Papers in Evolutionary Economic Geography

13.03

The properties of local knowledge bases and entrepreneurship: Evidence from Italian NUTS 3 regions

Alessandra Colombelli and Francesco Quatraro

Utrecht University
Urban & Regional research centre Utrecht

The properties of local knowledge bases and entrepreneurship: Evidence from Italian NUTS 3 regions[1].

Alessandra Colombelli[a,b,c] ad Francesco Quatraro[a,b*]

a) University of Nice Sophia Antipolis and CNRS-GREDEG
b) BRICK, Collegio Carlo Alberto
c) Dipartimento di Sistemi di Produzione ed Economia del'Azienda, Politecnico di Torino

ABSTRACT.

This paper investigates the relationship between the creation of new firms and the properties of the local knowledge bases, like coherence, cognitive distance and variety. By combining the literature on the knowledge spillovers of entrepreneurship and that on the recombinant knowledge approach, we posit that locally available knowledge matters to the entrepreneurial process, but the type of knowledge underlying theses dynamics deserve to be analyzed. The analysis is carried out on 104 Italian NUTS 3 regions observed over the time span 1995-2011. The results confirm that local knowledge is important, and suggest that the creation of new firms in Italy is associated to the exploitation of well established technological trajectories grounded on competences accumulated over time, rather than to the commercialization of brand new knowledge.

Keywords: Knowledge Coherence, Variety, Cognitive Distance, Italy, Knowledge-Spillovers Theory of Entrepreneurship, New Firms, Recombinant Knowledge.

JEL Classification Codes: L26, M13, R11, O33

*Corresponding Address: Francesco Quatraro, 250 rue Albert Einstein, 06560 Valbonne, France. Tel. +33495934373. Email: Francesco.quatraro@unice.fr.

[1] The authors acknowledge the funding of the European Union through the project WWWforEurope, grant agreement n. 290647, as well as the funding of the Collegio Carlo Alberto through the IPER project.

1 Introduction

A wide body of literature has emerged in the last decades investigating the issue of "entrepreneurship" from different perspectives. One of the reasons at the basis of such an interest lies in the belief that the creation of new firms is out of the main engines of economic growth (Vivarelli, 2013). Actually, according to the definition by Wennekers and Thurik (1999, p. 46–48) the entrepreneurial activity is "the manifest ability and willingness of individuals, on their own, in teams, within and outside existing organisations, to perceive and create new economic opportunities and to introduce their ideas in the market".

Thus, entrepreneurship has to do with novelty and change and involves a variety of entities both at micro and macro-level (Wennekers and Thurik 1999, Davidsson and Wiklund 2001). The relationship between entrepreneurial activities and economic performances however is not obvious, and is much related to the economic context in which the phenomenon takes place. Empirical analyses have addressed a wide range of dimensions related to the creation of new firms, so as to better appreciate both their influence on economic growth and the factors conducive to entrepreneurial activities. As is extensively discussed in Vivarelli (2013), microeconomic analyses have focused on the impact of firm size, credit rationing, education and learning dynamics, self-employment and innovation. On the other hand, the aggregate analyses of the topic has mostly focused on the shaping role of regional or national characteristics and the effects of the process of new firm formation on regional growth (Feldman, 2005; Acs et al., 2009; Delgado et al., 2010; Dejardin, 2011; Audretsch et al., 2012; Bishop, 2012; Qian et al., 2012).

This paper contributes the ongoing debate on the relationship between the features of local economic systems and new firm formation by investigating the specific influence of the characteristics of local technological knowledge. To this purpose, we will graft the knowledge spillovers theory of entrepreneurship (KSTE) onto the recombinant knowledge approach, and consider technological knowledge as the outcome of a combinatorial search activity carried out across a technological space in which combinable elements reside (Weitzman, 1998; Fleming, 2001; Fleming and Sorenson, 2001). In this direction we are able to specify a set of properties that can describe the internal structure of the local knowledge bases and that go beyond the traditional measure of knowledge capital stock. Indicators like knowledge coherence and knowledge variety can be calculated by exploiting the information contained in patent documents, and in particular by looking at the co-occurrence of technological classes which patents are assigned to (Saviotti, 2007; Quatraro, 2010).

Our analysis is focused on the patterns of new firm formation in Italian NUTS 3 regions (i.e. the “provincia” level) over the period 1995-2011. This appears an appropriate context for our analysis for different reasons. First, the close relationship between the entrepreneurial process and local economies calls for a focus on a sufficiently narrow definition of region. Second, the Italian economy appears to be stuck in mature industries and significantly late from a technological viewpoint, as compared to other most advanced countries (Quatraro, 2009a and b), so that our investigation will allow us to test the extent to which the relationship between the creation of new firms and technological knowledge is shaped by creative accumulation rather than creative destruction (Malerba and Orsenigo, 1997).

The results of the analysis confirms that local knowledge spillovers are important in shaping the entrepreneurial process. Moreover, when the characteristics of local knowledge bases are taken into account, the econometric analysis shows that knowledge coherence and variety exerts a positive influence on new firm formation, while cognitive distance negatively affects the rate of new firm creation. This suggests that in Italian regions entrepreneurship is mostly related to the exploitation of technological knowledge accumulated over time rather than to profiting from radical breakthroughs. The rest of the paper is organized as follows. Section 2 discusses the theoretical bases underpinning the relationship between entrepreneurship, local innovation and recombinant knowledge. In Section 3 we describe the data and the methodology, while in Section 4 we show the results of the econometric analysis. Finally, Section 5 provides the concluding remarks.

2 Entrepreneurship, local knowledge base and recombinant knowledge

New firms creation represents a crucial phenomenon in modern capitalist economies. Following Schumpeter (1911 and 1942), entrepreneurs are viewed as the main agents of innovation. Startup firms are all the more important in that they are likely to bring about innovations in the markets, above all when radical technologies are at stake, thus contributing economic growth (Aghion and Howitt, 1992; Wennekers and Thurik, 1999; Carree and Thurik, 2003; Audretsch et al., 2006; Friis et al., 2006).

Entrepreneurship is especially key to the process of economic development at the regional level. The emergence of entrepreneurial dynamics appears indeed to be geographically clustered, so that the local economy is likely to benefit from a self-enforcing process shaping regional comparative advantage (Feldman, 2001; Feldman et al. 2005). Despite this, empirical analyses of the link between entrepreneurship and regional dynamics have appeared only recently. On the one hand, a specific effort can be identified to assess the effects of entry dynamics on regional economic performances (see the special issue appeared in Small Business Economics in May 2011 'Entrepreneurial Dynamics and Regional Growth'). In this respect, new firm formation has been considered as a determinant of regional growth, cross-regional differences and regional employment dynamics.

On the other hand, both theoretical and empirical analyses have focused on the importance of the feature of local socio-economic systems to entrepreneurial dynamics. Feldman (2001) stresses the importance of the local availability of venture capital, supportive social capital, research universities and of support services to entrepreneurship. Audretsch et al (2012), following the Marshallian intuition, show that the local atmosphere shapes the process of entrepreneurship, above all in terms of regional regimes grounded on accumulated entrepreneurial culture. In the same direction, Qian et al. (2012) and Delgado et al. (2010) carry out empirical analyses of the impact of regional features in terms of knowledge and agglomeration on regional entrepreneurial dynamics. Stam (2007) argues that the interlink between regional contexts and the location choices of newborn firms evolves over firms' lifecycle, such that some local aspects, like the availability of an established network of relations, are more important in the early stages, while some others are important in later stages. All in all new firms appear to be strongly tied to local contexts and hardly decide to move abroad.

A more recent strand of literature has pointed to the importance of local knowledge bases to the entrepreneurial process. A key reference in this domain is the KSTE set forth by Acs et al. (2009). Such approach moves from a critique to endogenous growth theories, due to the fact that these latter, although in some cases are explicitly grounded on Schumpeter's legacy (Aghion and Howitt, 1992), fail to account for the essence of the Schumpeterian entrepreneur. In the KSTE entrepreneurs are the missing microeconomic link between the generation of new technological knowledge and economic growth (Audretsch, 1995). Entrepreneurs take advantage of the locally available knowledge to generate new economic opportunities. This implies a relationship between knowledge spillovers and entrepreneurial activity.

Empirical analyses have subsequent investigated and provided support to the impact of local knowledge spillovers on the entrepreneurial process, wherein the locally available stock of knowledge is the key variable and is usually proxied by R&D investments (Audretsch and

Keilbach, 2007) or by the research efforts carried out in the co-localized universities and research centres (Audretsch and Lehmann, 2005; Cassia, Colombelli, Paleari, 2009; Cassia and Colombelli, 2008).

More recently Bae and Koo (2008) and Bishop (2012) has noticed that not only the size of the knowledge stock, but also its nature is of some significance. Indeed the focus on knowledge stock implies an approach to technological knowledge as a homogenous good, neglecting the heterogeneity of competences behind its production and therefore its intrinsic heterogeneous nature. The analysis carried out by these authors focuses instead on the effects of knowledge diversity on new firm formation.

The issue of diversity has recently gained momentum in regional analyses as a consequence of the elaboration of an evolutionary approach to economic geography (Boschma and Frenken, 2007). In this framework, the accumulation of competences over time plays a key role in shaping the trajectories of regional development. The concept of regional branching identifies in this respect the emergence of new industrial activities out of the sectoral specialization emerged in the region in the course of time. Proximity matters not only from a geographical viewpoint, so that new variety in industrial activities is likely to be closely related to the activities already established in the region (Boschma, 2005; Boschma et al., 2013). In the same vein, the distinction between related and unrelated variety is also useful to clarify and disentangle the effects of classical Marshall-Arrow-Romer externalities from those of Jacobs' externalities (Frenken, Oort and Verburg, 2007).

The grafting of the KSTE onto the recombinant knowledge approach may be far reaching in enhancing the effects of the nature of local knowledge on new firm formation in an evolutionary perspective. The recombinant knowledge approach indeed a framework to represent the internal structure of regional knowledge bases as well as to enquire into the effects of its evolution. 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* and *proximity* of the technologies which knowledge bases are made of, as well as to the *variety* of the observed pairs of technologies. In view of this, the properties of knowledge structure may be made operative through the use of different methodologies, like social network analysis or the implementation of indicators based on co-occurrence matrixes in which rows and columns elements are bits of knowledge, while each cell reports the frequency with which each pair of technologies is observed.

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).

Actually, in the phase of emergence of new technologies one can more likely observe a decreasing average degree of technological complementarity in the region due to the introduction of technologies loosely related to the existing ones. In the same vein, the emergence phase is likely to be associated to increasing average technological distance and by the dominance of unrelated over related knowledge variety. The reverse applies instead in regional contexts characterized by the exploitation of established technologies. The link between technology lifecycle and entrepreneurship is however not obvious. On the one hand, Dejardin (2011) postulates that net entry rates may be linked to new products and emerging industries. On the other hand, Lumpkin and Dess (2001) stresses that the link between entrepreneurship and lifecycles are shaped by the intrinsic features of the entrepreneurs. Less proactive entrepreneurs are more likely to take advantage of established technological opportunities in mature industries, by taking market shares from an existing competitor, while more proactive entrepreneurs are more likely to benefit from emerging technologies in the earlier stages of the lifecycle.

In view of the arguments developed so far, we are now able to spell out the working hypotheses underlying the present analysis:

1. The entrepreneurial process is shaped by the local availability knowledge spillovers, in such a way that larger the amount of knowledge locally available, the higher the probability to observe new firms;
2. Not only the magnitude of local knowledge matters, but also its inherent heterogeneous nature. The structure of local knowledge bases may have differential effects on entreprenurship.
3. Entrepreneurship may be linked either to the early stages or the mature stages of the technology lifecycle, depending on the proactiveness degree of entrepreneurs. Thus the observed link between new firm formation and the properties of local knowledge bases in

terms of complementarity, proximity and variety in the aggregate may allow for making inferences on the nature of the local entrepreneurial culture.

3 Data, Variables and Methodology

3.1 The Data

In order to analyze the impact of the structure of local knowledge bases on the formation of new firms we matched the Patstat database updated to October 2011 with data provided by the Eurostat and NUTS3-level[2] data provided by the Italian institute of statistics (ISTAT), specifically the "Indicatori territoriali per le politiche di sviluppo" (local indicators for development policy) and the regional dataset on R&D expenditure. The Patstat database is a snapshot of the European Patent Office (EPO) master documentation database with worldwide coverage, containing tables including bibliographic data, citations and family links. These data combine both applications to the EPO and the application to the national patent offices, allowing to go back to 1920 for some patent authorities. This allows for overcoming the traditional limitation of EPO based longitudinal analysis due to its relatively young age.

Patent applications have been subsequently regionalized at the NUTS 3 level on the basis of inventors' addresses. Applications with more than one inventor residing in different regions have been assigned to each of the regions on the basis of the respective share. Our study is limited to the applications submitted in Italian regions, and uses International Patent Classification (IPC) maintained by the EPO to assign applications to technological classes.

3.2 The Variables

3.2.1 Dependent Variable

In order to implement our empirical analysis we took the (net) number of new businesses registering for value added tax (VAT). These data show some limitations insofar as only firms reaching a certain threshold level in terms of size are required to register for what. This is however

[2] The analysis covers the period 1995-2011. The Italian NUTS 3 classification changed in 2006 and 2009, when 4 and 3 new regions were added respectively. In order to ensure coherence in the dataset we used the before 2006 classification. This poses a problem only with respect to the Barletta-Andria-Trani region, which gathers together 7 municipalities that were previously part of the Bari province and 3 municipalities that were part of the Foggia province.

a problem common to all large datasets, which can be overcome only by implementing dedicated surveys, which however cannot have the same geographical coverage.

New firm formation at time t can be thought as a flow variable. In order obtain an index close to the (net) rate of new firm formation we divided it by the stock of firms observed in the area at the time t-1:

$$Entr_{i,t} = \left(\frac{NewFirm_{i,t}}{StockFirm_{i,t-1}}\right)$$

Where i is the NUTS3 region and t = [1995,2011] is the observed year. Figure 1 shows the distribution of firm's demography variables across Italian NUTS 3 regions.

>>> INSERT FIGURE 1 ABOUT HERE <<<

As is clear from all of the three diagrams (Top: ceased firms; Middle: new firms; Bottom: net entry), the regional distribution shows a rather low degree of spatial concentration. There is an evident area featured by high levels of new firm creation in between Lombardy and Veneto, while in the rest of Italy the evidence is somewhat scattered.

3.2.2 The Implementation of Knowledge Indicators

In Section 2 we have emphasized that a limited number of empirical analyses have focused on the impact of local conditions on entrepreneurial dynamics. The analysis conducted by Bishop (2012) is grounded on the measurement of regional knowledge diversity based on data on sectoral shares of employment to implement the informational entropy index. The idea is that each sector relies on specific competences, and thus sectoral data are indirect measures of the tacit knowledge observed in the region. Bae and Koo (2008) uses a more traditional approach to the measurement of knowledge, by looking at patent applications. They measure indeed diversity and relatedness relying respectively on an Herfindal index calculated on knowledge fields assigned by the USPTO and on patent citations.

In this paper we will follow an approach closer to this latter, in that we will use the information contained in patent documents[3] to calculate a number of variables that characterize the local knowledge base on the basis of the complementarity and similarity degree amongst its components.

[3]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

For what concerns the definition of the variables, let us start by the traditional concept of *knowledge stock* (KSTOCK). 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:

$$KSTOCK_{i,t} = \dot{h}_{i,t} + (1-\delta)KSTOCK_{i,t-1},$$

where $\dot{h}_{i,t}$ is the flow of patent applications and δ is the rate of obsolescence[4], where once again i is the region and t is the time period.

The implementation of knowledge characteristics proxying for variety, complementarity and similarity, rests on the recombinant knowledge approach. In order to provide an operational translation of such concepts 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. Each technological class j is linked to another class m when the same patent is assigned to both of them[5]. 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 document[6].

On this basis we calculated the following three key characteristics of firms' knowledge bases:

a) Knowledge variety (KV) measures the degree of technological diversification of the knowledge base. It is based on the informational entropy index.
b) Knowledge coherence (COH) measures the degree of complementarity among technologies.
c) Cognitive distance (CD) expresses the dissimilarities amongst different types of knowledge.

---

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).

[4] A similar approach is used by Soete et Patel (1985).

[5] In the calculations 4-digits technological classes have been used.

[6] It must be stressed that to compensate for intrinsic volatility of patenting behaviour, each patent application is made last five years in order to reduce the noise induced by changes in technological strategy.

3.2.2.1 Knowledge variety measured by the informational entropy index

Knowledge variety is measured using the information entropy index[7]. Entropy measures the degree of disorder or randomness of the system; systems characterized by high entropy are characterized by high degrees of uncertainty (Saviotti, 1988). The entropy index measures variety. Information entropy has some interesting properties (Frenken and Nuvolari, 2004) including multidimensionality.

Consider a pair of events (X_l, Y_j), and the probability of their co-occurrence p_{lj}. A two dimensional total variety (*TV*) measure can be expressed as follows:

$$KV \equiv H(X,Y) = \sum_l \sum_j p_{lj} \log_2 \left(\frac{1}{p_{lj}} \right) \tag{1}$$

Let the events X_l and Y_j be citation in a patent document of technological classes *l* and *j* respectively. Then p_{lj} is the probability that two technological classes *l* and *j* co-occur within the same patent. The measure of multidimensional entropy, therefore, focuses on the variety of co-occurrences or pairs of technological classes within patent applications.

The total index can be decomposed into 'within' and 'between' parts whenever the events being investigated can be aggregated into a smaller number of subsets. Within-entropy measures the average degree of disorder or variety within the subsets; 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 (Frenken and Nuvolari, 2004). Let the technologies *i* and *j* belong to the subsets *g* and *z* of the classification scheme respectively. If one allows $l \in S_g$ and $j \in S_z$ (g = 1,…,G; z = 1,…, Z), we can write:

$$P_{gz} = \sum_{l \in S_g} \sum_{j \in S_Z} p_{lj} \tag{1a}$$

Which is the probability to observe the couple lj in the subsets *g* and *z*, while the intra subsets variety can be measured as follows:

[7] For the sake of clarity the region and time indexes are omitted.

$$H_{gz} = \sum_{l \in S_g} \sum_{j \in S_z} \frac{p_{lj}}{P_{gz}} \log_2 \left(\frac{1}{p_{lj}/P_{gz}} \right) \tag{1b}$$

The (weighted) within-group entropy can be finally written as follows:

$$RKV \equiv \sum_{g=1}^{G} \sum_{z=1}^{Z} P_{gz} H_{gz} \tag{2}$$

Between group (or unrelated variety) can instead be calculated by using the following equation:

$$UKV \equiv H_Q = \sum_{g=1}^{G} \sum_{z=1}^{Z} P_{gz} \log_2 \frac{1}{P_{gz}} \tag{3}$$

According to the decomposition theorem, we can rewrite the total entropy *H(X,Y)* as follows:

$$KV = H_Q + \sum_{g=1}^{G} \sum_{z=1}^{Z} P_{gz} H_{gz} \tag{4}$$

When considering the International Patent Classification (IPC), the whole set of technological classes can be partitioned on the basis of macro technological fields. For example, two 4-digit technologies A61K and H04L belong respectively to the macro classes A and H. In our notation, H04L would be the technology *l* and H the macroset S_g. Similarly A61K would be the technology j and A the macroset S_Z.

Within-group entropy (or related variety) measures the degree of technological differentiation within the macro-field, while between-group variety (or unrelated variety) measures the degree of technological differentiation across macro-fields. The first term on the right-hand-side of equation (2) is the between-entropy, the second term is the (weighted) within-entropy.

We can label between- and within-entropy respectively as *unrelated technological variety* (*UTV*) and *related technological variety* (*RTV*), while total information entropy is referred to as *general technological variety* (Frenken et al., 2007; Boschma and Iammarino, 2009). This means that we consider variety as a global entity, but also as a new combination of existing bits of knowledge *versus* variety as a combination of new bits of knowledge. When variety is high (respectively low), this means that the search process has been extensive (respectively partial). When unrelated variety is high compared to related variety, the search process is based essentially on the combination of novel bits of knowledge rather than new combinations of existing bits of knowledge.[8]

3.2.2.2 The knowledge coherence index

Agents grounded in local contexts need to combine or integrate many different pieces of knowledge to produce a marketable output. Competitiveness requires new knowledge and knowledge about how to combine old and new pieces of knowledge. We calculate the coherence of NUTS3 regions' knowledge bases, defined as the average relatedness or complementarity of a technology chosen randomly within the firm's patent portfolio with respect to any other technology (Nesta and Saviotti, 2005, 2006; Nesta, 2008; Quatraro, 2010)[9].

Obtaining the knowledge coherence index requires a number of steps. First of all, we need to calculate the weighted average relatedness WAR_l of technology l with respect to all other technologies in the regional patent portfolio. This measure builds on the measure of *technological relatedness* τ_{lj} (Nesta and Saviotti, 2005, 2006). We start by calculating the relatedness matrix. The technological universe consists of k patent applications across all sampled firms. Let $P_{lk} = 1$ if the patent k is assigned the technology l [l= 1, …, n], and 0 otherwise. The total number of patents assigned to technology l is $O_l = \sum_k P_{lk}$. Similarly, the total number of patents assigned to technology j is $O_j = \sum_k P_{jk}$. Since two technologies can occur within the same patent, $O_l \cap O_j \neq$

[8] It must be noted that by measuring the degree of technological differentiation, the calculation of information entropy is affected by the number of technological classes observed, but not necessarily by the number of technological classes in the classification itself. Indeed, the introduction of new technological classes that are not observed does not affect the calculations in that they would be events with zero probability. Entropy rises or falls according to the number of technological classes that are actually observed in the patent sample. It reaches the maximum if all events are equiprobable, i.e. if all technological classes show the same relative frequency. If probabilities are unevenly distributed, one can have very low values of information entropy even if a very large number of technologies is observed.

[9] The function used to measure coherence is completely different from the one used to measure informational entropy. The fact that in both cases the co-occurrence of technological classes enters the calculations does not mean that both functions must lead to the same result. The informational entropy function measures the variety of the set, corresponding to the number of distinguishable entities it contains. The coherence function was introduced by Teece et al (1994) to measure the coherence of a firm based on its products. Nesta and Saviotti (2005, 2006) have subsequently adapted it to measure the coherence of the knowledge base of a firm. The coherence function measures the extent to which the distinguishable entities in the set (in our case the types of knowledge corresponding to different technological classes) are used together irrespective of the number of entities contained in the set. The two functions are in principle independent since they use the same type of data to calculate different properties of the same system. The mathematical independence of the two functions does not imply that the evolution of the corresponding properties is independent. Thus, if new technological classes are introduced into the knowledge base of a sector (an increase in the number of distinguishable entities of the set) there is no reason to expect the capacity of firms to combine the new types of knowledge to be created instantly. We expect that as new types of knowledge are introduced into the knowledge base of a sector, the firms will slowly learn to combine them thus leading to a temporary fall in coherence.

$\varnothing$, and thus the observed the number of observed co-occurrences of technologies l and j is $J_{lj} = \sum_k P_{lk} P_{jk}$. Applying this relationship to all possible pairs yields a square matrix Ω (n × n) in which the generic cell is the observed number of co-occurrences:

$$\Omega = \begin{bmatrix} J_{11} & & J_{l1} & & J_{n1} \\ \vdots & \ddots & & & \vdots \\ J_{1j} & & J_{lj} & & J_{nj} \\ \vdots & & & \ddots & \vdots \\ J_{1n} & \cdots & J_{\ln} & \cdots & J_{nn} \end{bmatrix} \tag{5}$$

We assume that the number x_{ij} of patents assigned to technologies i and j is a hypergeometric random variable of the mean and variance:

$$\mu_{lj} = E(X_{lj} = x) = \frac{O_i O_j}{K} \tag{6}$$

$$\sigma_{lj}^2 = \mu_{lj} \left(\frac{K - O_l}{K} \right) \left(\frac{K - O_j}{K - 1} \right) \tag{7}$$

If the observed number of co-occurrences J_{ij} is larger than the expected number of random co-occurrences μ_{ij}, then the two technologies are closely related: the fact that the two technologies occur together in the number of patents x_{ij} is not common or frequent. Hence, the measure of relatedness is given by the difference between the observed and the expected numbers of co-occurrences, weighted by their standard deviation:

$$\tau_{lj} = \frac{J_{lj} - \mu_{lj}}{\sigma_{lj}} \tag{8}$$

Note that this measure of relatedness has no lower or upper bounds: $\tau_{lj} \in \left] -\infty; +\infty \right[$. Moreover, the index shows a distribution similar to a t-test, so that if $\tau_{lj} \in \left] -1.96; +1.96 \right[$, we can safely assume the null hypothesis of non-relatedness of the two technologies i and j. The technological relatedness matrix Ω' can be considered a weighting scheme to evaluate the technological portfolio of regions.

Following Teece et al. (1994), WAR_l is defined as the degree to which technology l is related to all other technologies $j \in l$ in the region's patent portfolio, weighted by patent count P_{jt}:

$$\mathrm{WAR}_{\mathrm{lt}} = \frac{\sum_{\mathrm{j}\neq \mathrm{l}} \tau_{\mathrm{lj}} \mathrm{P}_{\mathrm{jt}}}{\sum_{\mathrm{j}\neq \mathrm{l}} \mathrm{P}_{\mathrm{jt}}} \tag{9}$$

Finally the coherence of the region's knowledge base at time *t* is defined as the weighted average of the WAR_{lt} measure:

$$COH_{\mathrm{t}} = \sum_{\mathrm{l}} \mathrm{WAR}_{\mathrm{lt}} \times \frac{\mathrm{P}_{\mathrm{lt}}}{\sum_{\mathrm{l}} \mathrm{P}_{\mathrm{lt}}} \tag{10}$$

Note that this index implemented by analysing the co-occurrence of technological classes within patent applications, measures the degree to which the services rendered by the co-occurring technologies are complementary, and is based on how frequently technological classes are combined in use. The relatedness measure τ_{lj} indicates that utilization of technology *l* implies use also of technology *j* 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 and marks a difference from entropy, which measures technological differentiation based on the probability distribution of pairs of technological classes across the patent sample[10].

If the coherence index is high, this means that the different pieces of knowledge have been well combined or integrated during the search process. Due to a learning dynamics, agents in the regions have increased capability to identify the bits of knowledge that are required jointly to obtain a given outcome. In a dynamic perspective, therefore, increasing values for knowledge coherence are likely to be associated with search behaviours mostly driven by organized search within well identified areas of the technological landscape. Conversely, decreasing values of knowledge coherence are likely to be related to search behaviours mostly driven by random screening across untried areas of the technological landscape in the quest for new and more profitable technological trajectories.

[10] To make it clear, informational entropy is a diversity measure which allows to accounting for variety, i.e. the number of categories into which system elements are apportioned, and balance, i.e. the distribution of system elements across categories. (Stirling, 2007). In this sense entropy does not say anything about the relationships between technological classes, but provides a measure of the diversity of technological co-occurrences, suggesting whether in a sector most of the observed co-occurrences focus on a specific couple or on the contrary whether the observed co-occurrences relate to a large number couples. In this framework, related and unrelated variety provide a measure of the extent to which observed variety applies to couples of technologies that belong to the same macro domain or to different macro-domains. One would expect established technologies to be characterized by relatively low variety of co-occurrences, insofar as the recombination focus on a relatively small numbers of technological classes that have proved to be particularly fertile. On a different ground, the coherence index is based on a normalized measure of how much each observed technology is complementary to all other technologies in the analyzed patents. In this sense it cannot be understood as a measure of diversity. The relatedness index indeed provides a measure of the degree to which two technologies are actually jointly used as compared to the expected joint utilization. The index allows to establishing a relationship of complementarity between the technologies in the analyzed patents. Based on the relatedness measure (tau), the coherence index provides an aggregate description of the degree to which the observed technologies in a given sector are complementary to one another.

3.2.2.3 *The cognitive distance index*

We need a measure of cognitive distance (Nooteboom, 2000) to describe the dissimilarities among different types of knowledge. A useful index of distance can be derived from *technological proximity* proposed by Jaffe (1986, 1989), who investigated the proximity of firms' technological portfolios. Breschi et al. (2003) adapted this index to measure the proximity between two technologies[11].

Let us recall that $P_{lk} = 1$ if the patent k is assigned the technology l [l= 1, …, n], and 0 otherwise. The total number of patents assigned to technology l is $O_l = \sum_k P_{lk}$. Similarly, the total number of patents assigned to technology j is $O_j = \sum_k P_{jk}$. We can, thus, indicate the number of patents that are classified in both technological fields l and j as: $V_{lj} = \sum_k P_{lk} P_{jk}$. By applying this count of joint occurrences to all possible pairs of classification codes, we obtain a square symmetrical matrix of co-occurrences whose generic cell V_{lj} reports the number of patent documents classified in both technological fields l and j.

Technologiocal proximity is proxied by the cosine index, which is calculated for a pair of technologies l and j as the angular separation or uncentred correlation of the vectors V_{lm} and V_{jm}. The similarity of technologies l and j can then be defined as follows:

$$S_{lj} = \frac{\sum_{m=1}^{n} V_{lm} V_{jm}}{\sqrt{\sum_{m=1}^{n} V_{lm}^2}\sqrt{\sum_{m=1}^{n} V_{jm}^2}} \tag{11}$$

The idea behind 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 m. Such measure is symmetric with respect to the direction linking technological classes, and it does not depend on the absolute size of technological

[11] Cognitive distance is the inverse of similarity or the equivalent of dissimilarity. The measure of similarity has been introduced by biologists and ecologists to measure the similarity of biological species and to understand to what extent they could contribute to biodiversity. The same measure has been applied by Jaffe (1986) to the similarity of technologies. It is not the only possible measure of similarity but it is the most frequently used one. The rational for its use is starts from the assumption that when two technologies, i and j, can be combined with a third technology k, they are similar. We call this measure cognitive distance both because the two terms are used as synonyms in the biological literature and, even more so, because cognitive distance is a concept used by Bart Nooteboom (2000) which has a number of very interesting implications for firm behavior and performance. In particular, the cognitive distance between different firms is expected to affect the probability that they form technological alliances. Intuitively, the need for a firm to learn a completely new technology (discontinuity) will lead to the incorporation into the firm's knowledge base of new patent classes, which would make the knowledge base recognizably different from what it was at previous times. The dissimilarity of the knowledge base can be expected to keep rising with respect to the pre-discontinuity knowledge base until the technology lifecycle has achieved maturity, at which stage the knowledge base of the firm will have stabilized, thus leading to a fall in cognitive distance.

field. The cosine index provides a measure of the similarity between two technological fields in terms of their mutual relationships with all the other fields. S_{lj} is the greater the more two technologies *l* and *j* co-occur with the same technologies. It is equal to one for pairs of technological fields with identical distribution of co-occurrences with all the other technological fields, while it goes to zero if vectors V_{lm} and V_{jm} are orthogonal (Breschi et al., 2003)[12]. Similarity between technological classes is thus calculated on the basis of their relative position in the technology space. The closer technologies are in the technology space, the higher is S_{lj} and the lower their cognitive distance (Engelsman and van Raan, 1991; Jaffe, 1986; Breschi et al., 2003).

The cognitive distance between j and l can be therefore measured as the complement of their index of technological proximity:

$$\mathrm{d}_{\mathrm{lj}} = 1 - \mathrm{S}_{\mathrm{lj}} \tag{12}$$

Having calculated the index for all possible pairs, it needs to be aggregated at the regional level to obtain a synthetic index of distance amongst the technologies in the firm's patent portfolio. This is done in two steps. First we compute the weighted average distance of technology *l*, i.e. the average distance of *l* from all other technologies.

$$\mathrm{WAD}_{\mathrm{lt}} = \frac{\sum_{\mathrm{j} \neq \mathrm{l}} \mathrm{d}_{\mathrm{lj}} \mathrm{P}_{\mathrm{jt}}}{\sum_{\mathrm{j} \neq \mathrm{l}} \mathrm{P}_{\mathrm{jt}}} \tag{13}$$

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

$$\mathrm{CD}_{\mathrm{t}} = \sum_{\mathrm{l}} \mathrm{WAD}_{\mathrm{lt}} \times \frac{\mathrm{P}_{\mathrm{lt}}}{\sum_{\mathrm{l}} \mathrm{P}_{\mathrm{lt}}} \tag{14}$$

The cognitive distance index measures the inverse of the similarity degree among technologies. When cognitive distance is high, this is an indication of the increased difficulty or cost the firm faces to learn the new type of knowledge which is located in a remote area of the technological space. Increased cognitive distance is related to the emergence of discontinuities associated with paradigmatic shifts in the sector knowledge base. It signals the combination of core technologies with unfamiliar technologies.

[12] For Engelsman and van Raan (1991), this approach produces meaningful results particularly at a 'macro' level, i.e. for mapping the entire domain of technology.

3.3 Methodology

The basic hypothesis spelt out in section 2 is that the properties of local knowledge bases exert an influence on the dynamics of new firm formation in view of the knowledge spillovers theory of entrepreneurship. In this direction the rate of creation of new firms is likely to be influenced by the variables described above, i.e. cognitive distance (CD), knowledge variety (KV, RKV, UKV) and knowledge coherence (COH). The test of such hypothesis needs for modelling the dependent variable $ENTR_{i,t}$ as a function of the characteristics of the knowledge base. The baseline specification would therefore be the following:

$$\ln\left(ENTR_{i,t}\right) = a + b_1 \ln KSTOCK_{i,t-k} + b_2 \ln CD_{i,t-k} + b_3 \ln COH_{i,t-k} + b_4 \ln KV_{i,t-k} + \rho_i + \sum \psi t + \varepsilon_{i,t} \qquad (15)$$

Where *KSTOCK* is the stock of patents observed in the region. The error term is decomposed in ρ_i and $\Sigma\psi t$, which are respectively region and time effects, and the error component ε_{it}. It must be noted that the variables proxying the characteristics of knowledge base are lagged five years in order to take into account the amount of time that is necessary for them to translate into an actual entrepreneurial process. Equation (15) can be estimated using traditional panel data techniques implementing the fixed effect estimator. It relates the rates of new firm creation to the characteristics of knowledge base. Covariates are lagged so as to minimize the risk of spurious relations. However, the features of local environments may take some time to exert an effect on entrepreneurial dynamics. For this reason we will allow for different lag specifications. Moreover, one needs also to control for the impact on the one hand of agglomeration economies, on the other hand of changing regional industrial specialization, so as to rule out the possibility that such effects are somehow captured by the knowledge-related variables. In view of this, we can write Equation (15) as follows:

$$\begin{aligned} \ln\left(ENTR_{i,t}\right) &= a + b_1 \ln KSTOCK_{i,t-k} + b_2 \ln CD_{i,t-k} + b_3 \ln COH_{i,t-k} + b_4 \ln KV_{i,t-k} + \\ &+ b_5 \ln R\&D_{t-k} + b_6 AGGL_{t-k} + b_7 LOQ_{t-k} + b_8 \ln UNEM_{i,t=1995} + b_9 AVBUSIZE_{t-k} + \rho_i + \psi t + \varepsilon_{i,t} \end{aligned} \qquad (16)$$

The rate of new firm formation depends now not only on local patent stock, variety, coherence and cognitive distance (respectively *KSTOCK*, *KV*, *COH* and *CD*). Following Acs et al. (2009), the effects of local knowledge spillovers are grasped by the intensity of R&D efforts. Moreover we also control for unemployment dynamics, which may affect the observed entrepreneurial behaviour. Following Crescenzi et al. (2007), the effects agglomeration economies are captured by the variable *AGGL*, which is calculated as the (log) ratio between regional population and size (square kilometres). The changing specialization is instead proxied by *LOQ*, i.e. the location quotient for manufacturing added value. Finally, as in Bishop (2012), we also control for the level of unemployment (*UNEM*) at the beginning of the period, and the average business size (*AVBUSIZE*) in the region. Table 1 provides a summary of variables definitions.

>>> INSERT TABLE 1 ABOUT HERE <<<

Table 2 reports instead the descriptive statistics concerning the variables used in the analysis after log transformation, while Table 3 shows the Spearman correlation coefficients amongst variable, so as to take into account for extreme values.

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

All in all the observed correlation coefficients, although almost always significant, do not raise particular concerns as the magnitude is not too high. The only exception is the stock of patents, which is highly correlated with the properties of local knowledge bases.

In addition to correlation, spatial dependence may also affect entrepreneurial dynamics. If spatial dependence is at stake, traditional econometric models may obtain biased results. In view of this, a new body of literature has recently developed, dealing with the identification of estimators able to account for both spatial dependence between the relationships between observations and spatial heterogeneity in the empirical model to be estimated. Former treatment of spatial econometric issues can be found in Anselin (1988), subsequently extended by Le Sage (1999).

The idea behind the concept of spatial dependence is straightforward. The properties of economic and social activities of an observed individual are likely to influence economic and social activities of neighbour individuals. Formally this relationship can be expressed as follows:

$$y_{i,t} = h(y_{j,t}),\ i = 1,\ldots,n,\ j \neq i \qquad \text{(xx)}$$

The dependence can therefore be among several observations. If this is the case, structural forms like equation (16) are likely to produce a bias in the estimation results. There are different ways to cope with this issue. In order to test whether spatial dependence affects our data, Lagrange-Multiplier tests are available which take into account the panel structure of the data (Elhorts, 2012). We implemented these tests to assess whether a spatial error model or a spatial autoregressive model are needed in this case.

>>> INSERT TABLE 4 ABOUT HERE <<<

Table 4 reports the results of such tests, conducted by using three different specifications of the spatial weighting matrix, i.e. contiguity matrix, 4 nearest neighbour and 8 nearest neighbour. As suggested by the diagrams showed in Figure 1, the entrepreneurial dynamics are featured by a low degree of spatial concentration. Indeed in all of the tests for spatial dependence we cannot reject the null hypothesis of non-spatial dependence at 5%.

4 Econometric results

The results of the econometric estimations of equation (16) are reported in Table 5. As specified in the previous section, we run different estimations with different lag specifications[13]. We show the results obtained by including the two-years lags of the covariates, as these are featured by the lowest Akaike index for each of the models. The first column report the results of the fixed-effect estimation including total knowledge variety. Consistently with the KSTE, the coefficient o regional R&D expenditure is positive and siginificant (1%). This supports therefore the idea that entrepreneurs create new firms by taking advantage of the locally available unexploited knowledge. We can interpret in the same direction the positive and significant coefficient on local knowledge stock. One can also interpret the coefficient on . For what concerns the properties of local knowledge bases, one can observe that the coefficient on knowledge coherence (COH) is positive and significant at 1%, the same way as the coefficient on cognitive distance (CD. The coefficient on variety is positive and significant at 10%.

>>> INSERT TABLE 5 ABOUT HERE <<<

These results taken together suggest that, while the KSTE holds, the entrepreneurial dynamics in Italian NUTS 3 regions are linked to mixed dynamics of local knowledge bases characterized by high degree of coherence and high degree of cognitive distance. The former suggests that new firms are likely to emerge out of established local technological trajectories grounded on the exploitation of technological competences accumulated over time However, the positive sign of cognitive distance suggests that a key condition to the creation of new firms is the local availability of complementary technological competences which span over a wide area of the technology landscape. In other words, new firms take advantage of knowledge spillovers within local contexts wherein the available knowledge base is characterized by high levels of integration as well as by high levels of dissimilarity. A narrow focus for search activities in the technological landscape can be detrimental to the creation of new firms. This interpretation is further supported by the positive sign on knowledge variety.

[13] We stopped at the third lag, due to data constraints. The results obtained by including the first or the third lag of the covariates do not yield significant changes.

The second column of table 5 reports instead of the estimation including related knowledge variety instead of total knowledge variety. The results are fairly similar to the previous ones. The coefficient on R&D and knowledge stock are still significant at 1% and. The coefficients of coherence (COH) and cognitive distance (CD) are still positive and significant at 1%. Differently from the previous estimation, related knowledge variety is not significant (although the coefficient is still positive). Once again, this evidence suggests that new firm formation in Italian provinces is associated to the exploitation of local knowledge bases which take advantage of learning and accumulated knowledge, to provide a guidance to search activities conducted across a wide and possibly distant area of the technology landscape. Regional innovating agents fishing in complementary but dissimilar (with respect to the accumulated competences) technology domains to generate new knowledge, are likely to create the conditions to foster the creation of new firms.

Column (3) reports the results of the estimations including unrelated knowledge variety. The signs of the coefficients are the same as the previous estimations, and (unrelated) variety is still not significant in this case. In column (4) we report instead the results of the estimations including both related and unrelated variety. Although these two latter may be characterized by a high degree of (negative) correlation, we nonetheless decided to run a regression which takes them into account jointly. The results are well in line with the previous evidence, indeed COH and CD are positive and significant at 1%, RKV and UKV are not significant. The results appear to be therefore robust to different specifications and suggest that the Italian context is characterized by a pattern of entrepreneurship, grounded on the exploitation of local knowledge opportuinities which are generated out of search activities counducted across complementary, although far away technology competences.

Finally column (6) provide estimations including the unemployment rate at the initial period. Since this latter is time-invariant, we implemented an OLS estimation, but including regional dummies. The sign and significance of the properties of the knowledge base are still consistent with the previous evidence, supporting their robustness. The coefficient on the unemployment rate is positive and significant, which is quite in line with the self-employment literature, according to which the creation of new firms can be an outcome of the economic agents' response to unemployment..

5 Conclusions

The issue of entrepreneurship has received increasing attention in the last decades, following the Schumpeterian view of the entrepreneur as an agent of change and an engine of economic growth. The literature on entrepreneurship is fairly large, ranging from micro-level analyses focusing on the idiosyncratic features of entrepreneurs to macro-level analyses focused on the relationship between the features of the local economy and the dynamics of new firm formation.

This paper aims to contributing this latter strand of analysis by investigating the effects of the characteristics of local knowledge bases on the rate of new firm creation. To this purpose we grafted the KSTE onto the recombinant knowledge approach and maintain that knowledge spillovers are important not only from a quantitative viewpoint, but also the nature of knowledge matters. We therefore derived a number of indexes proxying for the average degree of complementarity, similarity and variety of the technological competences residing in the region which are based on the information contained in patent applications.

The results of the empirical analysis are in line with previous literature on KSTE. Moreover, the effects of the properties of the local knowledge bases are pretty robust across different specifications, and allows for qualifying the argument put forth by the KSTE literature. Indeed, the evidence concerning entrepreneurial dynamics in Italian provinces suggests that the availability of local knowledge spillovers is not sufficient per se to lead the creation of new firms. If one looks at the properties of local knowledge bases, the rate of new firm formation appears to be fostered in contexts featured by knowledge stemming from search activities shaped by the accumulated competences and dispersed across a wide are of the technology landscape. New firms seem to emerge out of technological opportunities which are left unexploited by incumbents due to their relative distance from their core competencies.

Our results can bear some implications for regional technology policies. Indeed these latter usually aims at promoting local competitiveness through the support to local technology activities (Borras and Edquist, 2013). The choice of the correct policy mix should therefore take into account the differential effects that technological strategies may bear on incumbent firms with respect to prospective new firms. Both incumbents and prospective entrants may indeed play a key role for local competitiveness, such that policy measures should be grounded on the careful screening of local competitive advantages and devise a balanced mix of measures aimed at creating on the one hand the conditions to the creation of new firms and on the other hand at providing incumbent firms with exploitable knowledge.

6 References

Acs, Z.J., Anselin, L. and Varga, A., 2002, "Patents and innovation counts as measures of regional production of new knowledge", *Research Policy*, 31, 1069-1085.

Acs, Z.J., P. Braunerhjelm, D.B. Audretsch, and B. Carlsson (2009). The knowledge spillover theory of entrepreneurship. *Small Business Economics* 32, 15–30.

Aghion, P. and Howitt, P. (1992), A Model of Growth Through Creative Destruction, *Econometrica*, 60, 323-51.

Anselin, L., 1988. *Spatial Econometrics: Methods and models*. Kluwer, Dordrecht.

Audretsch, D.B. (1995), *Innovation and Industry Evolution*, Cambridge (Mass), MIT Press.

Audretsch, D.B., Falck, O., Feldman, M.P. and Heblich, S. (2012), Local Entrepreneurship in context, *Regional Studies*, 46, 379-389.

Audretsch, D.B. and Keilbach, M.C., (2007). The localisation of entrepreneurship capital: Evidence from Germany, *Papers in Regional Science*, 86, 351-365.

Audretsch, D.B., Keilbach, M.C. and Lehmann, E.E. (2006), *Entrepreneurship and Economic Growth*, Oxford, Oxford University Press.

Audretsch, D.B. and Lehmann, E.E. (2005) Does the knowledge spillover theory of entrepreneurship hold for regions?, *Research Policy*, 34, pp. 1191–1202

Bae, J., and J. Koo. 2008. The nature of local knowledge and firm formation. *Industrial and Corporate Change* 18, 1–24.

Bishop, P. (2012), Knowledge, diversity and entrepreneurship: a spatial analysis of new firm formation in Great Britain, *Entrepreneurship and Regional Development*, 24, 641-660.

Borras, S. and Edquist, C., (2013). The choice of innovation policy instruments. *Technology Forecasting and Social Change*, forthcoming.

Boschma, R.A., (2005). Proximity and innovation: A critical assessment. *Regional Studies* 39, 61-74.

Boschma, R.A., Frenken, K. (2007). A theoretical framework for economic geography: industrial dynamics and urban growth as a branching process. *Journal of Economic Geography* 7, 635-649.

Boschma, R.A. and Iammarino, S. (2009). Related Variety, Trade Linkages, and Regional Growth in Italy. *Economic Geography* 85, 289-311.

Boschma, R.A., Minondo, A., Navarro, M., (2013). The emergence of new industries at the regional level in Spain. A proximity approach based on product-relatedness. *Economic Geography*, 89, 29-51.

Breschi, S., Lissoni, F., and Malerba, F., 2003, “Knowledge relatedness in firm technological diversification”. *Research Policy*, 32, 69-97.

Carree, M. and Thurik, A.R., (2006), Understanding the Role of Entrepreneurship for Economic Growth, in Carree, M. and Thurik, A.R. (Eds.), *The Handbook of Entrepreneurship and Economic Growth*, Cheltenham, Elgar, ix-xix

Cassia, L., and A. Colombelli (2008). Do universities knowledge spillovers impact on new firm’s growth? Empirical evidence from UK, *The International Entrepreneurship and Management Journal* 4, 453–65.

Cassia, L., Colombelli, A and Paleari, S. (2008), Firms’ growth: Does the innovation system matter?, *Structural Change and Economic Dynamics*, 20, 211-220.

Davidsson P. and Wiklund J. (2001), Levels of Analysis in entrepreneurship research: Current research practice and suggestions for the future, *Entrepreneurship Theory and Practice*, 25(4), 81-99.

Delgado, M., Porter, M.E. and Stern, S. (2010), Clusters and Entrepreneurship, *Journal of Economic Geography*, 10, 495-518.

Dejardin, M. (2011). Linking net entry to regional economic growth, *Small Business Economics*, 36, 443-460.

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

Elhorst, J.P. (2012), Matlab software for spatial panels, *International Regional Science Review*, DOI: 10.1177/0160017612452429.

Engelsman, E.C. and van Raan, A.F.J, 1994, A patent-based cartography of technology, *Research Policy*, 23, 1-26.

Feldman, M. (2005), Creating a Cluster While Building a Firm: Entrepreneurs and the Formation of Industrial Clusters. *Regional Studies*, 39, 129-141.

Feldman, M. (2001), The entrepreneurial event revisited: Firm formation in regional context, *Industrial and Corporate Change*, 10, 861-891.

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

Fleming, L., Sorenson, O. (2001), Technology as a complex adaptive system: Evidence from patent data, *Research Policy* 30, 1019-1039.

Frenken, K., and Nuvolari, A., (2004). The early development of the steam engine: an evolutionary interpretation using complexity theory, *Industrial and Corporate Change*, 13, 419–450.

Frenken, K., van Oort F.G., Verburg, T., (2007). Related variety, unrelated variety and regional economic growth. *Regional Studies* 41, 685-697.

Jaffe, A.B, 1989, “Characterising the technological position of firms, with application to quantifying technological opportunity and research spillovers”, *Research Policy*, 18, 87–97.

Jaffe, A.B., 1986, “Technological opportunity and spillovers of R&D: evidence from firms’ patents, profits, and market value”, *American Economic Review*, 76, 984–1001.

Krafft, J., Quatraro, F., Saviotti, P.P., 2009. Evolution of the knowledge base in knowledge intensive industries, LEI and BRICK Working Paper no. 06/2009.

Le Sage, J.P., 1999. The theory and practice of spatial econometrics, Department of Economics, University of Toledo, available at www.spatial-econometrics.org.

Lumpkin, G.T. and Dess, G.G. (2001) Linking two dimensions of entrepreneurial orientation to firm performance: The moderating role of environment and industry life cycle. *Journal of Business Venturing*, 16, 429-451.

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

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

Nesta, L., Saviotti, P.P., 2006, “Firm knowledge and market value in biotechnology”, *Industrial and Corporate Change*, 15, 625–652.

Pavitt, K., 1985, “Patent statistics as indicators of innovative activities: possibilities and problems”, *Scientometrics*, 7, 77–99.

Quatraro, F., (2009a), The diffusion of regional innovation capabilities: Evidence from Italian patent data. *Regional Studies* 43, 1333-1348.

Quatraro, F., (2009b) Innovation, structural change and productivity growth. Evidence from Italian regions, 1980-2003. *Cambridge Journal of Economics* 33, 1001-1022.

Quatraro, F., (2010), Knowledge Coherence, Variety and Productivity Growth: Manufacturing Evidence from Italian Regions. *Research Policy* 39, 1289-1302.

Qian, H., Acs, Z.J. and Stough, R.R. (2012), Regional systems of entrepreneurship: The nexus of human capital, knowledge and new firm formation, *Journal of Economic Geography*, forthcoming, doi: 10.1093/jeg/lbs009.

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.

Soete, L. and Patel, P., 1985, “Recherche-developpement, importations de technologie et croissance économique: une tentative de comparaison internationale”, *Revue Economique*, 36, 975-1000.

Stam E. (2007), Why Butterflies Don't Leave. Locational Behavior of Entrepreneurial Firms, *Economic Geography*, 83, 27–50.

Teece, D., Rumelt, R., Dosi, G. And Winter, S., 1994, "Understanding Corporate Coherence: Theory and Evidence", *Journal of Economic Behavior and Organization*, 23, 1–30.

Vivarelli, M. (2013), Is entrepreneurship necessarily good? Microeconomic evidence from developed and developing countries, *Industrial and Corporate Change,* forthcoming.

Weitzman, M. L., 1998, "Recombinant growth", *Quarterly Journal of Economics*, 113, 331-360.

Wennekers, A.R.M., Thurik, A.R. (1999), Linking entrepreneurship and economic growth, *Small Business Economics*, 13, 27-55.

Table 1 - Description of the variables used in the analysis

Variable	Description
ENTR	logarithm of the ratio between new registered firms at time t and the stock of firms at time t-1 in region i
AGGL	logarithm of the ratio between population and the area of region i at time t
KSTOCK	logarithm of regional knowledge stock of region i at time t-5 (see details in the text)
COH	logarithm of knowledge coherence of region i at time t-5
KV	logarithm of knowledge variety of region i at time t-5
RKV	logarithm of related knowledge variety of region i at time t-5
UKV	logarithm of unrelated knowledge variety of region i at time t-5
CD	logarithm of cognitive distance of region i at time t-5
R&D	logarithm of the share of business expenditure in R&D of NUTS 2 region i at time t-1
UNEM	logarithm of unemployment rate at 1995
BUSIZE	Logarithm of the ratio between the regional number of employees and the stock of firms at the NUTS 3 level
Q	Logarithm of the location quotient of manufacturing employment at the NUTS 3 level

Table 2- Descriptive Statistics

variable	N	mean	min	max	sd	skewness	kurtosis
ENTR	1654	-2.186	-3.536	-1.632	0.130	-2.171	17.886
AGGL	2191	5.131	3.444	7.886	0.777	0.599	4.210
KSTOCK	2821	3.508	-1.300	9.100	1.887	0.037	2.645
COH	2494	2.836	-0.194	4.612	0.241	1.849	25.665
KV	2136	1.294	0.523	-0.209	2.224	-0.832	3.201
RKV	2014	0.891	-1.253	1.915	0.556	-0.799	3.560
UKV	1982	0.403	-1.387	1.064	0.371	-0.897	3.611
CD	2482	-0.303	-1.040	-0.089	0.078	-2.373	13.125
R&D	1743	-1.308	-3.457	0.000	0.741	-0.606	2.698
AVBUSIZE	1245	-5.317	-5.946	-4.554	0.159	-0.534	4.543
Q	1366	-0.140	-1.383	0.710	0.520	-0.469	2.227
UNEMP	1683	7.921	1.6	27.2	5.498	1.419	4.594

Table 3 - Spearman Correlation Coefficient

	ENTR	AGGL	KSTOCK	COH	KV	RKV	UKV	CD	R&D	UNEM	Q	AVBUSIZE
ENTR	1											
AGGL	0.022	1										
KSTOCK	-0.014	0.312*	1									
COH	0.069*	-0.102*	-0.044	1								
KV	-0.022	0.329*	0.908*	-0.096*	1							
RKV	0.001	0.313*	0.854*	-0.028	0.950*	1						
UKV	-0.072*	0.256*	0.578*	-0.246*	0.600*	0.346*	1					
CD	0.270*	0.014	-0.101*	0.007	-0.105*	-0.081*	-0.116*	1				
R&D	-0.075*	0.249*	0.346*	-0.155*	0.343*	0.351*	0.174*	-0.005	1			
UNEM	0.228*	0.155*	-0.432*	0.094*	-0.386*	-0.348*	-0.285*	0.252*	-0.401*	1		
Q	-0.134*	0.007	0.359*	-0.159*	0.371*	0.324*	0.282*	-0.197*	0.407*	-0.661*	1	
AVBUSIZE	-0.177*	-0.113*	0.366*	-0.144*	0.332*	0.269*	0.291*	-0.148*	0.078*	-0.464*	0.370*	1

Note : * indicates significance at 5% confidence level.

Table 4 – LM test of spatial dependence for panel data (Elhorst, 2012)

	Contiguity	4 nearest neighbour	8 nearest neighbour
LM test spatial lag (robust)	0.8384 (0.360)	3.8055 (0.051)	2.8594 (0.091)
LM test spatial error (robust)	0.2225 (0.637)	1.6494 (0.199)	0.6041 (0.437)

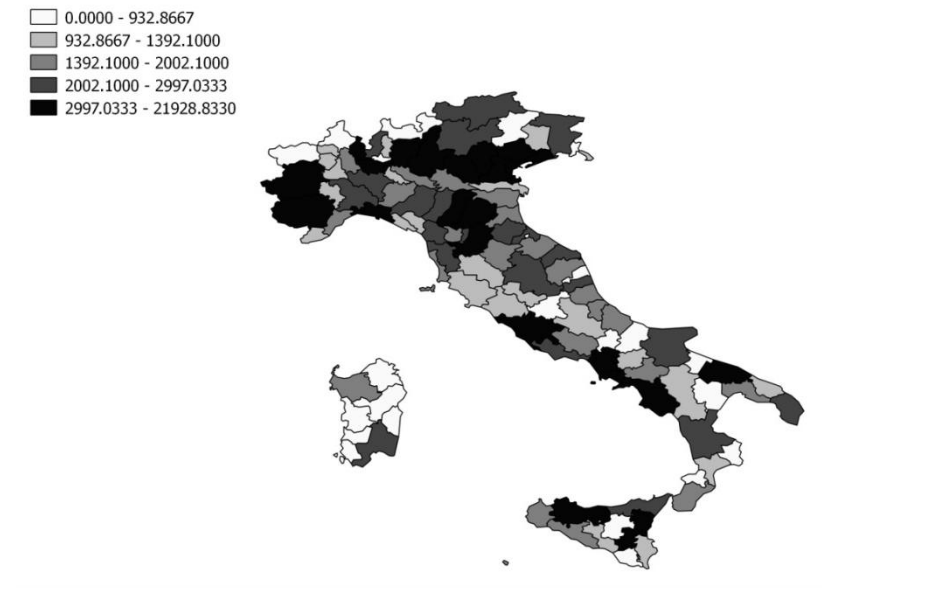
Note : H0: nonspatial model. The specification includes region and time fixed effects.

Table 5 - Econometric results, fixed effects estimations

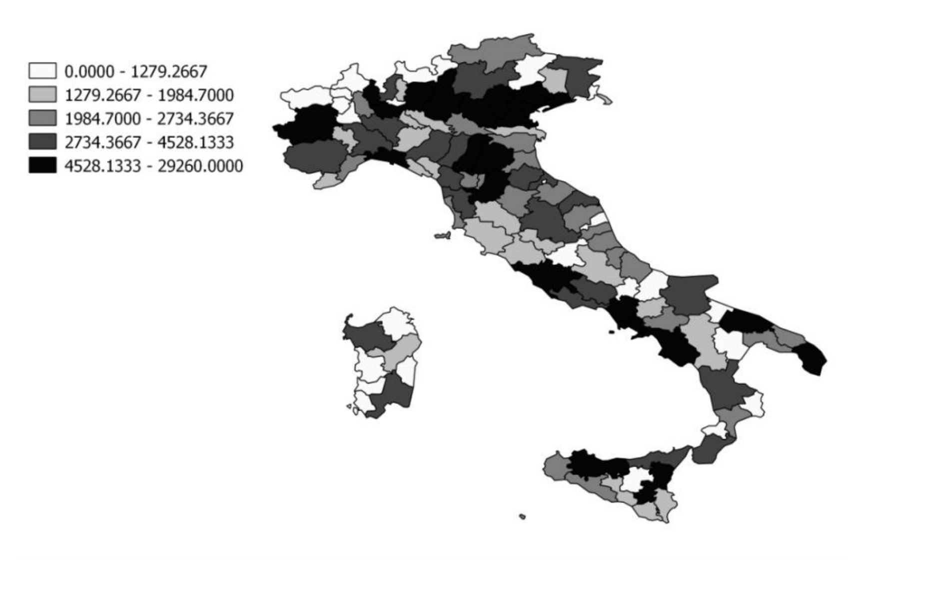
VARIABLES	(1) FE	(2) FE	(3) FE	(4) FE	(5) LSDV
L2.AGGL	-0.508** (0.200)	-0.622*** (0.210)	-0.532** (0.207)	-0.647*** (0.216)	-0.520*** (0.200)
L2.KSTOCK	0.0501*** (0.0161)	0.0567*** (0.0162)	0.0590*** (0.0153)	0.0539*** (0.0184)	0.0497*** (0.0162)
L2.COH	0.0871*** (0.0275)	0.107*** (0.0284)	0.101*** (0.0287)	0.126*** (0.0301)	0.0752*** (0.0283)
L2.KV	0.0255* (0.0144)				0.0300** (0.0147)
L2.RKV		0.0108 (0.0108)		0.0111 (0.0123)	
L2.UKV			0.0220 (0.0140)	0.0192 (0.0153)	
L2.CD	1.263*** (0.204)	1.200*** (0.214)	1.290*** (0.216)	1.265*** (0.226)	1.226*** (0.206)
L2.lnQ	0.0475 (0.0403)	0.0605 (0.0413)	0.0670 (0.0421)	0.0768* (0.0429)	0.0573 (0.0411)
L2.RD	0.0451*** (0.0131)	0.0455*** (0.0147)	0.0455*** (0.0139)	0.0447*** (0.0157)	0.0456*** (0.0139)
L2.AVBUSIZE	0.223*** (0.0717)	0.232*** (0.0726)	0.191*** (0.0736)	0.200*** (0.0745)	0.212*** (0.0720)
UNEM					0.0347*** (0.00974)
TREND	-0.0159*** (0.00182)	-0.0153*** (0.00195)	-0.0162*** (0.00183)	-0.0152*** (0.00203)	-0.0159*** (0.00184)
Constant	1.052 (1.164)	1.618 (1.220)	0.961 (1.182)	1.560 (1.232)	0.241 (0.828)
Observations	1,016	970	964	918	997
Log-likelihood	1022.714	982.986	982.099	946.406	1003.772
AIC	-2025.428	-1945.973	-1944.199	-1870.812	-1801.544
R-squared	0.200	0.200	0.199	0.206	0.401
Number of id	96	95	94	92	

Dependent variable: logarithm of the rate of new firm creation.
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

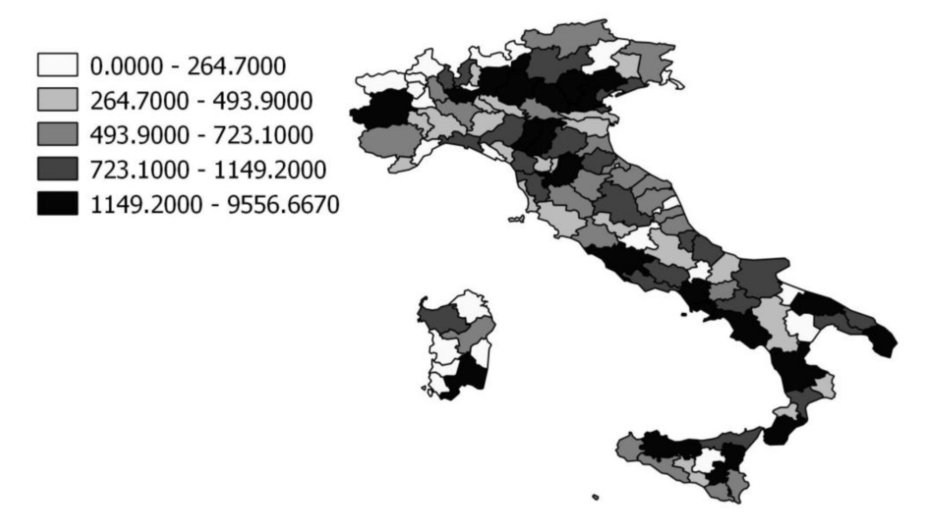
Figure 1 - Firms' Demography, Regional Breakdown (average values 2000-2005)



a) Ceased firms



b) New firms



c) Net entry