3rd edition of the International Conference
THE GOVERNANCE OF A COMPLEX WORLD

2014 Conference theme:
Smart, inclusive and sustainable

18-20 June, 2014
Torino, Italy

Innovative startups and local knowledge base: Evidence from Italian NUTS 3 regions

Alessandra Colombelli (Corresponding author)
DIGEP, Politecnico di Torino
Corso Duca degli Abruzzi, 24
10129 Torino, Italy
Phone +39 011 0907275
Fax +39 011 0907299
e-mail alessandra.colombelli@polito.it

Preliminary draft

1. Introduction

The need to target policy intervention towards innovative start-ups is widely acknowledged by both policy makers and scholars as a key measure of sustainable economic growth in the EU as well as overseas. Entrepreneurship, i.e. the process by which new enterprises are founded and become viable, is indeed essential to employment growth and job creation, wage growth, and wealth creation and is recognized as the key engines of economic activities, both for stagnant economies to recover and for emerging ones to sustain growth. In this context, at the end of 2012 a new regulation, providing for specific measures to foster the creation and development of innovative start-ups, was approved in Italy. Since then, at the end of 2013 more than 1500 innovative start-ups have registered at the Italian Chambers of Commerce.

This paper aims at investigating the relationship between the features of local economic systems, more precisely the specific influence of the characteristics of local knowledge bases, and the creation of innovative startups. To this purpose, we will graft the knowledge spillovers theory of entrepreneurship (KSTE) (Acs et al. 2009, 2013) 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, cognitive distance 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) by using the data on the creation of innovative startups within the framework of the new regulation. 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, so that our investigation will allow us to test the extent to which the relationship between the creation of innovative startups and technological knowledge is shaped by the phase of the regional technology lifecycle.

The results of the analysis shed light on the role of local knowledge spillovers in shaping the creation of innovative startups. Moreover, when the characteristics of local knowledge bases are taken into account, the econometric analysis can contribute to understand whether in Italian regions the creation of innovative startups is mostly related to the exploitation of technological knowledge accumulated over time or to the exploration of new research avenues.

2. Literature review

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).

The creation of new businesses 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 (Fritsch and Schindele, 2011).

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. Lee et al. (2004), drawing upon the notion of Jacobs' externalities, investigate the importance of social diversity and creativity to the formation of new firms. 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.

As far as the determinants of new firm creation are concerned , In in the textbook view originally put forward by Mansfield (1962), a queue of well-informed potential entrepreneurs is supposed to be waiting outside the market, and the expected level of profit is considered the trigger factor determining entry (see also Orr, 1974; Khemani and Shapiro, 1986). Moreover, new firm formation may be triggered also by other pull factors such as economic growth and high innovative potential (see see Acs and Audretsch, 1989a and 1989b; Geroski, 1995). In addition, Moreover, following authors such as Knight (1921), Schumpeter (1934 and 1939) and Oxenfeldt (1943), we are aware that important individual determinants may act as push factors and be related both to environmental circumstances and to the potential founder's personal characteristics. Within this framework, new firm formation can be modeledmodelled as an income choice based on a comparison between the wage earned in the previous job and the expected profit as an entrepreneur starting a new business in the same sector and in the same geographical area (see Creedy and Johnson, 1983; Vivarelli, 1991; Foti and Vivarelli, 1994; Audretsch, 1995; Geroski, 1995; Reynolds, 1997; Vivarelli, 2004). Pushing this argument further, founding a new firm may be an alternative to uncertain future career prospects, or even represents an 'escape from unemployment' (see Oxenfeldt, 1943; Evans and Leighton, 1990; Storey, 1991 and 1994). The empirical evidence suggesting the important role of job losses in fostering entry is indeed quite robust (see Storey and Jones, 1987; Santarelli, Carree and Verheul, 2009; Audretsch and Vivarelli, 1995 and 1996). Finally, the features of the industrial structure may also shape the dynamics of firm formation. For example, the industry minimum efficiency scale (MES) can represent an obstacle for new entrepreneurs (Acs and Audrestch, 1989b; Audretsch and Mahmood, 1995; Mata et al., 1995; Audretsch et al., 1999). Besides this, the sectoral composition of local economies is also a crucial factor (Quatraro and Vivarelli, 2013).

A more recent strand of literature has pointed to the importance of local knowledge spillovers 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, which 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 subsequently 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) have 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 variety 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.

In this direction, the grafting of the KSTE onto the recombinant knowledge approach may be far reaching in shedding further light on the effects of the nature of local knowledge on new firm formation in an evolutionary perspective (Weitzman, 1998; Fleming and Sorenson, 2001). According to this stream of research, knowledge stems from the combination of different technologies, so that the knowledge base of a firm (or of a sector or a region) can be represented as a web of connected elements (Krafft, Quatraro, Saviotti, 2014; Colombelli, Krafft Quatraro, 2014; Quatraro, 2010). 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. The recombinant approach allows therefore for qualifying the arguments put forth by the KSTE, by explicitly taking into account the relatedness, similarity and variety degree of the technological domains featuring the local knowledge base.

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. New firm creation is expected to be shaped by the relatedness, similarity and variety degree amongst the technological domains featuring the local knowledge base.

3. The Italian framework

At the end of 2012 the Italian Ministry of Economic Development approved a Law Decree on "Further urgent measures for Italy's economic growth", providing for specific measures which are aimed at promoting the creation and development of start-ups. This was the first time the Italian legislation took this kind of companies into consideration. The law recognizes that start-ups are important for the promotion of sustainable growth, technological development and employment, in particular youth employment, and aims at developing a new business culture, creating an environment which is more favorable to innovation, increasing social mobility, as well as attracting to Italy investments and talented people from abroad. Under this law, at the end of 2013 more than 1500 innovative start-ups registered at the Chambers of Commerce in Italy.

According to the Law Decree, in order to be included in the register of "innovative start-ups" and to benefit from governmental incentives, the new company needs to fulfill some requirements:

a) it must reside or be subject to taxation in Italy
b) it must have been established for no longer than 48 months
c) it has no turnover or has a turnover that does not exceed 5 million euros
d) is owned directly and for at least a 51% share by individuals, also in terms of voting rights
e) it does not distribute profits
f) it has to be clearly linked with innovation and technology.

According to the Law, a start-up is innovative if it satisfies at least one of the following criteria:

a) either 20% of its costs are related to R&D
b) at least one third of the team is made up of high qualified members. For high qualified members are intended all the people that either hold a PhD or are PhD candidates at an Italian or foreign university or have conducted research for at least three years or it is the owner or the licensee of a patent.

The incentives and benefits from support measures provided by the Law Decree include:

a) Start-ups can use specific flexible employment contracts.
b) Start-ups can remunerate their team members and the providers of external services, including lawyers and accountants, with stock options and work for equity respectively.
c) Priority credit access is granted to facilitations for the employment of highly qualified personnel in innovative start-ups.
d) Simplified and free-of-charge access for start-ups to the Fondo Centrale di Garanzia.
e) More help and support in the process of internationalization of startups, promoted by the Trade Promotion Agency ("ICE") .
f) Introducing a "fail fast" procedure: the aim of this measure is to avoid that the entrepreneur is "stuck" for ages because of the liquidation procedure and to allow him to start a new business project as soon as possible.

In addition to the above, the Italian Government is committed to increasing the resources available for venture capital and also provides some specific measures and incentives for incubators or accelerator that fulfill specific requirements concerning the start-up's physical structures, management, facilities and, above all, its track record.

4. Data, Variables and Methodology

2.1 The Data

Our sample includes 1694 innovative start-ups registered at the Chambers of Commerce in Italy. In particular, we restricted our analysis to companies that registered at the Italian Chamber of Commerce between 2009 and 2013 and that are also included in the "innovative start-ups" online directory. Table 1 and 2 report the breakdown of our sample by year of registration to the Chamber of Commerce and by industry.

In order to analyse the impact of the structure of local knowledge bases on the formation of new firms we matched the OECD RegPat Database (July 2012) with data provided by the Eurostat and NUTS3-level data provided by the Italian institute of statistics (ISTAT), specifically the "Indicatori territoriali per le politiche di sviluppo" (local indicators for development policy). The OECD RegPat is derived from the Patstat database, which ensures worldwide coverage, containing tables including patents bibliographic data, citations and family links. These data combine both applications to the EPO and the application to the national patent offices, allowing for going 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 are 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 by inventors residing in Italian regions, and uses International Patent Classification (IPC) maintained by the EPO to assign applications to technological classes.

2.2 The Variables

Our dependent variable is the count of innovative start-ups in each NUTS3 region.

The test of the KSTE traditionally adopts the local expenditure for research and development (R&D) as a proxy of the available pool of technological knowledge at the regional level (Acs et al., 2009). For the sake of comparison with these studies we also include R&D in the analysis, it should be desirable to use the same variable. Unfortunately, there are no available data concerning R&D expenditure at the NUTS 3 level in Italy. For this reason we adopt the local knowledge stock (KSTOCK), which is calculated by using patent applications as it follows. We apply 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, where once again i is the region and t is the time period.

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) use a more traditional approach to the measurement of knowledge, by looking at patent applications. They measure indeed diversity and relatedness relying respectively on the Herfindal index calculated on knowledge fields assigned by the USPTO and on patent citations.

In this paper we will follow an approach close to this latter, in that we will use the information contained in patent documents to calculate a number of variables that characterize the local knowledge base on the basis of the complementarity and similarity degree amongst its components. The implementation of knowledge characteristics proxying for variety, complementarity and similarity, rests on the recombinant knowledge approach. Details on the methodology and on its application to the sector, firm and region level can be found in Krafft, Quatraro, Saviotti (2014); Colombelli, Krafft and Quatraro (2013); Quatraro (2010).

We 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. 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 document .

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

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 average degree of complementarity among technologies making up the regional knowledge base.

c) Cognitive distance (CD) expresses the average degree of dissimilarity amongst different types of knowledge.

Besides the effects of the knowledge indicators, we also control for a number of factors that have proved to affect new firm formation in previous empirical settings. First of all, according to previous studies in this stream of literature, new firm formation may be triggered by pull factors such as high innovative potential (see see Acs and Audretsch, 1989a and 1989b; Geroski, 1995). For this reason we control for the effects of agglomeration economies (AGG), proxied by population density at the NUTS3 in the vector of control variables. A complementary measure of prospective economic benefits is also represented by the distance (DIST) of each province i from the administrative chief town of the NUTS2 region (Baptista and Mendonça, 2002; Bonaccorsi et al. 2013).

Moreover, agglomeration economies can also stem from the presence of other firms in the same place, which ensures to some extent the availability of local markets for intermediate goods. In this direction, we also added as a control variable the firm density (FIRMDENS), calculated as the ratio between the number of registered firms at time t in region i and the land use area

As the creation of new firms can be the outcome of an 'escape from unemployment' strategy, we control for the unemployment rate at the local NUTS 3 level (UNEM), calculated as the ration between the count of unemployed people and the count of individuals in the labour force at time t in region i.

Moreover, we calculated the numbers of incubators (INC) in each province. Actually, business incubators represent a key resource to the creation of new firms, which provide the conditions for successful undertakings and increase the survival likelihood (Colombo and Delmastro, 2002).

In order to control for the size of the region we also included per capita value added (VA) at time t in region i.

Finally, we add time dummies and, in order to address the issue of geographical differences in terms of new firm creation, dummy variables at the NUTS 1 level (i.e. Italian macro-areas).

Table 3 provides a summary of variables definitions.

2.3 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 (KSTE). 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 NISUi,t as a function of the characteristics of the knowledge base.

The baseline specification would therefore be the following:

$$\ln\left(NISU_{i,t}\right) = a + b_1 \ln KSTOCK_{i,t-3} + b_2 \ln CD_{i,t-3} + b_3 \ln COH_{i,t-3} + b_4 \ln KV_{i,t-3} + \tau_t + \rho_i + \varepsilon_{i,t}$$

(1)

As the features of local environments may take some time to exert an effect on entrepreneurial dynamics, we apply a 3 years lag to the explanatory variables.

However, 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 (1) as follows:

$$\ln\left(NISU_{i,t}\right) = a + b_1 \ln KSTOCK_{i,t-3} + b_2 \ln CD_{i,t-3} + b_3 \ln COH_{i,t-3} + b_4 \ln KV_{i,t-3} + b_5 \sum \ln C_{i,t-3} + \tau_t + \rho_i + \varepsilon_{i,t}$$

(2)

Where $\sum \ln C_{i,t-3}$ represents a set of control variables.

Equation 2 can be estimated using the negative binomial estimator. Indeed, the discrete and non-negative nature of the dependent variable suggests the adoption of estimation techniques for 'count data' models. Out of these models, the Poisson regression assumes that the dependent variable follows a Poisson distribution. The Poisson regressions model however assumes the equality between conditional variance and conditional mean in the distribution of the dependent variable. When this condition is not met, like in the present case, the negative binomial (NB) class of models is used, which permits over-dispersion. As suggested by the summary statistics reported in Table 4, our dependent variable appears to be overdispersed, so that the negative binomial estimator is expected to perform better than the Poisson one (Greene, 2003).

Table 4 reports instead the descriptive statistics concerning the variables used in the analysis after log transformation, while Table 5 shows the correlation matrix.

5. Econometric results

The results of the econometric estimations of equation (2) are reported in Table 6 and 7.

Consistently with the KSTE, the coefficient of regional KSTOCK is positive and significant. This supports therefore the idea that entrepreneurs create new firms by taking advantage of the locally available unexploited knowledge. For what concerns the properties of local knowledge bases, one can observe that the coefficient on cognitive distance (CD) is negative and significant, while the coefficient on variety is positive and significant. The same applies to the specification including related knowledge variety (RKV) and unrelated knowledge variety (UKV). The coefficient on the coherence is not significant.

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 negative sign of cognitive distance suggests that a key condition to the creation of new innovative start-ups is the local availability of similar technological competences. Moreover, the positive and significant coefficient of knowledge variety (KV) shows that the increase in the scope of the available competences is likely to favor the creation of new firms.

By looking at the correlation matrix reported in Table 5 it is clear that the very high correlation between knowledge stock and the three specification of knowledge variety (KV, RKV and UKV) may induce a bias in the results. For this reason in Table 7 we show the results of the estimation obtained by dropping knowledge stock from the vector of covariates. These results confirm the robustness of our analysis.

6. Conclusions

Innovative start-ups are considered as a powerful instrument for both stagnant economies to recover and developed ones to growth.

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 high technological diversification of the knowledge base and a high degree of similarity amongst different types of knowledge.

7. References (to be updated)

Acs, Z.J., Braunerhjelm, P., Audretsch, D.B., & Carlsson, B., (2009). The knowledge spillover theory of entrepreneurship, Small Business Economics, 32(1), 15–30.

Acs, Z.J., Audretsch, D.B., & Lehmann, E.E., (2013). The knowledge spillover theory of entrepreneurship. Small Business Economics, 41(4), 757-774.

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

Audretsch, D.B. and Mahmood, T. (1995), New Firm Survival: New Results Using a Hazard Function, Review of Economics and Statistics, 77, 97-103.

Audretsch, D.B., Santarelli, E. and Vivarelli, M. (1999), Start Up Size and Industrial Dynamics: Some Evidence from Italian Manufacturing, International Journal of Industrial Organization, 17, 965-83.

Audretsch, D.B. and Vivarelli, M. (1995), New Firm Formation in Italy, Economics Letters, 48, 77-81.

Audretsch, D.B. and Vivarelli, M. (1996), Determinants of New-Firm Startups in Italy, Empirica 23, 91-105.

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.

Bonaccorsi, A., Colombo, M.G., Guerini, M., Rossi-Lamastra, C. (2013), The impact of local and external university knowledge on the creation of knowledge-intensive firms: evidence from the Italian case, Small Business Economics.

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.

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

Evans, D.S. and Leighton L.S. (1989), Some Empirical Aspects of Entrepreneurship, American Economic Review, 79, 519-35.

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.

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.

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. (1934), The Theory of Economic Development, Cambridge (Mass.), Harvard University Press.

Schumpeter, J.A. (1939), Business Cycles: A Theoretical, Historical and Statistical Analysis of the Capitalist Process, New York, McGraw-Hill.

Vivarelli, M. (1991), The Birth of New Enterprises, Small Business Economics, 3, 215-23.

Vivarelli, M. (2004), Are All the Potential Entrepreneurs So Good?, Small Business Economics, 23, 41-9.

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

8. Appendix A – The Calculation of knowledge properties

6.1 Knowledge variety measured by the informational entropy index

Knowledge variety is measured using the information entropy index[1]. 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). 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). 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:

$$KV \equiv H(X,Y) = \sum_{l}\sum_{j} p_{lj} \log_2\left(\frac{1}{p_{lj}}\right) \tag{A1}$$

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:

$$P_{gz} = \sum_{l \in S_g}\sum_{j \in S_Z} p_{lj} \tag{A1a}$$

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

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:

$$H_{gz} = \sum_{l \in S_g} \sum_{j \in S_z} \frac{p_{lj}}{P_{gz}} \log_2 \left(\frac{1}{p_{lj}/P_{gz}} \right) \tag{A1b}$$

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

$$RKV \equiv \sum_{g=1}^{G} \sum_{z=1}^{Z} P_{gz} H_{gz} \tag{A2}$$

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

$$UKV \equiv H_Q = \sum_{g=1}^{G} \sum_{z=1}^{Z} P_{gz} \log_2 \frac{1}{P_{gz}} \tag{A3}$$

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

$$KV = H_Q + \sum_{g=1}^{G} \sum_{z=1}^{Z} P_{gz} H_{gz} \tag{A4}$$

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.

6.2 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).

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, $O_l \cap O_j \neq \varnothing$, 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:

$$\Omega = \begin{bmatrix} J_{11} & & J_{l1} & & J_{n1} \\ \vdots & \ddots & & & \vdots \\ J_{1j} & & J_{lj} & & J_{nj} \\ \vdots & & & \ddots & \vdots \\ J_{1n} & \cdots & J_{\ln} & \cdots & J_{nn} \end{bmatrix} \tag{A5}$$

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:

$$\mu_{lj} = E(X_{lj} = x) = \frac{O_i O_j}{K} \tag{A6}$$

$$\sigma_{lj}^2 = \mu_{lj}\left(\frac{K - O_l}{K}\right)\left(\frac{K - O_j}{K - 1}\right) \tag{A7}$$

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:

$$\tau_{lj} = \frac{J_{lj} - \mu_{lj}}{\sigma_{lj}} \tag{A8}$$

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}:

$$\mathrm{WAR}_{\mathrm{lt}} = \frac{\sum_{\mathrm{j \neq l}} \tau_{\mathrm{lj}} \mathrm{P}_{\mathrm{jt}}}{\sum_{\mathrm{j \neq l}} \mathrm{P}_{\mathrm{jt}}} \tag{A9}$$

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

$$COH_{\mathrm{t}} = \sum_{\mathrm{l}} \mathrm{WAR}_{\mathrm{lt}} \times \frac{\mathrm{P}_{\mathrm{lt}}}{\sum_{\mathrm{l}} \mathrm{P}_{\mathrm{lt}}} \tag{A10}$$

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.

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.

6.3 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.

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:

$$S_{lj} = \frac{\sum_{m=1}^{n} V_{lm} V_{jm}}{\sqrt{\sum_{m=1}^{n} V_{lm}^2} \sqrt{\sum_{m=1}^{n} V_{jm}^2}} \qquad \text{(A11)}$$

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 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)[2]. 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:

$$d_{lj} = 1 - S_{lj} \tag{A12}$$

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.

$$WAD_{lt} = \frac{\sum_{j \neq l} d_{lj} P_{jt}}{\sum_{j \neq l} P_{jt}} \tag{A13}$$

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:

$$CD_t = \sum_l WAD_{lt} \times \frac{P_{lt}}{\sum_l P_{lt}} \tag{A14}$$

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

[2] 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.

paradigmatic shifts in the sector knowledge base. It signals the combination of core technologies with unfamiliar technologies.

Table 1. Number of innovative start-ups by year, 2009-2013.

Year	%
2009	6%
2010	10%
2011	15%
2012	24%
2013	45%
Tot	**100%**

Table 2. Number of innovative start-ups by industry, 2009-2013

Industry	%
Manufacturing	17%
Services	81%
Other	2%
TOT	**100%**
HT	5%
MHT	8%
MLT	2%
LT	3%
KISA	14%
KISB	57%
KISC	0%
KISD	4%
LKISA	6%
LKISD	0%
Other	2%
TOT	**100%**

Table 3. Description of the variables used in the analysis

Variable	Description
NISU	Number of innovative start-ups at time t in region i
KSTOCK	logarithm of regional knowledge stock of region i
COH	logarithm of knowledge coherence of region i
KV	logarithm of knowledge variety of region i
RKV	logarithm of related knowledge variety of region i
UKV	logarithm of unrelated knowledge variety of region i
CD	logarithm of cognitive distance of region i
AGGL	logarithm of the ratio between population and the area (square-km) of region i
DIST	logarithm of the distance of each region i from the administrative chief town of the NUTS2 region
FIRMDENS	logarithm of the ratio between new registered firms and the local population at time t in region i
UNEM	logarithm of unemployment rate of region i
INC	logarithm of the number of incubators in region i
VA	logarithm of the ratio between value added and the local population at time t in region i

Table 4. Descriptive statistics

Variable	Obs	Mean	Std.Dev.	Min	Max
NISU	515	3.254	8.100	0	114
KSTOK	487	4.985	1.468	1.213	8.857
COH	490	1.148	0.227	0.273	1.971
CD	490	-0.663	0.130	-1.356	-0.337
KV	490	1.598	0.517	0.000	2.302
UKV	490	0.945	0.347	0.000	1.354
RKV	490	1.220	0.509	0.000	2.023
AGG	490	5.234	0.741	3.735	7.869
DIST	490	3.450	1.735	0.000	5.057
FIRMDENS	490	0.241	0.043	0.146	0.432
UNEM	490	0.028	0.013	0.009	0.074
INC	490	0.311	0.484	0.000	1.946
VA	490	0.022	0.005	0.011	0.040

Table 5. Correlation matrix

	KSTOCK	COH	CD	KV	UKV	RKV	UNEM	FIRMDENS	INC	DIST	AGG	VA
KSTOCK	1.0000											
COH	0.0668	1.0000										
CD	0.1608*	-0.1140	1.0000									
KV	0.8632*	0.0888	0.1149	1.0000								
UKV	0.7067*	0.0853	0.0623	0.8390*	1.0000							
RKV	0.8656*	0.0696	0.1342*	0.9405*	0.6285*	1.0000						
UNEM	-0.5576*	-0.2119*	-0.1101	-0.5142*	-0.5047*	-0.4587*	1.0000					
FIRMDENS	-0.6065*	-0.0524	-0.0723	-0.5836*	-0.5666*	-0.5314*	0.5890*	1.0000				
INC	0.4483*	-0.0294	0.2061*	0.3111*	0.2265*	0.3363*	-0.0927	-0.1741*	1.0000			
DIST	-0.2447*	0.0084	-0.1312*	-0.1296*	-0.1495*	-0.1242*	-0.0267	0.1094	-0.5158*	1.0000		
AGG	0.5424*	-0.0313	0.0001	0.4489*	0.3928*	0.4327*	-0.0922	-0.2684*	0.3105*	-0.3564*	1.0000	
VA	0.7635*	0.1504*	0.1467*	0.6456*	0.5255*	0.6489*	-0.7222*	-0.6745*	0.3555*	-0.1657*	0.2402*	1.0000

Table 6. Results, 2009-2013

Dep var NISU	(1)	(2)	(3)	(4)	(5)
L3.KSTOCK	0.5950***	0.4428***	0.5470***	0.4926***	0.4207***
	(0.0561)	(0.0848)	(0.0614)	(0.0844)	(0.0900)
L3.COH		-0.0555	-0.0219	-0.0230	-0.0505
		(0.2302)	(0.2296)	(0.2291)	(0.2306)
L3.CD		-1.0933***	-0.9961**	-1.0939***	-1.0624***
		(0.4051)	(0.4062)	(0.4067)	(0.4068)
L3.KV		0.6489**			
		(0.2635)			
L3.UKV			0.5129**		0.5398**
			(0.2306)		(0.2322)
L3.RKV				0.3994*	0.4377*
				(0.2269)	(0.2319)
L3. UNEM	15.2792**	18.4858***	20.3232***	17.2700***	19.4156***
	(6.0391)	(5.9658)	(6.0718)	(6.0288)	(6.0549)
L3.FIRMDEN	1.2896	2.2540	2.4578	1.5245	2.7080
	(1.7206)	(1.7766)	(1.8124)	(1.7483)	(1.8153)
L3. INC	0.3670***	0.4535***	0.4221***	0.4318***	0.4612***
	(0.1048)	(0.1064)	(0.1050)	(0.1067)	(0.1065)
L3.DIST	-0.1466***	-0.1568***	-0.1532***	-0.1579***	-0.1531***
	(0.0264)	(0.0266)	(0.0267)	(0.0267)	(0.0266)
L3.AGG	0.0001	-0.0119	-0.0275	-0.0041	-0.0128
	(0.0669)	(0.0668)	(0.0673)	(0.0676)	(0.0673)
L3.VA	24.9360	28.2418*	29.4695*	24.7960	31.3343*
	(17.1076)	(16.7307)	(16.9595)	(16.8001)	(16.8564)
_cons	-2.5217***	-3.9237***	-3.8646***	-3.3624***	-4.0265***
	(0.6766)	(0.8017)	(0.8070)	(0.7600)	(0.8092)
lnalpha					
_cons	-1.4155***	-1.5438***	-1.5260***	-1.5119***	-1.5488***
	(0.1907)	(0.1995)	(0.1976)	(0.1968)	(0.1994)
N	500	487	487	487	487
pseudo R^2	0.250	0.258	0.257	0.256	0.259
AIC	1687.2903	1649.7541	1651.1737	1653.0968	1649.5508
BIC	1758.9386	1733.5194	1734.9390	1736.8621	1737.5044

Table 7. Robustness check, 2009-2013

Dep var NISU	(1)	(2)	(3)	(4)
L3.COH	-0.0879	0.1271	-0.0069	-0.0666
	(0.2338)	(0.2431)	(0.2331)	(0.2341)
L3.CD	-1.2239***	-1.0085**	-1.3095***	-1.2049***
	(0.4049)	(0.4343)	(0.4074)	(0.4070)
L3.KV	1.7659***			
	(0.1920)			
L3.UKV		1.3903***		0.9531***
		(0.2275)		(0.2217)
L3.RKV			1.4223***	1.2540***
			(0.1606)	(0.1636)
L3. UNEM	18.0382***	21.5869***	14.1626**	18.5503***
	(6.0178)	(6.5700)	(6.1330)	(6.1035)
L3.FIRMDEN	3.5734**	3.7072*	1.8123	3.7805**
	(1.7818)	(1.9275)	(1.7802)	(1.8207)
L3. INC	0.6570***	0.7535***	0.6488***	0.6446***
	(0.1009)	(0.1099)	(0.1040)	(0.1010)
L3.DIST	-0.1505***	-0.1282***	-0.1551***	-0.1469***
	(0.0271)	(0.0294)	(0.0278)	(0.0272)
L3.AGG	0.0970	0.2254***	0.1524**	0.0929
	(0.0649)	(0.0689)	(0.0654)	(0.0648)
L3.VA	51.3629***	83.9943***	46.0952***	52.0177***
	(16.6393)	(18.7676)	(17.2821)	(16.8038)
_cons	-5.1498***	-5.2119***	-3.7570***	-4.8249***
	(0.7882)	(0.8869)	(0.7861)	(0.8091)
lnalpha				
_cons	-1.4434***	-1.0668***	-1.3366***	-1.4464***
	(0.1937)	(0.1645)	(0.1874)	(0.1926)
N	490	490	490	490
pseudo R^2	0.246	0.222	0.241	0.249
AIC	1678.0218	1729.9859	1690.2932	1672.9106
BIC	1757.7155	1809.6796	1769.9869	1756.7987