Hugar, L., (2005), *Agricultural Sustainability: Strategies for Assessment*, SAGE Publications, London.

[19] Williamson, O., (1996), *The Mechanisms of Governance*, Oxford University Press, New York.

MEASURING PERSISTENCE ON CONSUMPTION IN PORTUGAL

José **BELBUTE**, António **CALEIRO**
University of Évora, **Portugal**,
Department of Economics and CEFAGE-UÉ
jbelbute@uevora.pt, caleiro@uevora.pt

Abstract:
The paper deals with the detection and measurement of the level of persistence on aggregate private consumption in Portugal, USA, European Union and EuroZone as well as on some categories of aggregate consumption in Portugal. By the use of a non-parametric methodology applied to monthly data (1992-2007) it is concluded that aggregate consumption in Europe (both European Union and Euro Zone) is more persistent than in the USA and in Portugal. In particular, the relatively lower degree of persistence shown by the consumption in Portugal can be beneficial for the effectiveness of the countercyclical fiscal and monetary policies that are currently being implemented to overcome the current economic crisis. Our results also suggest that consumption of durables is less persistent, also being more volatile. This result is important in the explanation of the severity of the current economic crisis in Portugal.

Keywords: Consumption, Persistence, Portugal

JEL Classification: C14, C22, E21

1. Introduction and motivation

One of the most important stylized facts of the short-run business cycle is the relative smooth behavior of consumption relatively to gross domestic product (GDP). Consumption is less volatile than GDP and, in fact, it is the most stable component of aggregate demand. The life-cycle, permanent income and random walk literature explains this pattern by assuming that households try to smooth their life-time consumption path by adopting a forward looking behaviour. This behaviour in conjunction with rational expectations on the part of consumers makes it possible that current consumption expenditures incorporate anticipated future levels of income. In particular, the Hall (1978)'s consumption walk theory suggests that all future incomes are correctly anticipated and incorporated into current wealth so that current consumption fully reflects this information. Since all that it is known of the future is already taken into account in the present assessment of wealth, only true surprises can alter wealth and thus consumption. Put differently, in accordance to those theories, consumption displays some degree of inertia or persistence.

However the smoothness of consumption varies significantly by type of consumption. Typically, non-durables and services tend to exhibit a smooth(er) behaviour regardless the economic fluctuations but most of the business cycle fluctuation in consumption expenditures are due to durables.

In the current economic situation, which is characterized by some inertia of the economic aggregates at undesirable levels (for instance, low levels of growth and high levels of unemployment) the persistence that some relevant variables, namely consumption, exhibit is of obvious importance. This is so because, on the one hand, the presence of inertia can substantially change the response of households to a policy shock. This is particularly problematic for the formulation and the effectiveness of the present countercyclical policies that act through consumption. On the other hand, persistence can reduce the incidence, the length, and the severity of shocks and of changes of the economic conditions. Traditionally macroeconomic policies play the dominant role in smoothing the business cycle but the effectiveness of those policies depends upon the ability of the economic system to absorb (as displacement from the baseline) the shock and to return to the baseline, in short it depends on the economy's resilience. To the best of our knowledge, the analysis of persistence has focused on inflation and not on real variables such as consumption.

Persistence, broadly understood as a characteristic of a phenomenon that despite the actions taken in order to invert its evolution (or simple unfavorable states of nature), leads to an observation of some inertia in the evolution of that phenomenon, should be an interesting issue in Economics both from the theoretical and empirical points of view. In other words, persistence is the speed with which a variable returns to baseline (its previous level) after, say, a shock, i.e. some event (for instance, a

macroeconomic policy measure) that provoked an increase (or decrease) in that variable [Marques, (2004)].

There are two dimensions involved in the concept. First, the speed with which a variable respond to a shock and the speed with which it converges to its baseline. In particular, consumption persistence is associated with the speed with which consumption responds to a shock and with its convergence to its previous level. In this context, consumption is said to be (highly) persistent if, following a shock, it converges slowly to its previous level.

Plainly, the evolution of the economic variables is different under persistence and this fact cannot be ignored either from a theoretical approach or from a policy perspective. As a matter of fact, the literature on the economic importance of persistence is inexplicably scarce, taking into consideration that the first studies are from the beginnings of the 1980s and only quite recently authors showed a remarkable interest in the phenomenon. Clearly, this is in accordance to the analysis of the current situation in the world, in which the persistence of the recession is an issue to be tackled. Still, is also the almost inexplicable the lack of studies, especially through the use of appropriate techniques to detect and measure the level of persistence, aiming at testing the evidence of the phenomenon? In particular, for Portugal, to the best of our knowledge, there is no study of this nature.

The first studies that explicitly considered the importance of persistence were of macroeconomic nature. Following the seminal work of Taylor (1980), which has shown that staggered wage-setting can lead to persistence in employment after a temporary shock; the authors focused their attention in the causes of persistence in the major macroeconomic aggregates. For some time, both staggered wage-setting and staggered price-setting were considered as being similar in the process of generation of persistent real effects of monetary shocks. For instance, Rotemberg and Woodford (1997) argue that output persistence can be due to price staggering. However some other authors, namely Huang and Liu (2002), have argued that price staggering and wage staggering models are crucially different in such a way that the staggered price mechanism is not capable of generating persistence, while the staggered wage mechanism is indeed important in generating persistence. On the other hand, in Ascari (2003) it is argued that the ability of a model to produce output persistence is not due to price or wage staggering mechanisms but, in fact, is due to the particular characteristics of the model, namely the behaviour assumed by firms and by the labour force.

Further to the analysis of the causes, the consequences of persistence were analysed by several authors. For instance, in the standard real business cycle models, it was inherent a lack of propagation of causal effects of temporary shocks, such as policy measures, to output. Not surprisingly, some criticisms were made based upon the alleged inability of standard real business cycle models to reproduce the evolution of output shown in real world conditions [*see* Cogley and Nason, (1995)]. As a response to this critique, RBC models were augmented with persistence mechanisms capable of explaining the (strong) persistence of output that could be observed in reality [see, among others, Bouakez and Kano, (2006), and Maury and Tripier, (2003)]. In fact, this response did not close the debate, in which the possibility of monetary policy shocks affecting aggregate output is central. Indeed the persistence of shocks to aggregate output has been, still is (and most probably it will be for some time) one of the issues predominantly subject to investigation.

For the empirical evidence that monetary policy shocks can have permanent effects on aggregate output (or unemployment) there has been proposed some theoretical explanations, notably imperfect information about nominal fluctuations, namely about prices, and short-run nominal rigidities, such as sticky prices. For instance, considering nominal price stickiness and imperfect information, Kiley (2000) has shown that both factors allow nominal shocks to propagate in the cycle, but that only sticky prices propagate the real effects of nominal shocks. However, Wang and Wen (2006) argue that whether or not price rigidity is responsible for output persistence is not a theoretical question, but an empirical one.

Following another approach, Jonsson (1997), Lockwood (1997) and Svensson (1997), analysed the consequences of output or unemployment persistence on the establishment of inflation contracts. Other interesting consequence of output persistence is that it may turn upside down the political business cycle, which, in its typical form, is associated with depressions at the beginning of the mandate followed by pre-election inflationary expansions. This consequence on the pattern of the typical political business cycle is shown to exist by Gärtner (1996) who considers a model with adaptive expectations and a linear (in output)-quadratic (in inflation) policy objective function [see

also Caleiro, (2009)]. Furthermore, Gärtner (1999) also gives some credit to the output persistence hypothesis from an empirical point of view.

Quite recently, it was registered an increase of interest in analyzing the persistence of output, as well as of inflation, considering its relationship with other aspects such as the degree of openness of the economies [see Guender, (2006) for a recent analysis], the exchange-rate regime (for instance, Giugale and Korobow (2000) argue that the speed of recovery of real output following an interest rate shock is higher under a flexible exchange rate regime, than under a fixed exchange rate regime) or the structural change on the behaviour of consumers, firms or policy-makers. In particular, the causes, the patterns and the political implications of inflation persistence have been elected the main objective of the Inflation Persistence Network (IPN), an ECB team of economists undertaking joint research on inflation persistence in the euro area and in its member countries.

Having said that, it is important to mention, at this stage of our analysis, that the previous studies confirm the persistence of output (or unemployment) being an up-to-date relevant issue. Still, besides a lack in the methods of empirically measuring the persistence, a gap in the literature is evident. A microeconomic foundation of macroeconomic persistence, for instance in consumption, is missing, notwithstanding the existence of a well-built theoretical set of analyses based upon the hypothesis of intertemporal dependent preferences in the form of a process of habit-formation. Indeed, in a seminal work Dusenberry (1949) called the attention for the importance of past consumption on the current consumption of households. There are several ways to build intertemporal dependent preferences.

In the literature on macroeconomics (both closed and open) and finance the Ryder and Heal (1973) model for the so-called habit formation has been used to solve a number of puzzles (namely the well known "equity premium puzzle" – Constantinidies, 1990) as it conforms well to the short run low volatility of consumption. They showed that when instantaneous well-being is determined not only by the current level of consumption (the level effect) but also by it (average) past level (the habit or persistence effect) throughout a process of "learning-by-consuming", the intertemporal dependent preferences might be a sufficient reason to cause a cyclical behaviour of consumption along its time path. This hypothesis, built upon the importance of habits, has been tentatively used to explain the behaviour of the growth rate and of the savings rate during a recession [Carroll, (2000), Wendner, (2002)]. On the other hand, Belbute and Brito (2008) show that the presence of the inertial effect can not only lower the long run equilibrium level of natural capital and the growth rate of the economy, but also reduce the effectiveness of an environmental policy that is meant to improve environmental quality as well as sustainability.

The paper intents to measure the persistence in consumption by the use of statistical techniques, in a univariate approach, which are adequate to measure the degree of persistence over time. The paper has the following structure. Section 2 presents and justifies the methodology that will be used to measure the persistence on consumption, distinguishing the durable from the non-durable goods cases. Section 3 offers the results. Section 4 concludes.

2. The methodology to measure the persistence

As the previous section has shown, when distinguishing non-durable from durable goods, consumption in Portugal has gone through evolutions that are apparently distinct. A question that then comes up is the following: does this distinction has to do with the fact that consumption persistence is different in those two cases? We start giving an answer to this question by the use of statistical techniques, in a univariate approach, which are adequate to measure the degree of persistence over time.

Since some time ago, some authors have started to pay attention to persistence in (economic) time series as a phenomenon that reveals to be crucial to policy measures, namely at the inflation level. In fact, to the best of our knowledge, all the applications of the statistical techniques to measure the level of persistence have considered the inflation rate case [Hondroyiannis and Lazaretou, (2004); Levin and Piger, (2002); Marques, (2004); Minford *et al.*, (2004); and Pivetta and Reis, (2004)]. We propose to apply those statistical techniques, developed by Andrews and Chen (1994), Dias and Marques (2005) and Marques (2004), to the different kinds of consumption in Portugal. In this sense, the novelty in the approach is supposed to be a contribution to filling the gap in the literature on consumption.

Starting with a simple definition, consumption persistence is the speed with which consumption returns to baseline (its previous level) after, say a shock, *i.e.* some event (for instance, a macroeconomic policy measure) that provoked an increase (or decrease) in consumption. This definition, in other words, implies that the degree of consumption persistence is associated with the speed with which consumption responds to a shock. When the value is high, consumption responds quickly to a shock. On the contrary, when the value is small, the speed of adjustment by consumption is low. To put it clearer, a variable is said to be the more persistent the slower it converges or returns to its previous level, after the occurrence of a shock. Persistence is, thus, inversely related with the concept of mean reversion.

Quantifying the response of consumption to a shock is indeed important not only because it may allow assessing the effectiveness of economic policy measures but also because it may, indeed, show at what time is more essential to act, through those measures, in order to overwhelm a harmful effect of a shock over consumption. By definition, quantifying the response of consumption to shocks implies evaluating the persistence of consumption.

As the estimates of persistence at time *t* will express how long we expect that a shock to consumption will take to die off (if ever), given present and *past* consumption, authors have proposed to obtain those estimates by the use of *autoregressive models*. As it is well known, a univariate AR(k) process is characterised by the following expression:

$$f_t = \mu + \sum_{j=1}^{k} a_j f_{t-j} + \varepsilon_t \tag{1}$$

where f_t denotes the consumption rate at moment *t*, which is explained by a constant μ, by past values up to lag *k*, as well as by a number of other factors, whose effect is captured by the random variable ε_t. Plainly, (1) can also be written as:

$$\Delta f_t = \mu + \sum_{j=1}^{k-1} \delta_j \Delta f_{t-j} + (\rho - 1) f_{t-1} + \varepsilon_t \tag{2}$$

where

$$\rho = \sum_{j=1}^{k} \alpha_j \tag{3}$$

and

$$\delta_j = -\sum_{i=j+1}^{k} \alpha_i \tag{4}$$

In the context of the above model (1), or (2), persistence can be defined as the speed with which consumption converges to its previous level after a shock in the disturbance term that raises consumption at moment *t* by 1%.[1]

The techniques allowing for measuring the persistence are based on the analysis of the autoregressive coefficients α_j in (1) or (2), which are subject to a statistical estimation. Plainly, the most simple case of the models (1) or (2) is the so-called AR(1) model, that is:

$$f_j = \mu + \alpha_1 f_{t-j} + \varepsilon_t \tag{5}$$

[1] Given that the persistence is a long-run effect of a shock to consumption, this concept is intimately linked to a concept usually associated to autoregressive models such as (1) or (2), *i.e.* the impulse response function of consumption, which, in fact, is not a useful measure of persistence given its infinite length.

Clearly, the variable ε_t in this kind of models has a particular importance given that it may be associated with policy measures leading to a shock in the consumption rates. A positive shock, at moment *t*, will significantly last for future moments the higher is the autoregressive coefficient α_1. Following this approach, Andrews and Chen (1994) proposed the sum of the autoregressive coefficients, $\rho = \sum_{j=1}^{k} \alpha_j$, as a measure of persistence.[2] The rationale for this measure comes from realizing that for $|\rho| < 1$, the cumulative effect of a shock on consumption is given by $\frac{1}{1-\rho}$.

Unfortunately, the procedure above described is subject to a problem that is likely to occur in practice, in particular when analysing macroeconomic variables, which is the possible existence of factors leading to a non stationary behaviour of the time series as the result of, for instance, the existence of a trend in the data.[3] In fact, the existence of a unit root in the data generation process makes it impossible to accept the results from a traditional OLS estimation. This fact poses a problem from the viewpoint of measuring persistence in consumption rates, which means that we decided to follow another approach as it will be described below.

Recently, Marques (2004) has suggested a non-parametric measure of persistence, γ, based on the relationship between persistence and mean reversion. In particular, Marques (2004) suggested using the statistic:

$$\gamma = 1 - \frac{n}{T} \tag{6}$$

where *n* stands for the number of times the series crosses the mean during a time interval with $T + 1$ observations[4], to measure the absence of mean reversion of a given series, given that it may be seen as the unconditional probability of that given series not crossing its mean in period *t*.

As Dias and Marques (2005) have shown, there is a one-to-one correspondence between the sum of autoregressive coefficients, ρ, given by (3) and the non-parametric measure, γ, given by (6), when the data is generated by an AR(1) process, but such a one-to-one correspondence ceases to exist once higher order autoregressive processes are considered. In other words, only in the particular case of a first-order autoregressive model, AR(1), either one of the two measures can be used to quantify the level of persistence, as both transmit the same result, but as soon as higher order autoregressive models are considered, *i.e.*, AR(*k*) with $k \geq 2$, the monotonic relationship between ρ and γ no longer exists, therefore leading to possibly crucial differences when measuring persistence in the series.

Moreover, using the alternative measure of persistence, γ, given by (6), has some important advantages (Dias and Marques, 2004).[5] Given its nature, such measure of persistence does not impose the need to assume a particular specification for the data generation process, therefore does not require a model for the series under investigation to be specified and estimated.[6] This is so given that γ is indeed extracting all the information about the persistence from the data itself. As it measures how often the series reverts to its means and (high/low) persistence exactly means that, after a shock, the series reverts to or crosses its means more (seldom/frequently), one does not need to specify a particular form for the data generation process. To put differently, the less a time series cut its mean, the greater will be the degree of persistence and thus the higher the value of γ.

[2] Authors have, indeed, proposed other alternative measures of persistence, such as the largest autoregressive root, the spectrum at zero frequency, or the so called half-life. For a technical appraisal of these other measures see, for instance, Marques (2004) and Dias & Marques (2004).

[3] For instance, after the application of augmented Dickey-Fuller tests in order to test for the stationarity of the time series for consumption, we could not reject the existence of a unit root in all of them.

[4] The ratio *n/T* gives the degree of mean reversion.

[5] The statistical properties of γ are extensively analysed in Marques (2004) and Dias & Marques (2005).

[6] In technical terms, this means that the measure is expected to be robust against potential model misspecifications and given its non-parametric nature also against outliers in the data.

3. Consumption persistence in Portugal

3.1. An international perspective

Let us begin by considering quarterly data of aggregate consumption for Portugal, USA, European Union and EuroZone, for the period of 1995 – 2008 as plotted in Figure 1.

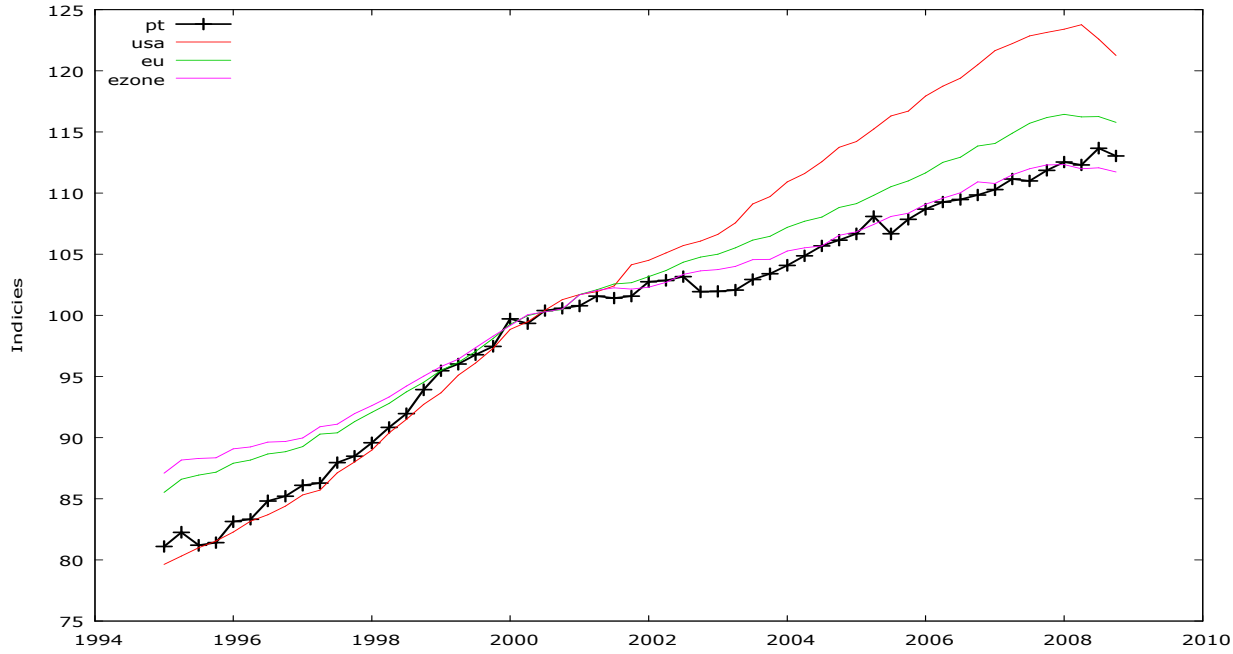


Figure 1. Total private Consumption for Portugal, USA, European Union and EuroZone
Quarterly Volume index, OECD reference year (2000), seasonally adjusted, (1995-2008)

Given that for all trajectories the hypothesis of non-stationarity cannot be rejected (either confirmed by the ADF tests and by the slow decline of all the ACF's - see the annex A), therefore making it impossible to rely upon the OLS estimation of AR models, it resulted clear that, in order to measure the persistence in the consumption rates, one should rely on the use of the non-parametric measure γ was given by (6).

Clearly, in order to compute the estimative for each kind of consumption, the mean of each series has to be computed. As suggested in Marques (2004), a time varying mean is more appropriate than the simple average for all the period under investigation. In our case we followed that suggestion by using the well known Hodrick-Prescott (HP) filter in order to compute the mean.

As it is well known, the HP filter defines the trend or mean, g_t, of a time series, f_t, as the solution to the minimisation problem:

$$\min_{g_t}\left\{\sum_{t=1}^{T}\left(f_t-g_t\right)^2+\lambda\sum_{t=2}^{T-1}\left[\left(g_{t+1}-g_t\right)-\left(g_t-g_{t-1}\right)\right]^2\right\} \qquad (7)$$

i.e. the HP-filter seeks to minimise the cyclical component $\left(f_t-g_t\right)$ subject to a smoothness condition reflected in the second term. The higher the parameter λ, the smoother will be the trend and the less deviations from trend will be penalised. In the limit, as λ goes to infinity, the filter will choose $\left(g_{t+1}-g_t\right)=\left(g_t-g_{t-1}\right)$, for $t=2,...,T-1$, which just amounts to a linear trend. Conversely, for $\lambda=0$, we get the original series.

Plainly, the HP-filter is a very flexible device since it allows us to approximate many commonly used filters by choosing appropriate values of λ. Given that the data is of quarterly frequency, authors

have suggested using values for λ around 1600.[7] In particular, Ravn and Uhlig (2002) suggest the use of $\lambda = 1600$. Considering this value, the results are shown in Figures 2 and 3.

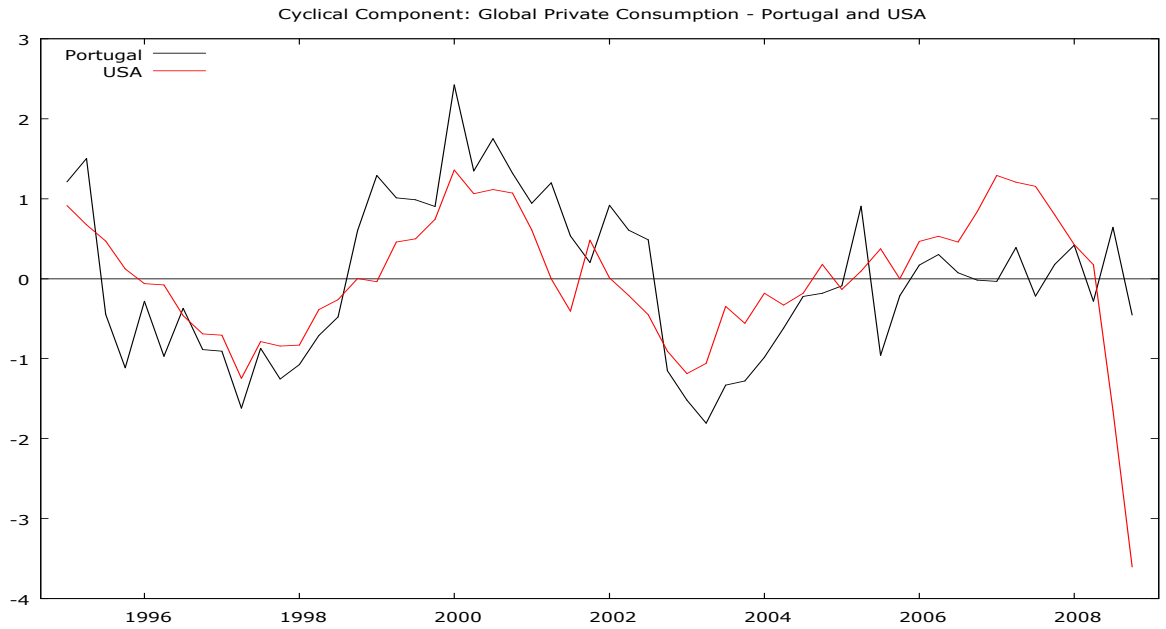


Figure 2. The cyclical component for total private consumption: Portugal and USA

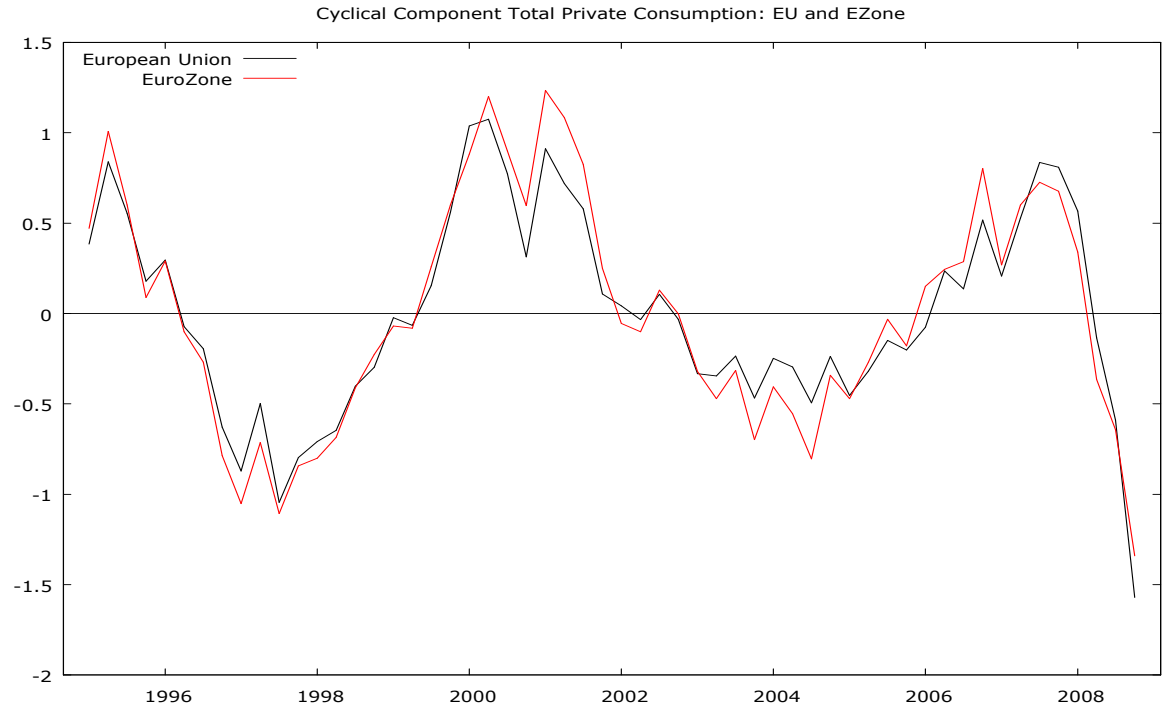


Figure 3. The cyclical component for total private consumption: Portugal and USA

For a better comparison see Figure 4.

[7] In order to check the robustness of the results we considered some other values when computing the estimates of γ. In qualitative terms the results were not changed.

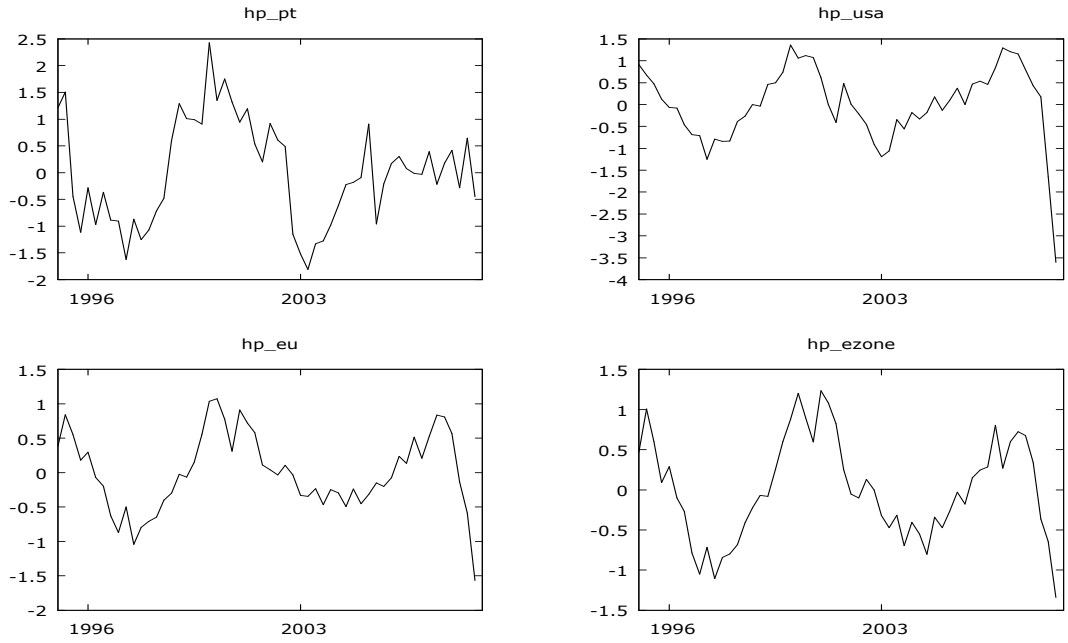


Figure 4. The cyclical component for total private consumption: Portugal and USA

Table 1 summarizes the main results for persistence. Clearly, aggregate private consumption in Portugal is less persistent (but more volatile) than in the USA, the Eurozone and the European Union meaning that after a shock, the Portuguese private consumption deviates quickly from its trend and tend to stay away from it (above or below) less time than in USA or EU/EZ. In the present context of world recession, this feature is quit relevant because in terms of economic forecasting, especially over the short run, persistence implies that one can fairly confidently predict what happens to consumption when it is below or above its trend. Being less persistent, one might expect that, in the short run, private consumption will stay above the trend less time in Portugal than in the USA, EU and EZone. The results also suggest that private consumption will stay above its trend more time in the European Union and in the Eurozone than in the USA.

Table 1. Persistence and Volatility in Aggregate private consumption

Country	n	γ	Volatility (1)
Portugal	13	0,7679	0,00778
European Union	7	0,8750	0,00349
EuroZone	7	0,8750	0,00373
USA	11	0,8036	0,00520

Volatility = Standard Error ($ln(x_t/x_{t-1})$)

3.2. How persistent is consumption of durable and non-durable goods in Portugal

Let us now consider data for consumption of durables and non-durables for the period 1992-2007 (data from the Portuguese Central Bank (BP) plotted in figure 5. As expected, durables are much more volatile than non-durables (and as well as aggregate private consumption). In fact, the standard deviation of the percentage deviations in non-durables is about 19% of that for durables and only about 11% of that for the aggregate private consumption.

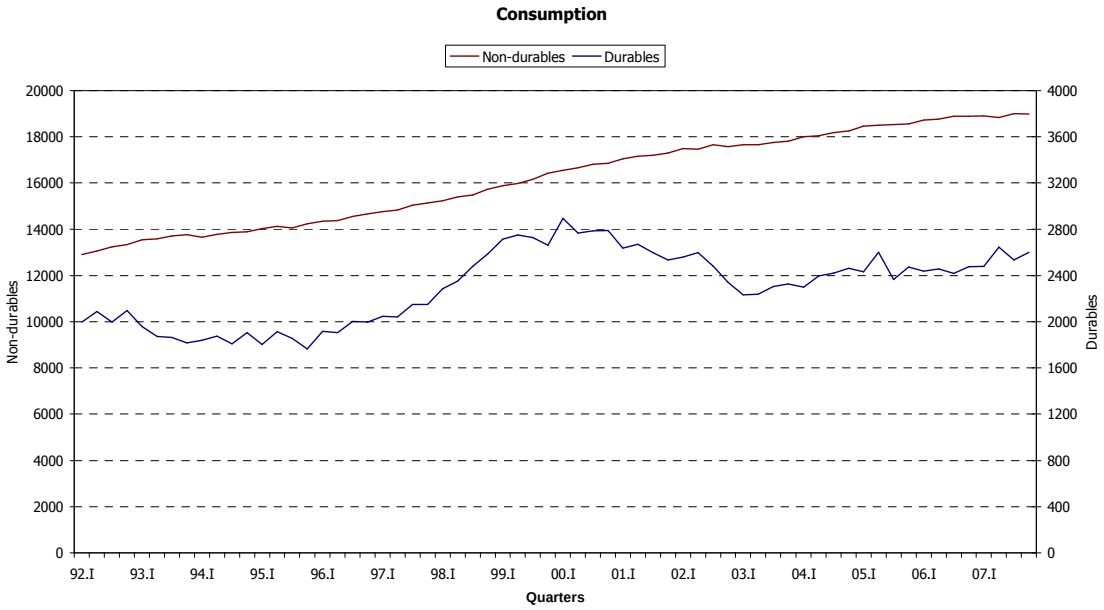


Figure 5. The dates

As with the total private consumption, also with all the disaggregated consumption trajectories the hypothesis of non-stationery cannot be rejected for both the ADF and the graphical inspection of the ACF's and PACF's tests (see the annex B). Again, under this circumstance, the measure of the persistence in the consumption rates in Portugal can be done by the use of the non-parametric measure γ as given by (6).

Figure 6 shows the cyclical component of durables and non-durables in order to facilitate the comparison. For the case of non-durables the cyclical component crosses the mean 10 times, which gives a measure of persistence, γ =0.84127. For the case of durables the cyclical component crosses the mean 12 times, which gives a measure of persistence, γ =0.809524.

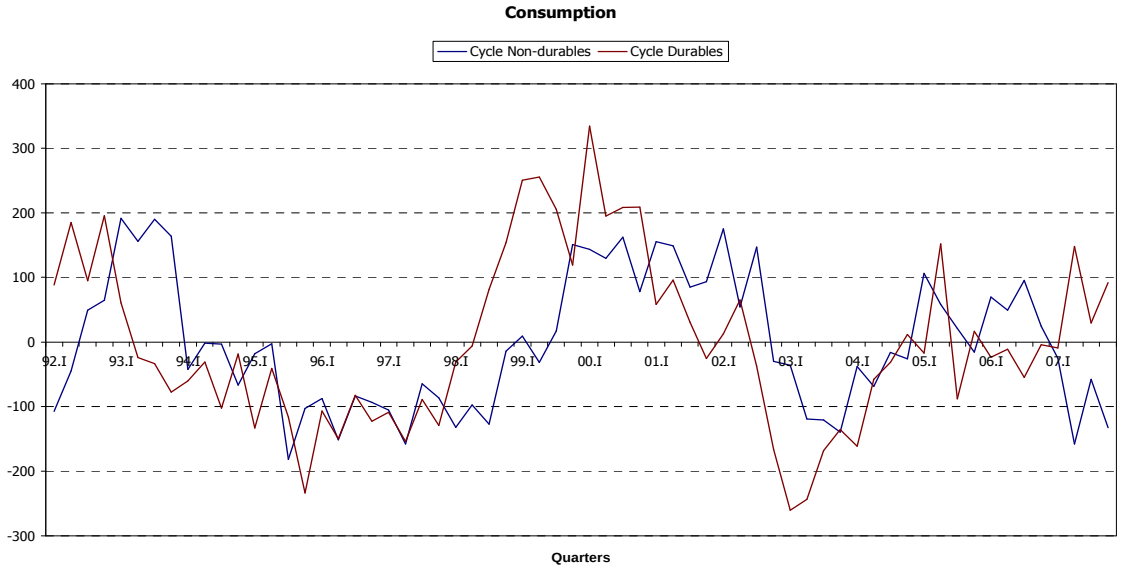


Figure 6. The cyclical component for consumption of durables and non-durables

Another source of the data makes it possible to consider consumption disaggregated into three categories: food, non-food and services and durables, for the period 1995 until 2007. Figures 7, 8 and 9 show the cyclical components of these categories.

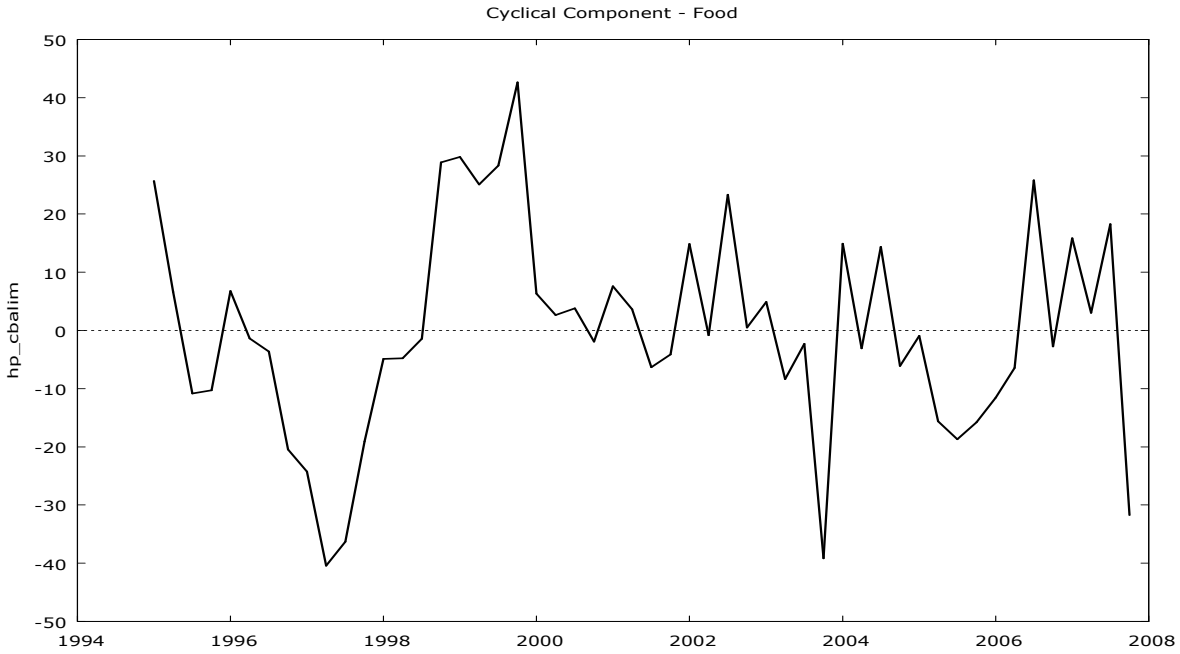


Figure 7. The cyclical component for consumption of food

For the case of food the cyclical component crosses the mean 19 times, which gives a measure of persistence, $\gamma = 0.6346$.

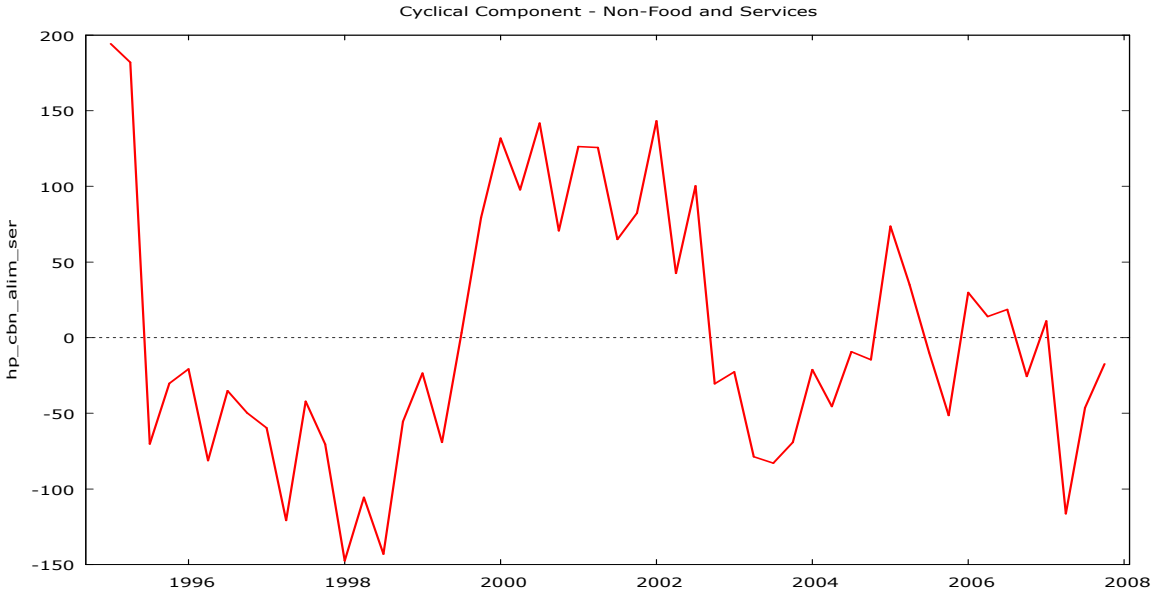


Figure 8. The cyclical component for consumption of non-food and services

For the case of non-food and services the cyclical component crosses the mean 9 times, which gives a measure of persistence, $\gamma = 0.8269$.

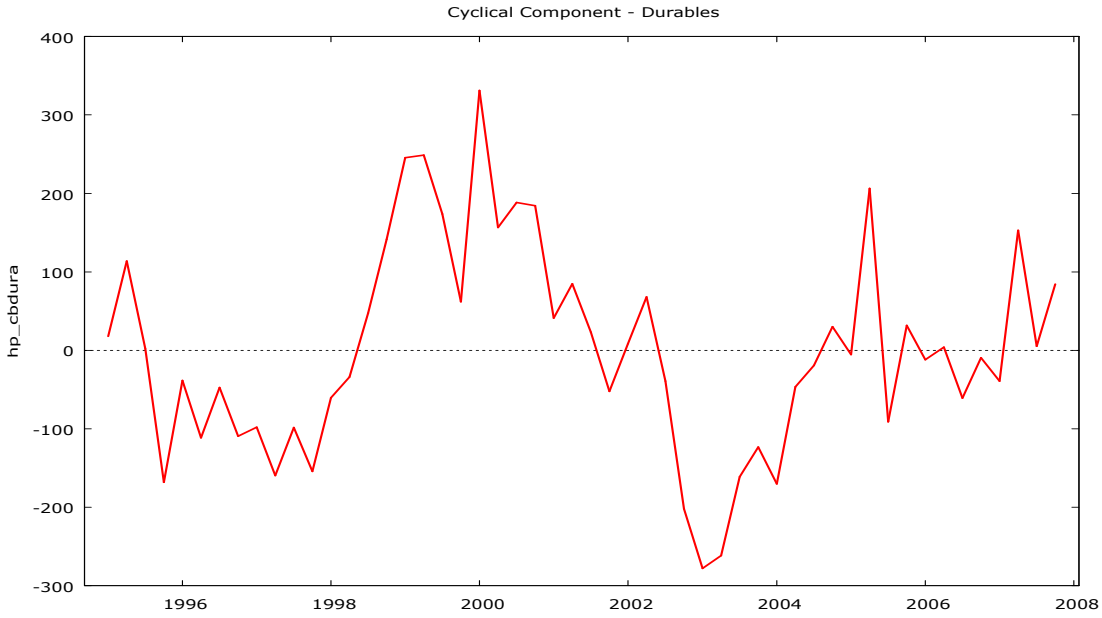


Figure 9. The cyclical component for consumption of durables

For the case of durables the cyclical component crosses the mean 14 times, which gives a measure of persistence, γ =0.7308.

To sum up, durables are more volatile and also less persistent. Note that non-durables volatility in BP data is different from the one in the INE data. Table 2 summarizes the results.

Table 2. A summary of the results

		Portuguese Central Bank (BP)			National Statistical Office (INE)		
		Volatility Index		γ	Volatility Index		γ
		V_1	V_2		V_1	V_2	
Aggregate Consumption			0,0065	0,8594		0,0107	0,8077
Non-durables	Food				0,0053	0,0051	0,6346
	Non-food and services	0,0052	0,0065	0,8413	0,0059	0,0071	0,8269
Durables		0,0404	0,0567	0,8095	0,0473	0,0549	0,7308

V_1 – Standard Error ($\ln(x_t/x_{t-1})$)

V_2– Standard Error [(x-x*)/x*] with x* being the trend.

4. Conclusions and directions for further research

This paper has explored the issue of consumption persistence in Portugal. One important stylized facts of the short-run business cycle is the relative smooth behaviour of consumption relatively to GDP. Indeed, private consumption is less volatile than GDP and, in fact, the most stable component of aggregate demand. Put differently, consumption exhibits a high degree of inertia or persistence.

Persistence is associated with the speed with which a variable responds to a *shock* and with its convergence (or its return) to its previous level. Consumption is said to be (highly) persistent if, following a shock, it converges slowly to its previous level. In other words, a high degree of persistence in consumption means that when consumption is above (or below) the trend, it tends to stay above (below) the trend for more time. This property is quite relevant for economic forecasting as it allow to confidently predicting the short run behaviour of consumption. However, countercyclical policies that act through the consumption may be ineffective if consumption is very persistent. Households may be fearful about the future, may identify uncertainty factors that might restrain their consumption expenditures or even look current fiscal policies as equivalent to future taxes.

Our conclusions suggest that aggregate consumption in Europe (both European Union and Euro Zone) is more persistent than in the USA and in Portugal. In the particular context of the world economic crisis, our results suggest that private consumption will stay less time above the trend in Portugal than in USA and than in European Union and/or European Zone. Furthermore, the relatively lower degree of persistence shown by the consumption in Portugal can be beneficial for the effectiveness of the countercyclical fiscal and monetary policies that are currently being implemented to overcome the current economic crisis.

When data is disaggregated between durables and non-durables, our results suggest that the consumption of durables are less persistent, being also more volatile as expected. This apparently non-intuitive result lies on the nature of the "consumption" of durables. In fact, as happens with capital goods, they yield a flow of consumption services over its lifetime (they are not consumed simultaneously with consumer spending and enjoyment). However, data includes "expenditures" on durable goods, not its services or enjoyment, whose pro-cyclically and high volatility nature is well known in the literature.

5. References

[1] Andrews, D., Chen, W.K., (1994), *Approximately Median-Unbiased Estimation of Autoregressive Models*, in: *Journal of Business and Economic Statistics*, Vol. 12, pp. 187 - 204.

[2] Ascari, G., (2003), *Price/Wage Staggering and Persistence: A Unifying Framework*, in: *Journal of Economic Surveys*, Vol. 17, pp. 511 - 540.

[3] Belbute, J., Brito, P., (2008), *On sustainable endogenous growth under intertemporally dependent preferences*, in: *ICFAI Journal of Environmental Economics*, Vol. 6, No. 4, pp. 7 - 28.

[4] Bouakez, H., Kano, T., (2006), *Learning-by-doing or habit formation?* in: *Review of Economic Dynamics*, Vol. 9, pp. 508 - 524.

[5] Caleiro, A., (2009), *How upside down are political business cycles when there is output persistence*, in: *Research in Economics*, Vol. 63, No. 1, pp. 22 - 26.

[6] Carroll, C., (2000), *Saving and growth with habit formation*, in: *American Economic Review*, Vol. 90, No. 3, pp. 341 - 355.

[7] Cogley, T., Nason, J., (1995), *Output dynamics in Real-Business-Cycle models*, in: *American Economic Review*, Vol. 85, pp. 492 - 511.

[8] Constantinides, G., (1990), *Habit formation: a resolution of the equity premium puzzles*, in: *Journal of Political Economy*, Vol. 98, 519 - 543.

[9] Dias, D., Marques, C.R., (2005), *Using Mean Reversion as a Measure of Persistence*, European Central Bank, Working Paper Series No. 450, March.

[10] Dusenberry, C., (1949), *Income-Consumption Relations and their Implications*, in: Metzler, *Income, Employment and Public Policy. Income, Saving and the Theory of Consumer Behaviour.* Cambridge: Harvard University Press.

[11] Gärtner, M., (1996), *Political business cycles when real activity is persistent*, in: *Journal of Macroeconomics*, Vol. 18, pp. 679 - 692.

[12] Gärtner, M., (1999), *The election cycle in the inflation bias: evidence from the G-7 countries*, in: *European Journal of Political Economy*, Vol. 15, pp. 705 - 725.

[13] Giugale, M.; Korobow, A., (2000), *Shock Persistence and the Choice of Foreign Exchange Regime: An Empirical Note from Mexico*, World Bank Policy Research Working Paper No. 2371.

[14] Guender, A., (2006), *Stabilising Properties of Discretionary Monetary Policies in a Small Open Economy*, in: *The Economic Journal*, Vol. 116, pp. 309 - 326.

[15] Hall, R., (1978), *Stochastic Implications of Life Cycle-Permanent Income Hypothesis: Theory and Evidence*, in: *Journal of Political Economy*, Vol. 86, pp. 971 - 987.

[16] Hondroyiannis, G.; Lazaretou, S., (2004), *Inflation Persistence during Periods of Structural Change: An assessment using Greek data*, European Central Bank, Working Paper Series No. 370, June.

[17] Huang, K., Liu, Z., (2002), *Staggered Price-Setting, Staggered Wage-Setting and Business Cycle Persistence*, in: *Journal of Monetary Economics*, Vol. 49, pp. 405 - 433.

[18] Jonsson, G., (1997), *Monetary politics and unemployment persistence*, in: *Journal of Monetary Economics*, Vol. 39, 303 - 325.

[19] Kiley, M., (2000), *Endogenous Price Stickiness and Business Cycle Persistence*, in: *Journal of Money, Credit and Banking*, Vol. 32, pp. 28 - 53.

[20] Levin, A.T., Piger, J.M., (2002), *Is Inflation Persistence Intrinsic In Industrial Economies?* Federal Reserve Bank of St. Louis, Working Paper 2002-023E, October.

[21] Lockwood, B., (1997), *State-contingent inflation contracts and unemployment persistence*, in: *Journal of Money, Credit, and Banking*, Vol. 29, pp. 286 - 299.

[22] Marques, C.R. (2004), *Inflation Persistence: Facts or Artefacts?* Banco de Portugal, Working Paper 8, June.

[23] Maury, T.P., Tripier, F., (2003), *Output persistence in human capital-based growth models*, in: *Economics Bulletin*, Vol. 5, pp. 1 - 8.

[24] Minford, P., Nowell, R.; Sofat, P.; Srinivasan, N., (2004), *UK Inflation Persistence: Policy or Nature*, Cardiff University, *mimeo*.

[25] Pivetta, F., Reis, R., (2004), *The Persistence of Inflation in the United States*, Harvard University, mimeo.

[26] Ravn, M.O., Uhlig, H., (2002), *On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations*, in: *The Review of Economics and Statistics*, Vol. 84, No. 2, May, pp. 371 - 376.

[27] Ryder, H., Heal, G., (1973), *Optimum growth with intertemporally dependent preferences*, in: *Review of Economic Studies*, Vol. 40, pp. 1 - 33.

[27] Rotemberg, J., Woodford, M., (1997), *An Optimization Based Econometric Framework for the Evaluation of Monetary Policy*, In: Rotemberg, J., Bernanke, B., (Eds.). *NBER Macroeconomics Annual 1997*, Cambridge (MA): The MIT Press, pp. 297 - 346.

[28] Svensson, L., (1997), *Optimal inflation targets, "conservative" central banks, and linear inflation contracts*, in: *American Economic Review*, Vol. 87, pp. 98 - 114.

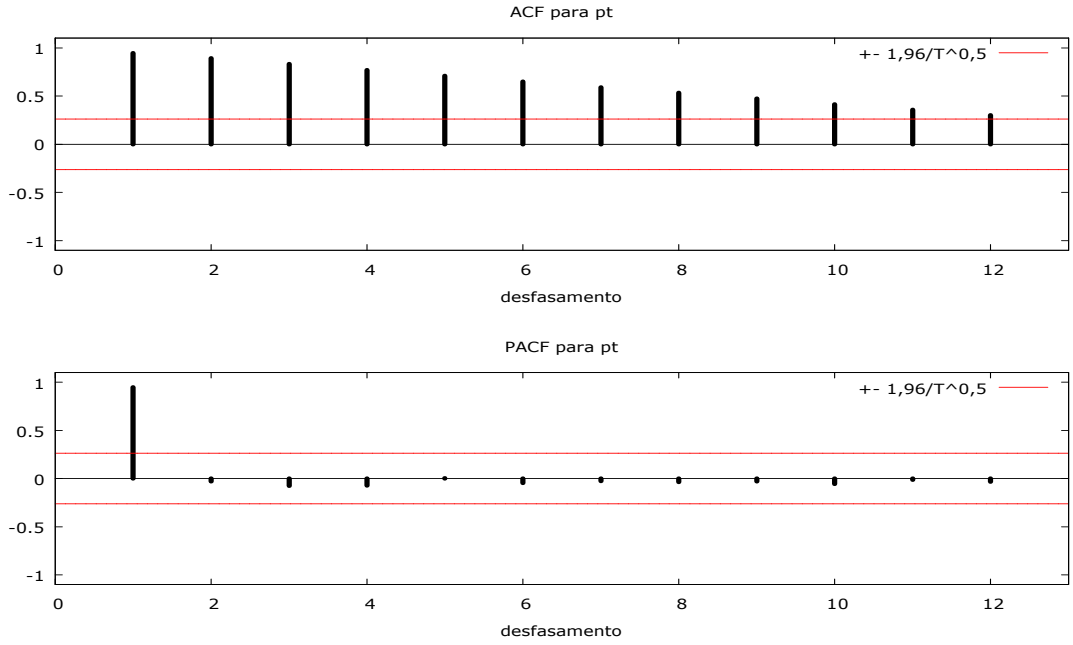
[29] Taylor, J., (1980), *Aggregate dynamics and staggered contracts*, in: *Journal of Political Economy*, Vol. 88, pp. 1 - 23.

[30] Wang, P.-f., Wen, Y., (2006), *Another look at sticky prices and output persistence*, in: *Journal of Economic Dynamics and Control*, Vol. 30, pp. 2533 - 2552.

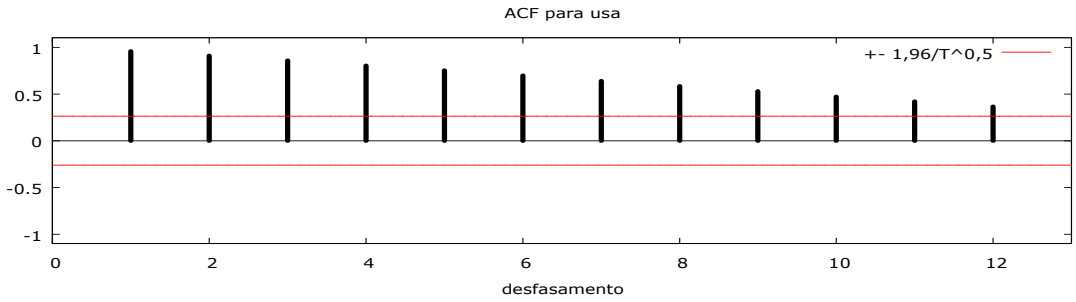
[31] Wendner, R. (2002), *Capital accumulation and habit formation*, in: *Economics Bulletin*, Vol. 4, pp. 1 - 10.

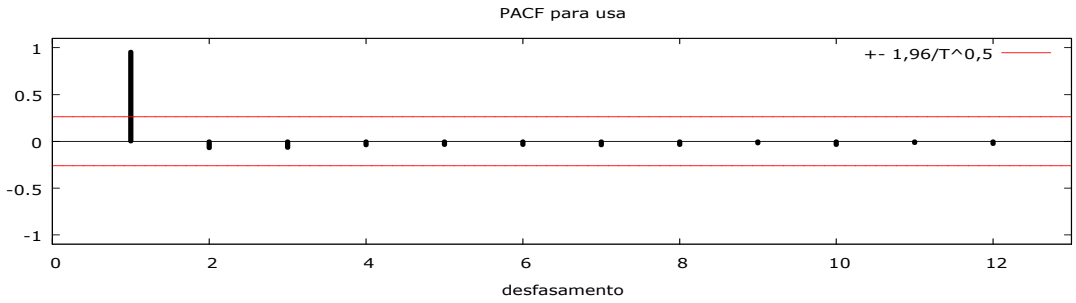
Annex A– The ACF and PACF of the series of total private consumption

a) Portugal

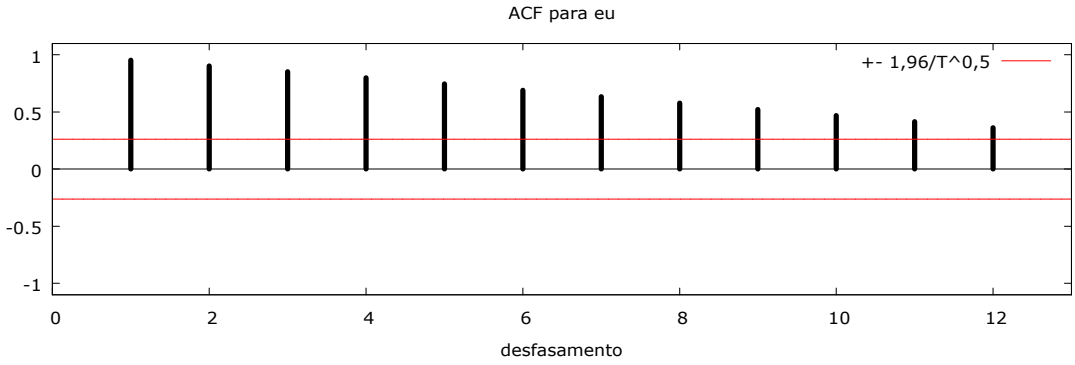


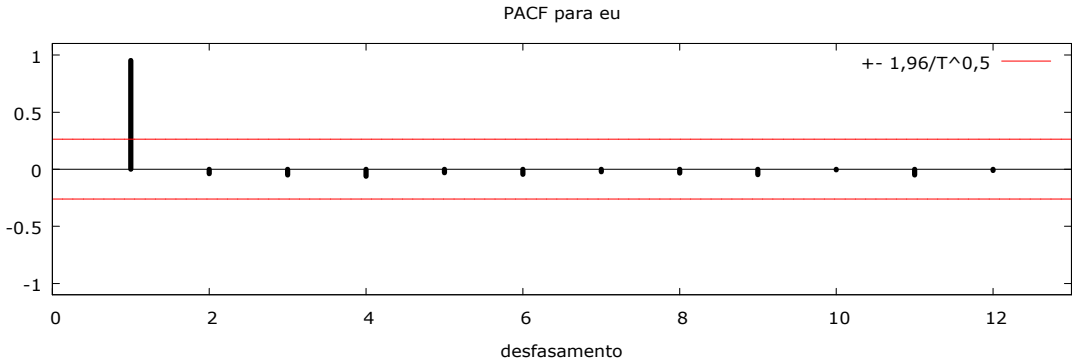
b) USA



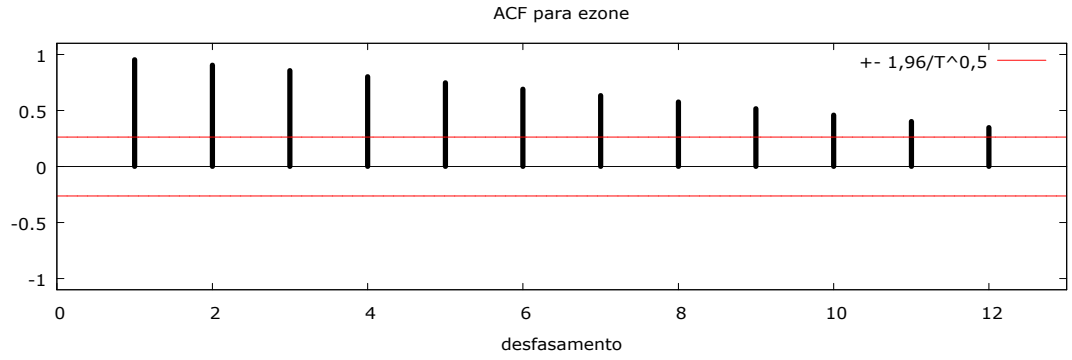


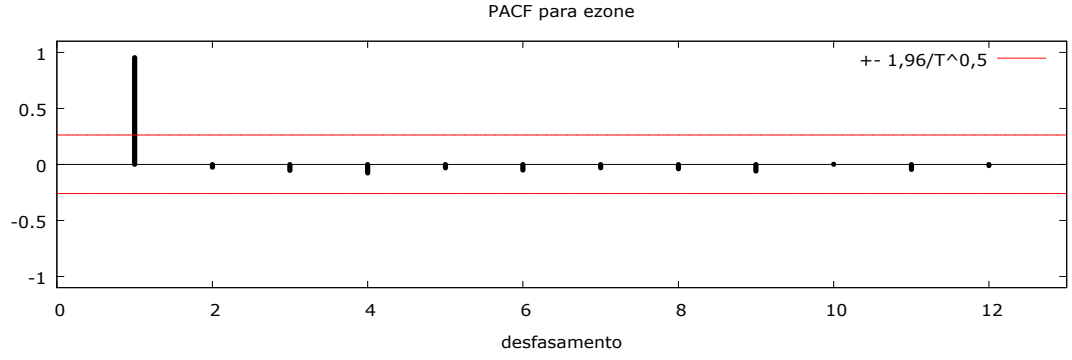
c) European Union


ACF para eu
+- 1,96/T^0,5
1
0.5
0
-0.5
-1
0
2
4
6
8
10
12
desfasamento


PACF para eu
+- 1,96/T^0,5
1
0.5
0
-0.5
-1
0
2
4
6
8
10
12
desfasamento

d) EuroZone


ACF para ezone
+- 1,96/T^0,5
1
0.5
0
-0.5
-1
0
2
4
6
8
10
12
desfasamento


PACF para ezone
+- 1,96/T^0,5
1
0.5
0
-0.5
-1
0
2
4
6
8
10
12
desfasamento

Annex B – The ACF and PACF of the series of disaggregated private consumption

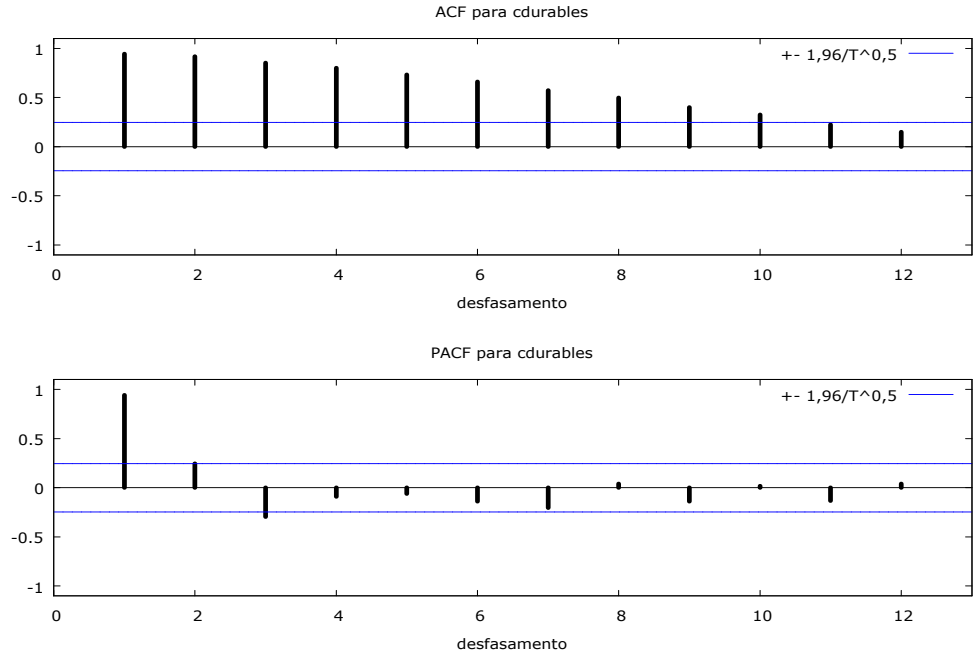


Figure 1. Durables

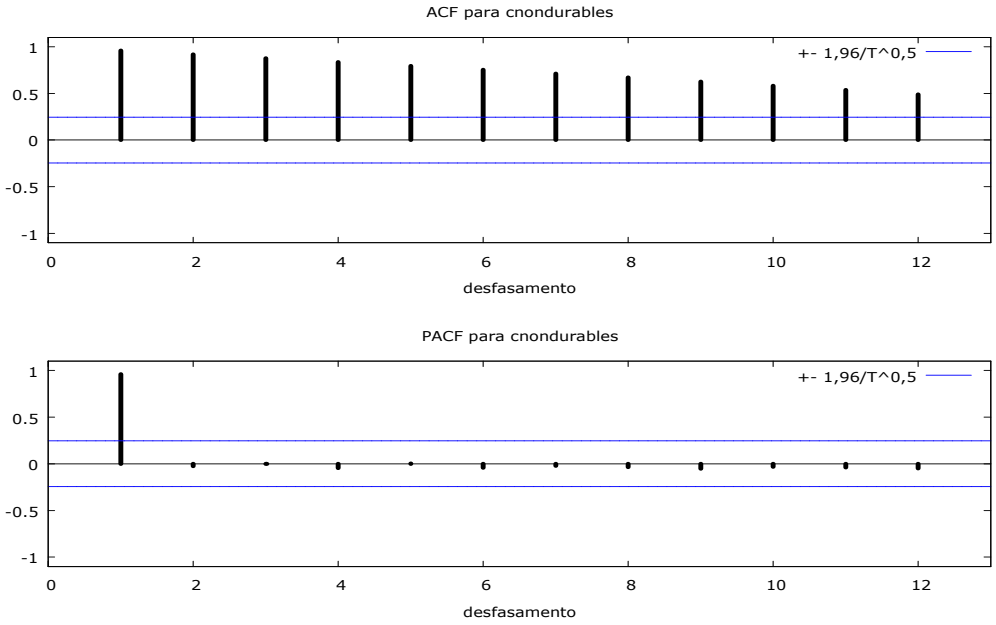


Figure 2. Non durables

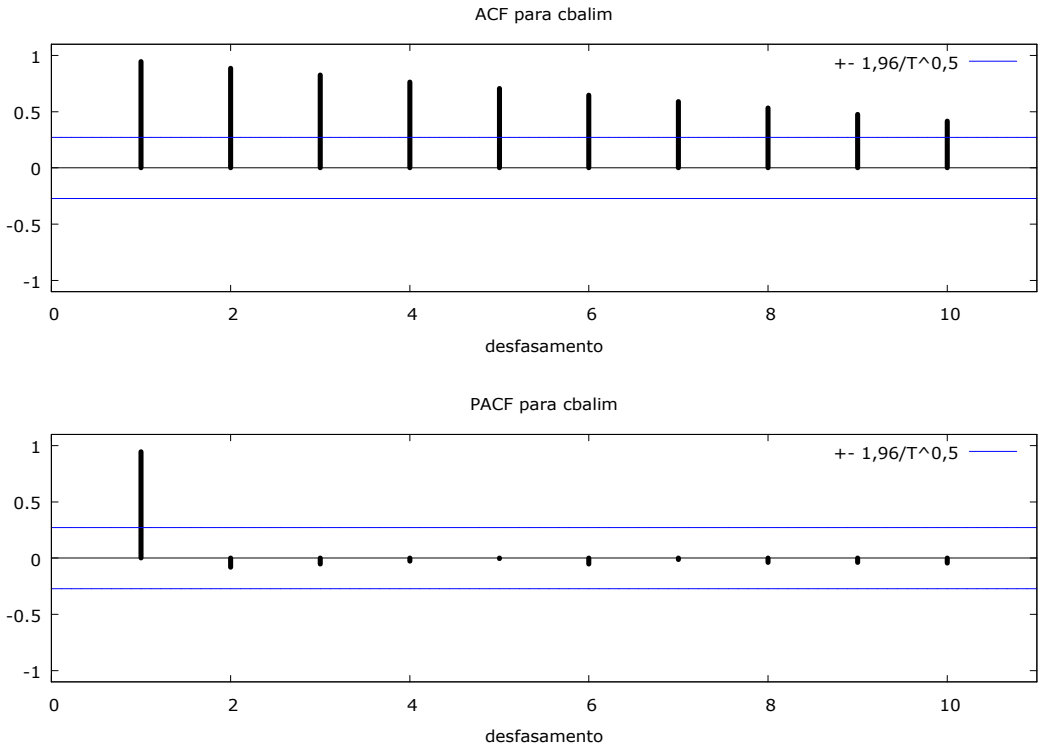


Figure 3. Food

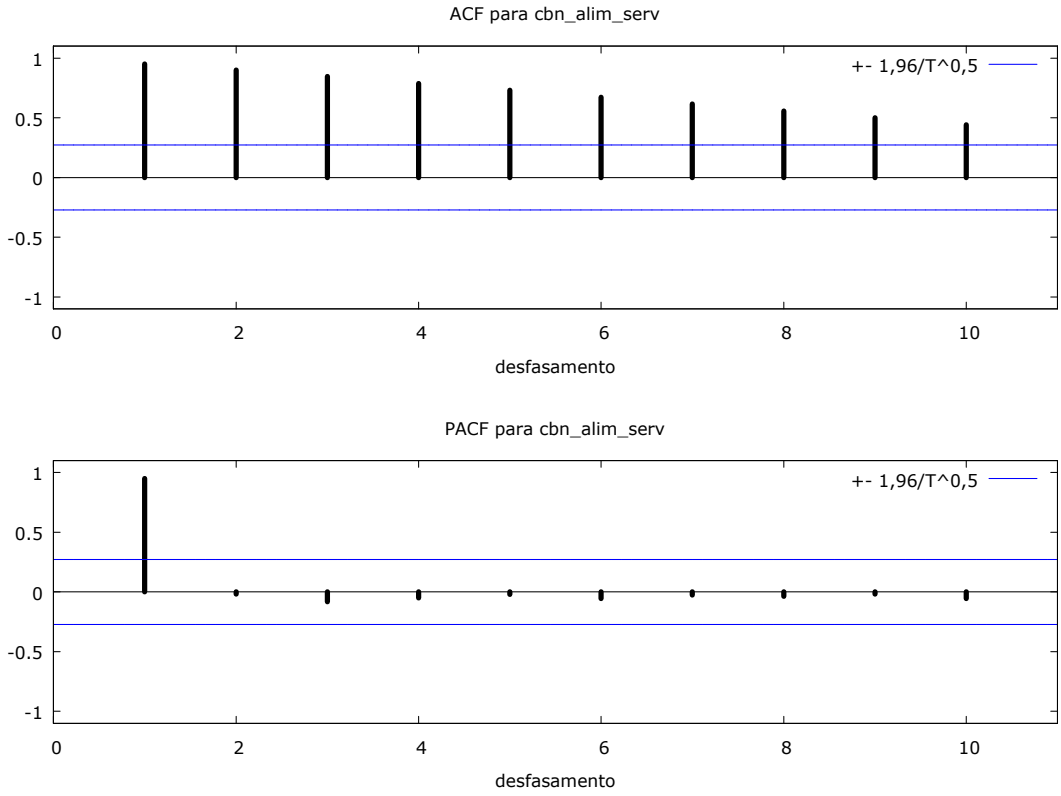


Figure 4. Non food and services

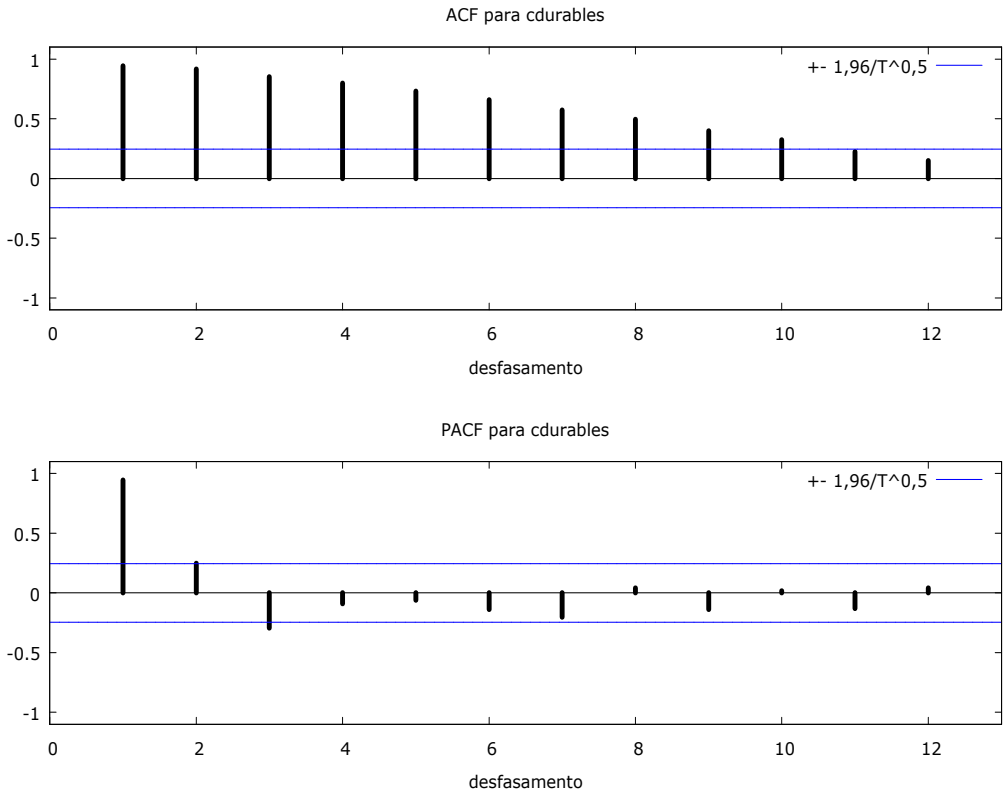


Figure 5. Durables

SOME ONTOLOGICAL ISSUES OF THE REA FRAMEWORK IN RELATION TO ENTERPRISE BUSINESS PROCESS

Frantisek **HUNKA**, Miroslav **HUCKA**
University of Ostrava, **Czech Republic**
frantisek.hunka@osu.cz
Josef **KASIK**
VSB-Technical University of Ostrava, **Czech Republic**
josef.kasik}@vsb.cz
Dominik **VYMETAL**
Silesian University in Opava, **Czech Republic**
vymetal@opf.slu.cz

Abstract:
The aim of the paper is to describe using REA framework to model enterprise planning not only at the operational level but also at the policy level. Using policy level enlarges the possibility of the models on the base of the REA framework because the policy level in this way represents metalevel of the model. The policy level of the REA framework itself is comprised both of the entities related by typification, grouping and policy relationships and of the Commitment entity with the fulfilment relationship. This entity may be viewed as either a sub layer or a middle layer of the REA framework. The Commitment entity belongs to the fundamental entities of the policy level but has some specifications that are expressed by the fulfilment relationship. This many-to-many relationship forms the link to the operational level. In the paper we discuss the problem and suggest some solution that moves the Commitment entity closer to the typification and grouping semantic abstractions.

Keywords: REA ontology, enterprise business process, semantic abstractions

JEL Classification: L15, L23, M11, O21

1. Introduction

Two core semantic abstractions represent policy level at the REA framework by [Geerts, McCarthy, (2006)]: *typification* and *grouping*. In short, the main task of the policy level is to give answer to the questions: *what could, should* or *must* occur. In general the policy level also gives the answers to the questions: *what is planned* or *scheduled*. The later questions are answered by the *Commitment* entity, which has the main relation to the *Event* entity at the operational level. At the first sight the *Commitment* entity contains planned or scheduled information while the *Event* entity comprises the actual information, which may be the same or different from the scheduled information.

From this point of view, it could seem that the typification semantic abstraction can be also applied in this case. However, the relationship between the *Commitment* entity and the *Event* entity has also broader meaning that is expressed by a many-to-many relationship. There are not only cases when a *Commitment* entity is related to one or more *Event* entities but there also exist cases when a single *Event* entity is related to several *Commitment* entities. This gives the reason, why the relationship between the *Commitment* entity and the *Event* entity is created by the fulfilment relationship. The grouping semantic abstraction is used when set-level characteristics are of interest and may even create an integral part together with the typification semantic abstraction. By this semantic abstraction a collection of individual entities may be specified with respect to some common properties.

2. Method of Advancement

Enterprise Planning Model, which is used for illustration of our approach, is composed of three parts: Material Purchase Model, Production Planning Model and Sales Model. We illustrate only Production Model in Figure 1, which is the most important part. In the schematic representation of the models we were also inspired by the Purchase Order pattern and the Schedule pattern described in [Hruby, (2006)].

The Production Planning Model consists of five decrement commitments: *Labor Requisition, Workplace Requisition, Tool Requisition, Part Requisition* and *Material Requisition* paired through conversion reciprocity with the increment commitment *Production Line*. The *Part Requisition*

commitment and *Material Requisition* commitment are promises by a Warehouse Clerk to make a specific amount of *Part Types* and *Material Type* available to the Worker. The *Tool Requisition* commitment is a promise by the Warehouse Clerk that tools of specified *Tool Types* will be available to the Worker, and a promise by the Worker to deliver the tools back. The *Labour Requisition* commitment is a promise by the Worker to the Supervisor to consume worker's Labour in a specific time. The *Workplace* commitment is a promise by the Supervisor to the Worker that a specified *Workplace* will be available to the Worker in specific time. Each commitment either uses reservation or consumes reservation of the adequate resource type. The operational level of the model is closely bounded with the policy level and contains five adequate Economic Events corresponding earlier described commitments. Resource types have their counterparts in the operational level in the form of resources.

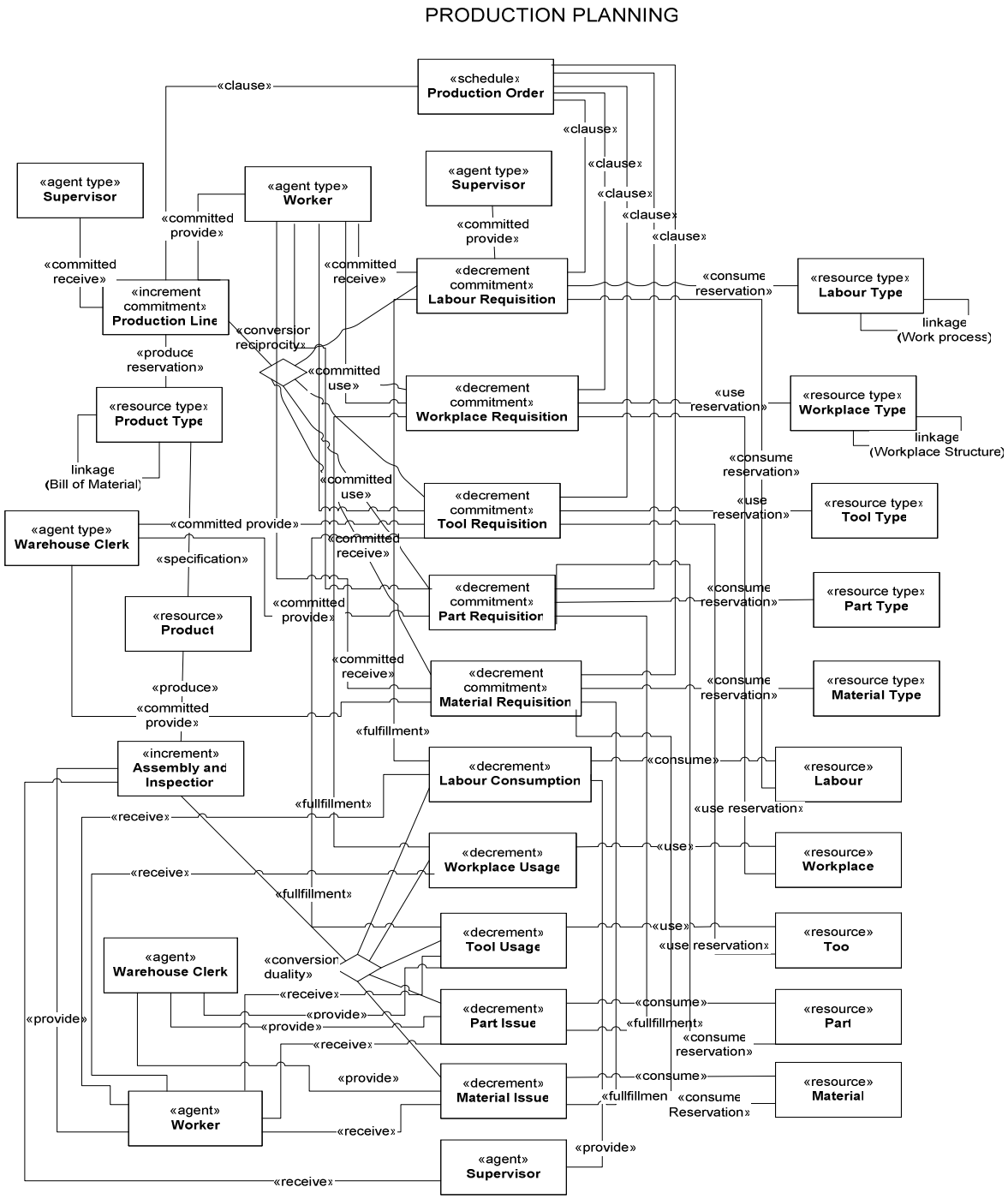


Figure 1. Production Planning Process

3. The Fulfilment Relation

In the presented results the *Commitment* entity plays an important role at the policy level of the REA framework. The current REA framework distinguishes two levels, the operation level (lower level) connected with the current facts that occurred in a company and the policy level (upper level) linked with the future activities and guidelines such as plans, commitments and control activities of the company. However, the policy level is not a homogenous one but is actually created by two "sublevels". One is represented by the *Commitment* entity with the fulfilment relationship, by which it is related to the operational level. The other represents entities created by the typification and grouping semantic abstractions, which use the typification, grouping and policy relationships. It can be said that two sublevels form "stratification" of the policy level.

By its character, the *Commitment* entity is semantically very close to the other "typed" entities in the policy level in the scope of the relationship to the operational level. The main difference between the fulfilment relationship and the typification relationship is cardinality. While the typification relationship represents a *one-to-many* relationship, the fulfilment relationship forms a *many-to-many* relationship. *Commitment Pattern* [Hruby, (2006)] describes it all in a more detailed way – see the Figure 2.

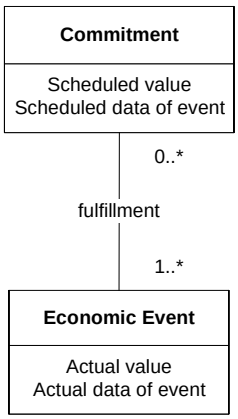


Figure 2. Commitment and Economic Event Pattern

Source: Hruby, P., (2006), *Model-Driven Design Using Business Patterns*. Springer-Verlag Berlin Heidelberg

This ambiguity between the *Commitment* and *Event* entities requires some specific solution in the form of unambiguous relationships. From the previous description emerges that the fulfilment relation encompasses two cases that do not occur simultaneously.

The first one happens when the *Commitment* entity is performed by one or more *Event* entities. In this particular case the fulfilment relationship can be replaced by the typification relationship.

The second one happens when the *Event* entity is performed by more *Commitment* entities. To make the relationship unambiguous an auxiliary entity has to be added to the current structure. In this case the *CommittedElement* entity is used to divide the *Commitment* entity into smaller parts so that each part corresponds only to one *Event* entity. The proposed solution is illustrated in the Figure 3, where we use {or} restriction to indicate that only one relationship is in force. The original fulfilment relation is replaced by the grouping and typification relationships.

Proposed modification illustrated in Figure 3 brings some drawbacks. The REA ontology does not know the {or} relationship and the *Economic Event* is related to the *Economic Commitment* by two relationships, which is a bit awkward. The other possibility that would improve the proposed modification solution should conform with the REA ontology and be simpler.

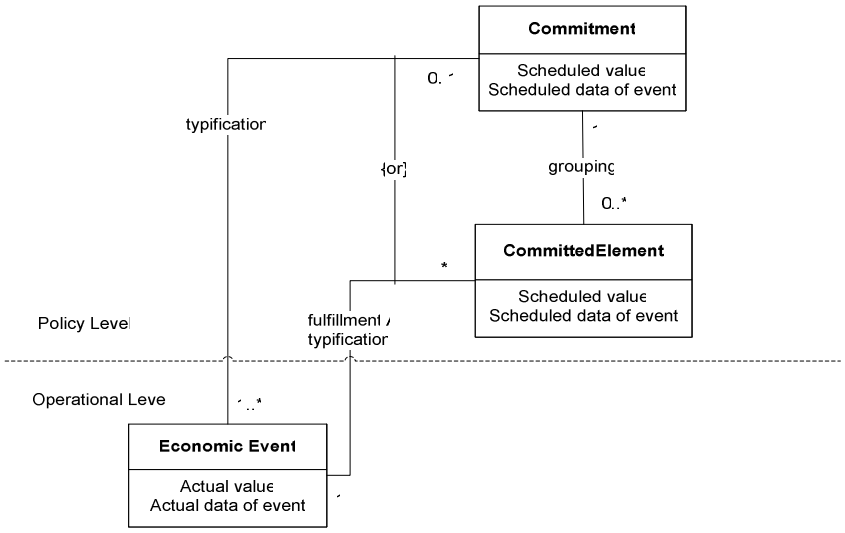


Figure 3. Proposed modifications of the Commitment and Economic Event

The typification relation is a very powerful tool that enables categorization of the typified entities. In the presented example it means that instances of the *Economic Event* can access data attributes of the instance of the *Commitment entity* that also comprises *CommittedElement* instances. The *CommittedElement* instances are stored in a collection that is represented by an attribute in the *Commitment entity*. The improved solution is in Figure 4.

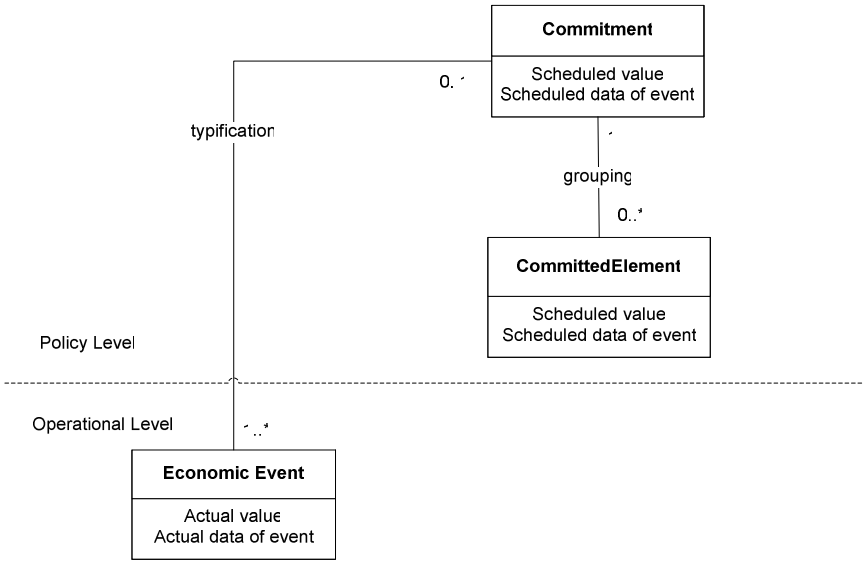


Figure 4. Improved proposal

4. Discussion with Further Improvement

However, the improved proposal diagram is rather isolated from the whole REA model. The proposal in the context of the REA model is illustrated in Figure 5.

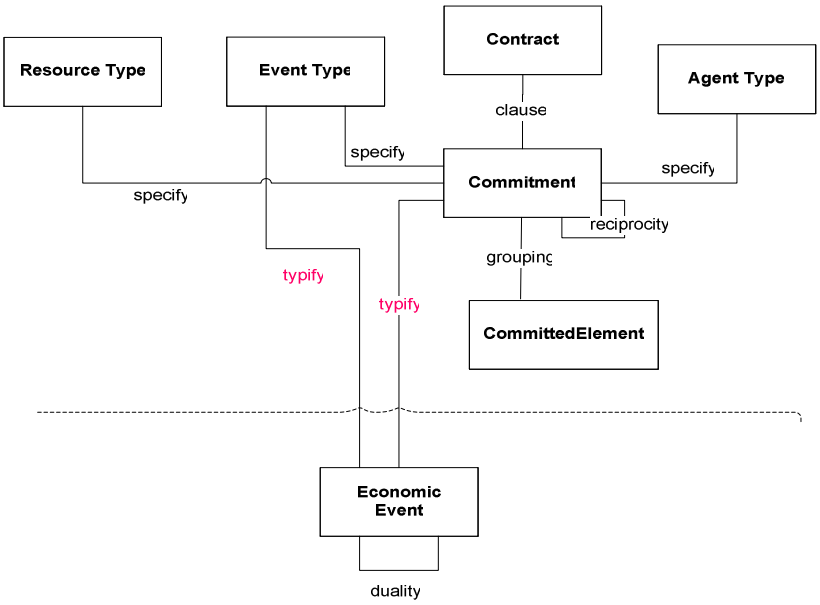


Figure 5. *Improved proposal* in context of the REA model

It is obvious from the Figure 5 that the *Economic Event* entity is related to the *Commitment* entity and to the *Event Type* entity through the typification relation. This solution however is not implementable by any software platform. One of the other improvements of this issue is to eliminate one of the typification relations. We will maintain the more important relationship, which is the relationship that relates the *Economic Event* and the *Economic Event Type* entities as we suppose to put the *Commitment* entity under the *Event Type* entity.

The other very important aspect that is necessary have in the mind during modification of the REA model is granularity of the entities. The other word for this activity can be gradual refinement of the requirements. What does it mean? The REA model should conform the gradual refinement of the requirements placed on the "planning – production" entities (*Event Type*, *Commitment*, *Event*). The *Event Type* entity represents normative rules simply recipe of the advancement. The *Commitment* entity gathers all planned requirements (information) – further adjustment of the original demands. Finally, the *Event* entity holds all actual information. This aspect should be fulfilled.

In the proposal that is stated we presume that in case when one *Event* entity fulfils several *Commitment* entities these *Commitment* entities have some properties common or very close. This precondition is frequently accomplished in reality. Therefore stratified typification between the *Event Type* entity, the *Commitment* entity and *Event* entity is created. This link ensures essential granularity that is necessary for the whole solution.

In case that one *Event* entity is related to many *Commitment* entities the common properties of these entities are placed in the Commitment entity and the different properties are stored in a collection of *CommittedElements* entities. The *Commitment* entity is related to the *CommittedElement* entities through grouping relationship (see Figure 6 and 7).

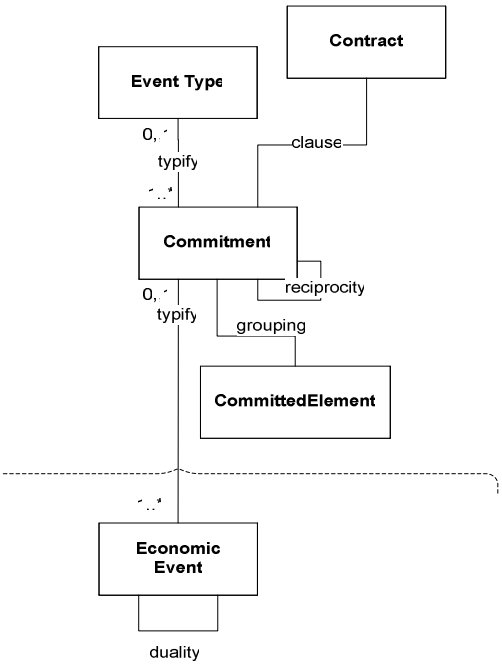


Figure 6. Partial view of the proposed solution

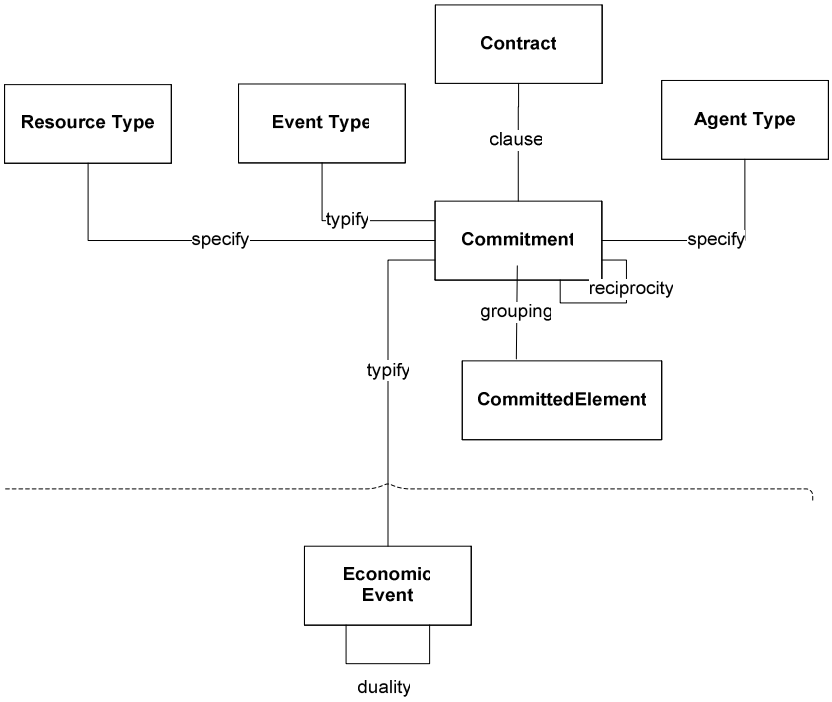


Figure 7. Whole view of the proposed solution

5. Conclusion

The typification and grouping semantic abstractions specify policy-level extension of the REA framework. These abstractions enable to work with the types of declared entities and with a special form of aggregation with set-level characteristics. The *Economic Commitment* entity with its fulfillment relationship stands a bit outside of the above mentioned abstractions. In the paper we tried to bring this entity closer to the typification and grouping semantic abstractions by introducing a new entity called *CommittedElement* and by replacing the fulfillment relation with the typification relation. In this way, the *Commitment* entity can be composed of the *CommittedElement* entities, which result

in the replacement the *many-to-many* relationship. The similar technique is utilized in the process of analysis and design of the information systems. While the relationship between analytical classes can have a many-to-many relationship, the relationship between designed classes is restricted only to one-to-many relationships, by Arlow and Neustadt (2002).

In the proposed solution a stratified typification relation is utilized to reserve the granularity of the entities that lie in the direction from planning towards production. It is also assumed that in case there are many *Commitment* entities related to one *Event* entity the *Commitment* entities have some common properties (information) stored in the *Commitment* entity. The specified properties are stored individually in the *CommittedEntities*.

6. Acknowledgements

The paper was supported by the grant reference No. 402/08/0277 provided by The Grant Agency of the Czech Republic.

7. References

[1] Geerts, GL, McCarthy, WE, (2006), *Policy-Level Specification in REA Enterprise Information Systems*, in: *Journal of Information Systems*, Vol. 20, No. 2, pp. 37-63.

[2] Geerts, GL, McCarthy, WE, (2000), *The Ontological Foundation of the REA Enterprise Information Systems*. Paper presented at the Annual Meeting of the American Accounting Association. Philadelphia, PA.

[3] Geerts, GL, McCarthy, WE, (2002), *An Ontological Analysis of the Primitives of the Extended REA Enterprise Information Architecture* at http://www.msu.edu/user/mccarth4/

[4] Hruby, P., (2006), *Model-Driven Design Using Business Patterns*. Springer-Verlag, Berlin, Heidelberg

[5] Martin, J., Odell, JJ, (1998), *Object-Oriented Methods: a Foundation*. Prentice Hall PTR Upper Saddle River, New Jersey

[6] Arlow, J., Neustadt, I., (2002), *UML and Unified Process: Practical Object-Oriented Analysis and Design*. Addison-Wesley UK

[7] Hunka, F., Hucka, M., Kasik, J., Vymetal, D., (2009), *Enterprise Planning Model Using REA Ontology*. At: http://vmbo.blogs.dsv.su.se/files/2009/02/stockholm_hunka.pdf

Journal of Applied Economic Sciences

AN UPDATE OF THE FRANCHISEE MOTIVATIONS: A STUDY IN SPAIN

Jose M. **RAMIREZ-HURTADO**
Pablo de Olavide University, Seville, **Spain**
jmramhur@upo.es
Bernardino **QUATTROCIOCCHI**
La Sapienza University of Rome, **Italy**
bernardino.quattrociocchi@uniroma1.it

Abstract:
This paper presents an empirical study of the motives for acquiring a franchise, based on a sample of franchisees operating in the Spanish market. The results are compared with the obtained in the previous studies. Although this topic has been studied previously, the spectacular progress of franchises throughout the world in the last decade may have changed franchisees' opinions about their motives. This manuscript establishes an update of motivational incentives for individuals to enter franchising from the previous studies. The results provide the following ranking: (1) Proven business format, (2) Start-up support, (3) Established name, (4) Training provided, (5) Faster development, and (6) Ongoing support. This study and the previous literature suggest there is a limited group of motives involved in the franchisee decision, but the ranking varies from study to study. It shows a different ranking of motives for selection of a franchisee from previous studies.

Keywords: franchising system, franchisee decision, motivation, incentives.

JEL Classification: L26, M12, M51, M54.

1. Introduction

Franchising is an excellent opportunity for personal and business development, and so many individuals seek information on how the system functions, its advantages and drawbacks before deciding to take on a franchise.

As franchising is applicable to a wide variety of sectors anyone interested in acquiring a franchise has a wide variety to choose from, such as catering, real estate, clothing and fashion, services, etc. Entering the franchise business, however, entails dealing with a wide variety of particularities, which requires a good grasp of the aspects involved [Ramirez, (2007)].

Elango and Fried (1997) provide an important review of the existing franchise literature. In relation to franchise system operations and more specifically in relation to franchisees, the above authors note the existence of studies analysing franchisor control [Carman and Klein, (1986)]. Other studies have attempted to analyse franchisee satisfaction [Lewis and Lambert, (1991); Hunt and Nevin, (1974); Schul *et al.*, (1995)] and franchisee cooperation [Anand, (1987); Anand and Stern, (1985); Guiltinan *et al.*, (1980); Hunt and Nevin, (1974); Sen, (1983)]. Elango and Fried (1997) also offer suggestions for future research lines.

Research has not kept pace with the significant developments in franchising. There is a relative shortage of information on the nature and scope of the franchise business globally [Hoffman and Preble, (2001)]. In this regard, very few studies have focused on analysing the motives for choosing the franchise system rather than setting up an independent company. In fact, even Elango and Fried (1997) do not suggest this topic in their literature review.

The work we present here attempts to contribute to the scanty literature on motives for choosing the franchise system. This paper describes the results of a survey of franchisees in Spain to establish the reasons for selection of a franchisee. The most recent study on this topic was by Guilloux *et al.* (2004) in France with data from 1999. There are also previous studies by Withane (1991) in Canada, Peterson and Dant (1990) in the USA, Knight (1986) also in Canada and the USA, Stanworth (1985) and Hough [Mendelsohn, (1985)]. Kaufmann and Stanworth (1995) provide another important study on the franchisee decision, also with data from the USA.

In recent years, franchising has become a very important part of many countries' economies. As time has gone by, more in-depth knowledge has been gained on a commercial system which is capturing the attention of business owners interested in turning their company into a franchisor, and individuals who wish to become franchisees.

Except for the work by Guilloux *et al.* (2004), most studies are more than 15 years old and that is why, in this work, we attempt to provide an update on the motives which lead a person to choose the franchise system rather than set up an independent company. We also want to determine whether there has been any significant variation in the ranking of criteria for the franchisee decision compared to previous studies.

Furthermore, this study is based on data from franchisees operating in the Spanish market. Studies from different countries are important for determining differences in the franchise decision. In addition to cultural differences between countries, there are also other differences in terms of aspects such as resources, infrastructures, history and traditions, politics, economic conditions and different legal frameworks [Preble and Hoffman, (1995); Sethi and Elango, (1995)], which may mean that franchisee motivations vary from one country to another. Another important contribution therefore, is that this study uses data from a country previously absent from the franchise literature, but which is among the leading countries in terms of franchise development.

According to data from the Spanish firm of consultants, Tormo and Associates (2007), in 2006 there were 902 franchises, representing 63,751 franchise establishments employing a significant number of people. In comparison to the retail trade in Spain, global turnover for franchising in 2006 was 17,585 million euros, or 14.68% of total retail trade turnover, (13.88% in 2005). Also in 2006, there were 63,751 franchise establishments representing 9.90% of all retail sales establishments, and in 2005 this figure was 8.97%. Thus, the franchise business is an important commercial system in the Spanish economy, enabling wealth to be generated through company creation, which in turn directly or indirectly, generates employment. Internationally, Spain ranks fifth, in terms of number of franchisors and franchise establishments.

In short, this study's main aim is to determine a ranking of franchisee motives in the Spanish market.

2. Previous research

Franchising is seen as an entrepreneurial option towards creating and developing ventures [Bygrave, (1997)]. Franchising is a commercial system for entrepreneurship in service industries to assemble resources in order to rapidly create large chains and gain first mover advantage [Michael, (2003)]. So, franchising can be seen as a method of entrepreneurial expansion [Nieman *et al.*, (2003)]. In fact, the entrepreneurs are more likely to franchise than to star an independent business [Williams, (1998)].

Hisrich *et al.* (2004) comment that an advantage of buying franchise is that the entrepreneur does not have to incur all the risks often associated with starting a business from scratch. Franchisees are contractually obliged to not deviate from the operational plans of the franchisor, inhibiting their creativity and proactivity.

From a strategic perspective, franchising is a popular method of leveraging company's intellectual capital by opening up new market with a highly motivated own-operator who will work harder than the average salaried manager to ensure the success of a particular location. The success of franchisees is the success of franchisors [Mendelsohn, (2003)].

The literature review points to a variation in the ranking of franchisee motivations, and consequently a variation in ranking by countries. The motivations that drive franchisees into choosing the franchise format may not be as homogeneous as supposed [Peterson and Dant, (1990)]. Table 1 shows the main results from previous studies. The study by Guilloux *et al.* (2004) found that the main franchisee motivation in France is advertising and start-up support. Potential franchisees need to be secured particularly about the question of start up support provide by the franchisor. The trade name and the possibility for development are the following motives for individuals to enter franchising. Comparing with previous studies, although the methodological approaches are not the same, these three reasons for choosing the franchise format receive a high level of assent. Withane (1991) reports that the main motive in Canada, is the proven business format. Many prospective franchisees choose to join franchises over starting independent businesses in order to take advantage of the established business format, goodwill, and start-up and on-going support system. It shows the importance of support like motivational incentive for individuals to enter franchising. Peterson and Dant (1990) found that the main motives in the USA are "training provisions", "greater independence", and "established name", while "lower operating costs" and "less management involvement" were rejected

as advantages. In a previous work Knight (1986) found that in Canada and the USA the main motive was an already established commercial name and the second was a greater job satisfaction than salaried employment. Stanworth (1985) describes the main motive as "national affiliation" which, according to Peterson and Dant (1990), can be considered equivalent to "established name" in the other studies and therefore coincides with Knight's (1986) first motive.

Finally, Hought (1986) found that the main motive was ongoing support. Thus the main criteria vary from one country to another and over time.

Table 1. Comparison of previous empirical studies

Ranking	Our Study 2007 (Spain)	Guilloux *et al* (2004) (French)	Withane (1991) (Canada)	Peterson and Dant (1990) (United States)	Knight (1986) (Canada and United States)	Stanworth (1985)	Hought (1986)
1	Proven business format	Advice and assistance before opening	Proven business format	Training provided	Established name	National affiliation	Franchise support
2	Startup support	Name	Less risky	Established name	Satisfaction	Franchisor support	Lacked experience
3	Established name	Possibility for development	Goodwill	Greater independence	Greater independence	Better investment	Established name
4	Training provided	Franchiser's support	Startup support	Better investment	Faster development		
5	Faster development	Profitability	Ongoing support	Lower development cost	Proven formula		
6	Ongoing support	Advertising	Quick start		Better investment		

Source: own design based on Guilloux *et al.* (2004) and Peterson and Dant (1990).

There are other additional criteria which are also considered important advantages of franchising. Izreali (1972) provided a fairly exhaustive list of 17 advantages of franchising. Since then, other authors have grouped these 17 franchising advantages into broader categories, for example Peterson and Dant's (1990) 7 category grouping.

Table 1 contains all the motives in the literature. There are, however, other motives which do not appear in the Table but were included in the studies by Stanworth (1985) and Hough (1986), despite the low percentages. Stanworth (1985) includes an opportunity to build a business for the future and/or family, greater job satisfaction and previous experience. Hough (1986) includes previous experience and pure change.

All these comments show that even if certain aspects remain constant, potential franchisees' criteria for choosing franchise evolve over time and that this should be scanned carefully [Guilloux *et al.*, (2004)].

3. Methodology

In late 2007, a questionnaire was sent to a convenience sample with 992 franchisees operating on the Spanish market. The franchisees were chosen from the *Franchise Guide* published on Tormo and Associates web site in 2007. Franchisees selected belonged to franchisor chains catalogued as interesting and profitable. Each franchisee was sent a questionnaire and asked to mark the most important criterion in their franchisee decision. The criteria which appeared in the questionnaire were obtained from the literature review. A total of 15 criteria were used, as some which used to be studied separately have been combined with others. There was also the opportunity to include a criterion not on the list. In addition to this information, franchisees were also asked to provide details of their educational level, business experience, the number of years they had owned the franchise and annual sales. A total of 220 valid responses were received. The franchisees traditionally do not reveal the information about the establishment. The franchisees think they are giving information that belongs to

the relation franchisor-franchisee. This explains the low number of responses and could be motivated by the franchisees perception that this action is not approved by the franchisors. Sample characteristics are shown in Table 2.

Table 2. Sample Characteristics

Characteristics	Percentage of Sample
Educational Level	
Without grade school	1.2
Basic school	8.3
High school	33.4
University degree	57.1
Prior Business Experience	
Without Experience	60.7
Previous Experience in Own Business	23.8
Previous Experience in Same Industry	15.5
Years Owning Current Franchise	
1 or less	53.6
2-3	32.2
4-5	7.1
over 5	7.1
Annual Sales (in Euro)	
less than 9.999	22.6
10,000-29,999	21.4
30,000-59,999	4.8
60,000-149,999	17.9
over 149,999	33.3

Over 90% of respondents had higher education, 8.3% had primary education and only 1.2% had no education. 60.7% of franchisees had no prior experience as a business owner, 23.8% had experience with their own business, 15.5% had experience in the same sector. Most of them, 85.8%, had owned the franchise for 3 years or less. In terms of sales, the largest percentage, 33.3%, had sales of over 150,000 euros.

4. Survey results and discussion

The main franchisee motives on the Spanish market are shown in Table 1. 15.0% of respondents considered that the main franchisee motive was that the business format is already proven. Close behind, 14.5% consider that the main franchisee motivation is start-up support. This is followed by the benefit of a well-known brand and notoriety, 12.3%; training provided 10.9%; faster development, 9.5%; and ongoing support, 7.7%.

These six motives are followed, in order of importance by, greater independence than being an employee in another business, job satisfaction, less risky than an independent business, lower start-up costs, franchise advertising, less daily involvement than in an own independent business, earn more money than in an independent business, the individual or family's previous experience and the fact that franchises are fashionable.

After analysing the most important franchisee motivations on the Spanish market, we examined the influence of the following variables: educational level, previous business experience, number of years as a franchisee and annual sales. This was done using a contingency table constructed on the basis of the motives and each of the variables. The possible influence of the different categories of variables on the motives indicated in Table 1 was determined by the Chi-square statistic. This statistic requires observations per cell to be equal to or greater than five. To minimize this effect we grouped some categories in some variables. Thus, in education level, we grouped the categories "no studies and primary studies", in the variable "number of years in the franchise" we considered the categories 1 year or less, 2-3 years and over 3 years. Finally for sales level, we considered the categories less than 29,999 euros, between 30,000 and 149,999 euros and over 149,999 euros. The results are shown in Tables 3, 4, 5 and 6.

4.1. Influence of educational level

The franchising system involves dealing with a variety of situations which require a good grasp of the business world and therefore it would seem that having the appropriate training is a huge advantage. Bates (1995) found that franchisees had significantly weaker educational backgrounds when compared to owners of non-franchised businesses, and thus were more susceptible to failure than independent business owners. Edens *et al.* (1976) consider that higher qualifications are appropriate for franchisees, and a primary education certificate would be the minimum acceptable.

Table 3. Influence of educational level on franchisee motives

Reasons	Minimum obligatory	High school	University degree	Chi-Square Statistic
		(Percentages)		
Proven business format	1.3	7.1	13.0	39.266 p<0.05
Startup support	2.6	7.8	10.4	
Established name	1.3	6.5	9.7	
Training provided	1.3	4.5	9.7	
Faster development	1.3	6.5	5.8	
Ongoing support	0.6	3.2	7.1	

Table 3 shows that the p-value for the Chi-square statistic is below 0.05, thus providing statistical evidence to reject the null hypothesis of independence between both variables. Therefore, the different categories of educational level have a significant influence on franchisees' opinion of their motivations. Particularly significant is the fact that individuals with a basic education consider that the most important motive is the start-up support. That is to be expected, because they have less training and need help to start up a business. The other motives follow a very similar order to that obtained in Table 1 for each educational level category.

4.2. Influence of prior business experience

Franchisees with prior business experience can be more critical when evaluating franchise advantages [Peterson and Dant, (1990)], and therefore their franchisee motivations may be different from those of inexperienced franchisees. Edens *et al.* (1976) consider that some prior business experience is appropriate for the ideal franchisee profile. In contrast to an independent business, where prior experience is vital for success, in the franchise system, franchisees' experience is not associated with their satisfaction [Hing, (1995)]. Pearsons with a history of self-employment will be more interested in becoming franchisees than will persons without such a history [Kaufmann and Stanworth, (1995)]. Often it is even preferable for the potential franchisee to have no prior experience [Stanworth, (1991)]. As Stanworth affirms, franchisor executives tend to prefer people from outside their industry with no preconceived ideas or bad habits which might interfere with the franchisor's training program or contaminate other franchisees. So, many franchisors will select franchisees who have no previous experience in the industry [McCosker and Frazer, (1998); Mendelsohn, (1993)] and some even actively recruit inexperienced franchisees because they feel they will be easier to indoctrinate into the system [Frazer, (2001)]. All this suggests that franchisees' opinions on their motivations may vary in relation to their previous experience.

Table 4. Influence of prior business experience on motives for become a franchisee

Reasons	No experience	Previous experience in own business	Previous experience in same industry	Chi-Square Statistic
		(Percentages)		
Proven business format	9.7	5.2	6.5	28.228 p<0.05
Startup support	14.3	3.2	3.2	
Established name	7.1	5.2	5.2	
Training provided	9.7	2.6	3.2	
Faster development	6.5	3.2	3.9	
Ongoing support	2.6	4.5	3.9	

As in the previous case, Table 2 shows that the p-value for the Chi-square statistic is less than 0.05 which provides statistical evidence to reject the null hypothesis about independence between variables. Therefore, the different prior experience categories significantly affect franchise motives. It should be emphasised that most inexperienced individuals consider that they became franchisees due to the start-up support. That is because they lack experience and so need help to start a business. However, those with experience attach little value to this motive. It should also be noted that many individuals with prior experience in a business other than a franchise significantly value ongoing support, even though this ranks at the bottom for the sample as a whole.

4.3. Influence of number of years in the franchise

As Peterson and Dant (1990) indicate, the longer franchisees have been in a franchise network, the greater their capacity and self-confidence in operating the sales outlet. For the same reason, the most senior franchisees have a great sense of personal power to "go it alone". In this case, some franchise advantages could be less important for them. Therefore, opinion on the motives for choosing the franchise system may vary in relation to number of years in the franchise.

Table 5. Influence of number of years on franchisee motivations

Reasons	1 or less	2-3	Over 3	Chi-Square Statistic
	Percentages			
Proven business format	8.4	7.8	5.2	
Startup support	11.0	6.5	3.2	
Established name	9.1	5.2	3.2	30.057
Training provided	9.7	3.2	2.6	p<0.05
Faster development	6.5	4.5	2.6	
Ongoing support	5.8	1.9	3.2	

As in the previous cases, Table 5 shows statistical evidence for rejecting the hypothesis of independence between both variables and so number of years in the franchise also has a significant influence on the opinion of franchise motives. Particularly outstanding is the fact that most franchisees who have been in the franchise for 1 year or less consider that the main franchisee motive is start-up support. This may be due to the fact that at the beginning the franchise system requires significant assistance in terms of infrastructure and financial capital. As the years go by, this motive becomes less important as the other two categories show. Also, for franchisees that have been in the franchise 1 year or less, training provided by the franchisor is very important. This motive becomes less important as the years go by and the franchisee begins to place more value on other aspects.

4.4. Influence of level of sales

If the level of sales is high then the franchisee may well consider he or she has considerable personal power. In this case, some franchise advantages may be less important for them. In contrast, when sales are low, franchisors may exercise greater control over franchisees, the franchisee's personal power is thereby diminished and the advantages of ongoing support are highly valued [Peterson and Dant, (1990)]. Thus, opinion on franchise motives may vary in relation to the level of sales.

Table 6. Influence of level of sales (in euros) on franchisee motivations

Reasons	Less than 29,999	30,000 – 149,999	Over 149,999	Chi-Square Statistic
	(Percentages)			
Proven business format	5.2	7.1	9.1	
Startup support	7.8	4.5	8.4	
Established name	5.8	4.5	7.1	41.751
Training provided	7.8	3.9	3.9	p<0.05
Faster development	4.5	1.9	7.1	
Ongoing support	3.2	2.6	5.2	

Analysis of the data in Table 6 shows that the p-value for the Chi-square statistic is below 0.05, providing statistical evidence to reject the null hypothesis of independence between both variables. Thus, the level of franchise sales also has a significant influence on opinion about franchisee motivations. Franchisors with sales below 29,999 euros do not consider that their main franchisee motivation is that franchises are a proven business format. This may be due to the fact that their annual income is not what they expected it to be. Unlike the previous opinion, the other individuals do consider this to be their main franchisee motivation. Another aspect worth highlighting is the fact that franchisees with an income of over 149,999 euros consider fast development to be very important, an aspect not highly valued by the other franchisees. This may be due to the fact that they have obtained high returns in a short time and therefore value this motive very positively.

In short, we have seen that educational level, prior business experience, number of years in the franchise and the level of sales significantly influence franchisees' opinion of their motives for choosing the franchise system rather than another business.

In order to examine this influence further, we looked at whether the ranking of motives obtained for the whole sample is the same for each of the above categories of variables. The ranking of motives per variable category is shown in Table 7.

Table 7. Comparison of ranking across variable categories

Reasons	Educational level			Prior business experience			Number of years in the franchise			Level of sales			FULL SAMPLE
	E1	E2	E3	P1	P2	P3	Y1*	Y2	Y3	S1	S2	S3	
Proven business format	2	2	1	2	1	1	4	1	1	3	1	1	1
Startup support	1	1	2	1	3	4	1	2	3	1	2	2	2
Established name	2	3	3	3	1	2	3	3	2	2	2	3	3
Training provided	2	4	3	2	4	4	2	4	4	1	3	5	4
Faster development	2	3	5	4	3	3	5	4	2	4	5	3	5
Ongoing support	3	5	4	5	2	3	6	5	3	5	4	4	6

E1: Until basic school.
E2: High school.
E3: University degree.
P1: Without experience.
P2: Previous experience in own business.
P3: Previous experience in same industry.
Y1: 1 or less.
Y2: 2-3.
Y3: Over 3.
S1: Less than 29,999.
S2: 30,000-149,000.
S3: Over 149,000.
*The only category with no tie on ranking.

In order to determine whether ranking is significantly different or not among the categories and the complete sample we found the Kendall tau coefficient which considers ties in the case of ordinal variables. The results are shown in Table 8.

Table 8. Kendall tau coefficient between categories and full sample

	E1	E2	E3	Full sample
E1	1.000	0.834** (0.000)	0.366 (0.087)	0.745** (0.000)
E2	0.834** (0.000)	1.000	0.271 (0.181)	0.662** (0.001)
E3	0.366 (0.087)	0.271 (0.181)	1.000	0.646** (0.001)
Full sample	0.745** (0.000)	0.662** (0.001)	0.646** (0.001)	1.000
	P1	P2	P3	Full sample
P1	1.000	-0.246 (0.221)	-0.200 (0.329)	0.817** (0.000)
P2	-0.246 (0.221)	1.000	0.825** (0.000)	-0.041 (0.839)
P3	-0.200 (0.329)	0.825** (0.000)	1.000	0.017 (0.933)
Full sample	0.817** (0.000)	-0.041 (0.839)	0.017 (0.933)	1.000
	Y1	Y2	Y3	Full sample
Y1	1.000	0.662** (0.001)	-0.110 (0.587)	0.738** (0.000)
Y2	0.662** (0.001)	1.000	0.325 (0.120)	0.941** (0.000)
Y3	-0.110 (0.587)	0.325 (0.120)	1.000	0.178 (0.381)
Full sample	0.738** (0.000)	0.941** (0.000)	0.178 (0.381)	1.000
	S1	S2	S3	Full sample
S1	1.000	0.355 (0.075)	0.176 (0.384)	0.709** (0.000)
S2	0.355 (0.075)	1.000	0.126 (0.534)	0.662** (0.001)
S3	0.176 (0.384)	0.126 (0.534)	1.000	0.499* (0.012)
Full sample	0.709** (0.000)	0.662** (0.001)	0.499* (0.012)	1.000

**Correlation is significant at 0.01.
*Correlation is significant at 0.05.

Educational level yields no significant difference in criteria ranking for the full sample, and a significant positive correlation ($p<0.01$). There is, however, a significant difference between the third category and the other categories for this variable. The main difference is in the motive "Faster development" which comes fifth in category E3 and fourth in categories E1 and E2.

The ranking for the variable experience in P2 and P3 differs significantly from that obtained for the full sample. In P2 and P3 the motive "Ongoing support" occupies second and third place respectively, while for the full sample it comes last. There are also differences in the ranking between P1 and P2; and P1 and P3.

In the case "number of years in the franchise", there are significant differences in ranking between category Y3 and the full sample. These differences are evident mainly in the two last motives for the full sample, "faster development" and "ongoing support", which occupy positions 2 and 3 respectively for franchisees with more than 3 years in the franchise system.

Finally, no significant differences are observed for sales level between the different variable categories and the full sample, although significant differences appear when the three variable categories are compared with each other. The most important difference is in the motive "training provided" which comes first in category S1, and 3rd and 5th respectively in categories S2 and S3.

In short, we have observed significant differences in the motivation ranking depending on the variable being studied.

5. Conclusions

In this work, we have determined the motives which lead an individual to choose the franchise system over other options. We consider that research on this topic, despite that already done, is scarce, especially in view of the significant recent growth in many countries which has influenced franchisees' opinions. The latest study in this field was done by Guilloux *et al.* (2004) with data from 1999. This study represents another recent step toward establishment the reasons for selection of a franchisee. It establishes an update of motivational incentives for individuals to enter franchising from the previous studies.

We have analysed franchisee motivations in a sample of 220 franchisees operating in Spain. The results rank the motives as follows: (1) Proven business format, (2) Start-up support, (3) Established name, (4) Training provided, (5) Faster development, and (6) Ongoing support. The paper shows a different ranking of motives for selection of a franchisee from previous studies. These motives also appear in the literature but in a different order. This study and the literature suggest there is a limited group of several motives which encourage individuals to become franchisees, but the ranking varies from study to study. As Peterson and Dant (1990) argued earlier, given the large number of situational, psychological, and economic-structural correlates, no single theory is likely to suffice. It appears very important to differentiate between startup and ongoing support. Startup support is more important than ongoing support for individuals to enter franchising.

In this study we have also examined whether educational level, prior business experience, number of years in the franchise and annual sales influence the above motives. The results show that all these variables have a significant influence on these motives. Similarly, differences have been found between the motivation ranking for some categories of variables and the motivation ranking for the complete sample.

This study should be extended to other countries, in particular those where the franchise system is less developed so that results may be compared. A further important aspect for research would be to find out how many motives actually influence the franchise decision.

The study sample was chosen on the basis of franchises catalogued as interesting and profitable by the consultancy firm Tormo and Associates. However, the results should be generalised with some caution as the sample design was not random, so the sample of franchisees may not be representative of the greater population. It would be very interesting to analyze the motives by type of franchisee and sector. Motives may be influenced by type of franchisee and the detailed nature of the sector in which the franchise is located.

6. References

[1] Anand, P., (1987), *Inducing Franchisees to Relinquish Control: An Attribution Analysis*, in: *Journal of Marketing Research*, 24(2), pp. 215-221.

[2] Anand, P. and Stern, L., (1985), *A Socio-psychological Explanation for Why Marketing Channel Members Relinquish Control*, in: *Journal of Marketing Research*, 22, pp. 365-376.

[3] Bygrave, W.D., (1997), *The Portable MBA in Entrepreneurship*, 2nd Edition, John Wiley: London.

[4] Carman, J.M. and Klein, T.A., (1986), *Power, Property, and Performance in Franchising: A Review and Integration of Theory*, in: *Research in Marketing*, 8, pp. 71-130.

[5] Edens, F.N., Donald, R.S. and Grider, D.G. Jr., (1976), *Franchisors Describe Ideal Franchisee*, in: *Journal of Small Business Management*, (July), pp. 39-47.

[6] Elango, B. and Fried, V.H., (1997), *Franchising Research: A Literature Review and Synthesis*, in: *Journal of Small Business Management*, 35(3), pp. 68-81.

[7] Frazer, L., (2001), *Causes of disruption to franchise operations*, in: *Journal of Business Research*, 54, pp. 227-234.

[8] Guilloux, V., Gauzente, C., Kalika, M. and Dubost, N., (2004), *How France's Potential Franchisees Reach Their Decisions: A Comparison with Franchiser' Perceptions*, in: *Journal of Small Business Management*, 42(2), pp. 218-224.

[9] Guiltinan, J.P., Rejab, I.B. and Rodgers, W.C., (1980), *Factors Influencing Coordination in a Franchise Channel*, in: *Journal of Retailing*, 56(3), pp. 41-58.

[10] Hing, N., (1995), *Franchisee satisfaction: contributors and consequences*, in: *Journal of Small Business Management*, 33(2), pp. 12-25.

[11] Hisrich, R.D., Peters, M.P. and Shepherd, D.A., (2004), *Entrepreneurship*, 6th edition, Irwin McGraw Hill: Boston.

[12] Hoffman, R.C. and Preble, J.F., (1991), *Franchising: Selecting a Strategy for Rapid Growth*, Long Range Planning, 24(4), pp. 74-85.

[13] Hough, J., (1986), *Power and authority and their consequences: a study of the relationships between franchiser and franchisee*, Faculty of Management Studies, Polytechnic of Central London, London. Unpublished PhD thesis.

[14] Hunt, S.D. and Nevin, J.R., (1974), *Power in a Channel of Distribution: Sources and Consequences*, in: *Journal of Marketing Research*, XI, pp. 186-193.

[15] Izreali, D., (1972), *Franchising and the Total Distribution System*, London, U.K.: Longman.

[16] Kaufmann, P. and Stanworth, J., (1995), *The Decision to Purchase a Franchise: A Study of Prospective Franchisees*, in: *Journal of Small Business Management*, 34(3), pp. 27-40.

[17] Knight, R., (1986), *Franchising from the Franchisor and Franchisee Points of View*, in: *Journal of Small Business Management*, 24(3), pp. 8-15.

[18] Lewis, M.C. and Lambert, D.M., (1991), *A Model of Channel Member Performance, Dependence, and Satisfaction*, in: *Journal of Retailing*, 67(2), pp. 205-225.

[19] McCosker, C. and Frazer, L., (1998), *Franchising Australia 1998: a survey of franchising practices and performance*, Toowoomba: USQ.

[20] Mendelsohn, M., (1985), *The Guide to Franchising*, Oxford, U.K.: Pergamon Press.

[21] Mendelsohn, M., (1993), *The Guide to Franchising*, 5th edition. Cassell: London.

[22] Mendelsohn, M., (2003), *The Guide to Franchising*, 6th edition. Thompson Learning: London.

[23] Michael, S.C., (2003), *First mover advantage though franchising*, Journal of Business Venturing, 18(1), pp. 61-80.

[24] Nieman, G.H., Hough, J. and Nieuwenhuizen, C., (2003), *Entrepreneurship: A South African perspective*. Pretoria: Van Schaik Publishers.

[25] Peterson, A. and Dant, R.P., (1990), *Perceived Advantages of Franchise Option from the Franchisee Perspective: Empirical Insights from a Service Franchise*, in: *Journal of Small Business Management*, 28(3), pp. 46-61.

[26] Preble, J. F. and Hoffman, R.C., (1995), *Franchising systems around the globe: A status report*, in: *Journal of Small Business Management*, 33(2), pp. 80-88.

[27] Ramirez, J.M., (2007), *Determinacion y estudio del perfil de franquiciado mediante la aplicacion de analisis conjunto*. Unpublished Doctoral Thesis (in spanish), Department of Economy, Quantitative Methods and Economy History, Pablo de Olavide University, Seville, Spain.

[28] Schul, P.L., Little, T.E. and Pride, W.M., (1985), *Chanel Climate: Its Impact on Channel Members' Satisfaction*, in: *Journal of Retailing*, 61(2), pp. 9-38.

[29] Sen, K.C., (1983), *The Use of Initial Fees and Royalties in Business-Format Franchising*, Managerial and Decision Economics, 14(2), pp. 175-190.

[30] Sethi, S. P. and Elango, B., (1995), *Does country of origin make a difference in MNC strategy: An exploratory study*, Paper presented at the Academy of International Business Meeting Seoul, Korea (November).

[31] Stanworth, J., (1985), The School of Management Studies of the Polytechnic of Central London, *The Guide to Franchising*, ed. Martin Mendelsohn, Oxford, U.K.: Pergamon Press.

[32] Stanworth, J., (1991), *Franchising and the Franchise Relationship*, in: *International Journal of Retail, Distribution and Consumer Research*, 1(2), pp. 175-199.

[33] Tormo and Associates, (2007), www.tormo.com

[34] Williams, D.L., (1998), *Why do entrepreneurs become franchisees? An empirical analysis of organizational choice*, in: *Journal of Business Venturing*, 14, pp. 103-124.

[35] Withane, S., (1991), *Franchising and Franchisee Behavior: An Examination of Opinions, Personal Characteristics, and Motives of Canadian Franchisee Entrepreneurs*, in: *Journal of Small Business Management*, 29(1), pp. 22-29.

THE EVOLUTION OF REAL GDP PER CAPITA IN DEVELOPED COUNTRIES

Ivan O. **KITOV**
Russian Academy of Sciences, Moscow, **Russia**
ikitov@mail.ru

Abstract:

The growth rate of real GDP per capita is represented as a sum of two components – a monotonically decreasing economic trend and fluctuations related to the change in some specific age population. The economic trend is modeled by an inverse function of real GDP per capita with a constant numerator. Statistical analysis data from 19 selected OECD countries for the period between 1950 and 2007 shows a very weak linear trend in the annual increment of GDP per capita for the largest economies: the USA, Japan, France, and Italy. The UK, Australia, and Canada show a larger positive linear trend in annual increments. The fluctuations around relevant mean increments are characterized by practically normal distribution (with Levy tails). Developing countries demonstrate annual GDP per capita increments far below those for the studied developed economies. This indicates an underperformance in spite of large relative growth rates.

Keywords: economic development, economic trend, business cycle, GDP per capita

JEL Classification: E32, O11, O57

1. Introduction

Real economic growth, as expressed by Gross Domestic Product, has been studied quantitatively since Simon Kuznets' works on accounting of national income and aggregate factor inputs in the 1950s. Hodrick and Prescott (1980) introduced a concept of two-component economic growth – an economic trend and a deviation or business cycle component. The trend component is responsible for the long-term growth and defines economic efficiency. In the long run, the deviation component of economic growth has to have a zero mean value. In 2004, Kydland and Prescott received the "Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel" for the study of "the driving forces behind business cycle", what demonstrates the importance of the best understanding of the economic growth processes and the explanation of the two-component behaviour.

Kydland and Prescott (1982), along with many other researchers, have proposed and studied exogenous shocks as the force driving fluctuations of real growth. Their research during the last 25 years has revealed numerous features of principal variables involved in the description of the economic growth. There are many problems left in the theory of economic growth. The study of Galor (2005) describes the evolution of income per capita since an epoch of Malthusian stagnation and discusses the process, which induced the transition to the current sustained economic growth in developed countries. It looks for a unifying theory accommodating various period of growth and based on solid micro foundations. This paper validates our model (2006) describing the evolution of economic trend in developed countries during the most recent period since 1950. We do not use any sophisticated technique of signal extraction, as proposed by Pedregal (2005), who explored two linear trend models with a nonlinear forecast function. In our framework, the long-term forecast is not limited in time since it is based on a constant annual increment of real GDP per capita. Obviously, our model does not allow an exponential growth path, unlike that presumed in the trend extracting procedures developed by Pollock (2007), and the business cycle has no upper limit to its frequency. We also test the trend forecast using actual time series.

Musso and Westermann (2005) analyzed the long-term evolution of real GDP and supply-side factors in view of potential interest from policymakers. There is a vivid discussion of the influence of modern information and communication technologies on productivity. The model we have developed denies any direct impact of such supply-side factors and roots the long-term economic growth in the rigid and hierarchical structure of income distribution [Kitov, (2008)]. This also means that externalities, including international ones, play no defining role in real economic growth contrary to the hybrid model propose by Klenow and Rodriguez-Clare (2005). We agree with Acemoglu and co-authors (2005), who stress the importance of economic institutions for real economic growth, and Jones (2005) elaborating on the input of intellectual achievements. There should be basic economic

institutions, which "determine the incentives of and the constraints on economic actors, and shape economic outcomes" and a developed framework for testing and implementation of modern ideas - technical, economic, social, etc. However, when established, the institutions are notable to provide any additional input to real economic growth.

In line with Jones and Manuelli (2005), we see the cause for the difference in per capita GDP between developed and developing countries in knowledge, or human capital, and its production and dissemination. The countries that have weaker systems of property rights, or higher wasteful taxation and spending policies, will tend to grow more slowly due to a strong bias of the distribution and evolution of personal incomes [Kitov, (2005)].

In this paper, only the evolution of annual increment of real GDP per capita in developed countries is modeled and no special econometric (statistical) techniques are used to validate the concept. The paper includes numerous figures, since the study is a purely empirical one and the best way to present quantitative results related to time series is to visualize them in form of time history. However, our model of real economic growth has been also tested for cointegration [Kitov, I., Kitov, O., Dolinskaya, S., (2007)]. A wide range of specific econometric tools developed for the investigation of real economic growth is presented by Durlauf and co-authors (2005).

Kitov (2006) proposed a model with the growth of real GDP dependent only on the change in a specific age cohort in the population and the attained level of real GDP per capita. According to this model, real GDP per capita has a constant growth increment and the observed fluctuations can be explained by the population component change. In developed countries, real GDP per capita has to grow linearly over time, if no large change in relevant specific age population is observed. Relative growth rate of real GDP per capita has to be an inverse function of the attained level of real GDP per capita with a potentially constant numerator for developed economies. This paper is devoted to validation of the model using GDP per capita and population data for some selected developed countries. Our principal purpose is to demonstrate the possibility to decompose GDP per capita growth into the two components.

2. The model and data

According to [Kitov, (2006)], there are two principal sources of the per capita GDP growth in the USA – the change in 9-year old population and the economic trend related to the measured GDP per capita level. The trend has the simplest form – no change in absolute growth (annual increment) values and is expressed by the following equation:

$$dG/dt = A \quad (1)$$

where G is the absolute value of real GDP per capita, A is a constant. The solution of this equation is as follows:

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

where $B=G(t_0)$, t_0 is the starting time of the studied period. Hence, evolution of real GDP per capita is represented by a straight line if the second factor of growth has no cumulative effect. As discussed below, only some developed countries are characterized by a significant influence of the second factor.

Then, relative growth rate can be expressed by the following equation:

$$dG/dtG = A/G(t) \quad (3)$$

Equation (3) indicates that the relative growth rate of per capita GDP is inversely proportional to the attained level of real GDP per capita, i.e. the observed growth rate should asymptotically decay to zero with increasing GDP per capita. On the other hand, the lower is the level the higher the growth rate. This inference might be a potential explanation for the concept of economic convergence. Relative growth rate must be higher in less developed countries, but the observed absolute gap in GDP per capita can not be overcome in future [Kitov, (2005)] unless some non-economic forces will disturb current status quo.

When considering real GDP per capita, one has to bear in mind the importance of a correction to

be applied to the per capita GDP values related to the difference between the total population and population of 16 years of age and above, as discussed in [Kitov, (2006)]. Only this economically active population should be considered when per capita values are calculated. By definition, Gross Domestic Income, which is equivalent to GDP, consists of the personal incomes obtained by the population of 15 years of age and over and corporate income, the corporations owned by the same population category. Thus, one can treat the published (original) readings of GDP per capita as biased and to be corrected for (multiplied by) the corresponding population ratio, i.e. the ratio of the total population and the population above 14 years of age.

Figure 1 shows the population ratio as obtained from the OECD population data (2006). When absent the missed readings are substitute with those for the closest year from above. Between 1955 and 2003, the ratio is characterized by an overall decrease with a slight increase demonstrated by some countries in the 1960s and 1970s. Currently, all the countries have the ratio below 1.3. In the 1950s, the ratio was above 1.3 for all the countries except Austria and Belgium. The last country met the decrease is Ireland - the drop started in 1980. Italy has had the lowermost ratio since 1970.

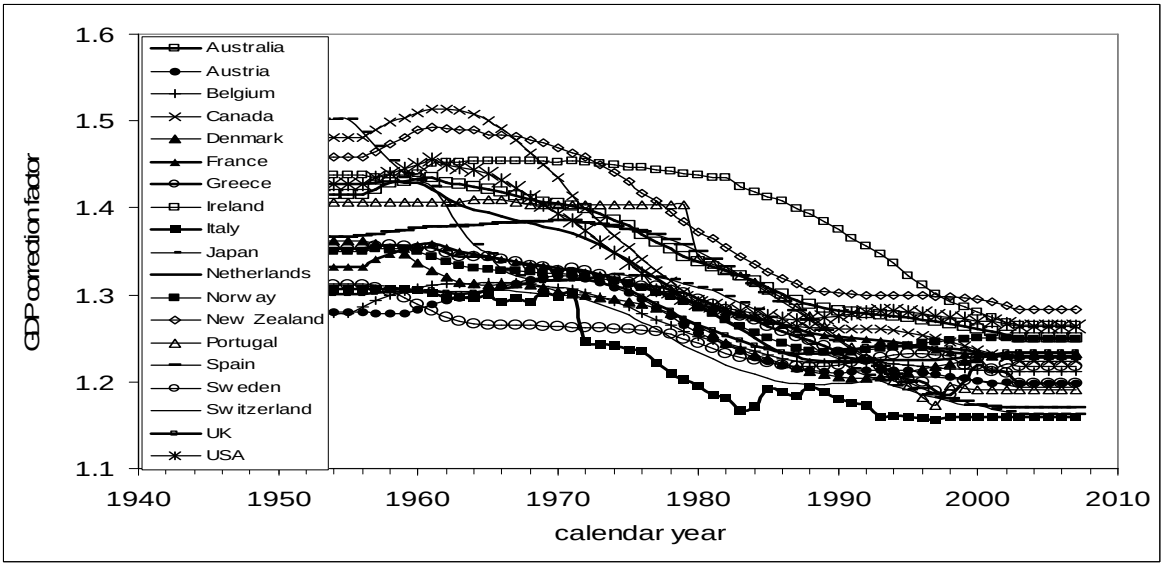


Figure 1. Evolution of a ratio of the total population and the population above 15 years of age for the selected OECD countries (2006).

High values of the ratio mean a relatively underestimated real GDP per capita and vice versa. A general feature of the curves is that after a small increase observed for some countries in the 1960s and 1970s the ratio decreases into the range between 1.3 and 1.15 in 2000. Thus the earlier GDP per capita values are relatively underestimated and the later readings are relatively overestimated. The longest period of a high ratio is observed in Ireland. Italy has a consistently low ratio. In the USA, the ratio drops from 1.45 in 1960 to 1.27 in 2003.

The decreasing ratio implies that the GDP per capita readings during the period between 1950 and 1970 are underestimated compared to those during the last 35 years. The larger is the total drop in the ratio during the entire period of the observation, the larger is the overall correction. In the study, the original and the corrected per capita GDP values are used and compared.

A cross-country comparison implies that GDP per capita is measured in the same currency units. There are two principal possibilities to reduce national readings of GDP per capita to some common scale: to use currency exchange rates or purchase power parities. In the study, we use the latter approach and data provided by the Conference Boars and Groningen Growth and Development Center (2008). For developed countries, two estimates of GDP per capita level are available: measured in 2007 US dollars, for which "EKS" purchasing power parities have been used and that expressed in 1990 US dollars, with the conversion at "Geary-Khamis" PPPs. These PPPs are obtained from the Organization for Economic Co-operation and Development (2005). Being an improvement on the previous dataset, the "EKS" PPPs are considered as more accurate and reliable. Amplitude of the change induced by the transition from "Geary-Khamis" PPPs to those of "EKS" is evaluated for the

counties under investigation. This change potentially characterizes uncertainty in the GDP per capita readings obtained with the PPP approach.

Only nineteen from thirty OECD member countries are analyzed. The selected countries meet some general criteria: 1) large economy size denominated in dollars; 2) continuous observations during the period between 1950 and 2007; 3) high level of real GDP per capita. According to the size criterion, small economies like Iceland and Luxembourg were excluded. When applied, the second criterion rejects Germany from the consideration. The third criterion has excluded such countries as Turkey, Poland and other new EU members. Finland and Korea have been excluded from the analysis with no reason at all.

Figure 2 illustrates the variations induced by corrections made for the "EKS" PPPs compared to those of "Geary-Khamis". The original values of the mean increment of GDP per capita ("EKS" and "Geary-Khamis") for every country are normalized to the corresponding values for the USA. The normalized values are consistently higher for the "EKS" PPPs, i.e. the GDP per capita values converted at "Geary-Khamis" PPPs were underestimated for all the countries. The difference varies with country and reaches 5% to 7% for Austria, Norway and Ireland. For the largest developed economies, the mean increments of GDP per capita expressed in 2007 dollars converge to that for the USA. We use the GDP per capita readings expressed in 2007 US dollars are used in the study. The only exception is the statistical description of the observed fluctuations.

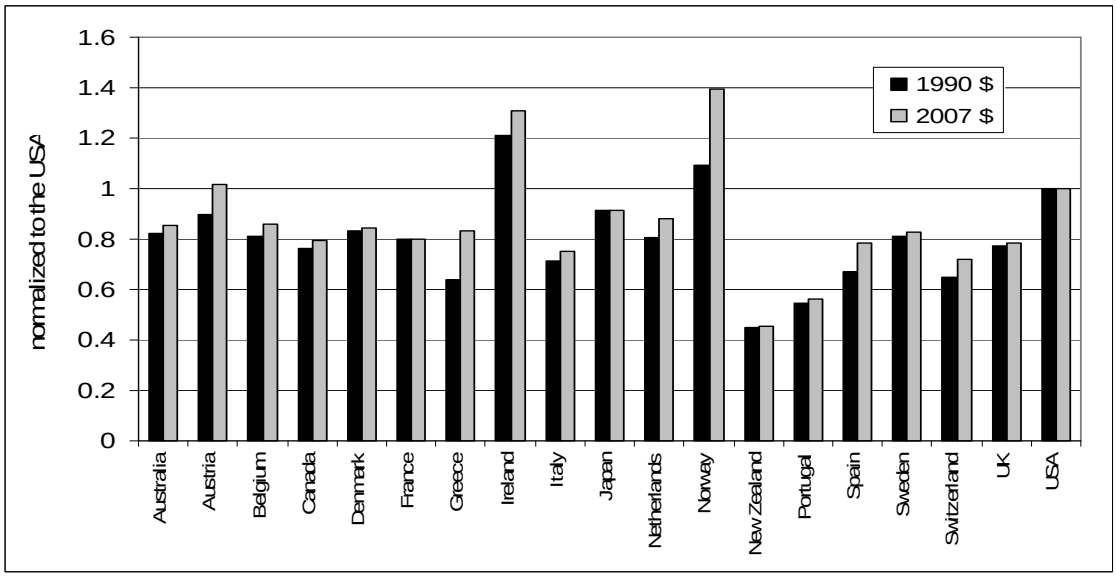


Figure 2. Comparison of two real GDP per capita data sets denominated in 1990 and 2002 dollars (2008), for which "Geary-Khamis" and "EKS" purchasing power parities have been used respectively.
Original values of the mean GDP per capita increment are normalized to the corresponding value for the USA. The normalized values are consistently higher for the "EKS" PPPs (except Canada), i.e. the GDP values converted at "Geary-Khamis" PPPs were underestimated for all the countries. The difference varies with country, however, and is larger than that between the original and corrected for the population values presented in Figure 3.

Figure 3 displays the averaged values of the annual GDP per capita increments denominated in 2007 US dollars for the period between 1950 and 2007. The original and corrected for the population ratio values for the nineteen countries are normalized to corresponding values for the USA. As before, this procedure allows a homogeneous comparison of the mean values. The corrected normalized values can be lower or higher than those for the original set. The sign of the change depends on the overall behavior of the population ratio during the entire period compared to that for the USA. Ireland, Austria and Norway are excellent examples of the originally underestimated GDP per capita values. Canada, Italy and Spain demonstrate an opposite behavior.

It is worth noting that the correction for population is of lower magnitude than that induced by the transition from the "Geary-Khamis" PPP to "EKS" one. The population correction is important, however, because it reduces potential uncertainty in the decomposition of the GDP per capita growth

into two components. The purchase power parity approach to the estimation of national GDP also needs some further improvements. Magnitude of the difference between the GDP per capita values converted at "Geary-Khamis" and "EKS" PPP sets is too high to believe that all the problems with the homogeneous and accurate cross-country comparison are resolved.

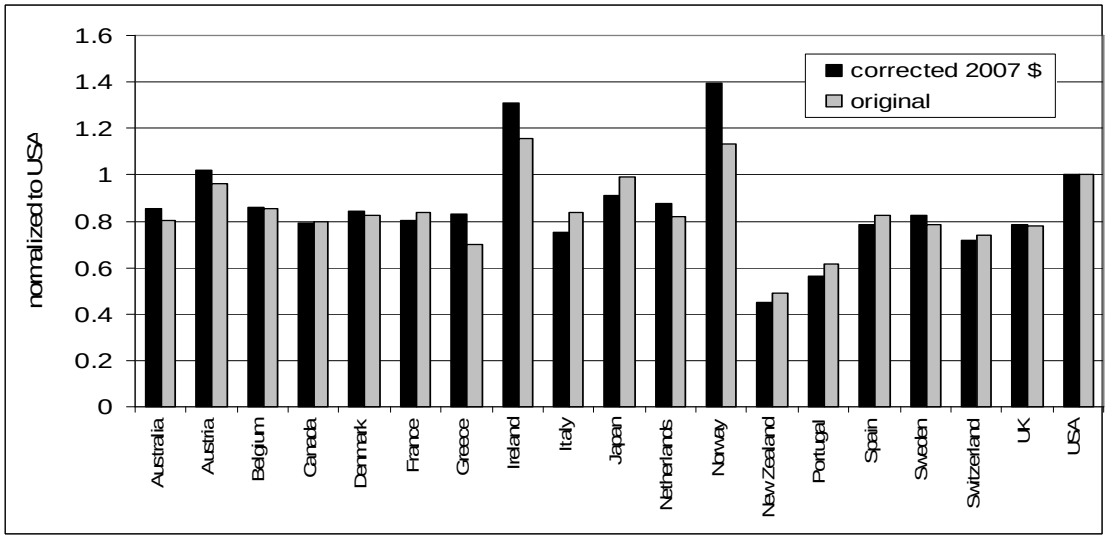


Figure 3. Comparison of the original and corrected for population mean values of GDP per capita increment expressed in 2002 dollars converted at "EKS" PPPs [2]

The values are normalized to the corresponding value for the USA for a homogeneous representation. There are countries with overestimated (where the corrected value is below the corresponding original value) and underestimated (opposite) values relative to the USA.

3. The trend in real GDP per capita

The nineteenth selected countries are presented in alphabetic order. Figure 4 shows the evolution of the annual increment of real GDP per capita for Australia as a function real GDP per capita for the original and corrected GDP readings. This is a natural visualization of Eq. (1). The population corrected values are connected by a solid line in order to highlight the evolution in time. Open circles represent the original measurements. In addition, three straight lines are drawn in the Figure. Bold line corresponds to the averaged increment of the population corrected GDP per capita for the entire period between 1950 and 2007. As a constant, this line is parallel to the x-axis. Second and third (solid) lines represent two linear regressions corresponding to the original and corrected data. Relationships for the regressions are also shown in the Figure; the lower one is always associated with the original GDP readings.

The model introduced in Section 2 implies that the mean value line has to coincide with the linear regression line, when the population induced component has a zero mean value. The observed fluctuations of the annual increment of GDP per capita are either predetermined or random ones, depending on the characteristics of the defining population changes. In terms of statistics, one could expect a normal distribution of the population changes. The number of processes affecting birth rate, mortality rate, and migration processes is very large and, according to the central limit theorem, leads to an approximately normal distribution of the deviations. However, random fluctuations of population do not presume the unpredictability of real economic growth. For example, the number of nine-year-olds in the USA, which are proved to be the driving force of the growth [Kitov, I., Kitov, O., Dolinskaya, S., (2007)], can be counted with any desirable accuracy. Statistical features of the increment are discussed later in Conclusion.

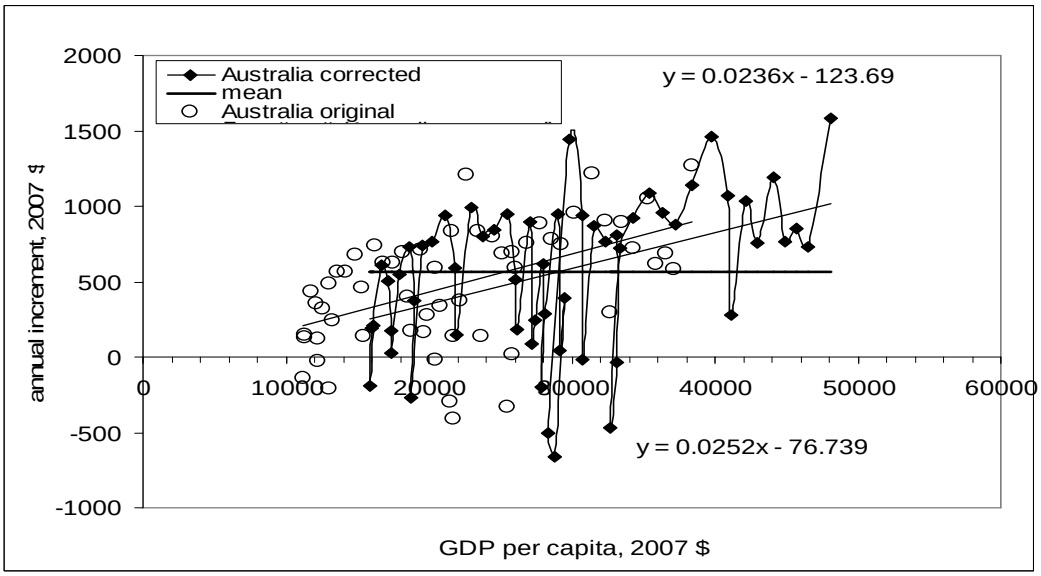


Figure 4. Increment of real GDP per capita (2007 $) vs. real GDP per capita in Australia for the period between 1950 and 2007

Two sets are presented - the original (open circles) and corrected for population (filled diamonds). Consequent values of the latter set are connected by a solid line for illustration of the evolution in time. Bold line represents the mean value of $565 for the population corrected data set. Two solid lines show linear regressions lines. The corresponding linear relationships are displayed, the lower relationship being associated with the original data set. The linear regression lines differ from that for the mean value.

Australia demonstrates a divergence between the regression lines and the mean value line. A positive linear trend (~0.01) has to indicate a more intensive growth of the specific age population in recent years compared to that in the 1950s and 1960s. This effect is observed also for other English-speaking countries under investigation. The average increment is $478 (2007 $). The largest deviation from the mean is -$1218. The linear trend coefficient is lower for the corrected data set than that for the original set. This is a common feature for almost all studied countries. Table 1 lists the mean values and regression coefficients for all country sets: the original and population corrected ones, converted at the “EKS” and “Geary-Khamis” PPPs.

Alphabetically, the following country is Austria. The average increment value for Austria is $570. This value is well above that for Australia. A prominent feature is an almost horizontal regression line for the population corrected data set with the slope of 0.006. Effectively, the mean line and the regression lines coincide, as predicted by (1). For the original set, the slope is slightly larger (0.009). Therefore, one can conclude that relevant specific age population has changed only marginally during the last 57 years (between 1951 and 2007). The largest fluctuation amplitude relative to the mean value was -$740 in 1993. In relative terms, such a deviation from the mean value is almost 4%.

The average increment in Belgium is $480 and the regressions are characterized by a positive linear trend that is higher for the original readings. The largest deviation from the mean value was -$790 in 1975. As in other countries, the negative deviations are usually sharp and deep somehow compensating longer periods of a weaker positive growth. The last twenty years have been relatively successful for Belgium. One can expect a compensating decrease, as was observed between $22.000 and $25.000.

Table 1. Mean values of GDP per capita increment for the original and population corrected readings in 1990 $ (converted at Geary-Khamis PPPs) and 2007 $ ("EKS" PPPs). Coefficients of linear regression (trend) are given for the original and corrected GDP per capita values denominated in 2007 $.

	original, 2007 $	corrected, 2007 $	trend original, $2007	trend corrected, $2007	original, 1990 $	corrected, 1990 $
Australia	478	565	0.025	0.024	317	374
Austria	570	675	0.009	0.006	347	410
Belgium	480	571	0.012	0.010	312	371
Canada	470	526	0.014	0.011	311	348
Denmark	473	560	0.012	0.014	322	381
France	443	531	0.0006	-0.003	305	365
Greece	469	553	0.022	0.018	248	292
Ireland	698	867	0.054	0.052	445	553
Italy	444	500	-0.003	-0.006	289	326
Japan	535	606	0.0001	-0.005	368	417
Netherlands	502	583	0.011	0.010	316	367
Norway	754	924	0.019	0.020	408	500
New Zealand	264	300	0.028	0.031	181	205
Portugal	323	373	0.008	0.004	216	249
Spain	457	522	0.014	0.007	269	307
Sweden	464	549	0.021	0.022	313	371
Switzerland	426	477	-0.010	-0.015	265	297
UK	431	520	0.026	0.026	292	353
USA	557	663	0.017	0.017	383	457

Canada is similar to Australia and is characterized by similar mean increment ($470), but lower trend coefficients (slopes). An important feature of the fluctuations in the Canadian economy is their amplitude reaching -$1391 from the average value – the highest among the studied economies. The two deep drops at $26000 and $31000 compensate a relatively successful history during the rest of the period.

Denmark and France (Figure 5) are similar in terms of weak linear trend, positive for Denmark and negative for France. Corresponding mean increments are also close - $473 and $443, respectively. If to neglect the slight slope observed in France, one can conclude that the observed fluctuations are characterized be a zero mean. Because of a limited time period of the observation, the trend values can be affected by side or truncation effects. The shape of fluctuations differs among the countries but longer periods of observations are necessary for suppressing the side effects.

The Greek economy had some hard years in its history but the last ten years were very successful. The economy is characterized by a relatively the mean increment of $469. The overall performance is expressed in an elevated linear trend. This country can not be used as an example of a developed economy maximizing its performance over years. A more prominent example of an excellent recovery gives Ireland with corresponding results displayed in Figure 6. A slow start was quickly compensated and the last twenty years of an extremely fast growth resulted in the leading position in the world economy with the mean increment $698. There are some doubts, however, that future will be so successful. Such a long and quick growth always ends up in a depression. This was observed in Japan and is related to the long-term decrease in the number of the specific age population [Kitov, I., (2006)]. Ireland has managed to increase birth rate for a very long period and has an age structure similar to that observed in Japan 20 years ago. The population distribution is currently peaked near 20 years with the defining age of 18 years. The years to come will demonstrate only decrease in the defining age population.

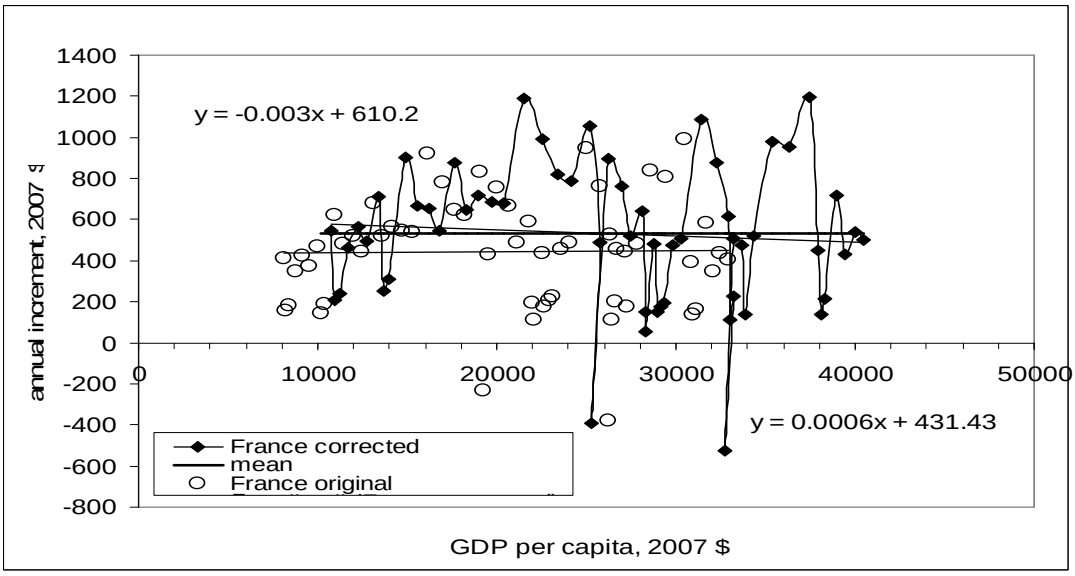


Figure 5. Same as in Figure 4 for France

The mean value is $443. The linear regression line for the original GDP values is practically parallel to the mean value line. The line for the corrected data set is characterized by a negative trend coefficient (-0.003).

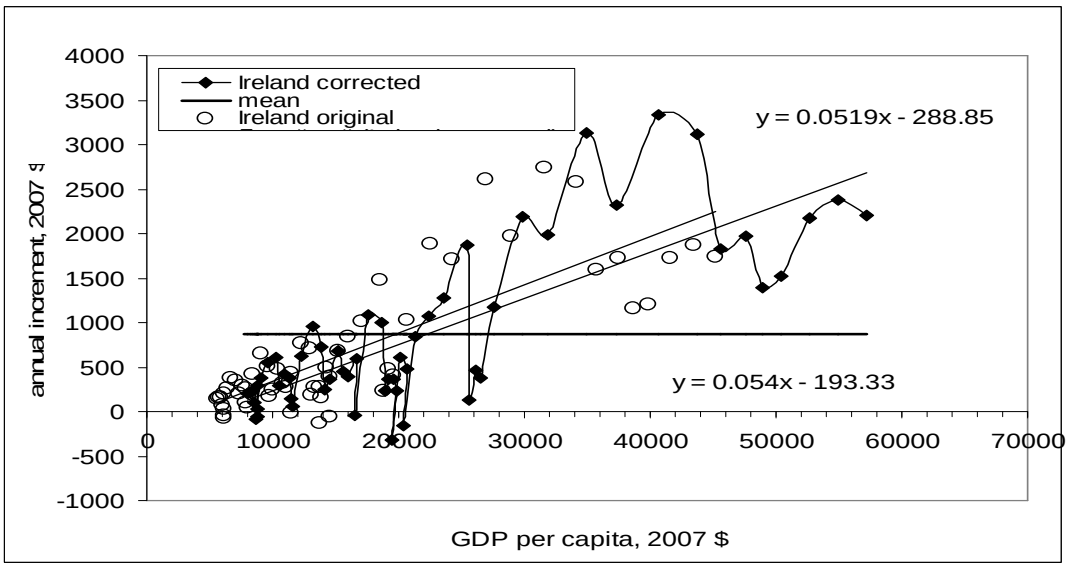


Figure 6. Same as in Figure 4 for Ireland

The mean value is $698. The growth of the real GDP per capita is outstanding during the last twenty years. There is a slight downward tendency during the last four years, however.

The next three countries are Italy, Japan, and Netherlands. Results for them are represented in Figures 7 through 9 and are similar to those for France – a weak positive or negative linear trend and the mean increment between $444 for Italy and $535 for Japan. These are also good examples of a zero linear trend in the history of GDP per capita increment.

Norway and New Zealand are very similar the pair Ireland/Greece. From the point of view of the current study they do not provide any additional insight into the GDP increment behaviour. Portugal is between Greece and New Zealand. Spain and Sweden are similar to other large European economies with a weak linear trend of the per capita GDP increment and the mean value around $450. Switzerland (Figure 10) had a decreasing increment which can be potentially explained by a permanent decrease in the young population portion.

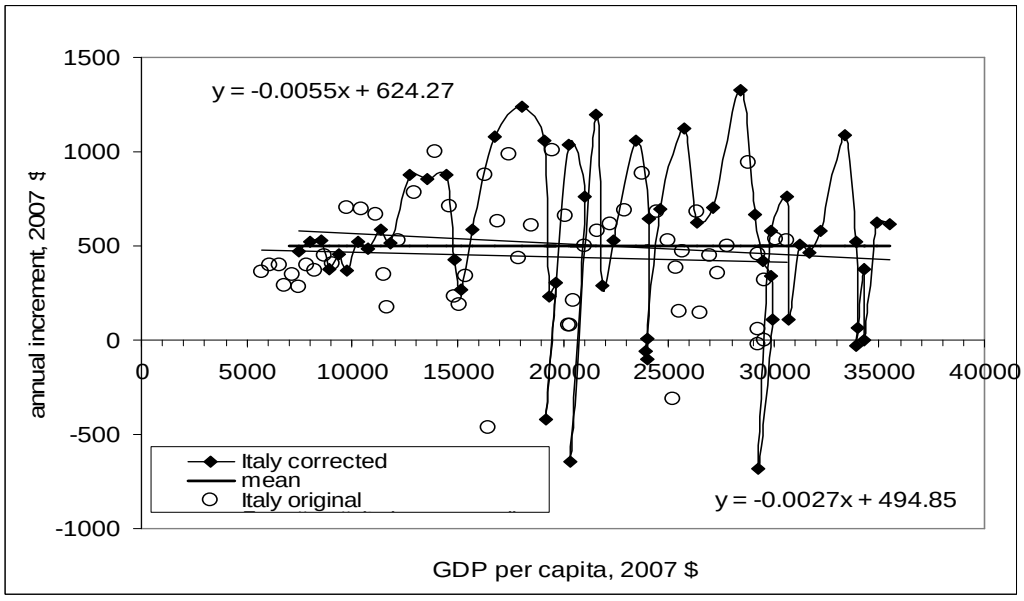


Figure 7. Same as in Figure 4 for Italy

Both the corrected and original GDP values produce a negative trend, the former being of a larger absolute value. Nevertheless, the lines are very close to those for the mean values.

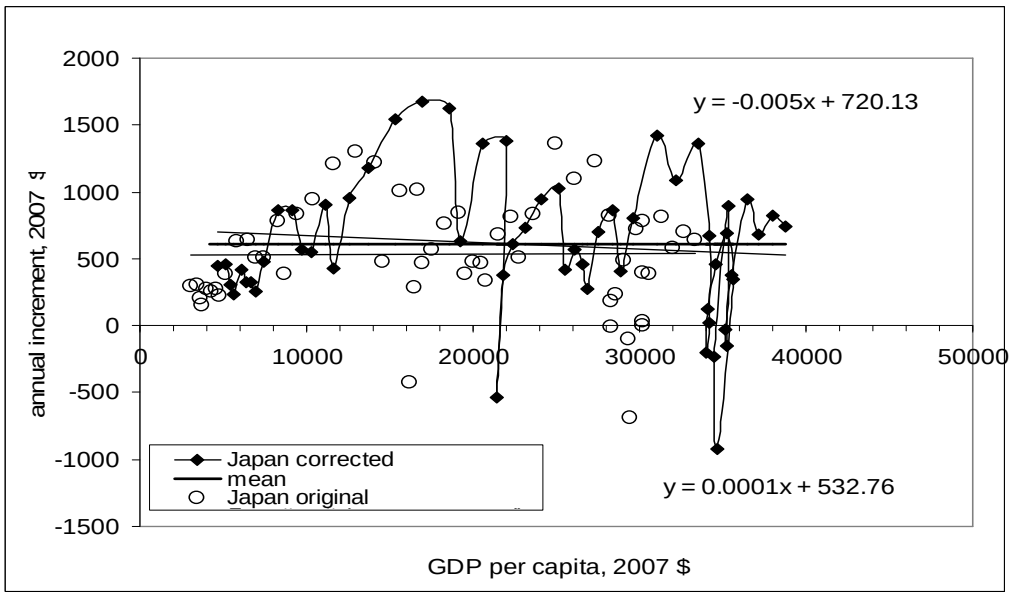


Figure 8. Same as in Figure 4 for Japan. The original linear regression line is parallel to x-axis. The corrected line is characterized by a negative trend. There were two periods of very quick growth between $12000 and $20000 and between $28000 and $33000. Both ended in periods of a low (sometimes - negative) growth rates. Same effect might be expected for Ireland.

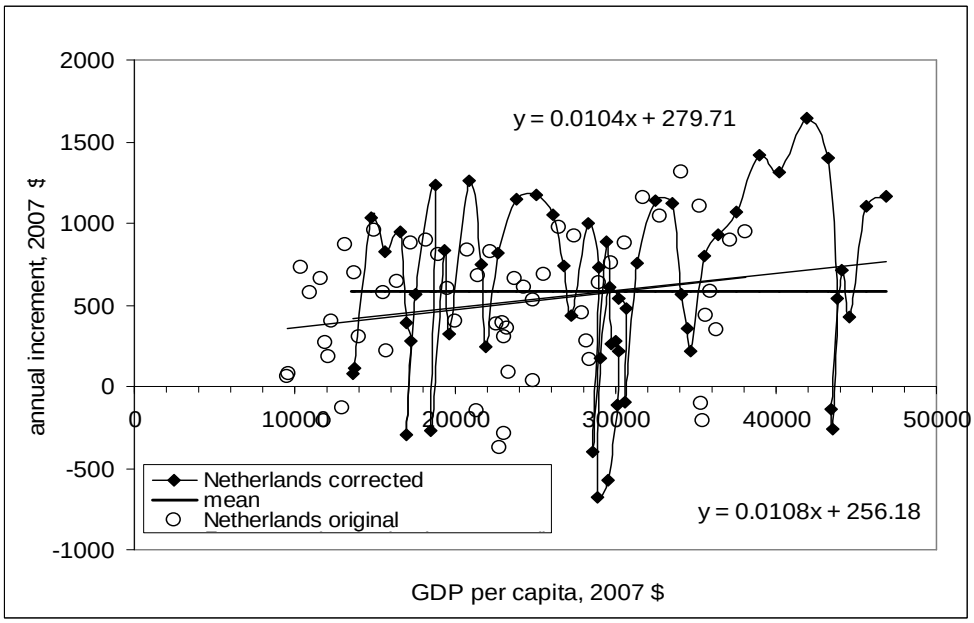


Figure 9 - Same as in Figure 4 for Netherlands. The mean value is $502.

The UK and USA differ only in the mean increment value: $431 and $557, respectively. Positive linear trend is relatively high for the both countries. The US trend is well explained by the change in the nine-year-old population [10]. When corrected for the integral nine-year-olds change between 1950 and 2007, the US mean value is only $462, i.e. in the tight group of the largest economies. The UK statistical agencies do not provide accurate population estimates for the entire period, but from the mean value one can assume that there was no significant increase in the number of nine-year-olds.

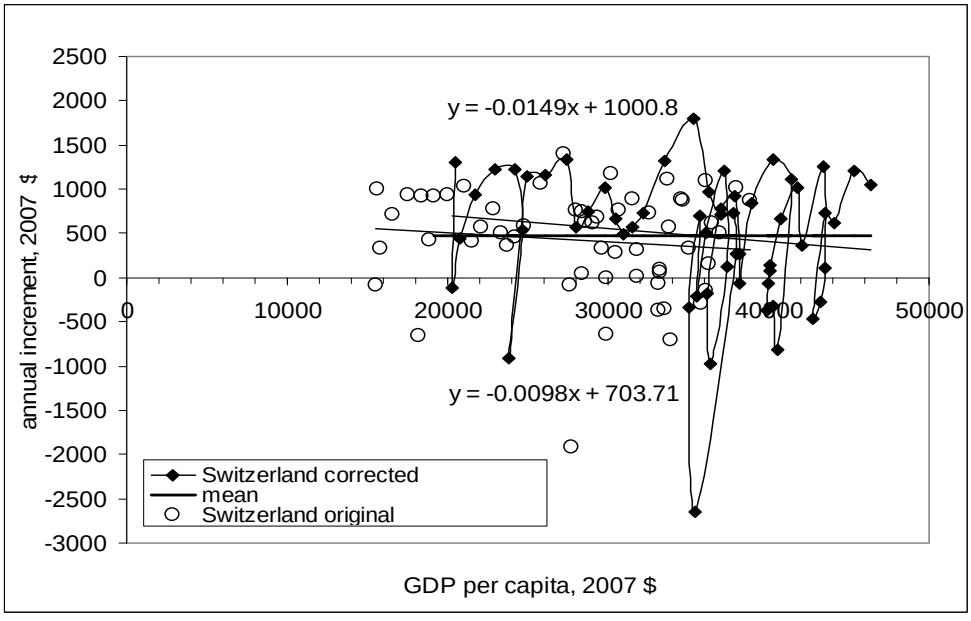


Figure 10. Same as in Figure 4 for Switzerland
The mean value is $426. The country shows a consistent negative trend in the GDP per capita annual increment.

4. Conclusion

The nineteen countries show various types of behaviour of GDP per capita during the period between 1950 and 2007. There are countries with a slightly negative trend of GDP per capita increment: France, Italy, Switzerland, and Japan. Despite the common negative trend the countries have quite different mean increments. Austria is characterized by an almost zero trend and has a large

mean value. One can count it in the club.

A majority of European countries including Belgium, Denmark, Greece, Netherlands, Portugal, Spain, and Sweden are characterized by a slightly positive trend. The mean increment value varies, however, from of $323 for Portugal to $502 in Netherlands. Greece and Portugal showed a weak growth in the beginning of the period, but have recovered to a normal pace. There are two outstanding European countries – Ireland and Norway. Their mean increment is very high, but the countries have a strong downward tendency during the last three to five years. One can expect them to follow the path of Japan – from a strong growth to a long period of stagnation. At the same time, the countries are small. Their influence on the world economy is negligible. Thus, we also deny the countries to influence our analysis of economic trend.

The studied English-speaking countries are characterized by a large positive trend, but should be separated into two groups. The first consists of only one member – New Zealand. The principal characteristic is a very poor performance during the entire period. The second includes Australia, Canada, the UK, and the USA. The mean increment for them is between $431 and $557.

The mean increment in the USA corrected for the total nine-year-olds change between 1950 and 2007 equal to 0.82 is only $557*0.82=$462. The mean value obtained with the correction is by about $100 lower than that without the correction. The accurate population estimates available for the USA allowed explanation for not only the trend, but also the largest fluctuations. Smaller deviations from the mean value are compatible with the characteristic noise of the population estimates and are not so well correlated. For France, this factor is 0.97. The author failed to find reliable data for the other countries under study.

We do not consider the countries with known political and economic problems in the past – Greece, Portugal, New Zealand. Overall, they demonstrated consistent underperformance. Switzerland surprisingly joins the club of weak growth, but the reason might be of a different nature – the decreasing population of the defining age.

The above analysis has revealed that the largest developed economies are characterized by very close values of the mean GDP per capita increment for the period between 1950 and 2007. The mean value defines the long-term economic trend. Thus, the countries are characterized by the same trend level not depending on the attained level of GDP per capita.

A different but important question is: What are statistical properties of the residual growth – fluctuations? In order to answer the question, frequency distributions in $200 (2007 $) wide bins were constructed for each of the original and population corrected data set. The obtained distributions are then approximated by normal distributions with (trail-and-error) parameters fit mainly the segment near corresponding centres.

Figure 11 shows the frequency distribution for the original GDP per capita readings as obtained using “EKS” PPPs. Amplitude of the fluctuations is measured from corresponding mean value for each of the nineteen countries. The distribution is very close to a normal one with the mean value and standard deviation of $0 (the mean value is subtracted) and $400, respectively. The approximating normal distribution is shown by open circles. Both tails of the real distribution are above the predicted values of the normal distribution. This effect is often observed in natural sciences and is associated with inaccurate measurements, limited amount of readings, and sometimes with action of some real factors. One can also suggest that the Levy distribution with heavy tails could better presents the observed tails.

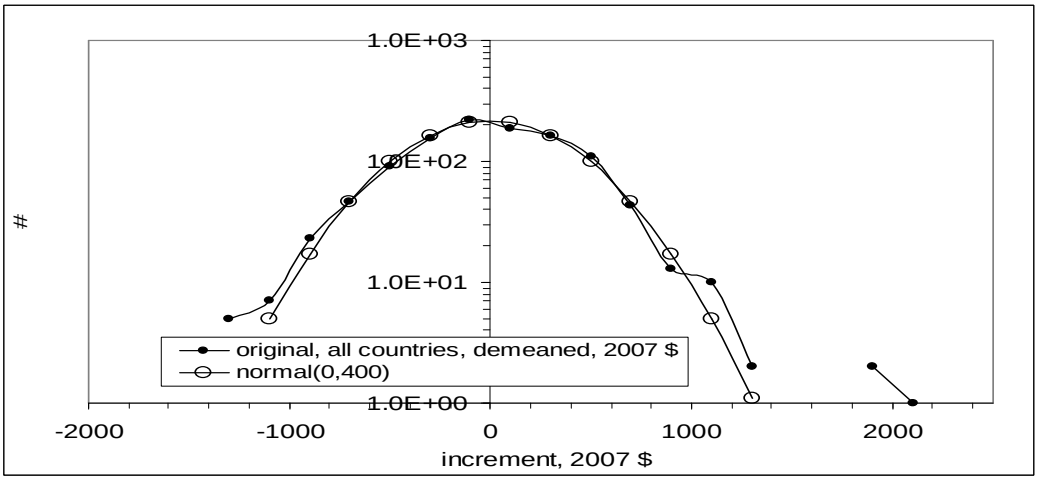


Figure 11. Frequency distribution of the GDP increments for the complete (19 countries) original data set. The GDP values are obtained at "EKS" PPPs. The mean values are extracted from corresponding increments resulting in a zero central value of the distribution. Normal distribution with a zero mean value and standard deviation of $400 is presented by open circles. The normal distribution is very close to the actual one, at least in the central zone.

Figure 12 displays the same curves for the population corrected GDP per capita. Due to relatively narrower bins ($200 original not equal to $200 corrected for the population ratio) the actual distribution is characterized by higher deviations from a normal distribution. At the same time, the central part of the actual distribution is still very close to the normal one.

From the above results, a successful large economy might be characterized by GDP increment randomly fluctuating around some constant level. It is very probable that the fluctuations are normally distributed. This hypothesis is supported by the above observations and the Jarque-Bera test. After removing five largest (in absolute sense) readings, i.e. the outliers likely belonging to some Levy distribution, from each side of the distribution in Figure 11, one can obtain skew ness of -0.095 and kurtosis 3.08. Hence, JB=1.86 for the set of 1073 readings instead of 1083 original readings. Critical JB value at 5% confidence level for two degrees of freedom is 5.99. The obtained JB < 5.99 and one cannot reject the null hypothesis that the data are from a normal distribution. The full set of 1083 readings is characterized by skew ness of 0.035 and kurtosis of 3.48, with JB=11. Obviously, the heavy tails affect the normality test.

There are numerous possibilities to improve convergence of the results if to obtain accurate population data and to enhance the PPP conversion procedure. The mean increment value ~$450 for all countries is a good starting point for calibrating the PPP methodology and evaluation of long term economic performance for developed countries.

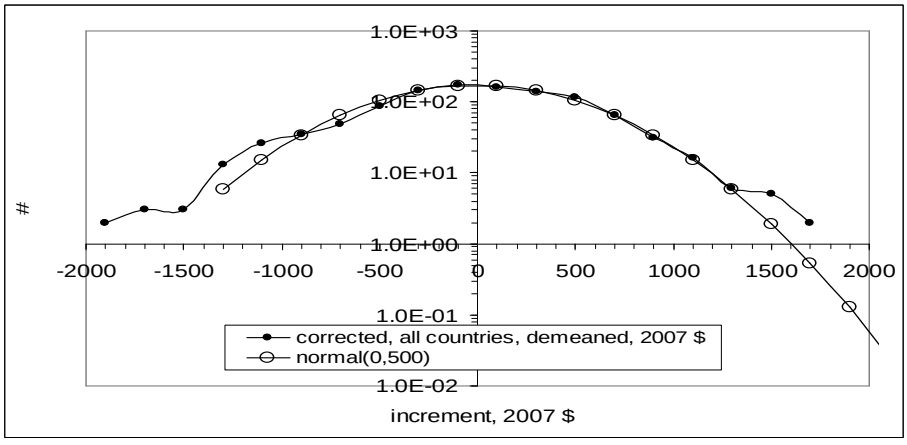


Figure 12. Same as in Figure 11 for the population corrected data set. Larger deviations from the normal distribution are observed

Developing counties also can be evaluated according to their compliance to the principal characteristic for developed countries. One may often hear about a "fast" growth of some developing countries like China and India. There is not criterion, however, to compare their growth rate to that expected in the USA, for example, at the same level of economic development. Using the mean increment, one can easily estimate the pace for any developing country compared to that observed in the developed world. For China, India and the (former) USSR, the increment evolution compared to that for France is represented in Figure 13. One can see that the countries demonstrate increments far below the French mean value (1990 dollars are used as the only available for all the studied countries). Having an intention to catch up a developed economy, any developing country has to analyze its time history of the GDP per capita increment [Kitov, I., (2005)]. No deficiency has to be allowed on the way to prosperity because any gap is created forever judging from the history of such successful developed countries as the USA, France and others.

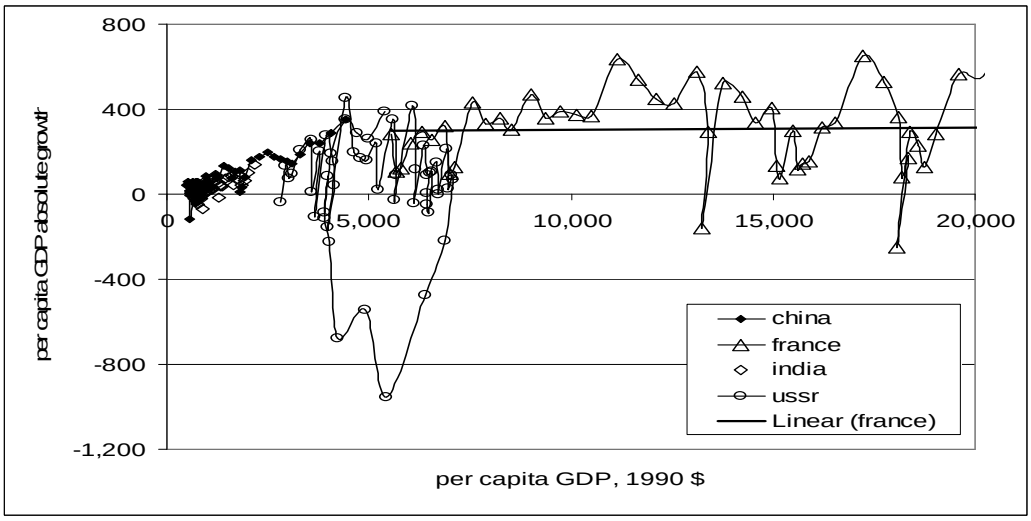


Figure 13. GDP per capita increment for China, India and the (former) USSR for the period between 1950 and 2007 compared to that for France.

GDP is expressed in 1990 $, the only available estimates for the non-OECD countries. India is far below the mean increment for France, but China has just reached the pace of leading developed countries. For the period of existence (between 1950 and 1990 in the study), the USSR was only about a quarter as effective as France.

5. References

[1] Acemoglu, D., Johnson, S., Robinson, J., (2005), *Institutions as the fundamental cause of long-run growth*, in: *Handbook of Economic Growth*, Eds. Aghion, P., Durlauf, S., Elsevier, North-Holland.

[2] The Conference Board and Groningen Growth and Development Centre, (2008), *Total Economy Database*, retrieved November 2008, http://www.conference-board.org/economics/database.cfm

[3] Durlauf, S., Johnson, P., Temple, J., (2005), *Growth econometrics*, in: *Handbook of Economic Growth*, Eds. Aghion, P., Durlauf, S., Elsevier, North-Holland.

[4] Galor, O., (2005), *From Stagnation to Growth: Unified Growth Theory*, in: *Handbook of Economic Growth*, Eds. Aghion, P., Durlauf, S., Elsevier, North-Holland.

[5] Hodrick, R., Prescott, E., (1980), *Postwar U.S. business cycles: an empirical investigation*, Discussion Paper, Northwestern University.

[6] Jones, C., (2005), *Growth and ideas*, in: *Handbook of Economic Growth*, Eds. Aghion, P., Durlauf, S., Elsevier, North-Holland.

[7] Jones, L., Manuelli, R., (2005), *Neoclassical models of endogenous growth: the effects of fiscal policy, innovation and fluctuations*, in: *Handbook of Economic Growth*, Eds. Aghion, P., Durlauf,S., Elsevier, North-Holland.

[8] Kydland, F., Prescott, E., (1982), *Time to Build and Aggregate Fluctuations*, in: *Econometrica*, *Econometric Society*, Vol. 50(6), pages 1345-70, November.

[9] Kitov, I., (2005), *Modelling the transition from a socialist to capitalist economic system*, MPRA Paper 2740, University Library of Munich, Germany, http://ideas.repec.org/p/pra/mprapa/2740.html

[10] Kitov, I., (2006), *GDP growth rate and population*, Working Papers 42, ECINEQ, Society for the Study of Economic Inequality, http://ideas.repec.org/p/inq/inqwps/ecineq2006-42.html

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

[12] 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/

[13] Kitov, I., Kitov, O., Dolinskaya, S., (2007), *Modelling real GDP per capita in the USA: cointegration test*, MPRA Paper 2739, University Library of Munich, Germany, http://ideas.repec.org/p/pra/mprapa/2739.html

[14] Klenow, P., Rodriguez-Clare, A., (2005), *Externalities and growth*, in: *Handbook of Economic Growth*, Eds. Aghion, P., Durlauf, S., Elsevier, North-Holland.

[15] Musso, A., Westermann, T., (2005), *Assessing potential output growth in the Euro area*, European Central Bank, Occasional Paper Series, No 22.

[16] Organization for Economic Co-operation and Development, (2005). *Purchasing Power Parities 2002*, Paris, January 2005, http://www.oecd.org/department/html

[17] Organization for Economic Co-operation and Development, (2006), *Corporate Data Environment, Labor Market Statistics*, DATA, User Queries, January 8, 2006, http://www1.oecd.org/scripts/cde/

[18] Pedregal, D., (2005). *Trend models for the prediction of economic cycles*, in: *Review on Economic Cycles*, International Association of Economic Cycles, Vol. 3(1), December

[19] Pollock, D., (2007). *Investigation economic trends and cycles*, University of Leicester, Working Paper No. 07/17, November 2007.

MODELING THE EFFECT OF BELIEF REVISIONS ON THE SUCCESS OF 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 provides a normative guideline regarding the successful formation of co-branding alliances for both academic researchers and practitioners. We use the expectancy-value model to quantify the mechanism of belief revision in co-branding. Starting from this, an existing mathematical model is adapted in order to investigate (1) the influence of belief revisions on the necessary condition of a successful co-branding alliance (i.e., a sufficient amount of required expansion for the partnering brands) and (2) the existence of an ideal situation that ensures the success. The resulting propositions show that belief revisions can affect a brand's intention with respect to a co-branding partnership. A simulation study demonstrates that an ideal situation exists when the partnering brands are similar in the magnitude of customers' belief revision, brand reputation, and customer loyalty. The present paper advances existing knowledge by relating the success of co-branding partnerships to consumer evaluations. Managerial implications and future research directions are also discussed.

Keywords: Belief revision, success of co-branding, consumer evaluations, mathematical modelling in marketing

JEL Classification: M31

1. Introduction

Over the past two decades co-branding has become an increasingly prevailing strategy for brand managers to leverage existing brand equities [e.g., the *Sony-Ericsson* mobile]. But 40 percent of these strategic alliances failed over a period of four years [cf. Doorley, (1993)]. Since consumer evaluation (i.e., attitudes and attribute beliefs) is regarded as the most important factor determining the success of co-branding [e.g., James, (2005), Hadjicharalambous, (2006)], most of the previous work has utilized the "attitudinal acceptance" of the co-brand and the allying (or partnering) brands [e.g., Simonin and Ruth, (1998), Desai and Keller, (2002), Rodrigue and Biswas, (2004), Walchli, (2007)] to measure the effectiveness of co-branding. However, analyzing the success of co-branding from this behavioural perspective has a critical limitation: the strategic intent (or interest) of a brand to form such an alliance is not fully considered.

To close this research gap, two economic theories – the signaling theory [e.g., Spence, (1973)] and the theory of inter-organizational exchange [e.g., Cook, (1977)] – have been applied to explain the function of the brand name [e.g., Rao and Ruekert, (1994), Rao *et al.*, (1999)] and to discuss the mutual benefits derived from the partnership, respectively [e.g., Bucklin and Sengupta, (1993), Venkatesh *et al.*, (2000)]. In this light, the term "success of co-branding" can be referred to as a "successful (alliance) formation" [Venkatesh *et al.*, (2000)]. However, analyses from this strategic (alliance) perspective are relatively scarce. Thus, the present study attempts to embellish the limited discussions in this field. In particular, we are going to incorporate a basic element in consumer evaluations, namely "attribute belief", which is considered an important aspect of the success of co-branding [Hillyer and Tikoo, (1995)] but the connection between the two has not yet been built up by marketing researchers.

Venkatesh *et al.* (2000)'s work is a good starting point for this purpose. From a strategic point of view, they provided a comprehensive analysis by considering both the effects of signaling and inter-organizational exchange. They assumed that a dynamic co-branding alliance is established to signal each brand's functional expertise. Furthermore, they claimed that the emergence of consumers' "preference change" between the allying brands (i.e., "shift-in preference") is indispensable, because preferences are considered to be the resource owned by each of the brands to be exchanged in the partnership. They argued that, eventually, the two players considered may have an endogenous competition on preferences and thus a certain amount of market expansion for the weak player is

required. Their study offers valuable insights into alliance success but disregards the behavioural contents (in particular the revision of attribute beliefs) behind "shift-in preference".

In comparison with Venkatesh *et al.* (2000), the present study addresses the mechanism of "belief revisions" in co-branding and examines the relation between belief revisions and the successful formation of a co-branding alliance. Accordingly, this paper aims to answer the following two questions:

(1) How can the belief revisions affect the necessary condition for a successful formation?

(2) Does an ideal situation exist that generally ensures a success?

By answering these questions the paper contributes to co-branding research in three ways. First, to the best of our knowledge, it is the first to build up a formal connection between the success of co-branding and consumer evaluations. Further, the present paper provides the rationales behind positive and negative belief revisions in co-branding. Finally, we use a mathematical modeling approach to analyze the relevant relations [Moorthy, (1993)], which is still less prevalent in this field [Huber, (2005)].

The remainder of this paper is organized as follows: Section 2 reviews the relevant literature. Then, in Section 3 we provide a brief description of the Venkatesh *et al.* (2000) model and formulate the mechanism of belief revisions in co-branding. Section 4 adapts the Venkatesh *et al.* (2000) model to offer two propositions regarding the impacts of the negative belief revisions on the necessary condition, and applies the analytical results to visualize the existence of an ideal situation by means of a simulation study. Section 5 finally discusses managerial implications as well as future research directions.

2. Related literature

2.1 Belief revisions in co-branding

Consumers' belief revision (i.e., belief dilution and enhancement) regarding the parent brands is a key issue in brand extension research. Existing studies report that, depending on category similarity or image consistency between the original and the extended products [Grime *et al.*, (2002)], the revision on the pre-extension beliefs about attribute (performance) levels can be negative or positive [e.g., Loken and Roedder John, (1993), Milberg, (2001)]. Indeed, the above process is related to the model of accommodation [Park *et al.*, (1993)]: consumers adapt their pre-extension beliefs to the new levels when they receive new but incongruent attribute information from the extended products. Since co-branding has been recognized as a sub-case of brand extension [Hadjicharalambous, (2006)], belief revisions can also occur when consumers evaluate the co-brand. However, different from brand extension, belief revisions in co-branding can be caused by the partnership [Hillyer and Tikoo, (1995)]. James (2005) further stated that belief revisions may result from the inconsistent attribute information of the co-brand and that incongruence may be the result of different perceptions of the allying brands. Recently, Geylani *et al.* (2008) concluded that the attribute levels of the allying brands can be enhanced, but attribute uncertainty may even be increased after co-branding. In sum, belief revisions in co-branding are also related to the accommodation model but the process is more complicated.

2.2 A specific type of co-branding alliance

This study focuses on a specific type of co-branding, namely the "functional co-branding alliance" [Cooke and Ryan, (2000)], which is established in order to offer a joint (or co-branded) product by integrating the advantageous product-related attributes from each of the allying brands. Before the alliance, both brands produce their products at the same step in the value chain within the same product category, and each brand can be distinguished by different attribute levels. In this aspect, the considered type of partnership is close to co-branding line extension [*cf.* Hadjicharalambous, (2006)]. In the following, we use two dimensions to categorize this type of alliance. The first dimension concerns the intended period and the number of new product releases; the second dimension deals with the purpose of the alliance [Desai and Keller, (2002)]. As shown in Table 1, the one to be analyzed in this paper is short-to-mid term cooperation with several new product releases by modifying the attribute levels of existing attributes of both brands (e.g., a co-branded pizza mixing existing attributes "good-taste" and "low-calories").

Table 1. A categorization of functional co-branding alliances

			Line extension	
			Modifying existing attributes	Adding new attributes
Intended period of alliances/No. of new product releases	Short-to-mid Term	Single release	An opera CD featuring *Paul Plishka* and *Placido Domingo* (mixing "Tenor" with "Bass" of male voice)	A one-time opera concert featuring *Placido Domingo* and *Whitney E. Houston* (mixing "male" with "female" voice)
		Multiple releases	Several releases of co-branded pizzas from brands *Appetite* and *Bio* (mixing "good-taste" with "low-calories")	A series of CDs featuring *Placido Domingo* and Whitney E. Houston
	Long term (usually a joint-venture)		*Fujitsu-Siemens*' PC products	*Sony-Ericsson*'s W-series music phones (adding *Sony*'s "Walkman" function)

Note: indicates the particular co-branding partnership to be analyzed in this paper.

3. An extension to Venkatesh *et al.* (2000)

3.1 Key elements of the Venkatesh *et al.* (2000) model

By referring to the Bass (1969) diffusion model, Venkatesh *et al.* (2000) built up a dynamic framework to investigate the necessary condition of alliance success. In this model, two brands *A* and *B* are the prospective partners in a mid-term co-branding alliance. Initially, at time $i = 1$ (with $i = 1,2,3,\ldots,I$) the market comprises two consumer segments of sizes $M_{A(1)}$ and $M_{B(1)}$ that prefer brand *A* and *B*, respectively. In the baseline situation (i.e., before the alliance is formed), each consumer at segment *A* (*B*) is assumed to adopt the product $J_{A(i)}$ ($J_{B(i)}$) at time *i*, and hence the potential market size for $J_{R(i)}$ is $M_{R(i)}$, where $M_{R(i)} = M_{R(1)}$ (*R* denotes brands with $R = \{A, B\}$).

If the alliance is in effect (see Fig. 1), it will release the *i*-th joint product $J_{AB(i)}$ at time *i* and each of the consumers who prefer *A* or *B* are assumed to adopt one unit of $J_{AB(i)}$ (accordingly, the total market size is at least $M_{A(1)} + M_{B(1)}$) during the intermediate period between time *i* and *i+1*. The authors further argued that, at time *i*+1, the consumers initially belonging to segment *A* (*B*) may change their preference to *B* (*A*) due to their consumption experiences from $J_{AB(i)}$. Therefore, the segment size of *A* can change from $M_{A(i)}$ to $(1-S_{AB})\times M_{A(i)} + S_{BA}\times M_{B(i)}$ and the segment size of *B* can change from $M_{B(i)}$ to $S_{AB}\times M_{A(i)} + (1-S_{BA})\times M_{B(i)}$, where S_{AB} and S_{BA} represent the proportions of consumers who shift their preference from *A* to *B* and from *B* to *A*, respectively (i.e., the shift-in ratios).

The Venkatesh *et al.* (2000) study is based on four main assumptions: (1) the segment are not overlapping and each consumer prefers only one brand at each time point, (2) the consumers will not shift their preferences to a third player within the relevant periods, (3) the shift-in ratios are modeled as time-independent variables, and (4) each of the players will split the (sales) revenue of the joint products according to its share of preference at each time point.

The authors further reported that, eventually, the share of preference will change from $M_{R(1)}/(M_{A(1)} + M_{B(1)})$ at the beginning of the alliance to an equilibrium level (hereafter, the equilibrium share) of $S_{BA}/(S_{AB} + S_{BA})$ for *A* and $S_{AB}/(S_{AB} + S_{BA})$ for *B*. That is, one of the brands can be a loser in the partnership when its equilibrium share is smaller than its initial share. In other words, the weak brand has to acquire more consumers from outside the alliance (requiring a certain amount of market expansion) to maintain its original revenue level. This type of market expansion is regarded as the necessary condition for the successful formation. As a consequence thereof, the

alliance may break up if the anticipated amount of expansion is not forthcoming. The sufficient amount of required expansion, denoted by ΔM_V, is expressed as [Venkatesh *et al.*, (2000)]:

$$\Delta M_V \geq Max\left\{M_{A(1)}\left(\frac{S_{AB}}{S_{BA}}\right)-M_{B(1)}, M_{B(1)}\left(\frac{S_{BA}}{S_{AB}}\right)-M_{A(1)}\right\}, \text{ for } S_{AB}, S_{BA} \neq 0. \quad (1)$$

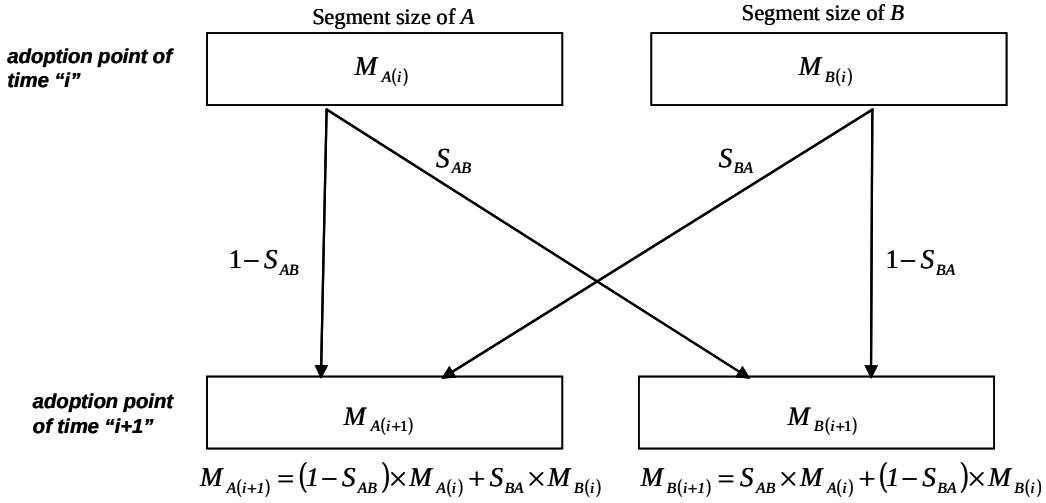


Figure 1. The evolution of alliance according to Venkatesh *et al.* (2000)

3.2 An extended model

We keep the above assumptions (1), (2), and (4) but additionally consider the shift-in ratio as a function of parameters that capture the mechanism of belief revisions. In the following we concretize this mechanism and re-examine equilibrium shares as well as the necessary condition for the successful formation.

3.2.1 Aspects of the market structure

At time *i*, each of the brands either releases its own product $J_{R(i)}$ or cooperates with each other for launching the *i*-th joint product $J_{AB(i)}$. Initially, the market comprises two segments of sizes $M_{R(1)}$ ($M_{R(1)} > 0$) that prefer *A* and *B*, respectively. Different from Venkatesh *et al.* (2000), we name the initial members of segment *A* (*B*) group *a* (*b*) and assume that *a* (*b*) is more familiar with *A* (*B*) than *B* (*A*) within the relevant time periods. In the broader sense group *a* (*b*) can be viewed as the loyal customers of *A* (*B*). If we use $M^G_{R(1)}$ to denote the size of group *G* (*G* = {a, *b*}), i.e. the members staying with brand *R* at time 1, then $M_{A(1)} = M^a_{A(1)}$ and $M_{B(1)} = M^b_{B(1)}$ holds.

Figure 2 once more illustrates the sequence of events according the above descriptions. The adapted model will specify the events that occur during the intermediate period between the first and second time points.

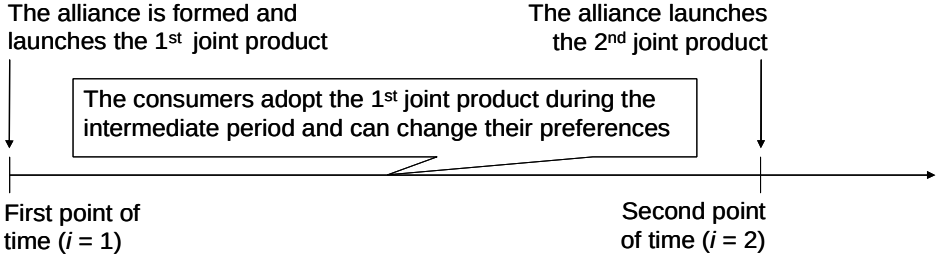


Figure 2. The sequence of events

In the following sections, we will deal with three types of attribute beliefs: (1) the pre-alliance (i = 1) beliefs of the partnering brands (hereafter, the pre-alliance beliefs), (2) the beliefs of the "first" co-branded product (hereafter, the co-branding beliefs), and (3) the post-alliance (i = 2) beliefs of the partnering brands (hereafter, the post-alliance beliefs).

3.2.2 Formation of initial preferences

We apply the expectancy-value model [Bass and Talarzyk, (1972), Fishbein and Ajzen, (1975)] to formulate preferences because it has been widely recommended for explaining preference formation [Agarwal and Malhotra, (2005)]. Two relevant product-related attributes, called x and y, are used to characterize brand A and B. The preference of group G at time i is formulated as a relative score composed of group G's relative weights of attribute importance $w^{K,G} > 0$ (K denotes attributes and $K = \{x, y\}$) and group G's belief of each attribute of each brand $P_{R(i)}^{K,G} > 0$ (for notational simplicity, we will not specify each element for the sets K, G, and R in the remaining of this paper). Group G's preference $\Phi_{R(i)}^{G}$ for each brand can be expressed as:

$$\Phi_{R(i)}^{G} = \sum_{K} w^{K,G} \times P_{R(i)}^{K,G} \ . \tag{2}$$

Practically, $w^{K,G}$ can be measured by asking a representative sample of consumers to divide 100 points between the two attributes, according to how important each attribute is to them. The number of points assigned to each attribute can be used as an indicator of the relative importance of that attribute [Wilkie and Pessemier, (1973); Mackenzie, (1986)]. Besides, $P_{R(i)}^{K,G}$ can be represented by the perceived levels that a specific attribute possesses [Wilkie and Pessemier, (1973)] and can be measured by rating scores in a fixed interval [e.g., from 0 to 100, see Geylani *et al.*, (2008)].

Let us further assume that the consumers belonging to different groups have identical pre-alliance beliefs (i.e., $P_{R(1)}^{K,G} = P_{R(1)}^{K}$ holds) and that attribute x is salient to A whereas y is salient to B. Hence, the initial attribute level of x (y) of A (B) can be assumed to be larger than the initial (i = 1) level of x (y) of B (A) [Geylani *et al.*, (2008)], i.e., $P_{A(1)}^{x} > P_{B(1)}^{x}$ and $P_{B(1)}^{y} > P_{A(1)}^{y}$. Besides, let D^{K} denote the initial attribute-level difference of attribute K between A and B, i.e., $D^{x} = P_{A(1)}^{x} - P_{B(1)}^{x}$ and $D^{y} = P_{B(1)}^{y} - P_{A(1)}^{y}$.

Furthermore, the differences of each attribute are assumed to be the same, i.e. $D = D^{x} = D^{y}$ applies. The above assumption is motivated by referring to Geylani *et al.* (2008, p.736), who also assumed an equal attribute-level difference in their experiment conditions. A positive initial attribute-level difference indicates a better product fit (in terms of attribute complementarily) [Park *et al.*, (1996), Geylani *et al.*, (2008)] but, however, also presents inconsistent attribution information to the consumers [Park *et al.*, (1996)]. Group G's relative weight of attribute importance of attribute K is quantified as $w^{K,G} \in (0,1)$ and we use the different relationships of $w^{K,G}$ to capture the between-group heterogeneity:

$$w^{x,a} > w^{y,a}, \text{ with } \sum_{K} w^{K,a} = 1, \tag{3}$$

$$w^{y,b} > w^{x,b}, \text{ with } \sum_{K} w^{K,b} = 1 \ . \tag{4}$$

That is, group a considers x to be more important, and group b concerns y more. Assuming that group G prefers the brand with the highest score, Eq. (2) to (4) can explain why group a (b)'s initial preference is A (B): $\Phi_{A(1)}^{a} > \Phi_{B(1)}^{a}$ and $\Phi_{B(1)}^{b} > \Phi_{A(1)}^{b}$.

3.2.3 Mechanism of belief revisions

A functional co-branding alliance has two effects on consumer evaluations: (1) the inconsistent attribute information causes confusions about the true levels of the co-branded products [Park *et al.*, (1996), Geylani *et al.*, (2008)] and (2) consumers use the co-branding beliefs to modify their pre-alliance beliefs [James, (2005), Geylani *et al.*, (2008)]. Both effects will be discussed in this subsection. The co-branding beliefs $J_{AB(1)}$ can be modelled as:

$$P^x_{AB(1)} = \lambda^x_A \times P^x_{A(1)} + \lambda^x_B \times P^x_{B(1)} + \varepsilon, \text{ where } \lambda^x_R \in [0,1] \text{ and } \sum_R \lambda^x_R = 1, \quad (5)$$

$$P^y_{AB(1)} = \lambda^y_A \times P^y_{A(1)} + \lambda^y_B \times P^y_{B(1)} + \varepsilon, \text{ where } \lambda^y_R \in [0,1] \text{ and } \sum_R \lambda^y_R = 1. \quad (6)$$

That is, by the theory of information integration [Anderson, (1981)], the pre-alliance beliefs are integrated into the co-branding beliefs [James, (2005), Geylani *et al.*, (2008)]. Therefore, in Equation (5) and (6), λ^K_R denotes the relative contributing weight of each attribute of each brand to the co-branding beliefs. Besides, a random term ε is added to represent the confusions about the true attribute levels (i.e., attribute uncertainty) of the first co-branded product. Hence, the co-branding beliefs are represented by the weighted average of the pre-alliance beliefs plus the confusion ε, which is assumed to be uniformly distributed on the interval $[-\theta,\theta]$. In a similar context, Geylani *et al.* (2008) also assumed that beliefs are symmetrically distributed around the mean. The symmetry of the co-branding beliefs [Equations (5) and (6)] is managed by assuming that the parameters of the uniform distributions are the "additive inverses" of each other. Besides, the uniform distribution is utilized to easily obtain analytical results on the shift-in ratios (i.e., $S^a_{AB(1)}$ and $S^b_{BA(1)}$).

Moreover, we assume that $\theta(D)$ is strictly increasing in *D* because confusions are positively related to the magnitude of the initial attribute-level difference [Geylani *et al.*, (2008)], i.e.,

$$\theta = \theta(D) = \delta D, \quad (7)$$

holds, where $\delta \in$ (0, 1/2) is a confusion parameter. Here, the upper limit of δ ensures that both *x* of *A* and *y* of *B* have a negative revision (see Equations (13) and (16)) and both *y* of *A* and *x* of *B* have a positive revision [see Equations (14) and (15)]. The rationale behind the negative and positive revisions will be provided in section 4.1.

If we posit that both brands contribute the same (i.e., λ^K_R = 1/2) to the co-branding beliefs, Equations (5) and (6) can be rewritten as

$$P^x_{AB(1)} = \frac{1}{2}\left(P^x_{A(1)} + P^x_{B(1)}\right) + \varepsilon, \quad (8)$$

$$P^y_{AB(1)} = \frac{1}{2}\left(P^y_{A(1)} + P^y_{B(1)}\right) + \varepsilon. \quad (9)$$

Assuming an equal attribute-level difference, $P^K_{AB(1)}$ in Equations (8) and (9) can be transformed into

$$P^x_{AB(1)} = P^x_{A(1)} - \frac{1}{2}D + \varepsilon = P^x_{B(1)} + \frac{1}{2}D + \varepsilon, \quad (10)$$

$$P^{y}_{AB(1)} = P^{y}_{A(1)} + \frac{1}{2}D + \varepsilon = P^{y}_{B(1)} - \frac{1}{2}D + \varepsilon . \quad (11)$$

Finally, the post-alliance beliefs of group *G* can be formulated as

$$P^{K,G}_{R(2)} = \gamma^{K,G}_{R} \times P^{K}_{AB(1)} + \left(1 - \gamma^{K,G}_{R}\right) \times P^{K}_{R(1)}, \text{ where } \gamma^{K,G}_{R} \in [0,1]. \quad (12)$$

Equation (12) is inspired by Geylani *et al.* (2008) and, accordingly, the updating weights $\gamma^{K,G}_{R}$ can be used to determine the degree of revision on each attribute of each brand over groups. Substituting $P^{K}_{AB(1)}$ from Equation (10) and (11) into Equation (12) yields the following relations which show the belief revisions of each attribute of each brand for each group:

$$P^{x,G}_{A(2)} = P^{x}_{A(1)} - \frac{1}{2}\gamma^{x,G}_{A} D + \gamma^{x,G}_{A}\varepsilon , \quad (13)$$

$$P^{y,G}_{A(2)} = P^{y}_{A(1)} + \frac{1}{2}\gamma^{y,G}_{A} D + \gamma^{y,G}_{A}\varepsilon , \quad (14)$$

$$P^{x,G}_{B(2)} = P^{x}_{B(1)} + \frac{1}{2}\gamma^{x,G}_{B} D + \gamma^{x,G}_{B}\varepsilon , \quad (15)$$

$$P^{y,G}_{B(2)} = P^{y}_{B(1)} - \frac{1}{2}\gamma^{y,G}_{B} D + \gamma^{y,G}_{B}\varepsilon . \quad (16)$$

3.2.4 Shift-in ratios

Assuming $S^{a}_{AB(1)}$ to be the probability of group *a*'s consumers shifting their preferences from *A* to *B* after having consumed the first joint product:

$$S^{a}_{AB(1)} = Pr\left(\Phi^{a}_{B(2)} > \Phi^{a}_{A(2)}\right), \text{ where } Pr(\bullet) \text{ is a suitable probability function,} \quad (17)$$

and, by substituting $\Phi^{a}_{R(i)}$ in Equation (17) by Equation (2), we get

$$S^{a}_{AB(1)} = Pr\left(w^{x,a}P^{x,a}_{B(2)} + w^{y,a}P^{y,a}_{B(2)} > w^{x,a}P^{x,a}_{A(2)} + w^{y,a}P^{y,a}_{A(2)}\right). \quad (18)$$

Similarly, $S^{b}_{BA(1)}$ denotes the probability of preference change for group *b* and is expressed as

$$S^{b}_{BA(1)} = Pr\left(\Phi^{b}_{A(2)} > \Phi^{b}_{B(2)}\right). \quad (19)$$

By replacing $\Phi^{b}_{R(i)}$ in Equation (19) by Equation (2), we get

$$S^{b}_{BA(1)} = Pr\left(w^{x,b}P^{x,b}_{A(2)} + w^{y,b}P^{y,b}_{A(2)} > w^{x,b}P^{x,b}_{B(2)} + w^{y,b}P^{y,b}_{B(2)}\right). \quad (20)$$

Indeed, $S^{a}_{AB(1)}$ and $S^{b}_{BA(1)}$ also represent the expected shift-in ratios of group *a* and *b*, respectively, because consumers belonging to the same group behave identically.

3.2.5 Equilibrium shares and necessary condition

Since we formulate the shift-in ratio as a function, our dynamical structure is different from Venkatesh *et al.* (2000) (cf. Figure 1). Figure 3 shows this dynamics. According to our setting, from *i* = 2, each segment consists of the members from *a* and *b*, two sub-segments (e.g., $M_{A(2)} = M^{a}_{A(2)} + M^{b}_{A(2)}$ and $M_{B(2)} = M^{a}_{B(2)} + M^{b}_{B(2)}$). Hence, our model can be considered to have

two independent dynamical systems – the evolutions of *a* (i.e., the state variables are $M^{a}_{A(i)}$ and $M^{a}_{B(i)}$) and *b* (i.e., the state variables are $M^{b}_{A(i)}$ and $M^{b}_{B(i)}$).

For example, as shown in Figure 3, the evolution of *a* during the second intermediate period (i.e., between time 2 and time 3) can be explained as follows: A proportion $S^{a}_{AB(2)}$ of the $M^{a}_{A(2)}$ consumers shift their preference to *B* while the remaining $\left(1-S^{a}_{AB(2)}\right)\times M^{a}_{A(2)}$ consumers still stay with *A*. By the same token, a proportion $S^{a}_{BA(2)}$ of the $M^{a}_{B(2)}$ consumers change their preference to *A* and a total amount of $\left(1-S^{a}_{BA(2)}\right)\times M^{a}_{B(2)}$ consumers stay with *B*. Finally, $M^{a}_{A(3)}$ and $M^{a}_{B(3)}$ will equal $\left(1-S^{a}_{AB(2)}\right)\times M^{a}_{A(2)}+S^{a}_{BA(2)}\times M^{a}_{B(2)}$ and $S^{a}_{AB(2)}\times M^{a}_{A(2)}+\left(1-S^{a}_{BA(2)}\right)\times M^{a}_{B(2)}$, respectively. The above process builds up also during the third intermediate period and hereafter. Therefore, to derive the equilibrium share of each brand, the steady states of $M^{a}_{A(i)}$, $M^{b}_{A(i)}$, $M^{a}_{B(i)}$, and $M^{b}_{B(i)}$ must be identified.

We now assume that initially two brands are equally reputed in terms of the segment size [Venkatesh *et al.*, (2000)], so $M_{A(1)}=M_{B(1)}=M$ holds. The equal level of reputation will be relaxed later in the simulation study. Besides, we assume that the belief revision is a one-shot event that occurs only when the customers adopt the first joint product (i.e., between time 1 and time 2). This assumption is based on the need of parsimony and the lack of a theoretical and empirical confirmation in the literature: we found that previous studies in co-branding discuss the belief revision only from the aspect of "static updating" {i.e., pre- and post-alliance; see [Hillyer and Tikoo, (1995), James, (2005), Geylani *et al.*, (2008)]}.

Based on the second assumption, the attribute level of each attribute of each brand will be fixed at I = 2 and, by applying Equations (18) and (20), $S^{a}_{AB(1)}=S^{a}_{AB(i)}=\left(1-S^{a}_{BA(i)}\right)$ and $S^{b}_{BA(1)}=S^{b}_{BA(i)}=\left(1-S^{b}_{AB(i)}\right)$ hold if $i \geq 2$. Hence, the equilibrium of $M^{a}_{A(i)}$, $M^{a}_{B(i)}$, $M^{b}_{A(i)}$, and $M^{b}_{B(i)}$ will be reached at time 2: The steady state of $M^{a}_{A(i)}$ and $M^{a}_{B(i)}$ is $\left(1-S^{a}_{AB(1)}\right)\times M$ and $S^{a}_{AB(1)}\times M$, respectively, whereas the steady state of $M^{b}_{A(i)}$ and $M^{b}_{B(i)}$ is $S^{b}_{BA(1)}\times M$ and $\left(1-S^{b}_{BA(1)}\right)\times M$, respectively. Finally, the equilibrium of $M_{A(i)}$ and $M_{B(i)}$ is $\left(1-S^{a}_{AB(1)}+S^{b}_{BA(1)}\right)\times M$ and $\left(S^{a}_{AB(1)}+1-S^{b}_{BA(1)}\right)\times M$, respectively, and thus the equilibrium share is $\left(1-S^{a}_{AB(1)}+S^{b}_{BA(1)}\right)/2$ for *A* and $\left(1+S^{a}_{AB(1)}-S^{b}_{BA(1)}\right)/2$ for *B*. So, following the logic of Venkatesh *et al.* (2000, p.25), the amount of required expansion will be at least $\left[2/\left(1-S^{a}_{AB(1)}+S^{b}_{BA(1)}\right)-2\right]\times M$ for *A* (denoted by ΔM_{A}) and $\left[2/\left(1+S^{a}_{AB(1)}-S^{b}_{BA(1)}\right)-2\right]\times M$ for *B* (denoted by ΔM_{B}). The proof of the amount of required expansion is available from the first author upon request. Accordingly, the necessary condition for the successful co-branding alliance (denoted by ΔM) reads

$$\Delta M \geq Max\left\{\left[2/\left(1-S^{a}_{AB(1)}+S^{b}_{BA(1)}\right)-2\right]\times M,\left[2/\left(1+S^{a}_{AB(1)}-S^{b}_{BA(1)}\right)-2\right]\times M\right\}. \qquad (21)$$

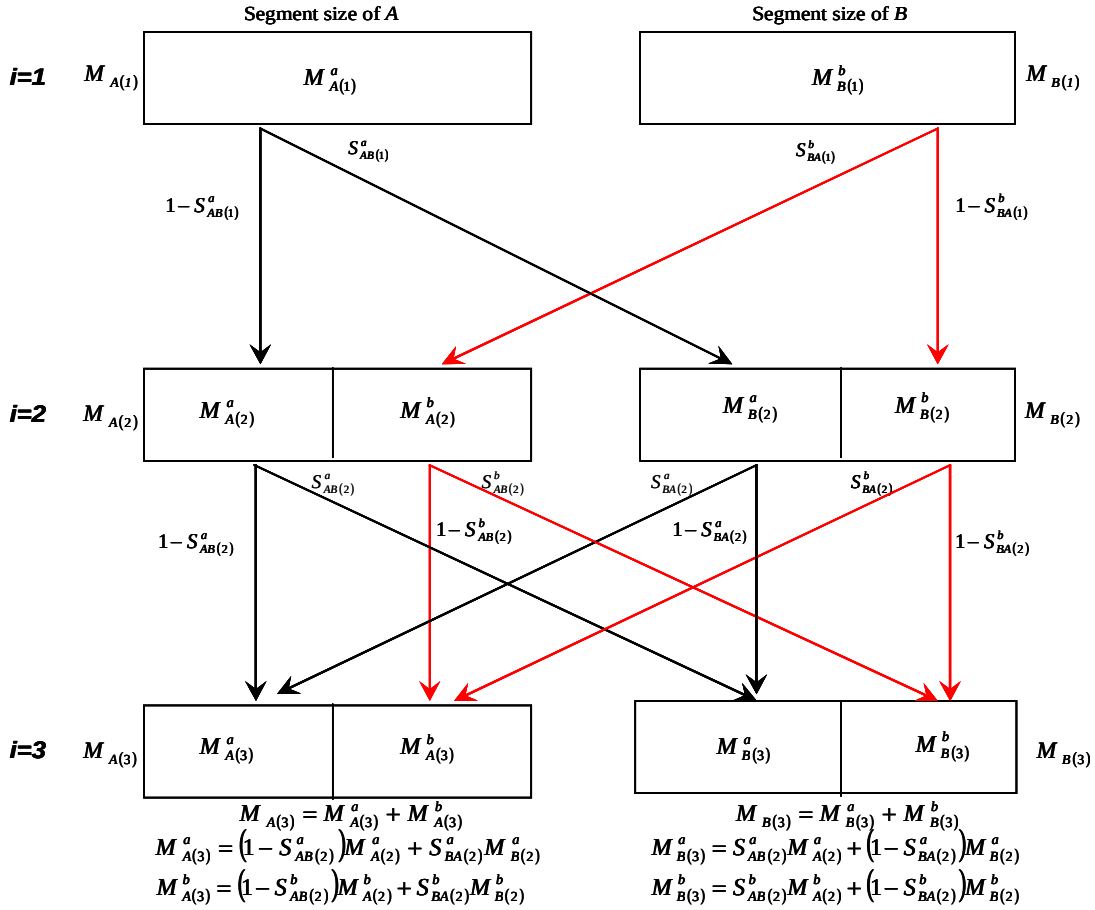


Figure 3. The evolution of alliance from time i = 1 to time i = 3

4. Propositions and simulation study

4.1 Propositions

Let us now get back to the shift-in ratios. Then, by substituting $P^{K,a}_{R(2)}$ ($P^{K,b}_{R(2)}$) from Equation (13) through (16) into Equation (18) and Equation (20), we obtain

$$S^{a}_{AB(1)} = Pr\left\{ w^{x,a} D\left[\frac{1}{2}\gamma^{x,a}_{A} + \frac{1}{2}\gamma^{x,a}_{B} - 1\right] + w^{y,a} D\left[1 - \frac{1}{2}\gamma^{y,a}_{A} - \frac{1}{2}\gamma^{y,a}_{B}\right] \right.$$
$$\left. > \varepsilon\left(w^{x,a}\left(\gamma^{x,a}_{A} - \gamma^{x,a}_{B}\right) + w^{y,a}\left(\gamma^{y,a}_{A} - \gamma^{y,a}_{B}\right)\right)\right\}, \qquad (22)$$

$$S^{b}_{BA(1)} = Pr\left\{ w^{x,b} D\left[1 - \frac{1}{2}\gamma^{x,b}_{A} - \frac{1}{2}\gamma^{x,b}_{B}\right] + w^{y,b} D\left[-1 + \frac{1}{2}\gamma^{y,b}_{A} + \frac{1}{2}\gamma^{y,b}_{B}\right] \right.$$
$$\left. > \varepsilon\left(w^{x,b}\left(\gamma^{x,b}_{B} - \gamma^{x,b}_{A}\right) + w^{y,b}\left(\gamma^{y,b}_{B} - \gamma^{y,b}_{A}\right)\right)\right\}. \qquad (23)$$

Let μ denote the ratio of relative weights of attribute importance (or consumer taste over the two attributes, [Hauser and Shugan, (1983)]) and suppose the following condition holds:

$$\mu = \left(w^{x,a}/w^{y,a}\right) = \left(w^{y,b}/w^{x,b}\right). \qquad (24)$$

Note that the equal ratio of two groups is a benchmark and will be relaxed later in the simulation study. By Equations (3) and (4), Equation (24) implies

$$\mu > 1. \qquad (25)$$

Equation (22) to (25) can now be used for proving some useful propositions.

Now we define the term "negative (positive) belief revisions". Compared with the pre-alliance beliefs of *x* (*y*) of *A* (*B*), the joint product is perceived to have a poorer attribute performance (cf. Equations (10) and (11)). Through the process of "accommodation" (see section 2.1), the pre-alliance beliefs about these two specific attributes may be diluted due to the inconsistency between the existing beliefs and the co-branding beliefs. We call this type of updating behaviour a "negative (belief) revision" [cf. Equations (13) and (16)]. On the contrary, a "positive (belief) revision" may exist [cf. Eq. (14) and (15)] on the pre-alliance beliefs about *y* (*x*) of *A* (*B*), because, in contrast to the pre-alliance beliefs of *y* (*x*) of *A* (*B*), the co-branded product is perceived to provide a better attribute performance. The above arguments also echo the result in Geylani *et al.* (2008) (see Fig. 1 in [Geylani *et al.*, (2008)]).

Besides, brand familiarity has been recognized as an important factor of moderating the belief (or attitude) updates [e.g., Simonin and Ruth, (1998), Sheinin, (2000), Grime *et al.*, (2002)], and therefore we discuss the impact of belief revisions under the following two cases.

Case 1: The consumers of each group are more sensitive to changes of the pre-alliance beliefs of their originally preferred brand.

Case 1 is inspired by Grime *et al.* (2002), who have inferred that a consumer with a higher level of familiarity with one brand tends to update that brand's initial beliefs *more* when she (he) receives inconsistent information from the (co-branded) extended products. Mathematically speaking, this implies

$$\gamma_A^{K,a} > \gamma_B^{K,a} \text{ and} \tag{26}$$

$$\gamma_B^{K,b} > \gamma_A^{K,b}. \tag{27}$$

Proposition 1 (2) illustrates the influence of *negative* revisions under the assumption of Case 1: When *A* (*B*)'s customers have a relative large negative updating on *A* (*B*), Proposition 1 and 2 can exist simultaneously.

Proposition 1: *Under certain conditions ($\gamma_A^{y,a} = \mu\gamma_B^{x,a}$ and $1 \geq \gamma_A^{x,a} > \gamma_A^{y,a} > \gamma_B^{x,a} > \gamma_B^{y,a} > 0$), brand A needs a larger amount of market expansion to form the alliance, ceteris paribus, when the difference between $\gamma_A^{x,a}$ and $\gamma_B^{y,a}$ increases.*

The intuition behind Proposition 1 is that group *a*'s relatively large negative revision on brand *A* can decline *A*'s intention (or interest) for (in) the alliance. Fig. 4 shows that when the customers of *A* (i.e., group *a*) have a relatively larger amount of negative revision on *A* than *B*, the pre-alliance belief of *x* of *A* will be diluted more. Consequently, a larger portion of *A*'s customers will shift their preference to *B* after co-branding. As argued by Venkatesh *et al.* (2000), *A* eventually has to require a relatively large amount of expansion for entering this partnership. Such a condition is a weak prospect for *A*.

Proof. By using Eq. (24) and (26), Eq. (22) can be rearranged as

$$S_{AB(1)}^{a} = Pr\left[\frac{w^{y,a}D\left(\frac{1}{2}\mu\gamma_A^{x,a} + \frac{1}{2}\mu\gamma_B^{x,a} - \mu + 1 - \frac{1}{2}\gamma_A^{y,a} - \frac{1}{2}\gamma_B^{y,a}\right)}{w^{y,a}\left(\mu\gamma_A^{x,a} - \mu\gamma_B^{x,a} + \gamma_A^{y,a} - \gamma_B^{y,a}\right)} > \varepsilon\right]. \tag{28}$$

By canceling out $w^{y,a}$ and assuming $1 \geq \gamma_A^{x,a} > \gamma_A^{y,a} > \gamma_B^{x,a} > \gamma_B^{y,a} > 0$ and $\gamma_A^{y,a} = \mu\gamma_B^{x,a}$, Eq. (28) can be rewritten as

$$S^{a}_{AB(1)} = Pr\left[D\left(\frac{1-\mu}{\mu\gamma^{x,a}_{A} - \gamma^{y,a}_{B}} + \frac{1}{2} \right) > \varepsilon \right]. \tag{29}$$

If we use ρ to represent $\left(\gamma^{x,a}_{A} - \gamma^{y,a}_{B}\right)$, Equation (29) can be expressed as follows:

$$S^{a}_{AB(1)} = Pr\left\{ D\left[\frac{1-\mu}{(\mu-1)\left(\rho + \gamma^{y,a}_{B}\right) + \rho} + \frac{1}{2} \right] > \varepsilon \right\}. \tag{30}$$

Furthermore, letting L be the term $D\left[\frac{(1-\mu)}{(\mu-1)\left(\rho + \gamma^{y,a}_{B}\right) + \rho} + \frac{1}{2} \right]$, we get $S^{a}_{AB(1)} = Pr(L > \varepsilon)$.

Since ε is uniformly distributed on the interval $[-\theta, \theta]$, we get

$$S^{a}_{AB(1)} = (L + \theta)/2\theta, \quad \text{for } -\theta < L < \theta\,. \tag{31}$$

Since $\partial S^{a}_{AB(1)} / \partial L > 0$ and $\partial L / \partial \rho > 0$,

$$\partial S^{a}_{AB(1)} / \partial \rho > 0\,. \tag{32}$$

The anticipated market expansion for brand *A* to forge the alliance (ΔM_A) is at least

$$\left[\frac{2}{1 - S^{a}_{AB(1)} + S^{b}_{BA(1)}} - 2 \right] M\,. \tag{33}$$

If all the other variables are fixed in Equation (33), one can easily confirm that the amount of anticipated expansion for *A* will increase as $\rho = \gamma^{x,a}_{A} - \gamma^{y,a}_{B}$ becomes larger. Q. E. D.

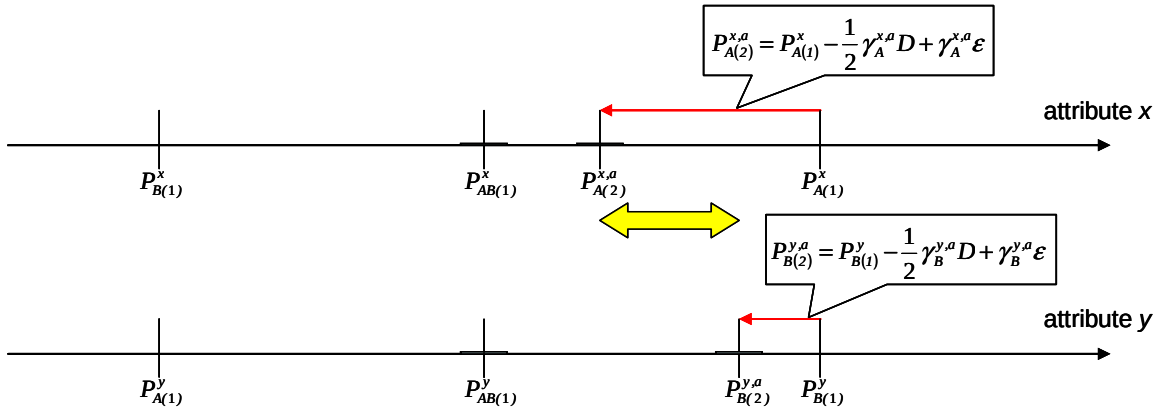


Figure 4. The negative revision of group *a*

Proposition 2: *Under certain conditions* ($\gamma^{x,b}_{B} = \mu\gamma^{y,b}_{A}$ *and* $1 \geq \gamma^{x,b}_{B} > \gamma^{y,b}_{B} > \gamma^{x,b}_{A} > \gamma^{y,b}_{A} > 0$), *brand B needs a larger amount of expansion to form the alliance, ceteris paribus, when the difference between* $\gamma^{y,b}_{B}$ *and* $\gamma^{x,b}_{A}$ *increases.*

The intuition of Proposition 2 is analogous to Proposition 1: group *b*'s relatively large negative revision on brand *B* can decline brand *B*'s interest in this partnership. The proof of Proposition 2 is available from the first author upon request because it is very similar to that of Proposition 1. Obviously, due to the complexity of the process of belief revisions in co-branding, a quantitative

prediction of the outcome of a co-branding partnership is only possible by applying the suggested modeling framework.

Case 2: The consumers of each group do not easily change the pre-alliance beliefs of their originally preferred brand.

Case 2 is contrary to Case 1 and is motivated by Sheinin (2000). Mathematically speaking, it implies

$$\gamma_A^{K,a} < \gamma_B^{K,a} \text{ and} \tag{34}$$

$$\gamma_B^{K,b} < \gamma_A^{K,b}. \tag{35}$$

Under this case, we can also offer two propositions to show the influences of the relative degree of positive revisions on each brand's intention for a partnership. But, since the argumentation is analogous to Proposition 1 and 2, details are skipped here but available from the first author upon request.

From brand manager's perspective, it might be interesting to get a sense of the amount of required expansions for both brands and to check when the required expansion is unlikely to occur [cf. Venkatesh *et al.*, (2000)]. Some simple simulations can serve this need.

4.2 Simulation study

A hypothetical co-branding alliance, formed by brands *Appetite* (*A*) and *Bio* (*B*), is used as an example in the following. The two equally reputed brands are assumed to release several co-branded pizzas, *Appetite-Bio*, on the market. At the beginning of the alliance, *Appetite* (*Bio*) has one group of loyal customers, group *a* (*b*), who are more familiar with *Appetite* (*Bio*) than *Bio* (*Appetite*). The two brands are assumed to be evaluated by two product-related attributes "good-taste" (*x*) and "low-calories" (*y*). Initially, *Appetite* (*Bio*) has a relatively high perceived attribute level on "good-taste" ("low-calories") while *Bio* (*Appetite*) has a relatively low level on the same attribute. This co-branding alliance also presents a better product-fit to the consumers (cf. section 3.2.2). As mentioned in section 3.2.3, after co-branding, the belief of "good-taste" ("low-calories") of *Appetite* and the belief of "low-calories" ("good-taste") of *Bio* will receive a negative (positive) revision.

Starting from the above scenario we will simulate the influences of group *a*(*b*)'s negative revisions on brand *A*(*B*)'s intention regarding a partnership simultaneously (which corresponds to the combination of Proposition 1 and 2). That is, we will utilize *a*'s updating weight of *x* of *A* ($\gamma_A^{x,a}$) and *b*'s updating weight of *y* of *B* ($\gamma_B^{y,b}$) as a set of input variables[1] and observe the corresponding changes of the necessary condition for the successful formation (i.e., the amount of required expansion, ΔM).

In short, the following three scenarios will demonstrate how the necessary condition for the successful formation is affected by the difference of the negative belief revisions between *Appetite* and *Bio* (caused by each brand's loyal customers). Furthermore, we also discuss the existence of an ideal situation.

Scenario 1 assumes that two groups have the same structure of parameters. Scenario 2 and 3 will relax the assumptions in our mathematical model (by assigning different parameter values over the two groups). By offsetting these limitations, our results will be become more robust and realistic.

Proposition 1 is used as an example to show how we select the value for each parameter. Actually, we separate the parameters involved in Proposition 1 into two categories. The first category is called the "brand characteristics" and is composed of the initial segment size of brand *A* ($M_{A(1)}$), the pre-alliance beliefs ($P_{R(1)}^K$), the initial attribute-level difference (D), and the co-branding beliefs

[1] To simulate Proposition 1, we let $\gamma_B^{y,a}$ be a parameter and employ $\gamma_A^{x,a}$ as the input variable to choose different values of $\rho = \gamma_A^{x,a} - \gamma_B^{y,a}$. Analogously, to simulate Proposition 2, we fix the value of $\gamma_A^{x,b}$ and use $\gamma_B^{y,b}$ as the input variable to select different values of $\gamma_B^{y,b} - \gamma_A^{x,b}$.

($P^{K}_{AB(1)}$). The second category is named the "consumer characteristics" and includes the relative weight of attribute importance ($w^{K,a}$), the confusion parameter (δ), and the updating weights ($\gamma^{K,a}_{R}$, excluding the variable $\gamma^{x,a}_{A}$).

For ease of calculation, we let $M_{A(1)}$ = 100 (for notational simplicity, hereafter we drop the time index of the market size in this section). $P^{x}_{A(1)}$ and $P^{y}_{B(1)}$ are set to 80 whereas $P^{y}_{A(1)}$ and $P^{x}_{B(1)}$ are set to 46 (the values are selected by referring to the experimental results of Geylani *et al.*, (2008, p.739). Moreover, according to Equation (8) and (9), $P^{K}_{AB(1)}$ is formulated as the sum of the midpoint between 46 and 80 and confusions. For the "consumer characteristics" category, the value of each parameter is chosen from a set. By Equations (3) and (24), μ is chosen from the set {1.1, 1.2, 1.3 and 1.4}.[2] Besides, δ is selected from the set {0.3, 0.333, 0.367, 0.4}.[3] Finally, to have a different range for ρ, we let $\gamma^{x,a}_{A}$ (hereafter, the negative updating weight of *A*) be the input variable and let $\gamma^{y,a}_{B}$ (hereafter, the negative updating weight of *B*) be a parameter chosen from the set {0.1, 0.2, 0.3}.

Note that for Proposition 1 (or Proposition 2) we will have 48 different examples (or different types of updating behaviours) if we use all combinations of the parameter sets listed above. Details about the 48 examples are available from the first author upon request.

4.2.1 Scenario 1: identical structure of parameters

This scenario shows the evolution of required expansion when the two groups have the same value for all the parameters, i.e., the same type of updating behaviour. In doing so, it will be more straightforward to visualize how the relative magnitude of each group's negative revisions ($\gamma^{x,a}_{A}$ and $\gamma^{y,b}_{B}$) influences the successful formation. Details on the parameters are provided in Table 2. Note that we exclude the cases where the values of $\gamma^{x,a}_{A}$ ($\gamma^{y,b}_{B}$) are smaller than 0.38, because in those cases the corresponding shift-in ratio is 0 and thus is out of consideration [cf. Venkatesh *et al.*, (2000)]. This setting also holds in scenario 2 and 3. Besides, we use our notations to replicate the simulations by Venkatesh *et al.* (2000) [see Figure 2A in Venkatesh *et al.*, (2000)] in Figure 5A and show the result of scenario 1 in Figure 5B. In addition, ΔM, is expressed as a percentage of the initial aggregated size of *Appetite* and *Bio* in Figure 5 (also Figure 6 and Figure 7). Details about the input and output variables in Figure 5 (also Figure 6 and Figure 7) are available from the first author upon request.

Venkatesh *et al.* (2000) showed that, when an equal shift-in ratio of both brands exists, i.e., if $S_{AB(1)} = S_{BA(1)}$ (presented by the black bullet points in Figure 5A) holds, each brand's "equilibrium share" (revenue) remains the same as initial levels and thus no brands act as a loser. We call this case an ideal situation (i.e., without required expansion, ΔM = 0). Our model addresses the importance of belief revisions: an ideal situation can only exist when the magnitude of the negative revisions of *Appetite*'s customer on *Appetite* and *Bio*'s customers on *Bio* is the same (i.e., $\gamma^{x,a}_{A} = \gamma^{y,b}_{B}$, presented by the black bullet points in Figure 5B). In brief, if the updating behaviour of both groups is the same, a similar magnitude of belief revisions can prevent both brands from being worse off in the alliance, thus achieving a successful formation.

[2] As mentioned in section 4.1 μ can also represent different levels of consumer taste over the two attributes. Since we do not want to address extreme consumer tastes, smaller values are considered.

[3] To select the values of the confusion parameter properly, we refer to Geylani et al. (2008), who showed that the standard deviation of consumers' confusion is reasonable between 5.88 and 7.85. We map those values into our setting and thus δ can be chosen from the set {0.3, 0.333, 0.367, 0.4} (i.e., the standard deviation of ε is equal to $(\delta D)/\sqrt{3}$, see Eq. (7)).

Table 2. Details on the parameters in Figure 5B

Group *a*	M	μ	δ	$\gamma_B^{y,a}$	$\gamma_A^{x,a}$
	100	1.4	0.4	0.1	[0.38, 1]
Group *b*	M	μ	δ	$\gamma_A^{x,b}$	$\gamma_B^{y,b}$
	100	1.4	0.4	0.1	[0.38, 1]

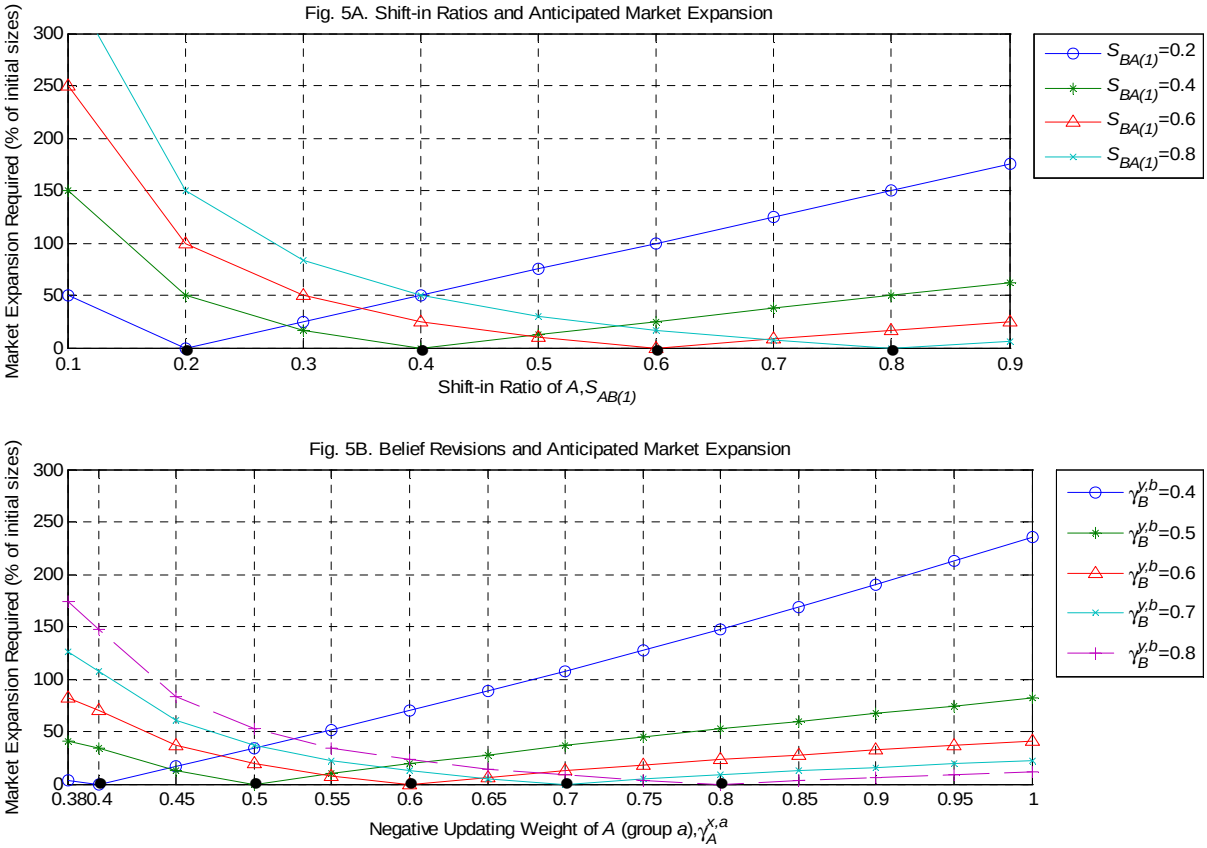


Figure 5. The evolution of anticipated market expansion

4.2.2 Scenario 2: different initial segment sizes

This scenario illustrates the evolution of required expansion when the allying brands' reputations are different (i.e., $M_{A(1)} \neq M_{B(1)}$, ceteris paribus). Details on the parameters are provided in Table 3. Figure 6 demonstrates the evolution of the amount of required expansion. In this scenario, an ideal situation does not exist **(**i.e., the respective curve does not reach the bottom line)[4] for the following conditions: $\gamma_B^{y,b} \geq 0.7$ in Figure 6A, $\gamma_B^{y,b} \geq 0.6$ in Figure 6B, and $\gamma_B^{y,b} \geq 0.5$ in Figure 6C.

[4] In Fig. 6 and 7, the curves with a kink (e.g., $\gamma_B^{y,b} \leq 0.6$ in Fig. 6A) are reaching the bottom (i.e., an ideal situation exist). However, we cannot always reach this specific point (e.g., for $\gamma_B^{y,b} = 0.4$, $\gamma_A^{x,a}$ is around 0.40387 in Fig. 6A) because the values of $\gamma_A^{x,a}$ are chosen with a step size of 0.05.

Table 3. Details on the parameters in scenario 2

Group *a*	M_A	μ	δ	$\gamma_B^{y,a}$	$\gamma_A^{x,a}$
	{75, 50, 25}	1.4	0.4	0.1	[0.38, 1]
Group *b*	M_B	μ	δ	$\gamma_A^{x,b}$	$\gamma_B^{y,b}$
	100	1.4	0.4	0.1	[0.38, 1]

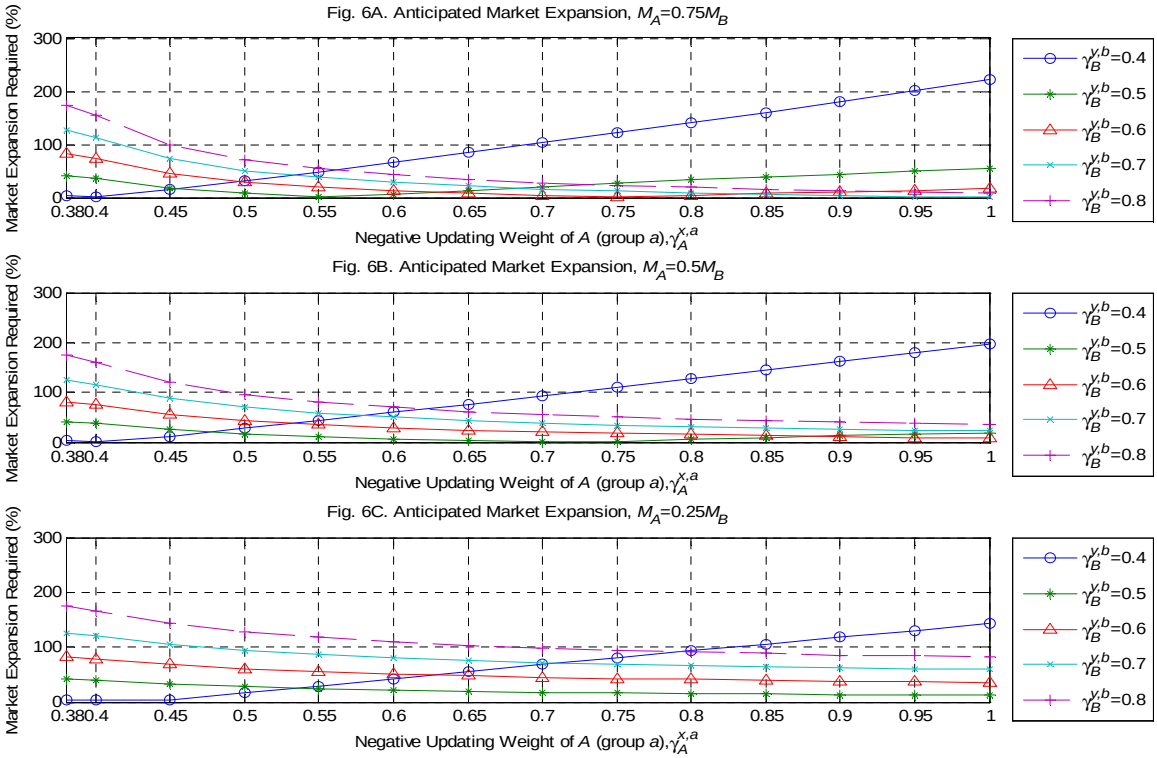


Figure 6. Results for scenario 2 (different initial segment sizes)

The non-existence of an ideal situation can be attributed to the incompatible brand reputations. For instance, when *Bio*'s customers have a considerable negative update on *Bio* (e.g., $\gamma_B^{y,b} \geq 0.6$ in Figure 6B), the equilibrium share of *Bio* in the alliance is always smaller than its initial level of 66.7%. Eventually, *Bio* must expand its market size if it partners with *Appetite*. On the other hand, *Appetite* benefits from the alliance as it can always achieve its initial preference share of 33.3%. Hence, this alliance is not appealing to *Bio* and it would definitely make a retreat. Thus, we can conclude that it is better for *Appetite* and *Bio* to have the same level of "reputation" because this helps both brands acquire a sufficient share from the alliance. In addition, a compatible brand reputation can also be thought of as one type of the "similar resource endowment" [cf. Bucklin and Sengupta, (1993)] since we assume that firms consider consumer preferences as the resource to be exchanged in a partnership.

4.2.3 Scenario 3: different relative weights of attribute importance

In Scenario 3, we allow the customers of *Appetite* and *Bio* to have different ratios of relative weights (i.e., $\mu^a \neq \mu^b$, ceteris paribus). By using the expectancy-value model, a larger ratio of relative weights may contribute to a higher level of attitudinal favorability of one brand (*cf*. Eq. (2) and (24)) and, as a consequence thereof, a higher degree of brand loyalty [Dyson *et al.*, (1996)]. So, the purpose of scenario 3 is to illustrate the evolution of required expansion when *Appetite*'s and *Bio*'s customers have different levels of loyalty to *Appetite* and *Bio*, respectively. Table 4 shows details on

the parameters in scenario 3, whereas Figure 7 presents the related simulation results. Similar to scenario 2, an ideal situation is not likely to occur in Figure 7B (when $\gamma_B^{y,b} \geq 0.7$) and in Figure 7C.

Table 4 - Details on the parameters in scenario 3

Group *a*	M	μ^a	δ	$\gamma_B^{y,a}$	$\gamma_A^{x,a}$
	100	1.4	0.4	0.1	[0.38, 1]
Group *b*	M	μ^b	δ	$\gamma_A^{x,b}$	$\gamma_B^{y,b}$
	100	{1.3, 1.2, 1.1}	0.4	0.1	[0.38, 1]

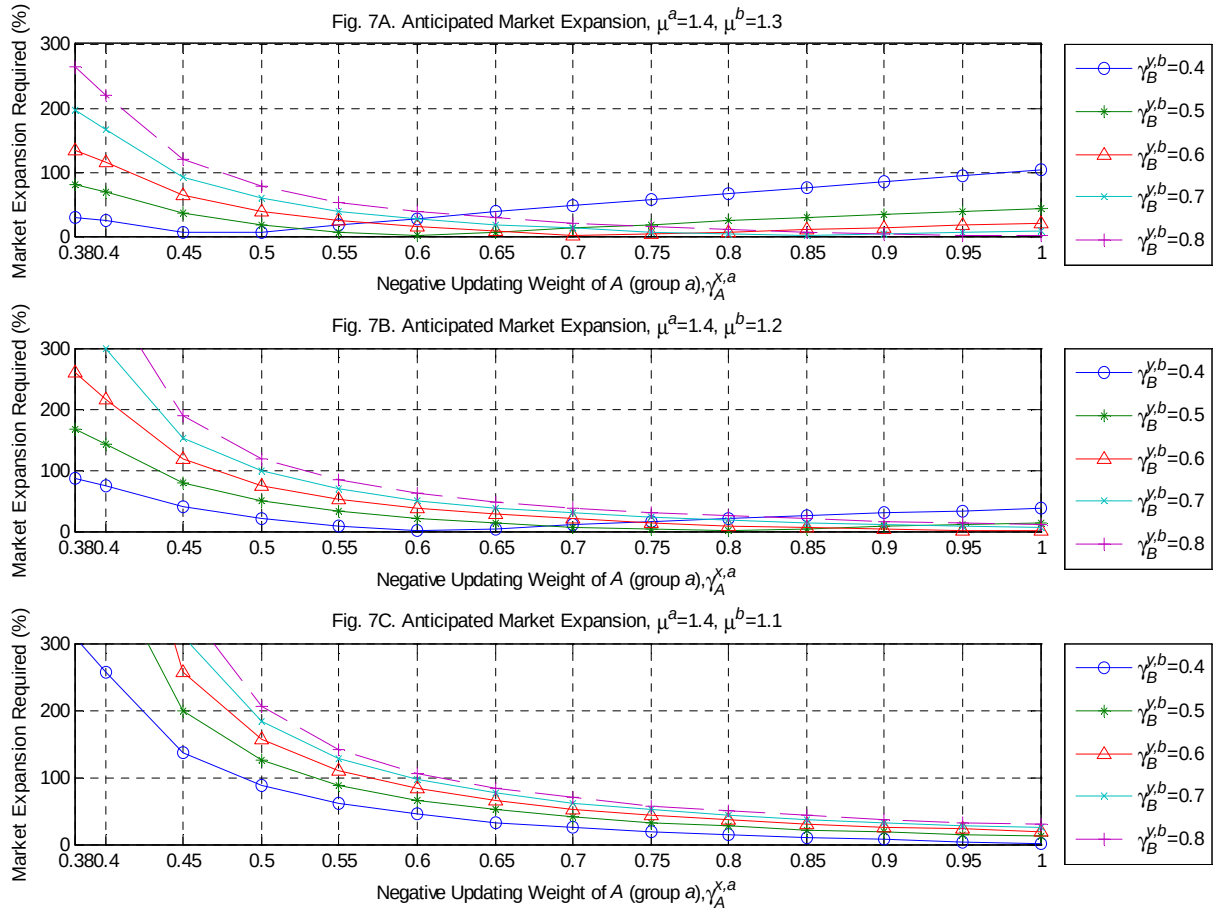


Figure 7. Results for scenario 3 (different relative weights of attribute importance)

In this scenario, the non-existence of an ideal situation is caused by the different levels of customer loyalty of *Appetite* and *Bio*. For example, *Bio* always loses a relatively large amount of its customers when, compared to *Appetite*, the loyalty level of its customers is relatively low (Figure 7C). In this case, *Bio* always has a shrinking equilibrium share (i.e., lower than 50%) if it allies with *Appetite*, and *Appetite* always dominates *Bio* by grabbing a larger equilibrium share (revenue) in the alliance. This might be a major reason for a failure of a partnership [Venkatesh *et al.*, (2000)]. Summing up, it is better for *Appetite* and *Bio* to have an equal level of customer loyalty.

5. Discussion

This paper provides normative guidelines for the successful formation of a co-branding alliance for both academic researchers and practitioners. Our results show that the relative magnitude of customers' belief dilutions on each of the allying brands ($\gamma_A^{x,a}$ and $\gamma_B^{y,b}$) may decline the partnering brands' intentions to ally an alliance [cf. Equations (32) and (33)]. That is, a co-branding alliance may

not be successfully formed even if the attribute complementarity exists (cf. section 3.2.2). Brand managers should also consider the more abstract level of consumer evaluations – namely belief revisions [cf. James, (2005)]. Our simulation study shows that the ideal situation can occur when both brands are similar with respect to the magnitude of customers' belief revision (scenario 1), brand reputation (scenario 2), and customer loyalty (scenario 3). In particular, we would like to emphasize the importance of a compatible "reputation" because it is related to a "free-riding" problem – a less-reputed brand may contribute less but gain more from its partner [e.g., Rao *et al.*, (1999]. In order to avoid this problem, brand managers should carefully check the quantity of loyal customers before initiating or entering a co-branding alliance. In addition, in order to achieve the same level of customer loyalty (μ), brand managers can use persuasive advertisements [Mackenzie, (1986)] to advocate the benefits brought by a specific attribute where one brand excels (e.g., *Appetite*'s good-taste or *Bio*'s low-calories).

There are three possible extensions to our work. First, we assume a static belief update. Future research can use the concept of "state dependence" in order to empirically test whether consumers have a dynamic updating behaviour [e.g., Erdem and Keane, (1996)] in the field of co-branding. Furthermore, we did not consider the theory of attitude accessibility [Fazio *et al.*, (1989)]. Park *et al.* (1996) have argued that an attribute with a larger salience can much easier be recalled from memory and can contribute more to the co-branding beliefs. In this case, the assumption that λ_R^x and λ_R^y are equal to $1/2$ should be relaxed. Finally, future research could measure the effectiveness of co-branding by the market performance. For example, a three-brand scenario – i.e., two allying brands and a competing brand – could be set up and the effects of belief revisions on the relative market share of the co-brand could then be examined. In doing so, we can offset an intrinsic limitation in our model – the customers of the allying brands' can also shift their preference to the competing brand and vice versa.

6. References

[1] Agarwal, J. and Malhotra, N. K., (2005), *An Integrated Model of Attitude and Affect: Theoretical Foundation and an Empirical Investigation*, in: *Journal of Business Research*, Vol. 58(4), pp.483-493.

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

[3] Bass, F. M. (1969), *A New Product Growth for Model Consumer Durables*, in: *Management Science*, Vol. 15(1), pp. 215-227.

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

[5] Bucklin, L. P. and Sengupta, S., (1993), *Organizing Successful Co-marketing Alliances*, in: *Journal of Marketing*, Vol. 57(2), pp. 32-46.

[6] Cook, K. S., (1977), *Exchange and Power in Networks Interorganizational Relations*, in: *The Sociological Quarterly*, Vol. 18(1), pp. 62-82.

[7] Cooke, S. and Ryan, P., (2000), *Global and Transnational Brand Alliances: A Conceptual Investigation of Typologies*, in: *Marketing in a Global Economy Proceedings*, Chicago: American Marketing Association, pp. 336-342.

[8] Desai, K. K., and Keller, K. L., (2002), *The Effects of Ingredient Branding Strategies on Host Brand Extendibility*, in: *Journal of Marketing*, Vol. 66(1), pp. 73-93.

[9] Doorley, T. L., (1993), *Teaming Up for Success*, in: *Business Quarterly*, Vol. 57, pp. 99-103.

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

[11] Erdem, T. and Keane, M. P. (1996), *Decision-making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets*, in: *Marketing Science*, Vol. 15(1), pp. 1-20.

[12] Fazio, R. H., Powell, M. and Willams, C., (1989), *The Role of Attitude Accessibility in the Attitude-to-behavior Process*, in: *Journal of Consumer Research*, Vol. 16(3), pp. 280-288.

[13] Fishbein, M. and Ajzen, I., (1975), *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Reading, MA: Addison-Wesley.

[14] Geylani, T., Inman, J. J. and Hofstede, F. T., (2008), *Image Reinforcement or Impairment: The Effects of Co-branding on Attribute Uncertainty*, in: *Marketing Science*, Vol. 27(4), pp. 730-744.

[15] Grime, I., Diamantopoulos, A. and Smith, G., (2002), *Consumer Evaluations of Extensions and Their Effects on the Core Brand: Key Issues and Research Propositions*, in: *European Journal of Marketing*, Vol. 36(11/12), pp. 1415-1438.

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

[17] Hauser, J. R. and Shugan, S. M., (1983), *Defensive Marketing Strategies*, in: *Marketing Science*, Vol. 2(4), pp. 319-360.

[18] Hillyer, C. and Tikoo, S., (1995), *Effect of Cobranding on Consumer Product Evaluations*, in: *Advances in Consumer Research*, Vol. 22(1), pp.123-127.

[19] Huber, J., (2005), *Co-Branding als Strategieoption der Markenpolitik*. Wiesbaden: Gabler.

[20] James, D., (2005), *Guilty through Association: Brand Association Transfer to Brand Alliances*, in: *Journal of Consumer Marketing*, Vol. 22(1), pp.14-24.

[21] Loken, B. and Roedder John, D. (1993), *Diluting Brand Beliefs: When Do Brand Extensions Have a Negative Impact*, in: *Journal of Marketing*, Vol. 57(3), pp. 71-84.

[22] Mackenzie, S. B., (1986), *The Role of Attention in Mediating the Effect of Advertising on Attribute Importance*, in: *Journal of Consumer Research*, Vol. 13(2), pp. 174-195.

[23] Milberg, S. J., (2001), *Positive Feedback Effects of Brand Extensions: Expanding Brand Meaning and the Range of Extendibility*, in: *Abante*, Vol. 4(1), pp. 3-35.

[24] Moorthy, K. S., (1993), *Theoretical Modeling in Marketing*, in: *Journal of Marketing*, Vol. 57(2), pp. 92-106.

[25] 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*, Vol. 33(4), pp. 453-466.

[26] Park, C. W., McCarthy, M. S., and Milberg, S. J., (1993), *The Effects of Direct and Associative Brand Extension Strategies on Consumer Response to Brand Extensions*, in: *Advances in Consumer Research*, Vol. 20(1), pp. 28-33.

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

[28] Rao, A. R., Qu, L. and Ruekert, R. W., (1999), *Signaling Unobservable Product Quality through a Brand Ally*, in: *Journal of Marketing Research*, Vol. 36(2), pp. 258-268.

[29] Rodrigue, C. S. and Biswas, A., (2004), *Brand Alliance Dependency and Exclusivity: An Empirical Investigation*, in: *Journal of Product and Brand Management*, Vol. 13(7), pp. 477-487.

[30] Sheinin, D. A., (2000), *The Effects of Experience with Brand Extensions on Parent Brand Knowledge*, in: *Journal of Business Research*, Vol. 49(1), pp. 47-55.

[31] 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*, Vol. 35(1), pp. 30-42.

[32] Spence, M., (1973), *Job Market Signaling*, in: *The Quarterly Journal of Economics*, Vol. 87(3), pp. 355-374.

[33] 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*, Vol. 17(1), pp. 3-31.

[34] Walchli, S. B., (2007), *The Effects of Between-partner Congruity on Consumer Evaluation of Co-branded Products*, in: *Psychology and Marketing*, Vol. 24(11), pp. 947-974.

[35] Wilkie, W. L., and Pessemier, E. A., (1973), *Issues in Marketing's Use of Multi-attribute Attitude Models*, in: *Journal of Marketing Research*, Vol. 10(4), pp. 428-441.

Journal of Applied Economic Sciences

SKILLS FOR TOURIST SERVICES

Andreea **MARIN-PANTELESCU**, Nicolae **LUPU**
Academy of Economic Studies – Bucharest, **Romania**
marindeea@yahoo.com, nicolae.lupu@com.ase.ro

Abstract:

Tourism is traditionally seen as a labour intensive industry, which employs large numbers of labour with relatively low levels of human capital. Tourism employment encompasses employment in such diverse sectors as accommodation, catering, attractions, airlines, souvenir shops, tour operation and travel agencies, transportation and so on. The different sectors of tourism are associated with wide range of occupations with diversity in human capital requirements.

The success of tourist destinations is strongly influenced by the quality of their human resources. An important part of the tourist experience comes from the human element, the quality of service, the attitude, responsiveness and helpfulness of the staff adds value to the whole tourist product.

Only competent and motivated employees can deliver high-quality tourism service and achieve competitive advantage for their firms and tourist destinations.

According to these facts, our paper will examine the major characteristics of tourism employment. It will also present a direct marketing research in a travel agency regarding the level of performance for the agency employees. It will study the relationship between the performance of the travel agency and the skills of its employees. Finally, some conclusions and practical solution will be presented with the purpose of raising the level of the human resource abilities and performance.

Key words: tourism, employment, skills, marketing, research

JEL Classification: L83, O14

1. Introduction

First the tourism progress presumes to insure the quality services. Of course, the material groundwork quality of the tourism activity is necessary to achieve this objective but only the competence, education and the personnel's passion for tourism constitute the total quality of the tourist product. Otherwise, in every field the investment realised for training and breaking-through the personnel constitute the main key thereby the companies are successful [Pender, (2005)]. The experience shows that where the tourism services are provided by trained, educated, respectful and specialized personnel the tourist offer is more demanded and it can obtain the custom fidelity [Baum, (2007)]. On vacation or in the case of a business travel, people appreciate a lot their receptivity to solicitations, the simplicity and the speed they check-in with at the hotel, the attendance up to the room, the luggage transportation, the publicity of the services offered and the presentation of the services from the room. In these conditions, the tourism personnel have to adapt to the customers' requires [Constantinescu, (2008)].

The mistakes, the oversights and the lack of politeness in tourism can effect in important financial loss to any tourism unit. More and more, it is admitted the fact that all the prosperity force of a tourism unit comes from the personnel's behaviour [Leslie, Russell, (2006)]. Therefore when a tourist product is conceived it is necessary to assure the training and the breaking-through of the personnel, including the organizing measures which define the recruiting policy, the preparation and the breaking-through of the personnel, the systems of the employees continuance, the solutions for the social problems, the control and the inspection of the tourism services offered.

2. The characteristics of the tourism personnel

According to The World Travel and Tourism Council, the contribution of the Travel and Tourism economy to total employment is expected to rise from 238,277,000 jobs in 2008, 8.4% of total employment, or 1 in every 11.9 jobs to 296,252,000 jobs, 9.2% of total employment or in 1 in every 10.8 jobs by 2018. Regarding Romania, according to The World Travel and Tourism Council, the contribution of the Travel and Tourism economy to total employment is expected to rise from 303,932 in 2008, to 376,428 by 2018 [Tourism Highlights, (2007)].

For characterizing the work force from tourism, together with the quantitative evolutions, the structural mutation has an important significance. The particularity of the tourism activity and its complexity are found in the multitude approach ways of the personnel structure.

The most common way of grouping the tourism workers is to fulfil the function that can be associated with an activity division, with the importance and the training level.

Concerning Romania, regarding the activity sector, we can talk about personnel engaged in hotels industry (45%), food industry (35%), tourism agencies (10%), tourism administration (7%) and other sectors of the tourism industry (3%) as in Figure1. [Lupu, (2005)]:

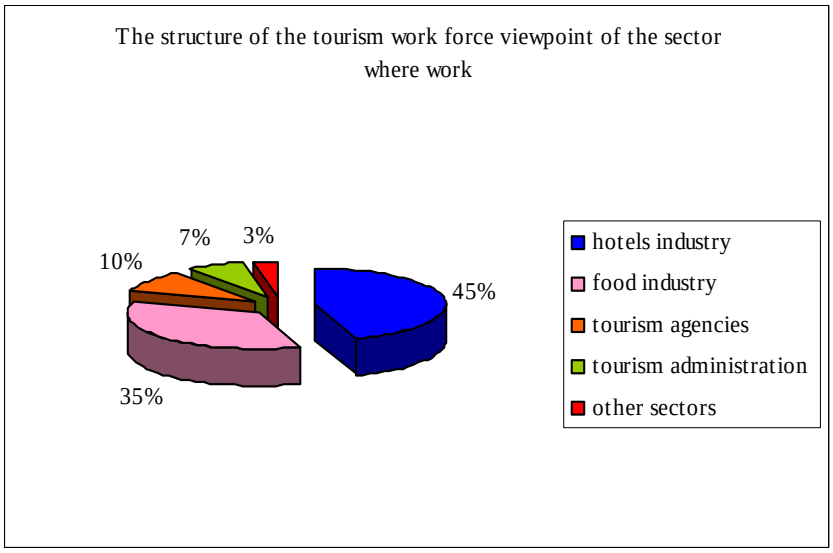


Figure 1. The structure of the tourism work force viewpoint of the sector where work

As you can observe the hotel and restaurant industries offer the most work places for the tourism personnel (about 80%). The reason why this is happening is because of the entrance of the biggest hotel chains such as Marriott, ACCOR, Howard-Johnson, and Ramada who need a lot of work force in our country.

The tourism agencies have known an emphatic development after the 90's, many businessmen starting an activity in this field.

Concerning the tourism administration, in our country the tourism activity is managed by the Ministry of Tourism which has five activity fields: vocational formation, strategy and programme, tourism promotion, international relations, licensing and controlling.

In the other sectors of the tourism industry we can mention the persons hired in the cultural-sportive activities, leisure and entertaining.

If we were to represent it in a grid of functions associated with the training level, the structure of the personnel from the point of view of vocational training would look like this (see the figure no.2) [Stănciulescu, Marin-Pantelescu, (2008)]:

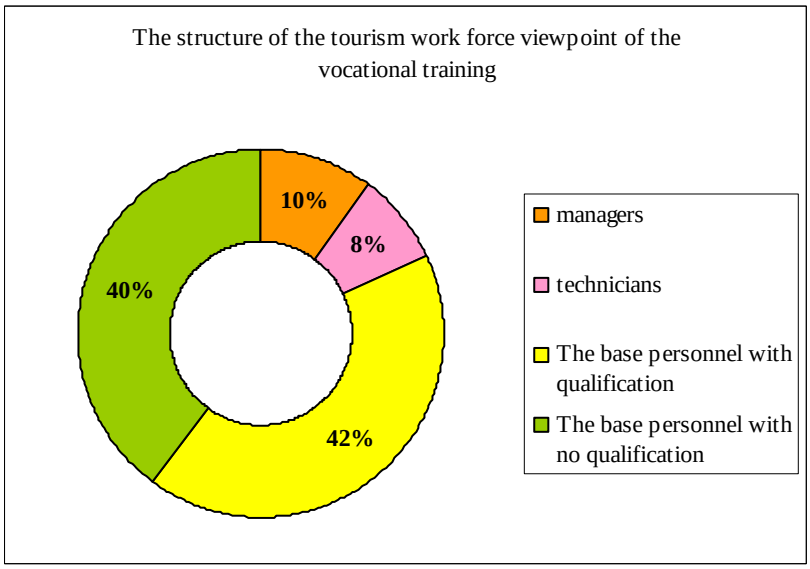


Figure 2. Structure of the personnel from the point of view of vocational training

The leading departments with superior knowledge – director, second director are only 10% of the total tourism personnel. This fact it is due to the big hotel chains that come with their own management and know-how which is used in our country.

The expert technicians that graduated of the postgraduate-schools, colleges and other forms of trainings represent only 8% of the total of the work force. At this chapter our country suffers the most, because the best expert technicians are going to other countries for a well paid job and a better standard of living.

The base personnel who has a middle and below average speciality qualification represents 42% of the total work force and the labourer base personnel with no qualification represents 40% of the total work force. It is noticed a lack of highly trained personnel due to the reduced investments for the training of the work force, due to the low payment and the seasonality of the tourism activity in our country which make the ones who are well trained to leave in foreign countries where they can develop a permanent tourism activity.

The professions from tourism, especially the vocations from the hotel field have *certain characteristics*, for example:

- *superior consumption of alive force* comparing with other economic branches; the necessary work ratio to product item is one of the highest/that suppose an important number of employees and it leads to a more reduced productivity; this situation is due to the fact that automation and mechanization have a limited sphere of applicability;
- *The big mobility of the work force*, seasonal activity leading to the fluctuation of the work force, that requires the use of the extra personnel when the tourism activity is intense;
- *Counter-time work schedule* compared to the usual work schedule (week-ends, official holidays, and personal holidays) brings about big difficulties in recruiting workforce, especially young one;
- *the wide dimension of a work day*, meaning that work in tourism assumes hiring the whole available time per day, effecting negative consequences on the familiar and the social plan;
- *Physical and nervous tiredness*, in the conglomerate periods with a lot of tourists at the peak season;
- *Psychological constraints*, meaning that the personnel who get in direct contact with the tourists has to have self control, good mood and patience;
- *a relatively high and complex level of training*, close to the one of the customers, especially for the employees who get in direct contact with the tourists (knowledge of one of the international languages to be able to provide useful information and a civilized behaviour);
- *High moral and material liability*, by involving the tourism employees in the process of serving the customers; material liability is determined by the material values that the employees take

care of and manage, to which we add the liability for the tourists' assets; moral liability refers to the quality of services or to the degree of satisfying the clients' needs, to the correct drawing up of the bill and keeping confidentiality concerning the clients.

All these characteristics of the tourism vocations have an important role in the recruiting and the hiring process as in the professional development process, too.

3. The tourism personnel requires

The tourism activity assumes a direct and continuous contact between the tourism personnel and the customers. That is why a very important constituent of the quality of the tourism services represents the personnel behaviour that creates the satisfaction and the fullness of the customer.

At the base of the professional behaviour of the tourism workers there are their personal qualities. This way in condition to be a great worker in tourism a person has to fulfil the next requires:

I. Physique inquires:
- Agreeable appearance with harmonious proportions between the parts of the body;
- Physical robust build;
- The gesture and movement harmony;
- Workmanship;
- Good articulation, pleasant voice;
- Good vision, perfect hearing, the sense of smelling;
- Resistance at physical effort, especially orthostatic;
- Healthy organism;

II. Intellectual inquires:
- Active capacity of leasing and understanding;
- Analysis and summarize capacity;
- The facility in assimilation the information;
- Good memory;
- Creative imagination;
- An easy mode of expression.

III. Training inquires:
- Professional qualification;
- General culture;
- A rich, correctly and proper vocabulary;
- PC knowledge;
- Foreign language knowledge.

IV. Psychical inquires:
- Observation spirit;
- Spontaneity;
- The capacity of working in stress conditions;
- Concentrating power and distributive attention;
- Sociability;
- Emotional stability;
- Calm, patience, tactful.

V. Moral inquires:
- Honesty, correctness;
- Responsibility;
- Politeness;
- Discretion;
- Loyalty;
- Punctuality;
- Self respect and to the others.

Depending on the specific of the activity *certain requests* are more important. For example:

↳ For *the reception chief*: the sense of managing, initiative, power of decision, and leader talent (next to the other specific qualities requested for reception workers);

↳ For *the receptionist*: communication abilities, emotional equilibrium, commercial spirit, intuition, dynamism, efficiency.

↳ For *the reservation worker*: very good knowledge of the hotel facilities and of the services offered, attention, rigour, communication abilities, efficiency, seriousness, responsibility;

↳ For *the pay office*: honesty, correctness, responsibility, attention, precision, rigour.

↳ For *the luggage man*: impressive and good looking appearance, a good physical condition, visual memory, watchfulness.

↳ For *the telephonist*: communication abilities, knowledge of the telephone links with the hotel departments, pleasant vocal timbre, and discretion.

↳ For *the chambermaid*: housekeeping, manual abilities, very good sight, correctness, discretion, honesty, distributive attention, good memory.

↳ For *the governess*: above the specific qualities for the subordinated personnel, managing spirit, methodical, leader talent, good taste, responsibility.

↳ For *the waiter*: good physical presence, elegancy, commercial spirit, observation spirit, very good memory, robust, patience, calm, inclination for customers working, courtesy, the sense of sight and hearing very well developed.

↳ In addition for the *quartering director*, the *dining room chief*: leading capacity, rigour, responsibility, leader talent.

↳ For *the cook*: good taste, a very well developed sense of smell, creativity, artistic sense, care for hygiene, neatness, responsibility.

↳ In addition for *the chef cook*: managing abilities, leading

Considering these inquiries, the tourism personnel will be able to make the tourist to feel better and earn his trust. Further he will develop a performance in the tourism activity with important earnings.

4. Case study: The relation between the tourism personnel and the performance of the tourism activity

The direct marketing research took place in Bucharest, between 1st of March – the 30th of April 2008, on a sample of 16 tourism agencies of tour-operators with close dimensions concerning the total assets and the number of the employees. The tourism agencies personnel were estimated, by the aspect of the professional characteristics, using the *professional graph of the tourism worker* (figure no.3) [Liu, Wall, (2006)].

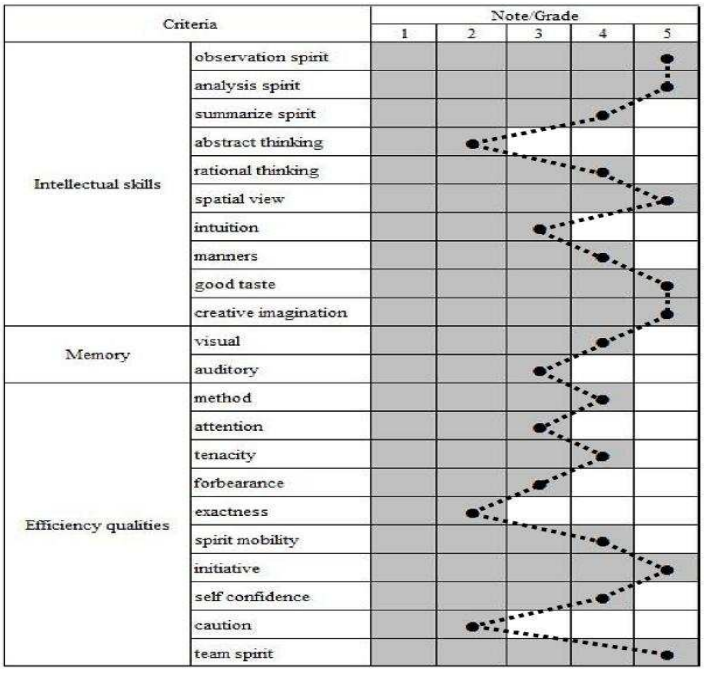


Figure 3. The professional graph of the tourism worker

The research results have shown that as part of the 9 tourism agencies the personnel characterises and the professional graph had a matching percentage of 96. At one criterion "*spatial view*" it was registered a score of 3 instead of 5 as it was required by the professional graph. The investigated travel agencies are in the table below and the 9 travel agencies that are corresponded to the professional graph are Atlantic Tour, Christian Tour, **Eximtur, Happy Tour,** J'info Tours, Jeka Turism, Marshall Tourism, **Paralela 45** and Romantic Travel.

Table 1. The market share of the investigated travel agencies

Nr.	*Investigated travel agencies*	*Market share %*
1	Ali Baba Tour	7,3
2	Atlantic Tour	5,1
3	Big Travel	4,9
4	Christian Tour	6,3
5	**Eximtur**	**10,1**
6	**Happy Tour**	**12,4**
7	Inter Tour Voyage	3,4
8	J'info Tours	9,8
9	Jeka Turism	1,2
10	Kartago Tour	1,3
11	Marshall Tourism	8,8
12	Milenium Tour	3,7
13	Omnia Turism	4,3
14	**Paralela 45**	**14,5**
15	Romantic Travel	5,1
16	Sind Romania	1,8
		100

The tourism agencies were investigated to the viewpoint of the turnover performance realised in April and May in 2008. Among the 9 tourism agencies it is observed a rise of the turnover from a month to another; otherwise the other tourism agencies register no winnings.

The *Paralela 45* travel agency had a turnover of *52 million euro* in 2008, becoming the leader of the Romanian travel agencies.

The second place on the Romanian travel agencies market is occupied by *Happy Tour* travel agency who registered in 2008 a turnover of *51 million euro*, with 5,5% less that the previsions made by their marketing department during the year .

The *Eximtur* travel agency is located on the third place in our top with a turnover of *40 million euro.*

The interview with the travel agencies personnel reviewed that customers are always asking for details, that they preferred travelling to destinations closer to home, more accessible and visiting friends and relatives rather then staying in a resort hotel. As a cause of the economic crisis the decline in average length of stay, as well expenditure is projected to be more pronounced than the decline in overall volume.

The travel agencies market place in Romania was expected to rise in 2008 with 20% (at the time we made the research), and to reach *800 million euro*.

Most of the travel agency managers sustained the fact that their employees are collaborating very good with customers and the customer relation management plan is extremely useful for them to gain more clients each year.

According to our research results we conclude with the fact that it is a direct and strong relation between the tourism personnel characterises and the performances of the tourism agencies.

5. Conclusions

The tourism firms have the objective of increasing the long time prosperity and this is why they have to pay special attention to the preparing of the human resources, especially the particular activity

abilities from tourism that will assure the professionalism of the future services. The differences that exist between the tourism companies are caused by their employees and the competition is not taking place only on the tourism products market, it starts with the work force market where the companies are in competition for their human resources.

In Romania the tourism has the potential to create work places at all of the society and economy levels, from the bar's personnel, chambermaid in hotels, guides, up to the top management in tourism. The tourism is intensive in work and it has to be headed to the quality, this way the sector would be competitive and productive. For all that, many times in our country the tourism is being seen as the last resource when it comes to pick a carrier because of the low salary, the demands for the personnel and the characteristics of the tourism work.

Considering that „politeness does not cost anything but buys everything", the personnel working in tourism should show more kindness, correctness and understanding towards the clients. Only in this way can the clients be made loyal, which leads to increased performances in tourism in the future.

6. References

[1] Pender, Lesley, (2005), *Marketing Management for Travel and Tourism*, Stanley Thorne Publishers, Great Britain

[2] Baum, Tom, (2007), *Human resources in tourism: Still waiting for change*, in: *Tourism Management*, Vol. 28, Issue 6, December 2007, pp. 1383-1399

[3] Constantinescu, Mădălina, (2008), *Knowledge Management through the Lens of Innovation and Labour Productivity in a Knowledge Based Economy*, in: *Journal of Applied Economic Sciences*, Volume III, Issue 2 (4), Summer 2008, ISSN 1834-6110, pp. 85-101

[4] Leslie, David, Russell, Hilary, (2006), *The importance of foreign language skills in the tourism sector: A comparative study of student perceptions in the UK and continental Europe*, in: *Tourism Management*, Vol. 27, Issue 6, December 2006, pp. 1397-1407

[5] Tourism Highlights,(2007), *Facts and Figures*, 2007 Edition, UNWTO

[6] Lupu, Nicolae, (2005), *Hotelul - Economie şi management*, Editura All Beck, Bucureşti

[7] Stănciulescu, Gabriela, Marin-Pantelescu, Andreea, (2008), *Worksheets and exercises for tourism operations techniques*, Editura ASE, Bucureşti

[8] Liu, Abby, Wall, Geoffrey, (2006), *Planning tourism employment: a developing country perspective*, *Tourism Management*, Vol. 27, Issue 1, February 2006, pp. 159-170

TOWARDS A GENERAL THEORY OF LIQUIDITY PREFERENCE

Fernando **MIERZEJEWSKI**
Katholieke Universiteit, **Belgium**
Faculty of Business and Economics, Leuven
femierze@lycos.com

Abstract:

A general theory of liquidity is proposed. The major hypothesis advanced in the paper is that individuals do face borrowing restrictions in capital markets. The value of portfolios combining risky assets and cash balances is then related to the price assigned in some market of deposit insurance, and is accordingly characterised by a method suggested by actuarial researchers — and commonly used by insurers and reinsurers. It is demonstrated that in this way, macroeconomics, financial economics and actuarial sciences fuse together in a unified theoretical framework, which can be applied as an alternative to the utility maximisation approach. Episodes of liquidity crises, which lack an explanation under classic economic theory, are meaningful within the new theoretical setting.

Keywords: Liquidity preference, Economic capital, Deposit insurance, Money demand, Monetary equilibrium, Quantity theory

JEL Classification: G11, G12, G3

1. Introduction

The paradigm of *perfect competition* has predominated in modern financial and economic theory. In this setting, individuals can always borrow the funds required to carry out their consumption and investment projects.

Most of the methods currently used to assess the price of investment portfolios are based on this assumption. At the macroeconomic level, the hypothesis supports the belief that financial institutions do efficiently allocate economic resources.

The empirical evidence, however, suggests that borrowing restrictions are always present in capital markets. A new theoretical setting that explicitly incorporates this condition is proposed in this paper.

2. Why a General Theory of Liquidity?

The presence of borrowing restrictions implies that individuals are not indifferent about the amount of cash holdings to maintain in their portfolios.

The problem is in the interest of private investors and financial conglomerates, for they are obliged to raise capital reserves in order to fix solvency requirements and to provide a guarantee to creditors and customers that they can honour their liabilities. It is also in the interest of insurers and reinsurers, who maintain cash provisions to execute the payments, promised in their issued policies.

In a more general context, the problem is connected to the *preference for liquidity* of individuals, i.e. the amount of balances they are willing to keep in the form of currency - or simply, the *money demand* of the economy.

When deciding the *target* interest rate (consistent with some predetermined levels of inflation and employment) and the corresponding money stock, central bankers must rely on estimations of the money demand - more precisely, they must rely on estimations of the *interest rate elasticity* of the money demand [see e.g. Blanchard, (2005), and Howells and Bain, (2005)]. Hence the central role the preference for liquidity plays in macroeconomics. In spite of this, no theoretical characterisation of this fundamental property has been yet provided.

In conclusion, the concept of *liquidity preference* is connected to problems precisely stated in different economic contexts, namely: the problem of capital allocation, traditionally solved with the aid of tools from financial economics; the determination of insurance *guaranty funds*, for which *actuarial* methods are provided; and finally, the conduction of monetary policy and the estimation of the money demand, for which linear models and statistical inference are used.

The deregulation policies worldwide implemented and gradually deepened during the last thirty years have blurred the distinction between banks, financial intermediaries and insurance companies, and have made the economy more sensible to the fluctuations of the whole financial system.

Such state of affairs shows the need of developing a *unified* framework for the characterisation of liquidity preference, or to put in more ambitious terms, it indicates the need of enunciating a *general theory of liquidity*.

I pretend to convince the reader that the aim can be accomplished by means of a method routinely used in actuarial practice. The method leads to an *optimal liquidity principle*, which is easy to implement, and which can be naturally aggregated to account for the behaviour of financial conglomerates, holdings, industries and the economy as a whole.

3. Borrowing Restrictions in Capital Markets

According to the Modigliani and Miller (1958) proposition, *rational* investors demand no cash provisions, for they are costly and do not affect the market value of aggregate portfolios. The theorem is a consequence of the hypothesis that in perfect markets financial securities and cash balances can be traded at any moment and without quantity restrictions. Then individuals can always remove the imbalances in their portfolios, for they can always borrow and lend *any* amount of capital at a fixed level of the interest rate.

However, financial firms and private investors do face borrowing restrictions in practice. As a matter of fact, lenders charge *premiums* on loans depending on the *credit quality* of their counterparts. Such premiums are incremented with borrowers' leverage ratios.

Borrowing restrictions are likely to appear in markets with *information asymmetries*, which may arise from two different sources: *agency costs* between shareholders and managers, and the *moral-hazard* implicit in the contracts established by firms with their customers.

The problems caused by *agency costs* include mismanagement and under-performance appearing when stockholders cannot fully observe the actions taken by managers - or when due to institutional rigidities they cannot promptly react to reverse undesired results. It can be also the case that managers behave poorly from the point of view of stockholders because they pursue strategies that maximise their own interest. This situation may be especially severe in companies where incentives are not properly established [see e.g. Fama, (1980), Tobin, (1982b), Barnea *et al.*, (1981), Jensen, (1986), and Merton and Perold, (1993)].

Moral-hazard, on the other hand, is induced by the fact that the portfolios held by financial institutions are not observed by their customers, who are thus unable to effectively assess the probability that their deposits will be returned in due time. As stated by Merton (1997), financial firms tend to be *opaque* institutions [see also Ross, (1989)].

Averse-to-risk customers will accordingly show preference for *guaranteed* deposits. Institutional guaranties usually take the form of cash holdings - to be delivered in case of default - that can be reinforced by a warrant issued by another institution or some governmental division, whose capacity and willingness to pay are beyond question. In other words, averse-to-risk customers agree to make deposits only if they are at least partially insured [see also Merton and Bodie, (1992)].

Merton (1974, 1977 and 1978) demonstrates that providing deposit insurance is equivalent to issue a *put option* on the value of the aggregate portfolio. The cost of deposit insurance can then be explicitly stated in terms of the volatility (i.e. the standard deviation) of the series of capital returns of the underlying portfolio.

But the hypothesis of continuous trading is also required in the derivation of the option pricing formula [Black and Scholes, (1973)]. In fact, if individuals can modify the composition of their portfolios at any moment, cash holding strategies can be always replicated by issuing and exchanging option contracts. In other words, *hedging*, *capital cushions* and *deposit insurance* are all perfect substitutes in the absence of borrowing restrictions.

The presence of borrowing restrictions is then reinforced by the fact that *non-standarised* policies are normally traded in insurance markets, which are assigned different prices depending on the information owned by the insurer and insured parties [see e.g. Venter, (1991), Wang *et al.*, (1997), and Goovaerts *et al.*, (2005)].

A new framework for the pricing of deposit insurance that effectively incorporates the presence of borrowing restrictions will be presented in the following section.

4. The Optimal Liquidity Principle

As stated by Tobin (1958), the problem of *liquidity-preference* is exclusively concerned with the determination of *optimal* combinations of risk and cash holdings.

More precisely, Tobin analyses the behaviour of a certain representative decision-maker who holds some aggregate exposure $\boldsymbol{X}$ and a cash balance that can be lent at the interest rate $\boldsymbol{r}$. Assuming that the underlying risk $\boldsymbol{X}$ is described by a *Gaussian* probability distribution and that both capital and securities can be traded without restrictions (in such a way that *every* combination of risk and cash can be attained by performing market operations), he demonstrates that the locus of *efficient* combinations of risk and cash is represented by a straight line in the plane of expected returns and standard deviations.

If additionally the *preferences* of the representative decision-maker regarding the different combined portfolios is characterised by a certain *utility function* (measuring the level of satisfaction that he or she obtains from the wealth produced by portfolios), every level of utility determines an *indifference curve* in the plane of expected returns and standard deviations.

Hence the *optimal* portfolio containing risk and cash holdings is characterised by the combination in the line of efficient portfolios that maximises the expected utility (defined in the mathematical sense) of the representative decision-maker. The optimal portfolio is thus determined at the point where the rate of variation of the expected return with respect to the standard deviation is equal to the marginal utility of substituting a unit of expected return by an additional unit of standard deviation [see equations (3.4), (3.5) and (3.7) in the paper of Tobin, (1958)].

A necessary condition to guarantee the existence of the optimal portfolio is that the representative decision-maker shows *aversion-to-risk*. On these grounds, Tobin regards liquidity preference as *behaviour towards risk*.

Within this context, the method of Tobin determines an *optimal liquidity principle*, which can be applied to describe the demand for liquidity under conditions of *perfect* competition. This method has been actually used to derive an expression for the money demand [see e.g. Holsmtrom and Tirole, (2000), Lucas, (2000), and Choi and Oh, (2003)].

A similar procedure can be followed to obtain a liquidity principle that explicitly incorporates the condition of *restricted borrowing*.

Indeed, notice that if $\boldsymbol{X}$ and λ respectively represent the percentage return of the underlying portfolio of securities and the proportion of funds invested in cash holdings, then the loss afforded at the end of the investment period (per unit of investment) can be expressed as:

$$\left(X+\lambda\right)_{-} = -\min\left(0, X+\lambda\right) \tag{1}$$

The burden of bankruptcy can be transferred to some insurance company provided that the insured party pays the *actuarial* price of the claim, equal to the expected value of the excess of loss [see Goovaerts *et al.*, (1984)]. Rational decision-makers must then choose the proportion of cash that minimises the cost of insurance plus the opportunity cost of capital:

$$\underset{\lambda}{Min}\; E_{\theta}\left[\left(X+\lambda\right)_{-}\right] + r\cdot\lambda \tag{2}$$

where θ denotes a parameter representing the expectations of the decision-maker.

The solution to the optimisation problem of Equation (2) is determined at the point where the marginal reduction in the insurance price, equal to the probability of default, is equal to the marginal cost of capital $\boldsymbol{r}$, i.e.:

$$T_{\theta,-X}\left(\lambda^{*}\right) = P_{\theta}\left\{X < -\lambda^{*}\right\} = r \tag{3}$$

The optimal proportion of cash is thereby characterised by an optimal exchange between a certain (non-random) cash flow and a flow of probability. A *well-defined* liquidity demand function is

obtained in this way, which is always inversely related to the level of the interest rate, for the inverse probability function is always inversely related to its argument:

$$\lambda^* = T^{-1}_{\theta,-X}(r) \tag{4}$$

A general theory of capital can be built on the grounds of Equation (4), which naturally extends the theoretical frameworks of the Modigliani-Miller proposition and the model of deposit insurance of Robert Merton [see Mierzejewski, (2008a)]. The rule can be applied to derive both *centralised* and *decentralised* mechanisms to allocate capital within financial conglomerates [Mierzejewski, (2006) and (2008b)].

More generally, the principle can be used to characterise the money demand of the economy and the monetary equilibrium. A theoretical basis is provided in this way to analyse the effect of monetary interventions in economies where individuals face borrowing restrictions.

Finally, the principle can be also applied to characterise the demand for cash balances in markets of short-term (interbank) loans. The market cost of capital (as determined by the market equilibrium) is then explicitly dependent on the statistical description of risks and the aggregate amounts of supplied and demanded balances.

5. At the Corporate Level

The problem of Equation (2) represents the trade-off faced by a decision-maker who has to decide between establishing an insurance contract (and paying the corresponding actuarial premium) on the one hand, and relying on borrowing and lending (at the interest rate ***r***) in some market of loans on the other.

Within multidivisional corporations, the necessity of keeping divisional cash holdings induces a loss at the aggregate level that can be measured explicitly.

Indeed, let $X_1,\ldots,X_n$ and $\lambda_1,\ldots,\lambda_n$ respectively denote the risks and cash proportions maintained by the divisions of some financial conglomerate, and let X and λ respectively denote the risk and the cash proportion maintained at the aggregate level. Therefore, if $\omega_1,\ldots,\omega_n$ represent the proportions of funds invested in divisions (with respect to the total amount of funds invested by the conglomerate), then the following inequality holds:

$$E_\theta\left[(X+\lambda)_-\right] \leq \sum_{i=1}^{n} \omega_i \cdot E_\theta\left[(X_i+\lambda_i)_-\right] \tag{5}$$

Therefore, as long as capital reserves are held at the divisional level — instead of maintaining a single aggregate balance at the headquarters of the conglomerate — central managers are obliged to minimise the sum of the insurance prices of the divisional claims, and not the insurance price of the aggregate claim, as would be their *first* choice:

$$\underset{\lambda_1,\ldots,\lambda_n}{Min} \sum_{i=1}^{n} \omega_i \cdot E_\theta\left[(X_i+\lambda_i)_-\right] \tag{6}$$

An optimal *centralised* allocation of capital is obtained in this way, which assigns the same cash proportions preferred by divisions when acting as stand-alone independent units, i.e.:

$$\lambda_i^* = T^{-1}_{\theta,-X_i}(r) \quad \forall i = 1,\ldots,n \tag{7}$$

Information asymmetries and differences in the attitude toward risk lead to discrepancies in the estimations of the informational parameter θ proposed by central and divisional managers. The optimal cash proportions determined by the centralised allocation of Equation (7) are thus expected to differ from the cash proportions that divisional managers would choose if they were allowed to decide independently.

An optimal *decentralised* mechanism can be implemented that leads to the same level of capital collected by the centralised allocation at the aggregate level, but which additionally allows central managers to measure the internal differences between the informational parameters.

The optimal decentralised mechanism is implemented in the following way. First central managers determine the internal price of capital and let subsidiaries to choose their cash proportions. Divisions are only allowed to invest their reserves at a current account contracted with the central administration. By comparing the actual cash proportions chosen by subsidiaries with the amounts obtained when applying the centralised allocation of Equation (7), central managers can estimate the aversion parameters of divisional managers [Mierzejewski, (2006) and (2008b)].

6. The Monetary Equilibrium

A fundamental macroeconomic relationship, already suggested by Keynes in his *General Theory* [1936, see the chapter devoted to the *psychological and business incentives to liquidity*; see also Keynes, (1937a), (1937b)], is the one relating the total *stock* of money ***M*** to the level of *nominal* output ***Y*** and interest rates [see e.g. Equation (6) in Friedman, (1970)]:

$$M = Y \cdot \lambda(r) = Py \cdot \lambda(r) \tag{8}$$

where ***P*** and ***y*** respectively denote the level of prices and *real* output, and where $\lambda(r)$ represents the *preference for liquidity* of the economy — i.e. the proportion of output that people maintain in the form cash holdings.

According to Equation (8), altering the money stock ***M*** necessarily implies that at least some of the variables ***P***, ***y*** or ***r*** must change in order to preserve the monetary equilibrium. Consequently, if the rate of change of the level of prices (i.e. the rate of *inflation*) and the rate of growth of real output were pegged to some predetermined levels, the monetary authority would be always able to induce some preferred level of the interest rate by supplying the right amount of money to the economy.

The efficacy of the mechanism depends, however, on the *flexibility* of prices and the *sensibility* of the liquidity-preference function $\lambda(r)$ with respect to the interest rate.

Indeed, if the liquidity-preference function were *perfectly elastic* with respect to the interest rate, then every variation in the amount of money would be completely absorbed by changes in the amount of balances held by the public — no matter the degree of price flexibility. Liquidity-preference is said to be *absolute* in this situation [see e.g. Tobin, (1947), (1972)]. By contrast, if prices were *flexible* and the liquidity-preference function was perfectly *inelastic*, then any variation in the money stock would induce price adjustments in the short-run — and output adjustments in the long-run [Friedman, (1966), (1970), (1971)].

Big controversy has arisen over this issue among economists, due to the consequences to the effectiveness of monetary policy in stimulating national output and reducing unemployment. Researchers and policy makers have been accordingly divided into two different schools: *Keynesians* or supporters of *fiscal* interventions on the one hand, and *monetarists* on the other [see e.g. Modigliani, (1977), and Tobin, (1981), (1993)].

In fact, assuming that the preference for liquidity is *absolute* (or nearly so) Keynesians conclude that money plays no role in the determination of the monetary equilibrium and hence, that only fiscal spending can stimulate the economy to attain *full employment*. This result can be formally obtained in a context of general equilibrium with the help of the *Hicksian IS-LM* model [see Tobin, (1947), (1972) and (1982a), and also Blanchard, (2005), for a presentation of the IS-LM model].

Monetarists alternatively claim that monetary policy does indeed affect real output. In reaching this conclusion, they assume that the opposite hypotheses hold, namely, that prices are *flexible* and liquidity-preference is *non-absolute* — in such a way that variations in the money stock are only *partially* absorbed by changes in the balances held at the aggregate level. Price adjustments are then expected to follow money stock variations in the short-run. As a consequence, production and spending can be encouraged in the short-run by increasing the amount of money in the economy. In the long-run, prices return to their original levels, but at a higher level of real output.

For this mechanism to work efficiently, the growth rate of prices must be pegged to some fixed level. On these grounds, monetarists claim that the major concern of governments must be the rate of *inflation* [Friedman, (1968), (1970), (1971)] — and not the levels of output and employment, as suggested by Keynessians.

A major issue behind the *monetary controversy* is then the empirical assumption over the elasticity of the liquidity-preference function with respect to the interest rate.

Let us investigate how the monetary equilibrium is determined when the preference for liquidity of the economy is characterised by the optimal liquidity principle of Equation (4).

Notice in the first place that if the series of capital returns of national output is described by a *Gaussian* probability distribution with *mean return* μ and *standard deviation* σ and people are *neutral* to risk (in such a way that distorted are equal to non-distorted expectations, i.e. $E_\theta[X] = E[X],\ \forall X$) then the liquidity principle explicitly depends on the risk-parameters:

$$\lambda_{\mu,\sigma}(r) = \sigma \cdot \Phi^{-1}(1-r) - \mu \tag{9}$$

where Φ denotes the cumulative probability distribution of a *standard* Gaussian random variable, whose mean return and standard deviation are respectively equal to zero and one.

Replacing Equation (9) into Equation (8) leads to the following alternative characterisation of the monetary equilibrium:

$$M = Py \cdot \left[\sigma \cdot \Phi^{-1}(1-r) - \mu\right] \tag{10}$$

Within this setting, the monetary equilibrium *not only* depends on the level of prices, real output and interest rates, but also on the risk-parameters μ and σ describing the series of percentage returns of nominal output. Accordingly, even if the rate of change of the level of prices is pegged to a fixed inflation target, the monetary authority *cannot* set the interest rate by simply controlling the money supply ***M***.

The conclusion is that neither *fiscal* nor *monetary* policy can be used alone to induce the economy to some predetermined equilibrium.

Furthermore, recall that the effects of monetary interventions over the monetary equilibrium can be assessed in terms of the *sensibility* or *semi-elasticity* of the demand for cash holdings with respect to the interest rate, a coefficient that is formally defined as:

$$\eta = \frac{1}{\lambda_{\mu,\sigma}(r)} \cdot \frac{d\lambda_{\mu,\sigma}(r)}{dr} \tag{11}$$

From Equations (9) and (11) we obtain:

$$\eta\left(r, \frac{\mu}{\sigma}\right) = \frac{-\sqrt{2\pi}}{\Phi^{-1}(1-r) - \dfrac{\mu}{\sigma}} \cdot \exp\left(\frac{\Phi^{-1}(1-r)^2}{2}\right) \tag{12}$$

Hence the semi-elasticity function may well be equal to infinite in some cases, or in other words, the preference for liquidity of the economy may well become *absolute* under certain circumstances. Actually,

$$\Phi^{-1}(1-r) \to \frac{\mu}{\sigma} \ \Rightarrow\ \eta \to -\infty \tag{13}$$

The states of the economy when this condition is satisfied are thereby regarded as *critical states*.

Therefore, although the semi-elasticity function does indeed remain *stable* over a broad set of combinations of the variables ***M***, ***P***, ***y***, ***r***, μ and σ, or in other words, although the economy can be

well represented by the *monetarist* paradigm within a wide class of sates, the economy can also evolve to states where *Keynesian* fears are confirmed and liquidity-preference becomes absolute.

As a conclusion, albeit monetary policy can sometimes effectively stimulate national output, it can become *suddenly* ineffective if certain paths are followed by the economy.

7. Efficient Markets and Liquidity Crises

Let ***M*** and ***L*** respectively denote the total supply of credit to a certain market of balances and the amount of funds invested on risk.

The equilibrium in the market of loans is attained when aggregate outflows and inflows of capital are equalised. The following relationship thereby describes the market equilibrium when the series of capital returns of the market portfolio follows a Gaussian probability distribution:

$$M = L \cdot \lambda_{\mu,\sigma}(r) \tag{14}$$

Within this context, the *discount* factor $\lambda_{\mu,\sigma}(r)$ explicitly represents the rate at which a unit of investment in the market portfolio is exchanged by a unit of capital, i.e. it represents the *market price of risk*, whilst the interest rate ***r*** represents the return accrued by a unit of capital invested in the market portfolio — or the *internal rate of return on risk.*

Variations in the credit supply ***M*** and the amount of funds ***L*** spent on securities produce two kind of adjustments in Equation (14): adjustments in the price of capital ***r***, and adjustments in the market price of risk that affect the risk-parameters μ and σ. In other words, the equilibrium in the market of balances simultaneously determines the equilibrium in two markets, the market of capital and the market of financial securities.

Two major consequences of the model must be emphasised. In the first place, notice that, from Equation (12), in Gaussian markets the semi-elasticity function η takes *positive* values under some combinations of the interest rate and the risk-parameters. This means that under certain circumstances individuals prefer to lend *all their balances* — and do not maintain cash reserves at all.

More precisely, we can state that people maintain reserves only when the condition $\Phi^{-1}(1-r)-\mu/\sigma > 0$ is satisfied. If instead $\Phi^{-1}(1-r)-\mu/\sigma \le 0$, individuals prefer to exclusively rely on capital markets.

The second major consequence is related to the fact that the magnitude of the semi-elasticity of the demand for balances is equal to infinite in those states when $\Phi^{-1}(1-r)-\mu/\sigma = 0$. People are willing to substitute all their risky assets for cash holdings under such circumstances. These can be regarded as *critical* states of capital markets, which can be corresponded to episodes of *liquidity crises*. Most notably, such critical states can be produced in the middle of a *bullish* trend — i.e. when the mean return of the market portfolio is *greater* than zero.

Although the possibility of the *spontaneous* appearance of *liquidity crises* in the model confirms a well-known and documented fact, it contradicts one fundamental paradigm that has determined the economic policies of many countries during the last quarter of century, namely, the *efficient market hypothesis*.

Recall that the *efficient market hypothesis*, proposed by Eugene Fama (1970, 1998), states that the prices at which securities are actually traded reflect all the available (and relevant) information. In other words, it claims that financial securities are always transacted at a *fair* price. As a consequence, it is impossible to *beat* or *outperform* markets, and hence, in particular, every kind of regulations and trading restrictions can only induce markets to *inefficiently* allocate resources.

Due to the efficient market hypothesis, many scholars have convinced themselves that if *liquidity crises* are observed, they still correspond to the *most efficient* state. Others have claimed that liquidity crises provide evidence that people do not behave *rationally* and hence, that *deregulated* markets do not always arrive to the *most efficient* equilibrium. This assumption is especially appealing, for according to another major economic principle, market prices necessarily reflect *economic fundamentals* — and it is difficult to accept that sudden contractions of the credit supply are the reflection of economic fundamentals.

From the model of equilibrium described by Equation (14), a different conclusion is obtained. Indeed, notice in the first place that every state of the market, including those states related to liquidity crisis - when the semi-elasticity function η is equal to *infinite* - represents the decisions taken by *rational* individuals, who pursue strategies that minimise the total cost of guaranteeing their underlying security portfolios, as stated in Equation (2). Within this context, liquidity crises provide no evidence of *irrationality*.

However, the fact that Equation (14) allows multiple combinations of transacted capital flows and interest rates, which are determined by the risk-parameters μ and σ, implies that *rationality* and *deregulation* are not sufficient conditions to ensure that markets behave *efficiently*. Within this theoretical setting, governments and regulatory authorities must induce markets (through persuasion and mandatory statements) to attain those states that are compatible with some predetermined level of economic performance.

We thus arrive to one of the main consequences of the model of equilibrium described by Equation (14), namely, that the *deregulation* of financial markets is *not necessarily* compatible with *financial stability* and *sustainable economic growth*.

8. Comparative Advantages of the Model and Some Comments on its Implementation

An appealing characteristic of the model presented in this paper is its exclusive dependence on *observable* variables, namely: the amount of money supplied by the central bank and the level of national output, as stated in Equation (10), or alternatively, the amount of funds spent on capital reserves and risky securities, as stated in Equation (14); the level of the opportunity cost of capital, and finally, the expected return and the standard deviation of the series of percentage variations of the level of income. This implies that the liquidity principle of Equation (9) can be effectively applied to describe the liquidity demand in any of the economic contexts previously proposed.

Firstly, at the corporate level, the principle can be regarded as an extension to the capital principle proposed by Merton and Perold (1993, see the technical appendix):

$$\lambda^{MP} = 0.4 \cdot \sigma \sqrt{T} \tag{15}$$

Then the principle of Merton and Perold represents a straight line in the plane of volatilities and capital proportions, which always intercepts the origin and has a constant slope. The principle of Equation (9) also represents a straight line in the plane of volatilities and capital proportions, though its intercept and slope are respectively determined by the expected return and the level of the interest rate. Consequently, a broader range of patterns can be described with the principle of Equation (9) than with the principle of Merton and Perold. In particular, some scenarios are possible with the principle of Equation (9), when the slope of the capital line tends to infinite, i.e. when decision-makers are willing to exchange *any* amount of capital in response to an infinitesimally small volatility movement [Mierzejewski, (2008a)].

Within the more general context of the determination of the liquidity preference of individuals, the superiority of the principle of Equation (9) over the liquidity principle of James Tobin (1958) is demonstrated by the fact that while the former explicitly determines the preferred amount of liquidity provisions as a function of the interest rate, the method of Tobin only provides an implicit rule, which not always leads to a closed expression.

Indeed, if the expected utility operator were used in Equation (2) to represent the price of the contract insuring the loss of the total portfolio — instead of the expected value of the excess of loss — then the following equation would be obtained from the first order condition:

$$\int_{-\infty}^{-\lambda} u'(x+\lambda) dF_X(x) - u(0) \cdot f_X(-\lambda) - r = 0 \tag{16}$$

where F_X and f_X respectively denote the *cumulative* and the *density* probability function of the random variable X (representing the series of percentage returns of the total portfolio) and where u' denotes the first derivative of the underlying utility function. Hence no expression is obtained

characterising the optimal liquidity principle — except for some restricted class of utility functions [see Chapter 2 in Mierzejewski, (2008b)].

As demonstrated, among others, by Lucas (2000), and Holmstrom and Tirole (2000), the utility maximisation approach provides a theoretical basis to derive expressions for the money demand.

As stated by Lucas (2000), such demand functions can be associated with *semi-log* and *log-log* specifications that are well known in macroeconomics. With the aid of these specifications, the observed patterns of real balances and interest rates can be satisfactorily explained. Lucas makes use of this theoretical framework to produce estimations of the gains obtained from reducing inflation.

The model of Lucas, however, depends on various specific assumptions about the economy and householders' behaviour — and also on some non-observable parameters. Besides, Lucas regards the money demand function as a *steady equilibrium relation*, which is accordingly associated with long-term fundamentals, not suited to account for short-term fluctuations.

Semi-log and *log-log* representations of the money demand can be also obtained from the liquidity principle of Equation (7), as long as the series of output percentage returns is respectively assumed to follow *Exponential* and *Pareto* probability distributions [Mierzejewski, (2008a)]. This means that the model proposed in this paper can be naturally integrated with the existing literature of monetary economics. Moreover, it requires of much fewer assumptions than the model of Lucas, and since it is explicitly formulated in terms of short-term output variations, it can effectively describe the effect of short-term contracts in capital markets.

In other words, although both the model of Lucas and the model proposed in this paper lead to the same (*semi-log* and *log-log*) functional relationships of the money demand, in the former setting the underlying parameters *exclusively* depends on long-term fundamentals, whilst in the later, they may depend *both* on long-term fundamentals and short-term fluctuations (a more detailed discussion on the subject can be found in Chapter 2 in Mierzejewski, 2008b).

Holmstrom and Tirole (2000), on the other hand, use the utility maximisation approach to derive the demand for liquidity of corporations faced to borrowing restrictions, which is similar from a broad perspective to the model proposed in this paper. Their model, however, requires of much more and stronger hypothesis.

Firstly, in the model of Holmstrom and Tirole (2000) individuals only invest in projects with *constant returns to scale*. Uncertainty is then introduced as the *probability of success* of the project, while *moral-hazard* is represented by the effort that managers exert to take better investment decisions. Notice that in this way both *uncertainty* and *moral-hazard* are measured by means of *non-observable* variables. Additionally, borrowers and lenders are assumed to be *neutral to risk*, and lenders obtain zero profit — i.e. they act in a competitive environment.

Finally, liquidity shocks are introduced in the model as *exogenous events* that affect the total benefit accrued by the conglomerate. The magnitude of the shock is distributed according to some probability distribution.

In the model proposed in this paper, by contrast, only *observable* variables intervene. Thus, uncertainty is associated with the randomness of the series of output variations, while moral-hazard is explicitly measured by the level of raised capital — on the grounds of a deposit insurance contract celebrated by the conglomerate with some insurance company [Mierzejewski, (2008a)].

Within this context, as explained in details in Section 7, liquidity shocks are *endogenous* events, interpreted as physical adjustments produced in markets with *rational* investors — who seek to minimise their total exposure. Liquidity shocks are thus corresponded to certain *critical states* characterised by certain combinations of the risk parameters and the interest rate. The aversion-to-risk of decision-makers plays thus no determinant role in the formation of liquidity crises, although it can exacerbate its effects under certain circumstances [see Chapter 3 in Mierzejewski, (2008b)].

In conclusion, it has been demonstrated that the new theoretical setting presented in this paper — for the characterisation of the demand for liquidity — can reproduce many of the results traditionally obtained by means of the utility maximisation approach. Besides, since *fewer assumptions* are required in the new theoretical framework, and only *closed* expressions are obtained, which exclusively depend on *observable* variables, we can regard the new approach as a *superior* alternative, both from the point of view of its theoretical and empirical possibilities.

9. Conclusions

A general theoretical framework has been presented to characterise the demand for liquidity in different economic contexts.

The major hypothesis advanced in the paper is that people do face *borrowing restrictions* in capital markets. Then individuals holding assets with random outputs are exposed to imbalances produced by contingent claims that are transacted in some market of deposit insurance.

The model provides a common base to describe the capital needs of consumers, firms, private investors and financial conglomerates, as well as the aggregate balance demanded by industries and the economy as a whole. Hence, it can be applied both at the corporate and macroeconomic level.

A *unified* theory of liquidity is thus obtained, which effectively fuses financial economics, macroeconomics and actuarial science.

Episodes of liquidity crises, which lack an explanation at the light of classic economic theory (influenced by the paradigm of perfect competition and the efficient markets hypothesis), are *meaningful* within the new theoretical setting.

10. References

[1] Barnea, A., Haugen, R.A. and Senbet, L.W., (1981), *An Equilibrium Analysis of Debt Financing Under Costly Tax Arbitrage and Agency Problems*, in: *The Journal of Finance*, Vol. 36 (3), pp. 569-581.

[2] Black, F. and Scholes, M.J., (1973), *The Pricing of Options and Corporate Liabilities*, in: *Journal of Political Economy*, Vol. 81 (3), pp. 637-654.

[3] Blanchard, O., (2005), *Macroeconomics*, Prentice Hall.

[4] Choi, W.G. and Oh, S., (2003), *A Money Demand Function with Output Uncertainty, Monetary Uncertainty, and Financial Innovations*, in: *Journal of Money, Credit and Banking*, Vol. 35 (5), pp. 685-709.

[5] Fama, E., (1970), *Efficient Capital Markets: a Review of Theory and Empirical Work*, in: *Journal of Finance* Vol. 25 (2), pp. 383-417.

[6] Fama, E., (1980), *Agency Problems and the Theory of the Firm*, in: *Journal of Political Economy*, Vol. 88 (2), pp. 288-307.

[7] Fama, E., (1998), *Market Efficiency, Long-Term Returns and Behavioural Finance*, in: *Journal of Financial Economics*, Vol. 49, pp. 283-306.

[8] Friedman, M., (1966), *Interest Rates and the Demand for Money*, in: *Journal of Law and Economics*, Vol. 9 (Oct. 1966), pp. 71-85.

[9] Friedman, M.. (1968). *The Role of Monetary Policy*, in: *The American Economic Review*, Vol. 58 (1), pp. 1-17.

[10] Friedman, M., (1970), *A Theoretical Framework for Monetary Analysis*, in: *The Journal of Political Economy*, Vol. 78 (2), pp. 193-238.

[11] Friedman, M., (1971), *A Monetary Theory of Nominal Income*, in: *The Journal of Political Economy* Vol. 79 (2), pp. 323-337.

[12] Goovaerts, M.J., Van Den Borre, E. and Laeven, R., (2005), *Managing Economic and Virtual Economic Capital within Financial Conglomerates*, in: *North American Actuarial Journal*, Vol. 9 (3), pp. 77-89.

[13] Holmstrom, B. and Tirole, J., (2000), *Liquidity and Risk Management*, in: *Journal of Money, Credit and Banking*, Vol. 32 (3), pp. 295-319.

[14] Howells, P. and Bain, K., (2005), *The Economics of Money, Banking and Finance*, Prentice Hall.

[15] Jensen, M.C., (1986), *Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers*, in: *The American Economic Review*, Vol. 76 (2), pp. 323-329

[16] Keynes, J.M., (1936), *The General Theory of Employment, Interest and Money*, London: Macmillan (reprinted 2007).

[17] Keynes, J.M., (1937a), *The General Theory of Employment*, in: *The Quarterly Journal of Economics* Vol. 51 (2), pp. 209-223.

[18] Keynes, J.M., (1937b), *The Ex-Ante Theory of the Rate of Interest*, in: *The Economic Journal,* Vol. 47 (188), pp. 663-669.

[19] Lucas, R.E., (2000), *Inflation and Welfare*, in: *Econometrica*, Vol. 68 (2), pp. 247-274.

[20] Merton, R.C., (1974), *On the Pricing of Corporate Debt: The Risk Structure of Interest Rates*, in: *Journal of Finance*, Vol. 29 (2), pp. 449-470.

[21] Merton, R.C., (1977), *An Analytic Derivation of the Cost of Deposit Insurance and Loan Guarantees: an Application of Modern Option Pricing Theory*, in: *Journal of Banking and Finance*, Vol. 1 (1), pp. 3-11.

[22] Merton, R.C., (1978), *On the Cost of Deposit Insurance when there are Surveillance Costs*, in: *Journal of Business,* Vol. 51 (3), pp. 439-452.

[23] Merton, R.C., (1997), *A Model of Contract Guarantees for Credit-Sensitive, Opaque Financial Intermediaries*, in: *European Finance Review*, Vol. 1 (1), pp. 1-13.

[24] Merton, R.C. and Bodie, Z., (1992), *On the Management of Financial Guarantees*, in: *Financial Management*, Vol. 21 (4), pp. 87-109.

[25] Merton, R.C. and Perold, A.F., (1993), *Theory of Risk Capital in Financial Firms*, in: *Journal of Applied Corporate Finance*, Vol. 5 (1), pp. 16-32.

[26] Mierzejewski, F., (2006), *Optimal Capital Allocation Confronting Bankruptcy and Agency Costs*, in: *Bank- en Financiewezen*, Vol. 2006 (2), pp. 72-77.

[27] Mierzejewski, F., (2008a), *The Economic Capital of Opaque Financial Institutions*, in: *Journal of Applied Economic Sciences*, Vol. III (Issue 3, Number 5), pp. 232-245.

[28] Mierzejewski, F., (2008b), *Essays on Liquidity-Preference in Markets with Borrowing Restrictions*. Ph.D. Thesis. Katholieke Universiteit Leuven. Available at https://repository.libis.kuleuven.be/dspace/handle/1979/2006

[29] Modigliani, F., (1977), *The Monetary Controversy or, Should We Forsake Stabilization Policies?*, in: *The American Economic Review*, Vol. 67 (2), pp. 1-19.

[30] Modigliani, F. and Miller, M.H., (1958), *The Cost of Capital, Corporation Finance and the Theory of Investment*, in: *The American Economic Review*, Vol. 48 (3), pp. 261-297.

[31] Ross, S.A., (1989), *Institutional Markets, Financial Marketing, and Financial Innovation*, in: *The Journal of Finance*, Vol. 44 (3), pp. 541-556.

[32] Tobin, J., (1947), *Liquidity Preference and Monetary Policy*, in: *The Review of Economic Statistics*, Vol. 29 (2), pp. 124-131.

[33] Tobin, J., (1958), *Liquidity Preference as Behaviour towards Risk*, in: *The Review of Economic Studies*, Vol. 25 (2), pp. 65-86.

[34] Tobin, J., (1972), *Friedman's Theoretical Framework*, in: *The Journal of Political Economy*, Vol. 80 (5), pp. 852-863.

[35] Tobin, J., (1981), *The Monetarist Counter-Revolution Today — an Appraisal*, in: *The Economic Journal*, Vol. 91 (361), pp. 29-42.

[36] Tobin, J., (1982a), *Money and Finance in the Macroeconomic Process*, in: *Journal of Money, Credit and Banking*, Vol. 14 (2), pp. 171-204.

[37] Tobin, J., (1982b), *The Commercial Banking Firm: a Simple Model*, in: *The Scandinavian Journal of Economics*, Vol. 84 (4), pp. 495-530.

[38] Tobin, J., (1993), B*usiness Cycles and Economic Growth: Current Controversies about Theory and Policy*, in: *Bulletin of the American Academy of Arts and Sciences*, Vol. 47 (3), pp. 33-47.

[39] Venter, G., (1991), *Premium Calculation Implications of Reinsurance without Arbitrage*, in: *Astin Bulletin*, Vol. 21 (2), pp. 223-230.

[40] Wang, S., Young, V. and Panjer, H., (1997), *Axiomatic Characterization of Insurance Prices*, in: *Insurance: Mathematics and Economics*, Vol. 21 (2), pp. 173-183.

ALGORITHM FOR PAYOFF CALCULATION FOR OPTION TRADING STRATEGIES USING VECTOR TERMINOLOGY

Sinha **PANKAJ**
University of Delhi, **Delhi, India**
pankaj-sinha@fms.edu
Johar **ARCHIT**
Institute of Technology, **New Delhi, India**
Department of Computer Engineering Netaji Subhas
architjohar@gmail.com

Abstract:
The aim of this paper is to develop an algorithm for calculating and plotting payoff of option strategies for a portfolio of path independent vanilla and exotic options. A general algorithm for calculating the vector matrix for any arbitrary combination strategy is also developed for some of the commonly option trading strategies.

Keywords: option trading strategy, payoff, vector, vanilla and exotic option.

JEL-Classification: C63, C88, C02, G00

1. Introduction

Hull (2009) discusses the payoffs for long and short positions in Call and Put options by using algebraic techniques. J.S. Chaput and L.H. Ederington (2003), Natenberg (1994) and Hull (2009) contain the bibliographies and survey of literature on the theoretical background of option strategies for the path independent vanilla and exotic options such as European, Bermuda, Forward Start, Digital/Binary and Quanto options. There are various open source option strategy calculators like "Option" [4] that only rely on algebraic, analytical and graph superposition techniques to plot graphs for overall profit/loss. We in this paper develop an algorithm using vector terminology to plot the final profit/loss graph of various option strategies.

2.1. Option strategies using vector notation

For a spot price S_T at time T and a strike price K, the payoff for a long position in call option is given by Max(S_T-K,0) and the payoff is Min(S_T-K,0) for the short position in the call option. Similarly the payoff for a long position in put is Max (K-S_T, 0) whereas it is Min (S_T-K, 0) for a short position in the put option. We can represent a vector payoff matrix for any option strategy as a 2xN matrix.

Vector	V_1	V_2	….	V_n
Strike Price	K_1	K_2	…..	K_n

In the above matrix the strike prices $K_1, K_2, \ldots K_n$ for combination of options are in the ascending order, i.e., $K_1<K_2<\ldots<K_n$. The vector V_i can be interpreted as slope of the payoff graph of option strategy. By default the smallest strike price is always taken to be zero i.e. $K_1=0$.The vector is always an integer in the interval $(-\infty, \infty)$ .We can interpret the above matrix in terms of slope of the profit/loss curve obtained for option strategies.

$$slope = \begin{cases} V_i \text{ , for } K_i < K < K_{i+1} \text{ and } i < n \\ V_i \text{, for } K > K_i \text{ and } i = n \end{cases}$$

Vector matrix for long and short position is given by

Long Position

V_1	V_2	….	V_n
K_1	K_2	…..	K_n

Short Position

$-V_1$	$-V_2$	….	$-V_n$
K_1	K_2	…..	K_n

Using the above vector notation we can represent long and short position in call option as under:

Long call

0	+1
0	K_1

Short Call

0	-1
0	K_1

For long position in call, profit/loss curve has two slopes 0 and +1 whereas for a short position the slope of profit/loss curve has two slopes 0 and –1.

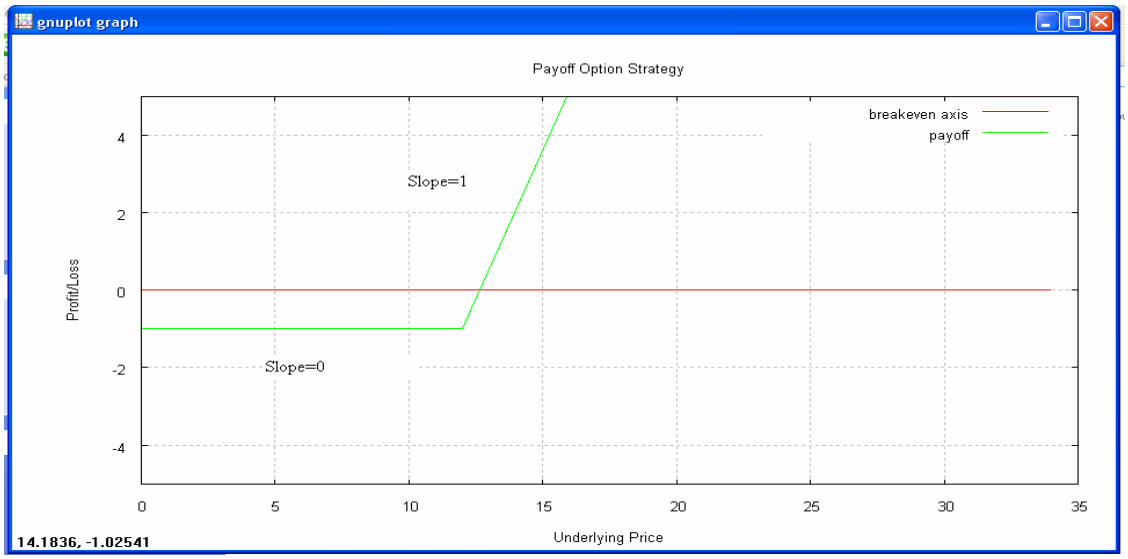


Figure 1. Long Position in Call Option

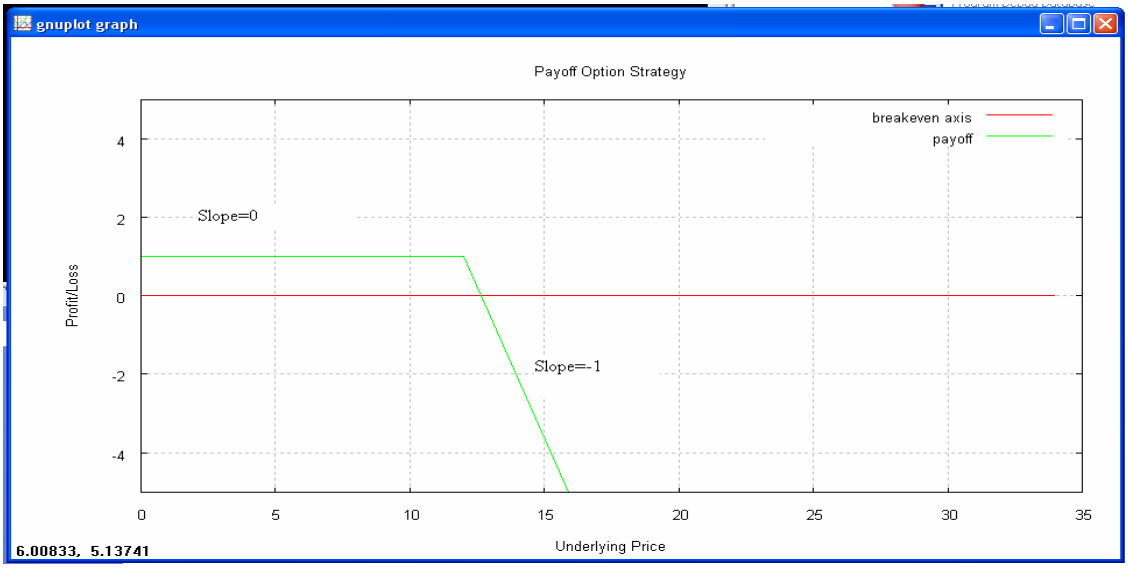


Figure 2. Short position in Call Option

Similarly, the vector matrix for long and short position in put options are:

Long Put

-1	0
0	K_1

Short Put

+1	0
0	K_1

For long position in stock, the slope of profit/loss curve is +1 and strike price is assumed to be zero whereas for short position in stock, the slope of profit/loss curve is -1 and strike price is assumed to be zero. The vector matrix notation is given as:

Long Stock

+1
0

Short Stock

-1
0

When we trade in n units of options using a particular option strategy, the entire vector row is multiplied by n.

$n*V_1$	$n*V_2$	….	$n*V_n$
K_1	K_2	…..	K_n

The data set for a portfolio using n option strategies can be represented as

Strategy 1

V_{11}	V_{12}
K_{11}	K_{12}

Strategy 2

V_{21}	V_{22}
K_{21}	K_{22}

…
…
…

Strategy i

V_{i1}	V_{i2}	….	V_{ij}	….
K_{i1}	K_{i2}	…..	K_{ij}	….

….
….
….

Strategy n

V_{n1}	V_{n2}	…	V_{nm}
K_{n1}	K_{n2}	….	K_{nm}

Note that the number of columns in each option strategy can be different. We can use the above-derived vector matrices to form profit/loss function for any combination of option strategies using the following algorithm:

Algorithm

To plot the overall payoff strategy we need the initial Y intercept of the strategy apart from the resultant vector matrix. This Y intercept can be calculated using matrices of length greater than one using the formula

```
Yint = ∑ ( -1*Vector(A[j])*Strike_price(A[j+1]) )
Yint = Yint + Net_Premium_Paid
```

```
Step 1
For I ← 1 to no_of_options
      For j ← 1 to length_of_option_matrix
Insert A[j] in Result_matrix in sorted increasing order on the basis of
Strike_price(A[j]).

Step 2
For k ← 1 to length_of_Result_matrix
      Vector(B[k])=0
      For I ← 1 to no_of_options
            For j ← 1 to length_of_option_matrix
                  If Strike_price(B[k]) = Strike_price(A[j])
                        Vector(B[k]) = Vector(B[k])+ Vector(A[j])
                  ElseIf j < length_of_option_matrix
If Strike_price(A[j]) < Strike_price(B[k]) < Strike_price(A[j+1])
                        Vector(B[k]) = Vector(B[k])+ Vector(A[j])
                  Else
                        Vector(B[k]) = Vector(B[k])+ Vector(A[j])

Step 3
For I ← 1 to no_of_options
      j=1
      If length_of_option_matrix > 1
Yint = Yint + -1 * Vector(A[j]) * Strike_price(A[j+1])
Yint = Yint + NetPremium

Step 4
For k ← 1 to length_of_Result_matrix – 1
Plot line with slope Vector(B[k]) and Y Intercept Yint
between points Strike_price(B[k]) and Strike_price(B[k+1])
ypoint=Vector(B[k])*( Strike_price(B[k+1]) – Strike_price(B[k]) ) +
Yint
Yint = ypoint -  Vector(B[k+1])* Strike_price(B[k+1]
k = length_of_Result_matrix
Plot line with slope Vector(B[k]) between points Strike_price(B[k]) and
infinity
```

The source code for the above algorithm is written and implemented on VC++.Net 2005 using open source graph plotting utility Gnuplot.

Illustration 1: An investor buys $3 put with strike price $35 and sells for $1 a put with a strike price of $30. (Example 10.2, page 224 given in Hull [1])

The above data can be represented as

Buy Put		+	Sell Put		=	Payoff(Bear Spread)		
-1	0		+1	0		0	-1	0
0	35		0	30		0	30	35

Initial Y intercept is -1*(-1*35) + -1*(1*30) – 3 + 1 = 35 -30 -3 + 1 = 3

One can use the following form to input the data of his/her option strategy:

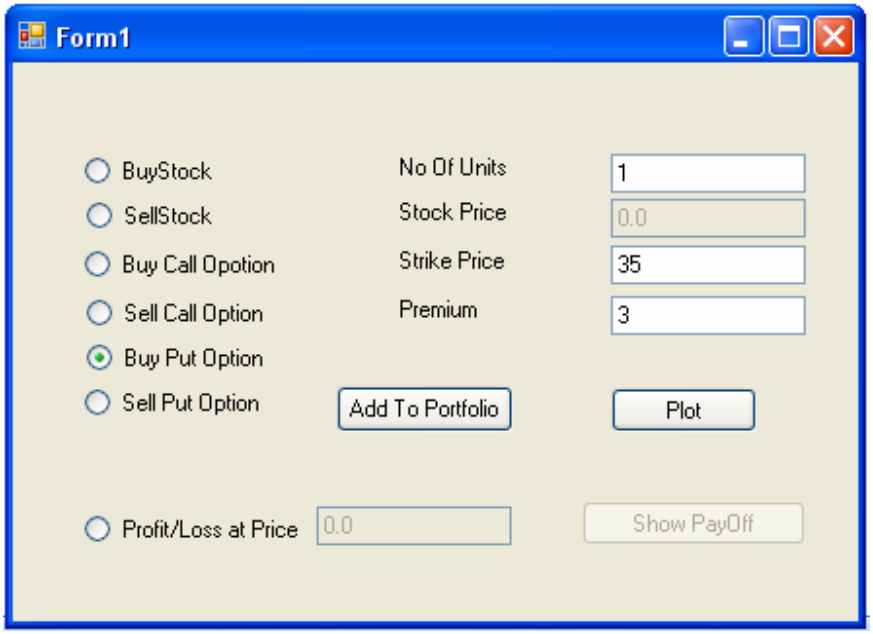


Figure 3. Input Screen

The following is the output of the final payoff of combination of option strategy in vector notation as discussed above.

Figure 4. Vector Payoff Matrix

The algorithm gives the following resultant profit/loss graph of the above combination of option strategies in the form of a bear put spread.

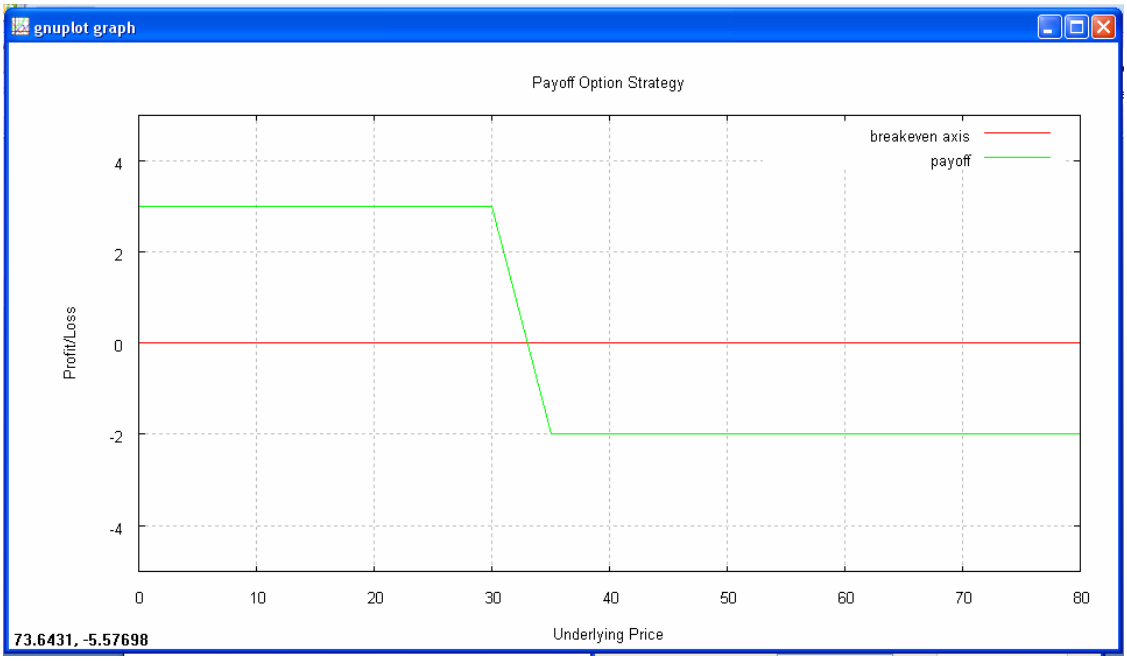


Figure 5. Payoff Graph

The loss is $2 if stock price is above $35 and the profit is $3 if stock price below $30.

2.2. Some More Complex Strategies

The following are the vector matrices for some of the commonly traded strategies:

Long Combo

($0<K_1<K_2$)

Sell Put		+	Buy Call		=	Long Combo		
+1	0		0	+1		+1	0	+1
0	K_1		0	K_2		0	K_1	K_2

Long Straddle

Buy Put		+	Buy Call		=	Long Straddle	
-1	0		0	+1		-1	+1
0	K_1		0	K_1		0	K_1

Short Straddle

The vector matrix of short straddle is negative of that of long straddle

+1	-1
0	K_1

Strip

Buy call		+	Buy 2 puts		=	Strip	
0	+1		-2	0		-2	+1
0	K_1		0	K_1		0	K_1

Strap

Buy 2 calls + Buy put = Strap

0	+2
0	K_1

+

-1	0
0	K_1

=

-1	+2
0	K_1

Long Strangle

($0<K_1<K_2$)

Buy put + Buy call = Long Strangle

-1	0
0	K_1

+

0	+1
0	K_2

=

-1	0	+1
0	K_1	K_2

Short Strangle

The vector matrix of short strangle is negative of that of short strangle. ($0<K_1<K_2$)

+1	0	-1
0	K_1	K_2

Collar

($0<K_1<K_2$)

Long Stock + Buy Put + Sell call =

+1
0

+

-1	0
0	K_1

+

0	-1
0	K_2

Collar

0	+1	0
0	K_1	K_2

Box Spread

($0<K_1<K_2$)

Buy Call + Sell call + Sell Put + Buy Put =

0	+1
0	K_1

+

0	-1
0	K_2

+

+1	0
0	K_1

+

-1	0
0	K_2

Box Spread

0	0	0
0	K_1	K_2

Long Call Butterfly

($0<K_1<K_2<K_3$)

Buy Call + Sell 2 call + Buy Call =

0	+1
0	K_1

+

0	-2
0	K_2

+

0	+1
0	K_3

Long Call Butterfly

0	+1	-1	0
0	K_1	K_2	K_3

Short Call Butterfly

The vector matrix of short call butterfly is negative of that of long call butterfly ($0<K_1<K_2<K_3$)

0	-1	+1	0
0	K_1	K_2	K_3

Long Call Condor

($0<K_1<K_2<K_3<K_4$)

Buy Call +

0	+1
0	K_1

Sell call +

0	-1
0	K_2

Sell Call +

0	-1
0	K_3

Buy Call =

0	+1
0	K_4

Long Call Condor

0	+1	0	-1	0
0	K_1	K_2	K_3	K_4

Short Call Condor

The vector matrix of short call condor is negative of that of long call condor ($0<K_1<K_2<K_3<K_4$)

0	-1	0	+1	0
0	K_1	K_2	K_3	K_4

Illustration 2: Let a certain stock be selling at $77. An investor feels that significant change in price is un-likely in the next 3 months. He observes market price of 3 month calls as

Strike Price($)	Call Price($)
75	12
80	8
85	5

The investor decided to go long in two calls each with strike price $75 and $85 and writes two calls with strike price $80. Payoff for different levels of stock prices is given as:

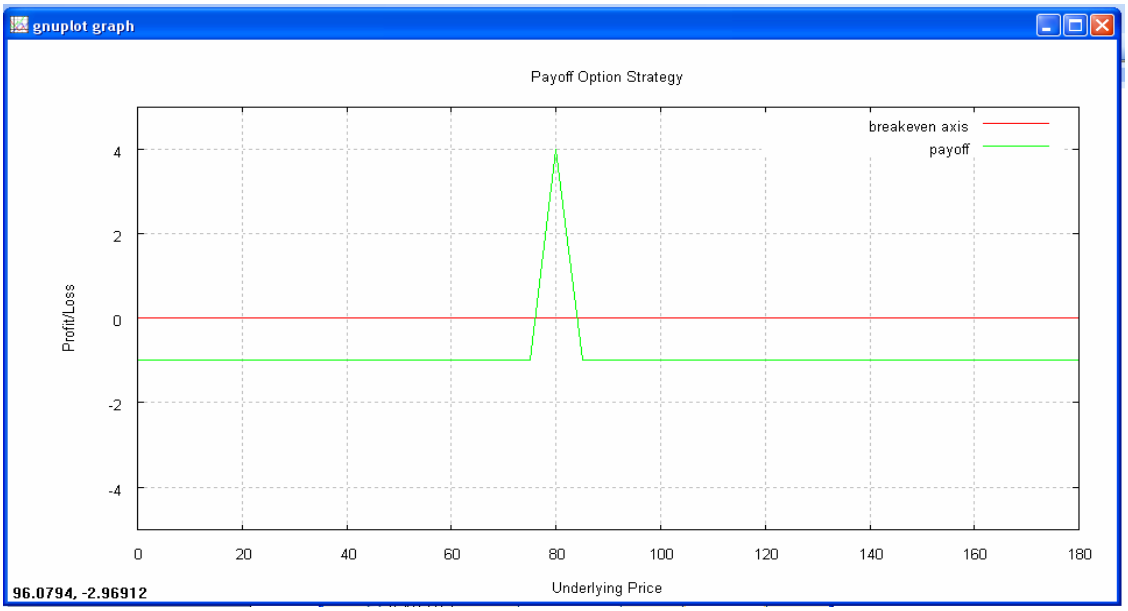


Figure 6. Payoff Graph

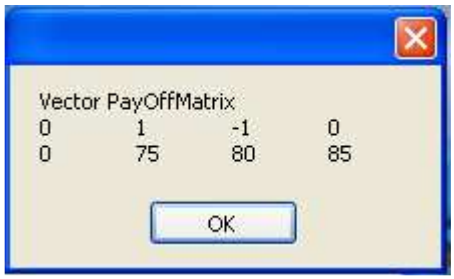


Figure 7. Vector Payoff Matrix

The profit /loss when stock price is at maturity is

Stock Price($)	Profit/Loss($)
65	-1
68	-1
73	-1
78	2
83	1

3. References

[1] Hull, J.C., (2009), *Options, Futures, and Other Derivatives*, Prentice Hall

[2] Natenberg, S., (1994), *Option Volatility and Pricing Strategies: Advanced Trading Techniques for Professionals* McGraw-Hill Professional Publishing

[3] Chaput, J. S. and Ederington L. H., (2003) *Option Spread and Combination Trading*, in: *Journal of Derivatives*, 10, 4, pp 70-88.

[4] http://sourceforge.net/projects/option

Journal of Applied Economic Sciences

SOME REMARKS ON THE EFFECTS OF ACTIVE LABOUR MARKET POLICIES IN POST-TRANSITION

Joanna **TYROWICZ**
jtyrowicz@wne.uw.edu.pl
University of Warsaw, National Bank of Poland, **Poland**
Piotr **WÓJCIK**
University of Warsaw, **Poland**
pwojcik@wne.uw.edu.pl

Abstract:
Social cohesion is the principal goal behind active labour market policies (ALMPs), including those financed at supra-national level, like the European Social Fund. In this paper we use NUTS4 level data on the local labour market dynamics in an attempt to verify direct and indirect effects of ALMPs. We use data for 2000-2007 for Poland, while this period comprised both stark increases and reductions in the unemployment rates. Over this time also the financing of ALMPs has been increased considerably, transforming both to higher intensity of ALMPs (wider coverage) and higher extensiveness of these activities (increase in per treatment cost). At the same time, these trends have transmitted into local context with highly heterogeneous composition of instruments used and actual coverage rates and costs. We implement seemingly unrelated regressions (SURE) approach to inquire the effects of ALMPs on inflows and outflows rates among Polish local labour markets.

Key words: unemployment dynamics, unemployment convergence, SURE, transition

JEL Codes: J64, E24, P21

1. Introduction

When evaluating the effectiveness of differentiated active labour market policies (ALMPs), one typically resorts to two approaches. Basing on individual data allows the estimation of the treatment effect for differentiated instruments, taking into account the developments in a control group. The literature in this field is vast, including, among others see: Sciulli (2005) as well as Destefanis and Fonseca (2007) for Italy, for the US or for Germany. This approach requires not only relatively detailed micro-level data, but also observing individuals after the completion of activisation programmes, which most transition countries lack in general.

The latter approach focuses on regional data instead. The obvious shortcoming is that either quite strong assumptions need to be made concerning the distribution of unemployed among regions (essentially imposing homogeneity during the estimation procedures), or one needs relatively large datasets and considerable heterogeneity to sustain underpinnings for policy implications of the findings[1]. On the other hand, an extensive theoretical framework for the effects of ALMPs on employment has been developed by Calmfors (1994), and recently in a stochastic framework by Lechner and Vazquez-Alvarez (2006)[2]. As stated by Hagen (2003), raising the efficiency of matching process is usually regarded as the main aim of ALMPs, and can be reached by adjusting the human capital of job seekers to the requirements of the labour market and by increasing the search intensity (as well as search capacity) of (former) programmes participants. These aims are especially pronounced in transition countries with large structural mismatches.

In this paper we use NUTS4 level data on the local labour market dynamics in an attempt to verify direct and indirect effects of ALMPs. We use data for 2000-2007 for Poland, while this period comprised both stark increases and reductions in the unemployment rates. Over this time also the financing of ALMPs has been increased considerably, transforming both to higher intensity of ALMPs (wider coverage) and higher extensiveness of these activities (increase in per treatment cost). At the same time, these trends have transmitted into local context with highly heterogeneous composition of instruments used and actual coverage rates and costs.

[1] Furthermore, regional data rarely allow measuring the scale of some negative indirect effects like displacement, substitution or redistribution effects.
[2] See also Calmfors and Skedinger (1995) as well as Calmfors et al (2002), while the empirical applications were extensively evaluated by Petrongolo and Pissarides (2001).

Under the conditions of stark unemployment rate disparities – as is the case of Poland – convergence necessitates that higher unemployment rate regions need to be characterised by relatively higher ratio of outflow and inflow rates. The algorithm allocating ALMPs financing across regions favours more troubled local labour market, giving a premium to higher than average unemployment rate, number of unemployed and worse than average structure of unemployed (e.g. share of long-term unemployed). If the ALMPs were efficient, we should observe positive impact on outflow ratios, while for the convergence it would be necessary that inflow/outflow ratios improve in more deprived regions. We have therefore two empirical aims: (i) verify whether the unemployment dynamics in more troubled regions permit catching up and (ii) asses whether the ALMPs actually contribute to alleviating the local labour market difficulties. To this end, we implement seemingly unrelated regressions (SURE) approach to inquire the effects of ALMPs on inflows and outflows rates among Polish local labour markets and compare the outflow/inflow ratio with a benchmark constructed as a counterfactual in these regressions.

The paper is organised as follows. In the next section we briefly review the literature concerning ALMPs as well as transition. In section 3 the situation in Poland and data are presented, while Section 4 discusses the method and the results. In the concluding section we derive policy implications of the findings.

2. Literature review

Unemployment dynamics at local level has received a lot of attention from the academia. Buettner (2007) compares empirical evidence on regional labour market flexibility in Europe (but uses different aggregation levels for different countries, which makes the results weaker). Marelli (2004) as well as Huber (2007) provide an overview of similarities and differences across European Union regions. In particular, it seems that CEE countries exhibit higher regional wage flexibility, Buettner (2007). At the same time, despite phenomenal migrations emerging after 2004, labour mobility is still assessed to be low (Kaczmarczyk and Tyrowicz 2008), while Fihel (2004) demonstrates that effectively in the local scale unemployment is not significant as pushing factor[3]. In the case of CEECs, the role of transition processes may indeed still be signifficant, Svejnar (2002a), while growing average job tenure as well as average time spent in unemployment or inactivity, Svejnar (2002b) were characteristic.

On the other hand, transition commenced in Poland in 1989, while after a decade another wave of massive unemployment sprung. While it is possible that some enterprises might have avoided the pains of restructuring in the early 1990s and were inevitably following these steps in the second part of this decade, typical market economy processes were already at play. These were indicated by educational boom (Poland has second highest tertiary education enrolment rates, after South Korea) as well as vanishing premium to being employed in a private sector, as argued by Saczuk and Tyrowicz (2009).

The effects of ALMPs in a transition context have been analysed already in mid 1990s, albeit with scarce data: including Boeri (1994), Lehman (1995), Burda (1996), Góra, Lehmann, Socha, and Sztanderska (1996), Kwiatkowski and Tokarski (1997) and Puhani (1999) as well as summary by Dar and Tznatatos (1999) as well as Martin (2000) and Martin and Grubb (2001)[4]. Typically, unlike micro-level studies, the findings were rather discouraging in terms of value for money or sometimes even lack of visible ALMPs effects. Frequent defence argument bases on the fact that some ALMPs effects take longer to appear or may not be discounted in the period of labour market contraction but will eventually boost employment with the change of business outlooks.

Vodopivec, Wörgötter and Raju (2003) review also the effects of the passive component of the labour market policies, finding some expected negative spillovers and interrelations between active and passive labour market policies. While a new wave of research sprung recently, incorporating Balkan and CIS countries with the availability of World Bank labour market surveys, the findings of

[3] All these issues have been surveyed, among others, by Huber (2007)
[4] The principal studies are reviewed in Munich, Svejnar and Terrel (2000)

the post-transition period are only slightly more discouraging. e.g. Vroman (2002), Godfray (2003), Betcherman, Olivas and Dar (2004), Hujer, Thomsen and Zeiss (2006), Fares and Tiongson (2007)[5].

In the empirical literature of unemployment rate characteristics, one can find a number of differentiated approaches towards the unemployment rate dynamics and persistence as well as distribution, according Decressin and Fatas (1995), Obstfeld and Peri (1998) or more recently, Armstrong and Taylor (2000). Perugini, Polinori and Signorelli (2005) use NUTS2 level data and inquire the regional differentiation of Poland and Italy. Marelli (2004) focuses on specialisation for NUTS2 EU regions with tripartite desegregation (industrial, agricultural and service sectors), but analyses predominantly income and economic convergence and not explicitly the underlying fundamentals[6]. In principle, however, convergence is necessitating relatively more favourable inflow/outflow ratios in relatively more troubled regions, which constitutes the main angle of this study. Theoretically, this approach builds on a model developed by Lechner and Vazquez-Alvarez (2005), which introduces exogenous – and potentially asymmetric – stochastic shock at the labour market into individual choices of effort and activity in the environment were skill improving costless training is available.

3. Data and the context of Poland

Transition from a centrally planned to a market economy typically involves massive layoffs and economic slowdown inhibiting vivid job creation, Grotkowska (2006). The situation in the early 1990s in Poland was no different, with unemployment rate increasing to the thresholds of above 10% in just two years and since then never fell below, Figure 1. There are some fundamental characteristics of the Polish labour market. Firstly, high unemployment is believed to have a structural character. Almost 70% of Polish unemployed have primary or vocational education only, frequently outdated professions or no longer applicable skills. Moreover, some surveys suggest that even roughly 50% of these individuals are reluctant to upgrade their qualifications, Tyrowicz (2006).

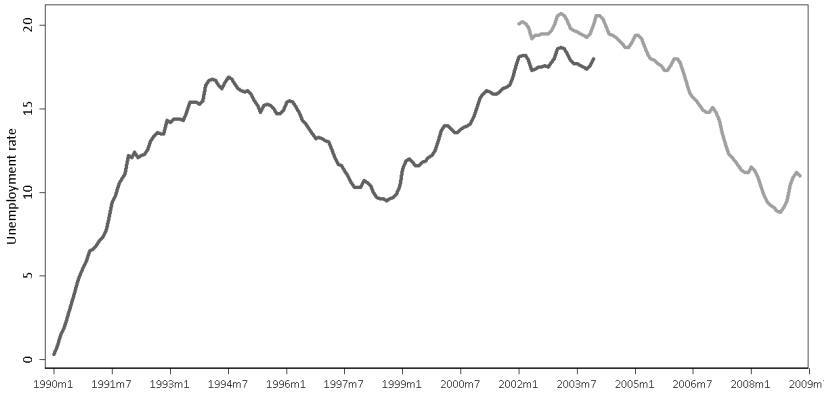


Figure 6. Unemployment rate evolution in Poland, 1990-2009[7]

Professional and geographic mobility is very low, while transitory migration of approximately 1 million Poles to Ireland, Sweden, Norway, Germany, UK and other EU countries concerns

[5] In addition, subsequent to the Hartz reforms in Germany there has been many studies, also at regional level, exploring the effects of these policies, including among others Hujer, Caliendo and Thomsen (2004), Csillag, Schneider, Uhlendorff and Zhao (2006) and many others.

[6] Overman and Puga (2002) perform conditional kernel density analyses of European unemployment rates taking into account the distributions of underlying fundamentals (eg. the skills, the regional specialisation as well as the growth rates of population and the labour force). Also KDE but in a different framework is applied by Tyrowicz and Wojcik (2007) for Poland and Tyrowicz and Wójcik (2009) for Czech Republic and Slovakia.

[7] In 2003 new national census data were made available, which revealed lower size of population and labour force, thus leading to updating upwards the unemployment rates by roughly 3 percentage points. Unfortunately, these were recalculated backwards only for 2002-2003 at national level and 2003 at local level. For subsequent computations, we have filtered the post Jan2003 data to avoid difficulties in statistical interpretation of the findings. Moreover, most of the variables used in the study are referenced to the number of unemployed at a local labour market rather than the unemployment rate. Consequently, the potential contamination of the dataset seems to have limited impact on the quality of the findings.

predominantly those aged under 30 years old (80%) and with relatively high skills (17% with a university degree). Long term unemployment rate is the highest in EU (currently app. 10%, EU average falls short of 4%). Employment in agriculture still exceeds 17%, which is extremely high by European standards, while half of the registered unemployed live in the rural areas. In addition, forecasts concerning the agricultural sector at large suggest, that due to increasing productivity, hidden over-employment in this sector will soon transform to *de facto* unemployment and/or premature labour market exits. Thus, low skills, low mobility and excessive employment in the agriculture are the main structural traits of the labour market problems in Poland.

Finally, this is not high unemployment that creates the main labour market challenge, but low activity and employment rates (currently at 56%, the lowest in EU25). Not only does this phenomenon hinder the economic growth processes, but also social security imbalances are reinforced (low number of working in comparison to social transfer recipients). Currently average exit age falls short of 58 years (with legal ages of 60 for women and 65 for men), while employment rate for 55+ age group amounts to only 28% (55 till retirement *de iure*). High unemployment rate among young workers (34% for workers fewer than 24 years) and highest gender gap in the 25-29 age groups suggest that entering the labour market – and commencing an adult life – is particularly difficult.

All these data show, that labour market policies oriented on stimulating employment rate should focus on activisation of youth (both male and female), female returnees (especially with none or little professional experience) and retaining 50+ and 55+ age groups. These is reflected as of 2004 in legislation, which specifies these groups as more vulnerable and requiring support from public employment services. Naturally, to assure efficiency, these policies should evolve in different directions depending on additional conditions, e.g. targeted group living in the rural areas, etc., which is not explicitly imposed by legislation.

On the contrary to the low employment and high unemployment rates, unregistered employment is of significant proportions, especially in the case of seasonal workers and supplementary income. With high tax wedge, low skilled positions are particularly strongly bound to demand unregistered labour. This is an important context for ALMPs efficiency for two main reasons. First, re-training in these domains may in fact be counter-effective, because unemployed after the programmes would be expected to assume positions with remuneration comparable to their pre-training shadow income. Secondly, some of the workers may find it more beneficial to remain at relatively low compensation in the shadow economy than to exhibit considerable effort to improve skills, because their net increase in remuneration might indeed be low. This last effect might be especially pronounced in the periods of labour market tightness. On the other hand, micro-level evidence suggests that on average workers in the shadow economy receive compensation lower than their counterparts with formal employment contracts, Cichocki and Tyrowicz (2009), which points to labour market segmentation and exclusion as important labour market governing rules.

3.1. The means of ALMPs

In Poland, Public Employment Services (PES) is the main actor in defining and implementing ALMPs. They are subordinate to the public administration (at both regional and local level), thus being only marginally responsive to central policies other than general legislation changes. They struggle with underinvestment both in terms of IT and in terms of HR (employment officers and managers constitute only 29% of the PES employees). On average there are 1600 unemployed per one job broker and over 4600 per one job counsellor.

ALMPs are financed from the Employment Fund (and so are financed passive labour market policies), which originates solely from employers contributions, currently at 0.22% of GDP (with the benefits its 0.8% of GDP). Poland is the only country in Europe with no budgetary contribution to ALMPs. At the same time, ALMPs instruments are highly regulated. They comprise subsidised employment (public work schemes – despite their low efficiency, strongly preferred by local authorities – and public service employment), on-the-job training and scholarships for youth, specific and general training (including the ability to navigate on the labour market). Finally, there are also instruments supporting self-employment (micro-enterprise development) and a subsidy for creating new jobs. As of 2004 European Social Fund means are used for employment policy as well, which allowed increasing the ALMPs spending from roughly PLN 0.4bln to PLN 2.2bln over a decade 1997-2006, approximately by 30% annually, while the proportion should continue in 2007 – 2013.

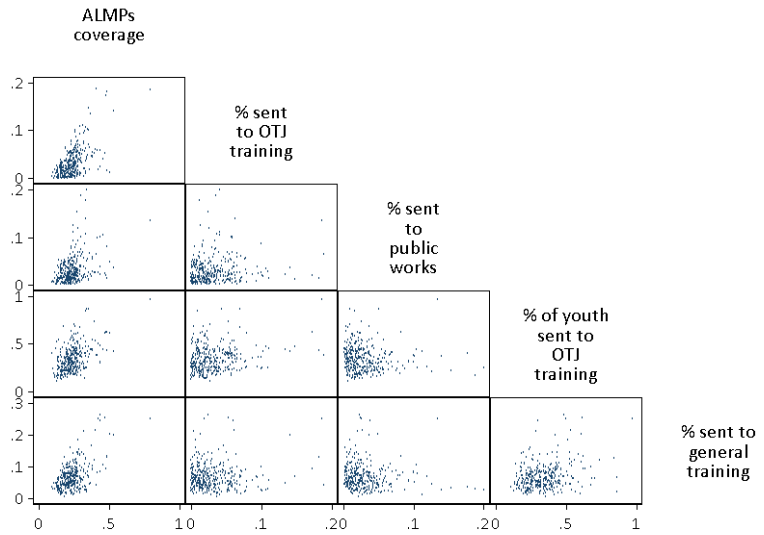


Figure 7. The heterogeneity in the use of instruments by PES in Poland (average over 2006)[8]

Increase in financing has been accompanied by changing the instruments to be used as of 2004[9], albeit in an extremely heterogeneous way. Figure 2 presents the correlation matrix for basic instruments. Visibly, there seems to be large differentiation in the use of instruments as well as overall coverage. Across NUTS4 units (*poviats*) coverage ranges between 5% and 45% of the unemployed population. With particular instruments, there is also a considerable heterogeneity, while it is not necessarily true that in regions with higher coverage rates, all instruments are used more extensively[10]. Importantly, the decisions about the use of instruments and coverage with reference to particular groups are at the discretion of a local labour market office, while the overall budget is the main constraint. NUTS2 authorities, who allocate funding among NUTS2 units sometimes – but not universally across Poland – require commitment to some minimum achievement levels in particular groups, but no targets are set. Thus, it seems that due to the institutional design, endogeneity between the characteristics of the local labour market and ALMPs financing should not be a big issue.

On the other hand, financing of ALMPs is distributed to NUTS2 regions according to an algorithm, which gives a premium to regions with higher than average number unemployment rate, number of unemployed and worse than average structure of unemployed (e.g. share of long-term unemployed). Consequently, the algorithm favours more troubled regional labour market, while NUTS2 level authorities frequently replicate this algorithm when dividing financing to NUTS4 units[11].

High variability of both coverage and instruments used – both the growing trend across time and the differentiation across regions – utters PES independence. On the other hand, unemployment rate evolutions across local labour markets are very heterogeneous too and with some steady and stark disparities. Tyrowicz and Wójcik (2007), using kernel density estimates, have demonstrated that the distributions of unemployment rate are essentially unaffected by general labour market trends, while Tyrowicz and Wójcik (2009) employed stochastic convergence concept and found that the effects of initial shock do not fade out at all in the case of majority of Polish local labour markets (while they are less persistent for Slovakia and definitely fade out for Czech Republic NUTS4 units). This

[8] In the case of on-the-job-training (OTJ training) for the youth, the share is specified by the number of under 25/27 years of age in this form of programes with reference to the number of youth registered as unemployed. The 25/27 years of age boundary is conditional on educational attainment – for university graduates it is 27 years, while for everyone else it is 25 years.

[9] Only as of mid-2004 reporting includes the usage of instruments and particular expenses allocated to different ALMPs. Therefore, in the remainder of the paper, whenever we demonstrate the instruments separately, it represents an average of instruments usage over 2007 for each NUTS4 unit (poviat). In the regressions, general estimates of coverage and expenses are used, without controlling for different instruments, as this would narrow the sample to only three years.

[10] Detailed histograms of the instruments usage are reported in Appendix of this paper.

[11] NUTS2 level labour offices (regional labour offices) do not implement any labour market instruments – either passive or active.

heterogeneity on both ends might indeed be causal (even with potential reverse causality), while the channel to best explore the potential links is through the inflows and outflows rates.

3.2. Inflows and outflows

Also the inflow and the outflow rates demonstrate high variation, both in time and across units, Figure 3. With reference to inflows and outflows from the unemployment pool, little analysis has been done so far at a local level, namely due data shortages. The importance of inflows and outflow rates for the determination of local unemployment rates and their convergence cannot be overstated, while it has also been recognised in the literature. For example, Newell and Pastore (1999) argue that it is the hazard of job loss differentiating for employees with longer tenure that drives the regional differences over the period of 1995-1999. Unfortunately, they use data for the former administrative structure, which forced them to essentially resort to 49 NUTS3 level, which at the time was not a policy relevant level. As of 1999, 380 NUTS4 units (*poviats*) were established, while policy is actually implemented at this level. The main reason why this differentiation is important for the analysis of ALMPs is that NUTS4 units do not exhibit any convergence whatsoever to NUTS2/3 nor to the national average over the 1999-2008 period, Tyrowicz and Wójcik (2009). Therefore, aggregation of data actually blurs the picture and may sometimes lead to misleading results.

Importantly, the ratio of inflows and outflows seems fairly stable across time. It has been observed at aggregate level that inflows/outflows ratio has been fairly constant throughout most of the period. Strawinski (2008) demonstrates even using labour force survey data that flows into unemployment do not reproduce the dynamic patterns observed in the unemployment rate, while the only flow that exhibits these dynamics is the one in the opposite direction. Also, throughout the entire period, inflows have been larger than outflows, while this finding from the aggregate level seems to hold also universally across all NUTS4 units.

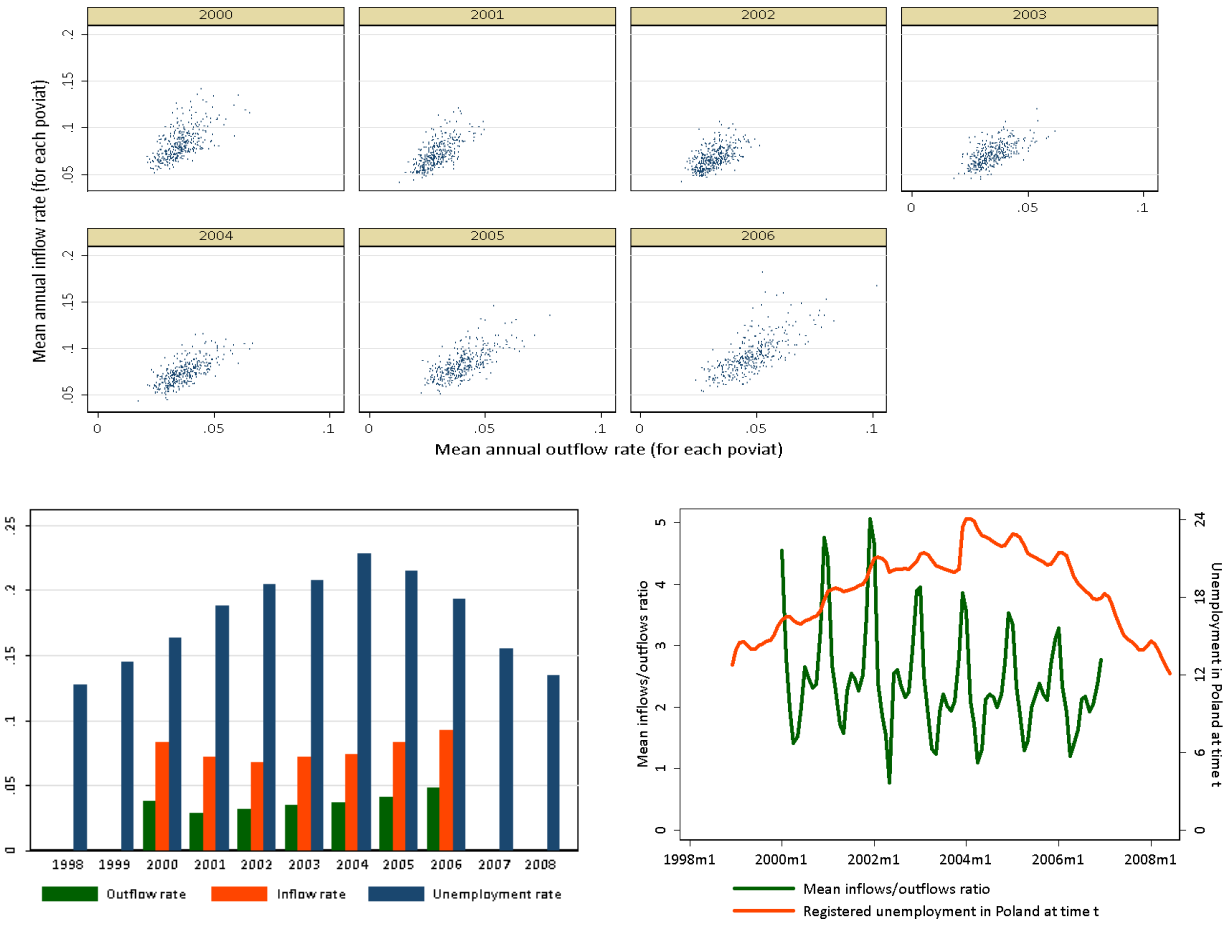


Figure 8. Aggregate (upper) and per *poviat* (lower left) inflow and outflow rates as well as inflow/outflow ratios (lower right)

Importantly, heterogeneity seems to shrink in the periods of worse labour market outlooks and expands over the up swings in the cycle. Notwithstanding, the values of inflows/outflows ratio throughout the entire period have been extremely both volatile and stable. Volatility is visible in extreme seasonality of this variable, ranging between 1 and as much as 5 over 2000-2002. This extreme range lowered slightly with the general improvement in the labour market outlooks, still however reaching even the level of 2.5-3 in the winter months. On the other hand, despite these improvements, lower boundary remained at constant level of slightly above unity. This stability demonstrates that even "in the good times" average ratio implies equalisation of inflows and outflows[12]. In other words, catching up would only be possible if high unemployment *poviats* had lower ratios as opposed to the more those in generally more favourable situation who should be characterised by higher ratios. This prediction will be tested with the use of actual as well as fitted ratio values.

To summarise, we use an extensive dataset for 380 Polish NUTS4 units over the times pan of 2000-2007[13]. Data come from the monthly reports of local labour offices to the Ministry of Labour and Social Affairs and is not available beyond these time boundaries. We matched these reports with the registered unemployment rate statistics provided also monthly by the Central Statistical Office at NUTS4 level (equivalent to the administrative area under the influence of particular labour office)[14]. Based on the raw data from monthly reports we have constructed a number of variables measuring or proxying the processes of interest. Unfortunately, we were unable to use detailed data on the use of particular labour market instruments, because data in this respect is only available for the last two years of the sample, which would limit considerably the scope of this study.

4. Methodology and results

We have two empirical aims: (i) verify whether the unemployment dynamics in more troubled regions permit catching up and (ii) asses whether the ALMPs actually contribute to alleviating the local labour market disparities. We approach these problems by analysing inflows and outflows rates as well as the interplay between the structural conditions, active labour market policies and the inflows/outflows ratio. The empirical strategy may be summarised as follows. We first inquire if – controlling for structural characteristics – heterogeneity of outflows rates is affected by the heterogeneous use of ALMPs across local labour offices in Poland. We measure the ALMPs by coverage (the share of unemployed in any form of treatment), intensity of treatment (the average cost of treatment per one person in treatment) and its extensiveness (the average cost of treatment per one unemployed). All these measures are constructed at NUTS4 level for each month of 2000-2007 time span. Naturally, because these are relatively high frequency data, past ALMPs spendings might affect the current outflows rate. Therefore, we have introduced the last variable proxying for the use of ALMPs, namely the share of spending by this particular NUTS4 unit in the national ALMPs expenses as of January of the particular year up to each consecutive month. To avoid the problems associated with the presence of large and small poviats in the sample, we have scaled this variable by the inverse of the share of unemployed registered in this particular local labour market in national unemployment pool.

Except for estimating the outflows equation, we intend to obtain a counter-factual inflows/outflows ratio, which necessitates the second regression. To this end, we implement seemingly unrelated regressions (SURE) approach. There are two main motivations for use of SURE. The first one is to gain efficiency in estimation by combining information on different equations[15]. SURE is based on the idea of a set of equations, where the disturbances are correlated across equations, in our case: local labour markets. Various methods have been employed to estimate such a set of equations in

[12] In principle, this necessitates that the eventual reductions in the unemployment rates are an effect of higher labor market exits (to inactivity) than entries (from schooling or inactivity).

[13] In fact, some of the NUTS4 units were only established as of January 2001, by separating cities and their surroundings into separate administrative units. Consequently, the number of units increased from roughly 350 to 380. In fact, our dataset contains approximately 420 units, because after the separation two distinct units should be analysed for the differences in unemployed characteristics and labour market opportunities.

[14] This is a unique dataset and is used for the first time, to the best of our knowledge.

[15] Zellner (1962) provided the seminal work in this area, and a thorough treatment is available in the book by Srivastava and Giles (1987). A recent survey can be found in Fiebig (2001).

an attempt to exploit the information in the correlated errors, either contemporaneously or autoregressively, in order to achieve greater efficiency in the estimates. If two estimated equations are linked by the potential correlation in their standard errors, typical OLS estimators will remain unbiased and consistent for each separate equation, but because the approach ignores the correlation of the disturbances the estimates will not be efficient. In this particular case – due to the considerable heterogeneity of analysed units, efficiency of estimators is actually a matter of concern.

In the first stage we estimate the following system of equations

$$INFLOWS_{i,t} = \alpha_i^{IN} + \beta^{IN} * \Sigma_{i,t} + \gamma^{IN} \Pi_{i,t} + \varepsilon_{i,t}$$

$$OUTFLOWS_{i,t} = \alpha_i^{OUT} + \beta^{OUT} * \Sigma_{i,t-1} + \gamma^{OUT} \Pi_{i,t-1} + \varepsilon_{i,t}^{OUT}$$

where $\Sigma_{i,t}$ denotes structural controls, including unemployment rate (both national average and local one), share of youth, females, long-term unemployed and those who live in the rural areas in the pool of unemployed registered in *poviat* i at time t. Similarly, $\Pi_{i,t}$ denotes policy variables, i.e. coverage (share of unemployed in any treatment), intensity (average cost per treated) and extensiveness (average cost per unemployed) of ALMPs in *poviat* i at time t. As has been suggested earlier, there may be also a kind of "cumulative" or long-term effect of ALMPs, which necessitates the inclusion of the share of spending by particular *poviat* i at time t in national spending since the beginning of each year. Including national average in the model plays the role of time fixed effects. At the same time, model comprises unit fixed effects (dummies for each *poviat*). Equations were estimated with the use of iterated SURE, which essentially employs maximum likelihood estimator instead of a standard OLS in each of the equations. Results are reported in Table 1.

Table 1. Estimation results

Models	**ISURE (MLE)**	
Equation	**Outflows**	**Inflows**
Variables	**(1)**	**(2)**
National unemployment at time t	0.000505***	-0.00051***
	(5.59e-05)	(8.18e-05)
Unemployment in *poviat* i at time t	-0.000122***	-0.00025***
	(1.19e-05)	(1.75e-05)
Lagged inflows rate		0.547***
		(0.00560)
Spending per treated	0.000206	
	(0.000555)	
Spending per unemployed	0.00134***	
	(0.000452)	
Share of unemployed in treatment	-0.00102**	
	(0.000456)	
Share of spending in national allocation	-0.133***	
	(0.0348)	
Share of unemployed living in rural areas	-0.00340***	
	(0.000614)	
Share of unemployed with no or low skills	0.117***	
	(0.0106)	
Share of unemployed under 25/27	0.0165***	
	(0.00165)	
Share of unemployed over 50/55+	0.00733***	
	(0.00157)	
Share of long term unemployed	-0.00829***	
	(0.00251)	
Share of females	0.0220***	
	(0.00250)	
Lagged share of females		0.0737***
		(0.00283)
Lagged share of elderly		0.00916***

		(0.00229)
Lagged share of LTU		-0.00745***
		(0.00128)
Lagged share of youth		-0.0190***
		(0.00237)
Lagged share of living in the rural areas		-0.000333
		(0.000925)
Lagged share with no or low skills		0.129***
		(0.0154)
Lagged outflows rate	0.562***	
	(0.00558)	
Constant	-0.00238	0.0134***
	(0.00202)	(0.00249)
Observations	22152	22152
R-squared	0.446	0.497
Breusch-Pagan test (H0: independence)	Rejection	

Source: Own computations based on local PES monthly reports to MLandSA over 2000-2007. ***, ** and * denote significance levels of 1%, 5% and 10% respectively

The performance of inflows equation is better than for the outflows, while both included lagged values (ADL specification) to immune the estimators of the potential hysteresis and persistence. All structural variables prove significant and have expected signs. Rural *poviats* are characterised by lower outflow rates (less dynamic labour markets), while inflow rates do not seem to be affected by the character of the local labour market. Higher share of unemployment among the youth and low-skilled coincides with higher outflow rates due to fact that typically dynamic labour markets exhibit more demand for labour, leaving in general less people without employment opportunities. Consequently, the share of youth and low-skilled among the unemployed needs to be relatively higher. This is consistent with finding a negative estimator of the share of long-term unemployed, since with the growth of this population among unemployed there are less chances of effectively putting PES beneficiaries in employment. Please, recall that outflow rate figures were on average relatively low, approximately 3.7% each month with the minimum of 0.02% and a maximum of 16%.

As to the inflow rates, past structure seems to be a pretty good predictor of the future inflows. Namely, lower shares among youth and LTU – both correlating with better employment opportunities – are associated with lower inflows. Similarly, higher shares of people without skills and those who live in the rural areas seem to be positively correlated with higher future inflows into employment. These phenomena have been addressed by researchers (e.g. Marody and Poleszczuk, 2008) and are believed to be linked by the "inheritance of unemployment" typical for some regions of Poland.

The results for policy variables are not speaking in favour of ALMPs efficiency hypothesis. Namely, intensity of treatment remains consistently insignificant irrespectively of specification. This suggests that high-cost treatments (e.g. trainings) do not seem to result in higher outflows rates. The accumulated effect is significant, but in fact negative – the more financing a particular region receives, the lower the impact on outflows into employment. Extensiveness measures are significant and positive; suggesting that availability of funds actually plays a positive role. At the same time, coverage is not significant – it does not seem to be "any" programme that matters. All in all, these results do not seem to confirm the hypothesis of the overall ALMPs efficiency. However, one could raise many doubts as to the reliability of the policy estimators in this equation. Namely, the accumulated effects may go beyond one year, while poor performance of policy variables may also follow from the fact that units are so heterogeneous in their structures and the use of instruments.

To this end, we have saved the predicted values of inflows and outflows rates and computed a "fitted inflows/outflows ratio" for each *poviat* i at each time t. More specifically, we have included only structural variables in the outflows equation (no $\Pi_{i,t}$ variables), which permits us to obtain a counterfactual outflow rate, had there been no ALMPs implemented. This obtained ratio is in some cases higher and in some cases lower than the actual inflows/outflows ratio. Below, using graphical analysis, we demonstrate whether the over/underperformance of the fitted ratio coincides with the use of ALMPs. Please note, that we do not actually predict the ratio, but separately in the SURE

framework the inflows rate and the outflows rate. The ratio is computed based on these two predictions, which implies it may be contaminated not by one but by two error terms[16]. By subtracting from the actual ratio the "fitted" one, we have constructed a simple measure which shows over/underperformance relative to a counterfactual benchmark. Namely, for the negative values of this "residual", we can state that actual ratio has been lower (i.e. labour market performance better), while the fitted value which is conditioned only on structural and macroeconomic factors is less favourable. Conversely, positive values of residual indicate that actually local labour market performed worse than the model would have predicted. The purpose of this exercise is following: if we can demonstrate that the negative values of the residual are associated with the use of ALMPs, we find indirect positive effect of activisation efforts by PES. In other words, we seek negative coefficients on policy variables in the "residual" regression. This question is approached both graphically and by a robust panel data fixed effect model. More specifically, we estimate the following equation:

$$RESIDUAL_{i,t} = \alpha_i + \beta_i * \Sigma_{i,t} + \gamma_i \Pi_{i,t} + \varepsilon_{i,t}$$

with the notation used before. As control factors in $\Sigma_{i,t}$ we have included national and local unemployment rate and structural characteristics. The model was estimated as OLS with robust standard errors, as panel-corrected heteroscedasticity and autocorrelation consistent standard errors and as GLS. Results are reported in Table 2.

Table 2. Indirect effects of ALMPs on local labour markets performance

Variables	(1)	(2)	(3)
Local unemployment rate	-0.0254***	-0.0141***	-0.0117***
	(0.00571)	(0.00113)	(0.00330)
National unemployment rate	0.0155**	0.0177***	0.238***
	(0.00786)	(0.00541)	(0.0589)
Structural controls	Included and significant	Included and significant	Included and significant
Spending per one unemployed	-0.662***	-0.134***	-0.442
	(0.0522)	(0.0435)	(0.359)
Share of spending in total national ALMPs spending	-56.49***	-19.36***	-11.40
	(12.68)	(3.336)	(9.172)
Share of people in any treatment	0.270***	0.230***	0.524**
	(0.0435)	(0.0439)	(0.230)
Spending per one person in any treatment	0.142**	0.114**	0.267
	(0.0553)	(0.0537)	(0.427)
Observations	22 827	22 827	22 827
R-squared	0.26	n.a.	0.30
X^2 statistic	47.01	661.47	11.07
Method of estimation	FE OLS	FE GLS	FE PCSE

Source: Own computations based on local PES monthly reports to MLandSA over 2000-2007. *Note*: panel-corrected standard error (PCSE) estimates for linear cross-sectional time-series models where the parameters are estimated by Prais-Winsten regression. PCSE estimation allows effectiveness even in the presence of AR(1) autocorrelation within panels and cross-sectional correlation and heteroscedasticity across panels. Robust standard errors reported. Year dummies significant (not reported, available upon request) Structural estimators included but not reported, available upon request, individual effects included but not reported. Constant included but not reported. ***, ** and * denote significance levels of 1%, 5% and 10% respectively

[16] Appendix presents the graphs (by year) of the actual versus fitted ratios. There do not seem to be large discrepancies in the relation between them across time, which implies the model captures large swings in the labour market outlooks we observed over the analysed period.

Results are fairly consistent, although the performance of PCSE estimator is lowest, while the relatively large size of the standard errors makes some of the variables insignificant. Still, the signs and the order of magnitude remain unaffected by the method of estimation. While positive indirect effect of ALMPs on local labour markets performance would require the estimated parameters to be negative, only extensiveness measure (spending per one unemployed) and long-term effect measure (share of spending in total national ALMPs spending) fall short of zero. Both coverage and intensity measures have positive sings. While for the two latter measures graphical analysis provides no further insights, the reliability of the first two is somewhat undermined. Figure 4 scatters the policy measures against the actual values or residuals across time and *poviats*. Namely, the negative coefficients found on extensiveness and accumulated variables seem to both follow from a small group of outliers with very high values of these predictors. Most of the dots are spread flat with heterogeneity understandably increasing in the proximity of 0 values for the residual. Graphical inspection does not seem to reveal any actual difference between the two groups of measures in Table 2.

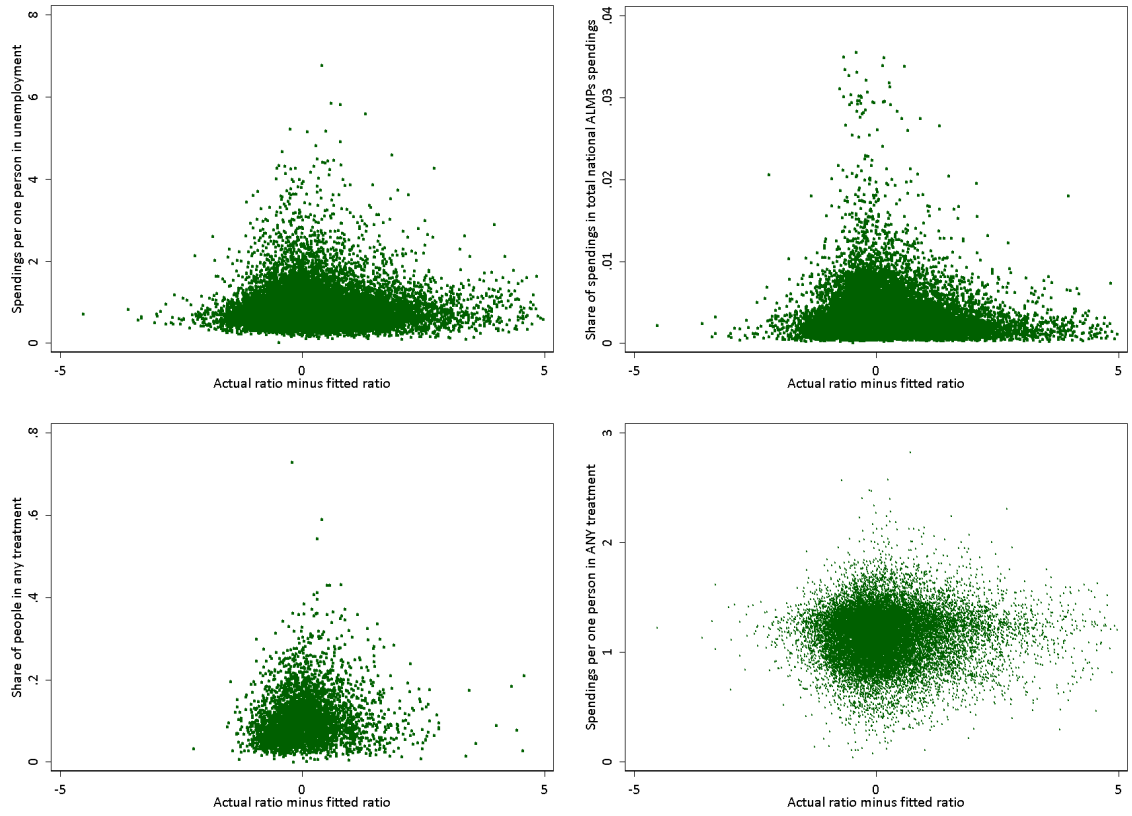


Figure 9. Indirect effects of ALMPs on local labour market performance

Summarising, the literature traditionally assumes that diversification of the use and coverage of cohesion policies provides sufficient variation to derive conclusions regarding the effectiveness of cohesion efforts. Our main empirical purpose here was to inquire the whether convergence may be achieved with the use of ALMPs by inquiring the effect of ALMPs on outflows rate (directly) and inflows/outflows ratio (indirectly). We used policy relevant NUTS4 level data, since actual labour market policies - with special emphasis on the active ones - are performed at exactly this level. Time span in this study allows covering both the up and the down cycles in labour market conditions, which guarantees that the results are not trend driven.

We found that even if statistically one can demonstrate the relevance of ALMPs for unemployment disparities alleviation, these results are not reliable (dependence on outliers) and point largely to the inefficiency of ALMPs implementation at the local level. Neither coverage nor the cost of treatment affect significantly outflows rate. They remain insignificant when the indirect approach is used as well. We confirmed significance (and adequate signs) on the extensiveness and accumulation

measures (average cost per unemployed and share of spending in national ALMPs spending), but these proxy more for the finance availability than for the actual policies. One could derive a conclusion that local PES with more resources tends to over perform the majority, but this conclusion is weak in a sense that the means of over performance (higher coverage or higher cost of treatment) were not confirmed empirically.

5. Conclusions

Transition economies typically experienced rapid growth of the unemployment rates due to profound restructuring. Naturally, these processes affected local labour markets asymmetrically, since regions were diversified with respect to industry composition and economic outlooks. Tyrowicz and Wojcik (2009) demonstrate that diverging unemployment rates' patterns seem nested in the data for transition countries. This paper demonstrates that much of the observed effect currently may be attributed to the lack of ALMPs effectiveness, i.e. not the consequence of transition hardship, but the mistakes made right now.

Financing ALMPs plays an important role in improving the management of ALMPs in general. There is still a lot to be done to assure sensible labour market interventions in order to appropriately respond to the key labour market difficulties and challenges, not allowing any important risk groups to fall out of the horizon. Thus, one has to derive conclusions from the experience of implementing ALMPs so far by approaching the following issues. It seems that one can attribute these findings to the usual suspects found frequently in the literature. Ability to diagnose and forecast in a longer-term perspective taking the view of differentiated groups (and stakeholders!) on the labour market is especially viable in local context, thus providing a challenge for PES, but also to local authorities as well as some other institutions that can affect employment policies. Creating framework for mutual responsibility in labour market policies shared by differentiated stakeholders (local authorities, PES, educational institutions, employers, NGOs, etc) seems key but also difficult. From the other end, there seems to be a need for the programming of the labour market policies with the orientation to efficiency. The urging need for individualisation in projects and labour market services in general is crucial for efficient treatment. In general, ALMPs need to be put into management-by-objectives framework. Finally, diversity of risk groups needs to be reflected in the diversity of tools and instruments – problems and challenges of the Polish labour market are highly differentiated.

All these problems need to be viewed in core-periphery dichotomy with special focus on rural areas. All these conclusions and recommendations point to the direction of making better use of ALMPs financing in employment policies as well as increasing the efficiency of ALMPs.

6. Acknowledgements

Authors would like to thank Joseph Zweimuller and participants of NoEG 2009 for extremely valuable comments. Participants of various seminars at Paris I - Sorbonne, Indiana University and the University of Warsaw as well as NEM 2007 in Tallinn, CAPE 2007 in Nuremberg and XXII AIEL in Naples and MET-EACES in London conferences. We are also indebted to two anonymous referees. The remaining errors are, of course, ours. Part of the work has been performed while Joanna Tyrowicz was a Visiting Researcher at IZA in Bonn and at Columbia University in New York, their support is gratefully acknowledged. Usual disclaimer applies.

7. Bibliography

[1] Armstrong, H. and Taylor, J., (2000), *Regional Economics and Policy*, 3rd edn, Blackwell.

[2] Boeri, T. and Terrell, K, (2002), *Institutional Determinants of Labor Reallocation in Transition,* in: *Journal of Economic Perspectives,* 16(1), pp.51–76.

[3] Breitung, J. and Meyer, W., (1994), T*esting for Unit Roots in Panel Data: Are Wages on Different Bargaining Levels Cointegrated?*, Technical Report 4

[4] Boeri, T., (1994), *Labour Market Flows and the Persistence of Unemployment in Central and Eastern Europe*, in: *OECD* (ed.): *Unemployment in Transition Countries: Transient or Persistent?* pp. 13-56, Paris: OECD.

[5] Boeri, T., (1996), *Unemployment Outflows and the Scope of Labour Market Policies in Central and Eastern Europe*, in: *OECD* (ed.): *Lessons from Labour Market Policies in Transition Countries*, pp. 41-70, Paris: OECD.

[6] Betcherman, G., Olivas, K. and Dar, A., (2004), *Impacts of Active Labor Market Programs: New Evidence from Evaluations with Particular Attention to Developing and Transition Countries*, World Bank Social Protection Discussion Paper Series, Discussion Paper No 0402

[7] Buettner, T., (2007), *Unemployment Disparities and Regional Wage Flexibility: Comparing EU Members and EU-accession Countries*, Empirica 34, pp. 287–297.

[8] Calmfors, L., (1993), *Centralisation of Wage Bargaining and Macroeconomic Performance: A Survey*, OECD Economics Department Working Papers 131, OECD Economics Department.

[9] Calmfors, L., Forslund, A. and Hemström, M., (2002), *Does Active Labour Market Policy Work?* Lessons from the Swedish Experiences, Seminar Papers 700, Stockholm University, Institute for International Economic Studies.

[10] Calmfors, L. and Skedinger, P., (1995), Does Active Labour Market Policy Increase Employment? Theoretical Considerations and Some Empirical Evidence from Sweden, in: *Oxford Review of Economic Policy*, 11(1), pp. 91-109.

[11] Camarero, M., Carrion-i Silvestre, J. L. and Tamarit, C., (2006), *Testing for Hysteresis in Unemployment in OECD Countries: New Evidence Using Stationarity Panel Tests with Breaks*, in: *Oxford Bulletin of Economics and Statistics*, 68(2), pp. 167–182.

[12] Carlino, G. A. and Mills, L. O., (1993), *Are U.S. Regional Incomes Converging? A Time Series Analysis*, in: *Journal of Monetary Economics*, 32(2), pp. 335–346.

[13] Cichocki S. and Tyrowicz, J., (2009), *Shadow Employment In Transition: A Matter of Choice or No Choice*, WNE UW Working Paper No 7(17)

[14] Csillag, M., Schneider, H., Uhlendorff, A and Zhao, Z., (2006), *The Impact of Labor Market Reform on the Effectiveness of Training Programs in Germany*, Institut Zur Zukunft Der Arbeit, IZA Discussion Paper No

[15] Dar, A. and Tzannatos, Z., (1999), *Active Labor Market Programs: A Review of the Evidence from Evaluations*, Social Protection Discussion Paper No.9901.

[16] Decressin, J. and Fatas, A., (1995), *Regional Labour Market Dynamics in Europe*, in: *European Economic Review*, 39(9), pp. 1627-1655

[17] Dixon, R., Shepherd, D. and Thomson, J., (2001), *Regional Unemployment Disparities in Australia*, Regional Studies 35(2), pp. 93–102.

[18] Fiebig, D. G., (2001), Seemingly Unrelated Regression, in Baltagi, B. ed, A Companion to Theoretical Econometrics

[19] Ferragina, A. M. and Pastore, F., (2008), *Mind the Gap: Unemployment in the New EU Regions*, in: *Journal of Economic Surveys* 22(1), pp. 73–113, Backwell Publishers, 0101-121.

[20] Fares, J. and Tiongson E. R., (2008), Youth Unemployment, Labor Market Transitions, and Scarring: Evidence from Bosnia and Herzegovina, 2001-2004, *mimeo*, World Bank

[21] Flek, V., Galuscak, K., Gottvald, J., Hurnik, J., Jurajda, S. and Navrati, D., (2004), *Anatomy of the Czech Labour Market: From Over-Employment to Under-Employment in Ten Years?* Working Papers 2004/07, Czech National Bank, Research Department.

[22] Godfray, M., (2003), *Youth Employment Policy in Developing and Transition Countries - Prevention as well as Cure*, World Bank Social Protection Discussion Paper Series, DP No 0320

[23] Gomes, F. A. R. and da Silva, C. G., (2006), *Hysteresis vs. Nairu and Convergence vs. Divergence: The Behavior of Regional Unemployment Rates in Brazil*, Technical report.

[24] Góra, M. and Lehman, H., (1995), *How Divergent is Regional Labour Market Adjustment in Poland?*, in S. Scarpetta and A. Wörgötter (eds), The Regional Dimension of Unemployment in Transition Countries. A Challenge for Labour Market and Social Policies, OECD

[25] Góra, M., Lehmann, H., Socha, M. and Sztanderska U., (1996), *Labour Market Policies in Poland*, in: OECD (ed.): Lessons from Labour Market Policies in Transition Countries, pp. 151-176, Paris: OECD.

[26] Gray, D., (2004), *Persistent Regional Unemployment Differentials Revisited*, Regional Studies 38(2), pp. 167–176.

[27] Grotkowska, G., (2006), *The Case of Poland - Recent Changes of Non-standard Employment and Labour Market Flexibility*, in C. Kohler, K. Junge, T. Schroder and O. Struck (eds), Trends in Employment Stability and Labour Market Segmentation, SFB 580 Mitteilungen, Universitat Jena.

[28] Grubb, W. and Ryan, P., (2000), *The Roles of Evaluation for Vocational Education and Training*. ILO, Geneva.

[29] Hagen, T., (2003), *Three Approaches to the Evaluation of Active Labour Market Policy in East Germany Using Regional Data*, Technical report.

[30] Huber, P., (2007), *Regional Labour Market Developments in Transition: A Survey of the Empirical Literature*, in: *European Journal of Comparative Economics* 4(2), pp. 263 – 298.

[31] Hujer, R., Caliendo, M. and Thomsen, S., (2004), *New evidence on the effects of job creation schemes in Germany – a matching approach with threefold heterogeneity*, Research in Economics: An International Review of Economics 58, pp. 257-302.

[32] Hujer, R., Thomsen, S.L. and Zeiss, Ch., (2006), *The Effects of Short-Term Training Measures On the Individual Unemployment Duration in West Germany*, *mimeo*, ZEW Manheim

[33] Im, K. S., Pesaran, M. H. and Shin, Y., (2003), *Testing for Unit Roots in Heterogeneous Panels*, in: *Journal of Econometrics*, 115(1), pp. 53–74.

[34] Kaczmarczyk, P. and Tyrowicz, J., (2008), *Współczesne procesy migracyjne w Polsce (Contemporaneous Migration Processes from Poland)*, Fundacja Inicjatyw Społeczno-Ekonomicznych, Warszawa.

[35] Kwiatkowski, E. and Tokarski, T., (1997), *Active Labour Market Policies in Poland*, An Augmented Matching Function Approach, Institute of Economics, University of Lodz, mimeo.

[36] Kwiatkowski, E. and Kubiak, P., (1998), *On Some Determinants of Regional Unemployment in Poland in Transition*, Institute of Economics, University of Lodz, mimeo.

[37] Lechner, M., Vazquez-Alvarez, R., (2005), *Stochastic labour market shocks and Active labour market policies: a theoretical and empirical analysis*, *mimeo*

[38] Lehmann, H., (1995), *Active Labor Market Policies in the OECD and in Selected Transition Economies*, World Bank Policy Research Working Paper No.1502, Washington D.C.

[39] Lehmann, H. and Walsh, P., (1998), *Gradual Restructuring and Structural Unemployment in Poland: A Legacy of Central Planning. mimeo*, LICOS, Centre for Transition Economies, Katholieke Universiteit Leuven, Leuven.

[40] Levin, A., Lin, C.-F. and Chu, C.-S., (2002), *Unit Root Tests in Panel Data: Asymptotic and Finite-sample Properties*, in: *Journal of Econometrics* 108(1), pp.1–24.

[41] Maddala, G. S. and Wu, S., (1999), *A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test*, in: *Oxford Bulletin of Economics and Statistics* 61(0), pp.631–52

[42] Marelli, E., (2004), *Evolution of employment structures and regional specialisation in the EU*, in: *Economic Systems* 28(1), pp. 35-59.

[43] Martin, E., (2006), *The Impact Upon Employment of the Decline in Danube Transport*, in: *South-East Europe Review* 1(1), pp. 81–88.

[44] Martin, J., (2000), *What Works among Labor Market Policies: Evidence from OECD Countries' Experiences*, in: *OECD Economic Studies*, No.30.

[45] Martin, J.P. and Grubb, D., (2001), *What works and for whom: a review of OECD countries' experiences with active labour market policies*, IFAU Working Paper 2001:14.

[46] Newell, A. and Pastore, F., (1999), *Structural Unemployment and Structural Change in Poland*, in: *Studi Economici*, Vol. 69(3), pp. 81–100.

[47] Münich, D., Svejnar, J. and Terrell, K., (2000), *The Worker-Firm Matching in the Transition: (Why) Are the Czechs More Successful than Others?*, Working Paper 107, The William Davidson Institute

[48] Ng, S. and Perron, P., (1995), *Estimation and Inference in Nearly Unbalanced, Nearly Cointegrated Systems*, in: *Cahiers de recherche 9534*, Centre interuniversitaire de recherche en économie quantitative, CIREQ.

[49] O' Leary, C., Nesporova, A. and Samorodov, A., (2001), *Manual on Evaluation of Labor Market Policies in Transition Economies.* International Labour Office

[50] Obstfeld, M. and Peri, G., (1998), *Regional nonadjustment and fiscal policy: lessons from EMU*, NBER Working Paper 6431, NBER.

[51] van Ours, J. and Vodopivec, M. (2005), *How Changes in Benefits Entitlement Affect Job-Finding: Lessons from the Slovenian "Experiment"*, World Bank Employment Policy Primer, April 2005

[52] Overman, H. and Puga, D., (2002), *Regional Unemployment Clusters*, in: *Economic Policy*, Vol. 17(34), pp. 117-147.

[53] Pehkonen, J. and Tervo, H., (1998), *Persistence and Turnover in Regional Unemployment Disparities*, in: *Regional Studies*, Vol. 32(5), pp. 445–458.

[54] Perugini, C., Polinori, P. and Signorelli, M., (2005), *GDP Growth and Employment in EU's Member Regions: Convergence Dynamics and Differences for Poland and Italy*, EACES Discussion Paper 2, EACES.

[55] Petrongolo, B. and Pissarides, C. A., (2001), *Looking Into the Black Box: A Survey of the Matching Function*, in: *Journal of Economic Literature*, Vol. 39(2), pp. 390-431.

[56] Saczuk, K. and Tyrowicz, J., (2009), *Wage Inequalities Explosion Or Is It?* National Bank of Poland, *mimeo*

[57] Sala-i Martin, X., (2002), *15 Years of New Growth Economics: What Have We Learnt?*, Technical report.

[58] Sarno, L. and Taylor, M. P., (1998), *Real Interest Rates, Liquidity Constraints and Financial Deregulation: Private Consumption Behavior in the U.K*, in: *Journal of Macroeconomics*, Vol. 20(2), pp. 221–242.

[59] Scarpetta, S. and Huber, P., (1995), Regional Economic Structures and Unemployment in Central and Eastern Europe. An Attempt to Identify Common Patterns, The Regional Dimension of Unemployment in Transition Countries. A Challenge for Labour Market and Social Policies, OECD.

[60] Sciulli, D., (2005), *Making the Italian Labor Market More Flexible: An Evaluation of the Treu Reform*, mimeo

[61] Silverman, B., (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall.

[62] Srivastava, V. K. and Giles, D. E. A., (1987), *Seemingly Unrelated Regression Equations Models*, New York: Marcel Dekker Inc..

[63] Svejnar, J., (2002a), *Labour Market Flexibility in Central and East Europe*, William Davidson Institute Working Paper, William Davidson Institute.

[64] Svejnar, J., (2002b), *Transition Economies: Performance and Challenges*, in: *Journal of Economic Perspectives*, Vol. 16(1), pp. 3-28.

[65] Tyrowicz, J., (2006), *Relations Between Public Employment Services and Labour Market Organisations: Results of a Quantitative Survey*, in: *NGOs On the Labour Market in Poland*, ed: Michał Boni, Ilona Gosk, Joanna Tyrowicz, Bartłomiej Piotrowski, Kuba Wygnański, Fundacja Inicjatyw Społeczno-Ekonomicznych, Warsaw, 2006 (in Polish)

[66] Tyrowicz, J. and Wójcik, P., (2007), *Regionalna konwergencja bezrobocia w Polsce*, Gospodarka Narodowa, 5(2007)

[67] Tyrowicz, J. and Wójcik, P., (2009), *Unemployment Convergence in Transition*, in: E. Marelli and M. Signorelli (eds), *Growth and Structural Features of Transition*, London Palgrave Macmillan

[68] Vodopivec, M., Wörgötter, A. and Raju, D., (2003), *Unemployment Benefit Systems in Central and Eastern Europe: A Review of the 1990s*, March 2003, World Bank Social Protection Discussion Paper Series No. 0310

[69] Vroman, W., (2002), *Unemployment Insurance and Unemployment Assistance: A Comparison*, World Bank Social Protection Discussion Paper, DP No 0203

[70] Zellner, A., (1962), *An Efficient Method of Estimating Seemingly Unrelated Regression Equations and Tests of Aggregation Bias*, in: *Journal of the American Statistical Association*, Vol. 57, pp. 500-509.

Appendix

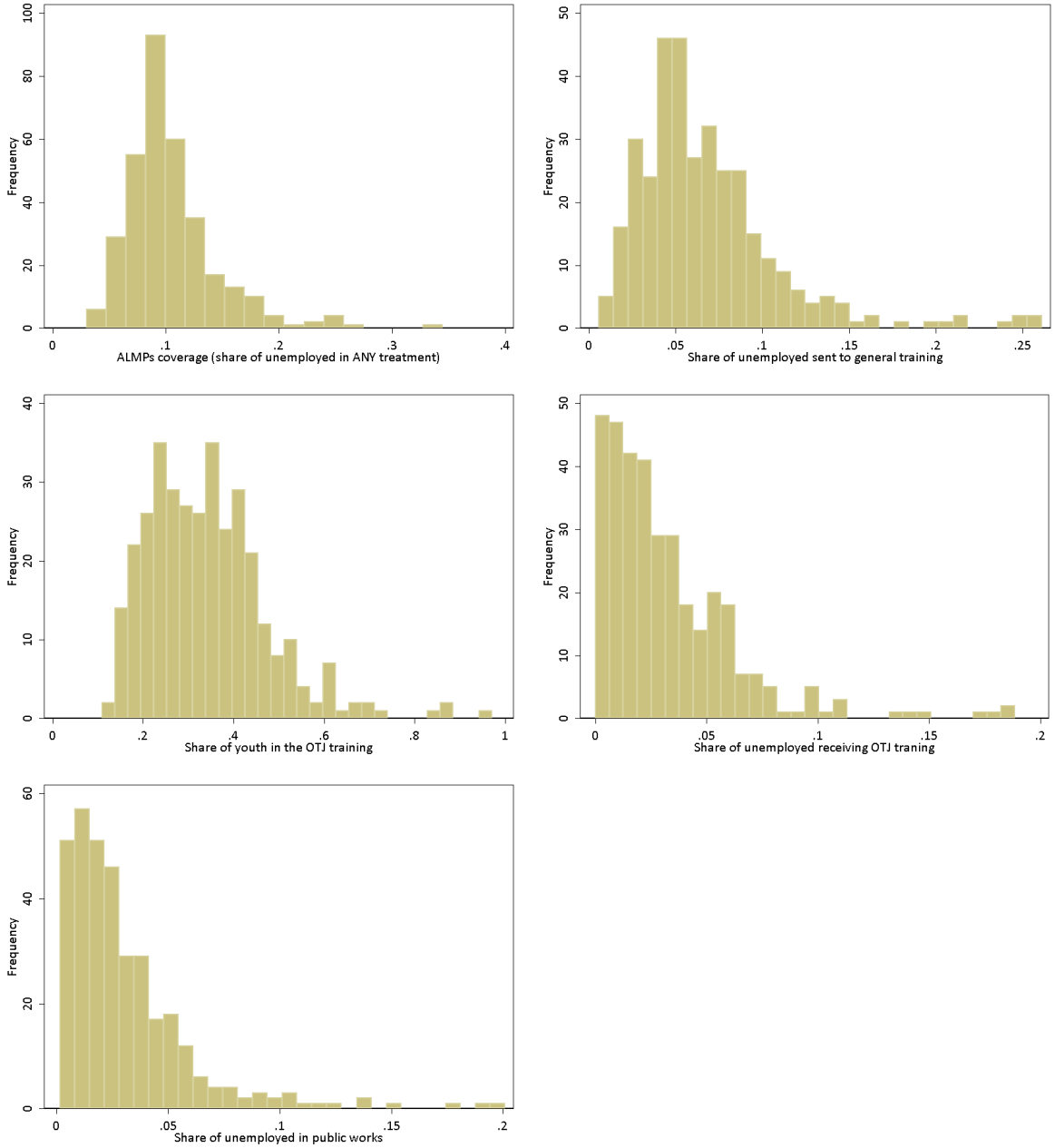


Figure 10. Distributions of ALMPs instruments usage

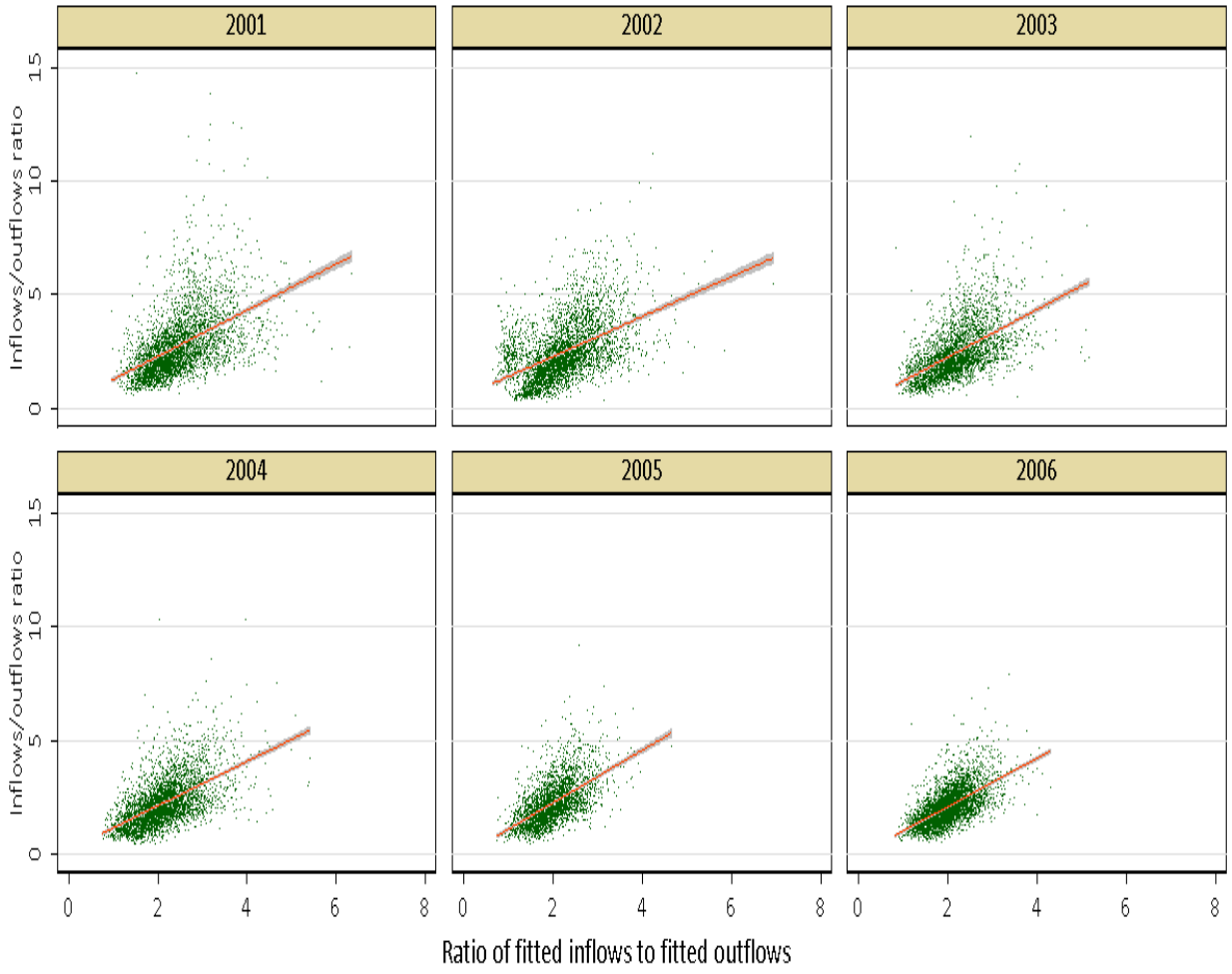


Figure 11. Actual versus fitted inflows/outflows ratios

Journal of Applied Economic Sciences

CARRY TRADE FUNDAMENTALS AND THE FINANCIAL CRISIS 2007-2010

Claus **VISTESEN**
Global Economy Matters
clausvistesen@gmail.com

Abstract:
This paper takes the form of an event study surrounding the current financial crisis. It proposes a theoretical relationship which can be used to model traditional carry trade crosses on a daily return basis as a negative function of equity returns and a positive function of market volatility. In order to test this theory, an Arbitrage Pricing Theory framework is adopted which is used to estimate the factor betas of carry trade crosses with respect to equity returns and market volatility. It is shown how the variation in the currency crosses explained by the functional relationship as well as the estimated factor betas have increased significantly in relation to the financial crisis. The results indicate that low yielding currencies (the JPY and CHF) can be successfully modelled as a negative function of equity returns and a positive function of volatility in the market. The results furthermore underpin studies that have shown how carry trading activity is highly sensitive towards sudden sparks of volatility and risk aversion, and thus how carry trade fundamentals are time varying.

Keywords: International finance, carry trading, financial crisis, currencies,

JEL Classification: F3, F31, G15

1. Introduction

One of the most vexing features of today's international financial markets is the carry trade phenomenon which exploits wide global interest rate differentials to earn the spread between low yielding and high yielding currencies. Carry trading consequently violates one of the few fundamental theories we have to explain currency markets; the uncovered interest rate parity (UIP). The UIP states that the expected change in the spot rate must reflect the interest differential between the two currencies. The theory predicts that the country with the high interest rate will see its currency depreciate (i.e. as it is assumed ex ante that the higher interest rate is a compensation for this depreciation). In formal terms:

$$(E)\Delta S = (1+i_h)/(1+i_f) \quad (1)$$

Where $i_{h,} i_f$ are interest rates in "home" and "foreign" respectively? Regarding the UIP, Bilson (1981) is often referred to as the initial study to reject the hypothesis, but also Meese and Rogoff (1983) and Longworth (1981) provide evidence to reject it. However, the evidence against the UIP is not entirely uniform. Chinn and Meredith (2004) manage to differentiate the conclusions from the main bulk of the literature. In their 2004 IMF staff paper, they consequently find that the UIP holds over longer time horizons. Furthermore, they show how failure of UIP to hold in the short run can be attributed to the interaction between shocks on the exchange rate market and endogenous monetary policy reactions.

Under the conditions of the UIP, the interest rate differential should be exactly offset by a change in the spot rate over the investment period in question. In this regard, the mechanics of the carry trade are interesting in the sense that a vigorous pursuit of carry trade by investors can turn into a self-fulfilling violation of the UIP; something which Plantin and Shin (2008) have coined as self-reinforcing arbitrage Brière and Drut (2009). In this way, the pursuit of carry trade will tend to keep low yielding currencies from appreciating against high yielding currencies since the aforementioned are being sold in the carry trade transaction itself. Moreover, many investors don't actually need to perform the carry trades per se,[1] but simply latch on to the trade in the sense that they, in the spot market, sell the most common funding carry trade currencies (CHF and JPY) against the most

[1] E.g. through constructing money market instrument portfolios in high interest rate currencies with borrowed funds in low interest rate currencies.

common (and liquid) high yielder; for example Gagnon and Chaboud (2007) find evidence of carry trading behaviour with respect of the JPY. Specifically, it is the effects and determinants of this latter strategy, or piggy backing if you will, which is of interest to this paper.

It is clear that such activity cannot be expected to create positive returns on a consistent basis, and periods of volatility and sudden reversals of asset prices can prove devastating for carry trade investors since positions are often highly leveraged Brière and Drut (2009). Nevertheless, and given the lingering persistence of wide global interest rate differentials some scholars have attempted to account for the ability to make consistent profits from carry trading. In Olmo and Pilbeam (2008) carry trading is however not found to yield excess returns for the most common carry trading crosses. Curiously, the authors do find excess returns in the context of the GBP/USD cross which is somewhat odd given that interest rate differentials between the US and UK tend to be significantly narrower than other potentially more "juicy' trades". Brière and Drut (2009) specifically show how fundamental strategies based e.g. on PPP tend to outperform carry trade strategies in the context of crises. These results are mirrored by Corcoran (2009) who shows, in an arbitrage-pricing-theory (APT) framework, how excess carry trade returns earned by a US investor investing in foreign money market instruments (t-bills) are explained by equity market and exchange rate volatility. This also supports studies by Brunnermeier et al (2008) and Farhi and Gabaix (2008) who show how currency crashes, and essentially sovereign defaults in the context of highly leveraged *high interest rate* economies, can explain carry trade risk premiums.

This paper does not directly attempt to qualify these studies but rather assume, ex ante, that carry trading exists as an integral part of market practice and discourse. As such, it is of less importance to the conclusions of this paper that carry trading works (i.e. earns excess returns) than it is important to assume that investors act according to the tenets of carry trading. Specifically, this study takes the form of an event study surrounding the current financial and economic crisis that has gripped global markets.

This opens the door for an investigation of one of the interesting derivative effects from carry trading activity. One question which thus seems pertinent is the extent to which carry trading activity as measured by movements in the most common funding currencies can say something about general market conditions. Clearly and assuming that carry trading does not create positive returns on an universally consistent basis it would be interesting to gauge the extent to which shifts in *"carry trading behavior"* coincides with other changes in the market. This is exactly what this paper sets out to examine in the context of the credit turmoil and thus to pin down the notion of carry trade fundamentals. In doing so, it is however important to point out that this paper firmly inserts itself in the tradition of the most recent studies on carry trading activity. These studies are Corcoran (2009) which shows how returns on carry trade are approximated through equity and exchange rate volatility, Cairns *et al.* (2007) which shows how "low yielders" can be modeled as a positive function of volatility, and finally; Kohler (2007) and Brière and Drut (2009) who show how equities can be modeled as negative beta assets to low yielders. The crucial point however to emphasize is that this paper attempts to model exchange rates as a function of volatility and equity returns and how this might have changed in the context of the current financial crisis. As such, this paper follows the same path as Christiansen *et al.* (2009) which presents an econometric model to suggest that carry trade crosses and strategies are subject to time-varying systematic risk or more specifically that the fundamentals of carry trade strategies change with market conditions.

The paper proceeds as follows. Section two presents the theoretical framework, section three presents the estimation and results, section four discusses the results and section five concludes.

2. Theoretical Framework

Since the end of July 2007 equity markets across the global have weakened significantly and given recent forecasts as e.g. the one propounded in the IMF's 2009 World Economic Outlook, we are going to be stuck in the mire for some time.

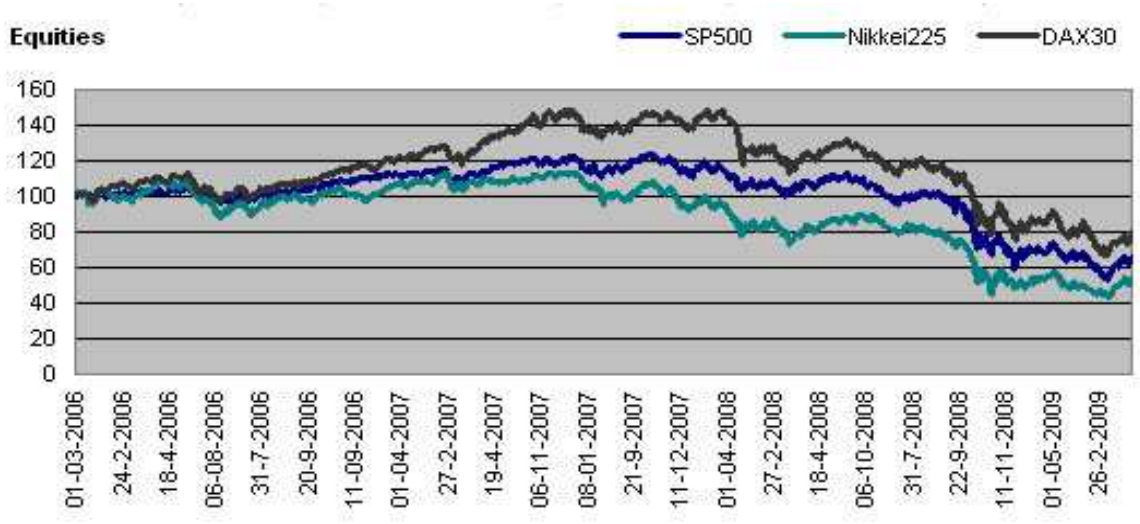


Figure 1

Source: Thomson Data stream

In the context of the credit turmoil, this has led to a discourse surrounding *unwinding* of risky carry trade positions. One key element in this discourse is how the funding currencies for carry trades (here, the JPY and CHF) are being coined as risk sentiment gauges, and thus measures of risk in the market place. The *unwinding* effect in this regard would then, in part, be conjured by investors' and traders' abandonment of highly leveraged spot market positions against the CHF and JPY. One way to operationalize this would be to narrate the CHF and JPY as the famous canaries whose demise were used by coal miners in the 19th century Britain to gauge when it was time to get out of the mine due to the presence of toxic gasses. In this way, CHF and JPY crosses can equally be seen as canaries in the context of financial markets whereby a sudden spike of volatility or a downward correction in risky assets is followed by an appreciation of the funding carry trade currencies as positions are unwound. Formally, the mechanics of such movements would suggest a negative correlation between the CHF and JPY and risky assets which would follow the results in Corcoran (2009), Kohler (2007), Brière and Drut (2009), and Cairns *et al.* (2007). Moreover, this would also suggest that we should have observed a strengthening across the board of the low yielding currencies since August 2007. This however is not uniformly so, as can be seen below.

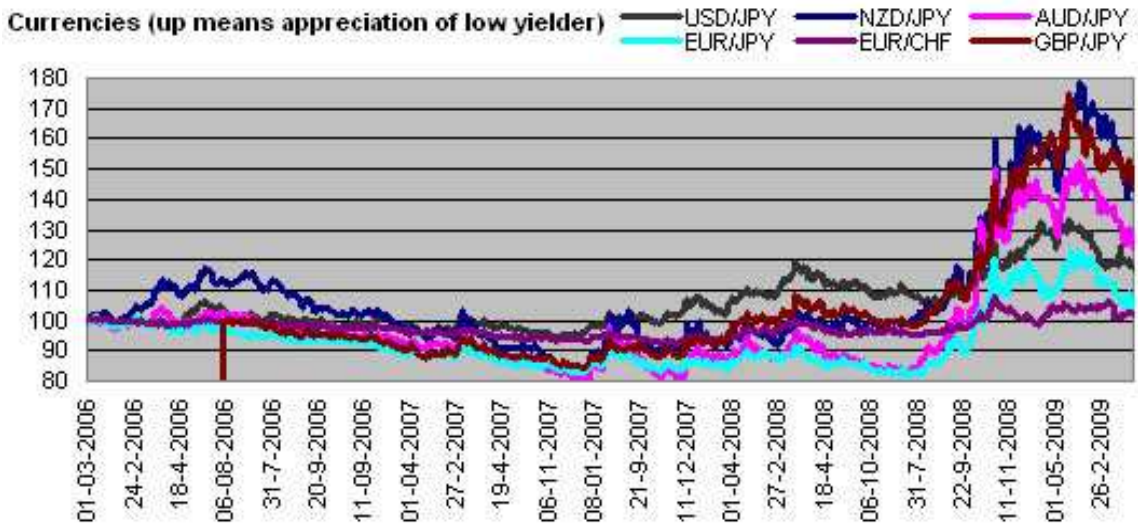


Figure 2

Source: Thomson Datastream

As can readily be observed, the beginning of the credit turmoil has seen significant divergence between the JPY and CHF crosses. Yet, this is merely if we look at the levels of the time series. If we look at the daily trend there thus seems to be considerable negative co-movement between equities and the low yielders (in level form). In fact, if we home in on the two graphs above even a scant glance suggest a negative correlation between equities and low yielding currencies. It is exactly this tendency which is of interest in the present context.

Also, if we turn the attention to volatility let us first confirm the fact that volatility has increased markedly since the credit turmoil took hold in august 2007.

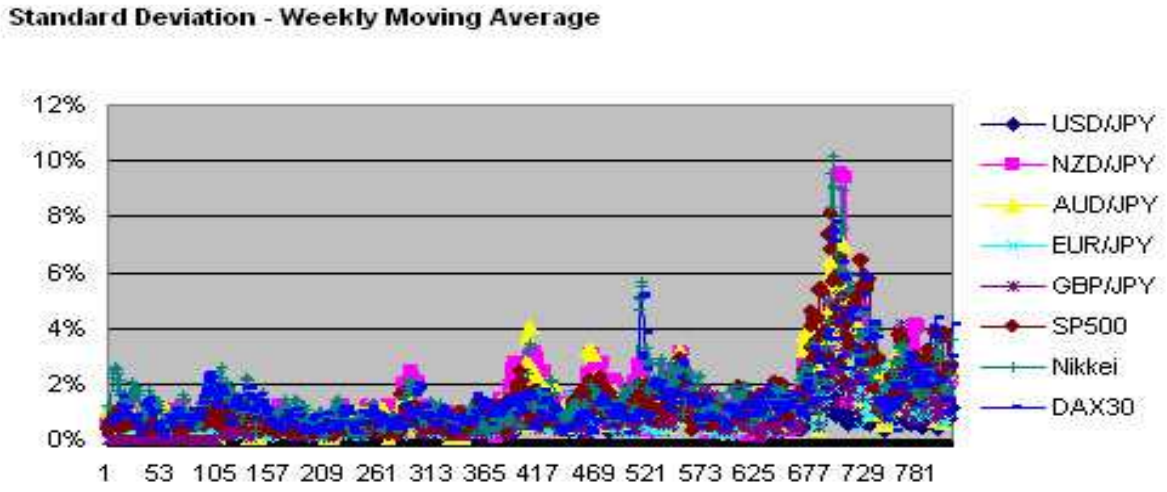


Figure 3

Source: Thomson Data stream and CBOE (Jan 2006 – April 2009)

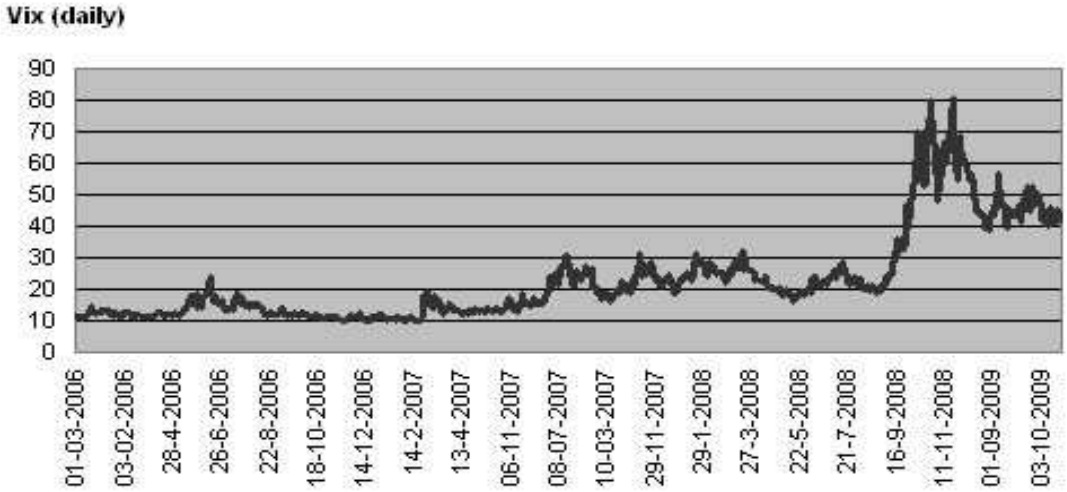


Figure 4

Source: CBOE

Both exchange rate volatility, idiosyncratic equity volatility, and general market volatility as measured by the VIX have increased significantly. If we focus the attention on the VIX, and use 01-08-2007 as the starting point of the crisis[2], the result is very clear.

[2] This data will be used as a breaking point throughout.

Table 1

Vix[3]	
Mean(1)	13,15320707
SD(1)	2,448010521
Mean(2)	32,00909953
SD(2)	14,57665003

Source: CBOE, own calculations

Consequently, both the mean value and standard deviation of the measure, which can be interpreted as a second derivative effect, have spiked significantly in a post crisis. This suggests that both the *level* and *variation* of volatility have increased. Following the theme of the present study one would expect low-yielding currencies to exhibit a positive relationship with volatility.

Notional evidence of carry trade dynamics is easy to find. Daily readership of Bloomberg's financial news stream will thus often present market participants with headlines such as *Yen Falls as Asian Stock Gains Boost Confidence in Carry Trades*[4], which is indicative of the relationship described above. Moreover and apart from an account of the theatricals of financial markets such reports also highlight two other points. First of all, it indicates that the argument upon which this paper builds its case is already formalized in the daily market discourse. Secondly, it suggests that the relationship is one which, at the very least, can be tracked on a short term frequency basis. Consequently, this paper studies daily returns within a, for traditional empirical purposes, relatively short period.

Following the points above the inquiry begins with the following expression for the functional relationship between the return of a funding currency in a carry trade transaction (the JPY and CHF in this case).

$$R_{fx} = f\left(-R_e, \sigma\right) \tag{2}$$

where the subscript "fx" indicates that the left hand side is an exchange rate. In order for the expression above to make intuitively sense the currency pair should be quoted as number of high yielding currency to low (i.e. directly). Thus, if the USD/JPY is traditionally quoted as amount of JPY to USD (e.g. 110), the expression used here will be (1/[USD/JPY])[5] in order to convey the idea of the low yielders as negative beta assets at the same time as they are a positive function of volatility. Theoretical impetus for the choice of this functional form can be found in Zimmerman *et al.* (2003) who point towards two important points. One the one hand they detail how stock market volatility is higher in down periods (bad news spawn more volatility than comparable good news). Given that volatility is supposed to adversely affect carry trade returns this supports the findings by Brière and Drut (2009) and Corcoran (2009). Secondly, they also question the merits of international diversification by showing that in down periods when volatility is high and when economic activity is shrinking, we also observe a significant increase in correlation amongst international securities Zimmerman *et al.* (2003).

There may be reason to believe that this functional form has general validity across time, but in the context of the present study we can amend the expression in one crucial way. Consequently, and bearing in mind that this study takes the form of an event study in the sense that it studies pre and post crisis dynamics, we can deduce the following expression:

$$E\left\langle R_{fx} \middle| \theta \right\rangle = f\left(-R_e, \sigma\right) \tag{3}$$

[3] Where (1) means period 1 before the crisis and (2) indicates period 2 after the crisis set in.
[4] 2008 Bloomberg News Article
[5] i.e. amount of USD per JPY

Consequently, the functional form of the expectation of the return of a low yielder in a carry trade transaction becomes conditional on the value of (θ). The parameter (vector) θ indicates that we are in a crisis. Clearly, the vector θ is rather innocuous in the present context and will not be subject to direct analysis, but following the remarks above it must incorporate measures such as volatility, equity returns, as well as real economic variables all imbued in order to identify a period of recession or crisis.

To operationalize the proposition above, this paper follows the intuition from Arbitrage Pricing Theory Ross (1976) and the one adopted in Cocoran (2009) by letting the return on a carry cross (quoted directly) to be modelled as a linear combination of k factors.

$$\begin{aligned} R_{fx} &= \alpha_i + \beta_{i1} I_1 + \beta_{i2} I_2 + ... + \beta_{ij} I_j + e_{ij} \\ E[e_i e_j] &= 0 \end{aligned} \tag{4}$$

In our present cast, the proposed functional form will be the following;

$$R_{fx} = \alpha_i + \beta_{i1} I_1 + \beta_{i2} I_2 + e_{ij} \tag{5}$$

With:

R_{fx} equal to the return on a low yielding carry trade currency (e.g. a long USD/JPY position when quoted directly).

α_i equal to the expected value of R_{fx} if the risk factors are equal to 0. In this case and with the method adopted here of using first differences of daily values $E(\alpha_i) = 0$; I assume mean reversion in the first difference.

I_1 is equal to the return vector of an equity index.

I_2 is equal to the vector of the VIX (high value) in changes.

In a standard APT framework and following Cocoran (2009) one would first estimate the factor betas using the approach of Fama and Macbeth (1973) through time series regression and then move into the cross-section in order to estimate the factor prices (risk premiums). In this study the focus will be on the first stage, as it were, of this approach and thus the value of the factor betas. This leads to the estimation of the following equation.

$$\ln\left(\frac{\gamma_t}{\gamma_{t-1}}\right) = \alpha_0 + \beta_1 \ln\left(\frac{R_{mt}}{R_{mt-1}}\right) + \beta_2 \ln\left(\frac{Vix_t}{Vix_{t-1}}\right) + e_t \tag{6}$$

Which we can rewrite as:

$$\Delta\gamma = \alpha_0 + \beta_1 \Delta R_m + \beta_2 \Delta Vix + e_t \tag{7}$$

The functional form which incorporates the variables in changes (continuous compound) is chosen in order to avoid stationarity issues when performing time series analysis on level form variables. The value for the VIX is the change in the value of the *high value* on a daily basis. This is used in order to capture the peak level of volatility in the VIX and whether the carry currency pairs react to sharp reversals in implied market volatility.

Since this paper studies the relationships sketched above in relation to an event in the form of the current crisis, the stability of the proposed relationship will also be investigated. It is thus interesting for this study to break up the expression above into one in a pre crisis framework and one in a post crisis framework. This takes us into the world of econometric tests for parameter stability Chow (1960), Gujarati (2003) and Greene (2003, pp. 130-147).

A first simple test involves the entire estimation of the regression following Chow (1960) and indicates whether there has been a structural break in the parameters without telling us which of the estimated parameters that have changed. Consider consequently the following approach Gujarati (2003) and assume the generic regression for the whole period as stated above and then amend it with the following regressions for period one and two respectively;

$$\Delta\gamma^{*} = \alpha_0^{*} + \beta_1^{*}\Delta R_m^{\ *} + \beta_2^{\ *}\Delta Vix^{*} + e_t^{*} \text{ and } \Delta\gamma^{\cdot\cdot} = \alpha_0^{\cdot\cdot} + \beta_1^{\cdot\cdot}\Delta R_m^{\ \cdot\cdot} + \beta_2^{\cdot\cdot}\Delta Vix^{\cdot\cdot} + e_t^{\cdot\cdot} \qquad (8)$$

Where (*) indicates a regression for period 1 and (¨) indicates a regression for period 2. The mechanics of the Chow Test assumes that $\alpha_0 = \alpha_0{*} = \alpha_0^{\ \cdot\cdot}$ in all three estimations but also more importantly that $\beta_1 = \beta_1{*} = \beta_1^{\cdot\cdot}$ as well as $\beta_2 = \beta_2{*} = \beta_2^{\cdot\cdot}$. In performing the Chow Test we test whether the residual sum of squares (RSS) from the original regression is statistically different from the sum of the RSS from the two period regressions. Formally, the test is conducted by calculating the following F-value:

$$F = \frac{\left(RSS_R - RSS_{UR}\right)/K}{\left(RSS_{UR}\right)/\left(n_1 + n_2 - 2k\right)} \sim F_{[k,(n_1+n_2-2k)]} \qquad (9)$$

Where RSS_R is the residual sum of squares from the original full sample size regression and RSS_{UR} is the sum of residual sum of squares from the two separate period regressions. If the F statistic is sufficiently large, we reject the null of no structural break.

Another more rigorous approach is to follow Gujarati (2003) and Greene (2003) and apply dummy variables to check which of the parameters that change and how much. In this way, I specify the following regression to be estimated.

$$Y = \alpha_0 + \beta_1\left(\alpha_1 D_t\right) + \beta_2\Delta R_m + \beta_3\Delta Vix + \beta_4\left(D_t\Delta R_m\right) + \beta_5\left(D_t\Delta Vix\right) + e_t \qquad (10)$$

Where D_t is dummy variable which takes the value of 0 if we are in period one (pre-crisis) and 1 if we are in period 2 (post crisis). An estimated parameter for β_1, β_4 or β_5 significantly different from 0 indicates a structural break for the beta value of the intercept, market return, and volatility respectively. In this case, the new parameter coefficient estimated for period 2 will be given by $\alpha_0 + \beta_1$ for the intercept, $\beta_2 + \beta_4$ for the market return, and $\beta_3 + \beta_5$ for volatility Gujarati (2003). This approach allows us to scrutinize specific change in parameters across periods and is a valuable addition to the observation of changes in the overall coefficient of determination (R-sq) of the regression across periods.

3. Estimation and Results

Thomson Data stream was used to pull data on 6 currency pairs considered to be traditional carry trade crosses. Of the six, one CHF crosses and five JPY crosses have been used.[6] Furthermore three major stock indices from three main regions in the form of the SP500, the Nikkei 225 and the DAX30 were chosen as the market(s). As for the term for the volatility term it will be proxies through the use of The CBOE Volatility Index (VIX) which is a measure of market volatility calculated through the use of options on the SP500[7]. The data series consists of daily values (returns) of the seven currency crosses and the three stock market indices from 01-03-2006 to 04-02-2009 of a total of 817 daily observations[8]. These data sets form the basis of the estimation below.

[6] USD/JPY, NZD/JPY, AUD/JPY, EUR/JPY, EUR/CHF and GBP/JPY

[7] Daily data was obtained from the Chicago Board of Options Exchange's website (daily values at high).

[8] Since the VIX does not display observations on all the sample days, all time series have been cut to fit the schedule of the VIX.

In order to set the stage for the estimations above it would be interesting initially to have a look at simple correlations (of the time series in changes) and see whether these confirm the theoretical framework described above. Specifically, it is interesting to observe whether there has been a change in a post crisis perspective. This initial evidence seems to provide a solid foundation for the hypotheses stated (see appendix). If we look at the full sample, all currency crosses are positively correlated with the VIX index and this correlation has increased markedly in a post crisis perspective. The mean increase in correlation with the VIX for all currency pairs, in a post crisis perspective, is a sound 173%. In terms of the currency pairs' correlation with the equity indices it is, for the most part, negative. Only the NZD/JPY's and AUD/JPY's positive correlation with the SP500 cloud the picture. In a post crisis perspective however, the results are unequivocal with the negative correlation for all currencies, except the NDZ/JPY and AUD/JPY, having increased on average with 258%, 125% and 152% for the Nikkei 225, Dax30 and SP500[9] respectively.

After these initial results, we turn to the estimation of the following relationship using OLS.

$$\Delta\gamma = \alpha_0 + \beta_1 \Delta R_m + \beta_2 \Delta Vix \tag{11}$$

Thus, the estimation of the currency crosses' factor betas shall be approximated by the equation above for a total of 18 regressions (3 stock market indices, 6 currency crosses and one volatility parameter). In the expression above, the estimated parameters (β_1, β_2) will be the main result to gauge. Given the theme of the present study and the fact that all currencies are quoted directly one would expect negative signs for β_1 and positive sign for β_2. First, the full sample regressions will be reported and then the investigation turns to the split dataset and the tests for structural stability.

The results for 18 regressions are reported in the tables in the appendix. In general, the full sample regressions yield results which are strongly indicative of the theoretical relationship noted above. Yet, the initial results from the full sample regressions are misleading. An important initial observation in this regard is the increase in the models' R-sq values across periods. In percentage points[10] the average increase in R-square values is 14%, 27% and 17% for the regressions including the SP500, the Nikkei 225 and the Dax3030 respectively. This suggests, with some force, how the proposed relationship is particularly strong in a context of a financial and economic crisis. All R-square values calculated in a post-crisis perspective are significant at 1% (which was not always the case in the pre-crisis regressions), and their values indicate a relatively strong explanatory power. Especially, there are 13 regressions in the post-crisis context which have R-square values above 0.2 which, in the present context, must be considered a strong result since we are dealing with first differenced daily time series.

Turning to the estimated coefficients and the idea of the currency crosses as negative beta assets to equities as well as the hypothesis that they can be modelled as a positive function of volatility, the waters get increasingly muddier.

With regards to the Nikkei 225 and the DAX30 the factor prices of the currency crosses all correspond with the theoretical framework as they have negative beta values which increase markedly in the second period estimations. The results are more disappointing for the SP500 in this regard where only the USD/JPY and GBP/JPY conform to the relationship proposed with negative beta values that are higher (and statistically significant) in the second period estimation. In terms of the estimations in relation to the VIX, the results are strong and unequivocal. In the full sample regression most currency pairs are successfully modelled as a positive function of volatility which is consistent with carry trade fundamentals in which investors buy into relative low yielding currencies (unwinding carry trade positions) when volatility spikes. This result is intensified when we look at the difference between period one and two. Both in connection to the level of statistical significance and in relation to the value (and signs) of the estimated coefficients do we observe an increased strength in the models' ability to model the currency pairs as a positive function of volatility. The only exceptions

[9] Excluding the NZD/JPY and AUD/JPY since these do not exhibit a negative correlation with the SP500 in the first place.

[10] Since by definition; 0<r-sq<1.

here are the regressions for the NZD/JPY and AUD/JPY in relation to the Nikkei 225 where the parameter estimated for the VIX is not statistically significant.

In summary, there appears to be strong evidence for the proposed theoretical relationship above in which, conditional on crisis dynamics, relative low yielding currencies can be modelled as negative beta assets to equities and positive functions of volatility. In order however to quantify this result, the investigation now turns to the examination of parameter stability across the two periods.

As a first approximation, the *chow test* Chow (1960), Gujarati (2003) and Greene (2003) will be performed based on the F-test showed above. As noted, RSS_R is the residual sum of squares from the original full sample size regression and RSS_{UR} is the sum of residual sum of squares from the two separate period regressions. $(n_1 + n_2 - 2k)$ is equal to (395+420)-(2*3) = 809 and the critical values of the F is 2.1, 2.61 and 3.78 for 10%, 5% and 1% level of significance respectively. The null is that there is no structural break which means that a significant F-value would indicate that a structural break is present as per reference to rejection of the null. In the table below the computed F value is shown for all the 18 regressions.

Table 2. Chow test stats

Chow-test stats[11]	USD/JPY	NZD/JPY	AUD/JPY	EUR/JPY	GBP/JPY	EUR/CHF
Sp500, VIX	579.99	573.23	578.31	582.96	514.45	564.09
Nikkei 225, VIX	581.97	553.70	585.75	582.99	510.86	559.04
Dax3030, VIX	576.93	555.66	559.85	567.30	501.07	552.62

The F-statistics computed above strongly support the results of a structural break in the regressions around at the advent of credit crisis. They are consequently all well within the confines of statistical significance at 1%.

These F-statistics however tell us nothing about which of the estimated parameters that have changed. This is of interest in the present context since we have two explanatory variables (equity returns and the VIX) and it would be useful to know which of these two variables that is to blame, as it were, for the structural break. Moreover, it would be nice to rule out the possibility that the structural break is due entirely to a change in the level of the currencies, which would be captured by a significant change in the intercept. In this way, we proceed with the following estimation.

$$Y = \alpha_0 + \beta_1(\alpha_1 D_t) + \beta_2 \Delta R_m + \beta_3 \Delta Vix + \beta_4 (D_t \Delta R_m) + \beta_5 (D_t \Delta Vix) + e_t \qquad (12)$$

The output of these regressions is reported in its entirety in the appendix and by nature, it is a bit difficult to get an immediate overview.[12] The following points are worth paying attention to. First of all, all the intercepts and the respective period dummies used to capture any structural break due to a change in the average daily change of the currencies are statistically insignificant.[13] This is in line with the expectations noted above.

In terms of regressions where both the parameters for the VIX and the equity index are significant, there are 7. These are the USD/JPY, the EUR/JPY, and GBP/JPY to the DAX30 and Nikkei 225 respectively as well as the EUR/CHF to the Nikkei 225. This indicates that the effect from changes in volatility and equity returns have been greater in a post-crisis perspective. In these regressions the average increase in the beta parameter for the VIX is 0.03 and for the equity dummies the number is -0.11 for the Nikkei 225 and 0.15 for the DAX30. These numbers may appear small, but it is worth remembering in this case that we are talking about daily returns and thus an interval where small changes have a potentially high impact. In terms of the SP500, the results are poor in so far as goes the fact that none of regressions exhibit statistically significant dummies for both the VIX and the

[11] The Vix is of course included in all these regressions too.

[12] With 18 regressions consisting each of 6 explanatory variables there are 108 parameters to deal with. As such, the reader is advised to read the whole paper before digging into the specifics of this regression output.

[13] Except for the GBP/JPY to the Nikkei 225 and Vix, but since the second period intercept has a p-value of more than 0.1 I do not consider this to be a credible result.

equity indices. In fact, none of the regressions show a significant increase in the beta value for the equity index whereas, in many of the cases, the VIX dummy variable is significant. This suggests that the relationship between the SP500 and the carry trade crosses in question here have not changed much even if the models' ability to explain the variation (the R-sq) has indeed increased.

The results for the VIX dummy are, in general, strong. Only in two of the 18 regressions do we observe that there has *not* been a structural break in the estimated coefficient for the VIX. This indicates that the effect from changes in volatility on the currency crosses and thus a carry trade position has increased significantly since the advent of the credit crisis. The result is less robust for the equity parameters although it seems that, in the cases where the dummies are significant, the change is relatively high. Consider for example the dummy for the Nikkei 225 index to the NZD/JPY, AUD/JPY and EUR/JPY which shows that the beta value of these currencies to the Nikkei 225 have increased (in negative values) by 0.235, 0.364, and 0.189 respectively. In general, the results concerning structural breaks with respect to equity betas are unequivocal in the sense that the dummies for the Nikkei 225 are all significant whereas the picture is more clouded for the DAX30 and SP500. This indicates that the findings by Hau, H, and H, Rey (2006) whereby higher returns on domestic equity market are associated with a depreciation of the home currency are perhaps showing up in these estimations.

4. Discussion

The estimation above presents several interesting results. As a first initial summary the results significantly underpin the theoretical framework sketched earlier. Not only do the vast majority of the currency crosses exhibit negative beta values to the three main stock indices but also, at the same time, they can be modelled as positive functions of market volatility.

In terms of the differentiation between the two periods and thus the real objective of this study, the results are quite strong. It is important, I think, in this respect to point to the fact that the r-square values for period 2 are markedly higher than in period 1. Given that the present study deals with daily returns it strongly suggests that that the proposed relationship has intensified in strength after the financial crisis took hold. This supports the findings of Christiansen *et al.* (2009) that the strength of carry trade fundamentals is time varying.

It is also important to point out that the tests for structural break do not test for the strength of the relation as measured by the R-sq, but rather the value of the estimated parameters. This investigation produced decidedly murkier results, but still indicates that key relationships have intensified. Especially, it seems as if the carry crosses' functional relationship with volatility has increased significantly. Also, all the currency crosses' negative relationship with the Nikkei 225 index has increased in a post-crisis perspective. Consequently, the results which show carry trade currency pairs as negative beta assets seem particularly strong in the context of the Nikkei 225 index. However, it is also clear that if we look at the full sample period, not only the JPY crosses show negatively significant beta values to the Nikkei but so do the EUR/CHF. This strong result is echoed with the DAX30 where strong results are presented for other currency indices than the EUR/JPY and EUR/CHF. In relation to the SP500 the results were somewhat more meagre with the notable exception of the USD/JPY which has exhibited a strong structural break around the summer 2007. In overall terms, one could distinguish between the currency pairs by looking at their respective coefficients of determination. In this way, some of the models clearly offer a higher degree of explanatory power and thus, by derivative, a more believable act for some currency pairs as negative beta assets and positive functions of volatility. Examples here would be GBP/JPY and EUR/CHF to the DAX30, the EUR/CHF, AUD/JPY, and NZD/JPY to the Nikkei 225 as well as the USD/JPY to the SP500 and DAX3030.

Here, at the brink of the paper, (at least) three overall questions impose. The first is the question of structural stability of beta values or more specifically the sign of the estimated parameters. The second is the dodgier question of causality between currency pair and equity index and the third relates to the statistical issue of heteroscedasticity in the regressions.

On the first question this paper clearly falls outside the norm as it takes the form of an event study with daily returns over a relatively short time span. Considerable ink has been devoted by finance scholars in determining the estimation period which best approximates a stabile beta value (using the CAPM). At a first glance such studies are not directly replicable in the present context. In

this way, this study uses an APT framework to investigate the factor betas of currencies. However and as a rough approximation, results from the CAPM literature have shown that 4-6 years (about 300 observations with weekly returns) provide the strongest result Alexander and Chervany (1980). It has also been shown how extreme betas are shown to be less stable over time than betas drifting closer to the mean Alexander and Chervany (1980). The methodological glitches notwithstanding, it is interesting in the context of the present study. As such, one should be careful making general extrapolations on the basis of the findings above. On the other hand though, and given the strength of the results, effort should be put into pinning down which of these relationships hold up for scrutiny over time. Special attention should be devoted to pinning down the relationship $E\left\langle R_{fx} \middle| \theta \right\rangle = f(-R_e, \sigma)$ and what actually constitutes a reasonable proxy for the vector (θ). Given the theme of this study, volatility clearly seems to be a key variable. Finally, the stability of the relationship should also be held up against the findings by Christiansen *et al.* (2009) and thus the time varying aspect of the functional relationship.

Turning to the issue of causality, it is ironic that this paper initially was drafted with a model in which the currency crosses were used to model the equity returns. In this way, it would perhaps be best to leave this issue alone all together. One can consequently always quibble about causality in the context of statistical analysis even to such an extent to make the actual results secondary to the inquiry. This mistake will not be made here. In the regressions estimated above the idea has been to model carry trade crosses as a function of a number of carry trade fundamentals that were postulated. However, this does not mean that one could not achieve interesting results by switching the order of variables. Granger causality tests (1969) could of course be performed to formally ascertain the arrows of causality but in essence, the Granger test itself says very little about what really constitutes causality more than it merely provides a binary analysis of what affects what.

Finally, there is the issue of heteroscedasticity which seems to be an inbuilt issue of this study's methodology. The problem with heteroscedasticity in the context of OLS estimation and the Gaussian linear model is well known as it can create biased estimates of the beta parameters and underestimate the standard errors depending on the severity of the residuals' unequal variance. Consider consequently the regression framework estimated above through 2.7.

In order for the Chow test to be strictly valid and following Gujarati (2003) and Greene (2003) a prerequisite is that:

$$V\left(e_t^{*}\right) = V\left(e_t^{\cdot\cdot}\right) \tag{13}$$

However, given that the nature of the theoretical framework itself is built on the premise that volatility in one period is larger (different) than in the other, the issue here becomes a rather difficult one to deal with directly. In this way, a central prerequisite for this study will almost always be:

$$E\left[V\left(e_t^{*}\right)\right] \neq E\left[V\left(e_t^{\cdot\cdot}\right)\right] \tag{14}$$

This means de-facto presence of unequal variance in the two sub-periods. Gujarati (2003) performs a simple test to check whether there is a statistically significant difference between the variance of the residuals in the two estimation periods of the trial example.[14] The computed F-stat is found to reject the null of equal variance and thus the Chow test should not be used. Still, Gujarati (2003) is not adamant that this poses a serious issue. This is echoed in Greene (2003) where it is argued that as long as the sample size is large enough, unequal variance should not pose a major issue. Moreover in the present study all p-values, standard errors, and f-stats are highly significant to support the proposed relationship which should give us some confidence despite the obvious methodological and, as it were, practical issue with heteroscedasticity. The individual scholar should decide whether she believes that the method above can be applied or whether more elaborate techniques should be deployed to test for structural breaks in the estimated time series.

[14] GDP regressed on income and savings.

5. Conclusion

The principles of carry trading and how to bet against the theory of uncovered interest rate parity are well known. Moreover, carry trading and the effect of investors pursuing it, have almost turned in to an urban legend on financial markets where many derivative effects of *"carry trading behaviour"* are cited. This paper has attempted to scrutinize and essentially pin down the idea of carry trade fundamentals in relation to the financial crisis which gripped the global economy in 2007. Using an Arbitrage Pricing Theory framework it has been shown how the factor betas of carry trade currencies with respect to equity returns and market volatility have changed with the advent of this crisis. It has furthermore been shown how the strength of the variation between carry trades and equity returns as well as market volatility have increased markedly in the context of the financial crisis. The results indicate that low yielding currencies (the JPY and CHF) can be successfully modelled as a negative function of equity returns and a positive function of volatility in the market.

It has consequently been suggested how the JPY and CHF, often cited as the traditional funding currencies in carry trades, exhibit strong negative correlations and factor betas to equities (SP500, Nikkei 225 and DAX30) and positive factor betas to market volatility measured by the VIX. This lends evidence to the idea of the CHF and JPY as risk sentiment gauges and how this relationship strengthens in the context of a period of heightened volatility. In this regard it is important to watch the currency pairs with significant negative beta values in relation to equities and positive beta values for volatility; (GBP/JPY and EUR/CHF to the DAX30, the EUR/CHF, AUD/JPY, and NZD/JPY to the Nikkei 225 as well as the USD/JPY to the SP500 and DAX3030).

The key point to take away from this study is that the financial crisis has intensified the link between carry trade currencies and risky assets as well as volatility. However, it is equally important to emphasize how carry trade strategies will be especially sensitive to reversals in the context of a financial crisis Brière and Drut (2009). This also means that while it may seem tempting to hedge equity positions through long positions in carry trade currencies one has to be careful of reversals and the fact that these fundamentals are ultimately time varying.

Further studies on this topic should attempt to widen the time span of the sample to gauge the general validity of the results and thus follow in the steps of Christiansen *et al.* (2009) as well as attempt to make forecasts of daily exchange rate and/or stock returns based on the relationships cited above.

6. References

[1] Gordon, A.J., Chervany, L.N., (1980), *On the Estimation and Stability of Beta*, in: *The Journal of Financial and Quantitative Analysis*, Volume 15, no. 1 (march 1980), pp. 123-137.

[2] Bilson, J., (1980), *The Speculative Efficiency Hypothesis*, in: *Journal of Business*, Volume 54 (June), pp. 433-451.

[3] Brière, Drut (2009), *The Revenge of Purchasing Power Parity on Carry Trades During Crises*, CEB Working Paper No. 09/013 Feb. 2009.

[4] Brunnermeier, M., Nagel, S., Pedersen, L., (2008*), Carry Trades and Currency Crashes*, NBER Working Paper No. 14473, November.

[5] Menzie, C.D., Guy, M., (2004), *Monetary Policy and Long Run Horizon Uncovered Interest Parity*, IMF Staff Paper, Volume 51, no 3.

[6] Chow, C.G., (1960), *Tests of Equality between Sets of Coefficients in Two Linear Regressions*, in: *Econometrica*, Volume 28, no. 3, pp. 591-605.

[7] Christiansen, C., Ranaldo, A., Söderlind, P., (2009), T*he Time-Varying Systematic Risk of Carry Trade Strategies*, CREATES Research Paper 2009-15.

[8] Corcoran, A., (2009), *The Determinants of Carry Trade Risk Premia*, IIS discussion paper no. 287.

[9] De Santis, A.R., Sarno, L., (2008), Assessing *the Benefits of International Portfolio Diversification in Bonds and Stocks*, ECB Working Paper no. 883 (March 2008).

[10] Fama, E., MacBeth, J, (1973), *Risk, Return, and Equilibrium: Empirical Tests*, in: *Journal of Political Economy*, Volume 81, Issue 3, pp. 607-636.

[11] Farhi, E., Gabaix, X., (2008), *Rare disasters and exchange rates*, Working paper, Harvard University and NYU Stern.

[12] Gagnon, J.E., Chaboud, A., (2007), *What can the data tell us about carry trades in Japanese yen*, FRB International Finance Discussion Paper No. 899.

[13] Granger, C.W.J., (1969), *Investigating Causal Relations by Econometric Models and Cross Spectral Methods*, in: *Econometrica*, July 1969, pp. 424-438.

[14] Greene, W.H., (2003), *Econometric Analysis*, 5th edition Prentice Hall.

[15] Gujarati, D.N., (2003), *Basic Econometrics*, 4th edition McGraw Hill.

[16] Hau, H, Rey, H, (2006), *Exchange Rates, Equity Prices, and Capital Flows*, Review of Financial Studies 19; 273-317.

[17] IMF, (2009), *World Economic Outlook 2009*, IMF publication.

[18] Kohler, D., (2007), *Carry Trades: Betting Against Safe Haven*, University of St.Gallen Discussion Paper no. 2007-12.

[19] Longworth, D., (1981), *Testing the Efficiency of the Canadian–U.S. Exchange Market under the Assumption of No Risk Premium*, in: *Journal of Finance*, Volume 36, No. 1, pp. 43–9.

[20] Meese, R., (1989), *Empirical Assessment of Foreign Currency Risk Premiums*, in: *Financial Risk: Theory, Evidence, and Implications*, ed. by Courtenay Stone (Boston: Kluwer Academic Publications).

[21] Olmo, J., Pilbeam, K., (2008), *The Profitability of Carry Trades, Annals of Finance* 11 April 2008.

[22] Orléan, A., (1999), *Le pouvoir de la finance*, Odile Jacob Editions.

[23] Plantin G., Shin, H.S., (2008), *Carry Trades and Speculative Dynamics, a*vailable at SSRN: http://ssrn.com/abstract=898412.

[24] Rogoff, K., (1983), *Empirical Exchange Rate Models of the Seventies*, in: *Journal of International Economics*, Volume 14 (February), pp. 3–24.

[25] Ross, S., (1976), *The arbitrage theory of capital asset pricing*, in: *Journal of Economic Theory*, 13 (3): 341–360.

[26] White, S., Cutler, K.M., (2008), *Yen Falls as Asian Stock Gains Boost Confidence in Carry Trades*, 16th June 2008 Bloomberg News Article.

[27] Zimmermann, H. *et al.*, (2003), *Global Asset Allocation: New Methods and Applications (chapter 3)*, John Wiley and Sons inc.

7. Appendix – Data and Graphs

Correlation Matrices

Full Sample						
#	USD/JPY	NZD/JPY	AUD/JPY	EUR/JPY	GBP/JPY	EUR/CHF
Vix	0,355256366	0,174129	0,240750341	0,326564	0,392726518	0,32993936
Sp500	-0,300479864	0,115559	-0,053187031	-0,19866	-0,350007209	-0,2241003
Nikkei 225	-0,293748069	-0,39113	-0,640958967	-0,55027	-0,441874591	-0,5312755
Dax3030	-0,450682853	-0,13498	-0,295130437	-0,43626	-0,559866936	-0,4160669

Period 1						
#	USD/JPY	NZD/JPY	AUD/JPY	EUR/JPY	GBP/JPY	EUR/CHF
Vix	0,149672151	0,066999	0,082318735	0,173884	0,209176906	0,27332206
Sp500	-0,097482843	0,023719	0,034686339	-0,11193	-0,105107472	-0,17254
Nikkei 225	-0,042092326	-0,1417	-0,282688015	-0,20346	-0,161343266	-0,31571
Dax3030	-0,124319825	-0,14025	-0,114981163	-0,2116	-0,282909743	-0,3489963

Period 2						
#	USD/JPY	NZD/JPY	AUD/JPY	EUR/JPY	GBP/JPY	EUR/CHF
Vix	0,468596586	0,228258	0,317089617	0,410511	0,477242619	0,38218628
Sp500	-0,337082447	0,129386	-0,059964892	-0,20638	-0,36486902	-0,2287173
Nikkei 225	-0,350858839	-0,43308	-0,696440933	-0,60522	-0,467573305	-0,5659824
Dax3030	-0,527008823	-0,13142	-0,322330276	-0,47228	-0,588362883	-0,4266842

Factor Betas

Factor Betas estimates are tested against the null that the parameter is equal to 0. The intercept is excluded as it is insignificant for all the regressions (according to expectations). As for level of significance for the individual parameters, we have * for 1%, ** for 5 %, and *** for 10%; no asterisk indicate a failure to reject the null. The parameter VIX is naturally included three times for each of the three groups since it is included as a variable in three different regression contexts. Note that because of data retrieval issues, the sample size for the GBP/JPY is reduced to 710 observations for the full sample regression and 289 and 420 observations for the period 1 and period 2 regressions respectively.

Full Sample						
#	USD/JPY	NZD/JPY	AUD/JPY	EUR/JPY	GBP/JPY	EUR/CHF
Sp500	-0.074*	0.238*	0.070**	-0.031	-0.133*	-0.021**
Vix(SP500)	0.031*	0.071*	0.063*	0.043*	0.049*	0.018*
Nikkei 225	-0.073*	-0.304*	-0.488*	-0.241*	-0.193*	-0.100*
Vix(Nikkei 225)	0.032*	0.010	0.003	0.021*	0.045*	0.010*
Dax3030	-0.167*	-0.043	-0.214*	-0.216*	-0.337*	-0.085*
Vix(Dax3030)	0.014*	0.035*	0.022**	0.014**	0.014**	0.007*

Period 1						
#	USD/JPY	NZD/JPY	AUD/JPY	EUR/JPY	GBP/JPY	EUR/CHF
Sp500	-0.003	0.139**	0.154**	-0.002	0.030	0.001
Vix(SP500)	0.011**	0.018***	0.020*	0.013*	0.017*	0.009*
Nikkei 225	-0.002	-0.100*	-0.177*	-0.071*	-0.052**	-0.049*
Vix(Nikkei 225)	0.012*	0.005	0.002	0.010*	0.013*	0.007*
Dax3030	-0.022	-0.151**	-0.080	-0.083*	-0.132*	-0.064*
Vix(Dax3030)	0.009***	-0.007	0.001	0.004	0.003	0.002*

Period 2						
#	USD/JPY	NZD/JPY	AUD/JPY	EUR/JPY	GBP/JPY	EUR/CHF
Sp500	-0.054*	0.293*	0.116**	0.001	-0.104*	-0.011
Vix(SP500)	0.053*	0.117*	0.110*	0.074*	0.082*	0.028*
Nikkei 225	-0.068*	-0.335*	-0.541*	-0.261*	-0.185*	-0.106*
Vix(Nikkei 225)	0.052*	0.016	0.007	0.034*	0.070*	0.013*
Dax3030	-0.167*	0.021	-0.190*	-0.209*	-0.325*	-0.079*
Vix(Dax3030)	0.030*	0.072*	0.054*	0.034*	0.036*	0.015*

R-square values for the 18 regressions above:

Full Sample	USD/JPY	NZD/JPY	AUD/JPY	EUR/JPY	GBP/JPY	EUR/CHF
Sp500, VIX	0.148*	0.081*	0.063*	0.109*	0.189*	0.115*
Nikkei 225, VIX	0.158*	0.154*	0.411*	0.322*	0.256*	0.305*
Dax3030, VIX	0.213*	0.032*	0.093*	0.196*	0.318*	0.182*

Period 1	USD/JPY	NZD/JPY	AUD/JPY	EUR/JPY	GBP/JPY	EUR/CHF
Sp500, VIX	0.022**	0.012***	0.020**	0.030*	0.045*	0.075*
Nikkei 225, VIX	0.022**	0.021**	0.080*	0.058*	0.055*	0.140*
Dax3030, VIX	0.023*	0.021**	0.013***	0.047*	0.081*	0.124*

Period 2	USD/JPY	NZD/JPY	AUD/JPY	EUR/JPY	GBP/JPY	EUR/CHF
Sp500, VIX	0.233*	0.132*	0.114*	0.169*	0.248*	0.148*
Nikkei 225, VIX	0.248*	0.190*	0.486*	0.395*	0.313*	0.345*
Dax3030, VIX	0.308*	0.052*	0.125*	0.244*	0.365*	0.203*

Dummy Regressions

USD/JPY - SP500				
Parameters	Coefficients	SE	T-stat	P-value
Intercept	0.000	0.000	-0.178	0.859
Period Dummy	0.000	0.000	0.717	0.474
Change SP500	-0.003	0.065	-0.043	0.966
Change VIX (high)	0.011	0.007	1.667	0.096
Dummy*Vix	0.042	0.009	4.743	0.000
Dummy*Sp500	-0.051	0.067	-0.765	0.445

USD/JPY - Nikkei 225				
Parameters	Coefficients	SE	T-stat	P-value
Intercept	0.000	0.000	-0.182	0.855
Period Dummy	0.000	0.000	0.657	0.511
Change Nikkei 225	-0.002	0.031	-0.081	0.936
Change VIX (high)	0.012	0.005	2.123	0.034
Dummy*Vix	0.040	0.008	5.348	0.000
Dummy*Nikkei 225225	-0.065	0.034	-1.911	0.056

USD/JPY - Dax30				
Parameters	Coefficients	SE	T-stat	P-value
Intercept	0.000	0.000	-0.123	0.902
Period Dummy	0.000	0.000	0.455	0.649
Change DAX30	-0.022	0.046	-0.483	0.629
Change VIX (high)	0.009	0.007	1.359	0.175
Dummy*Vix	0.021	0.009	2.261	0.024
Dummy*Dax3030	-0.145	0.050	-2.937	0.003

NZD/JPY - SP500				
Parameters	Coefficients	SE	T-stat	P-value
Intercept	0.000	0.001	-0.495	0.621
Period Dummy	0.002	0.001	1.665	0.096
Change SP500	0.139	0.144	0.963	0.336
Change VIX (high)	0.018	0.015	1.192	0.234
Dummy*Vix	0.099	0.020	5.047	0.000
Dummy*Sp500	0.154	0.149	1.039	0.299

NZD/JPY - Nikkei 225				
Parameters	Coefficients	SE	T-stat	P-value
Intercept	0.000	0.001	-0.394	0.694
Period Dummy	0.001	0.001	0.857	0.392
Change Nikkei 225	-0.100	0.067	-1.491	0.136
Change VIX (high)	0.005	0.012	0.383	0.702
Dummy*Vix	0.011	0.016	0.693	0.489
Dummy*Nikkei 225225	-0.235	0.074	-3.182	0.002

NZD/JPY - DAX30				
Parameters	Coefficients	SE	T-stat	P-value
Intercept	0.000	0.001	-0.203	0.839
Period Dummy	0.001	0.001	1.121	0.262
Change DAX30	-0.151	0.109	-1.390	0.165
Change VIX (high)	-0.007	0.017	-0.403	0.687
Dummy*Vix	0.078	0.022	3.605	0.000
Dummy*Dax3030	0.172	0.118	1.457	0.145

AUD/JPY - SP500				
Parameters	Coefficients	SE	T-stat	P-value
Intercept	0.000	0.001	-0.638	0.524
Period Dummy	0.001	0.001	1.336	0.182
Change SP500	0.154	0.138	1.117	0.264
Change VIX (high)	0.020	0.015	1.340	0.181
Dummy*Vix	0.091	0.019	4.849	0.000
Dummy*Sp500	-0.038	0.142	-0.268	0.789

AUD/JPY - Nikkei 225				
Parameters	Coefficients	SE	T-stat	P-value
Intercept	0.000	0.001	-0.620	0.535
Period Dummy	0.000	0.001	0.456	0.648
Change Nikkei 225	-0.177	0.052	-3.397	0.001
Change VIX (high)	0.002	0.009	0.166	0.868
Dummy*Vix	0.006	0.013	0.439	0.661
Dummy*Nikkei 225225	-0.364	0.058	-6.328	0.000

AUD/JPY - DAX30				
Parameters	Coefficients	SE	T-stat	P-value
Intercept	0.000	0.001	-0.428	0.669
Period Dummy	0.001	0.001	0.861	0.389
Change DAX30	-0.080	0.100	-0.802	0.423
Change VIX (high)	0.001	0.015	0.065	0.948
Dummy*Vix	0.053	0.020	2.660	0.008
Dummy*Dax3030	-0.110	0.109	-1.016	0.310

EUR/JPY - SP500				
Parameters	Coefficients	SE	T-stat	P-value
Intercept	0.000	0.000	-0.872	0.383
Period Dummy	0.001	0.001	1.216	0.224
Change SP500	-0.002	0.085	-0.018	0.986
Change VIX (high)	0.013	0.009	1.408	0.160
Dummy*Vix	0.061	0.011	5.345	0.000
Dummy*Sp500	0.002	0.087	0.024	0.981

EUR/JPY - Nikkei 225				
Parameters	Coefficients	SE	T-stat	P-value
Intercept	0.000	0.000	-0.970	0.332
Period Dummy	0.000	0.001	0.707	0.480
Change Nikkei 225	-0.071	0.035	-2.011	0.045
Change VIX (high)	0.010	0.006	1.550	0.122
Dummy*Vix	0.024	0.009	2.773	0.006
Dummy*Nikkei 225225	-0.189	0.039	-4.857	0.000

EUR/JPY - DAX30				
Parameters	Coefficients	SE	T-stat	P-value
Intercept	0.000	0.000	-0.718	0.473
Period Dummy	0.000	0.001	0.766	0.444
Change DAX30	-0.083	0.059	-1.410	0.159
Change VIX (high)	0.004	0.009	0.466	0.641
Dummy*Vix	0.029	0.012	2.485	0.013
Dummy*Dax3030	-0.125	0.064	-1.955	0.051

EUR/CHF - SP500				
Parameters	Coefficients	SE	T-stat	P-value
Intercept	0.000	0.000	-0.793	0.428
Period Dummy	0.000	0.000	1.045	0.296
Change SP500	0.001	0.037	0.018	0.985
Change VIX (high)	0.009	0.004	2.262	0.024
Dummy*Vix	0.019	0.005	3.789	0.000
Dummy*Sp500	-0.012	0.038	-0.313	0.754

EUR/CHF - Nikkei 225				
Parameters	Coefficients	SE	T-stat	P-value
Intercept	0.000	0.000	-0.839	0.401
Period Dummy	0.000	0.000	0.585	0.559
Change Nikkei 225	-0.049	0.016	-3.078	0.002
Change VIX (high)	0.007	0.003	2.424	0.016
Dummy*Vix	0.007	0.004	1.744	0.082
Dummy*Nikkei 225225	-0.057	0.017	-3.289	0.001

EUR/CHF - DAX30				
Parameters	Coefficients	SE	T-stat	P-value
Intercept	0.000	0.000	-0.484	0.629
Period Dummy	0.000	0.000	0.577	0.564
Change DAX30	-0.064	0.026	-2.459	0.014
Change VIX (high)	0.002	0.004	0.568	0.571
Dummy*Vix	0.012	0.005	2.367	0.018
Dummy*Dax3030	-0.015	0.028	-0.511	0.609

GBP/JPY - SP500				
Parameters	Coefficients	SE	T-stat	P-value
Intercept	0.000	0.001	-0.813	0.417
Period Dummy	0.001	0.001	1.830	0.068
Change SP500	0.030	0.110	0.274	0.784
Change VIX (high)	0.017	0.011	1.495	0.135
Dummy*Vix	0.065	0.014	4.605	0.000
Dummy*Sp500	-0.134	0.113	-1.190	0.234

GBP/JPY - Nikkei 225				
Parameters	Coefficients	SE	T-stat	P-value
Intercept	0.000	0.001	-0.787	0.431
Period Dummy	0.001	0.001	1.680	0.093
Change Nikkei 225	-0.052	0.057	-0.914	0.361
Change VIX (high)	0.013	0.009	1.451	0.147
Dummy*Vix	0.058	0.012	4.969	0.000
Dummy*Nikkei 225225	-0.132	0.061	-2.176	0.030

GBP/JPY - DAX30				
Parameters	Coefficients	SE	T-stat	P-value
Intercept	0.000	0.001	-0.566	0.572
Period Dummy	0.001	0.001	1.430	0.153
Change DAX30	-0.132	0.078	-1.689	0.092
Change VIX (high)	0.003	0.011	0.234	0.815
Dummy*Vix	0.034	0.014	2.429	0.015
Dummy*Dax3030	-0.193	0.083	-2.320	0.021

Journal of Applied Economic Sciences

PUBLIC SPENDING AND REGIONAL CONVERGENCE IN ITALY

Daniele **VITTORIO**
University *Magna Graecia* of Catanzaro, **Italy**
v.daniele@unicz.it

Abstract:

This paper examines the relationship between public spending and regional growth in Italy in the period 1996-2007. The main results suggest that, in the period examined, a phase of σ and β-convergence in productivity took place. Despite the profound regional disparities existing in Italy, the allocation of public spending calculated in per capita terms has favoured the most developed regions. When the entire sample of twenty regions is considered, no correlation between capital expenditure and productive growth is found. When the sample is split into Northern and Southern regions, differences in the links between public spending and productivity growth can be noted, however: only in the first group of regions, in fact, a significant, positive relationship between expenditure for development and growth can be found.

Keywords: Italy, regional convergence, development policy, public spending.

JEL Classification: 018, R 38, R 58.

1. Introduction

The aim of this paper is to examine the relationship between public spending and regional growth in Italy in the period 1996-2007. The empirical analysis, based on a standard framework of conditional convergence, uses data for different categories of public expenditure: capital, current and for development. The paper is structured as follows. The second section introduces a short literature review on the role of recent regional development policies in Italy. The third section takes a look at the regional convergence in per capita GDP and in productivity and, through a panel dataset, investigates the relationship between public spending and growth. Some conclusive remarks follow.

The main results suggest that, in the period examined, a phase of absolute convergence in productivity took place. Despite the profound regional disparities existing in Italy, the allocation of public spending in per capita terms has favoured the most developed regions. When the entire sample of twenty regions is considered, no correlation between capital expenditure and productivity growth is found. When the sample is split into Northern and Southern regions, differences can be noted, however: only in the first group of regions, in fact, a significant, positive relationship between expenditure for development and growth can be found.

2. Recent policies and regional growth

During the 1990s, in Italy, the regional development policy underwent a profound change. In 1992 ceases the "Extraordinary Interventions" for the development of the *Mezzogiorno*[1] that, for over forty years, constituted the framework within which the measures for less developed areas where implemented. In the second half of the 1990s, a new phase of regional policy replaced the previous one. This so-called "new regional development policy" was formulated in the context of the EU's policy scheme and characterised by a bottom-up approach and a multi-level governance system [Loddo, (2004); Chiri (2006)].

The policy measures adopted are diverse. In particular, some measures are included in "Negotiation planning", such as Territorial Pacts for local development, Planning contracts for industrial investments and Area contracts for the implementation of new business initiatives in some circumscribed areas, notably those facing employment crises. Other measures added to these: in particular, incentives and subsidies for investments (provided through the financial incentive scheme in which the main instrument is the law no. 488/92) and the initiatives included in the EU regional policy for the less developed areas.

At a distance of more then ten years from its commencement, the evaluation of the results obtained by the new development policy has brought different, often opposing, interpretations [Rossi,

[1] The well-known term "Mezzogiorno" is referred to the eight Southern Italian regions, historically less developed.

(2005); Viesti, (2009)]. According to some scholars [Rossi, (2004); Atella, (2004)], this policy would have substantially failed in its objective to reduce the gap between the North and the South appearing as "staggeringly wasteful" for its inefficiency in the use of public resources. According to others, however, this policy would have gained valuable results; nevertheless some limitations primarily derived from the lack of resources actually available for the less developed regions [Viesti, (2009)]. In a less sophisticated version, the argument according to which the Southern regions received and substantially wasted a huge amount of financial resources finds a wide consensus in Italian public opinion and among politicians. Very frequently, however, these arguments are not supported by data or any empirical evidence.

The effectiveness of public policy to promote regional economic convergence is a theme extensively examined in the literature. For instance, several studies have been devoted to investigate have evaluated the impact of EU regional policies – in particular structural funds — on growth. The results reached are sometimes controversial. For instance, some authors find that the European structural and cohesion funds have had a positive impact on regional economic growth [Cappelen *et al*., (2003); Beugelsdijk, Eijffinger, (2005); Checherita, (2009)] while others reached the conclusion that EU regional policies principally serve for redistribution purposes, motivated by political reasons, but they have no effect in fostering economic growth [Boldrin and Canova, (2001); Dall'Erba *et al.*, (2007)].

Relatively few studies have examined the case of Italy. For instance, Coppola and Destefanis (2007) have investigated the effects of EU structural funds in the period 1989-2003 for the main economic sectors, finding a positive but weak fund effect on capital accumulation and total factor productivity. Percoco (1999) studied the impact of EU regional policy in the programming period 1994-1999, showing a high variation in the rate of growth induced by structural funds. This study suggests that the effects have been related with the efficiency in the allocation of financial resources from regional administrations. Aiello and Pupo (2009) estimated the effects of structural funds on regional economic growth in the period 1980-2007. The analysis is carried out in a model in which funds are considered as an explanatory variable in the convergence equation. The authors found the effects of funds were mainly redistributive, and they did not contribute to the factors that influence long run regional economic growth.

With regard to the policy implemented in the period 1996-2007 — the "new development policy" — studies have evaluated the results obtained from some instruments, such as the Territorial pacts [Accetturo e De Blasio, (2007)], the Programming contracts for industrial development [Giunta e Florio, (2002); Bianchi, (2007)] or from the incentives to investments [Bronzini e De Blasio, (2006)]. The results emerging from these studies are not univocal. If, as in the case of Territorial pacts for local development, the overall results appear very modest, for other kinds of intervention, such as the Programming contracts, the evaluations show a different situation, in which there have been some successful results. On the basis of these studies it appears very hard to take conclusive stock of the new development policy. In the light of this debate, the analysis concerning the evolution of regional disparities and the regional distribution of public expenditure can offer some policy implications.

3. The empirical analysis

3.1. Methodology and data

The relationship between public spending and regional economic growth is analysed in the standard framework of conditional convergence [Barro and Sala-i-Martin, (1991)]. The estimations are based on the equation 1:

$$\left(\ln y_{i,T} - \ln y_{i,T}\right)\frac{1}{T} = \alpha_i + \beta_1 y_{i,t} + \beta_2 PS_{i,t} + \beta_3 X_{i,t-1} + \varepsilon_{i,t} \quad (1)$$

with: $\varepsilon_{i,t} = \lambda_i + u_{i,t}$

in which the dependent variable is the rate of growth of per capita (or per worker) GDP, $y_{i,t}$ the log of the same variable in the region *i* at time *t*, $PS_{i,t}$ is the share of public spending on GDP, and $\mathbf{X}_{i,t-1}$ a

set of variables that captures the "fundamentals" of the economy, that is those characteristics that have a permanent effect on its growth rate.

Data consist in a balanced panel and a fixed effect (FE) model is used for estimations. The advantages of using a panel data model are several. Studies show in fact, how the cross-section approach leads to a downward bias of the convergence coefficient. The reason is that the steady state of an economy is affected by a number of factors that cross-section analysis tends to neglect, suffering an omitted-variable bias. These regional-specific unobservable factors can be modelled through panel data techniques [Tondl, (2001); Islam, (2003); Durlauf *et al.*, (2005)].

Data on public expenditure refer to the period 1996-2007 — for which homogeneous time-series are available — and derived from Regional Public Accounts (RPA), a detailed database, published by the Italian Ministry for the Economy, which measures public financial flows at the territorial level. The RPA database can be used to analyse different two reference universes: general government and the "public sector" which includes general government plus enterprises subject to the direct or indirect control of public entities. For each macro-category of expenditure series for individual spending chapters are available: for instance, the expenditure for development contains different capital expenditure categories. For the high degree of disaggregation, RPA accounts constitute a complete dataset for measuring and evaluating public policies at the regional level [De Luca *et al.*, (2005)].

Since the variables potentially correlated with regional economic growth are numerous, to select those to be included as control, the results of some recent studies that employ the Bayesian approach to check the robustness of the explanatory variables used in the cross-section studies on economic growth [Doppelhofer *et al*, (2000), Fernández *et al.*, (2001)] were considered. The Bayesian technique was also recently used in regional studies. For example, examining the Spanish case, León-Gonzàlez and Montolio (2004), found that the initial level of per capita GDP and some types of private and public investment are strongly related to growth. Other variables, such as human capital proxies and the sectoral composition of production — measured by the relative share of agriculture and industry product — have a relatively high inclusion probability. On the basis of these studies, the model estimated included the subsequent control variables: the share of gross fixed investment on regional GDP (*Investment*); a proxy of human capital, given by the average number of years of schooling for each region (*School*); the shares of agricultural (*Agriculture*) and industrial (*Industry*) employment in total regional employment. Data on regional GDP, investment and sectoral employment are derived from the Regional accounts database published by Istat (2005; 2007), while the years of schooling are calculated on the basis of estimation contained in Ciccone (2004).

3.2. Evidence on convergence

The literature on convergence among the Italian regions is very extensive [Paci e Pigliaru, (1996); Terrasi, (1999); Ciriaci, (2001); Vamvakidis, (2003); Maffezzoli, (2006); Magrini, (2007); Daniele, (2008)]. Even if the methods and the period examined in these studies are diverse, main findings show how a process of absolute β-convergence in per capita GDP took place during the period 1960-1975, while in the subsequent period regional disparities remained stable or, in certain phases, increased. Only in recent years, notably during the second half of the 1990s, a process of convergence in productivity and, to a lesser extent, in per capita GDP has taken place.

As a first step in the empirical analysis, convergence across GDP per capita and per worker is examined. The examination covers the period 1980-2007, in order to offer a relative long term picture of the evolution of regional disparities.

Figure 1 illustrates the correlation between the level of development of the Italian regions relative to the national average in 1980 and 2007. The picture displays a substantial stability in the distribution of per capita GDP (r^2 = 0, 90) even if it is possible to note some changes in the relative position of some regions, in particular those in the Mezzogiorno.

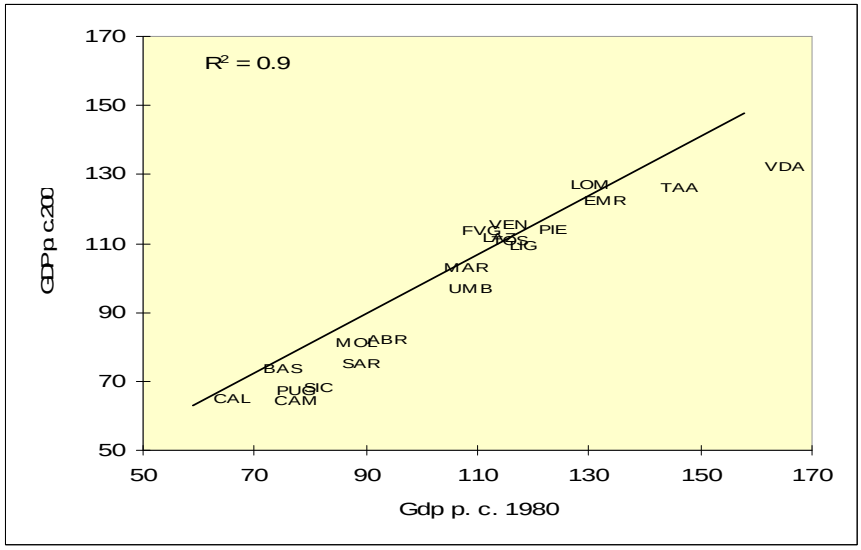


Figure 1. Correlation in regional per capita GDP 1980 and 2007 (Italy = 100)

Source: Calculations on Istat, *Regional Economic Accounts*.

The process of σ-convergence is illustrated in Figure 2 that presents the dynamics over time of the dispersion (variance) of regional per capita GDP. The dispersion increased during the period 1980-95, in which a phase of σ-divergence occurred, while a slight reduction can be observed for the subsequent years. Overall, the degree of dispersion remains roughly the same. To examine the proximate forces at the basis of this dynamics, it is possible to decompose the variance of per capita GDP according to the following identity:

$$\operatorname{var}\ln(Y/P) = \operatorname{var}\ln(Y/E) + \operatorname{var}\ln(E/P) + 2\operatorname{cov}[\ln(Y/E), \ln(E/P)] \qquad (2)$$

in which Y is the GDP, P population and E the employment. Figure 3 shows the variance of productivity and unemployment rate. It is easy to observe how the phase of σ-divergence in GDP per capita is mainly explained by the increase in the dispersion of regional employment rates. The variance of productivity exhibits a decreasing trend between 1980 to 1999, remaining roughly stable in the subsequent period. This description suggests that the Italian regions converged mainly in terms of productivity, while the convergence in per capita GDP was a weaker process.

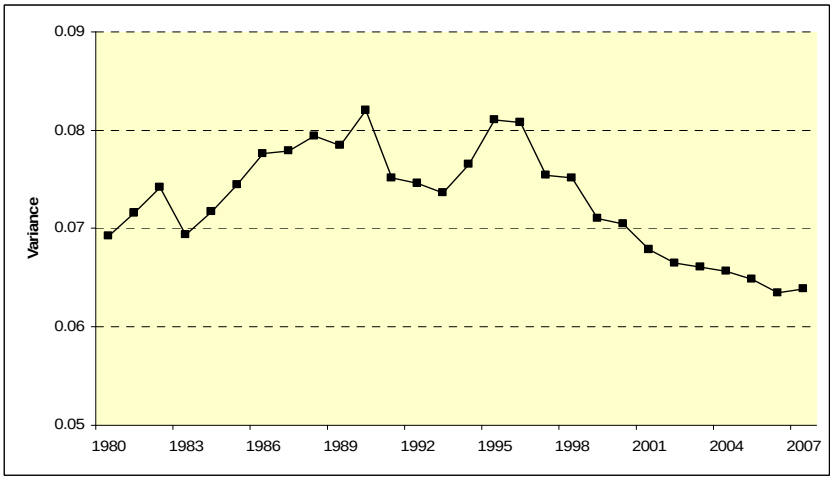


Figure 2. Variance of per capita GDP among Italian regions, 1980-2007

Source: Calculations on Istat, *Regional Economic Accounts*.

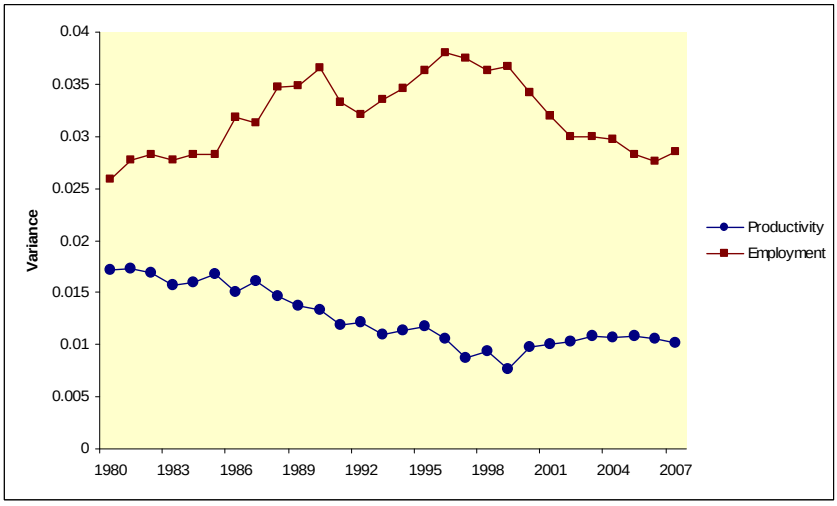


Figure 3. Variance of GDP per worker and of employment rates 1980-2007.

Source: Calculations on Istat, *Regional Economic Accounts*.

Table 1 report the results of regressions of the growth rates on the initial levels of GDP per capita and labour productivity. Both in pooled OLS and fixed effects (FE) estimations the β coefficient is negative and significant, but the value of r^2 indicates that the model is not robust for per capita GDP, while it has a sufficient explanatory power for productivity. The values of the F statistic and of the Hausman's test show, in fact, how the FE model is consistent, while the OLS is not adequate for the nature of the data.

Table 1. Absolute β-convergence

	GDP per capita growth		Productivity growth	
	Pooled OLS	FE	Pooled OLS	FE
Const	0.0779**	0.2812**	0.3039**	0,4441**
	(3.822)	(7.467)	(11.12)	(13,67)
Log *y*	-0.0065**	-0.0278**	-0.0277**	-0,0411**
	(-3.114)	(-7.061)	(-10.65)	(-13,27)
F stat.	*1.10 [0.36]*		*0.61 [0.88]*	
Hausman		*12.0 [0.00]*		*6.44 [0.01]*
N	*100*	*100*	*100*	*100*
Adj. R^2	*0.03*	*0.05*	*0.20*	*0,13*
lnL	*326.8*	*338.6*	*329.3*	*336,2*

5-years annual growth rates. T-statistics in parentheses. * indicates significance at the 10 percent level; ** indicates significance at the 5 percent level. For *F* and Hausman's tests, p-value are reported in squared parentheses; a low p-value counts against the null hypothesis that the random effects model is consistent, in favour of the fixed effects model.

3.3. Public spending and regional growth

Before analysing the role of public spending, territorial distribution is examined. Data show that the regions of the Centre-North, where 64% of Italian population live, received over 71% of total expenditure, while the South just 28,5% (Table 2). The disaggregation of expenditure in macro-categories confirms that the Centre-North area received the largest part of spending flows (Table 3), both current and for development expenditure in particular. Significant differences can be observed even when the expenditure is calculated in per capita terms. An inhabitant of the Mezzogiorno area

received, on average, 9,103 euro of current and 1,666 euro of capital expenditure, while an inhabitant of the Centre-North area, 13,078 and 2,022 euro respectively (Table 4). In per capita terms, the distribution of public spending has, clearly, privileged the more developed part of Italy.

Table 2. Centre-North and Mezzogiorno: main indicators and public spending (Italy = 100)

	Population	GDP	GDP per capita	Total public expenditure
Centre-North	64,1	75,9	117,1	71,5
Mezzogiorno	35,9	24,1	67,8	28,5

Source: Calculations on Istat and RPA Database.

Table 3. Territorial distribution of public expenditure, 1996-2007 (%)

Categories of expenditure	Centre-North	South
Capital	68,5	31,5
Current	72,0	28,0
Investment	71,2	28,8
Development	66,2	33,8

Calculated on cumulated values in current prices. Data are referred to the "enlarged public sector".
Source: Calculations on RPA Database.

Table 4. Public expenditure, average 1996-2007

	As a % of GDP		Per capita	
	Centro-Nord	Mezzogiorno	Centro-Nord	Mezzogiorno
Correnti	41.0	50.0	13,078	9,103
Capitale	4,5	7.5	2,022	1,666
Investimenti	2.1	3.2	958	693
Sviluppo	3.4	6.5	1,227	1,121

In current prices. Data are referred to the "enlarged public sector".
Source: Calculations on RPA Database.

Figure 4 illustrates the relationship between public expenditure for development and the per capita GDP in the twenty Italian regions. It is easy to see that there is no correlation between the two variables: in the considered decade three Northern Italian regions with "Special statute", (Valle d'Aosta, Trentino Alto Adige and Friuli Venezia Giulia) received the highest share of spending flows; other regions, such as Sicily, Campania, Apulia and Calabria (the less developed) the lowest share. The regional distribution of public expenditure offers some important elements for consideration of the implementation of regional policy. In a nation in which profound regional development disparities exist, the distribution of financial resources aimed at territorial re-balancing should privilege the areas lagging behind. It does not seem to be the case of Italy, despite the fact that the programming documents and Development plans established that 45% of the public spending total should have gone to the Mezzogiorno area [Viesti (2009)].

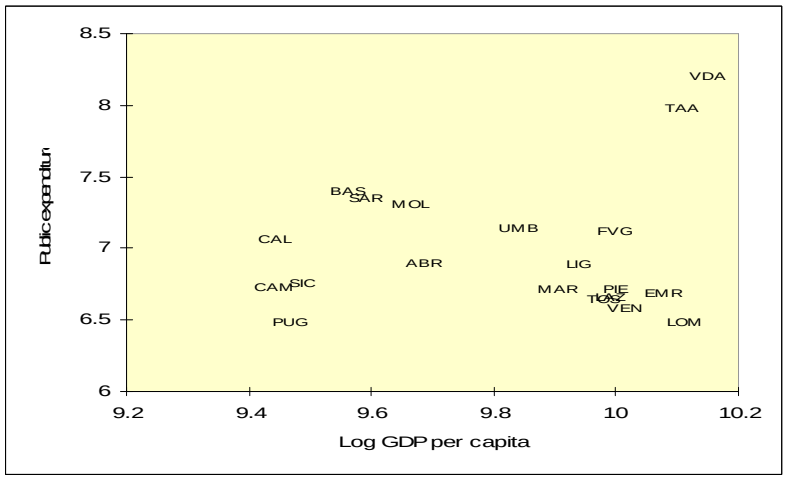


Figure 4. Public expenditure for development and per capita GDP – average 1996-2007. Source: Calculations on Istat and RPA Database.

Table 5 reports the results of FE estimations, in which the dependent variable is the rate of growth of labour productivity and different categories of expenditure are considered. The results suggest the existence of conditional convergence. Both the capital and development expenditure categories seem not to have influenced regional growth, while the current expenditure exhibits a positive and significant correlation. Regions with a large agricultural employment share have grown faster: this result is consistent with the hypothesis of conditional convergence, given that in the less developed regions the share of the agricultural sector is, on average, larger.

Table 5. Public spending and productivity growth: all regions

	(1)	(2)	(3)	(4)	(5)	(6)
Const	3.382**	3.013**	3.374**	3.989**	3.614**	3.981**
	(8.01)	(8.70)	(7.91)	(8.52)	(11.80)	(8.64)
Log *y*	-0.3191**	-0.2901**	-0.3183**	-0.3863**	-0.3506**	-0.3858**
	(-8.12)	(-8.74)	(-8.01)	(-8.90)	(-12.24)	(-9.01)
Development exp.	0.1305			0.1538		
	(0.91)			(0.87)		
Current exp.		0.1283**			0.0915**	
		(4.64)			(2.50)	
Capital exp.			0.1421			0.1718
			(1.00)			(1.01)
Investment				-0.2346**	-0.1290**	-0.2386**
				(-2.20)	(-1.99)	(-2.32)
Agricolture				0.0915**	0.0889**	0.0913**
				(2.64)	(2.91)	(2.66)
Industry				-0.0836	-0.1008	-0.0836
				(-1.08)	(-1.32)	(-1.11)
School				0.0913**	0.0684**	0.0919**
				(6.13)	(4.61)	(6.30)
Hausman	*54.9*	*68.7*	*55.3*	*81.3*	*87.6*	*82.2*
	[0.00]	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*
N	*100*	*100*	*100*	*100*	*100*	*100*
Adj. R^2	*0.30*	*0.37*	*0.30*	*0.39*	*0.41*	*0.39*
lnL	*330.5*	*335.7*	*330.6*	*340.1*	*341.7*	*340.4*

Two-years annual growth. Fixed effect estimates. T-statistics in parentheses. * indicates significance at the 10 percent level; ** indicates significance at the 5 percent level. For Hausman's tests p-values are reported in squared parentheses; a low p-value counts against the null hypothesis that the random effects model is consistent, in favour of the fixed effects model.

Different results are obtained when the sample of twenty regions is split into groups. Table 6 reports the results for the twelve Centre-North regions. It is possible to observe a positive and significant correlation between the three categories of expenditure and growth, even when the control variables are included. In the Southern regions the results presented in Table 7 show quite a different situation: in fact while in the basic specifications it is possible to find a significant relationship between public spending and growth (with development and capital expenditure with negative coefficients), when controls are inserted the correlations are weak. In all estimations, the Hausman's test confirms that FE model is consistent with the nature of dataset.

Table 6. Public spending and productivity growth 1996-2006: Centre-North regions

	(1)	(2)	(3)	(4)	(5)	(6)
Const	2.830**	2.309**	2.804**	2.961**	3.143**	2.968**
	(8.79)	(5.30)	(8.90)	(6.32)	(9.70)	(6.45)
Log *y*	-0.2669**	-0.2211**	-0.2644**	-0.2760**	-0.3067**	-0.2766**
	(-8.70)	(-5.41)	(-8.80)	(-5.87)	(-9.74)	(-5.97)
Development exp.	0.3709**			0.3873**		
	(3.18)			(2.22)		
Current exp.		0.1097**			0.1079**	
		(2.39)			(2.27)	
Capital exp.			0.3843**			0.3924**
			(3.18)			(2.16)
Investment				-0.2882*	-0.2702*	-0.2764*
				(-1.83)	(-1.98)	(-1.81)
Agricolture				-0.1374	-0.0644	-0.1332
				(-0.88)	(-0.55)	(-0.85)
Industry				-0.0786	-0.0617	-0.0842
				(-0.74)	(-0.60)	(-0.78)
School				0.0437**	0.0853**	0.0446**
				(2.58)	(5.94)	(2.67)
Hausman	*36.3*	*19.1*	*36.2*	*37.6*	*27.3*	*37.2*
	[0.00]	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*
n	*60*	*60*	*60*	*60*	*60*	*60*
Adj. R^2	*0.31*	*0.20*	*0.31*	*0.30*	*0.22*	*0.30*
lnL	*210.5*	*206.1*	*210.5*	*213.0*	*209.7*	*212.8*

Two-years annual growth. Fixed effect estimates. T-statistics in parentheses. * indicates significance at the 10 percent level; ** indicates significance at the 5 percent level. For Hausman's tests p-values are reported in squared parentheses; a low p-value counts against the null hypothesis that the random effects model is consistent, in favour of the fixed effects model.

The results obtained can be explained in the light of previous literature on the effects of regional policy in Italy. Firstly, *caeteris paribus*, the impact of financial resources on productivity growth depends not only on their quantity, but also on the efficiency of their allocation. In general, this concerns different resources, both those aimed at physical capital accumulation and those devoted to human capital formation.

Table 7. Public spending and productivity growth 1996-2006: Southern regions

	(1)	(2)	(3)	(4)	(5)	(6)
const	4.145**	3.757**	4.161**	4.731**	4.058**	4.703**
	(6.77)	(15.49)	(6.75)	(6.75)	(20.21)	(6.95)
Log *y*	-0.3910**	-0.3647**	-0.3928**	-0.4503**	-0.3919**	-0.4498**
	(-6.80)	(-16.05)	(-6.76)	(-8.88)	(-22.34)	(-9.04)
Development exp.	-0.3181**			-0.2039		
	(-3.00)			(-0.64)		

Current exp.		0.1401**			0.0965	
		(5.24)			(1.68)	
Capital exp.			-0.2884**			-0.1416
			(-3.61)			(-0.49)
Investment				-0.1154	-0.1123*	-0.1341
				(-0.84)	(-1.94)	(-1.07)
Agricolture				0.0897**	0.0761**	0.093**
				(3.62)	(2.99)	(3.53)
Industry				-0.2413	-0.1544	-0.2273
				(-1.20)	(-1.37)	(-1.20)
School				0.0938**	0.0569**	0.0976**
				(4.22)	(2.05)	(4.57)
Hausman	*22.5*	*38.5*	*21.6*	*40.0*	*43.9*	*36.2*
	[0.00]	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*	*[0.00]*
n	*40*	*40*	*40*	*40*	*40*	*40*
Adj. R^2	*0.45*	*0.51*	*0.44*	*0.48*	*0.49*	*0.47*
lnL	*129.9*	*132.2*	*129.5*	*133.8*	*134.2*	*133.5*

Two-years annual growth. Fixed effect estimates. T-statistics in parentheses. * indicates significance at the 10 percent level; ** indicates significance at the 5 percent level. For Hausman's tests p-values are reported in squared parentheses; a low p-value counts against the null hypothesis that the random effects model is consistent, in favour of the fixed effects model.

In the case of Italy, some studies proved that the intervention policies for the development of the Mezzogiorno area had indirect effects – such as the reinforcement of rent-seeking behaviour - that notably limited the policies' effectiveness [Del Monte e Giannola, (1997)]. Differences between North and South were also found for human capital allocation among sectors, with a large prevalence of less productive activities (such as those in the public sector) in the Mezzogiorno [Di Liberto (2001)]. Secondly, great waste and inefficiency in the use of financial resources and in public investments are documented [Rossi, (2005)]. Finally, as shown by a study of Marrocu and Paci (2005) different impacts of the productive inputs exist between the North and the South. These authors estimated a production function for Italy for the period 1996-2003, showing that the stock of public capital exhibits different degrees of elasticity for the two macro-areas of the Country. More specifically, the economic infrastructures, which account for the largest proportion of the public stock, were much more productive in the South compared to the rest of the Country. In contrast, all the other types of public intervention showed a negative impact in the South and a low impact in the Centre-North. The cited analysis leads to the conclusion that in the Mezzogiorno the regional and local administrations are much less efficient in delivering public funds than in the rest of Italy.

4. Conclusive remarks

This paper has examined the evolution of regional disparities in Italy and the relationship between public spending and productivity growth. The results obtained can be summarised as follows. During the period 1980-2007, regions converged significantly in productivity levels but much less so in per capita GDP. Despite the profound regional disparities existing among the Italian regions, the distribution of public spending has not privileged the less developed areas. In per capita terms, in fact, the Centre-North received a higher public expenditure than the South. Differences in advantages for the less developed regions can be found if the expenditure is considered as a share of GDP.

The regression analysis shows how the impact of public spending on growth differs, dependent upon the regions and the categories of expenditure. When the entire sample is considered, a positive relationship between current expenditure and productivity growth is found. If the sample is split into Northern and Southern regions, the estimates lead to different results. In the first group, composed of the most developed Italian regions, results show a positive and significant correlation between capital public expenditure and growth, while in the less developed Southern regions productivity growth

results as only weakly linked to current expenditure. Despite the fact that the analysis refers to a relatively short period of time, it is coherent with the findings of previous studies and tends to suggest the existence of differences, not only in the amount of financial resources devoted to the North and the South of Italy, but also in the efficiency of their allocation.

5. References

[1] Aiello F., Pupo, V., (2009), *The Impact of Structural Funds in Italy*, MPRA, Paper nr. 14429, http://mpra.ub.uni-muenchen.de/14429.

[2] Atella, V., (2005), *«Buche» di keynesiana memoria,* in: N. Rossi, *Mediterraneo del Nord*. Laterza, Roma-Bari, pp. 137-152.

[3] Beugelddijk, M., Eijffinger, S.C.W., (2005), *The Effectiveness of Structural Policy in the European Union: An Empirical Analysis for the EU-15 in 1995-2001*, in: *Journal of Common Market Studies*, Vol. 43, N. 1, pp. 37-51.

[4] Boldrin, M., Canova, F., (2001), *Inequality and convergence in Europe's regions: reconsidering European regional policies*, in: *Economic Policy*, Vol. 32, pp. 207-253

[5] Bronzini, R., De Blasio, G., (2006), *Una valutazione degli incentivi pubblici agli investimenti*, in: *Rivista Italiana degli Economisti*, N. 3, pp. 331-362.

[6] Cappelen, A., Castellacci, F., Fagerberg, J., Verspagen, B., (2003), *Regional Support, Growth and Convergence in the EU*, in: *Journal of Common Market Studies*, Vol. 41, N. 4, pp. 621-44.

[7] Checherita, C., Nickel, C., Rother, P., (2009), *The Role of Fiscal Transfers for Regional Economic Convergence in Europe*, European Central Bank, Working Paper Series, nr. 1029, March.

[8] Chiri, S., (2006), *Le politiche per il Mezzogiorno dopo il 1992,* in: L. Cannari e F. Panetta (eds.), *Il sistema finanziario e il Mezzogiorno*, Cacucci, Bari.

[9] Ciccone, A., (2004), *Human Capital as a Factor of Growth and Employment at the Regional level: The Case of Italy*, Report for the European Commission, DG for Employment and Social Affairs, Brussels.

[10] Ciriaci, D., (2001), *Convergenza e dualismo: Nord e Sud tra il 1970 e il 1999*, in: *Rivista Economica del Mezzogiorno*, nr. 4, pp. 763-803.

[11] Coppola, G., Destefanis, S., (2007), *Fondi strutturali, produttività e occupazione. Uno studio sulle regioni italiane*, in: *Rivista di economia e statistica del territorio*, nr. 2, pp. 191-212.

[12] Daniele, V., (2008), *Divari di sviluppo e crescita regionale in Italia 1980-2004. Alcuni aspetti empirici*, in: *Rassegna Economica*, nr. 1, pp. 259-286.

[13] Dall'Erba, S., Guillain, R., Le Gallo, J., (2007), *Impact of Structural Funds on Regional Growth: How to Reconsider a 7 Year-Old Black-Box*, The University of Arizona, Discussion Papers, GRD 06-07.

[14] De Luca, S., Nusperli, F., Sferrazzo, A., Tancredi, A., Volpe, M., (2007), *Measuring the Results of Public Intervention: Data for Evaluating the Territorial Impact of Policy*, in: *Scienze Regionali, Italian journal of regional science*, Vol. 2, anno 2007.

[15] Del Monte, A., Giannola, A., (1997), *Istituzioni economiche e Mezzogiorno. Analisi delle politiche di sviluppo*, NIS, Roma.

[16] Di Liberto, A., (2001), *Stock di capitale umano e crescita nelle regioni italiane: un approccio panel*, in: *Politica Economica*, Vol.17, nr.2, pp.159-184.

[17] Doppelhofer, G., Miller, R.I., Sala-i-Martin, X., (2000), *Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach*, NBER Working Paper 7750.

[18] Durlauf, S.N., Johnson, P.A., Temple, J.R.W., (2005), *Growth Econometrics*, in P. Aghion e S. Durlauf (eds.), *Handbook of Economic Growth*, Vol. 1, Elsevier, North-Holland.

[19] Fernández, C., Ley E., Steel, M.F.J., (2001), *Model Uncertainty in Cross-Country Growth Regressions*, in: *Journal of Applied Econometrics*, Vol. 16, nr. 5, pp. 563-576.

[20] Giunta, A., Florio, M., (2002), *L'esperienza dei contratti di programma: una valutazione a metà percorso*, in: *Rivista di Economia e politica industriale*, Vol. 23, nr. 2, pp. 319-358.

[21] Islam, N., (2003), *What have we learnt from the Convergence Debate?* in: *Journal of Economic Surveys*, Vol. 17, pp. 309-6.

[22] León-Gonzàlez, R., Montolio, D., (2004), *Growth, convergence and public investment. A Bayesian model averaging approach*, in: *Applied Economics*, Vol. 36, nr. 17, pp. 1925-1936.

[23] Loddo S., (2006), *Structural Funds and Regional Convergence in Italy*, CRENoS, Working paper, 2006/03.

[24] Maffezzoli, M., (2006), *Convergence Across Italian Regions and the Role of Technological Catch-Up*, in: *Topics in Macroeconomics*, 6(1), Article 15.

[25] Marrocu, E., Paci, R., (2006), *The effects of public capital on the productivity of the Italian regions*, Crenos, Working paper 2006/13.

[26] Magrini, S., (2007), *Analysing Convergence through the Distribution Dynamics Approach: Why and How?*, 2007, Università Ca' Foscari, DSE Working Paper 13.

[27] Paci, R., Pigliaru, F., (1996), *β-convergence and/or Structural Change? Evidence from the Italian Regions*, CRENoS, Working paper 96/07.

[28] Percoco, M., (2005), *The Impact of Structural Funds on Italian Mezzogiorno*, in: *Région et Développement*, nr. 21, pp. 141-153.

[29] Perugini, C., Signorelli, M., (2005), *Employment Differences, Convergences and Similarities in Italian Provinces*, Quaderni del Dipartimento di Economia, Finanza e Statistica 07/2005, Università di Perugia, Dipartimento Economia, Finanza e Statistica.

[30] Rossi, N., (2005), *Mediterraneo del nord. Un'altra idea del Mezzogiorno*, Laterza, Roma-Bari.

[31] Sala-i-Martin, X., (1996), *Regional Cohesion: Evidence and Theories of Regional Growth and Convergence*, in: European Economic Review, Vol. 40, nr. 6, pp. 1325-1352.

[32] Terrasi, M., (1999), *Convergence and divergence across Italian regions*, in: *The Annals of Regional Sciences*, Vol. 33, nr. 4, pp. 491-510.

[33] Tondl, G., (2001), *Convergence after Divergence? Regional Growth in Europe*, Springer, Wien/Heidelberg/New York, 2001.

[34] Vamvakidis, A., (2003), *Regional Convergence in Italy:* 1960-2002, in: *IMF*, Italy: Selected Issues, IMF Country Report, nr. 3/352, Washington, pp. 31-56.

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 Economics

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, IndexCopernicus 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:

July 25th, 2009 – Deadline for payments
August 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 25. 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
SWIFT Code RZBRROBU

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 scanned copy of Payment Form (Bank Transfer) by email to: office_conference@yahoo.com