**THE KNOWLEDGE COST FUNCTION**

**THE KNOWLEDGE COST FUNCTION**

**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), Collegio Carlo Alberto
and
University of Nice Sophia Antipolis, CNRS-GREDEG

ABSTRACT. This paper contributes the economics of knowledge with the analysis of the knowledge cost function and sheds light on the determinants of the large variance in the cost of knowledge 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 knowledge. The empirical section is based upon a panel of companies listed on UK and the main continental Europe financial markets (Germany, France, Italy and the Netherlands) 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 the Marshallian

hypothesis that the size and the composition of the stock of external knowledge play a key role in reducing the actual cost of the generation of new technological knowledge at the firm level. The results shed a new light about the Schumpeterian hypothesis. The evidence suggests, in fact, that the size of the stock of internal knowledge helps reducing the costs of knowledge, while they increase along with the size of R&D expenditures and employment.

JEL CODES: O30

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

## 1. INTRODUCTION

The study of the cost of knowledge 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 and analyze the knowledge cost function.

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 knowledge differ widely across firms. Some firms are able to generate new technological knowledge 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 knowledge differ and their variance becomes a fascinating area of research. The new appreciation of the role of knowledge indivisibility in the generation of new knowledge enables to better grasping the specific effects of knowledge externalities and knowledge cumulability on the costs of knowledge (Antonelli and Colombelli, 2015).

The rest of the paper is structured as it follows. Section 2 recalls the recent advances 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 knowledge. Section 3 provides an empirical investigation based on the econometric estimate of a knowledge cost function based upon a panel of companies listed on UK and the main continental Europe financial markets (Germany, France, Italy and the Netherlands) for the period 1995 – 2006, for which information about patents have been gathered. The conclusions summarize the results and discuss the implications of the analysis.

## 2. KNOWLEDGE AS AN INPUT AND OUTPUT

After a long period of time during which the early economics of knowledge has investigated in depth the determinants and the effects of the characteristics of knowledge as a good - with special attention to its limited appropriability, non-rivalry in use and non-tradeability - the new economics of knowledge pays much attention to the characteristics of the knowledge generation process. In this context, it has grasped the implications of another bundle of characteristics of knowledge as an economic good that received lesser attention: knowledge indivisibility, in terms of both knowledge cumulability and knowledge complementarity. The twin character of knowledge - at the same time an input and an output - and its limited exhaustibility enable to grasp a key aspect of the knowledge generation process. Its generation consists in the recombination of knowledge items that enter the process as inputs (Weitzman, 1996 and 1998).

Because of knowledge complementarity and knowledge cumulability, next to current R&D activities, both the 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 both the introduction of an innovation and the generation of further knowledge (David, 1993). 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 knowledge. Specifically the study of the knowledge cost function helps grasping to what extent the cost of knowledge is affected by the availability of the full range of the inputs and their costs (Antonelli and David, 2015).

As soon as it becomes evident that R&D activities are not the single input into the knowledge generation process (Gunday et al., 2011), the stocks and composition of existing knowledge both internal and external to each firm, as indispensable and strictly complementary inputs, acquire a new relevance (Antonelli and Colombelli, 2015). Knowledge inputs such as the amount of external knowledge that can be accessed by firms to generate new knowledge are distributed unevenly across space. Major institutional and structural characteristics affect the actual amount of external knowledge that each firm can access and use as an input. The costs of these inputs differ in turn because of the variance in the access conditions to the external knowledge available (Cohen and Levinthal, 1989 and 1990) and because of the different characteristics of the local pools of external knowledge (Saviotti, 2007; Quatraro, 2010 and 2012). 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). Knowledge inputs and outputs vary across firms also because firms differ in their specific competence in managing the knowledge generation process (Nelson, 1982).

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; Smit et al., 2013; Marrocu et al., 2012).

The composition of the knowledge pools to which co-localized firms have access also plays an important role in assessing the levels of absorption activities (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; Krafft, Quatraro, Saviotti, 2009; Quatraro 2010).

The interplay between the stock and composition of internal knowledge, which also increases a firm's absorptive capacity, and the actual levels of knowledge externalities helps increasing the amount of knowledge that each firm can generate with a given amount of R&D activities and competence acquired by means of internal learning processes.

The analysis of a knowledge cost function that takes into account the role of the internal stocks of knowledge and of 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 evidence provided controversial evidence on the actual advantages of large firms with respect to smaller ones in the rates of generation of technological knowledge and the eventual introduction of innovations. The results of the empirical studies in different sectors, historic periods, countries and regions have not provided conclusive evidence (Link, 1980; Link and Siegel, 2007). The recent advances of the economics of knowledge enable to focus the Schumpeterian hypotheses on the knowledge generation activity and on the long lasting effects of the limited divisibility and exhaustibility of knowledge. The Schumpeterian hypothesis, in other words, would apply only to the size of the stock of knowledge and not to the sheer size of firms in terms of employment

Following this approach, we put forward the specific hypothesis that the size of firms exerts negative - cost-reducing - effects when it is measured in terms of internal knowledge stock rather than in terms of sheer size. For a given size in terms of employment, firms with a larger stock of internal knowledge have lower unit knowledge costs than firms with smaller internal stocks. The advantage of incumbents, in other words, stems specifically from the effects of knowledge cumulability and non-exhaustibility and is specific to the size of the stock of knowledge.

b) the Marshallian hypothesis. According to the standard application in the economics of knowledge of the Marshallian hypothesis, firms located in large industrial districts with a strong knowledge base have better chances to access knowledge spillovers and feed their own knowledge generation process. In large districts with a rich knowledge base firms have better access to external knowledge and can substitute it to expensive R&D activities. Knowledge externalities are pecuniary rather than pure: relevant search and absorption costs are necessary in order to use knowledge spilling from third parties as an input into the generation of new knowledge. For this reason, not only the size of the knowledge base, but also its nature is of some significance. Indeed, the generation of technological knowledge stems from a variety of competences: knowledge is not a homogenous good and therefore its intrinsic heterogeneous nature cannot be neglected. We thus argue that not only the amount of knowledge available at the local level, also the characteristics of that knowledge have an impact on the costs of knowledge. More precisely, the larger is the size and the coherence of the local knowledge pools and its complementarity with the internal knowledge base and the lower the search costs (Antonelli, 2008). On the other side, the higher is the variety and the dissimilarity in the combination of technologies in the firm region the higher is the cost associated to the firm knowledge output. When the local knowledge which firms may access is distributed across a wide range of technology domains and is featured by technologies which are far away from one another in the technological space, the absorption costs increase. Following these arguments along the lines of the Marshallian tradition of analysis, we put forward two hypotheses: i) we expect a negative correlation between the

size of the regional stock of knowledge and the firm cost of knowledge; ii) knowledge externalities are all the more effective the larger are the levels of coherence and the lower the levels of variety and dissimilarity of the local knowledge base. We thus expect that the unit knowledge costs decrease with the regional levels of knowledge coherence and increase with the regional levels of variety and dissimilarity.

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 knowledge base of each firms in terms of the levels of the stock of knowledge in the generation of new knowledge, and the identification of the key role of knowledge external to each firm but available in regional proximity. Specifically we expect that unit knowledge costs are lower the larger is the size of the stock of internal knowledge , and the larger is the pool of external knowledge that firms can access and its consistence with the stock of internal knowledge.

## 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 following European markets: UK, Germany, France, Italy and the Netherlands. These countries where selected not only for their economic size and importance but also as they represent the main

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

European financial markets. The variety in terms of size, sectors, regions and countries of this set of companies seems to provide a reliable representation of the European business sector. The IPER database has been built by matching information from multiple sources of data. Our main source of 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. The use of patents as the single indicator of the knowledge output is, indeed, a limit of the analysis. A large literature has identified the limits of patents as the exclusive source of information on the actual amount of knowledge generated: not all firms patent their 'inventions'; small firms rely less than large ones on patents to increase the appropriability of their inventions; patents are used more to secure property rights of inventions that apply to product rather than process inventions; firms in fashion industries rarely patent their distinctive knowledge. The awareness of these limits has not prevented the use of patents in the large empirical literature that relies on the legacy of Zvi Griliches (1984 and 1990).

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 31.6% and 45.4% of observations, respectively. Medium low and low technology firms account for 4.5% and 8.9% respectively, while knowledge intensive firms represent 9.4% of observations.

*Table 1 about here*

**3.2 The Econometric Analysis: Methodology and Variables**

The econometric analysis is organized on a baseline equation and a number of complementary specifications that explore in detail the different facets of the basic hypotheses. Our baseline estimating equation is the following (2):

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

In equation (2) all explanatory variables are lagged one year so as to mitigate endogeneity problems. Given the panel nature of our dataset and to control for unobserved firm-specific characteristics, Equation (2) has been estimated using a fixed effects estimator. The Hausman test confirms that fixed effects perform better than the random effects estimator.

In Equation (2) the dependent variable for the firm *i* at time *t* is the cost of knowledge output measured by the logarithm of the ratio between the firm current R&D expenditures and the number of patents delivered. This measure is a good proxy of the actual cost for producing new technological knowledge. Yet, it is worth noting that the cost of external interactions based on knowledge is not directly accounted for. The unit cost of knowledge is explained by two sets of independent variables that are respectively: A) the knowledge base of each firm as defined by the internal expenses in R&D, the size of the internal knowledge stock and the age of the firm and B) the size the local pools of external knowledge and the composition of in terms of variety, complementarity and similarity (dissimilarity). As to the latter set of variables, variety aims to capture the technological differentiation within the knowledge base of a region; coherence measures the extent to which the pieces of knowledge that firms combine to generate new technological knowledge are complementary to one another; finally, similarity (dissimilarity) measures the extent to which the pieces of knowledge used by firms are close (distant) one another in the technology space (Krafft et al. 2014).

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 (*R&Dexp*) to total assets (in logarithms). In order to appreciate the effects of the stocks of internal knowledge of firms, the model includes the variable *PStock* measured as the ratio between the number of patents held by each firm (*CumPatStock*) and total assets (in logarithms). *CumPatStock*

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]:

$$CumPatStock_{it-1} = \dot{h}_{it-1} + (1-\sigma)CumPatStock_{it-2} \quad (3)$$

where $\dot{h}_{it-1}$ is the flow of patent applications (in logarithm) and $\delta$ is the rate of obsolescence. Next to the stock of patents we include the age of firms. This variable aims at grasping the effects of the accumulation of competence by means of learning processes. *Age* is measured by the years since foundation and is expressed in logarithm; it aims at grasping the effects of the accumulation of tacit knowledge and competence, based upon learning processes that do affect the generation of patentable knowledge.

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, in the second basket of variables, we include first a variable aimed at grasping the effects of the size of the knowledge pools into which firms are embedded: *RegPStock*, that is the log of patents stock 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* (i.e. (PIM).

Next, we include other variables proxying for variety, complementarity and similarity. These indicators rest on the recombinant knowledge approach. In order to provide an operational translation of such concepts one needs to identify both a proxy for the bits of knowledge and a proxy for the elements that make their structure. We consider patents as a proxy for knowledge, and then look at

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

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, all these variables are expressed in logarithms:

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). Unrelated variety measures the technological diversification of the knowledge base which is likely to be affected by radically new type of knowledge, while related variety measures the technological diversification of the knowledge base which is likely to be affected by incremental recombination of already existing types of knowledge.

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 inclusion 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 Antonelli, Krafft, Quatraro, 2010). 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 (Saviotti, 2004 and 2007; Colombelli, Krafft, Quatraro, 2013; Quatraro, 2010). We finally include 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. Moreover, as a further robustness

check, we also estimated additional models including firm *Size* among the covariates. The inclusion of a variable that accounts for the sheer size (i.e. the logarithm of employees number for firm i at time t-1) enables to appreciate the estimated parameters as the direct effect of the variables proxying for the internal knowledge base, after taking into account the effects of the size of the firm.

For each variable the measurement method is defined in Table 2, while descriptive statistics are reported in Table 3. The correlation matrix for the extended model can be found in Table 4. As reported in the table, correlations among some independent variables are relatively high. In particular, *RegPStock* is highly correlated with the three knowledge variety measures. Not surprisingly, the different measures of knowledge variety are highly correlated one each other. To further detect multicollinearity among covariates, we also checked the variance-inflation factor (VIF) for each covariate. If *RegPStock* is regressed on all the other covariates, including each of the three knowledge variety measures in different regressions, the VIF assumes values in the range 1.35-1.38, much less than the accepted cut-off value of 10 (Neter et al., 1990). Finally, when *RegPStock* is regressed on all the other covariates, including both *RegRTV* and *RegUTV*, the VIF value equals 1.43. Yet, in our empirical analysis we ran different regression models. First, the three specifications of knowledge variety are included in different regression models. Subsequently, we include the two components of knowledge variety (*RegRTV* and *RegUTV*) in the same model. Moreover, we ran different regression models excluding knowledge stock from the vector of covariates. Finally, also a relatively high correlation is observed between internal R&D and the stock of patents. However, if *R&D* and *PStock* are regressed on all the other covariates the VIF assumes values in the range 2.61-2.84, much less than the cut-off value of 10. Yet, we also ran different regression models excluding *R&D* from the vector of covariates as a robustness check.

*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 Hausman test, comparing the results obtained 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 theinternal activities performed by each firm in terms of R&D expenditure, patents stock and age. 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 about the internal stock of knowledge help to confirm and qualify the Schumpeterian hypothesis. The stock of patents (*PStock*) of each firm exerts in fact a strong negative and significant effect ($p<0.01$ in all estimations) on the costs of knowledge. This is fully consistent with expectations, as the dependent variable is a measure of the unit costs of knowldge, which is likely to decrease as the stock of internal knowledge that firms can mobilize and use to generate new technological knowledge increases, other things being equal. Knowledge cumulability and non exhaustibility exert a strong non-ergodic effect that favors incumbents that can rely upon their internal knowledge in the recombinant knowledge generation process. We interpret these results in light of the Schumpeterian hypothesis. The characteristics of knowledge, combined with the recombinant feature of the knowledge generation process, favor incumbents that can prolong through time the benefits of earlier 'inventions'.

The intensity of R&D expenses is positively and significantly related with the cost of knowledge 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 size of the estimated parameter however seems most important. It is consistently much lower than 1 with a range comprised between 0,190 and 0.211 according to the different specifications. This suggests that the unit costs of knowledge increase, albeit much less than proportionately, with the levels of R&D intensity. The results of Table 10 (see below, more in detail) where the base line model is implemented with the absolute levels of R&D expenses and the explicit integration of the size of firms, provide further evidence that confirm our new interpretation of the Schumpeterian hypothesis: while the size of the stock of internal knowledge helps reducing the unit cost of knowledge, the size of R&D expenditures and the absolute size of firms in terms of employment have a positive impact on the unit cost of knowledge. For a given size of the stock of internal knowledge larger firms have higher knowledge unit costs than smaller firms.

The results of the other variables are most important as they confirm the Marshallian hypotheses that, knowledge costs decline with the size of the local pools of external knowledge. Moreover they confirm that, not only the size of the knowledge base, but also its nature is of some significance.

The results of the variables that account for the size and composition of the regional knowledge base differ whether they concern the size of the external stock, measured using the stock of patents of the firms localized in the region or the knowledge structure in terms of variety (*RegTV*), complementarity (*RegCD*) and similarity (*RegCOH*). If we focus on column 2, results show that the size of the regional knowledge stock (*RegPStock*) exerts a negative and significant effect on the cost of knowledge. This would suggest that 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, results show that *RegTV* is positively and significantly related to the firm cost of knowledge. Let us recall that this index provides a measure of the diversification 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 knowledge output. This might be due to the fact that firms need to put higher efforts in trying and 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 (as shown in columns from 3 to 5). 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 knowledge increases as an effect of the higher volume of resources that the firm needs to commit in order to search and absorb the locally available external knowledge.

The evidence concerning the effect of regional cognitive distance confirms such result. 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. Yet, *WRegPStock* turns out not to be significantly related to the cost of knowledge.

*Table 6 about here*

To further check the robustness of our analysis and to explore the different facets of the hypothesis we run additional models. Table 7 shows results for the equation excluding *RegPStock* from the covariates. Indeed, by looking at Table 4, one may notice that such variable has high correlation with regional total, related and unrelated variety. This may affect the significance level of *RegPStock*. For this reason, we also run the regressions by dropping regional knowledge stock, so as to check the robustness of the results concerning the variety measures. The results actually do not change.

In order to control for the relatively high correlation between *R&D* and *PStock*, Table 8 reports results for the regression models which exclude R&D from the vector of covariates. The results confirm the robustness of our model.

Finally, as a further robustness check and to better test the Schumpeterian hypothesis about the positive effects of the size of firms, we also estimated additional models including firm *Size* among the covariates. Table 9 shows results for an extended version of our baseline model which, next to R&D intensity, includes firm *Size* as an independent variable. Here, again, the estimated parameter of R&D intensity, now taking into account the inclusion of the variable *Size*, is positive albeit well below 1. The results of Table 10 provide the definitive test of our hypothesis and conclude the investigation about the effects of the size of the firm on the cost of knowldge. The results for this alternative specification - which includes firm *Size* and where R&D expenditures and patent stocks

are measured in absolute terms instead that being measured as a ratio over total assets - confirm that even when the size of the firm (in terms of employment) is directly included in the estimation model, the value of the estimated parameter of R&D expenditures in absolute terms is positive albeit well <1. With respect to the results of the estimates of the base line model conducted on the R&D intensity the increase is confirmed although the estimated parameter now varies in a range comprised between 0.288 and 0.312. Also *Size* is found to be positive related to the cost of knowledge.

The results of both the variables that accounts for the size of the firm and its R&D expenditures in absolute terms confirm our interpretation of the Schumpeterian hypothesis. The size of firms helps reducing knowledge unit costs only if it is measured in terms of the stock of internal knowledge. When the size is measured in terms of the amount of R&D expenditures and employment, under the control of specific variables that account for the size of the internal knowledge stock, knowledge unit costs increase. The Schumpeterian hypothesis is not confirmed with respect to the sheer size when knowledge generation costs are considered.

These results lead to articulate the distinction between knowledge generation and knowledge exploitation (March, 1991). Further work might investigate the relationship between the sheer size of firms and knowledge exploitation. Our results suggest that for given levels of the internal stock of knowledge larger firms are less efficient in the generation of new knowledge. It becomes most important to understand whether larger firms might be more efficient in the exploitation of knowledge

## 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 widely. 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 knowledge 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 knowledge. 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 hypotheses according to which knowledge externalities exert positive effects according not only to the density of the local pools of knowledge, but also to their levels of coherence.

The empirical analysis of the costs of knowledge, based upon a panel of companies listed on UK and the main continental Europe financial markets (Germany, France, Italy and the Netherlands) 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 internal knowledge as well as the stock and the composition of external knowledge.

The results confirm that the size and the composition of the stock of internal 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 knowledge. These results are important as they contribute to cast a new perspective about the Schumpeterian hypothesis. The size of firms exerts positive –reducing- effects on knowledge unit costs only when it is measured by the stock of knowledge. The sheer size of firms, in terms of R&D expenditures and employment, under the control of the size of the internal stock of knowledge, exerts positive – increasing - effects on knowledge unit costs. For given levels

of the size of the stock of internal knowledge, larger firms have larger knowledge unit costs than small firms. The sheer size of firms does not help reducing knowledge costs generation. This result pushes to reformulate the Schumpeterian hypothesis introducing the distinction between knowledge generation and exploitation. We have demonstrated that the sheer size does not help increasing the cost-efficiency of firms: additional work should be made to investigate whether the sheer size of firms may favor the exploitation of knowledge.

The results about the role of the size and composition of external knowledge fully confirm the Marshallian hypothesis, stressing the important role of the composition of the local knowledge pools.

These results 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 knowledge generation 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 are not consistent with the competences accumulated in the region are likely to increase the average level of unrelated variety and dissimilarity, and as a consequence, increase the average cost per patent. The implementation of technology policies that focus the local knowledge endowments and try to upgrade it may be more effective than the pursuit of technological goals that are unrelated with the local pools of competence.

From the managerial viewpoint, the results of our analysis confirm the intuition of Edith Penrose about the central role of the stock of knowledge internal to each firm. The results confirm that also the composition of the bundle of technological activities carried out at the local level plays an important role. This has two important implications for decision-making. First, footloose firms should

take into account, in their decisions concerning the location of their R&D laboratories, the local mix of technological competences, so as to select sites with the size and the mix of the local knowledge pools that are more consistent with the technological strategies of the firm. . Second, firms, with a given rooted location, should choose, among different possible technological strategies, those that are more compatible and consistent with the specific composition of the local knowledge pools. The location in areas featured by a bundle of technological competencies consistent with the innovation strategies of the firm is indeed likely to make the search process for new combinations of technologies more effective and hence new knowledge less costly.

**Acknowledgements.** The authors acknowledge the useful comments of the Editor and two anonymous referees as well as the financial support of the European Union D.G. Research with the Grant number 266959 to the research project 'Policy Incentives for the Creation of Knowledge: Methods and Evidence' (PICK-ME), within the context Cooperation Program / Theme 8 / Socio-economic Sciences and Humanities (SSH), and the institutional support of the Collegio Carlo Alberto.

## 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. (2008), Pecuniary knowledge externalities: The convergence of directed technological change and the emergence of innovation systems, *Industrial and Corporate Change*, 17, 1049-1070.

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

Antonelli. C., Krafft, J., Quatraro, F. (2010), Recombinant knowledge and growth: The case of ICTs, *Structural Change and Economic Dynamics* 21, 50-69.

Antonelli, C., David, P.A. (eds.) (2015), *The economics of knowledge and knowledge driven economy*, Routledge, London. forthcoming

Antonelli, C., Colombelli, A. (2015), External and internal knowledge in the knowledge generation function, *Industry and Innovation*, forthcoming,

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., 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. (1984), Patent Statistics as Economic Indicators: A Survey, in *R&D and Productivity: The Econometric Evidence*, University of Chicago Press for the NBER, pages 287-342.

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 94, 25-43.*

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.

Krafft, J., Quatraro, F. and Saviotti, P.P. (2009), The evolution of knowledge base in biotechnology: A social network analysis, *Economics of Innovation and New Technology*, 2011, 20(5), 445-475.

Krafft, J., Quatraro, F. and Saviotti, P.P. (2014), Knowledge characteristics and the dynamics of technological alliances in pharmaceuticals: Empirical evidence from Europe, US and Japan, *Journal of Evolutionary Economics*, 24 (3), 587-622.

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

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.

Quatraro, F. (2012), *The Economics of Structural Change in Knowledge*, Routledge, London.

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*, forthcoming

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* | *Percent* | *Cum.* |
|---|---|---|
| *High-technology manufactures - HT* | 31.6 | 31.6 |
| *Medium-high technology manufactures - MHT* | 45.4 | 77.0 |
| *Medium-low technology manufactures - MLT* | 4.5 | 81.5 |
| *Low technology manufactures - LT* | 8.9 | 90.4 |
| *Knowledge intensive sectors - KIS* | 9.4 | 99.8 |
| *Less knowledge intensive sectors - LKIS* | 0.2 | 100.0 |
| ***Total*** | **100.0** | |

**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 |
| *R&Dexp* | Log (R&D) for firm i at time t-1 |
| *PStock* | Log of (Patents stock (PIM) / Total assets) for firm i at time t-1 |
| *CumPatStock* | Log of (Patents stock (PIM)) for firm i at time t-1 |
| *Age* | Log of years since foundation for firm i at time t-1 |
| *Size* | Log of employees number 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* | 870 | 9.330 | 1.725 | 2.996 | 15.547 |
| *R&D* | 870 | -3.340 | 1.097 | -7.777 | 0.420 |
| *R&Dexp* | 870 | 10.888 | 2.146 | 2.996 | 15.824 |
| *PStock* | 870 | -11.247 | 1.914 | -18.732 | -6.587 |
| *CumPatStock* | 870 | 2.981 | 1.756 | -0.650 | 7.519 |
| *Age* | 870 | 3.425 | 1.185 | 0 | 5.541 |
| *Size* | 854 | 8.920 | 2.296 | 1.386 | 13.090 |
| *RegPStock* | 870 | 8.845 | 1.347 | 4.853 | 10.892 |
| *WRegPStock* | 870 | 7.627 | 0.268 | 6.988 | 8.268 |
| *RegTV* | 870 | 2.182 | 0.131 | 1.653 | 2.397 |
| *RegRTV* | 870 | 1.882 | 0.155 | 1.232 | 2.129 |
| *RegUTV* | 870 | 0.822 | 0.110 | 0.269 | 0.991 |
| *RegCD* | 870 | -0.264 | 0.020 | -0.368 | -0.223 |
| *RegCOH* | 870 | 1.782 | 0.546 | 0.660 | 3.846 |

**Table 4 Correlation matrix**

| | *PCOST* | *R&D* | *PStock* | *Age* | *RegPStock* | *WRegPStock* | *RegTV* | *RegRTV* | *RegUTV* | *RegCD* | *RegCOH* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *PCOST* | 1.000 | | | | | | | | | | |
| *R&D* | 0.041 | 1.000 | | | | | | | | | |
| *PStock* | -0.628 | 0.493 | 1.000 | | | | | | | | |
| *Age* | 0.126 | -0.207 | -0.188 | 1.000 | | | | | | | |
| *RegPStock* | 0.224 | -0.063 | -0.162 | 0.034 | 1.000 | | | | | | |
| *WRegPStock* | 0.025 | 0.033 | 0.215 | -0.095 | -0.045 | 1.000 | | | | | |
| *RegTV* | 0.149 | 0.019 | -0.107 | -0.020 | 0.821 | 0.025 | 1.000 | | | | |
| *RegRTV* | 0.149 | -0.002 | -0.119 | -0.015 | 0.814 | 0.056 | 0.982 | 1.000 | | | |
| *RegUTV* | 0.097 | 0.098 | -0.014 | -0.033 | 0.513 | -0.116 | 0.660 | 0.509 | 1.000 | | |
| *RegCD* | 0.175 | -0.159 | -0.064 | 0.010 | -0.039 | 0.461 | -0.153 | -0.118 | -0.224 | 1.000 | |
| *RegCOH* | 0.084 | 0.052 | -0.023 | -0.029 | 0.288 | -0.267 | -0.125 | -0.146 | 0.019 | -0.120 | 1.000 |

**Table 5 Results - Baseline model**

| Fixed effects | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| VARIABLES | *PCost* | *PCost* | *PCost* | *PCost* | *PCost* |
| *R&D* | 0.211*** | 0.197** | 0.208*** | 0.196** | 0.190** |
| | (0.0763) | (0.0768) | (0.0766) | (0.0769) | (0.0770) |
| *PStock* | -0.489*** | -0.463*** | -0.463*** | -0.471*** | -0.468*** |
| | (0.0605) | (0.0613) | (0.0615) | (0.0614) | (0.0614) |
| *Age* | 0.0145 | 0.0971 | 0.0787 | 0.0985 | 0.110 |
| | (0.160) | (0.163) | (0.163) | (0.163) | (0.163) |
| *RegPStock* | | -0.631* | -0.557 | -0.331 | -0.496 |
| | | (0.374) | (0.379) | (0.356) | (0.379) |
| *RegTV* | | 2.512** | | | |
| | | (1.152) | | | |
| *RegRTV* | | | 1.306 | | 1.151 |
| | | | (0.920) | | (0.921) |
| *RegUTV* | | | | 1.837** | 1.745** |
| | | | | (0.866) | (0.869) |
| *RegCD* | | 8.910** | 8.625** | 9.248** | 9.236** |
| | | (4.156) | (4.162) | (4.166) | (4.164) |
| *RegCOH* | | 0.275 | 0.203 | 0.0814 | 0.216 |
| | | (0.238) | (0.240) | (0.214) | (0.240) |
| Constant | 6.951*** | 8.953** | 11.45*** | 10.40*** | 9.602** |
| | (0.881) | (3.987) | (3.722) | (3.773) | (3.826) |
| Observations | 870 | 870 | 870 | 870 | 870 |
| R-squared | 0.386 | 0.395 | 0.393 | 0.395 | 0.396 |
| Number of id | 171 | 171 | 171 | 171 | 171 |

Standard errors in parentheses

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

**Table 6 Results - Extended model**

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| VARIABLES | *PCost* | *PCost* | *PCost* | *PCost* | *PCost* |
| *R&D* | 0.211*** | 0.192** | 0.201*** | 0.186** | 0.183** |
| | (0.0763) | (0.0772) | (0.0771) | (0.0774) | (0.0775) |
| *PStock* | -0.489*** | -0.459*** | -0.458*** | -0.465*** | -0.464*** |
| | (0.0605) | (0.0616) | (0.0617) | (0.0616) | (0.0616) |
| *Age* | 0.0145 | 0.0904 | 0.0711 | 0.0910 | 0.103 |
| | (0.160) | (0.163) | (0.163) | (0.163) | (0.163) |
| *RegPStock* | | -0.582 | -0.494 | -0.288 | -0.440 |
| | | (0.381) | (0.386) | (0.358) | (0.386) |
| *WRegPStock* | | -1.513 | -1.918 | -2.260 | -1.736 |
| | | (2.259) | (2.262) | (2.203) | (2.259) |
| *RegTV* | | 2.332** | | | |
| | | (1.184) | | | |
| *RegRTV* | | | 1.126 | | 0.990 |
| | | | (0.944) | | (0.945) |
| *RegUTV* | | | | 1.785** | 1.718** |
| | | | | (0.868) | (0.870) |
| *RegCD* | | 8.297* | 7.872* | 8.347* | 8.545** |
| | | (4.257) | (4.256) | (4.257) | (4.261) |
| *RegCOH* | | 0.258 | 0.180 | 0.0782 | 0.194 |
| | | (0.239) | (0.242) | (0.214) | (0.241) |
| Constant | 6.951*** | 20.65 | 26.12 | 27.57 | 22.90 |
| | (0.881) | (17.93) | (17.69) | (17.16) | (17.72) |
| Observations | 870 | 870 | 870 | 870 | 870 |
| R-squared | 0.386 | 0.395 | 0.393 | 0.396 | 0.397 |
| Number of id | 171 | 171 | 171 | 171 | 171 |

Standard errors in parentheses

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

**Table 7 Alternative specifications**

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| VARIABLES | *PCost* | *PCost* | *PCost* | *PCost* | *PCost* |
| *R&D* | 0.211*** | 0.197** | 0.207*** | 0.193** | 0.188** |
| | (0.0763) | (0.0769) | (0.0767) | (0.0768) | (0.0770) |
| *PStock* | -0.489*** | -0.475*** | -0.473*** | -0.478*** | -0.478*** |
| | (0.0605) | (0.0610) | (0.0611) | (0.0610) | (0.0610) |
| *Age* | 0.0145 | 0.0518 | 0.0411 | 0.0777 | 0.0786 |
| | (0.160) | (0.161) | (0.161) | (0.161) | (0.161) |
| *RegTV* | | 1.897* | | | |
| | | (1.094) | | | |
| *RegRTV* | | | 0.841 | | 0.731 |
| | | | (0.865) | | (0.864) |
| *RegUTV* | | | | 1.881** | 1.837** |
| | | | | (0.865) | (0.867) |
| *RegCD* | | 7.752* | 7.679* | 8.633** | 8.431** |
| | | (4.104) | (4.116) | (4.113) | (4.120) |
| *RegCOH* | | 0.182 | 0.115 | 0.0603 | 0.139 |
| | | (0.231) | (0.233) | (0.213) | (0.232) |
| Constant | 6.951*** | 4.470 | 7.192*** | 7.243*** | 5.733** |
| | (0.881) | (2.974) | (2.337) | (1.651) | (2.431) |
| Observations | 870 | 870 | 870 | 870 | 870 |
| R-squared | 0.386 | 0.393 | 0.391 | 0.394 | 0.395 |
| Number of id | 171 | 171 | 171 | 171 | 171 |

Standard errors in parentheses

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

**Table 8 Alternative specifications**

| Fixed effects | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| VARIABLES | *PCost* | *PCost* | *PCost* | *PCost* | *PCost* |
| *PStock* | -0.446*** | -0.424*** | -0.421*** | -0.433*** | -0.431*** |
| | (0.0587) | (0.0596) | (0.0598) | (0.0598) | (0.0598) |
| *Age* | -0.00375 | 0.0818 | 0.0598 | 0.0841 | 0.0978 |
| | (0.161) | (0.163) | (0.163) | (0.163) | (0.164) |
| *RegPStock* | | -0.632* | -0.548 | -0.292 | -0.479 |
| | | (0.376) | (0.381) | (0.357) | (0.380) |
| *RegTV* | | 2.846** | | | |
| | | (1.149) | | | |
| *RegRTV* | | | 1.487 | | 1.291 |
| | | | (0.922) | | (0.923) |
| *RegUTV* | | | | 2.109** | 1.997** |
| | | | | (0.863) | (0.866) |
| *RegCD* | | 8.573** | 8.226** | 8.969** | 8.965** |
| | | (4.171) | (4.179) | (4.181) | (4.178) |
| *RegCOH* | | 0.340 | 0.261 | 0.120 | 0.269 |
| | | (0.237) | (0.240) | (0.215) | (0.240) |
| Constant | 6.774*** | 7.877** | 10.67*** | 9.495** | 8.625** |
| | (0.883) | (3.981) | (3.728) | (3.772) | (3.820) |
| Observations | 870 | 870 | 870 | 870 | 870 |
| R-squared | 0.380 | 0.389 | 0.386 | 0.389 | 0.391 |
| Number of id | 171 | 171 | 171 | 171 | 171 |

**Table 9 Alternative specifications**

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| VARIABLES | *PCost* | *PCost* | *PCost* | *PCost* | *PCost* |
| *R&D* | 0.211*** | 0.195** | 0.207*** | 0.197** | 0.189** |
| | (0.0794) | (0.0798) | (0.0796) | (0.0798) | (0.0800) |
| *PStock* | -0.476*** | -0.444*** | -0.444*** | -0.454*** | -0.449*** |
| | (0.0647) | (0.0660) | (0.0661) | (0.0660) | (0.0660) |
| *Age* | -0.0991 | 0.00278 | -0.0175 | 0.00831 | 0.0197 |
| | (0.182) | (0.185) | (0.185) | (0.186) | (0.186) |
| *Size* | 0.242** | 0.250** | 0.247** | 0.238** | 0.247** |
| | (0.117) | (0.117) | (0.118) | (0.117) | (0.117) |
| *RegPStock* | | -0.743* | -0.669* | -0.423 | -0.614 |
| | | (0.387) | (0.392) | (0.368) | (0.392) |
| *RegTV* | | 2.713** | | | |
| | | (1.167) | | | |
| *RegRTV* | | | 1.469 | | 1.313 |
| | | | (0.932) | | (0.933) |
| *RegUTV* | | | | 1.868** | 1.764** |
| | | | | (0.878) | (0.880) |
| *RegCD* | | 8.274** | 7.949* | 8.661** | 8.626** |
| | | (4.211) | (4.218) | (4.225) | (4.222) |
| *RegCOH* | | 0.278 | 0.207 | 0.0708 | 0.223 |
| | | (0.244) | (0.246) | (0.221) | (0.246) |
| Constant | 5.314*** | 7.648* | 10.29*** | 9.439** | 8.481** |
| | (1.206) | (4.149) | (3.889) | (3.928) | (3.984) |
| Observations | 854 | 854 | 854 | 854 | 854 |
| R-squared | 0.386 | 0.395 | 0.393 | 0.395 | 0.396 |
| Number of id | 171 | 171 | 171 | 171 | 171 |

Standard errors in parentheses

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

**Table 10 Alternative specifications**

| Fixed effects | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| VARIABLES | *PCost* | *PCost* | *PCost* | *PCost* | *PCost* |
| *R&Dexp* | 0.312*** | 0.296*** | 0.309*** | 0.293*** | 0.288*** |
| | (0.0780) | (0.0784) | (0.0781) | (0.0788) | (0.0788) |
| *CumPatStock* | -0.436*** | -0.398*** | -0.395*** | -0.413*** | -0.407*** |
| | (0.0720) | (0.0737) | (0.0739) | (0.0740) | (0.0741) |
| *Age* | -0.0863 | 0.0222 | 0.00277 | 0.0267 | 0.0377 |
| | (0.183) | (0.186) | (0.186) | (0.187) | (0.187) |
| *Size* | 0.315** | 0.312** | 0.302** | 0.307** | 0.316** |
| | (0.123) | (0.124) | (0.124) | (0.124) | (0.124) |
| *RegPStock* | | -0.812** | -0.748* | -0.505 | -0.687* |
| | | (0.391) | (0.396) | (0.372) | (0.397) |
| *RegTV* | | 2.541** | | | |
| | | (1.173) | | | |
| *RegRTV* | | | 1.382 | | 1.233 |
| | | | (0.936) | | (0.938) |
| *RegUTV* | | | | 1.723* | 1.621* |
| | | | | (0.889) | (0.892) |
| *RegCD* | | 9.419** | 9.101** | 9.758** | 9.744** |
| | | (4.211) | (4.216) | (4.225) | (4.223) |
| *RegCOH* | | 0.258 | 0.194 | 0.0614 | 0.205 |
| | | (0.245) | (0.248) | (0.222) | (0.247) |
| Constant | 7.204*** | 10.67*** | 13.05*** | 12.42*** | 11.57*** |
| | (1.193) | (4.001) | (3.748) | (3.776) | (3.829) |
| Observations | 854 | 854 | 854 | 854 | 854 |
| R-squared | 0.380 | 0.389 | 0.387 | 0.389 | 0.390 |
| Number of id | 171 | 171 | 171 | 171 | 171 |

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

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

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

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

The total index can be decomposed into 'within' and 'between' parts whenever the events being investigated can be aggregated into a smaller number of subsets. Within-entropy measures the average degree of disorder or variety within the subsets; between-entropy focuses on the subsets, measuring the variety across them.

It can be easily shown that the decomposition theorem holds also for the multidimensional case (Frenken and Nuvolari, 2004). Let the technologies *i* and *j* belong to the subsets *g* and *z* of the classification scheme respectively. If one allows $l \in S_g$ and $j \in S_z$ ($g = 1,\ldots,G$; $z = 1,\ldots, Z$), we can write:

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

[3] 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:

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

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.[4]

### 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; Nesta, 2008; Quatraro, 2010)[5].

---

[4] 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.

[5] 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

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* $\tau_{lj}$ (Nesta and Saviotti, 2005). 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

---

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) 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.

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 $\mu_{ij}$, 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 $P_{jt}$:

$$WAR_{lt} = \frac{\sum_{j \neq l} \tau_{lj} P_{jt}}{\sum_{j \neq l} P_{jt}} \quad (9)$$

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

$$COH_t = \sum_l WAR_{lt} \times \frac{P_{lt}}{\sum_l P_{lt}} \quad (10)$$

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 $\tau_{lj}$ 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[6].

[6] To make it clear, informational entropy is a diversity measure which allows to accounting for variety, i.e. the number of categories into which system elements are apportioned, and balance, i.e. the distribution of system elements across categories. (Stirling, 2007). In this sense entropy does not say anything about the relationships between technological classes, but provides a measure of the diversity of technological co-occurrences, suggesting whether in a sector most of the observed co-occurrences focus on a specific couple or on the contrary whether the observed co-occurrences relate to a large number couples. In this framework, related and unrelated variety provide a measure of the extent to which observed variety applies to couples of technologies that belong to the same macro domain or to different macro-domains. One would expect established technologies to be characterized by relatively low variety of co-occurrences, insofar as the recombination focus on a relatively small numbers of technological classes that have proved to be particularly fertile. On a different ground,

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 portfolios. Breschi et al. (2003) adapted this index to measure the proximity between two technologies[7].

---

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.

[7] 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

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:

---

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.

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

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)[8]. 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 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.

$$WAD_{lt} = \frac{\sum_{j \neq l} d_{lj} P_{jt}}{\sum_{j \neq l} P_{jt}} \qquad (13)$$

[8] 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.

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:

$$CD_t = \sum_l WAD_{lt} \times \frac{P_{lt}}{\sum_l P_{lt}} \tag{14}$$

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.