

High Growth Firms and Technological Knowledge: Do gazelles follow exploration or exploitation strategies?¹

Alessandra Colombelli^{a,b} – Jackie Krafft^b – Francesco Quatraro^{b*}

- a) Dipartimento di Sistemi di Produzione ed Economia dell’Azienda, Politecnico di Torino
- b) University of Nice Sophia Antipolis and CNRS-GREDEG

ABSTRACT.

This paper analyzes the contribution of high-growth firms to the process of knowledge creation. We articulate a demand-pull innovation framework in which knowledge creation is driven by sales growth, and knowledge stems from creative recombination. Given the established literature on high growth firms and economic growth, we wonder whether gazelles follow patterns of knowledge creation mostly dominated by exploration or exploitation strategies. To this purpose, we derive indicators able to describe the structure of knowledge and qualify firms’ innovation strategies. The empirical results suggest that the reality is richer than the interpretative frameworks. Increasing growth rates are indeed associated to exploration strategies, supporting the idea that high growth firms are key actors in the creation of new technological knowledge. But in the meantime, firms showing growth rates significantly higher than the average are able to command the exploration strategies by constraining them within the boundaries of familiar technological competences, suggesting that the exploration process is less random than anticipated. We end up with the result that high growth firms, and especially gazelles, follow predominantly an exploration strategy, but with the characteristics of an organized search which is often more observed in an exploitation strategy.

Keywords: Gazelles; Recombinant Knowledge, Schumpeterian innovation patterns

JEL Classification Codes: L20, L10, O32

*Corresponding address : University of Nice Sophia Antipolis, CNRS-GREDEG. 250, rue Albert Einstein, 06560 Valbonne, France. Email: francesco.quatraro@unice.fr

¹ A preliminary version of this paper has been discussed at the European Meeting on Applied Evolutionary Economics (EMAEE) held in Pisa in February 2011, the Workshop on “High-Growth Firms” organized at the RATIO Institute in Stockholm on the 19th-20th May 2011, the second conference organized by the Journal of Industrial and Business Economics in Parma on the 20th-21st June 2011, the seminar series of the GREDEG-CNRS, University of Nice and at the SEIKE seminar series organized at the BRICK, Collegio Carlo Alberto. We wish to thank Cristiano Antonelli, Giulio Bottazzi, Alex Coad, Lionel Nesta, Marco Vivarelli and Pier Paolo Patrucco for their useful comments. 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 predominantly a 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 Hözl, 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 investigate. 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, 2003 and 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 either small or large firms as well as either mature 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 rethink 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 Hözl (2011), the empirical studies of firms' growth concern 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 conducted mostly 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 to deal with the issue of heterogeneous impacts of explanatory variables across different classes. Within this strands of analysis, Coad and Rao (2008) propose an alternative way to proxy innovation and analyze its effects on firms' growth, and find that innovation is of crucial importance for high-growth firms. On a different ground, Coad and Rao (2010) show 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 in R&D than others, supporting the conclusion that gazelles derive much of their success from the exploitation of local comparative advantages.

Most importantly, the studies on the relationship between high-growth and innovation use 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 (see 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 is still 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. Within this line of reasoning, 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 particular, we wonder to what extent gazelles can be thought as featuring the population of firms in sectors dominated by Mark I or Mark II Schumpeterian patterns of innovation (Malerba and Orsenigo, 1995, 1997). The issue of the degree to which gazelles can be considered as hybrids in terms of innovation patterns represents therefore the key research question of the current paper. Our results support the idea that gazelles do not strictly belong to one mode or the other, but rather appear as a mixed population of firms adopting a combination of exploration attitude with organized search strategies.

We argue that a dynamic view of Schumpeterian patterns of innovation has to be developed in order to understand how a gazelle can jump from one category to the other, or extend the characteristics of one category closer to the other category. In order to do that, we focus predominantly on the type of strategy that high growth firms and especially gazelles develop (exploration versus exploitation) when they create new technological knowledge, and how they perform this strategy (i.e. in a random versus more organized way). From this framework, we produce empirical evidence and draw policy implications.

For this purpose, we will graft into a demand-pull innovation background 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, we wonder whether gazelles are really key actors in the creation of technological variety within the economy, bearing the risk of exploring the technology landscape towards areas far from their core competencies and therefore with uncertain outcomes. In our paper, gazelles will be identified as firms showing average sales growth rates higher than 20% over the observed time span, 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?

The literature on the inducement of innovation is broad and rich of empirical contributions. Such literature is traditionally divided into two large sets, one emphasizing the importance of the accumulation of skills and scientific knowledge as main drivers, the other emphasizing the role of economic mechanisms on the demand side. The former approach is usually referred to as technology-push, while the second is demand-pull. The adoption of this latter framework in the analysis of the differential innovation behavior of high growth firms can be far reaching, above all when combined with the collective knowledge approach and the literature on Schumpeterian patterns of innovation.

The pioneer of the demand-pull approach in its modern form is undoubtedly Jacob Schmookler². He observed how series on technology creation as proxied by patent applications tend to follow series on output (Schmookler, 1954 and 1962). The suggested interpretation of this evidence was grounded on the idea that “more money will be available for invention when the industry’s sales are high than when they are low. Increased sales imply that both the producing firms and their employees will be in a better position than before to bear the expenses of invention” (Schmookler, 1962: p.17). In this framework, the ability to finance the activities of knowledge creation plays a central role (Schmookler, 1966). The empirical analyses of the effects of firms’ performances on knowledge creation have however been confined to the level of innovation, without any attempt to qualify the patterns of innovation (Griliches and Schmookler, 1963; Scherer, 1982).

² Of course, the seeds of such argument can be dated back to as a remote contribution as that of Adam Smith (1776), who emphasized the indirect effects on increasing demand on technological change through the positive effects of division of labour. Such argument has been further developed and integrated by Marshall (1890) and Young (1928).

In this respect, the identification of two distinct Schumpeterian patterns of innovation by Malerba and Orsenigo (1995 and 1997) provides useful insights. On the one hand, Schumpeter Mark I is characterized by creative destruction, ease of entry and appearance of new firms based on business opportunities, which challenge incumbents and continuously disrupt the current ways of production, organization and distribution. On the other hand, Schumpeter Mark II is characterized by 'creative accumulation', the relevance of industrial R&D labs and the key role of large firms. The authors 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 existing technological premises.

In this direction we wonder whether gazelles may be identified as a phenomenon characterizing the technological regime typical of the Schumpeter Mark I or rather of the Schumpeter Mark II.

Therefore, within this framework we seek to investigate the role that high-growth firms play in the advancement of technological knowledge, and above all the way they do it, i.e. by trying to understand if they operate mostly according to the 'widening' pattern characterizing Schumpeter Mark I, or according to the 'deepening' pattern typical of Schumpeter Mark II.

Now, we argue that 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 it 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 for 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 how we can characterize the internal structure of the knowledge base. Basically, this characterization takes into account 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 that lead us to derive three main properties of knowledge structure at a general level:

- **Variety** is 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.
- **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.

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 approach captures both the cumulative character of knowledge creation and the key role played by the properties describing knowledge structure, as well as the possible link 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 combination of the demand-pull framework with the recombinant knowledge approach and the analysis of the Schumpeterian patterns of gazelles' innovation activities allows us to refine our main working hypotheses as follows.

Sales growth are a key factor engendering high levels of innovations. For this reason, gazelles are expected to be characterized by demand-driven dynamics of knowledge creation. This latter can be in turn implemented by means of search behaviors aimed either at widening or deepening the technological competences of the firm. Our main research question is whether the important contribution of gazelles to economic growth can be ascribed to the undertaking of search behaviors typical of a Schumpeter Mark I pattern of innovation activities, or possibly to a Schumpeter Mark II. According to the first pattern, the positive impact of high-growth firms would be due to their capacity to undertake search behaviors directed towards the exploration of untried technological fields, so as to widen the existing knowledge base initially in a rather random way. The extension of the knowledge base would indeed be possible only by going beyond the fences of what firms already

know. Exploration would tend therefore to be a key part of the destructive creativity of gazelles within the 'widening' pattern. The search behavior of high-growth firms would be expected to depart to some extent from the established trajectories to discover new fields in the technology landscape so as to enlarge their search scope. According to the second pattern, search behaviors would be more focused on a range of 'successful' technological fields, leading to a deepening of the existing knowledge base. Exploitation is intended to combine knowledge in a more organized way, and high growth firms could play an active role in it.

We can try to map the different paths that gazelles may follow (see Figure 1), by distinguishing i) the strategies that they elaborate (exploration versus exploitation) and ii) the way in which they predominantly implement these strategies (random search versus organized search). We end up thus with a fourparts diagram, with one horizontal axis (strategies) and one vertical axis (type of search).

On this diagram, we can visualize the typical Schumpeterian Mark 1 and Mark 2 patterns of innovation (1st and 3rd quarter). In Mark 1, firms are depicted as developing different characteristics of a product innovation, in a situation of high uncertainty that imply the predominance of a trial and error system of development; while in Mark 2, firms take advantage of their experience and lower uncertainty to select the most successful ways to produce innovation. After a period of exploration where firms try any possible combination to produce innovation, then comes a situation of more stabilized choices around a smaller set of possibilities.

>>> INSERT FIGURE 1 ABOUT HERE <<<

Now, the interest in gazelles is that they do not necessarily belong to pure models of innovation patterns. Due to their multifaceted characteristics in terms of size, innovation behavior, etc., it is necessary to consider them in a dynamic framework where they may evolve from one model to the other (say for instance from Mark 1 to Mark 2), or extend one model to some characteristics that usually belong to the other model. For instance, gazelles may be small firms, highly oriented towards a model of innovation by exploration, but the manner in which they follow this strategy may be more organized than predicted in Mark 1. Alternatively, one may also find gazelles that are large in size, and experimenting predominantly an exploitation strategy (i.e. Mark 2 characteristics), but with some random screening properties in the way they combine pieces of knowledge that are usually exclusively attributed to Mark 1.

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 other relevant variables, 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 digits level (in Appendix B we provide the sectoral concordance table used to link the two classifications).

Our final dataset consists of an unbalanced panel of 335 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 30% and 37% observations, respectively. Medium low and low technology include 3% and 10% while knowledge intensive sectors cover about 7% of observations. Finally, each of the other economic groups includes around or less than 10% observations.

>>>INSERT TABLE 1 ABOUT HERE<<<

Since in this paper we are interested in the dynamic aspects of the relationship between sales and knowledge creation, we will use growth rates of all the relevant variables. On a general level, growth rates are defined as follows:

$$Growth_{i,t} = \ln(X_{i,t}) - \ln(X_{i,t-1}) \quad (1)$$

Where X is measured in terms of sales; knowledge capital stock; knowledge coherence; cognitive distance; knowledge variety; related knowledge variety; unrelated knowledge variety. All these variables, which will be introduced better explained in what follows and in appendix A, are calculated for firm i at time t . Following previous empirical works (Bottazzi et al, 2010; Coad, 2011), the growth rates distributions have been normalized around zero in each year by removing means as follows:

$$s_{i,t} = Growth_{i,t} - \frac{1}{N} \sum_{i=1}^N Growth_{i,t} \quad (2)$$

Where N stands for the total number of firms in the sample. This procedure effectively removes average time trends common to all the firms caused by factors such as inflation and business cycles.

Figure 2 shows the distribution of firms' growth rates for all the relevant variables. 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).

>>>INSERT FIGURE 2 ABOUT HERE<<<

Such evidence suggests that standard regression estimators, like ordinary least squares (OLS), assuming Gaussian residuals may perform poorly if applied to these data. To cope with this, a viable and increasingly used alternative consists of implementing the least absolute deviation (LAD) techniques, which are based on the minimization of the absolute deviation from the median rather than the squares of the deviation from the mean.

Figure 3 illustrate the distribution of firm's sales growth by macro-sector (see Appendix B for the details of macro-sectors definition). 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, while knowledge intensive sectors show highly dispersed growth rates.

>>>INSERT FIGURE 3 ABOUT HERE<<<

Descriptive statistics for the knowledge indicators and the other variables included in our model are shown in Table 2. The reported variables are growth rates normalized according to Equation 2. The values on kurtosis as well as those on the percentiles confirm that growth rates are characterized by highly skewed distributions.

>>>INSERT TABLE 2 AND TABLE 3 ABOUT HERE<<<

In Table 3 we show instead the matrix of correlations among the variables we use in the empirical exercise, marking a significance level of 1%. Although some significant pattern of correlation can be identified, these involve mostly the variety-related variables, and in any case the coefficients are not high enough to generate too severe concerns (except for the correlation of the three variety measures that, as one could expect, are characterized by non-negligible correlations).

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, 2003 and 2009). Moreover, previous works find that growth rates are autocorrelated.

However, the originality of this paper is that we propose to reverse the line of reasoning, by adopting a demand-pull approach in which sales growth provide the incentive to commit resources to knowledge creation activities. In this direction our empirical strategy is different from the empirical exercises in the field following the seminal contributions by Griliches and Schmookler (1963) and Scherer (1982). We indeed carry out a direct test of the effect of sales growth rates on knowledge creation emphasizing thus the demand pull side of innovation. Moreover, another element of originality lies in the fact that we are not that much interested in understanding whether increasing sales do affect the level of knowledge creation, but we are mostly focused on the qualitative aspects of the knowledge creation process, by resting on the properties of knowledge structure (i.e. variety, coherence and similarity), the empirical implementation of which will be described in the next section. This differentiates the analysis provided in this paper both from the contributions within the

Schmooklerian tradition, and from those contributions aiming at emphasizing the relative weak effect of firms' sales on their R&D intensity (see for example Pakes and Schankerman, 1977 and 1984).

Given the purposes of this paper, the question remains as to which extent knowledge is a determinant of firms' growth or the reverse. Moreover, the properties of knowledge structure may be related with one another and with the level of knowledge creation. For this reason an empirical strategy looking at the coevolution of the series can be the most useful in order not to impose any aprioristic relationship amongst the variables at stake. In order to better grasp the potential co-evolutionary patterns of such interdependent variable we implement the analysis within the context of a (reduced form) vector autoregression (VAR) model (Coad, 2010). In order to introduce the VAR model, let us recall the generic operational definition of the variables we will use in the analysis s_{it} , i.e. the growth rate detrended through normalization. The baseline VAR model can be then written as follows:

$$s_{i,t} = \alpha + \beta s_{i,t-1} + \varepsilon_{i,t} \quad (3)$$

Where s_{it} is an $m \times 1$ vector of random variables for firm i at time t , β is an $m \times m$ matrix of slope coefficients that are to be estimated. In our particular case $m=7$ and corresponds to the vector [sales growth (i,t), growth of knowledge capital (i,t), coherence growth (i,t), growth of cognitive distance (i,t), variety growth (i,t), related variety growth (i,t), unrelated variety growth (i,t)]. ε is an $m \times 1$ vector of disturbances.

Since we are interested in verifying the differential impact of high growth firms, we resorted to indicator variables. In particular, we first build a simple dummy variable, which we called HGF, which identify gazelles as those firms showing average growth rates of at least 20% over the whole observed period³. This allows to investigating any shift in the intercept in the estimated equations when gazelles are concerned. We then calculated the interaction variable [HGF \times sales growth(t-1)], which allows to detecting any modification in the slope coefficient of sales growth as far as gazelles are concerned.

It must be noted that we do not include any individual dummies in the analysis. Even though unobserved heterogeneity due, for example, to sector of activity, location, etc. , can have important effects on the estimation results, the inclusion of individual dummies along with lagged variables may engender some biases for fixed-effect estimation of dynamic panel-data models, a problem known as Nickell-bias. Some alternative approaches relate to the use of instrumental variable (IV) or GMM estimators (Blundell and Bond, 1998). The main problem with this lies in the difficulty to find out good instruments, which is particularly hard when dealing with growth rates. When instruments are weak, IV estimation of panel VAR thus leads to imprecise estimates. Binder et al. (2005) propose instead a panel VAR model including firm-specific effects, which is however based on the assumption of normally distributed errors, which is not the case for what concerns the growth rates of the variables used in our regressions.

³ Unfortunately, due to data limitations, firms' age is not available, and thus this information cannot be used in the definition of high growth firms.

Since we are dealing with growth rates, instead of levels, we can maintain that any firm-specific component has been largely removed. Moreover, we follow the wide body of literature on the analysis of firms' growth rates stating that the non-Gaussian nature of growth rate residuals are a far more important econometric problem deserving careful attention.

In view of this, equation (3) is estimated via reduced form VARs, so as to avoid any ex ante definition of the causal structure on the relationships between the variables. Such reduced forms VARs in fact correspond to a series of m individual regressions. Given the distributional properties of the variables, we will prefer the implementation of LAD estimators.

3.3 The Implementation of Knowledge Indicators

For what concerns the definition of the knowledge related variables, 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 obsolescence⁴.

The implementation of knowledge characteristics proxying for variety, coherence and similarity, 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 avenue⁵. 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 patent document, we will refer to the link between j and m as the co-occurrence of both of them within the same patent document⁶. On this basis we calculated the following three key characteristics of firms' knowledge bases (see the appendix A for the methodological details):

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

⁵ 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).

⁶ It must be stressed that to compensate for intrinsic volatility of patenting behaviour, each patent application is made last five years.

- a) Knowledge variety (KV) measures the degree of technological diversification of the knowledge base. It is based on the information entropy index, and it can be decomposed in related knowledge variety (RKV) and unrelated knowledge variety (UKV).
- b) Knowledge coherence (COH) measures the degree of complementarity among technologies.
- c) Cognitive distance (CD) expresses the dissimilarities amongst different types of knowledge.

The adoption of these variables marks an important step forward in the operational translation of knowledge creation processes. In particular, they allow for a better appreciation of the collective dimension of knowledge dynamics. Knowledge is indeed viewed as the outcome of a combinatorial activity in which intentional and unintentional exchange among innovating agents provides the access to external knowledge inputs (Fleming and et al., 2007). The network dynamics of innovating agents provide the basis for the emergence of new technological knowledge, which is in turn represented as an organic structure, characterized by elementary units and by the connections amongst them. The use of such variables implies therefore a mapping between technology as an act and technology as an artefact (Arthur, 2009; Lane et al., 2009; Krafft and Quatraro, 2011). Co-occurrences matrixes are very similar to design structure matrixes (DSM) (Baldwin and Clark, 2000; Murmann and Frenken, 2006; Baldwin, 2007), in that they can be thought as adjacency matrixes in which we are interested not only in the link between the elements, but also by the frequency with which such links are observed.

In other words these measures capture the design complexity of knowledge structure, and allow for featuring the innovation behaviour of firms, as well as its evolution, in relation with the changing architecture of such structure (Henderson and Clark, 1990; Murmann and Frenken, 2006). In this perspective, an increase in knowledge coherence is likely to signal the adoption of an exploitation strategy, while a decrease is linked to exploration strategies. Increasing values of cognitive distance are instead related to random screening across the technology landscape, while decreasing cognitive distance is more likely to be linked to organized search behaviour. Knowledge variety is likely to increase in any case when new combinations are introduced in the system. However the balance between related and unrelated variety should be such that the related one is likely to dominate during exploitation phases, while the unrelated one gains more weight in the exploration strategies (Krafft, Quatraro, Saviotti, 2009). We can thus enrich our interpretative scheme proposed in Fig. 1, so as to obtain a guide to interpret the innovation behaviour of high growth firms.

>>> INSERT FIGURE 4 ABOUT HERE <<<

With this tool at hand, we can now move to the presentation and discussion of the results of the empirical analysis.

4 Empirical results and Discussion

We report in Table 4 the results of the VAR estimation for the baseline model. Each row corresponds to a regression making the system of the vector autoregression model. Thus, the columns provide us with the differential effect of each (lagged) explanatory variable on each dependent variable. For the purposes of the present paper the first column represents the most important element, in that it

shows the effect of lagged sales growth rates on all the seven variables in the model. However, before focusing on this column, it is fair to note that all the coefficients along the diagonal of the matrix are statistically significant. Moreover, they are almost all negative, with the exception of sales and knowledge capital. This would suggest that while little persistence can be found in the case of the growth of knowledge properties, we can observe some persistence at work with sales and even more with knowledge capital. This result is not in line with previous works (Coad, 2007 and 2011), where the lagged values of sales growth proved to negatively affect the present values. However, this difference can be due to the peculiarity of the sampled firms, which are all listed firms on the main European markets, and therefore likely interested by positive performances on financial markets.

>>> INSERT TABLE 4 ABOUT HERE <<<

For what concerns the coefficients in the first column, we can observe first of all a positive relationship between lagged sales growth and knowledge capital. This is a further support to the demand pull hypothesis according to which increasing sales provide firms with the necessary amount of resources to commit to the production of technological knowledge. When we move to the effects on the properties of knowledge structure, we find that sales growth are negatively related to knowledge coherence and cognitive distance, while positively with total variety. This suggests that the more firms grow, the more they are likely to introduce variety in their technology portfolio. Such technological variety is characterized by low levels of complementarity (coherence), which support the idea of an exploration behavior, but at the same time low levels of dissimilarity. This would suggest that increasing sales growth is associated by exploration behaviours which are directed towards the discovery of new complementary fields, which are however not too far from the technological competences of the firm.

In order to understand if any peculiar pattern can be devised by high-growth firms, table 5 includes in the dummy HGF as a regressor in all the VAR equations. Thus now we are interested in the first two columns of the table. At a general level the results do not vary that much for what the patterns of persistence of the observed variables. It is interesting to note that also the effects of sales growth on the present patterns are similar to those showed in the previous table. Increasing growth rates stimulated increasing growth rates of knowledge capital and of knowledge variety, which in turn is achieved by adopting search strategies directed towards the exploration of new complementary fields which are however consistent with the technological capabilities of the firm.

>>> INSERT TABLE 5 ABOUT HERE <<<

When we look at the HGF dummy, we see that, of course, the coefficient on sales growth is positive and significant. The same applies for what concerns growth rates of knowledge capital, suggesting that the economic influences on knowledge creation operating via the demand side are even stronger in the case of gazelles. When we move to the effects on the properties of knowledge structure, we find the dummy is significant only in the case of cognitive distance, where the coefficient is negative. This suggests that, provided that increasing growth rates are associated to a reduction of cognitive distance, high growth firms tend to adopt search strategies characterized on average by lower levels of dissimilarity. The dummy does not show any other significant effects on knowledge variables.

The picture is further enriched when we consider the interaction variable [sales growth(t-1)×HGF], as we do in Table 6. Once again, the inclusion of such variable does not change very much the patterns of persistence of the growth rates for the variables under scrutiny, according to which the coefficient relating the lagged to the present value of each of the properties of knowledge structure is negative and significant, while a positive and significant sign is found for knowledge capital. The most evident change concerns the coefficient for sales growth, which is again positive but not significant. Evidently, the significance of the coefficient is absorbed by the interaction variable (col 1, row 1).

>>> INSERT TABLE 6 ABOUT HERE <<<

The effect of lagged sales growth on knowledge is in line with the previous estimates, suggesting that increasing growth rates are likely to provide economic incentives to the commitment of resources to knowledge creation activities. Even in this case, the patterns of effects on the properties of the knowledge base confirm that increasing growth rates lead to the introduction of knowledge variety through exploration strategies directed towards the widening of complementary technologies (coherence) but similar to the competences in the technology portfolio of the firm. This interpretation is also supported by the positive and significant coefficient on related variety.

When we focus the attention on gazelles however, we find out interesting qualifications of this pattern. It is worth recalling that the inclusion of the interaction variable allows to appreciating the changes in the slope coefficient of sales growth when high growth firms are taken into account. The coefficient on knowledge coherence is positive and significant, although it is not such that it offsets the negative value of the coefficient of sales growth. The net effect of sales growth in the case of high growth firms remains therefore negative, but significantly lower (in absolute value) than the overall effect. The value and the negative sign on the variety variables also suggest that increasing growth rates in the group of gazelles are associated with a substantially stable technological variety.

In order to obtain the full picture of the differential patterns of knowledge creation in high growth firms, in Table 7 we report the results of the estimations including both the dummy and the interaction variable. As far as the effects of sales growth are concerned (col 1), we find out once again that to some extent persistence is at work in our sample. The effect on knowledge capital confirms to be positive and significant, providing further support to the existence of positive inducement dynamics on the demand-side. For what concerns the properties of knowledge structure, the sign on the coefficients for knowledge coherence and cognitive distance are both negative and significant, while those on variety and related variety are positive and significant. This corroborates the idea that increasing growth rates favor the introduction of technological variety channeled by a search behavior aiming at exploring different complementary domains, which are however not too distant from the competences firms have developed in their technological portfolio.

>>> INSERT TABLE 7 ABOUT HERE <<<

Let us now turn again to the subgroup of high growth firms, by focusing on columns 2 and 3. The coefficients on the dummy variable suggest that these engender higher growth rates of knowledge capital. Moreover, and most importantly, on average gazelles are characterized by lower growth rates of coherence, which is consistent with the tendency to exploration behaviors, although the negative coefficient on sales growth is mitigated by the one on the interaction variables. That is to say that high growth firms show on average a tendency to exploration, although in that subgroup

increasing growth rates do not lead to sensible variations in such a behavior. The value and the negative sign on the variety variables also confirm that increasing growth rates in the group of gazelles are associated with a substantially stable technological variety.

The picture emerging from the analysis supports the idea that gazelles can be hardly entirely placed within the Schumpeter Mark I or the Mark II innovation pattern. On the whole, they seem to be interested by an hybrid pattern of search behavior, characterized by both attitude towards exploration and the implementation of organized search strategies. Thus gazelles would be better featured by an organized exploration pattern, and placed in the bottom-left quadrant of the diagram proposed in Fig. 5. In this perspective, on average increasing sales growth rates stimulated the creation of new technological knowledge, by adopting search behaviors more directed to the screening of new complementary fields which are compatible with exploration strategies. However such screening is not that random, as it seems that fast growing firms prefer to move within a reasonable distance with respect to their technological competences. In the group of gazelles this pattern of behavior is even more marked, as these appear to be characterized on average by lower growth rates of coherence and cognitive distance, as well as of related variety. We thus end up with the result that high growth firms, and especially gazelles, follow predominantly an exploration strategy, but with the characteristics of an organized search which is often more observed in an exploitation strategy.

5 Conclusions

The main objective of this paper was to enquire into the differential contribution of high growth firms to the process of knowledge creation. In particular, we grafted the literature on Schumpeterian patterns of innovation into a demand-pull framework *à la* Schmookler in which the growth of sales provides the stimulus to create new technological knowledge. We thus wondered whether gazelles are more likely to undertake Mark I or Mark II patterns of knowledge creation.

The process of firms' growth has long attracted the attention of economists. More recently, a new strand of literature emerged focusing 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 investigation 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 emphasizes their crucial contribution to aggregate economic performances. We framed our analysis in a demand-pull innovation framework, in which increasing sales growth are viewed as key inducement to the creation of new technological knowledge. Instead of focusing merely on the level of knowledge creation, we wondered if high growth firms are characterized by peculiar patterns of demand-driven knowledge creation. In particular we wanted to investigate whether their search behaviour could be characterized mostly as of random screening or organized search. 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. variety

(related and unrelated), coherence, and similarity, which in turn can be usefully employed to distinguish between 'random screening' and 'organized search' strategies.

The analysis has been carried out by using data on listed companies and patent applications. We have implemented a series of VARs estimated by means of LAD estimators, including a dummy and interaction variable to detect differential performances of gazelles. The empirical results suggest that within the group of high growth firms, increasing growth rates of sales not only stimulate the creation of new technological knowledge, but also drive search behaviours characterized by the screening of complementary fields across the technology landscape, without however moving too far from the established technological competences of the firm.

Such results represent the first attempt to investigate the contribution of gazelles to the process of knowledge creation, and certainly 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, we think that such evidence provides an interesting basis for further investigations as well as for their policy implications. The common wisdom, according to which the proceeding on a higher growth path may be detrimental or costly at some stage in terms of innovation, is not confirmed in our work. In their process of growth, firms constantly experience a search process where new combinations are generated from their existing competences, and there is even some robust evidence that gazelles are also more innovative than other types of firms, even by controlling the randomness in their search process. In terms of policy, this gives some important arguments to consider higher growth firms as a distinctive population, outperforming both in terms of sales and in terms of innovation, thus deserving a specific attention. Our results call for the integration of innovation policies with industrial policies directed towards the support of high-growth firms so as, first, to strengthen the dynamics that lead gazelles to be innovative and, second, to take stock of their ability widening the knowledge base to increase further their performance in terms of sales. 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.

Finally, it would be indeed central to rethink the allocation mechanisms of government funding, going beyond the focus on firms' size to emphasize the important of growth dynamics. The issue here is plausibly to explore incentives and tax credit towards high growth firms which are transversal to all industries, and supporting an innovation policy that could be classified as an horizontal industrial policy. 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, exploration based on the screening of more complementary fields.

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, C., Krafft, J. and Quatraro, F., 2010, Recombinant Knowledge and Growth: The Case of ICTs, *Structural Change and Economic Dynamics*, 21(1), 50-69.
- Arthur, W.B., 2009, *The Nature of Technology. What It Is and How It Evolves*. New York, Free Press.
- Attaran, 1985, Industrial diversity and economic performance in U.S. areas. *The Annals of Regional Science* 20, pp. 44-54.
- Baldwin, C. Y., 2007, Where do transactions come from? Modularity, transactions, and the boundaries of firms, *Industrial and Corporate Change*, 17, 155-195.
- Baldwin, C. Y. and Carl, K.B., 2000, *Design Rules, Volume I, The power of Modularity*. Cambridge MA, MIT Press.
- Birch, 1979, *The job generation process*. Cambridge MA, MIT Press.
- Birch, 1981, Who creates jobs?, *The Public interest*, 65(fall), 3-14.
- Boschma R. and Iammarino, S., 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, SantAnna School of Advanced Studies, Pisa.
- Bottazzi, G. and Secchi, A., 2006, Explaining the Distribution of Firms Growth Rates, *Rand Journal of Economics*, 37, 234-263.
- Breschi S., Lissoni, F. and Malerba, F., 2003, Knowledge relatedness in firm technological diversification, *Research Policy*, 32, 69-97.
- Coad, A. and Hözl, W.; 2011, Firms growth: empirical analysis, in M. Dietrich and J. Krafft, *Handbook on the economics and theory of the firm*, Cheltenham: Edward Elgar.
- Coad, A. and Rao, R., 2008, Innovation and firm growth in high tech sectors: a quantile regression approach, *Research Policy*, 37(4), 633-648.
- Coad, A. and Rao, R., 2010, Firm growth and R&D expenditure, *Economics of Innovation and New Technology*, 19(2), 127-145.
- Coad, A., 2010, Exploring the processes of firm growth: evidence from a vector auto-regression, *Industrial and Corporate Change*, 19, 1677-1703.
- Coad, A., 2007, A Closer Look at Serial Growth Rate Correlation, *Review of Industrial Organization*, 31(1), 69-82.
- Daunfelt et al., 2010, When is Gibrats law a law?, Ration working papers 158, The Ratio Institute.
- Dosi, G., 1982, Technological paradigms and technological trajectories, *Research Policy*, 11, 147-162.
- Fleming, L., 2001, Recombinant Uncertainty in Technological Search, *Management Science* 47(1), 117-132.
- Fleming, L., Mingo, S. and Chen, D., 2007, Collaborative brokerage, generative creativity and creative success, *Administrative Science Quarterly*, 52, 443-475.

- Frenken et al., 2007, Related Variety, Unrelated Variety and Regional Economic Growth, *Regional Studies*, 41(5), 685-97.
- Frenken, K. and Nuvolari, A., 2004, Entropy Statistics as a Framework to Analyse Technological Evolution, in J. Foster and W. Hözl (eds), *Applied Evolutionary Economics and Complex Systems*. Edward Elgar: Cheltenham, U.K. and Northampton, Mass.
- Gibrat, 1931, Les inégalités économiques. Recueil Sirey, Paris.
- Griliches, Z., 1979, Issues in assessing the contribution of research and development to productivity growth. *The Bell Journal of Economics*. 10 92-116.
- Griliches, Z., 1990, Patent statistics as economic indicators: a survey, *Journal of Economic Literature*, 28, 1661-1707.
- Henderson, R.M and Clark, K.B, 1990, Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms, *Administrative Science Quarterly*, 35, 9-30.
- Henrekson and Johansson, 2010, Gazelles as Job Creators - A Survey and Interpretation of the Evidence, *Small Business Economics*, 35(2), 227-244.
- Hözl, W., 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.
- Hözl, W., 2010, The Economics of Entrepreneurship Policy: Introduction to the Special Issue, *Journal of Industry, Competition and Trade*, 10(3), 187-197.
- Jaffe, A., 1986, Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value, *American Economic Review*, 76(5), 984-1001.
- Jaffe, A., 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.
- Krafft, J. and Quatraro, F., 2011, The dynamics of technological knowledge: from linearity to recombination. In Antonelli, C. (ed) *Handbook on the Economic Complexity of Technological Change*. Cheltenham, Edward Elgar.
- Krafft, J., Quatraro, F. and Saviotti, P.P., 2009. Evolution of the knowledge base in knowledge intensive sectors. LEI-BRICK Working Paper no 06/2009.
- Krafft, J., Quatraro, F. and Saviotti, P.P., 2011, The knowledge base evolution in biotechnology: A social network analysis, *Economics of Innovation and New Technology*, 20, 445-477.
- Lane, D.A., van Der Leeuw, S.E., Pumain, D., West, G. (eds.), 2009, *Complexity perspectives in innovation and social change*. Springer, Berlin.
- Lee, 2010, A theory of firm growth: learning capability, knowledge threshold, and patterns of growth, *Research Policy*, 39, 278-289.
- Lotti, F., Santarelli, E. and Vivarelli, M., 2003, Does Gibrats Law Hold Among Young, Small Firms?, *Journal of Evolutionary Economics*, 13, 213-35
- Lotti, F., Santarelli, E. and Vivarelli, M., 2009, Defending Gibrat's Law as a Long-Run Regularity, *Small Business Economics*, 32, 31-44
- Malerba, F. and Orsenigo, L., 1995, Schumpeterian Patterns of Innovation, *Cambridge Journal of Economics*, Oxford University Press, vol. 19(1), pages 47-65, February.

- Malerba, F. and Orsenigo, L., 1997, The Dynamics and Evolution of Industries, *Industrial and Corporate Change*, Oxford University Press, vol. 5(1), pages 51-87.
- Mansfield, E., 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.
- Murmann, J.P. and Frenken, K., 2006, Towards a systematic framework for research on dominant designs, technological innovations, and industrial change, *Research Policy*, 35, 925-952.
- Nelson, R.R. and Winter, S.W., 1982, *An Evolutionary Theory of Economic Change*, Cambridge, Harvard University Press.
- Nelson, R.R., 1990, Capitalism 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, L. and Saviotti, P.P., 2006, . Firm Knowledge and Market Value in Biotechnology. *Industrial and Corporate Change*. 15(4) 625-52.
- Nesta, L. 2008, Knowledge and productivity in the worlds largest manufacturing corporations. *Journal of Economic Behavior and Organization*. 67 886-902.
- Nightingale, P., 1998, A cognitive model of innovation, *Research Policy*, 27, 689-709.
- Nooteboom, B., 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, K., 1985, Patent statistics as indicators of innovative activities: Possibilities and problems. *Scientometrics* 7, 77-99.
- Quatraro, F., 2010, Knowledge Coherence, Variety and Productivity Growth: Manufacturing Evidence from Italian Regions. *Research Policy* 39, 1289-1302.
- Saviotti, P.P., 1988, Information, variety and entropy in technoeconomic development, *Research Policy*, 17(2), 89-103.
- Saviotti, P.P., 2004, Considerations about the production and utilization of knowledge, *Journal of Institutional and Theoretical Economics*, 160, 100-121.
- Saviotti, P.P., 2007, On the dynamics of generation and utilisation of knowledge: The local character of knowledge, *Structural Change and Economic Dynamics*, 18, 387-408.
- Schumpeter, J.A., 1912, *The Theory of Economic Development*, Harvard University Press, Cambridge.
- Schumpeter, J.A., 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.

APPENDIX A – The properties of knowledge structure

Knowledge Variety

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):

$$H(X, Y) = \sum_{j=1}^q \sum_{m=1}^w p_{jm} \log_2 \left(\frac{1}{p_{jm}} \right) \quad (A1)$$

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 $m \in S_z$ ($g = 1, \dots, G$; $z = 1, \dots, Z$), we can rewrite $H(X, Y)$ as follows:

$$H(X, Y) = H_Q + \sum_{g=1}^G \sum_{z=1}^Z P_{gz} H_{gz} \quad (A2)$$

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:

$$H_Q = \sum_{g=1}^G \sum_{z=1}^Z P_{gz} \log_2 \frac{1}{P_{gz}} \quad (\text{A2a})$$

$$P_{gz} = \sum_{j \in S_g} \sum_{m \in S_z} P_{jm} \quad (\text{A2b})$$

$$H_{gz} = \sum_{j \in S_g} \sum_{m \in S_z} \frac{P_{ij}}{P_{gz}} \log_2 \left(\frac{1}{P_{jm} / P_{gz}} \right) \quad (\text{A2c})$$

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).

Knowledge Coherence

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} :

$$WAR_{jit} = \frac{\sum_{m \neq j} \tau_{jm} P_{mit}}{\sum_{m \neq i} P_{mit}} \quad (A3)$$

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

$$R_{it} = \sum_{j \neq m} WAR_{jit} \times \frac{P_{jit}}{\sum_j P_{jit}} \quad (A4)$$

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.

Cognitive Distance

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-centred correlation of the vectors V_{lk} and V_{jk} . The similarity of technologies l and j can then be defined as follows:

$$S_{lj} = \frac{\sum_{k=1}^n V_{lk} V_{jk}}{\sqrt{\sum_{k=1}^n V_{lk}^2} \sqrt{\sum_{k=1}^n V_{jk}^2}} \quad (A5)$$

The idea underlying the calculation of this index is that two technologies j and l 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:

$$d_{lj} = 1 - S_{lj} \quad (A6)$$

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.

$$WAD_{it} = \frac{\sum_{j \neq i} d_{ij} P_{jit}}{\sum_{j \neq i} P_{jit}} \quad (A7)$$

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:

$$CD_t = \sum_i WAD_{it} \times \frac{P_{lit}}{\sum_i P_{lit}} \quad (A8)$$

Figure 1

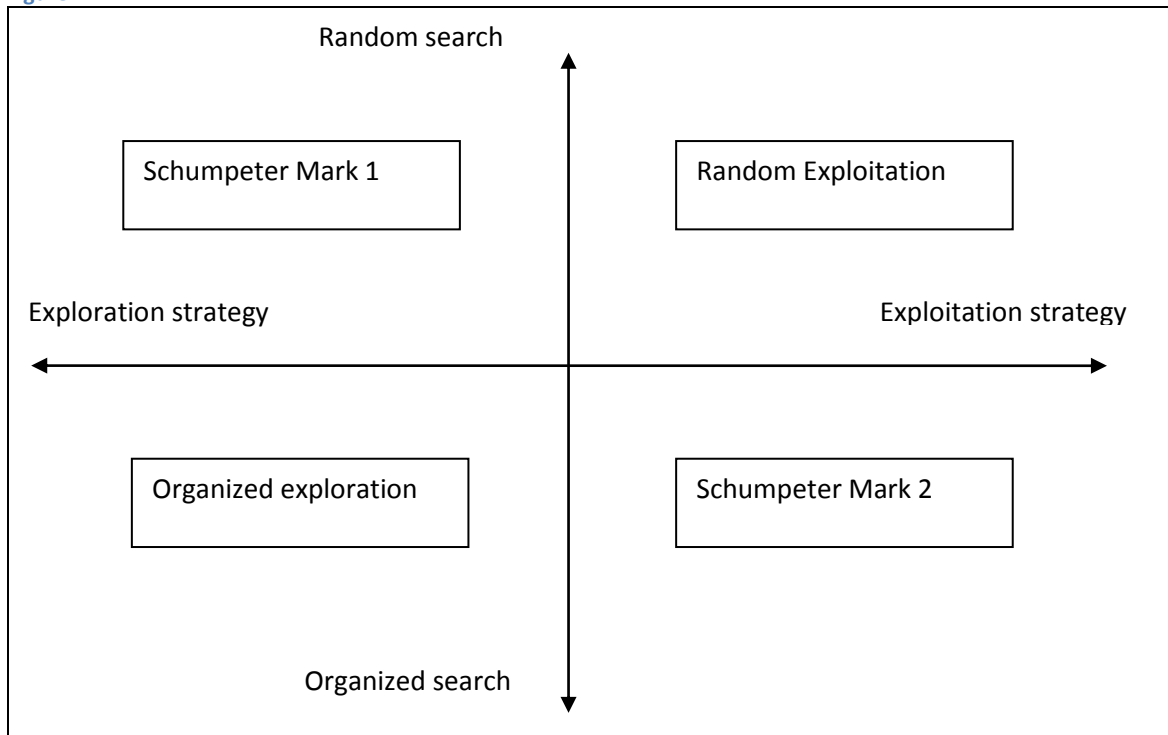


Figure 2 – Kernel density estimation of growth rates distribution of the main variables

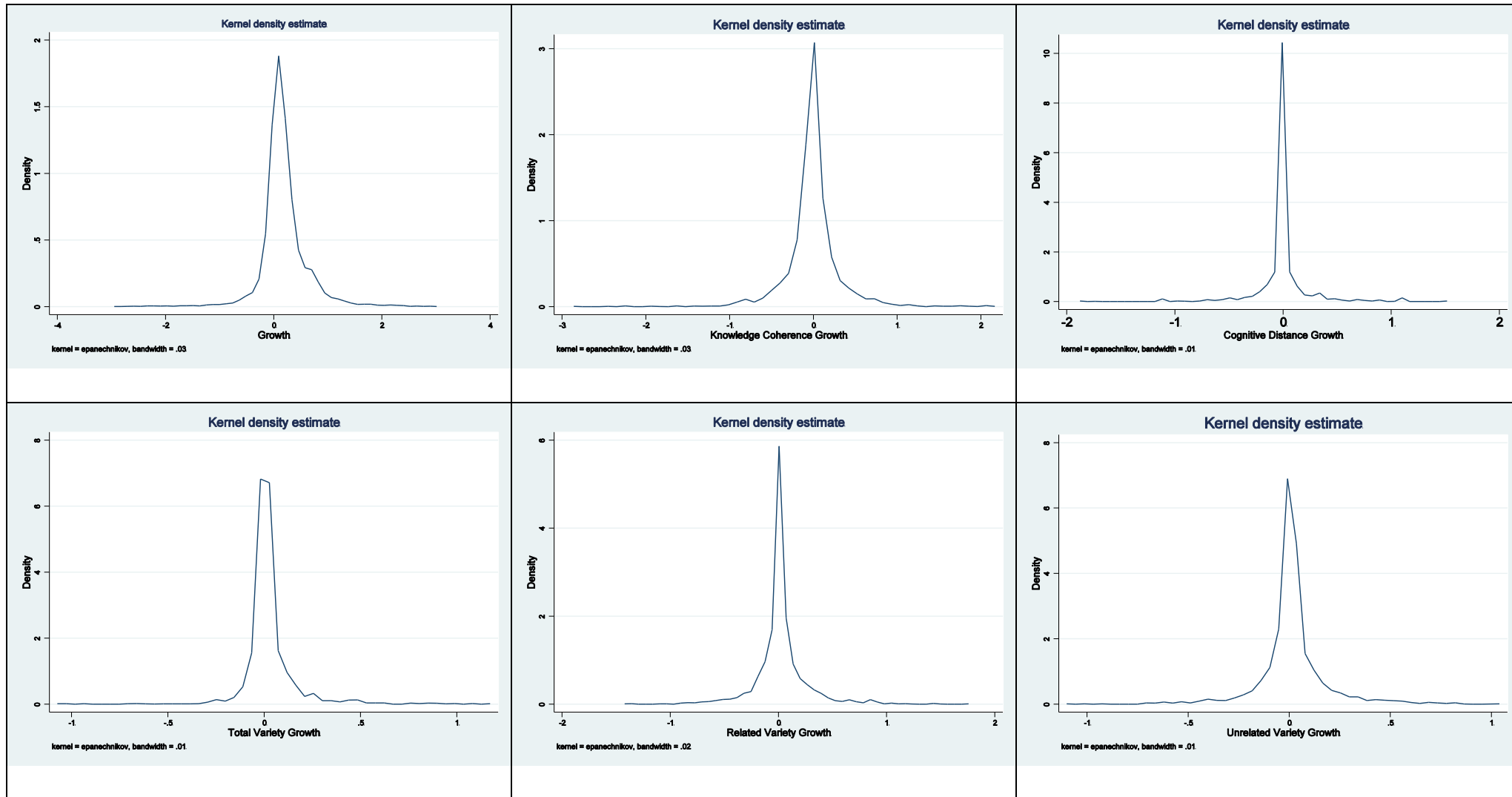
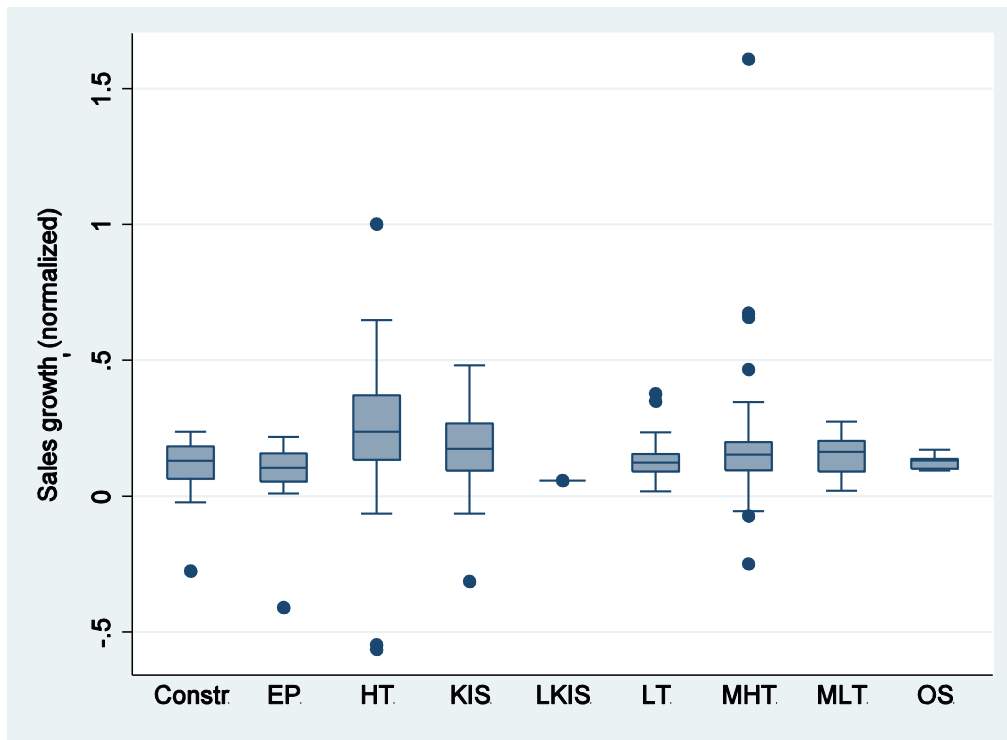


Figure 3 – Box plot of sales growth by macro-sector



Note: See Appendix B for the definition of macro-sectors.

Figure 4

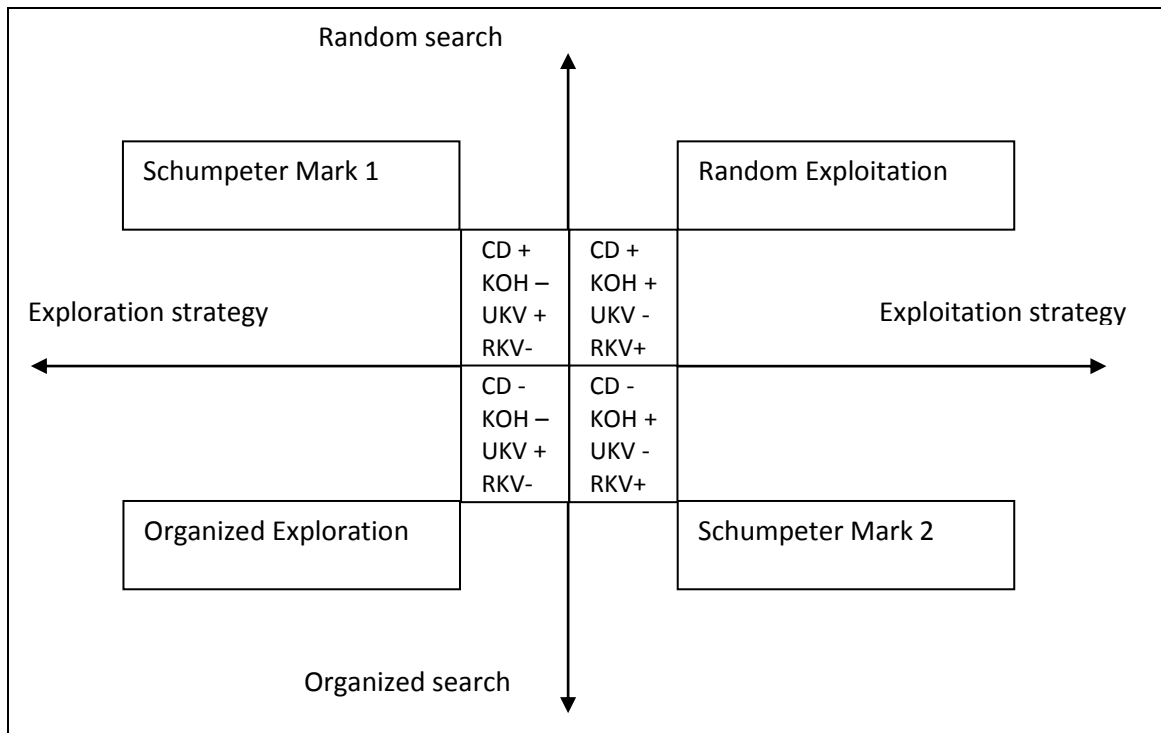


Table 1 – Descriptive statistics by macro-sector, 1988-2005

	Freq.	Percent	Cum.
HT	102	30.45	30.45
MHT	123	36.72	67.17
MLT	11	3.28	70.45
LT	34	10.15	80.6
KIS	26	7.76	88.36
LKIS	1	0.30	88.66
OS	5	1.49	90.15
Constr	23	6.87	97.02
EP	10	2.99	100
Total	335	100.00	

Table 2 - Descriptive statistics (All variables are expressed in normalized growth rates according to Eq. 2).

	Std. Dev.	kurtosis	p10	p25	p50	p75	p90	Obs.
Sales Growth	0.395	33.500	-0.099	0.012	0.112	0.250	0.557	2819
Knowledge Coherence	0.353	33.390	-0.288	-0.104	-0.007	0.089	0.285	2819
Knowledge Capital	0.178	13.146	-0.096	-0.032	0.041	0.138	0.272	2819
Cognitive Distance	0.242	11.531	-0.169	-0.032	-0.006	0.029	0.177	1448
Related Variety	0.132	25.337	-0.058	-0.018	0.004	0.033	0.103	2554
Unrelated Variety	0.167	10.986	-0.128	-0.031	0.006	0.043	0.151	2394
Knowledge Variety	0.231	10.983	-0.180	-0.049	0.001	0.067	0.235	2287

Table 3- Correlation Matrix

	Sales Growth	Knowledge Coherence	Knowledge Capital	Cognitive Distance	Related Variety	Unrelated Variety	Knowledge Variety
Sales Growth	1.000						
Kn. Coherence	0.009	1.000					
Knowledge Capital	0.073*	0.043	1.000				
Cognitive Distance	0.002	0.041	0.058	1.000			
Related Variety	-0.003	-0.047	0.393*	-0.072*	1.000		
Unrelated Variety	-0.020	-0.046	0.198*	-0.083*	0.478*	1.000	
Knowledge Variety	0.036	-0.020	0.213*	0.022	0.376*	-0.428	1.000

Table 4 – Results of VAR estimation, one-year lag. Baseline model (All variables are expressed in normalized growth rates according to Eq. 2).

	Sales Growth(t-1)	Knowledge Coherence (t-1)	Knowledge Capital (t-1)	Cognitive Distance (t-1)	Related Variety (t-1)	Unrelated Variety (t-1)	Knowledge Variety (t-1)	N. Obs.
Sales Growth	.101*** (.008)	.007 (.012)	-.103*** (.028)	.031* (.016)	-.026 (.039)	-.003 (.055)	.091 (.104)	1366
Knowledge Coherence	-.021*** (.007)	-.303*** (.009)	.035 (.022)	.016 (.013)	.072** (.031)	.113*** (.044)	-.282*** (.082)	1366
Knowledge Capital	.013*** (.003)	.006 (.005)	.699*** (.011)	.0001 (.006)	-.037** (.015)	-.044** (.022)	.119*** (.041)	1366
Cognitive Distance	-.008*** (.003)	.002 (.004)	.011 (.009)	-.013** (.005)	.004 (.014)	-.012 (.019)	.057 (.037)	1288
Related Variety	.0004 (.005)	.010 (.007)	.130*** (.017)	.020** (.010)	-.240*** (.024)	-.021 (.033)	.195*** (.063)	1366
Unrelated Variety	-.00006 (.004)	-.0001 (.005)	.004 (.013)	.003 (.008)	.042** (.019)	-.072*** (.027)	.015 (.051)	1366
Knowledge Variety	.005** (.002)	.001 (.003)	.081*** (.008)	-.010** (.004)	.026** (.011)	.036** (.016)	-.239*** (.030)	1366

Note: standard errors between parentheses. p<0.1; ** : p<0.05; *** : p<0.01.

Table 5 - Results of VAR estimation. One-year lag. Model including a dummy for HGFs (All variables are expressed in normalized growth rates according to Eq. 2).

	Sales Growth(t-1)	HGF (dummy)	Knowledge Coherence (t-1)	Knowledge Capital (t-1)	Cognitive Distance (t-1)	Related Variety (t-1)	Unrelated Variety (t-1)	Knowledge Variety (t-1)	N. Obs.
Sales Growth	.063*** (.009)	.218*** (.009)	.030** (.012)	-.097*** (.029)	.001 (.017)	.049 (.040)	.050 (.057)	-.066 (.107)	1366
Knowledge Coherence	-.020*** (.006)	-.008 (.006)	-.299*** (.008)	.031* (.019)	.018* (.011)	.062** (.026)	.094*** (.038)	-.262*** (.071)	1366
Knowledge Capital	.0137*** (.004)	.009** (.004)	.006 (.005)	.694*** (.012)	-.003 (.007)	-.025 (.018)	-.030 (.025)	.092** (.047)	1366
Cognitive Distance	-.009*** (.003)	-.005* (.003)	.002 (.004)	.011 (.009)	-.011** (.006)	-.006 (.014)	-.028 (.020)	.093** (.038)	1228
Related Variety	-.0002 (.005)	.001 (.006)	.010 (.007)	.129*** (.018)	.020* (.010)	-.236*** (.026)	-.018 (.036)	.187*** (.068)	1366
Unrelated Variety	-.0007 (.004)	.005 (.004)	.0004 (.006)	.010 (.014)	.0014 (.008)	.051*** (.020)	-.065** (.029)	-.011 (.054)	1366
Knowledge Variety	.005** (.002)	.002 (.003)	.001 (.003)	.081*** (.008)	-.009** (.004)	.025** (.011)	.035** (.016)	-.235*** (.031)	1366

Note: standard errors between parentheses. p<0.1; ** : p<0.05; *** : p<0.01.

Table 6 - Results of VAR estimation. One-year lag. Model including the interaction term (All variables are expressed in normalized growth rates according to Eq. 2).

	Sales Growth(t-1)	HGF*Growth	Knowledge Coherence (t-1)	Knowledge Capital (t-1)	Cognitive Distance (t-1)	Related Variety (t-1)	Unrelated Variety (t-1)	Knowledge Variety (t-1)	N. Obs.
Sales Growth	.005 (.009)	.329*** (.017)	.018 (.011)	-.132*** (.026)	.013 (.015)	.004 (.037)	.021 (.053)	.056 (.098)	1366
Knowledge Coherence	-.035*** (.006)	.026** (.013)	-.303*** (.009)	.037* (.021)	.019 (.012)	.071** (.029)	.107*** (.041)	-.276*** (.077)	1366
Knowledge Capital	.010*** (.004)	.006 (.008)	.004 (.005)	.698*** (.013)	-.001 (.007)	-.038** (.018)	-.049** (.025)	.120*** (.048)	1366
Cognitive Distance	-.006* (.003)	-.008 (.006)	.003 (.004)	.008 (.010)	-.014** (.006)	.001 (.014)	-.016 (.020)	.067* (.038)	1228
Related Variety	.012* (.006)	-.022* (.012)	.011 (.007)	.132*** (.018)	.012 (.011)	-.232*** (.026)	-.018 (.037)	.180*** (.069)	1366
Unrelated Variety	-.002 (.004)	.011 (.008)	.00002 (.005)	.007 (.013)	.004 (.007)	.040** (.018)	-.074*** (.026)	.021 (.049)	1366
Knowledge Variety	.009*** (.002)	-.009* (.005)	.001 (.003)	.080*** (.008)	-.013*** (.004)	.029*** (.011)	.040*** (.016)	-.249*** (.030)	1366

Note: standard errors between parentheses. p<0.1; ** : p<0.05; *** : p<0.01.

Table 7 - Results of VAR estimation. One-year lag. Model including both the HGFs dummy and the interaction term (All variables are expressed in normalized growth rates according to Eq. 2).

	Sales Growth(t-1)	HGF (dummy)	HGF*Growth	Knowledge Coherence (t-1)	Knowledge Capital (t-1)	Cognitive Distance (t-1)	Related Variety (t-1)	Unrelated Variety (t-1)	Knowledge Variety (t-1)	N. Obs.
Sales Growth	.075*** (.011)	.232*** (.010)	-.043** (.021)	.034*** (.013)	-.097*** (.030)	.001 (.017)	.056 (.042)	.058 (.060)	-.080 (.112)	1366
Knowledge Coherence	-.035*** (.007)	-.013* (.008)	.031** (.015)	-.307*** (.010)	.031 (.022)	.018 (.013)	.059* (.031)	.097** (.045)	-.250*** (.085)	1366
Knowledge Capital	.017*** (.004)	.009** (.004*)	-.005 (.009)	.007 (.005)	.698*** (.012)	-.003 (.007)	-.028* (.017)	-.031 (.025)	.094** (.046)	1366
Cognitive Distance	-.007** (.003)	-.004 (.003)	-.004 (.007)	.003 (.004)	.010 (.010)	-.012** (.006)	-.004 (.015)	-.027 (.021)	.089** (.040)	1228
Related Variety	.013** (.006)	.005 (.006)	-.027** (.013)	.011 (.007)	.135*** (.018)	.023** (.010)	-.231*** (.025)	-.018 (.036)	.162** (.068)	1366
Unrelated Variety	-.0009 (.005)	.004 (.005)	.004 (.010)	.0003 (.006)	.008 (.015)	.002 (.009)	.044** (.021)	-.063** (.030)	-.002 (.056)	1366
Knowledge Variety	.009*** (.003)	.002 (.003)	-.009* (.006)	.001 (.003)	.078*** (.008)	-.010** (.005)	.031*** (.012)	.043*** (.017)	-.244*** (.032)	1366

Note: standard errors between parentheses. $p < 0.1$; ** : $p < 0.05$; *** : $p < 0.01$.

Appendix B - Sectoral classification and concordance

Macro sectors	Sector	STAN (ISIC 3)	Datastream
High-technology manufactures HT	Pharmaceuticals	2423	4577
	Office, accounting and computing machinery	30	9572, 9574
	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
Medium-high technology manuf. MHT	Chemicals excluding pharmaceuticals	24ex2423	1353, 1357
	Machinery and equipment, n.e.c.	29	573, 583, 2757
	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
Medium-low technology manuf. MLT	Coke, refined petroleum products and nuclear fuel	23	533, 537, 577, 587
	Rubber, plastics products and other non-metallic mineral products	25-26	2353, 2723, 3357
	Basic metals and fabricated metal products	27-28	1753, 1755, 1757
Low technology manufactures LT	Food products and beverages	15	3533, 3535, 3537, 3577
	Tobacco products	16	3785
	Textiles, textile products, leather and footwear	17-19	3763, 3765
	Pulp, paper and paper products	21	1737
	Printing and publishing	22	5557
	Manufacturing nec and recycling	36-37	2727, 3724, 3726, 3767
Knowledge intensive sectors KIS	Post and telecommunications	64	5553, 6535, 6575
	Financial intermediation (excl insurance, pension)	65	8355, 8773, 8779
	Insurance and pension funding	66	8532, 8534, 8536, 8538, 8575
	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 LKIS	Wholesale, trade (excl. Motor vehicles)	51	2797, 5379
	Retail trade; repair of household goods	52	5333, 5337, 5371, 5373, 5375
	Hotels and restaurants	55	5753, 5757
Other services OS	Transport and storage	60-63	2771, 2773, 2775, 2777, 2779, 5751, 5759
	Community social and personal services	75-99	5377
Energy producing activities EP	Mining, quarrying of energy producing materials	10-12	1771
	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