THE COST OF KNOWLEDGE

Cristiano Antonelli
Dipartimento di Economia, Università di Torino
and
BRICK (Bureau of Research in Innovation Complexity and Knowledge), Collegio Carlo Alberto

and

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
and
BRICK (Bureau of Research in Innovation Complexity and Knowledge), University of Nice Sophia Antipolis, CNRS-GREDEG

ABSTRACT. This paper contributes the economics of knowledge and innovation with the analysis of the knowledge cost function and sheds light on the determinants of the large variance in the cost of innovation across firms. The amount and the structure of external knowledge and the internal stocks of knowledge that firms can access and use in the generation of new technological knowledge help firms to reduce the costs of innovations. The empirical section is based upon a panel of companies listed on UK and the main continental Europe financial markets (Germany, France and Italy) for the period 1995 – 2006, for which information about patents have been gathered. The econometric analysis of the costs of knowledge considers the unit costs of patents on the right hand side, and on the left hand side next to R&D expenditures, the stock of knowledge internal and external to each firm. In order to articulate the different facets of the external knowledge that is made accessible by proximity with firms co-localized in the same region (NUTS2), we further include other variables proxying for regional variety, complementarity and similarity. The results confirm that the stock of internal knowledge and the access to external knowledge play a key role in reducing the actual cost of the generation of new technological knowledge at the firm level.

JEL CODES: O30

KEY WORDS: KNOWLEDGE AS AN OUTPUT; KNOWLEDGE STOCK; EXTERNAL KNOWLEDGE.

1. INTRODUCTION

The identification of the knowledge generation function has been a major progress in the economics of knowledge (Weitzman, 1998, Crépon, Duguet, Mairesse, 1998). Finally technological knowledge can be analyzed as the output of a dedicated economic activity. Working along these lines increasing evidence shows that the unit costs of innovations differ widely across firms. Some firms are able to introduce an innovation with low levels of current expenditures in R&D. Others experience very high levels of current expenditures. As a matter of fact the costs of innovations differ and their variance becomes a fascinating area of research. The study of the cost of innovation seems an important area of investigation that has received, so far, quite surprisingly, very little attention. After the introduction of the knowledge generation function it is now necessary to introduce the knowledge cost function.

The new appreciation of the role of knowledge indivisibility in the generation of new knowledge leads to better grasping the specific effects of knowledge externalities and knowledge cumulability on the costs of knowledge. Knowledge cumulability and knowledge externalities in fact do affect the costs of innovations. The rest of the paper is structured as it follows. Section 2 recalls the recent acquisitions of the new economics of knowledge and applies them to grasping the determinants of the heterogeneity of firms in terms of unit costs of their innovations. Section 3 provides an empirical investigation based on the econometric estimate of an innovation cost function based upon a panel of companies listed on UK and the main continental Europe financial markets (Germany, France and Italy) for the period 1995 – 2006, for which information about patents have been gathered. The conclusions summarize the results.

2. INNOVATION AS AN INPUT AND OUTPUT

Building upon the arrovian foundations of the analysis of knowledge as an economic good, the new economics of knowledge has made it possible to grasp the dynamic characteristics of the generation of knowledge (Nelson, 1959; Arrow, 1962 and 1969). After a long period of time during which the early economics of knowledge has investigated in depth the determinants and the effects of the limited appropriability of knowledge as an economic good, the new economic of knowledge paid much attention to elaborating the implications of knowledge indivisibility in terms of both knowledge cumulability and knowledge complementarity. Knowledge is at the same time an input and an output. Its generation consists in the recombination of knowledge items that enter the process as inputs (Nelson, 1982). Because of knowledge complementarity and knowledge cumulability, next to current R&D activities both external knowledge generated by third parties but not fully appropriated and the internal stocks of knowledge generated by each firm in the past, are now recognized as relevant inputs into the generation of knowledge as an output. The knowledge generated as the output of a dedicate activity is itself a necessary condition and hence an input for the introduction of an innovation (David, 1993).

In turn the introduction of an innovation is the result of complex process that mobilizes the competence of the firm, the amount of external knowledge that can be used, and relies on the dedicated resources that are specifically invested. The introduction of an innovation requires the generation of new knowledge. The introduction of an innovation is actually the output of an activity that entails many different inputs and a specific competence (Lööf and Heshmati, 2002). These distinctions make clear that much more attention is necessary to use the standard indicators. Some of the traditional indicators such as patents and innovation counts actually measure the output of the innovation process. Others, namely R&D, are more precisely input measures.

This has led to the analysis of the generation of technological knowledge as a specific economic activity (Crépon, Duguet, Mairesse, 1998 ; Nesta, Saviotti, 2005; Lööf and Heshmati, 2002). A second important step in this enquiry can be done with the analysis of the knowledge cost function. This approach enables to identify the determinants of the great variance in the costs of innovations.

Specifically the study of the knowledge cost function helps grasping to what extent the cost of knowledge is affected by the costs of inputs.

As soon as it becomes evident that R&D are not the single input into the innovation process (Gunday et al., 2011), in fact, the actual access conditions to the other knowledge inputs acquire a new relevance. The other knowledge inputs such as the amount of external knowledge that can be accessed by firms to generate new knowledge and eventually introduce new technologies are distributed unevenly across space. Major institutional and structural characteristics affect the actual amount of external knowledge that each firm can use as an input (Cohen and Levinthal, 1989). The costs of inputs may differ in turn because of the variance in the access conditions to the external knowledge available and because of the different characteristics of the local pools of external knowledge. For the same token firms differ widely with respect to the size and the characteristics of the stocks of internal knowledge that can be used to generate new knowledge (Jones, 1995). Input and output do not coincide especially when firms differ in their specific competence in managing the innovation process.

The inclusion in the knowledge cost function of these variables stems from the identification of the recombinatorial character of the knowledge generation process and enables to appreciate the role of knowledge indivisibility, as articulated in knowledge cumulability and knowledge complementarity in its generation (Weitzman, 1996 and 1998). Let us consider them in turn.

Knowledge cumulability – and its limited exhaustibility – implies that the stock of existing knowledge can be used again and again and plays a central role as an input into the generation of new knowledge. The stock of knowledge qualifies and identifies the knowledge base of each firm. The inclusion of this variable enables to grasp the path dependent character of the knowledge generation. The generation of new technological knowledge at each point in time, by each agent, in fact, is strongly influenced by the accumulation of knowledge in the past. The current levels of R&D expenditures of each agent do play a role but only in a context that is shaped by the past of each firm (Antonelli, 2011; Belenzon, 2012).

The appreciation of knowledge complementarity enables to put in context the role of knowledge externalities. A large literature had explored the role of technological spillovers as a major input into the generation of new technological knowledge (Colombelli et al., 2013). In this approach external knowledge plays an important and yet supplementary role in the generation of new technological knowledge (Griliches, 1979, 1990, 1992). Moreover its recipients are mainly viewed as the passive beneficiary of knowledge leaking from other firms (Feldman, 1999). A large body of empirical evidence has subsequently confirmed that external knowledge is an essential input into the generation of new knowledge (Adams, 1990) The interplay between internal knowledge, which also increases a firm's absorptive capacity, and knowledge externalities may increase the firm's propensity to innovate (Smit et al., 2013) and productivity (Marrocu et al., 2012).

The characteristics of the regional context into which firms are located play an important role in assessing the actual weight of absorption costs. The composition of the knowledge pools to which co-localized firms have access plays an important role (Grillitsch et al., 2013; Camagni and Capello, 2013). Technological knowledge cannot be regarded as a homogeneous pile but rather as a composite bundle of highly differentiated and idiosyncratic elements that are qualified by specific relations of interdependence and interoperability. This approach enables to identify the extent to which the generation of new technological knowledge in a field depends upon the contributions of knowledge inputs stemming from other fields: a new knowledge item exhibits high levels of compositeness when it relies upon a large number of other knowledge fields (Antonelli, 2011). The quality of the local pools of knowledge in other words matters as well as its sheer size. The larger is the coherence of the local knowledge base and shorter is the distance between different types of knowledge, the higher is the probability that they can be combined together (Saviotti, 2004 and 2007; Kraft, Quatraro, Saviotti, 2009; Quatraro 2010).

The analysis of a knowledge cost function that takes into account the role of the internal stocks of knowledge and the local pools of external knowledge enables to consider again and yet from quite a different perspective two standard assumptions of the economics of innovations i.e. the well-known Schumpeterian and Marshallian hypotheses. Let us consider them in turn.

a) the Schumpeterian hypothesis. Joel Mokyr (1990:267) has recently masterly summarized Schumpeterian hypothesis as follows 'large firms with considerable market power, rather than perfectly competitive firms are the 'most powerful engine of technological progress''. Schumpeter with his *Capitalism, Socialism and Democracy* went actually so far as to claim that perfect competition is not only impossible but inferior' (Schumpeter 1942:106). The Schumpeterian hypothesis has fed a long lasting theoretical debate and the large empirical provided controversial evidence on the actual advantages of large firms with respect to smaller ones in the rates of introduction of innovations. The results of the empirical studies in different sectors, historic periods, countries and regions have not provided conclusive evidence (Link and Siegel, 2007). According to our approach, we articulate the more specific hypothesis, that firms with a larger stock of internal knowledge have indeed an important advantage on firms with no internal stocks. The advantage in other words stems specifically from the effects of knowledge cumulability and are specific to the size of the stock of knowledge

b) the Marshallian hypothesis. According to the Marshallian hypothesis, firms located in large industrial districts have better chances to access knowledge spillovers and feed their own knowledge generation costs. In large districts firms have better access to external knowledge and can substitute expensive R&D activities for the cheap external one. The size of the district favours the introduction of technological innovations at low costs. According to our approach, based upon the analysis of the composition of the local pools of knowledge that yield knowledge externalities, knowledge externalities are all the more effective, the larger are the levels of coherence of the local knowledge base.

The following knowledge cost function (1) provides the general frame of our approach:

$$CK_{it}=(R\&D_{it}KNOWLEDGEBASE_{it}\ EXTERNAL\ KNOWLEDGE_{it}) \qquad (1)$$

Equation (1) provides a suitable specification of the knowledge cost function, that accommodates, next to the role of R&D expenditures, the appreciation of the burden of knowledge associated with the knowledge base of each firms in terms of the levels of the stock of knowledge in the generation of new knowledge, the identification of the key role of knowledge external to each firm but available in regional proximity.

3. EMPIRICAL EVIDENCE

3.1 Dataset

Our source of data is the IPER[1] database, which collects information on 3382 active companies listed on the main European markets (UK, Germany, France, Italy and the Netherlands). The IPER database has been built by matching information from multiple sources of data. Our main source of

[1] The implementation of the IPER database has been financed by the Collegio Carlo Alberto, under the IPER project.

market and accounting data is Thomson Datastream, which delivers worldwide economic and financial time series data. To obtain additional relevant variables, we include in the dataset information collected from AMADEUS by Bureau Van Dijk, which contains financial information on European companies. In order to match information from the two databases described above, we used the ISIN code, the International Securities Identification Number (ISIN) which uniquely identifies a security.

We also use data from the OECD REGPAT database, which provides regional information on the addresses of patent applicants and inventors as well as on technological classes cited in patents granted by the European Patent Office (EPO) and the World Intellectual Property Organization (WIPO), under the Patent Co-operation Treaty (PCT), from 1978 to 2006. In order to match the firm level data with data on patents, we draw on the work of Thoma et al. (2010), which develops a method for harmonization and combination of large-scale patent and trademark datasets with other sources of data, through standardization of applicant and inventor names. The new evidence about the actual meaning of patent citations, often included by patent officers to better specify the borders of the domain of the intellectual property right, rather than its quality, suggests to use the raw evidence of the number of patents with no attempt to try and elaborate misleading quality indicators (Van Zeebroeck, 2011; Van Zeebroeck and van Pottelsberghe 2011).

Finally, we pooled the dataset by adding industry level information from the STAN database, which provides information at the industry level for the OECD countries. As STAN is based on the ISIC revision 3 sectoral classifications and Thomson Datastream uses the four digit level ICB industry classification, we provide in Appendix A the sectoral concordance table used to link the two classifications.

Our final dataset includes active companies listed on the main European financial market that submitted at least one patent application to the EPO in the period analysed. Table 1 reports the sample distribution by macro-sector classes. High and medium-high technology firms account for around 32% and 45% of observations, respectively. Medium low and low technology firms account for 4% and 9% respectively, while knowledge intensive firms represent some 10% of observations.

Table 1 about here

3.2 Methodology and Variables

Our estimating equation is the following (2):

$$PCost_{it} = \beta_1 + \beta_2 R\&D_{it-1} + \beta_3 PStock_{it-1} + \beta_4 Size_{it-1} + \beta_5 RegPStock_{it-1} + \beta_6 RegTV_{it-1} + \beta_7 RegCD_{it-1} + \beta_8 RegCOH_{it-1_i} + \sum \rho_i + \sum \psi_t + \varepsilon_{it} \quad (2)$$

Equation (2) has been estimated using a fixed effects estimator.

In Equation (2) the dependent variable for the firm *i* at time *t* is the cost of knowledge output measured by the ratio of the firm current R&D expenditures to the number of patents delivered and it is explained by three set of independent variables that are respectively the internal expenses in R&D, the burden of the knowledge base of each firm as defined by the size of the knowledge stock, and the composition of external knowledge in terms of variety, complementarity and similarity.

More precisely, on the right hand side, the first set of variables considers *R&D*, i.e. the current research efforts and activities funded by each firm at time t-1, measured as the ratio of R&D expenditures to total assets in logarithms. In order to appreciate the effects of the stocks of internal knowledge of firms, we have included the variable *PStock* measured in terms of the number of patents held by each firm. This is computed by applying the permanent inventory method (PIM) to patent applications. We calculate it as the cumulated stock of patent applications using a rate of obsolescence of 15% per annum[2]:

$$PStock_{it} = \dot{h}_{it} + (1 - \sigma)PStock_{it-1} \tag{3}$$

where $\dot{h}_{it}$ is the flow of patent applications and δ is the rate of obsolescence. Moreover, we include the variable *Size*, measured as the log of employees' number for firm i at time t-1.

We do include other variables to articulate the different facets of the knowledge that is external to each firm and made accessible by proximity with firms co-localized in the same region. The first variable is *RegPStock*, that is the log of patents stock (PIM) in the same region (NUTS2) of firm i at time t-1. The method used for computing this variable is the same used for *PStock*.

We further include other variables proxying for variety, complementarity and similarity. These indicators rest on the recombinant knowledge approach. In order to provide an operational

[2] A 15% obsolescence rate is the most common value used in the literature (see, for example, Nesta, 2008; Colombelli *et al.* 2013, 2014). As a robustness check we also experimented with alternative obsolescence rates. We found that the obsolescence rate value makes little difference in empirical estimations.

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. 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, i.e. the nodes of the network representation of recombinant knowledge. 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 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. We thus include in equation (2) *RegTV*, as a measure of the regional total variety, *RegRTV* and *RegUTV*, measuring the related and unrelated variety respectively, (see Appendix B for the methodological details).

b) Knowledge coherence (COH) measures the degree of complementarity among technologies. It is measured by means of the *RegCOH* index (see Appendix B).

c) Cognitive distance (CD) expresses the dissimilarities amongst different types of knowledge and is measured using the *RegCD* variable (see Appendix B).

The adoption of these variables marks an important step forward in the operational translation of knowledge creation processes. In particular, they allow for a better appreciation of the collective dimension of knowledge dynamics. Knowledge is indeed viewed as the outcome of a combinatorial activity in which intentional and unintentional exchange among innovating agents provides the access to external knowledge inputs (Fleming and et al., 2007).

The recombinant knowledge approach provides indeed a framework to represent the internal structure of regional knowledge bases as well as to enquire into the effects of their 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.

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 the regional technological trajectory (Dosi, 1982; March, 1991; Saviotti, 2004 and 2007; Colombelli, Krafft, Quatraro, 2013; Quatraro, 2010).

We finally include *Size*, measured by means of the employees' number, to control for firm size, and time dummies in order to control for time effects.

In order to check further the robustness of our empirical analysis with respect to the role of the external knowledge, we also estimated an extended model including the patenting activities of firms localized outside the firm's region (*WRegPStock*) Here *WRegPStock* aims at capturing the role of the sources of external knowledge that are far away from firm i. The variable *WRegPStock* has been computed as the log of patents stock (PIM) in the NUTS2 regions of the EU-24 member states, weighted using a row-normalized inverse distance matrix so as to appreciate the contribution of knowledge produced in regions close to firm's i region at time t-1.

For each variable the measurement method is defined in Table 2, while descriptive statistics are reported in Table 3. The correlation matrix can be found in Table 4.

Table 2, 3 and 4 about here

3.3 Results

The results of the fixed effects regression estimations for Equation (2) are reported in Table 5. The Hausmann test, comparing the results yield with the fixed effects model with those obtained from the random effects regression model, indicates that the fixed effects model is a better fit for our regressions. In order to cope with multicollinearity among the knowledge-related variables, column 1 shows the results for the baseline equation that only includes variables measuring the internal activities performed by each firm in terms of R&D expenditure and patents stock. Columns 2 to 5 include also the variables proxying for the size and composition of the external pool of knowledge. More precisely, the results of the model including the *RegTV* variable are presented in column 2. Columns 3 and 4 show the results for the *RegRTV* and *RegUTV* variables, respectively, while column 5 includes the two latter variables in the same model.

Table 5 about here

The results of the econometric exercise confirm the hypotheses that, next to R&D expenses, the knowledge cost function depends upon the role of both the internal stocks of knowledge and the local pools of external knowledge.

The intensity of R&D expenses is positively and significantly related with the cost of patents in all the estimations. This is quite in line with the expectations, being R&D intensity a measure of the technological efforts of the firm. The stock of patents (*PStock*) of each firm exerts instead a strong negative and significant effect ($p<0.01$ in all estimations) on the costs of patents. This is also consistent with expectations, as the dependent variable is a measure of the unit costs of patents, which is likely to decrease as patents increase, other things being equal.

The results of the variables that account for the regional knowledge base differ whether they concern the size, measured using the stock of patents or the knowledge structure in terms of variety (*RegTV*), complementarity (*RegCD*) and similarity (*RegCOH*). In particular, the size of the regional knowledge stock (*RegPStock*) exerts a negative and significant effect on the cost of knowledge in two out of four estimations (column 2 and 3)[3]. This would suggest that in contexts wherein companies that can access large pools of external knowledge save on the costs of their internal knowledge generating activities. As far as knowledge variety is concerned, the results reported in column 2 show that the *RegTV* is positively and significantly related to the firm cost of patents. Let us recall that this index provides a measure of the differentiation of observed combinations of technologies in regions' knowledge bases. The results thus indicate that the higher is the variety in the combination of technologies in the firm region the higher is the cost associated to the firm innovation output. This might be due to the fact that firms need to put higher efforts in trying and

[3] In columns (4) and (5) the regional knowledge stock is no longer significant. However, by looking at table 4, one may notice that such variable has high correlation with regional related and unrelated variety. This may affect the significance level of *RegPStock*. For the same reason, we also run the regressions in columns (4) and (5) by dropping regional knowledge stock, so as to check the robustness of the results concerning related and unrelated variety. The results actually do not change. The tables are available from the authors.

experimenting new combinations of technologies distributed across a wide range of technology domains. When we disentangle the effects of related and unrelated variety we find that only the latter (*RegUTV*) is significant. The procedure by which the index is derived (see Appendix B) reveals that the concepts of 'related' and 'unrelated' variety refer basically to the belonging of technologies to the same technological domain, as defined by the classification system used (in our case the International Patent Classification). The positive and significant impact of *RegUTV* on the cost of knowledge would imply that an increase in the regional variety of technologies that belong to very different technological domains is likely to increase the costs of knowledge generating activities at the firm level. The unit cost of patents increases as an effect of the higher volume of resources that the firm needs to commit in order to better absorb the locally available external knowledge.

Such result is confirmed by the evidence concerning the effect of regional cognitive distance. The coefficient is indeed positive and significant across all of the four models in which it is included. The cognitive distance may be interpreted as an index of the average dissimilarity amongst the different technological competences that make up the regional knowledge base. When the local knowledge which firms may access, is featured by technologies which are far away from one another in the technological space, firms need to strengthen their absorptive capacity by widening the scope of technological domains that they can master in order to take advantages of knowledge spillovers in the generation of new knowledge. This implies increasing volumes of firm-level R&D expenditures per single patent.

As a robustness check, we further estimated an extended model including the patenting activities of firms localized outside the firm's region (*WRegPStock*). Table 6 reports the results of the fixed effects regression estimations for the equations including *WRegPStock*. These results confirm the robustness of our analysis as regards the variables included in the baseline model. *WRegPStock* turns out to be negatively and significantly related to the cost of knowledge in all regressions. This result confirms that firms are able to absorb, not only knowledge generated in close proximity to the firm, but also external knowledge generated very far from it and, consequently, to save on the costs of their internal knowledge generating activities.

Table 6 about here

4. CONCLUSIONS AND IMPLICATIONS FOR FURTHER RESEARCH

The economics of knowledge has made a major progress with the identification of the knowledge generation function. This empirical evidence has shown that the relationship between inputs and outputs of the innovative activity across firms exhibits a huge variance. With given levels of R&D inputs, the actual amount of knowledge generated by each firm differs. A second, important step along this line of analysis can be done with the analysis of the knowledge cost function. This approach can help understanding why the cost of innovation is far from homogeneous. This evidence has been rarely detected in the literature and poorly investigated.

The study of the knowledge cost function enables to analyze the role of the different cost items that concur to the definition of the knowledge output. This innovative approach enables to explore in a novel perspective two important hypotheses that are at the core of the economics of innovation. Namely the so-called Schumpeterian hypothesis according to which firms with larger stock of internal knowledge are superior in the generation of new knowledge and the so-called Marshallian hypothesis according to which knowledge externalities exert positive effective according not only to the density but also to the levels of coherence of the local pools of knowledge.

The empirical analysis of the costs of innovation, based upon a panel of companies listed on UK and the main continental Europe financial markets (Germany, France and Italy) for the period 1995 – 2006, for which information about patents have been gathered, has considered the unit costs of patents on the right hand side, and on the left hand side next to R&D expenditures, the stock of knowledge internal and external to each firm. The results confirm that the stock of internal knowledge and the access to external knowledge play a key role in assessing the actual capability of each firm to generate new technological knowledge and hence in reducing the costs of innovation.

The results of our analysis also bear important implications for technology policy at the regional level as well as for the strategic management of the firm. Technology policy represents indeed one of the key levers that policymakers may use to trigger local development. Due to the collective and systemic nature of innovation activities, the choice of the correct policy mix is of crucial importance. The promotion of specific technological domains at the local level may affect the effectiveness of knowledge generation processes of incumbents firms. In this direction the attempts to foster the emergence of technologies which break the competences accumulated in the region are likely to increase the average level of unrelated variety and cognitive distance, and as a consequence, increase the average cost per patent.

On the firm's side, the composition of technological activities at the local level becomes a key variable that firms should take into account in their decisions concerning the location of their R&D laboratories. The location in areas featured by established technological trajectories is indeed likely to make the search process for new combinations of technologies less costly for innovating firms as compared to the location in areas marked by initial stages of new technological trajectories.

Acknowledgements. This paper contributes the research project ‘Policy Incentives for the Creation of Knowledge: Methods and Evidence’ (PICK-ME), funded by the European Union D.G. The institutional support of the Collegio Carlo Alberto under the IPER project is acknowledged.

5. REFERENCES

Adams, J. D, (1990), Fundamental stocks of knowledge and productivity growth, *Journal of Political Economy*, 98, 673-702.

Anderson, P.W., Arrow, K.J., Pines, D. (eds.) (1988), *The economy as an evolving complex system*, Addison Wesley, Redwood.

Antonelli, C. (ed.) (2011), *Handbook on the economic complexity of technological change*, Edward Elgar, Cheltenham.

Antonelli, Patrucco, P.P., Quatraro, F. (2011), Productivity growth and pecuniary knowledge externalities: An empirical analysis of agglomeration economies in European regions*, Economic Geography* 87, 23-50

Arrow, K. J. (1962), Economic welfare and the allocation of resources for invention, in Nelson, R. R. (ed.) *The rate and direction of inventive activity: Economic and social factors,* Princeton University Press for N.B.E.R., Princeton, pp. 609-625.

Arrow, K. J. (1969), Classificatory notes on the production and transmission of technical knowledge, *American Economic Review* 59, 29-35.

Belenzon, S. (2012), Cumulative innovation and market value: Evidence from patent citations, *Economic Journal* 122, 265–285.

Camagni, R., Capello, R. (2013), Regional innovation patterns and the eu regional policy reform:

Toward smart innovation policies, *Growth and Change*, 44 (2) , pp. 355-389.

Cohen, W.M., Levinthal, D.A. (1989), Innovation and learning: The two faces of R&D*, Economic Journal* 99, 569-596.

Cohen, W.M., Levinthal, D.A. (1990), Absorptive capacity: A new perspective on learning and innovation, *Administrative Science Quarterly* 35, 128-152.

Colombelli, A., Krafft, J. and Quatraro, F. 2013. Properties of knowledge base and firm survival: Evidence from a sample of French manufacturing firms, *Technological Forecasting and Social Change*, 80, 1469-1483.

Colombelli A., Krafft J., Quatraro F. (2014), High growth firms and knowledge structure: evidence from a sample of listed companies, Industrial and Corporate Change, 23 (1), 261-291.

Colombelli A., Foddi M., Paci R. (2013), "Scientific regions" in Capello R., Lenzi C. (eds.) *Territorial Patterns of Innovation an Inquiry on the Knowledge Economy in European Regions*, 43-69, Routledge.

Crépon, B., Duguet, E., Mairesse, J. (1998), Research and development, innovation and productivity: An econometric analysis at the firm level, *Economics of Innovation and New Technology* 7, 115–158.

David, P.A. (1993), Knowledge property and the system dynamics of technological change, *Proceedings of the World Bank Annual Conference on Development Economics,* The World Bank,Washington.

Feldman, M. A. (1999), The new economics of innovation, spillovers and agglomeration: A review of empirical studies, *Economics of Innovation and New Technology* 8, 5-25.

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

Griliches, Zvi. 1990. Patent statistics as economic indicators: A survey. *Journal of Economic Literature*, Vol. 28, pp. 1661-1707.

Griliches, Z. (1992), The search for R&D spillovers, *Scandinavian Journal of Economics* 94, Supplement: 29-47.

Grillitsch, M., Tödtling, F., Höglinger, C. (2013), Variety in knowledge sourcing, geography and innovation: Evidence from the ICT sector in Austria, *Papers in Regional Science*.

Gunday, G., Ulusoy, G., Kilic, K., Alpkan, L. (2011), Effects of innovation types on firm performance, International Journal of Production Economics, 133(2), 662-676.

Kraft, J., Quatraro, F. and Saviotti, P.P. (2009), The knowledge base evolution in biotechnology: A social network analysis, *Economics of Innovation and New Technology*

Jones, C. (1995), R&D based models of economic growth, *Journal of Political Economy* 103, 759-784.

Jones, F.B. (2009), The burden of knowledge and the death of the renaissance man: Is innovation getting harder? *Review of Economics and Statistics* 76, 283–317.

Lane, D. A. et al. (2009), *Complexity perspectives in innovation and social change*, Springer, Berlin.

Link, A. N. (1980), Firm size and efficient entrepreneurial activity: A reformulation of the Schumpeter hypothesis, *Journal of Political Economy*, 88, 771-782.

Link, A., Siegel, D. (2007*), Innovation, entrepreneurship, and technological change*, Oxford University Press, Oxford.

Lööf, H., Heshmati, A. (2002), Knowledge capital and performance heterogeneity: A firm-level innovation study, International Journal of Production Economics, Vol. 76 (1), 61-85.

March, J.C. (1991), Exploration and exploitation in organizing learning, *Organization Science* 2, 71-87.

Marrocu, E., Paci, R., Pontis, M. (2012), Intangible capital and firms' productivity, *Industrial and Corporate Change,* 21 (2) , 377-402.

Mokyr, J. (1990), *The lever of riches*, New York and Oxford, Oxford University Press.

Nelson, R.R. (1959), The simple economics of basic scientific research, *Journal of Political Economy* 67, 297-306.

Nelson, R.R. (1982), The role of knowledge in R&D efficiency, *Quarterly Journal of Economics* 97, 453-470.

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

Nesta, L., Saviotti, P.P. (2005) Coherence of the knowledge base and the firm's innovative performance: Evidence from the pharmaceutical industry, *Journal of Industrial Economics* 53, 123-142.

Quatraro, F. (2010), Knowledge coherence variety and productivity growth: Manufacturing evidence from Italian regions, *Research Policy* 39, 1289-1302.

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. (1942), *Capitalism, Socialism, and Democracy*, New York: Harper and Row.

Smit, M.J., Abreu, M.A., de Groot, H.L. (2013), Micro-evidence on the determinants of innovation in the Netherlands: The relative importance of absorptive capacity and agglomeration externalities, *Papers in Regional Science*.

Van Zeebroeck, N. (2011), The puzzle of patent value indicators, *Economics of Innovation and New Technology* 20, 33–62.

Van Zeebroeck, N and B. van Pottelsberghe (2011), The vulnerability of patent value determinants, *Economics of Innovation and New Technology* 20, 283–308.

Weitzman, M. L. (1996), Hybridizing growth theory, *American Economic Review* 86, 207-212.

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

Table 1 Sample distribution in macrosectors

Macro-sector	*Freq.*	*Percent*	*Cum.*
HT	283	31.83	31.83
MHT	403	45.33	77.17
MLT	39	4.39	81.55
LT	77	8.66	90.21
KIS	85	9.56	99.78
LKIS	2	0.22	100.00
Total	**889**	**100.00**	

Table 2 Variables measurement method

VARIABLES	
PCOST	Log (R&D / N Patents) for firm i at time t
R&D	Log (R&D / Total assets) for firm i at time t-1
Size	Log of employees number for firm i at time t-1
PStock	Log of patents stock (PIM) for firm i at time t-1
RegPStock	Log of patents stock (PIM) in the same region (NUTS2) of firm i at time t-1
WRegPStock	Log of patents stock (PIM) belonging to EU-24 member states other than that of firm i at time t-1, weighted using a row-normalized inverse distance matrix
RegTV	Log of total variety in the region (NUTS2) of firm i at time t-1
RegRTV	Log of related variety in the region (NUTS2) of firm i at time t-1
RegUTV	Log of unrelated variety in the region (NUTS2) of firm i at time t-1
RegCD	Log of cognitive distance in the region (NUTS2) of firm i at time t-1
RegCOH	Log of knowledge coherence in the region (NUTS2) of firm i at time t-1

Table 3 Descriptive statistics

Variable	*Obs*	*Mean*	*Std.Dev.*	*Min*	*Max*
PCOST	891	9.304	1.704	2.996	15.547
R&D	891	-3.350	1.127	-7.777	0.420
Size	891	8.894	2.324	1.386	13.090
PStock	891	3.005	1.747	-0.650	7.519
RegPStock	891	8.841	1.338	4.853	10.892
WRegPStock	891	7.632	0.265	6.988	8.268
RegTV	891	2.182	0.131	1.653	2.397
RegRTV	891	1.882	0.156	1.232	2.129
RegUTV	891	0.824	0.110	0.269	0.991
RegCD	891	-0.263	0.020	-0.368	-0.223
RegCOH	891	1.786	0.541	0.660	3.846

Table 4 Correlation matrix

	PCOST	*R&D*	*Size*	*PStock*	*RegPCap*	*WRegPStock*	*RegTV*	*RegRTV*	*RegUTV*	*RegCD*	*RegCOH*
PCOST	1.000										
R&D	0.029	1.000									
Size	0.555	-0.385	1.000								
PStock	0.134	-0.005	0.578	1.000							
RegPStock	0.218	-0.070	0.270	0.248	1.000						
WRegPStock	0.027	0.038	-0.055	0.207	-0.045	1.000					
RegTV	0.143	0.008	0.219	0.202	0.821	0.026	1.000				
RegRTV	0.143	-0.011	0.247	0.225	0.814	0.055	0.983	1.000			
RegUTV	0.091	0.084	0.015	0.015	0.520	-0.107	0.667	0.519	1.000		
RegCD	0.174	-0.133	0.013	0.046	-0.052	0.448	-0.172	-0.139	-0.229	1.000	
RegCOH	0.082	0.079	-0.009	0.040	0.292	-0.262	-0.117	-0.138	0.023	-0.107	1.000

Table 5 Results

Fixed effects	(1)	(2)	(3)	(4)	(5)
VARIABLES	*PCost*	*PCost*	*PCost*	*PCost*	*PCost*
R&D	0.143*	0.136*	0.146**	0.133*	0.129*
	(0.0736)	(0.0736)	(0.0736)	(0.0739)	(0.0739)
Size	0.340***	0.346***	0.340***	0.341***	0.348***
	(0.107)	(0.107)	(0.108)	(0.107)	(0.107)
PStock	-0.396***	-0.361***	-0.356***	-0.375***	-0.371***
	(0.0712)	(0.0730)	(0.0732)	(0.0734)	(0.0734)
RegPStock		-0.783**	-0.714*	-0.443	-0.632
		(0.386)	(0.391)	(0.369)	(0.392)
RegTV		2.811**			
		(1.161)			
RegRTV			1.500		1.311
			(0.931)		(0.933)
RegUTV				1.964**	1.843**
				(0.880)	(0.883)
RegCD		10.40**	10.11**	10.63**	10.66**
		(4.135)	(4.142)	(4.143)	(4.141)
RegCOH		0.282	0.208	0.0623	0.214
		(0.243)	(0.247)	(0.221)	(0.246)
Constant	10.44***	13.38***	16.17***	15.21***	14.26***
	(0.987)	(4.027)	(3.756)	(3.800)	(3.856)
Observations	891	891	891	891	891
R-squared	0.354	0.365	0.362	0.365	0.366
Number of id	179	179	179	179	179

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 6 Results

	(1)	(2)	(3)	(4)	(5)
VARIABLES	*PCost*	*PCost*	*PCost*	*PCost*	*PCost*
R&D	0.143*	0.125*	0.132*	0.118	0.117
	(0.0736)	(0.0738)	(0.0738)	(0.0740)	(0.0740)
Size	0.340***	0.357***	0.352***	0.356***	0.360***
	(0.107)	(0.107)	(0.108)	(0.107)	(0.107)
PStock	-0.396***	-0.357***	-0.353***	-0.370***	-0.368***
	(0.0712)	(0.0729)	(0.0731)	(0.0732)	(0.0733)
RegPStock		-0.660*	-0.572	-0.354	-0.497
		(0.391)	(0.397)	(0.370)	(0.398)
WRegPStock		-4.024*	-4.436*	-4.778**	-4.299*
		(2.297)	(2.301)	(2.244)	(2.297)
RegTV		2.359**			
		(1.188)			
RegRTV			1.101		0.929
			(0.952)		(0.954)
RegUTV				1.871**	1.795**
				(0.879)	(0.882)
RegCD		8.767**	8.348**	8.733**	8.941**
		(4.233)	(4.233)	(4.228)	(4.234)
RegCOH		0.239	0.159	0.0581	0.166
		(0.244)	(0.247)	(0.221)	(0.247)
Constant	10.44***	44.24**	49.81***	51.19***	46.92***
	(0.987)	(18.07)	(17.85)	(17.32)	(17.86)
Observations	891	891	891	891	891
R-squared	0.354	0.368	0.366	0.369	0.370
Number of id	179	179	179	179	179

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Appendix A

Sectoral classification and concordance

Macro sectors	Sector	STAN (ISIC 3)	Datastream
High-technology manufactures HT	Pharmaceuticals	2423	4577
	Office, accounting and computing machinery	30	9572, 9574
	Radio, television and communication equipment	32	2737, 3743, 3745,3747,9576, 9578
	Medical, precision and optical instruments	33	4535, 4537, 4573
	Aircraft and spacecraft	353	2713, 2717
Medium-high technology manuf. MHT	Chemicals excluding pharmaceuticals	24ex2423	1353, 1357
	Machinery and equipment, n.e.c.	29	573, 583, 2757
	Electrical machinery and apparatus, nec	31	2733, 3722
	Motor vehicles, trailers and semi-trailers and other transport equipment, aircraft excluded	34, 351, 352-359	2753, 3353, 3355
Medium-low technology manuf. MLT	Coke, refined petroleum products and nuclear fuel	23	533, 537, 577, 587
	Rubber, plastics products and other non-metallic mineral products	25-26	2353, 2723, 3357
	Basic metals and fabricated metal products	27-28	1753, 1755, 1757
Low technology manufactures LT	Food products and beverages	15	3533, 3535, 3537, 3577
	Tobacco products	16	3785
	Textiles, textile products, leather and footwear	17-19	3763, 3765
	Pulp, paper and paper products	21	1737
	Printing and publishing	22	5557
	Manufacturing nec and recycling	36-37	2727, 3724, 3726, 3767
Knowledge intensive sectors KIS	Post and telecommunications	64	5553, 6535, 6575
	Financial intermediation (excl insurance, pension)	65	8355, 8773, 8779
	Insurance and pension funding	66	8532, 8534, 8536, 8538, 8575
	Activities related to financial intermediation	67	8775, 8777, 8985, 8995
	Real estate activities	70	8633, 8637, 8671, 8672, 8673, 8674, 8675, 8676, 8677, 8771
	Renting of m&eq and other business activities	71-74	2791, 2793, 2795, 2799, 5555, 9533, 9535, 9537
	Health and social work	85	4533
	Recreational cultural and sporting activities	92	5752, 5755

Appendix B

Knowledge variety measured by the informational entropy index

Knowledge variety is measured using the information entropy index[4]. 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:

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

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 $j \in S_z$ ($g = 1,\ldots,G$; $z = 1,\ldots, Z$), we can write:

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

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

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:

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

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.[5]

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

[5] 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.

randomly within the firm's patent portfolio with respect to any other technology (Nesta and Saviotti, 2005, 2006; Nesta, 2008; Quatraro, 2010)[6].

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

[6] 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.

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

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

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}:

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

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

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

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[7].

[7] 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

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.

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

---

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.

portfolios. Breschi et al. (2003) adapted this index to measure the proximity between two technologies[8].

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

---

[8] 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.

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:

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

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 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)[9]. 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:

$$d_{lj} = 1 - S_{lj} \qquad (12)$$

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

[9] 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.

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.

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

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:

$$\mathrm{CD}_t = \sum_l \mathrm{WAD}_{lt} \times \frac{P_{lt}}{\sum_l P_{lt}} \tag{14}$$

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.