High Growth Firms and Knowledge Structure: Evidence from a Sample of Listed Companies[1]

Alessandra Colombelli[a,b,c] – Jackie Krafft[c] – Francesco Quatraro[b,c]

PRELIMINARY DRAFT

(do not quote without authors' permission)

a) CRENoS, University of Cagliari
b) BRICK, Collegio Carlo Alberto
c) CNRS, GREDEG and University of Nice Sophia Antipolis

ABSTRACT.

This paper analyzes the contribution of high-growth firms to the process of knowledge creation. We articulate a framework in which knowledge stems from creative recombination and gazelles appear as the main actors of the Schumpeterian Mark I pattern of innovation, characterized by creative destruction and widening of the knowledge base. We derive indicators able to describe the structure of knowledge and feature firms' innovation strategies as either 'random screening' or 'organized search'. Empirical results confirm that out of the firms in our sample gazelles are those undertaking search activities in fields often loosely related to their existing competences, contributing to increase the variety in the knowledge base and to open up new technological trajectories. Gazelles therefore appear to be as genuine agents of creative destruction, creating the condition for restless growth.

Keywords: Gazelles; Recombinant Knowledge, Schumpeterian innovation patterns

JEL Classification Codes: L20, L10, 032

[1] The authors acknowledge the financial support of the European Union D.G. Research with the Grant number 266959 to the research project 'Policy Incentives for the Creation of Knowledge: Methods and Evidence' (PICK-ME), within the context Cooperation Program / Theme 8 / Socio-economic Sciences and Humanities (SSH).

1 Introduction

The process of firms' growth has long attracted the attention of economists. Most of empirical investigations in this field draw in fact on the seminal work by Gibrat (1931), who proposed that the growth of firms is a predominantly random process.

In the recent years the analysis of such topic gained momentum, with particular attention to the distributional properties of firms' growth rates, as well as to their persistence over time and to the investigation of their determinants (Bottazzi and Secchi, 2006; Coad, 2007; Coad and Hölzl, 2011, Parker et al., 2010; Acs and Mueller, 2008; Lee, 2010).

The understanding of the determinants of firms' growth represents a somewhat uncertain field to explore. First of all, it requires going beyond the traditional representation of the growth process as a purely stochastic one. Although the debate on the validity of the Gibrat's Law is still open, it is widely recognized that it cannot be assumed as a general law and that its validity cannot be taken as granted *ex ante* (see Lotti, Santarelli and Vivarelli, 2009). Moreover, it is quite hard to find out empirical regularities across sectors and countries concerning the effects of the most relevant variables on firms' growth.

More recently, the focus of empirical analyses on the determinants of firms' growth shifted to the understanding of the determinants of growth rates far higher than the average. As Henrekson and Johansson (2010) point out, the emergence of such field of enquiry derived from Birch's contributions, in which high growth firms, defined 'gazelles', are indicated as the main source of job creation in the economic system (Birch, 1979 and 1981). Gazelles represent companies that have achieved a minimum of certain percentage of sales growth each year, they can be small firm or young firms, and they often come from high technology sectors. These different definitions tend to illuminate different aspects of what a high growth firm really is, but they converge in saying that they deserve a focus in the analysis since they represent the most dynamic population of firms in the economy. For this reason the phenomenon of gazelles is important not only from the economic but also from the policy perspective. The understanding of the conditions under which firms become gazelles as well as of the channels through which they contribute to the dynamics of aggregate economic growth may indeed help policymakers to devise targeted supporting policy measures.

The focus on gazelles on the one hand engendered a repositioning of the analyses of firms' growth process, while on the other hand contributed to put the debate on the role of small firms in the process of economic growth. It is indeed clear from the studies conducted so far that, while gazelles are overrepresented in small size classes, there are also large firms contributing to a large extent to employment growth. Other regularities concern the young age of high growth firms and their somewhat even distribution across high-tech and low-tech industries (Henrekson and Johansson, 2010).

As noted by Coad and Hölzl (2011), the empirical studies of firms' growth concerns mainly the analysis of the distribution of growth rates, the enquiry of firms' growth determinants and the assessment of the contribution of gazelles to the process of economic growth.

The investigation of the relationship between innovation and faster rates of growth has received some attention only in the last years. Such studies have been mostly conducted within empirical settings drawing upon quantile regressions. These allow indeed to conduct regressions taking into account the uneven distribution of firms across different growth rates classes, and hence the issue of heterogeneous impacts of explanatory variables across different classes. Within this strands of analysis, Coad and Rao (2008) proposed an alternative way to proxy innovation and analyze its effects on firms' growth, and found that innovation is of crucial importance for high-growth firms. On a different ground, Coad and Rao (2010) shows that R&D is more likely to be influenced by growth than to influence it. Hölzl (2010) uses the CIS data and proposes a comparative analysis across 16 countries of the R&D behavior of gazelles, according to whether they are located in countries on the technological frontier or not. He finds that high-growth firms in advanced countries invest more than the others in R&D, supporting the conclusion that gazelles derive much of their success from the exploitation of local comparative advantages.

Most importantly, the studies concerning the relationship between high-growth and innovation uses firms' growth as a dependent variable. In other words, they assume the economic importance of gazelles and try and understand which could be the main factors affecting their outperforming behavior. When other dependent variables are taken into account, like the R&D in Coad and Rao (2010), the implementation of quantile regressions assign firms to different classes according the growth rate of R&D expenditure, and not on the basis of firms' growth. Therefore it becomes difficult to understand the contribution of gazelles to innovation dynamics.

In this paper we aim at filling this gap by investigating the differential contribution of high-growth firms to the creation of technological knowledge. This is largely motivated by the fact that the literature on gazelles emphasizes that the bulk of their economic contribution is due to the process of creative destruction that they are able to engender. The net job creation ascribed to high growth firms stems from an endless dynamics process in which new opportunities emerge and are likely to replace obsolete activities (Hölzl, 2009, 2010; Henrekson and Johansson, 2010; Daunfelt et al., 2010).

In doing so, we will adopt an approach to technological knowledge which allows to emphasizing its collective and recombinant nature, as well as to identifying key properties able to characterize innovation strategies as of random screening or organized search (Krafft, Quatraro and Saviotti, 2009). While this approach has been successfully implanted to analyze productivity performances at different levels (Nesta, 2008; Quatraro, 2010; Antonelli, Krafft, Quatraro, 2010), there are no contributions yet in the literature that have used it in the investigation of high-growth firms.

The emphasis on the process of creative destruction indeed recalls the concept of variety and selection, as well as the different innovation strategies of firms. In this direction, the main hypothesis we spell out in this paper is that gazelles are key in creating technological variety within the economy. They bear the risk of exploring the technology landscape towards areas far from their core competencies, therefore with uncertain outcomes but also more chances to produce radical innovations. This is the main mechanisms by which they contribute the process of restless economic growth. In our paper, gazelles will be identified as firms within the highest growth quantile, they are not necessarily small, nor young, and they come from every sector in the economy.

The rest of the paper is organized as follows. Section 2 presents the theoretical underpinnings of the analysis, and outlines the working hypotheses. In Section 3 we describe the data and the methodology, with particular emphasis on the implementation of knowledge related indicators. In section 4 we present and discuss the empirical results. Finally, Section 5 concludes.

2 High-growth firms and technological knowledge: a Schumpeterian story

Joseph Schumpeter is undoubtedly the economist who has influenced the most the development of the field of economics of innovation. Since his early works he has indeed underlined the importance of innovation dynamics to the process of economic development in capitalistic systems (Schumpeter, 1912 and 1942). Schumpeter's contributions became then the pillars of the evolutionary approach to technological change (Nelson and Winter, 1982; Nelson, 1990), and paved also the way to the investigation of the features of the so-called "Schumpeterian patterns" of technological change, according to which two different technological regimes can be identified, i.e. the "Schumpeter Mark I" and the "Schumpeter Mark II" (Malerba and Orsenigo, 1995 and 1997).

The integration of such Schumpeterian perspectives into the analysis of the contribution of gazelles to the economic system may be far reaching. We have already outlined the suggested interpretation provided by previous studies, according to which the positive effects of high growth firms to economic growth lie in the process of creative destruction. This concept is at the core of Schumpeter's theory of economic change. In *Capitalism, Socialism and Democracy* indeed Schumpeter argues: "The fundamental impulse that sets and keeps the capitalist engine in motion comes from the new consumers' goods, the new methods of production or transportation, the new markets, the new forms of industrial organization that capitalist enterprise creates. […] This process of Creative Destruction is the essential fact about capitalism. It is what capitalism consists in and what every capitalist concern has got to live in" (Schumpeter, 1942, pp. 82-83). Such competitive dynamics are clearly characterized by the emergence of new means of productions and new products that are likely to replace the old ones.

Creative destruction is also a distinctive feature of the pattern of innovative activities that Malerba and Orsenigo (1997) call Schumpeter Mark I. In particular such pattern is also characterized by ease of entry, the appearance of new firms based on business opportunities, which challenge incumbents and continuously disrupt the current ways of production, organization and distribution. On the contrary, the Mark II pattern is characterized by 'creative accumulation', the relevance of industrial R&D labs and the key role of large firms. They also label the two patterns as 'widening' and 'deepening'. The former is related to an innovative base which is continuously growing, while the latter are characterized by accumulation strategies based on the existing technological premises (Malerba and Orsenigo, 1995).

In this direction gazelles may well be identified as a phenomenon characterizing the technological regime typical of the Schumpeter Mark I. This is likely to characterize any kind of industry in such a regime, as high-growth firms are quite evenly distributed across sectors. Moreover, the literature emphasizes that firms do not remain necessarily gazelle for their entire life, and this is consistent with the evidence according to which the evolution of industries may be such that the Schumpeterian Mark I pattern of innovative activities may turn into the Schumpeter Mark II.

Conversely, firms characterized by average growth rates are not expected to contribute that much the net job creation and economic development. They may be viewed as typical of the 'deepening' phase characterizing the Mark II. They exploit the existing knowledge base to generate incremental innovations that may have a very small impact on the economy at the aggregate level.

Therefore, within this framework we hypothesize that high-growth firms play a key role in the advancement of technological knowledge, in that they are the main players of the 'widening' pattern characterizing the Schumpeter Mark I, while 'average-growing' firms are mostly typical of the 'deepening' pattern.

However, the investigation of the relationships between high-growth firms and the dynamics of technological knowledge within the perspective outlined so far may take a great advantage from the recent theories on knowledge creation. One of the main problems that have indeed characterized the analysis of the effects of innovation on growth lies in the difficulty of finding a reliable proxy for innovation activities (Coad and Rao, 2008).

Traditional approaches to technological knowledge have mostly represented as a homogeneous stock, as if it were the outcome of a quite uniform and fluid process of accumulation made possible by R&D investments, the same way as capital stock (Griliches, 1979; Mansfield, 1980). Such kind of representation is hardly useful to investigate the nature of firms' search strategies, as it only allows evaluating it from a quantitative rather than a qualitative viewpoint.

More recently, an increasingly share of scholars in the economics of innovation has elaborated theoretical approaches wherein the process of knowledge production is viewed as the outcome of a recombination process (Weitzmann, 1998; Kauffman, 1993). The creation of new knowledge is represented as a search process across a set of alternative components that can be combined one another. A crucial role is played here by the cognitive mechanisms underlying the search process aimed at exploring the knowledge space so as to identify the pieces that might possibly be combined together. The set of potentially combinable pieces turns out to be a subset of the whole knowledge space. Search is supposed to be local rather than global, while the degree of localness appears to be the outcome of cognitive, social and technological influences. The ability to engage in a search process within spaces that are distant from the original starting point is likely to generate breakthroughs stemming from the combination of brand new components (Nightingale, 1998; Fleming, 2001).

Based on these achievements, we can introduce the concept of knowledge structure. If knowledge stems from the combination of different technologies, knowledge structure can be represented as a web of connected elements. The nodes of this network stand for the elements of the knowledge space that may be combined with one another, while the links represent their actual combinations. The frequency with which two technologies are combined together provides useful information on the basis of which one can characterize the internal structure of the knowledge base according to the average degree of complementarity of the technologies which knowledge bases are made of, as well as to the variety of the observed pairs of technologies. In particular, we can identify at least three main properties of knowledge structure at a general level:

- **Coherence** can be defined as the extent to which the pieces of knowledge that agents within the sector combine to create new knowledge are complementary one another.

- **Similarity** (or dissimilarity) refers to the extent to which the pieces of knowledge used in the sector are close one another in the technology space.

- **Variety** is instead related to the technological differentiation within the knowledge base, in particular with respect to the diverse possible combinations of pieces of knowledge in the sector, from the creation of a radically new type of knowledge to the more incremental recombination of already existing types of knowledge.

The dynamics of technological knowledge can therefore be understood as the patterns of change in its own internal structure, i.e. in the patterns of recombination across the elements in the knowledge space. This allows for qualifying both the cumulative character of knowledge creation and the key role played by the properties describing knowledge structure, as well as for linking them to the relative stage of development of a technological trajectory (Dosi, 1982; Saviotti, 2004 and 2007; Krafft, Quatraro and Saviotti, 2009).

This approach allows for better qualifying a key distinction concerning innovation strategies, i.e. the one between exploration and exploitation (March, 1991). The view of knowledge as an outcome of a recombination activity allows for the introduction of two nested dimensions, defined according to the degree to which agents decide to rely either on exploration or exploitation or on a combination of both. To this purpose concepts like search depth and search scope have been introduced (Katila and Ahuja, 2002). The former refers to degree to which agents intend to draw upon their prior knowledge, while the latter refers to the degree to which agent intend to rely on the exploration of new areas in the knowledge space.

The grafting of the recombinant knowledge approach into the analysis of the Schumpeterian patterns of gazelles' innovation activities allows us to refine our main working hypothesis as follows.

The main channel through which gazelles contribute the process of economic growth is the mechanism of creative destruction. This in turn is a key feature of the Schumpeter Mark I pattern of innovation activities. The positive impact of high-growth firms is therefore due to their capacity to undertake search behaviors directed towards the exploration of untried technological fields, so as to widen the existing knowledge base. The extension of the knowledge base is indeed possible only by going beyond the fences of what firms already know. Exploration is therefore a key part of the destructive creativity of gazelles within the 'widening' pattern. The search behavior of high-growth firms is therefore expected to depart from the established trajectories to discover new fields in the technology landscape so as to enlarge their search scope. Conversely, average-growth firms are those engaged in strategies that keep them within the comfortable boundaries of their established knowledge base. They organize their search activities by exploiting learning dynamics.

In view of this, we turn now to describe the data and the methodology we will use to provide an operational definition of the concept of recombinant knowledge as well as of the properties of knowledge structure and to be used to featuring the search behavior of high-growth firms.

3 Data and Methodology

3.1 Dataset

The dataset used in this paper is an unbalanced panel of firms which are publicly traded in UK, Germany, France, Sweden, Italy and Netherlands. Our prime source of data for both market and accounting data is Thomson Datastream. In order to include relevant variables like year of foundation and zip code, we pooled the dataset by adding also information collected from AMADEUS by Bureau Van Dijk. For all the countries, the period of observations goes from 1988 to 2005. Other fundamental sources of data are the OECD REGPAT database and the OECD-EPO citations database. The former presents patent data that have been linked to regions utilizing the addresses of the applicants and inventors. The latter covers the citations associated to all patent applications published by EPO and WIPO, under the Patent Co-operation Treaty (PCT), from their introduction in 1978 up until 2006.

In order to match the database collecting information at the firm level and those including patents data we refer to the work by Thoma et al. (2010) that develops a method for the harmonization and combination of large-scale patent and trademark datasets with each other and other sources of data through the standardization of applicant and inventor's names.

We finally pooled the dataset by adding also information at the industry level from the OECD STAN database. As STAN uses the ISIC revision 3 sectoral classification while Thomson Datastream uses the ICB industry classification at the four digit level, in Appendix A we provide the sectoral concordance table used to link the two classifications.

Our final dataset consist of an unbalanced panel of 316 active companies that are listed on the main European financial market and have applied for more than one patent at the European Patent Office over the period under scrutiny. Table 1 reports the sample distribution by macro-sector. High and medium-high technology firms are highly represented in our sample covering about 24% and 39% observations, respectively. Medium low and low technology include 9% and 11% while knowledge intensive sectors cover about 6% of observations. Finally, each of the other economic groups includes around or less than 3% observations.

>>>INSERT TABLE 1 ABOUT HERE<<<

Figure 1 shows the distribution of firms' growth rates. We define firm's rate of growth as:

$$Growth_{i,t} = s_{i,t} - s_{i,t-1} \qquad (1)$$

where $s_{i,t}$ is the logarithm of firm size at time t and $s_{i,t-1}$ its lagged value.

>>>INSERT FIGURE 1 ABOUT HERE<<<

As evidenced by the figure, the empirical distribution of the growth rates for our sample seems closer to a Laplacian than to a Gaussian distribution. This is in line with previous studies analysing the distribution of firm growth rates (Bottazzi et al. 2007; Bottazzi and Secchi 2003; Castaldi and Dosi 2009).

In Figure 2 we show the sample average growth rate by year. On average the growth rate of sample companies follows an increasing trend in the period going from 1990 and 2000 while it decrease in period 2000-2003.

>>>INSERT FIGURE 2 ABOUT HERE<<<

Figure 3 illustrate the distribution of firm's growth by macro-sector. The diagrams show that firms' rate of growth is highly dispersed in high-tech sectors and that the dispersion decreases going from high-tech to low-tech sectors. The same apply to service sectors, where knowledge intensive sectors show higher dispersed growth rate than less knowledge intensive ones.

>>>INSERT FIGURE 3 ABOUT HERE<<<

3.2 Methodology

Most of empirical works analysing the determinants of firms' growth are based on the Gibrat's Law, which holds that firm growth is independent from size. Yet, a number of recent studies reveals departure from this law and it is widely recognized that it cannot be assumed as a general law and that its validity cannot be taken as granted *ex ante* (see Lotti, Santarelli and Vivarelli, 2009). Moreover, previous works find that growth rates are autocorrelated. Yet, in this paper we investigate the differential contribution of high growth firms to the creation of technological knowledge. For this reason, we model the relationship between firms's knowledge structure and growth as follows:

$$d\log x_{i,t}/dt = \lambda_1 + \lambda_2 \log x_{i,t-1} + \lambda_3 Growth_{i,t-n} + \lambda_3 \log Size_{it-1} + \omega_j + \psi_t + \varepsilon_{it} \qquad (2)$$

Where x_{it} is a vector of variables describing the structure of firms' knowledge, which will be introduced in the next Section, growth is the log difference of deflated sales and logSize represents the level of sales in logarithm. We further includes a vector of sectoral dummies (ω_j) and a vector of year dummies (ψ_t) controlling for macro cyclical and time effects .

3.3 The Implementation of Knowledge Indicators

The implementation of knowledge indicators rests on the recombinant knowledge approach. In order to provide an operational translation of such variables one needs to identify both a proxy for the bits of knowledge and a proxy for the elements that make their structure. For example one could take scientific publications as a proxy for knowledge, and look either at keywords or at scientific classification (like the JEL code for economists) as a proxy for the constituting elements of the knowledge structure. Alternatively, one may consider patents as a proxy for knowledge, and then look at technological classes to which patents are assigned as the constituting elements of its structure, i.e. the nodes of the network representation of recombinant knowledge. In this paper we will follow this latter avenue[2]. Each technological class *j* is linked to another class *m* when the same patent is assigned to both of them. The higher is the number of patents jointly assigned to classes *j* and *m*, the stronger is this link. Since technological classes attributed to patents are reported in the

[2] The limits of patent statistics as indicators of technological activities are well known. The main drawbacks can be summarized in their sector-specificity, the existence of non patentable innovations and the fact that they are not the only protecting tool. Moreover the propensity to patent tends to vary over time as a function of the cost of patenting, and it is more likely to feature large firms (Pavitt, 1985; Griliches, 1990). Nevertheless, previous studies highlighted the usefulness of patents as measures of production of new knowledge. Such studies show that patents represent very reliable proxies for knowledge and innovation, as compared to analyses drawing upon surveys directly investigating the dynamics of process and product innovation (Acs et al., 2002). Besides the debate about patents as an output rather than an input of innovation activities, empirical analyses showed that patents and R&D are dominated by a contemporaneous relationship, providing further support to the use of patents as a good proxy of technological activities (Hall et al., 1986). Moreover, it is worth stressing that our analysis focuses on the dynamics of manufacturing sectors.

patent document, we will refer to the link between *j* and *m* as the co-occurrence of both of them within the same patent document[3]. We may now turn to explain how knowledge characteristics may be translated into computable variables.

1) Let us start by the traditional firm's knowledge stock. This is computed by applying the permanent inventory method to patent applications. We calculated it as the cumulated stock of past patent applications using a rate of obsolescence of 15% per annum: $E_{i,t} = \dot{h}_{i,t} + (1-\delta)E_{i,t-1}$, where $\dot{h}_{i,t}$ is the flow of patent applications and δ is the rate of obsolescence[4].

2) As for the properties of knowledge we are interested in, we decided to measure variety in firms' knowledge base by using the information entropy index. Entropy measures the degree of disorder or randomness of the system, so that systems characterized by high entropy will also be characterized by a high degree of uncertainty (Saviotti, 1988).

   Such index was introduced to economic analysis by Theil (1967). Its earlier applications aimed at measuring the diversity degree of industrial activity (or of a sample of firms within an industry) against a uniform distribution of economic activities in all sectors, or among firms (Attaran, 1985; Frenken et al., 2007; Boschma and Iammarino, 2009).

   Differently from common measures of variety and concentration, the information entropy has some interesting properties (Frenken and Nuvolari, 2004). An important feature of the entropy measure, which we will exploit in our analysis, is its multidimensional extension. Consider a pair of events (X_j, Y_m), and the probability of co-occurrence of both of them p_{jm}. A two dimensional (total) entropy measure can be expressed as follows (firm and time subscripts are omitted for the sake of clarity):

$$H(X,Y) = \sum_{j=1}^{q}\sum_{m=1}^{w} p_{jm} \log_2\left(\frac{1}{p_{jm}}\right) \tag{3}$$

   If one considers p_{jm} to be the probability that two technological classes *j* and *m* co-occur within the same patent, then the measure of multidimensional entropy focuses on the variety of co-occurrences of technological classes within firms' patents portfolios.

   Moreover, the total index can be decomposed in a "within" and a "between" part anytime the events to be investigated can be aggregated in a smaller numbers of subsets. Within-entropy measures the average degree of disorder or variety within the subsets, while between-entropy focuses on the subsets measuring the variety across them. It can be easily shown that the decomposition theorem holds also for the multidimensional case. Hence if one allows $j \in S_g$ and $m \in S_z$ (g = 1,…,G; z = 1,…, Z), we can rewrite *H(X,Y)* as follows:

---

[3] It must be stressed that to compensate for intrinsic volatility of patenting behaviour, each patent application is made last five years.

[4] Different depreciation rates have been implemented, which provided basically similar results.

$$H(X,Y) = H_Q + \sum_{g=1}^{G}\sum_{z=1}^{Z} P_{gz} H_{gz} \tag{4}$$

Where the first term of the right-hand-side is the between-group entropy and the second term is the (weighted) within-group entropy. In particular:

$$H_Q = \sum_{g=1}^{G}\sum_{z=1}^{Z} P_{gz} \log_2 \frac{1}{P_{gz}} \tag{4a}$$

$$P_{gz} = \sum_{j\in S_g} \sum_{m\in S_Z} p_{jm} \tag{4b}$$

$$H_{gz} = \sum_{j\in S_g} \sum_{m\in S_z} \frac{p_{ij}}{P_{gz}} \log_2\left(\frac{1}{p_{jm}/P_{gz}}\right) \tag{4c}$$

Following Frenken et al. (2007), we can refer to between-group and within-group entropy respectively as *unrelated technological variety* (*UTV*) and *related technological variety* (*RTV*), while total information entropy is referred to as *general technological variety (TV)*. The distinction between related and unrelated variety is based on the assumption that any pair of entities included in the former generally are more closely related, or more similar to any pair of entities included in the latter. This assumption is reasonable when a given type of entity (patent, industrial sector, trade categories etc.) is organized according to a hierarchical classification. In this case each class at a given level of aggregation contains "smaller" classes, which, in turn contain yet "smaller" classes. Here, small refers to a low level of aggregation.

We can reasonably expect then that the average pair of entities at a given level of aggregation will be more similar than the average pair of entities at a higher level of aggregation. Thus, what we call related variety is measured at a lower level of aggregation (3 digit class within a 1 digit macro-class) than unrelated variety (across 1 digit macro-classes). This distinction is important because we can expect unrelated (or inter-group) variety to negatively affect productivity growth, while related (or intra-group) variety is expected to be positively related to productivity growth. Moreover, the evolution of total variety is heavily influenced by the relative dynamics of related and unrelated variety, such that if unrelated variety is dominant the effects of total variety on productivity growth can be expected to be negative, while the opposite holds if related technological variety dominates the total index (Krafft, Quatraro, Saviotti, 2011).

3) Third, we calculated the *coherence* (*R*) of firms' knowledge base, defined as the average complementarity of any technology randomly chosen within the firm's portfolio with respect to any other technology (Nesta and Saviotti, 2005 and 2006; Nesta, 2008).

To yield the knowledge coherence index, a number of steps are required. In what follows we will describe how to obtain the index at the firm level. First of all, one should calculate the weighted average relatedness WAR_i of technology *i* with respect to all other technologies present within the sector. Such a measure builds upon the measure of technological relatedness τ, which is introduced in Appendix A. Following Teece et al. (1994), WAR_j is defined as the degree to which technology *j* is related to all other technologies $m \neq j$ within the firm *i*, weighted by patent count P_{mit}:

$$WAR_{jit} = \frac{\sum_{m \neq j} \tau_{jm} P_{mit}}{\sum_{m \neq i} P_{mit}} \tag{5}$$

Finally the coherence of knowledge base within the firm is defined as weighted average of the WAR_{jit} measure:

$$R_{it} = \sum_{j \neq m} WAR_{jit} \times \frac{P_{jit}}{\sum_j P_{jit}} \tag{6}$$

This measure captures the degree to which technologies making up the firm's knowledge base are complementary one another. The relatedness measure τ_{jm} indicates indeed that the utilization of technology *j* implies that of technology *m* in order to perform specific functions that are not reducible to their independent use. This makes the coherence index appropriate for the purposes of this study.

4) We finally implement a measure of knowledge similarity, as proxied by cognitive distance (Nooteboom, 2000), which is able to express the dissimilarities amongst different types of knowledge. A useful index of distance can be derived from the measure of *technological proximity*. Originally proposed by Jaffe (1986 and 1989), who investigated the proximity of firms' technological portfolios. Subsequently Breschi et al. (2003) adapted the index in order to measure the proximity, or relatedness, between two technologies. The idea is that each firm is characterized by a vector *V* of the *k* technologies that occur in its patents. Knowledge similarity can first be calculated for a pair of technologies *l* and *j* as the angular separation or un-cented correlation of the vectors V_{lk} and V_{jk}. The similarity of technologies l and j can then be defined as follows:

$$S_{lj} = \frac{\sum_{k=1}^{n} V_{lk} V_{jk}}{\sqrt{\sum_{k=1}^{n} V_{lk}^2} \sqrt{\sum_{k=1}^{n} V_{jk}^2}} \tag{7}$$

The idea underlying the calculation of this index is that two technologies j and i are similar to the extent that they co-occur with a third technology k. The cognitive distance between j and l is the complement of their index of the similarity:

$$d_{lj} = 1 - S_{lj} \tag{8}$$

Once the index is calculated for all possible pairs, it needs to be aggregated at the firm level to obtain a synthetic index of technological distance. This can be done in two steps. First of all one can compute the weighted average distance of technology *i*, i.e. the average distance of *i* from all other technologies.

$$WAD_{lt} = \frac{\sum_{j \neq l} d_{lj} P_{jit}}{\sum_{j \neq l} P_{jit}} \tag{9}$$

Where P_j is the number of patents in which the technology *j* is observed. Now the average cognitive distance at time *t* is obtained as follows:

$$CD_t = \sum_l WAD_{lit} \times \frac{P_{lit}}{\sum_l P_{lit}} \tag{10}$$

Descriptive statistics for the knowledge indicators described so far and the other variables included in our model are shown in Table 2. In the next section we will provide the empirical results obtained by estimating Equation (2).

>>>INSERT TABLE 2 ABOUT HERE<<<

4 Empirical results and Discussion

The main hypothesis of this paper is that the economic importance of gazelles is to be ascribed, at least in part, to their contribution to the process of knowledge generation. In particular, we maintain that gazelles as agents of creative destruction are the main actors of the Schumpeterian Mark I pattern of innovation. In view of this, they undertake search strategies allowing for the widening of the existing knowledge base, characterized by the exploration of the technological landscape towards unfamiliar areas that are more likely to bring about disruptive technological change.

Tables 3 and 4 provide the results of the empirical analyses carried out to test such hypothesis, on the basis of Equation (2). In particular, we regressed the growth rates of our properties of the knowledge base, i.e. coherence, cognitive distance and variety (related and unrelated) against the growth rates of sales and a set of control variables. We first carried out our regressions on the pooled sample, and then repeated them by assigning firms to different percentiles, identified on the basis of sales growth.

>>>INSERT TABLE 3 ABOUT HERE<<<

Table 3 shows the results which take into account the effect of the first lag of sales growth rate. Interestingly enough, growth rate do not seem to affect the growth rate of knowledge coherence within this time span. On the contrary, sales growth exhibits a relationship with cognitive distance, although only for firms in the second and the third percentiles. In particular, for firms in the first quartile an acceleration of the sales growth is related to the slackening of cognitive distance growth.

In other words, the enhancement of the pace of slow-growing firms is related to a search behavior conducted within the boundaries of existing competences, that do not make cognitive distance increase. When we move to the subsequent quartile the sign of the impact of growth on cognitive distance changes. However, the effect of growth on such variable is not statistically significant for high growth firms.

The results are quite interesting for what concerns the effects of growth on technological variety. Indeed the effect of growth on total variety is positive and significant on the pooled regression. When we look at the decomposition in quartiles, we find that sales growth is positively related to variety growth only for firms in the higher quartiles, i.e. for gazelles. This is largely supportive that gazelles are responsible for enlarging the scope of the knowledge bases. Moreover, when we decompose total variety in related and unrelated components, the positive effect of sales growth on total variety appears to be driven by unrelated variety. We can therefore maintain that gazelles are not only responsible of the increase of technological variety, but that they are also likely to boost technological variety by introducing unrelated technological fields within the knowledge base. They are therefore likely to bring about disruptive technological change, based on unfamiliar knowledge combinations which destroy the obsolete competences. Such result is also confirmed by the evidence concerning related variety. In this case, sales growth has a positive only for firms in the lower quartiles, signaling that slow-growing firms are able to introduce some variety only if new combinations fall within the realm of well-know technological avenues.

Table 4 shows instead the results of the estimations of Equation (2) carried out by using a 2-years lag for sales growth. The evidence concerning the coherence index is fairly supportive of our hypothesis. Knowledge coherence is indeed expected to fall as an effect of exploration strategies in which firms direct their innovation efforts towards knowledge that exhibit lower degree of complementarity with the existing knowledge base. By enlarging the search scope they necessarily try and combine pieces of knowledge loosely related to their established competences. The negative and significant sign linking sales growth to coherence growth can therefore be interpreted in this direction. We are not in the position to push the interpretation too far. The results support the idea that for high-growth firms, an acceleration of growth rates is related to the slackening of coherence growth rates. This means that coherence falls, but this does not imply that coherence becomes negative, a situation that will be more difficult to manage.

>>>INERT TABLE 4 ABOUT HERE<<<

For what concerns cognitive distance, we find that growth is positive and weakly significant only in the pooled regressions, while it does not show any significance when we look at the different quartiles. A somewhat similar situation applies for what concerns technological variety. It is interesting to note instead that growth rates positively affect related technological variety for firms in the lower quartile. Once again this supports the idea that slow-growing firms enhance their performances by searching in the neighborhood of their technological competences.

5 Conclusions

The process of firms' growth has long attracted the attention of economists. More recently, a new strand of literature emerged focused on the analysis of high-growth firms, also defined as gazelles, due to the increasing evidence about their exceptional contribution to aggregate economic growth.

However, few empirical studies can be found on the relationship between high-growth firms and innovation. The existing literature in this framework has mostly analyzed innovation as a determinant of high-growth rates. No investigations can instead be found concerning gazelles' contribution to the process of knowledge creation. In this paper we have tried to fill this gap.

The bulk of the literature on gazelles identify in the process of creative destruction the main channel through which they contribute to aggregate economic performances. Such emphasis inevitably evokes Schumpeter's contribution to the economics of innovation. In this direction we have recalled the two Schumpeterian patterns of innovation activities, i.e. the Mark I and Mark II, proposing that gazelles are likely to play as main actors within the Mark I regime, characterized by creative destruction and the widening of the knowledge base. In such a regime disruptive technological change is more likely to happen, creating new business opportunities and replacing the old ones. Gazelles are therefore supposed to undertake innovation strategies oriented toward the exploration of the technology landscape beyond the fences of their established competences, so as to enlarge their search scope. The grafting of recombinant knowledge theory in this framework allowed us to propose the concept of knowledge structure as characterized by three key properties, i.e. coherence, similarity and variety (related and unrelated), which in turn can be usefully employed to distinguish between 'random screening' and 'organized search' strategies.

Such hypotheses have been tested by using data about listed companies and patent applications. We have regressed the growth rates of the properties of the knowledge base against the firms' sales growth, by using two different lag specifications. We have found that the 1-year lag of growth process of high-growth firms is more likely to positive affect the growth of technological variety, and in particular of unrelated technological variety, while the 2-years lag of firms' growth negatively affect the growth rates of knowledge coherence.

Such results represent the first attempt to investigate the contribution of gazelles to the process of knowledge creation, and certainly they require further refinement. In particular it may be worth analyzing such relationships by portioning the sample according to different definitions of the quartiles. Moreover, the kinds of empirical implementation of knowledge coherence like the one used in this paper has been recently criticized by Bottazzi and Pirino (2010), and it would be interesting to try the corrected index they suggest to check if our results still hold. Moreover, it would also be interesting to further check for the robustness of the results, by implementing different estimators, accounting for the distribution of explanatory variables as well as for the impact of outliers.

All in all, although preliminary, we think that such evidence provides an interesting basis for further investigations as well as for their policy implications. Our results call in fact for the integration of innovation policies with industrial policies directed towards the support of high-growth firms. Innovation policies are indeed often evoked as strategic tools to foster economic growth, by placing particular importance, on the one hand, on the interactive dynamics of knowledge production, hence implementing conditions fostering the creation of clusters, and, on the other hand, on the identification of key sectors. These are important issues and we think that our analysis adds another important dimension there. It would be indeed useful to rethink the allocation mechanisms of government funding, going beyond the focus on firms' size to emphasize the important of growth dynamics. High-growth firms should therefore be the target of innovation policies aiming at fostering the exploration of new technological fields susceptible to provide the basis for the elaboration of new business opportunities.

6 References

Acs et al., 2002, "Patents and Innovation Counts as Measures of Regional Production of New Knowledge," *Research Policy*, 2002, 31(7), 1069-1085.

Acs and Mueller, "Employment effect of business dynamics: mice, gazelles and elephants", Small Business Economics, 30, 85-100.

Antonelli, Krafft, Quatraro, 2010, "Recombinant Knowledge and Growth: The Case of ICTs", *Structural Change and Economic Dynamics*, 21(1), 50-69.

Attaran, 1985, Industrial diversity and economic performance in U.S. areas. *The Annals of Regional Science* 20, pp. 44-54.

Birch, 1979, The job generation process, Cambridge MA: MIT Press.

Birch, 1981, "Who creates jobs?", The Public interest, 65(fall), 3-14.

Boschma and Iammarino, 2009, Related variety, trade linkages, and regional growth in Italy. *Economic Geography* 85, 289-311.

Bottazzi, G. And Pirino, D., 2010, Measuring Industry Relatedness and Corporate Coherence, LEM Working Paper 10/2010, Sant'Anna School of Advanced Studies, Pisa.

Bottazzi and Secchi, 2006, "Explaining the Distribution of Firms Growth Rates," Rand Journal of Economics, 37, 234–263.

Breschi et al., 2003, 'Knowledge relatedness in firm technological diversification', *Research Policy*, 32, 69-97.

Coad and Hölzl, 2011, "Firms growth: empirical analysis", in M. Dietrich and J. Krafft, Handbook on the economics and theory of the firm, Cheltenham: Edward Elgar.

Coad and Rao, 2008, "Innovation and firm growth in high tech sectors: a quantile regression approach", Research Policy, 37(4), 633-648.

Coad and Rao, 2010, "Firm growth and R&D expenditure," Economics of Innovation and New Technology, 19(2), 127–145.

Coad, 2007, "A Closer Look at Serial Growth Rate Correlation," Review of Industrial Organization, 31(1), 69–82.

Daunfelt et al., 2010, "When is Gibrat's law a law?", Ration working papers 158, The Ratio Institute.

Dosi, 1982, "Technological paradigms and technological trajectories", *Research Policy*, 11, 147–162.

Fleming, 2001, "Recombinant Uncertainty in Technological Search", *Management Science* 47(1), 117-132.

Frenken et al., 2007, 'Related Variety, Unrelated Variety and Regional Economic Growth', *Regional Studies*, 41(5), 685-97.

Frenken and Nuvolari, 2004, 'Entropy Statistics as a Framework to Analyse Technological Evolution', in J. Foster and W. Hölzl (eds), *Applied Evolutionary Economics and Complex Systems*. Edward Elgar: Cheltenham, U.K. and Northampton, Mass.

Gibrat, 1931, Les inégalités économiques. Recueil Sirey, Paris.

Griliches, 1979, . Issues in assessing the contribution of research and development to productivity growth. *The Bell Journal of Economics*. 10 92-116.

Griliches, 1990, ‘Patent statistics as economic indicators: a survey’, *Journal of Economic Literature*, 28, 1661-1707.

Henrekson and Johansson, 2010, “Gazelles as Job Creators - A Survey and Interpretation of the Evidence,” *Small Business Economics*, 35(2), 227-244.

Hölzl, 2009, "Is the R&D behaviour of fast-growing SMEs different? Evidence from CIS III data for 16 countries", *Small Business Economics*, 33(1), 59-75.

Hölzl, 2010, "The Economics of Entrepreneurship Policy: Introduction to the Special Issue," *Journal of Industry, Competition and Trade*, 10(3), 187-197.

Jaffe, 1986, ‘Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value’, *American Economic Review*, 76(5), 984-1001.

Jaffe, 1989, ‘Real Effects of Academic Research’, *American Economic Review*, 79(5), 957-70.

Katila and Ahuja, 2002, “Something old, something new: a Longitudinal Study of Search Behavior and New Product Introduction”, *Academy of Management Journal*, 45(6), 1183-1194.

Kauffman, 1993, *Origins of order: Self-Organization and selection in evolution*, Oxford University Press, Oxford.

Klepper, S. (1997), ‘Industry Life Cycles’, *Industrial and Corporate Change*, 6(1), 145-81.

Krafft, Quatraro and Saviotti, 2009, 2009. Evolution of the knowledge base in knowledge intensive sectors. LEI-BRICK Working Paper no 06/2009.

Krafft, Quatraro, Saviotti, 2011, The knowledge base evolution in biotechnology: A social network analysis”, *Economics of Innovation and New Technology*, accepted for publication, forthcoming.

Lee, 2010, “A theory of firm growth: learning capability, knowledge threshold, and patterns of growth”, *Research Policy*, 39, 278-289.

Lotti, Santarelli and Vivarelli, 2009,

Malerba and Orsenigo, 1995, "Schumpeterian Patterns of Innovation," *Cambridge Journal of Economics*, Oxford University Press, vol. 19(1), pages 47-65, February.

Malerba and Orsenigo, 1997, "The Dynamics and Evolution of Industries," *Industrial and Corporate Change*, Oxford University Press, vol. 5(1), pages 51-87.

Mansfield, 1980, Basic research and productivity increase in manufacturing, *American Economic Review*, 70, 863-73.

March, 1991, ‘Exploration and exploitation in organizational learning’, *Organization Science*, 2(1), 71-87.

Nelson and Winter, 1982, *An Evolutionary Theory of Economic Change*, Cambridge, Harvard University Press.

Nelson, 1990, Capitalis as engine of progress, *Research Policy*, 19, 193-214.

Nesta and Saviotti, 2005, Coherence of the Knowledge Base and the Firm's Innovative Performance: Evidence from the U.S. Pharmaceutical Industry. *Journal of Industrial Economics.* 53(1) 123-42.

Nesta and Saviotti, 2006, . Firm Knowledge and Market Value in Biotechnology. *Industrial and Corporate Change*. 15(4) 625-52.

Nesta, 2008, Knowledge and productivity in the world’s largest manufacturing corporations. *Journal of Economic Behavior and Organization*. 67 886-902.

Nightingale, 1998, A cognitive model of innovation, *Research Policy*, 27, 689-709.

Nooteboom, 2000, *Learning and innovation in organizations and economies*, Oxford: Oxford University Press.

Parker et al., 2010, "What happens to gazelles? The importance of dynamic management strategy", Small Business Economics, 35, 203-226.

Pavitt, 1985, Patent statistics as indicators of innovative activities: Possibilities and problems. *Scientometrics* 7, 77-99.

Quatraro, 2010, "ICT Capital and Services Complementarities. The Italian Evidence, *Applied Economics*, 2010, DOI: 10.1080/00036840903299805.

Saviotti, 1988, 'Information, variety and entropy in technoeconomic development', *Research Policy*, 17(2), 89-103.

Saviotti, 2004, 'Considerations about the production and utilization of knowledge', *Journal of Institutional and Theoretical Economics*, 160, 100-121.

Saviotti, 2007, On the dynamics of generation and utilisation of knowledge: The local character of knowledge', *Structural Change and Economic Dynamics*, 18, 387-408.

Schumpeter, 1912, *The Theory of Economic Development*, Harvard University Press, Cambridge.

Schumpeter, 1942, *Capitalism, Socialism and Democracy*, Harper and Row, New York.

Teece et al., 1994, 'Understanding Corporate Coherence: Theory and Evidence', *Journal of Economic Behavior and Organization*, 23(1), 1-30.

Theil, 1967, *Economics and Information Theory*. Amsterdam: North Holland, University Press: Oxford.

Weitzmann, 1998, "Recombinant growth", *Quarterly Journal of Economics,* 113, 331-360.

Figure 1 Kernel density distribution of growth

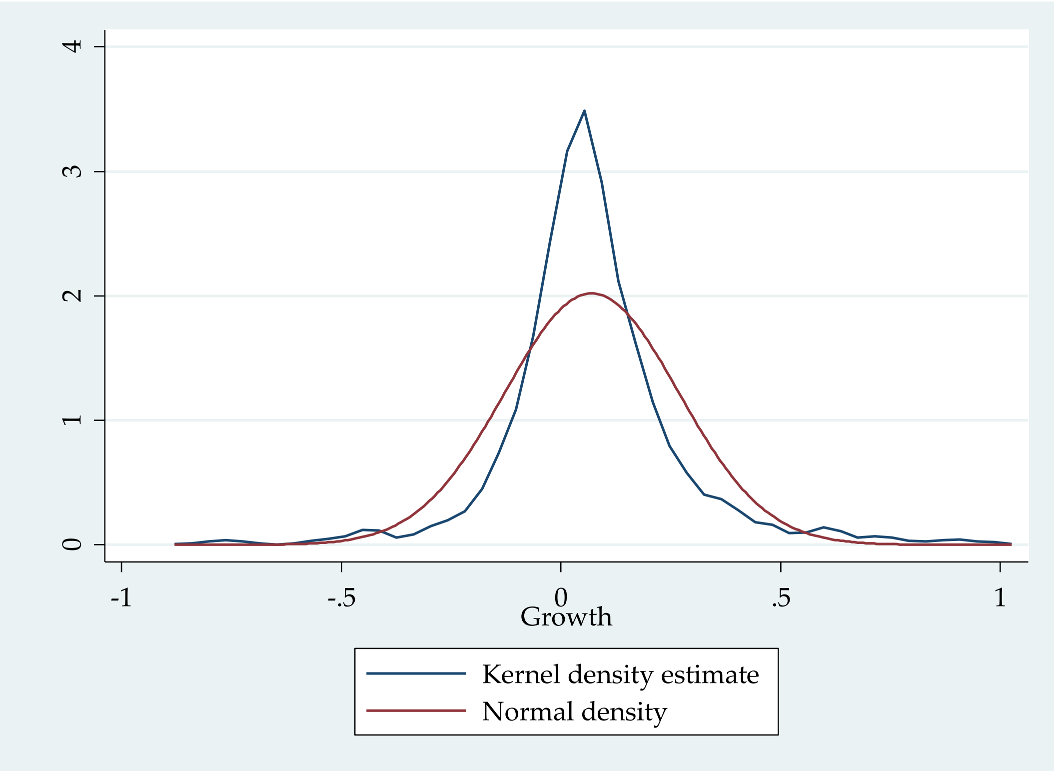

Figure 2 Sample average growth by year

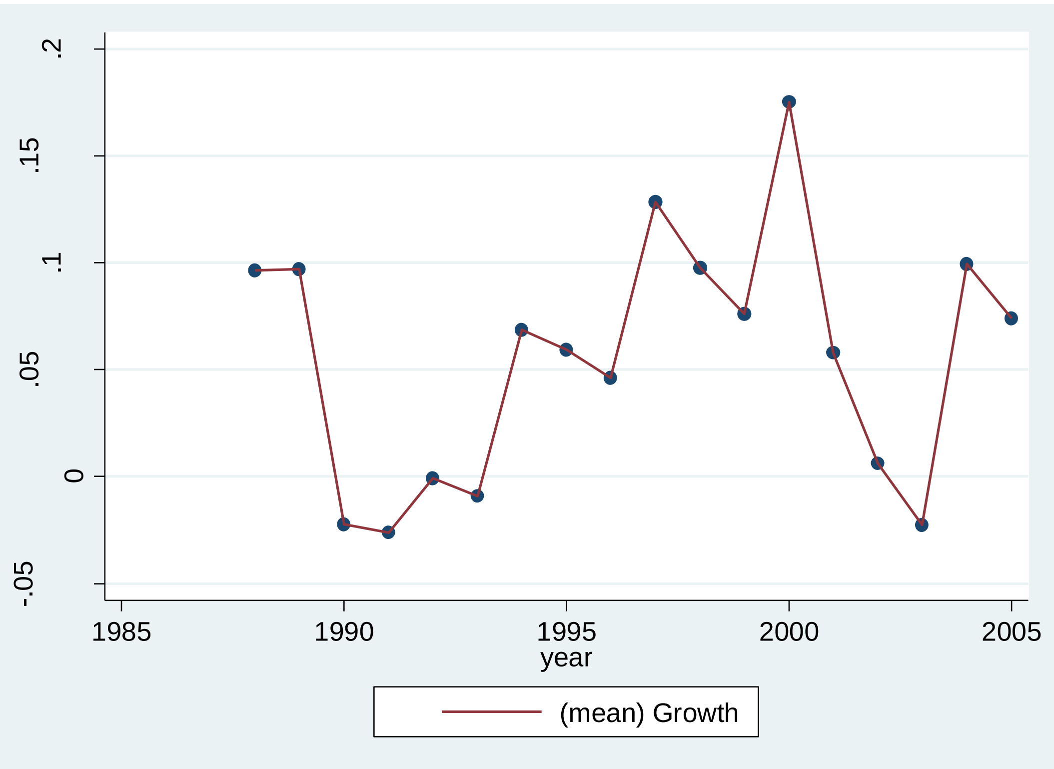

Figure 3 Box plot of growth by macro-sector

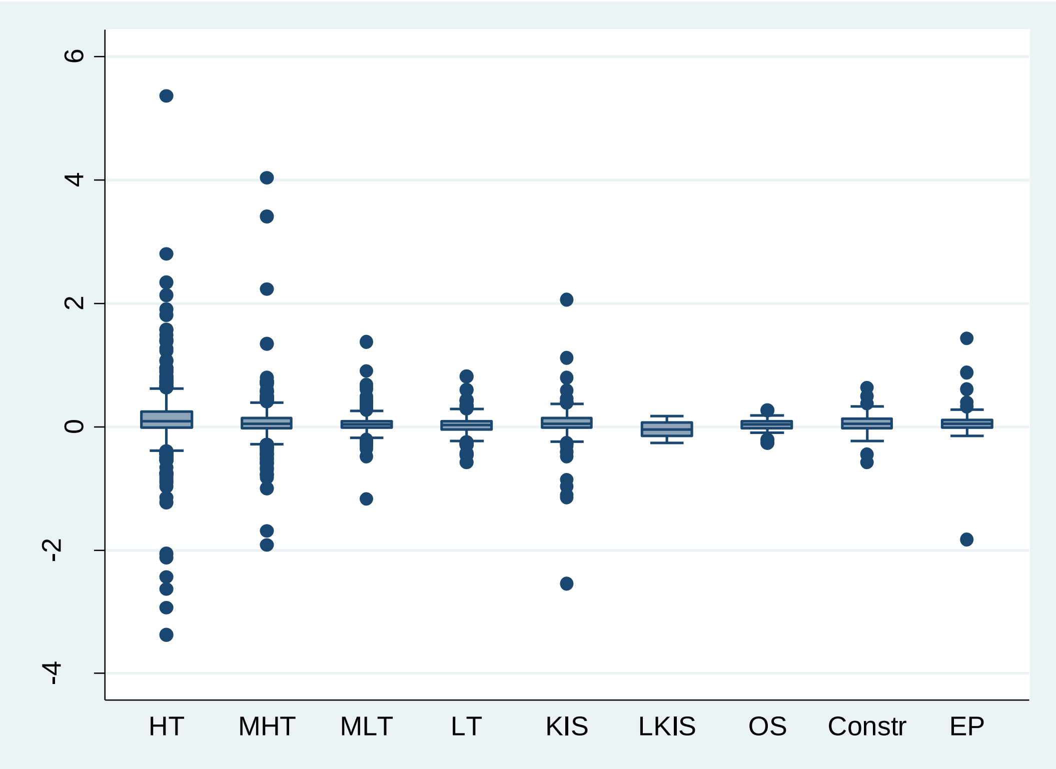

Table 1 Descriptive statistics by macro-sector, 1988-2005

	Freq.	Percent	Cum.
HT	655	24.30	24.30
MHT	1,066	39.54	63.84
MLT	252	9.35	73.18
LT	306	11.35	84.53
KIS	177	6.57	91.10
LKIS	16	0.59	91.69
OS	43	1.59	93.29
Constr	99	3.67	96.96
EP	82	3.04	100.00
Total	2,696	100.00	

Table 2 Descriptive statistics

Variable	Obs	Mean	Std.Dev.	Min	Max
Growth	2696	0.065	0.326	-3.369	5.361
Size	2696	6424190	1.66e	30.779	1.80e
logSize	2696	13.529	2.446	3.427	19.006
KCAP	2576	538.967	2363.487	3.7	30176.2
KOH	2576	14.195	21.906	-11.957	195.183
dlog KOH	2576	-0.0120	0.356	-5.223	3.797
CD	1401	0.0183	0.0336	0	0.307
dlogCD	1317	0.00564	0.239	-1.115	1.524
TV	2544	3.355	1.962	0	9.262
dlogtv	2343	0.00898	0.128	-1.098	1.200
RTV	2544	1.624	1.310	0	6.140
dlogrtv	2107	0.0117	0.216	-1.386	1.0779
UTV	2544	1.731	0.957	0	4.379
dlogutv	2195	0.00427	0.165	-1.0859	1.0482

Table 3 – Growth and Knowledge Structure (1-year lagged growth rates)

Dependent variable dlogKOH/dt						
	OLS	FE	1st quartile	2^{nd} quartile	3rd quartile	4th quartile
logKOH(t-1)	-.223***	-.742***	-.190***	-.249***	-.233***	-.204***
	(.011)	(.019)	(.025)	(.022)	(.021)	(.026)
logKCAP(t-1)	.012***	.035**	.024**	.014*	.017**	-.002
	(.004)	(.014)	(.010)	(.008)	(.008)	(.012)
Growth(t-1)	.016	.026*	-.011	.069	.009	.008
	(.016)	(.015)	(.025)	(.049)	(.044)	(.028)
logSize(t-1)	-.011***	-.046	-.016**	-.021***	-.009	-.005
	(.003)	(.085)	(.007)	(.007)	(.006)	(.007)
N obs	2805	2805	549	875	869	512
R-sq.	0.134	0.386	0.158	0.169	0.158	0.148
Dependent variable dlogCD/dt						
	OLS	FE	1st quartile	2^{nd} quartile	3rd quartile	4th quartile
logCD(t-1)	-.289***	-.487***	-.218***	-.374***	-.143***	-.387***
	(.019)	(.028)	(.051)	(.035)	(.039)	(.041)
logKCAP(t-1)	-.001	.031*	.006	-.00004	-.005	.005
	(.004)	(.019)	(.009	(.008)	(.007)	(.011)
Growth(t-1)	-.0005	.055***	-.039**	.102***	-.032	-.010
	(.014)	(.019)	(.020)	(.037)	(.039)	(.029)
logSize(t-1)	-.002	-.024	-.001	-.007	-.001	-.003
	(.003)	(.017)	(.006)	(.007)	(.007)	(.006)
N obs	1439	1439	255	466	463	255
R-sq.	0.141	0.229	0.191	0.235	0.069	0.362
Dependent variable dlogTV/dt						
	OLS	FE	1st quartile	2^{nd} quartile	3rd quartile	4th quartile
logTV(t-1)	-.061***	-.369***	-.071***	-.035***	-.078***	-.052***
	(.007)	(.017)	(.019)	(.014)	(.015)	(.019)
logKCAP(t-1)	.004*	.025***	.001	.002	.007*	-.001
	(.002)	(.007)	(.006)	(.003)	(.004)	(.006)
Growth(t-1)	.016**	.002	.006	.019	.009	.032***
	(.007)	(.006)	(.015)	(.017)	(.018)	(.012)
logSize(t-1)	.002*	.011**	.002	.002	.006**	.00006
	(.001)	(.005)	(.004)	(.002)	(.003)	(.003)
N obs	1961	1961	371	624	611	355
R-sq.	0.082	0.244	0.168	0.105	0.119	0.105
Dependent variable dlogUTV/dt						
	OLS	FE	1st quartile	2^{nd} quartile	3rd quartile	4th quartile
logUTV(t-1)	-.094***	-.468***	-.107***	-.079***	-.085***	-.124***
	(.009)	(.018)	(.022)	(.018)	(.017)	(.025)
logKCAP(t-1)	.008***	.043***	.001	.007	.011***	.008
	(.002)	(.009)	(.006)	(.005)	(.004)	(.007)
Growth(t-1)	.018**	.008	.022	-.046**	.006	.052***
	(.009)	(.008)	(.016)	(.023)	(.024)	(.016)
logSize(t-1)	-.0003	.015*	-.0009	-.002	.0005	.002
	(.002)	(.008)	(.004)	(.004)	(.004)	(.004)
N obs	1870	1870	354	593	592	331
R-sq.	0.071	0.257	0.152	0.083	0.103	0.152

Dependent variable dlogRTV/dt						
	OLS	FE	1st quartile	2^{nd} quartile	3rd quartile	4th quartile
logRTV(t-1)	-.093*** (.009)	-.534*** (.018)	-.096*** (.021)	-.089*** (.018)	-.100*** (.018)	-.087*** (.027)
logKCAP(t-1)	.010*** (.004)	.036*** (.012)	.011 (.008)	.013** (.006)	.005 (.006)	.003 (.011)
Growth(t-1)	.012 (.013)	.005 (.011)	.009 (.022)	.075*** (.029)	.013 (.039)	-.015 (.028)
logSize(t-1)	.007*** (.002)	-.0009 (.010)	.008 (.006)	.015*** (.005)	.014*** (.005)	-.001 (.006)
N obs	1819	1819	340	591	563	325
R-sq.	0.074	0.338	0.125	0.099	0.134	0.129
Note : *** : p<0.01 ; ** : p<0.05 ; * : p<0.1 ; all regressions include time and sector dummies.						

Table 4 - Growth and Knowledge Structure (2-year lagged growth rates)

Dependent variable dlogKOH/dt						
	OLS	FE	1st quartile	2^{nd} quartile	3rd quartile	4th quartile
logKOH(t-1)	-.230***	-.751***	-.222***	-.244***	-.208***	-.237***
	(.012)	(.020)	(.027)	(.023)	(.021)	(.026)
logKCAP(t-1)	.011**	.042***	.0213*	.0129	.015**	.004
	(.004)	(.016)	(.011)	(.009)	(.007)	(.012)
Growth(t-2)	-.027	-.008	-.049	.069	.001	-.071**
	(.019)	(.019)	(.038)	(.065)	(.032)	(.035)
logSize(t-1)	-.010***	-.009	-.013*	-.021***	-.011*	-.003
	(.003)	(.014)	(.008)	(.008)	(.006)	(.007)
N obs	2576	2576	476	768	822	510
R-sq.	0.140	0.393	0.194	0.160	0.141	0.188
Dependent variable dlogCD/dt						
	OLS	FE	1st quartile	2^{nd} quartile	3rd quartile	4th quartile
logCD(t-1)	-.278***		-.196***	-.381***	-.108***	-.374***
	(.020)		(.052)	(.037)	(.040)	(.041)
logKCAP(t-1)	-.001		.004	-.003	-.007	.004
	(.004)		(.009)	(.008)	(.008)	(.010)
Growth(t-2)	.029*		.027	.078	.026	.003
	(.017)		(.028)	(.056)	(.029)	(.036)
logSize(t-1)	$2.70 \cdot 10^{-6}$		.002	-.004	.001	-.001
	(.003)		(.007)	(.007)	(.007)	(.006)
N obs	1317		220	400	435	262
R-sq.	0.1386		0.200	0.257	0.066	0.325
Dependent variable dlogTV/dt						
	OLS	FE	1st quartile	2^{nd} quartile	3rd quartile	4th quartile
logTV(t-1)	-.041***	-.385***	-.038**	-.018	-.065***	-.039**
	(.008)	(.018)	(.017)	(.014)	(.017)	(.019)
logKCAP(t-1)	.0001	.026***	-.002	-.002	.003	-.003
	(.002)	(.008)	(.006)	(.004)	(.006)	(.008)
Growth(t-2)	.006	-.013*	.021	.032	-.016	.008
	(.007)	(.007)	(.015)	(.004)	(.014)	(.013)
logSize(t-1)	.002*	.014***	.0004	.003	.008***	.0003
	(.001)	(.005)	(.003)	(.002)	(.003)	(.003)
N obs	2343	2343	433	710	742	458
R-sq.	0.058	0.255	0.132	0.094	0.081	0.072
Dependent variable dlogUTV/dt						
	OLS	FE	1st quartile	2^{nd} quartile	3rd quartile	4th quartile
logUTV(t-1)	-.106***	-.478***	-.102***	-.103***	-.106***	-.114***
	(.010)	(.019)	(.021)	(.019)	(.017)	(.024)
logKCAP(t-1)	.010***	.040***	.0002	.014***	.011**	.008
	(.003)	(.009)	(.006)	(.005)	(.005)	(.008)
Growth(t-2)	-.011	-.021**	.006	-.021	-.059***	.016
	(.009)	(.010)	(.018)	(.031)	(.017)	(.017)
logSize(t-1)	-.0003	.020**	-.001	-.002	.003	.002
	(.002)	(.008)	(.003)	(.004)	(.003)	(.003)
N obs	2195	2195	401	668	704	422
R-sq.	0.070	0.268	0.159	0.084	0.107	0.1254

Dependent variable dlogRTV/dt						
	OLS		1st quartile	2^{nd} quartile	3rd quartile	4th quartile
logRTV(t-1)	-.102*** (.010)	-.518*** (.018)	-.093*** (.021)	-.096*** (.018)	-.105*** (.019)	-.109*** (.025)
logKCAP(t-1)	.017*** (.004)	.040*** (.013)	.018* (.009)	.007 (.007)	.021*** (.008)	.020* (.012)
Growth(t-2)	.030** (.013)	.024* (.013)	.008 (.028)	.164*** (.041)	.038* (.022)	-.005 (.027)
logSize(t-1)	.006*** (.002)	-.006 (.010)	.007 (.005)	.014*** (.005)	.008* (.005)	-.002 (.005)
N obs	2107	2107	382	644	674	407
R-sq.	0.068	0.312	0.118	0.118	0.0915	0.137
Note : *** : p<0.01 ; ** : p<0.05 ; * : p<0.1 ; all regressions include time and sector dummies.						

Appendix A - Sectoral classification and concordance

Macro sectors	Sector	STAN (ISIC 3)	Datastream
	Pharmaceuticals	2423	4577
High-technology manufactures	Office, accounting and computing machinery	30	9572, 9574
HT	Radio, television and communication equipment	32	2737, 3743, 3745,3747,9576, 9578
	Medical, precision and optical instruments	33	4535, 4537, 4573
	Aircraft and spacecraft	353	2713, 2717
	Chemicals excluding pharmaceuticals	24ex2423	1353, 1357
Medium-high technology manuf.	Machinery and equipment, n.e.c.	29	573, 583, 2757
MHT	Electrical machinery and apparatus, nec	31	2733, 3722
	Motor vehicles, trailers and semi-trailers and other transport equipment, aircraft excluded	34, 351, 352-359	2753, 3353, 3355
	Coke, refined petroleum products and nuclear fuel	23	533, 537, 577, 587
Medium-low technology manuf.	Rubber, plastics products and other non-metallic mineral products	25-26	2353, 2723, 3357
MLT	Basic metals and fabricated metal products	27-28	1753, 1755, 1757
	Food products and beverages	15	3533, 3535, 3537, 3577
	Tobacco products	16	3785
Low technology manufactures	Textiles, textile products, leather and footwear	17-19	3763, 3765
LT	Pulp, paper and paper products	21	1737
	Printing and publishing	22	5557
	Manufacturing nec and recycling	36-37	2727, 3724, 3726, 3767
	Post and telecommunications	64	5553, 6535, 6575
	Financial intermediation (excl insurance, pension)	65	8355, 8773, 8779
Knowledge intensive sectors	Insurance and pension funding	66	8532, 8534, 8536, 8538, 8575
KIS	Activities related to financial intermediation	67	8775, 8777, 8985, 8995
	Real estate activities	70	8633, 8637, 8671, 8672, 8673, 8674, 8675, 8676, 8677, 8771
	Renting of m&eq and other business activities	71-74	2791, 2793, 2795, 2799, 5555, 9533, 9535, 9537
	Health and social work	85	4533
	Recreational cultural and sporting activities	92	5752, 5755
Less knowledge intensive sectors	Wholesale, trade (excl. Motor vehicles)	51	2797, 5379
LKIS	Retail trade; repair of household goods	52	5333, 5337, 5371, 5373, 5375
	Hotels and restaurants	55	5753, 5757
Other services	Transport and storage	60-63	2771, 2773, 2775, 2777, 2779, 5751, 5759
OS	Community social and personal services	75-99	5377
Energy producing activities	Mining, quarrying of energy producing materials	10-12	1771
EP	Mining, quarrying (excl energy)	13-14	1773, 1775, 1777, 1779
	Electricity, gas, and water supply	40-41	7535, 7537, 7573, 7575, 7577
Constr	Construction	45	2357, 3728