Knowledge cumulability and complementarity in the knowledge generation function

Cristiano Antonelli[ab] and Alessandra Colombelli [ac*]

[a]*BRICK (Bureau of Research in Innovation Complexity and Knowledge), Collegio Carlo Alberto;* [b]*Dipartimento di Economia, Università di Torino;* [c]*DIGEP, Politecnico di Torino*

* Corresponding author. Email: alessandra.colombelli@polito.it

Abstract

This paper explores the role of external knowledge and internal stocks of knowledge in the generation of new technological knowledge. It relies on the notion of recombination and brings together three concepts: the appreciation of current expenses in R&D activities; the analysis of the role of the stock of knowledge composition; the identification of the role of external knowledge available in the regional proximity. The empirical section is based upon a panel of companies listed on the main European financial markets for the period 1995–2006. The econometric analysis considers patents as a measure of the knowledge output and, on the right 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 knowledge.

JEL codes: O30

Key words: knowledge generation function; knowledge stock; external knowledge; path dependence.

1. Introduction

This paper contributes the analysis of the process by means of which new technological knowledge is being generated. We define the object of our analysis as the knowledge generation function, as distinct from the knowledge production function. The notion of knowledge production function applies to standard production functions where knowledge is considered explicitly as an input. On the output side next to alternative measures of output (sales, value added), a variety of performance indicators such as labor or total factor productivity have been considered. The knowledge generation function, instead, considers the activities that make possible the generation of new knowledge.

The knowledge production function is one the pillars of the applied economics of innovation (Griliches 1979, 1990, 1992; Romer, 1990; Link and Siegel, 2007). It has been widely applied in variety of contexts including firms, regions, industries and countries. In the knowledge production function approach innovations as measured by such indicators as R&D expenses, patents and innovation counts either enter directly, next to capital and labor, a production function or indirectly with a two-step procedure in a model that estimates its effects on the general efficiency of the same production function. The evidence confirms that innovation is a major input into the production of other goods and is strongly associated to the rates of increase of total factor productivity and in general to economic performances performances (Cohen, 2010; Hall, Mairesse and Mohnen, 2010).

The notion of knowledge generation function, instead, studies specifically the direct relations between the inputs that make possible generation of knowledge as an output. The knowledge generation function was born in the context of the efforts to appreciate the contribution of knowledge to the generation of other goods and became a full-fledged knowledge generation function, as we suggest to identify it, only when both the input and the output are refereed to direct measures of knowledge, rather than other economic variables.

From this viewpoint the knowledge generation function, as distinct from the knowledge production function, can be considered a direct consequence of the intuition of Zvi Griliches about the role of knowlege spillovers. It becomes progressively clear the need to include a broader array of inputs on both the left and the right hand side of the specification so as to broaden and deepen the analysis of the relationship between knowledge inputs and knowledge outputs that was originally framed simplistically in terms respectively of R&D activities and patents (Levin et al. 1987).

The notion of knowledge generation function had been first introduced in the growth literature by Phelps (1966) who called it "technology function" and "effective research function". Gomulka (1970) refereed to the direct relationship between knowledge inputs and knowledge outputs as the "production function of innovations". Jones (1995) uses the term "the idea production function". The knowledge generation function had received little empirical support at the aggregate level and for quite a long time no evidence at the disaggregate level supported the approach.

The contribution by Crépon, Duguet and Mairesse (1998) marks a major progress in the empirical analysis of this approach from many viewpoints. To start with, it provides the first econometric analysis of the knowledge generation function at the firm level. Second, it provides a broad econometric model into which the relationship between knowledge inputs and knowledge outputs is estimated within a four equation model able to assess in parallel the effects of R&D expenditures not only on innovation counts but also on labor productivity and total factor productivity.

Since then a tiny but growing empirical literature has explored the characteristics of the knowledge generation function assessing the role of different measures and proxies for both knowledge inputs and knowledge outputs. Parisi, Schiantarelli and Sembenelli (1996) have explored the likelihood that R&D expenditures affect the introduction of product innovations, as distinct from process innovations, using the European Community Innovation Survey. The first empirical estimates of the knowledge generation function are quite simplistic and use R&D expenditures as the key input. Furman, Porter and Stern (2002), show that the differences in the levels of innovation activity across countries is explained by the differences in the level of inputs such as R&D manpower and spending invested in the generation of innovations. The empirical framework of the knowledge generation functions has been applied not only at the country level, but also at the regional level with interesting results

(Fritsch (2002); O´ hUallachain and Leslie, 2007) with an elementary frame where the R&D expenditure is the single input and the patents granted to a given region are the output. Nesta and Saviotti (2005) make a major innovation in this methodology implementing the empirical analysis of the determinants of the generation of new knowledge at the firm level with the inclusion of the characteristics of the stock of knowledge as a central input into the generation of new knowledge analyzing the relationship between the coherence of the knowledge base and the innovative performance of U.S. pharmaceutical firms measured by means of citation-weighted patent count.

In the while, the economics of knowledge has made important progress in the inquiry about characteristics of knowledge as an economic good and its generation process. Nelson (1982) made an important contribution to implement the knowledge generation function approach stressing the need to consider explicitly knowledge as the output of a dedicated activity and take into account of the variety of inputs, complementary to R&D expenditures that make possible the generation of new knowledge.

David (1993) marks the final step in the process stressing that knowledge is at the same time an input in the generation of new knowledge and in the production of all the other goods, but it is also and primarily the output of a dedicated activity. Existing knowledge plays a key role as an input into the generation of new knowledge. The new understanding of the knowledge generation as a recombinant process where existing knowledge items enter as inputs shed new light on the role of knowledge indivisibility (Weitzman, 1996). After much attention paid to the notion of knowledge non-appropriability, other key characteristics such as non-exhaustibility, cumulability and complementarity that stem from its intrinsic indivisibility need to be fully appreciated.

This paper contributes the empirical analysis of the knowledge generation function. It applies and implements the notion of recombination to grasp the specific characteristics of the knowledge generation process. The appreciation of the generation of new technological knowledge as a recombination process that consists in the reorganization and reconfiguration of the relations among existing knowledge items enables to better appreciate the effects of knowledge indivisibility, as articulated in knowledge complementarity and knowledge cumulability in the generation of new knowledge so as to grasp the path dependent character of the knowledge generation process and its strong systemic nature. 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 and by the characteristics of the system into which each firm is embedded. The current levels of research and development expenditures of each agent do play a role but only in a context that is shaped by the past of each firm and by its localization (Antonelli, 2011).

The rest of the paper is structured as it follows. The next section 2 provides a synthetic account of the notion of knowledge recombination and explores its implications in the identification of the role of knowledge cumulability and knowledge complementarity. A novel specification of the knowledge generation function is the result of the analysis. Section 3 presents the data and the econometric procedure elaborated to test the model elaborated in section 2. The conclusions summarize the results of the analysis and explore the implications.

2. The recombinant generation of new technological knowledge

The generation of knowledge is characterized by specific attributes: knowledge is at the same time the output of a specific activity and an essential input into the generation of new knowledge. The arrovian analysis of knowledge as an economic good plays a key role in this context (Arrow, 1962 and 1969; Nelson, 1959).

The recombinant knowledge approach has paved the way to elaborate a new frame of analysis able to accommodate the central role of existing knowledge as an input into the generation of new knowledge. As Weitzman recalls, "when research is applied, new ideas arise out of existing ideas in some kind of cumulative interactive process that intuitively has a different feel from prospecting for petroleum" (Weitzman, 1996:209). This insight has led to the so-called recombinant approach: new ideas are generated by means of the recombination of existing ideas under the constraint of diminishing returns to scale in performing the R&D activities that are necessary to apply new ideas to economic activities. The generation of new knowledge stems from the search and identification of elements of knowledge that had been already generated for other purposes and yet reveals characteristics and properties that had not been previously considered. The search for existing knowledge items both internal to the firm, stored in the stock of competence and knowledge accumulated in the past and external to each firm, their screening and re-assessment leads to their subsequent active inclusion and integration as elements of the recombination process (Weitzman 1996 and 1998; Fleming and Sorenson 2001).

The recombinant approach enables to appreciate the central role of two important inputs into the generation of new technological knowledge such as the knowledge base of each firm as qualified by stock of knowledge that each firm possess, and the knowledge that are external and yet highly complementary to the research activities undertaken by each firm.

The crucial role of the stock of knowledge as an input into the generation of new knowledge enables to grasp its non-ergodic character. Knowledge accumulated in the past exerts a strong influence in the future generation of new knowledge. Past knowledge, however, is not the single, deterministic factor: current efforts in terms of R&D activities and access to high quality pool of external knowledge, in fact, may alter the amount of knowledge that each firm is able to generate at each point in time. The path dependent character of the knowledge generation process, as distinct from the deterministic past dependence, can be fully appreciated when the role of the stock of knowledge is put in the broader context of a multi-inputs knowledge generation function (Antonelli, Crespi, Scellato, 2012)

The stock of knowledge qualifies and identifies the knowledge base of each firm. Its composition plays an important role. 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. The quality of the stock of knowledge in other words matters as well as its sheer size. The shorter is the distance between different types of knowledge, the higher the probability that they can be combined together. Furthermore, this representation provides the basis to move to empirical analyses by constructing an image of the knowledge base as a network in which the nodes are constituted by units of knowledge at a given level of aggregation. Such empirical investigations can be successfully conducted by exploiting information contained in patent documents (Saviotti, 2004 and 2007; Krafft, Quatraro, Saviotti, 2009; Quatraro 2010; Colombelli, Krafft, Quatraro, 2011).

A large literature has explored the role of technological spillovers as a major input into the generation of new technological knowledge. In this approach external knowledge plays an important and yet supplementary role in the generation of new technological knowledge. Moreover its recipients are mainly viewed as the passive beneficiary of knowledge leaking from other firms (Feldman, 1999). The notion of pecuniary knowledge externalities has been implemented to better appreciate the active role that perspective users of external knowledge need to undertake in order to acquire external knowledge, now regarded as a necessary and complementary input that cannot be fully substituted by other knowledge inputs (Antonelli, 2009, 2011). A large body of empirical evidence confirms that external knowledge is an essential input into the generation of new knowledge. At each point in time, no agent possesses all the knowledge inputs. Yet agents need to access the variety of knowledge items that are available in the system and as such are being possessed and used by the other firms that belong to the system. The search of external knowledge is necessary and its acquisition is the result of an intentional activity. The access to external knowledge is possible only if dedicated resources are invested to search, screen, interact to understand, access and eventually recombine the external units of knowledge with the internal ones. Pecuniary knowledge externalities are found in specific and fertile regions of the knowledge and regional space where inputs of useful knowledge can be accessed at low cost, below equilibrium levels (Colombelli and von Tunzelmann 2011).

The integration of these issues into the recombinant approach to the generation of technological knowledge enables to lay down our basic argument. Technological knowledge is at the same time an input and an output and it is the result of an intentional economic action. Technological knowledge is localized in the accumulated competence of firms and in the knowledge space into which firms are rooted. New knowledge can be generated, by means of the recombination of existing knowledge items, when, where and if:

A) an intentional action directed to its generation is undertaken. New technological knowledge does not fall like manna from heaven. The generation of knowledge requires an active and explicit action. Research and development activities are necessary to activate the recombination process. This is not their single function. Current research and development activities are also necessary to access, learn and absorb the stock of existing knowledge within the firm and the flows of external knowledge generated by third parties. Current research and development activities are necessary also to track the records, entertain, retrieve, and eventually use again the knowledge that has been produced in the past and is stored in the stock of knowledge and competence that each firm has accumulated. Current research and development activities moreover are necessary to access, learn and absorb the knowledge that is external to each firm: this contrasts the passive attitude that would characterize the perspective users of technological spillovers. In sum current research and development expenditures are necessary not only to perform the recombination but also to feed it with the stock of knowledge internal to each firm and with the access to external knowledge being generated by other firms.

B) the knowledge base of each firm is identified and the role of previous knowledge is fully appreciated. The knowledge base of a firm is identified by size and the compositeness of the stock of knowledge that each firm has been able to generate in the past. The knowledge base exerts its positive effects in the long run and enters directly as an input the knowledge generation function. According to our interpretative framework, the knowledge generation is a non-ergodic process where history matters as it helps building higher and higher levels of competence and innovative capability. The process is path-dependent as opposed to past-dependent, however, as it is influenced by the events that take place along the process such as changes in R&D strategies and in the external context.

C) the effects of pecuniary knowledge externalities stemming from the amount of knowledge being generated in the proximity of each firm are appreciated. Pecuniary knowledge externalities take place when access conditions to the local pools of knowledge make possible the actual use of external knowledge in the generation of new technological knowledge at costs that below equilibrium levels. Each agent has access only to local knowledge interactions and externalities, i.e. no agent knows what every other agent in the system at large knows. Because of the localized character of knowledge externalities and interaction, proximity in a multidimensional space, in terms of distance among agents and their density, matters. Agents are localized within networks of transactions and interactions that are specific subsets of the broader array of knowledge externalities, interactions and transactions that take place in the system. The wider and easier is the access to the local pools of knowledge and the larger is the amount of technological knowledge that each firm is able to generate, for given levels and composition of the internal stock of knowledge and the amount of current efforts in R&D activities.

The following knowledge generation function (1) provides the general frame of our approach. Here the dependent variable for the firm i at time t is the knowledge output and it is explained by three independent variables that are respectively the internal expenses in R&D, the knowledge base of each firm as defined by the size and the composition of the knowledge stock, and the external knowledge:

$$TK_{it} = (R\&D_{it-1}, KNOWLEDGE\ BASE_{it-1}, EXTERNAL\ KNOWLEDGE_{it-1})^{\alpha} \qquad (1)$$

Equation (1) provides a suitable specification of the knowledge generation function, that accommodates, next to the role of R&D activities, the appreciation of the role played by the knowledge base of each firms in terms of the levels and the composition 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 and enables to grasp their non-ergodic effects. The flow of knowledge generated at each point in time by each firm, in fact, adds on to its stock of knowledge, increasing, with due depreciation rates, its capability to generate additional flows of knowledge in the future.

This specification marks an important progress with respect to previous specifications of the knowledge generation function as it takes into account both the size of the knowledge base – with the inclusion of the stock of patents – and the characteristics of the knowledge base – with the inclusion of the measure of its cognitive distance.

3. Dataset

The dataset is an unbalanced panel of publicly traded firms in UK, Germany, France, Italy and the Netherlands. Our main source of data is Thomson Datastream. To obtain additional relevant variables, we include in the dataset information collected from AMADEUS by Bureau Van Dijk. The period of observation for all the countries examined is 1995 to 2006. 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. We pooled the dataset by adding industry level information from the OECD STAN database.

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 33% and 46% of observations, respectively. Medium low and low technology firms account for 5% and 8% respectively, while knowledge intensive firms represent some 8% of observations.

4. Methodology and variables

Our analysis is based on the knowledge generation function equation. The general knowledge generation function (1) enables to specify an econometric model where all terms enter in logarithmic form to account for the multiplicative relationship that has been retained. The knowledge output is measured in terms of patents and it is explained by three sets of independent time varying variables that identify the specific relevant characteristics of the knowledge base and external knowledge respectively, as it follows in equation (2):

$$TK_{it} = \beta_1 + \beta_2 R\&D_{it-1} + \beta_3 KSTOCK_{it-1} + \beta_4 CD_{it-1} + \beta_5 EXTK_{it-1} + \rho_i + \sum \psi_t + \varepsilon_{it} \quad (2)$$

Here on the right hand side, the first set considers just R&D, i.e. the current research efforts and activities funded by each firm at time *t*. The original specification of the knowledge generation function implemented by Griliches (1990) and Crepon, Duguet and Mairesse (1998) would not consider other variables on the right hand side. We do include instead other variables to articulate the different facets of the knowledge base, taking into account the contribution of both the stock and the composition of knowledge, and of the knowledge that is external to each firm, taking into account the flows of knowledge that, because of proximity, are likely to occur between firms co-localized in the same region.

In order to appreciate the effects of the stocks of internal knowledge of firms, we have included the variable *KSTOCK* measured in terms of the number of patents held by each firm. This is computed by applying the permanent inventory method to patent applications. We calculate it as the cumulated stock of patent applications using a rate of obsolescence of 15% per annum:

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

where $\dot{h}_{it}$ is the flow of patent applications and δ is the rate of obsolescence. To better screen the role of the stock of knowledge as distinct from the sheer size effect, the variable *R&D* is specified both in absolute terms and in relative ones as the ratio of R&D expenditures to total assets.

The variables *CD* accounts for the composition of the stock of knowledge internal to each firm and qualify its knowledge base in terms of cognitive distance (CD). Following the recombinant knowledge approach, this index expresses knowledge dissimilarities amongst different types of knowledge. (See next section 4.1 for the specification and measure of this variable).

The third set of variables accounts for the contribution of the knowledge that is external to each firm at time t but made accessible by proximity. Here *EXTK* measures the patenting activities of firms localized within the same region and as such can produce positive Jacobs pecuniary knowledge externalities mainly based upon the mobility of skilled personnel and more generally of the array of knowledge interactions that make cheaper the access and use of external knowledge.

We finally included both sectoral and time dummies in order to control for industrial and time effects. 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.

To better appreciate the effects of the knowledge base we have also specified the following equations:

$$TK_{it} = \beta_1 + \beta_2 R\&D_{it-1} + \beta_3 KSTOCK_{it-1} + \beta_4 KSTOCK_{it-1} * CD_{it-1} + \beta_5 EXTK_{it-1} + \\ + \rho_i + \sum \psi_t + \varepsilon_{it} \quad (4)$$

$$TK_{it} = \beta_1 + \beta_2 R\&D_{it-1} + \beta_3 KSTOCK_{it-1} * CD_{it-1} + \beta_4 EXTK_{it-1} + \rho_i + \sum \psi_t + \varepsilon_{it} \quad (5)$$

Where the knowledge base enters directly the equation as a single variable specified as the multiplicative interaction of the stock of patents and their cognitive distance.

4.1 The cognitive distance index

To describe the composition of the stock of knowledge of each firm we measure the dissimilarities among different types of knowledge (Nooteboom, 2000). 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, Lissoni and Malerba (2003) adapted this index to measure the proximity or relatedness between two technologies.

We define 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 $O_l = \sum_{lk} P_{lk}$. Similarly, the total number of patents assigned to technology j is

$O_j = \sum_k P_{jk}$. We can, thus, indicate the number of patents that are classified in both technological fields l and j as: $V_{lj} = \sum_k P_{lk} P_{jk}$. By applying this count of joint occurrences to all possible pairs of classification codes, we obtain a square symmetrical matrix of co-occurrences whose generic cell V_{lj} reports the number of patent documents classified in both technological fields l and j.

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

$$S_{lj} = \frac{\sum_{m=1}^{n} V_{lm} V_{jm}}{\sqrt{\sum_{m=1}^{n} V_{lm}^2} \sqrt{\sum_{m=1}^{n} V_{jm}^2}} \quad (6)$$

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, Lissoni and Malerba, 2003). 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 (Breschi, Lissoni and Malerba, 2003; Engelsman and van Raan, 1994; Jaffe, 1986).

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

$$d_{lj} = 1 - S_{lj} \tag{7}$$

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

$$\mathrm{WAD}_{lt} = \frac{\sum_{j \neq l} d_{lj} P_{jt}}{\sum_{j \neq l} P_{jt}} \tag{8}$$

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:

$$\mathrm{CD}_t = \sum_l \mathrm{WAD}_{lt} \times \frac{P_{lt}}{\sum_l P_{lt}} \tag{9}$$

5. Results

The dependent variable in equations 2, 4 and 5 is a count variable and so they can be estimated by means of either a Poisson or a negative binomial model. Since our dependent variable is over-dispersed, as showed in Table 3 by the fact that its variance is far larger than the mean, the negative binomial estimator seems to be more appropriate. Moreover, since firms included in our sample belong to all industrial sectors, they show a different patenting behaviour. The histogram of the dependent variable suggests that the number of zeros might be excessive (Figure 1). For this reason, equations 2, 4 and 5 can be estimated by a zero-inflated regression model. Zero-inflated models attempt to account for excess zeros. Two kinds of zeros are thought to exist in the data, "true zeros" and "excess zeros". Zero-inflated models estimate two equations simultaneously, one for the count model and one for the excess zeros. Zero-inflated regression models might be a good option if there are more zeros than would be expected by either a Poisson or negative binomial model. We thus finally use a zero-inflated negative binomial regression estimator.

Table 5 reports the results of the zero-inflated negative binomial regression estimations for the baseline model (column 1 and 4) and the extended ones that include the interaction variable (column 2-3 and 5-6). The Vuong test, comparing the zero-inflated models with the negative binomial regression model, indicates that the zero-inflated negative binomial is a better fit than the standard negative binomial in most of our regressions[1].

[1] In one out of six regressions the z-value is not significant. We also estimated this model using a standard negative binomial and results do not show relevant changes.

The results of the econometric exercise confirm our hypotheses[2]. The results of the variables that account for the knowledge base differ whether they concern the stock of patents or their cognitive distance. The stock of patents (*KSTOCK*) of each firm exerts a strong positive and significant effect ($p<0.01$ in all estimations) on the output in terms of patents of the knowledge generation function. Its elasticity, according to estimated parameter, significantly close to unity, confirms that the stock of knowledge exerts its influence on the current capability of the firm to generate new knowledge.

The composition of the stock of knowledge, as measured by its cognitive distance (*CD*), exerts a negative and significant effect in the baseline model (column 1). This means that when firms focus their search activity in a region of the knowledge space that is close to their accumulated competences, they are more likely to get to the successful generation of new knowledge. However, cognitive distance does not exert a significant effect in the model where R&D is measured in absolute terms (column 4). This

[2] Since the explanatory variables are expressed in log while the dependent variable is measured in absolute levels, the coefficients can be interpreted as semi-elasticities.

result may be influenced by the heterogeneity of firms considered. In some circumstances, such as in the automobile and engineering industries, the technological variety actually helps the generation of new knowledge at the intersection of electronics, informatics and mechanics. In others, such as in chemistry, pharmaceutics and biotechnology, the cognitive distance of the knowledge stock limits the actual efficiency of the knowledge generation process.

The negative and significant results of the knowledge base, specified as the multiplicative relation between the stock of patents and their cognitive distance, when it enters the econometric model together with the stock of patents possessed by each firm (columns 2 and 5) or without it (columns 3 and 6), however, may suggest that stocks of patents with high levels of similarity exert positive effects on the generation of new knowledge.

The results of the R&D variable shed a new light on the effects of the size of firms, with respect to the specific relationship between the size of the input, in terms of the amount of current efforts, and the size of the output. Indeed R&D activities contribute with a significant effect the generation of new knowledge, but the value of the elasticity, according to the estimated parameter, comprised between around 0.1 when *R&D* is measured in absolute terms, and 0.08-0.1 when *R&D* is measured in terms of intensity tells that the flow of new patents increase at a rate that much less than proportional to the increase of the current efforts of firms in terms of R&D activities. Firms benefit of the advantages of the internal stock of patents but can increase their stock of patents, with current R&D activities, only to a limited extent.

Finally the positive and significant role of external knowledge is confirmed by the results where *EXTK* is always significant and positive. Again, as already noticed with respect to R&D expenditures, the elasticity of external knowledge – as an input – to the knowledge output is quite small. The localization in a good knowledge pool helps increasing the knowledge output, but only to a limited extent.

The limited magnitude of current R&D efforts and external knowledge appears all the more relevant when compared to the strong effects of the stock of patents. The comparative assessment of the different inputs of the knowledge generation process stresses the relevance of the historic accumulation of competence and internal knowledge in the capability to generate new knowledge. The persistent character of the knowledge generation process is fully confirmed (Antonelli, Crespi, Scellato, 2012).

6. Conclusions

The knowledge generation function is an important tool that enables to open the black box of the knowledge generation process. The arrovian analysis of the characteristics of knowledge as an economic good has important implications to understand the specificities of the knowledge generation process. The key attributes of indivisibility and limited appropriability identified, especially when the twin character of knowledge that is at the same time the output of the generation process and an input into the following generation process, are fully grasped and their implications receive proper appreciation.

The analysis of knowledge indivisibility has made it possible to identify the two key dimensions of knowledge cumulability and knowledge complementarity. The efficient and effective generation of new knowledge at time *t* is possible only standing upon the shoulders of the technological knowledge that has been generated until that time. Finally, since no agent can command all the technological knowledge that has been generated at each point in time and the limited appropriability of technological knowledge engenders the possibility to access and use the knowledge that has been generated by third parties, knowledge external to each firm plays a key role in the generation of new technological knowledge.

The early specifications of the knowledge generation function did not pay attention to the characteristics of knowledge as an economic good and could not take advantage of their important implications to understand the dynamic process that make the generation of new technological knowledge possible. The results of our empirical analysis confirm that the output of knowledge is generated not only by means of R&D expenditures: other key inputs enter the knowledge generation function. Our results suggest the strong and positive role of the stock of knowledge possessed by each firm, and the key role of proximity. External knowledge plays a role as an input when it is embedded in firms that are co-localized in regional proximity. Proximity matters as well with respect to the components of the knowledge base: the higher the similarity of the knowledge base and the larger the output in terms of new knowledge.

These results have important implications on two counts. First they confirm that the generation of technological knowledge is a historic process characterized by clear elements of path dependent persistence. The specific endowment cumulated in the past through time and represented by the stock of technological knowledge and its composition exerts long-term effects on the actual capability of firms to generate new technological knowledge. These effects can be altered and affected by the current effects of the efforts of research and development expenditures that each firm is able to fund and perform at each point in time and by the quality of the pools of external knowledge to which each firm can access. The clear importance of both R&D activities and external knowledge confirms that the process is non-ergodic, but not past- dependent. The initial conditions affect the process but its direction and speed can be significantly changed by events that occur along the process.

The second important implication of our analysis concerns the role of external knowledge. The appreciation of the key role played by external knowledge enables to fully understand the systemic conditions that shape and make the generation of new technological knowledge possible. The generation of new technological knowledge is influenced by individual characteristics of each firm such as the past accumulation of knowledge and the current commitment of resources to research, but requires the access to complementary external knowledge that is commanded by other firms co-localized. The generation of technological knowledge cannot be regarded as the result of a stand-alone activity but rather as the product of a collective process. This leads to the identification of innovation as an emergent property of a system. The characteristics of the system are crucial to assess the amount and the characteristics of the knowledge being generated.

These results are important for their implications for both public policy and corporate strategy. The appreciation of the strong complementarity between knowledge externalities and current inputs call attention on the fact that the actual amount of knowledge that a firm is able to generate with given levels of R&D activities is heavily influenced by the knowledge stock of co-localized firms. The appreciation of the role of the stock of knowledge and its characteristics stresses the need to manage carefully its evolution and sheds a new light on the implications of the non-ergodic features of the generation of technological knowledge: past knowledge generating activities do affect the efficiency of current activities.

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. A preliminary version of the paper has been presented to the conference “The governance of a complex world” held in Nice, 1-3 November 2012.

Notes on contributors

Cristiano Antonelli holds the chair of Political Economy of the University of Torino. He is editor of the journal Economics of Innovation and New Technology and the Director of BRICK (Bureau of Research on Innovation, Complexity and Knowledge) of the Collegio Carlo Alberto. He is the author of more than 160 journal articles and books.

Alessandra Colombelli is assistant professor at the Politecnico of Torino, where she teaches Strategy, Technology and Marketing. She is research associate at BRICK (Bureau of Research on Innovation, Complexity and Knowledge) of the Collegio Carlo Alberto and the CNRS-GREDEG, University of Nice Sophia Antipolis. She is the author of several scientific articles in the areas of economics of innovation and management of technology.

7. References

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

Antonelli, C., Crespi, F., Scellato, G. 2012. Internal and external factors in innovation persistence, *Economics of Innovation and New Technology*, 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.

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

Cohen, W.M. (2010), Fifty years of empirical studies of innovative activity and performance, in Hall, B.H. Rosenberg, N. (eds.), *Handbook of the economics of innovation*, Elsevier, Amsterdam, pp. 131- 213.

Colombelli, A., Krafft, J. and Quatraro, F. 2011. High Growth Firms and Technological Knowledge: Do gazelles follow exploration or exploitation strategies?, LEI&BRICK Working Paper n. 14/2011, University of Torino.

Colombelli A., von Tunzelmann G.N. 2011. The persistence of innovation and path dependence, in Antonelli C. (ed.) *Handbook on the economic complexity of technological change*, Edward Elgar, Cheltenham, pp. 105-120.

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.

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

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.

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

Fritsch, M. 2002, Measuring the quality of regional innovation systems: A knowledge production function approach, *International Regional Science Review* 25, 86-101.

Furman, J. L., Porter, M.E., Stern, S. 2002, The determinants of national innovative capacity. *Research Policy* 31, 899–933.

Gomulka, S. 1970. Extensions of 'The golden rule of research' of Phelps, *Review of Economic Studies* 37, 73-93 .

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

Griliches, Z. 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.

Griliches, Z., Mairesse, J. 1984. Productivity and R&D at the firm level. In Z. Griliches (Ed.), *R&D, patents, and productivity*. Chicago: Chicago University Press.

Hall, B.H., Mairesse, J. and Mohnen, P. 2010, Measuring the returns to R&D, in Hall, B.H. Rosenberg, N. (eds.), *Handbook of the economics of innovation*, Elsevier, Amsterdam, pp. 1034-1074

Jaffe, A. 1986. Technological opportunity and spillovers of R&D: Evidence from firms' patents, profits, and market value, *American Economic Review*, 76(5), 984-1001.

Jaffe, A. 1989. Real Effects of Academic Research, *American Economic Review*, 79(5), 957-70.

Krafft, J., Quatraro, F. and Saviotti, P.P. 2009. The evolution of the knowledge base in biotechnology: Social network analysis of biotechnology, *Economics of Innovation and New Technology* 20, 445-475.

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

Levin, R.C., Klevorick, A. K., Nelson, R.R. and Winter, S.G. 1987. Appropriating the returns from industrial research and development, *Brookings Papers on Economic Activity*, *Economic Studies Program, The Brookings Institution*, 18, 783-832.

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

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., 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.

Nooteboom, B. 2000. *Learning and innovation in organizations and economies*, Oxford: Oxford University Press.

O´ hUallachain, B., Leslie, T.F. 2007. Rethinking the regional knowledge production function, *Journal of Economic Geography* 7, 737–752.

Parisi, M. L., Schiantarelli, F., Sembenelli, A. 2006. Productivity innovation and R&D: Micro evidence for Italy, *European Economic Review*, vol. 50, pp. 2037-2061.

Phelps, E.S. 1966, Models of technological progress and the golden rule of research, *Review of Economic Studies* 33, 133-145.

Romer, P.M. 1990, Endogenous technological change, *Journal of Political Economy* 98, S71-102

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.

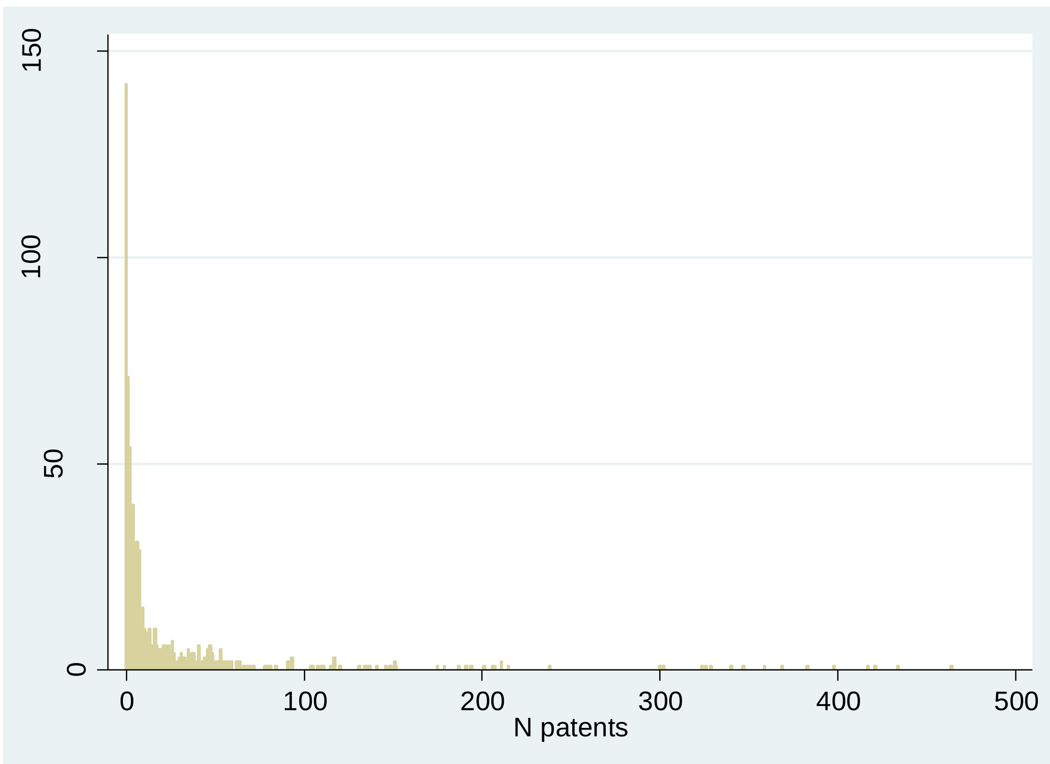
Saviotti, P. P. 2007. On the dynamics of generation and utilisation of knowledge: The local character of knowledge, *Structural Change and Economic Dynamics* 18, 387-408.

Thoma, G., Torrisi, S., Gambardella, A., Guellec, D., Hall, B.H. and Haroff, D., 2010. Harmonizing and combining large datasets – an application to firm-level patent and accounting data, NBER Working Paper 15851.

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.

Figure 1. Histogram of the dependent variable (Number of patents)



Macro-sector	Freq.	Percent	Cum.
HT	239	33.29	33.29
KIS	58	8.08	41.36
LT	58	8.08	49.44
MHT	328	45.68	95.13
MLT	35	4.87	100.00
Total	718	100.00	

Table 1. Sample distribution in macrosectors

	Variable	Measurement method
Technological knowledge	*TK*	No. patents for firm i at time t
R&D expenses	*R&DAbs*	Log R&D for firm i at time t-1
R&D intensity	*R&D*	Log (R&D / Total assets) for firm i at time t-1
External knowledge	*EXTK*	Log of No. patents in the same region (NUTS2) of firm i at time t-1
Knowledge stock	*KSTOCK*	Log of patents stocks (PIM) for firm i at time t-1
Cognitive distance	*CD*	Log of cognitive distance of firm i at time t-1

Table 2. Variables measurement method

Table 3. Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
TK	718	25.989	61.818	0.000	464.000
R&D	718	-3.214	1.070	-7.592	-0.251
R&DAbs	718	11.301	2.065	2.996	15.824
EXTK	718	6.616	1.391	2.079	8.705
KSTOCK	718	3.564	1.582	-0.163	7.519
CD	718	-4.271	1.199	-8.817	-1.179

Table 4. Correlation matrix

	TK	*R&D*	*R&DAbs*	*EXTK*	*KSTOCK*	*CD*
TK	1.00					
R&D	0.02	1.00				
R&DAbs	0.44	-0.05	1.00			
EXTK	0.10	0.01	0.24	1.00		
KSTOCK	0.57	-0.18	0.71	0.27	1.00	
CD	-0.31	0.19	-0.36	-0.18	-0.41	1.00

Table 5. Results

	(1)	(2)	(3)	(4)	(5)	(6)
			Estimator ZINB			
VARIABLES			*Dep.var. TK*			
R&D	0.0838**	0.0823**	0.112**			
	(0.0385)	(0.0383)	(0.0503)			
R&DAbs				0.102***	0.101***	0.137***
				(0.0195)	(0.0196)	(0.0236)
EXTK	0.0643***	0.0651***	0.0694**	0.0568**	0.0572**	0.0569**
	(0.0236)	(0.0234)	(0.0281)	(0.0231)	(0.0229)	(0.0268)
KSTOCK	0.932***	0.850***		0.846***	0.811***	
	(0.0227)	(0.0562)		(0.0272)	(0.0551)	
CD	-0.0554*			-0.0321		
	(0.0316)			(0.0305)		
*CD * KSTOCK*		-0.0166*	-0.146***		-0.00719	-0.124***
		(0.00881)	(0.00475)		(0.00877)	(0.00515)
Constant	-4.521***	-4.257***	-2.665***	-5.502***	-5.379***	-4.503***
	(0.318)	(0.325)	(0.352)	(0.309)	(0.329)	(0.368)
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Sectoral dummies	Yes	Yes	Yes	Yes	Yes	Yes
Inflate Macro-sector	0.116	0.116	-0.183	0.190	0.188	0.0549
	(0.186)	(0.186)	(0.234)	(0.184)	(0.184)	(0.217)
Constant	-3.823***	-3.820***	-2.835***	-4.097***	-4.088***	-3.643***
	(1.007)	(1.006)	(0.964)	(1.021)	(1.018)	(1.141)
lnalpha	-1.084***	-1.084***	-0.559***	-1.159***	-1.158***	-0.662***
	(0.0907)	(0.0906)	(0.0810)	(0.0911)	(0.0910)	(0.0829)
Vuong test z	1.86**	1.85**	1.19	2.10**	2.10**	1.34*
Observations	718	718	718	718	718	718

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1