Knowledge, Jacobs externalities and innovation performance in European regions

Published version:

DOI:10.1080/00343404.2016.1217405

(Article begins on next page)

Knowledge Composition, Jacobs Externalities and Innovation Performance in European Regions

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

Francesco Crespi
Dipartimento di Economia, Università Roma Tre
BRICK (Bureau of Research in Complexity, Knowledge, Innovation), Collegio Carlo Alberto.

Christian Mongeau
Dipartimento di Economia, Università Roma Tre
BRICK (Bureau of Research in Complexity, Knowledge, Innovation), Collegio Carlo Alberto.

Giuseppe Scellato
Politecnico di Torino
BRICK (Bureau of Research in Complexity, Knowledge, Innovation), Collegio Carlo Alberto.

Abstract

This paper analyses the role of the composition of the regional stock of knowledge in explaining innovation performance. The paper provides three main contributions. First, it investigates the relevance of Jacobs knowledge externalities in characterizing the technological capabilities at the regional level. Second, it applies the Hidalgo-Hausmann (HH) methodology to analyze knowledge composition by looking at patent data of 214 regions, located in 27 state members of the European Union (EU) during the years 1994-2008. Third, it econometrically assesses the role of knowledge base composition in a knowledge generation function. The results of the empirical analysis confirm that the characterization of regional knowledge base through the HH indicators provides interesting information to understanding its composition and to qualify it as a provider of the Jacobs knowledge externalities that account for the dynamics of regional innovative performance.

1. Introduction

Economic literature has repeatedly tried to qualify the composition of an economic system and its consequences in terms of performance and dynamics. At the macro-level, according to Grossman and Helpman (1991) and Aghion and Howitt (1998), the variety of inputs and outputs that a system is able to produce positively affects its total factor productivity. Following this new emphasis of the economics of growth on the role of system composition a stream of investigations at the regional, industrial and technological levels have flourished (see for instance Baptista and Swann, 1998; Frenken et al., 2007; Boschma and Iammarino, 2009; Galliano et al., 2014). Moreover, Glaeser, Kallal, Scheinkman and Shleifer (1992) showed that the variety of activities is one of the main determinants of the growth of cities.

This paper implements this line of research and focuses on the regional level of analysis, which in the light of the literature on Regional Systems of Innovations appears to be particularly suited for our purposes (Cooke et al., 1997; Asheim and Coenen, 2005). In particular, we focus our analysis on the stock of knowledge embedded in each regional system as an endogenous endowment and explore the effects not only of its size, but also of its composition on the generation of technological knowledge. Our analysis provides three main contributions. First, we discuss the notion of Jacobs knowledge externalities in the framework of the recombinant generation of technological knowledge at the regional level (Boschma, 2005; Usai, 2011; Antonelli, Patrucco and Quatraro, 2011). Second, we apply the approach used for qualifying the knowledge composition of economic systems developed by Hidalgo and Hausmann (2008 and 2009; HH hereafter) to a set of European regions, using patent data to map technological capabilities. In particular, the analyzed European regions have been mapped against three measures of the knowledge base: *variety*, *ubiquity* and *complexity*. The first measure captures the degree of regional diversification in terms of technological competences, while *ubiquity* accounts for the level of diffusion, and hence commonality, of the technologies in which a region is specialized. Finally, the *complexity* measure provides a synthetic indicator that maps different regions according to their ability in specializing in sophisticated, and thus more rare, technologies emerging where large number of high skilled individuals and specific technological competences are available.

The last contribution of the present study consists in providing empirical evidence on the relevance of the proposed indicators in explaining knowledge generation processes at the regional level.

The paper is organized as follows. Section 2 provides the theoretical background for the analysis. Section 3 explains the data and methodology applied to measure the characteristics of the knowledge bases at the regional level. Section 4 provides a descriptive analysis of the distribution and dynamics of the relevant indicators across different regions. Section 5 presents an econometric test on the potential relevance of the identified indicators in the generation of new knowledge at the regional level. The last section summarizes the main results and discusses the implications of the present analysis for future research.

2. Jacobs externalities and the composition of the knowledge base

Jacobs externalities are well known to regional economists. Named after Jane Jacobs, they identify the positive effects exerted on economic growth by the composition of the -heterogeneous- activities that cluster in a geographic space. Jacobs externalities differ from Marshallian externalities. Although they both account for the positive effects of the regional clustering of activities, Marshallian externalities focus on the positive effects exerted by the sheer size and density of the cluster, while Jacobs externalities focus on the composition of the activities that cluster.

Jacobs knowledge externalities are found when the composition of the knowledge base of an economic system exhibits high levels of organized complexity and, as such, is able to provide cheaper access to and use conditions of the stock of quasi-public knowledge that is necessary for the recombinant generation of technological knowledge. The generation of technological knowledge, in fact, consists in the recombination of heterogeneous and complementary knowledge items. When, because of the variety, rarity and relatedness of the components of the local knowledge base, they are available at low absorption costs, Jacobs knowledge externalities exert their positive –pecuniary- effects reducing the costs of knowledge both as input and output.

A range of approaches and methodologies have been implemented to explore the characteristics of the composition of an economic system, and to identify their causes and effects. Let us consider them individually.

Knowledge Variety

Variety has been the first aspect of the composition of the system that has attracted the attention, and it has been defined and qualified in terms of products, technologies and industries.

By focusing the analysis on products, a distinction has been made between related variety i.e. variety of the products and activities within industrial sectors and technological classes, and unrelated variety i.e. variety across sectors (Frenken and Boschma, 2007; Frenken et al., 2007; Saviotti and Frenken, 2008). According to the rich evidence collected at the regional level (Boschma and Iammarino, 2009; Neffke et al., 2011; Boschma, et al., 2013), variety, qualified in terms of product relatedness, plays a positive role in favoring the emergence of new industries.

The analysis of variety in terms of technologies was pioneered by Archibugi and Pianta (1992) and Pianta and Meliciani (1996) that investigated technological variety by means of an international analysis of the distribution of patents across technological classes. They pointed out that the variety of the knowledge base and the advance of countries are strongly associated with a non-linear relationship. Similar results have been found by Imbs and Wacziarg (2003), who suggest that the relationship between the sectoral concentration of economic systems and output follows a U-shaped pattern. Excess knowledge differentiation is not likely to exert positive effects on the innovative capabilities of local firms (Noteboom, 2000; Boschma et al., 2014; Rigby, 2015). Hence, Jacobs externalities have a limit. Excess diversification and dispersion at the regional level limit the working of Edgeworth complementarities, and reduce the benefits of the generation and exploitation of technological knowledge. However, next to the levels of related variety other aspects may play a role in shaping Jacobs knowledge externalities. The actual composition of a bundle, in fact, differs not only with respect to the levels of variety of the components, but also in terms of their *interrelatedness*, *coherence* and relative *rarity*.

Knowledge interrelatedness

The input-output matrices of industrial systems differ in terms of the set of complementary industries that use each other outputs to produce the final goods is complete. The differences in matrices completeness and hence in knowledge interrelatedness do play a role in assessing the levels of productivity (Leoncini, et al., (1996); Gehringer, (2011a and b, 2012).

In parallel, network analysis contributes the analysis of knowledge interrelatedness with the investigation of the architectural composition of the knowledge base of aggregate economic systems, individual industries and firms. Some systems are characterized by higher levels of interrelatedness than others: patents and inventors cluster around nodes that have a central position in centered networks. Flat knowledge bases can be found at the other extreme when the distribution of links is symmetric and dispersed (Duguet and MacGarvie, 2005; Han and Park, 2006; Mina, Ramlogan, Tampubolon, Metcalfe, 2007; Sternitzke, Bartkowski, Schramm, 2008).

With respect to these analyses, an important step forward was made by Nesta and Saviotti (2005 and 2006) who study the composition of a system with an *ex-post* analysis of the actual distribution of the different bundles. This approach allowed to investigate knowledge composition in terms of its *coherence* through the analysis of co-occurrence: the frequency with which two knowledge items are found together.

Knowledge coherence

Nesta and Saviotti (2005 and 2006) provide a major contribution to the analysis of the composition of the system with the notion of technological coherence. Technological coherence measures the average technological proximity of patents proxied by the number of co-occurrences in different technological classes of the patents belonging to a bundle. Technological coherence explores a new aspect of the composition of a system that is quite different from the related variety as it is able to appreciate the levels of complementarity of the components. According to a large number of empirical studies that have implemented this approach, the performances of a technological system are better for high levels of technological coherence (Antonelli, Krafft, Quatraro, 2010; Quatraro, 2010; Colombelli, Krafft and Quatraro, 2013).

Knowledge rarity

The qualification of the composition of a system in terms of the *ex-post* revealed rarity of its constituent elements enables a further major step. Bundles that are made of items that happen to be frequently associated are more homogeneous than bundles that include rare items. The frequency of their occurrences is in fact low. The notion of rarity begins to play a role. Not only the number of the activities and the related knowledge bases matter, but also, and above all, so does their relative scarcity. Hence, the composition of the bundle of activities that are likely to engender high-level Jacobs externalities can be qualified by their relative scarcity (Hidalgo and Hausmann, 2008 and 2009).

The appreciation of the levels of ubiquity of the activities that are part of a system provides the opportunity to integrate the different aspects of the composition that have been identified in the literature in a single framework. A good composition able to yield strong Jacobs externalities can be identified from the levels of inclusion of rare activities. In this context, the knowledge composition measures that can be derived by applying the HH framework seem most appropriate to grasp the pecuniary effects of the organized complexity of a system in terms of Jacobs knowledge externalities.

Table 1 summarizes this approach. When the variety of the bundle of activities is high, but it is able to include only ubiquitous products and competencies, the levels of Jacobs externalities are low. When the variety of the bundle of activities is high and the bundle includes rare items, there is strong likelihood that the levels of Jacobs externalities are high. When the variety of the bundle is low and includes only ubiquitous items, the levels of Jacobs externalities are deemed to stay low. When, finally, the variety of the bundle is small, but includes rare items, the levels of Jacobs externalities are likely to exhibit high levels of variance because, on the one hand, the limited variety reduces the working of the recombinant generation of technological knowledge, but can yield rare combinations that characterize the generation of radical new knowledge that yield high profits and total factor productivity increases with positive effects on output growth.

[TABLE 1 AROUND HERE]

Building on this analytical framework we want to assess, firstly, if the knowledge composition indicators proposed by HH represent a meaningful and informative measure for classifying and understanding regional performances. Secondly, building on the path breaking explorations of Pakes and Griliches (1984) and Jaffe (1986), we aim to test the role of knowledge composition in the generation of technological knowledge. With respect to this latter aspect, so far the literature has analytically assessed the causes of the positive effects of the (Marshallian) knowledge externalities that stem from the *size* of the existing knowledge base (Adams, 1990 and 2006; Weitzman, 1996, 1998; Arthur, 2009; Antonelli and Colombelli, 2015). In the present analysis we claim that, next to Marshall knowledge externalities, Jacobs knowledge externalities that capture the composition and the organized complexity of the stock of quasi-public knowledge, do play a role in shaping the knowledge generation at the regional level and that this effect can be grasped by analyzing the enhancing role played by qualified variety in knowledge composition (Antonelli and David, 2016).

3. Measuring knowledge base composition

3.1 Methodology

In order to analyze the composition of the knowledge structure and technological capabilities of European regions, the indicators proposed by Hidalgo and Hausmann (2008 and 2009) have been adopted with the aim of capturing the degree of diversification, the average ubiquity and the complexity of innovation activities. The HH method has been applied to regional patent portfolios data, based on the location of the inventors. The use of international patent classification allows developing a fine-grained analysis of the technological composition of innovation activities carried out in a specific area.

The peculiar aspect of the HH method is that it makes no use of ex-ante technological distances to qualify the degree of diversification of the knowledge base. Such distances are commonly computed on large samples of patents and are by definition generated irrespectively of the geographic distribution of the patents. On the contrary, the HH method derives implicitly such patterns from the empirical observation of the distribution of patenting activities across regions. In this perspective the method appears to be superior in capturing the localised dynamics of knowledge generation: it is the degree of geographical clustering of different activities that reveals the properties of knowledge. Hence, the outcome of the HH methods encompasses jointly diversity in the technological and geographical space. The method can be regarded as a bottom-up approach in which the observed evolution of the specializations of innovation systems provides hints on the actual complementarities among technological domain, rather than a top-down approach in which the structure of interdependency (or relatedness) between technologies are pre-defined on pure technological evidence.

A preliminary step, before the "Method of Reflections" developed by Hidalgo and Hausmann (2008 and 2009) is applied, consists in identifying whether region $r = 1, \ldots, R$ is specialized in technology $t = 1, \ldots, T$.[1]

Even though various indicators of specialization are available (Archibugi and Pianta, 1992), we followed HH and computed a *Revealed Technological Advantage* index (RTA), which is defined as:

$$RTA_{rt} = \frac{P_{rt}/(\sum_{r=1}^{R} P_{rt})}{\sum_{t=1}^{T}(P_{rt})/(\sum_{t=1}^{T}\sum_{r=1}^{R}(P_{rt}))} \tag{1}$$

where P_{rt} is the number of patents of region r in patent class t, R is the number of regions, and T is the number of technological fields. RTA is the share of patents in technology t of region r normalized by the share across all technologies. Thus, it follows that $RTA_{rt} = 1$ represents a threshold of specialization: when $RTA_{rt} > 1$, region r is considered to be specialized in technology t.

The next step is to define a "specialization matrix" $\mathbb{M}$ as a binary-valued matrix, in which the rows represent regions and the columns represent technologies, whose generic element (r, t) is equal to 1 if region r is specialized in technology t:[2]

[1] A notational *caveat*: while Hidalgo and Hausmann use p to indicate "products", we prefer to use t (for technological fields that are based on patent class).

[2] In the definition of the generic element of $\mathbb{M}$, HH give a value of 1 also when $RTA_{rt} = 1$. We preferred to assign a unitary value only if RTA_{rt} is strictly greater than 1, as $RTA_{rt} = 1$. This different approach is simply a clarification, as in our database no unitary *RTA* index is present.

$$\mathbb{M}(r,t)=\begin{cases}1 & \text{if} \quad RTA_{rt}>1\\ 0 & \text{if} \quad RTA_{rt}\leq 1\end{cases} \tag{2}$$

The Method of Reflections uses the specialization matrix $\mathbb{M}$ in an iterative way in order to obtain various indicators that represent different aspects of the specialization structure of the regions and technologies. Considering I_X as a unitary vector of dimension $X = R, T$, the initial conditions can be computed as

$$\vec{K}_{r,0}=\mathbb{M}I_R \tag{3}$$

$$\vec{K}_{t,0}=\mathbb{M}'I_T \tag{4}$$

which represent the summation of $\mathbb{M}$ by rows and columns, respectively. In particular, the former vector, $\vec{K}_{r,0}$, has a generic i-th element given by $K_{r,0}(i)=\sum_{t=1}^{T}\mathbb{M}(i,t)$ which indicates *variety*, i.e., the number of technologies in which region i is specialized; $\vec{K}_{t,0}$, on the other hand, has as generic j-th element $K_{t,0}(j)=\sum_{r}\mathbb{M}(r,j)$ which represents the *ubiquity* of a specific technology, i.e., the number of regions specialized in technology j.

Given $\mathbb{M}$ and the starting values, $\vec{K}_{r,0}$ and $\vec{K}_{t,0}$, the iterative formulae of the Method of Reflections are:

$$\vec{K}_{r,n}=\text{diag}\left(1/\vec{K}_{r,0}\right)\mathbb{M}\vec{K}_{t,n-1} \tag{5}$$

$$\vec{K}_{t,n}=\text{diag}(1/\vec{K}_{t,0})\mathbb{M}'\vec{K}_{r,n-1} \tag{6}$$

where n indicates the iteration number ($n > 0$), $\text{diag}(1/\vec{K}_{x,0})$ is a diagonal matrix of dimension $X = R,T,$ whose elements are given by the $(1/K_{x,0}(1), 1/K_{x,0}(2), \ldots, 1/K_{x,0}(X))$ vector for $x = r, t$.

The various iterations of $\vec{K}_{r,i}$ and $\vec{K}_{t,i}$ for $n = 1,\ldots, N$ provide interesting properties of the specialization patterns within and across regions. If $n = 1$, the generic i-th element of $\vec{K}_{r,1}$ is

$$K_{r,1}(i) = \frac{1}{K_{r,0}(i)} \mathbb{M}(i,t)'\vec{K}_{t,0} \tag{7}$$

which represents the *average ubiquity* of the technologies in which region i is specialized, while the generic j-th element of $\vec{K}_{t,1}$ is

$$K_{t,1}(j) = \frac{1}{K_{t,0}(j)} \mathbb{M}(r,j)\vec{K}_{r,0}{}' \tag{8}$$

which represents the *average variety* of the regions that are specialized in technology j. For $n = 2$, the i-th element of $\vec{K}_{r,2}$ measures the *average variety of regions with a technological structure similar to region i*, while the j-th element of $\vec{K}_{t,2}$denotes the *average ubiquity of the technologies of the regions that are specialized in technology j*. For a discussion on the meaning of $\vec{K}_{r,n}$ and $\vec{K}_{t,n}$ for $n \geq 3$, see Hidalgo and Hausmann (2009; Suplementary Material).

By focusing only on $\vec{K}_{r,n}$, it is easy to see that its n-th iteration, for $n \geq 2$, can be written as

$$\vec{K}_{r,n} = \widehat{\mathbb{M}}\vec{K}_{t,n-2} \tag{9}$$

where

$$\widehat{\mathbb{M}} = \mathrm{diag}\big(1/\vec{K}_{r,0}\big)\mathbb{M}\mathrm{diag}(1/\vec{K}_{t,0})\mathbb{M}' \tag{10}$$

The previous matrix plays an important role in the construction of the *Technology Complexity Index* (TCI) that synthetically measures the complexity of regional technological capabilities. Such a measure is computed by normalizing $\vec{K}$, which is the eigenvector associated to the second highest eigenvalue of matrix $\widehat{\mathbb{M}}$:[3]

$$\overrightarrow{TCI} = \frac{\vec{K} - \text{average}(\vec{K})}{\text{stdev}(\vec{K})} \tag{11}$$

The said eigenvector captures the largest amount of variance in the system and is the adaptation to the case of technologies of the Economic Complexity Index (ECI) proposed by Hidalgo and Hausmann (2008 and 2009).

3.2 Dataset description

The dataset consists of a large collection of patent applications at the European Patent Office that have been regionalized according to the corresponding addresses of the inventors[4]. We dropped regions belonging to the first quintile, in terms of patent applications, during the 1994-2008 period[5]. The final database covers the inventive activity of 214 geographical regions, located in 26 members state of the European Union (EU) over the 1994-2008 time interval. Patent applications have been collected and elaborated through the REGPAT database. REGPAT is built upon the Worldwide Statistical Patent Database (PATSTAT) published by the European Patent Office (EPO) twice a year and the OECD Patent Database, which relies on the EPO Bibliographic Database and Abstracts (EBD). Each record provides detailed information concerning the application and priority dates of the corresponding patent, the assignee and inventor's names, their addresses, the country identifier, the region

[3] The first eigenvector is a vector of ones, and has thus been excluded as it is not informative.

[4] The use of patent data in innovation studies is affected by some well-known limitations. In particular, we are aware that the use of patent data might lead to an underestimation of the contribution to the overall innovation output in a specific region from smaller companies relying on informal incremental innovation models, firms operating in the service sectors as well as firms relying on alternative approaches to protect the value of innovation.

[5] We chose to exclude the regions in the first quintile given that the number of patents in these regions is extremely limited and this could bias the analysis.

codes, the application and publication numbers. Patents have been assigned to regions on the basis of the addresses of the inventors[6].

Each patent application reports one or more International Patent Classification (IPC) codes. Such codes are the basis for our analysis of the technological composition of the regional portfolios. We used 4-digit IPC codes obtaining a mapping of all patents into 623 technology areas. Hence, for each region we have computed the number of patent application in each of the 623 technology areas by year. Patents with more than one IPC codes have been double counted.

4. Descriptive analysis

In this section we illustrate and discuss the descriptive evidence on the distribution of the indicators of knowledge composition among the analyzed regions.

In the graphs we make use of the following indicators: a) Kr0 – a measure of *variety* across technologies of the patent portfolio of a region; b) Kr1 – a measure of the average *ubiquity* of the technologies in which a region is specialized, which provides information about how common are the technological competencies of one region compared to other regions; c) TCI – the measure of technological *complexity* derived through the method of reflection as illustrated in the previous section 3.1.

We have grouped regions in clusters of high-, medium-, and low-performance regions. The performance metric is based on the 2013 Regional Competitiveness Index (RCI). The RCI is calculated for all the EU28 NUTS2 regions and summarizes the information coming from a set of 73 variables that pertain to the following "pillars": institutions, macroeconomic stability, infrastructures, health, quality of primary and secondary education, higher education, training and lifelong learning, labour market efficiency, market size, technological readiness, business sophistication, innovation (see Annoni and Dijkstra, 2013). In particular we have classified as high- and low-performance regions those belonging respectively to the

[6] Patents can be allocated to geographical zones according to the corresponding addresses of either their applicants or inventors. The latter approach provides a better representation of the actual location of R&D labs and is not influenced by patent application policies adopted by firms.

upper and lower 30th percentiles of the distribution of the overall RCI, calculated on all 272 regions, while the remaining regions are classified as medium-performance regions.[7]

Fig. 1 shows the distribution of Kr0 against Kr1 by considering their averages in the 1994-1998, 1999-2003, and 2004-2008 sub periods. The two indicators are negatively correlated, suggesting that the regions with highly diversified technological capabilities are those that are able to generate new knowledge in sectors where only a limited number of regions are capable of producing innovations. In this respect, the graphs show the existence of polarization phenomena across different groups of regions, with the high-performance regions being concentrated in the lower-right side of the graph and low-performance regions positioned mostly in the upper left-hand side of the figure. As clearly visible in Fig. 1, the negative relationship between Kr0 and Kr1 is confirmed and even reinforced over time.

[FIGURE 1 AROUND HERE]

In Fig. 2 and Fig. 3 we further explore the dynamics of the indicators during the observed years. In particular, Fig. 2 reports the variation of Kr0 from 1994-1998 to 2004-2008 with respect to its initial value. The graph shows a process of convergence among regions in terms of technological diversification. Less diversified regions in the first period of observation show the highest growth of the index. Interestingly, the variance in the growth of Kr0 is quite relevant in the group of high performing regions, suggesting the presence of a strong heterogeneity in the diversification patterns of the "strongest" group of regions.

[FIGURE 2 AROUND HERE]

[7] The RCI data can be accessed at: http://ec.europa.eu/eurostat/statistics-explained/images/3/3c/Focus_on_competitiveness_RYB2014.xlsx

Fig. 3 instead shows an overall tendency toward an increase in the indicator measuring the average ubiquity of the technological activities of regions. This is an obvious outcome considering that, as evidenced by the dynamics of the diversification indicators, the observed regions have generally experienced entry into new technological domains so that the command of a fraction of previously “rare” technological capabilities becomes more common across regions.

The graph shows that the growth in average ubiquity of the medium-performance regions is highly dispersed, showing an increase in the relative commonality of their patenting activities. The distribution of the high-performance regions is instead more concentrated and evenly distributed with respect to the average dynamics of the population. Finally, data seems to reveal that within the group of low-performing regions, the increase of average ubiquity is greater for those regions that at the beginning of the observation period were characterized by lower ubiquity levels and vice versa.

[FIGURE 3 AROUND HERE]

Finally, Fig. 4 shows the distribution of European regions with respect to the measure of technological complexity associated with regional technological capabilities. On average, the highest levels of the complexity index are found in regions belonging to the high-performance group, with a strong degree of persistence in time. On the contrary, the variance in time in the level of technological complexity of the low-performance regions is greater, with the majority of regions belonging to this group experiencing an increase in this indicator with respect to the initial level.

[FIGURE 4 AROUND HERE]

The overall descriptive evidence on the correlations between the knowledge base characteristics and a broadly defined indicator of economic performance and competitiveness has revealed some interesting issues.

First, the proposed indicators appear to provide relevant and consistent information about the effects of the evolution of the knowledge bases on the competitiveness of regions. On average, high performing regions show a specialization into more technologies that have a lower degree of ubiquity, indicating their ability to command more rare technological capabilities. Still, during the observed years there has been considerable entry by a subset of lagging regions into new technological domains. However, we observe that on average high performing regions have been able to consistently keep more complex specialization patterns. We suggest that such evidence exactly points to the relevance of the composition of the portfolio of technologies at the region level to support further innovation development. The increasing technological interdependencies in new products and the hybridization of related scientific domains engenders a persistent advantage for those regions that have brought together diverse knowledge pools rather than pursuing competitive advantage through the specialization into fewer technological domains. In the following section we further explore this issue by applying to the data an econometric approach that controls for region level inputs and spatial dynamics.

5. Econometric analysis

In this section, we test whether the composition of the knowledge base of a specific region displays a significant correlation with its capability to generate further innovations. In line with our theoretical discussion on the role of Jacobs externalities in the knowledge generation function, the approach is meant to test whether – after accounting for the time invariant specific effects and R&D inputs – it is still possible to identify a positive marginal effect of the composition of the knowledge base on innovation performance. In the baseline model specification, the regional innovation output is regressed against a set of region-level controls and two indicators of the properties of the regional knowledge base, according to the following specification:

$$PAT_{i,t} = \beta_1 GERD_{i,t-1} + \beta_2 Pop_{i,t-1} + \beta_3 HiTechEmpl_{i,t-1} + \beta_4 Kr0/Kr1_{i,t-1} + \beta_5 TCI_{i,t-1} + \varepsilon_{i,t}$$

where the regional innovation output is measured by means of the log of the number of new patent applications by inventors located in region *i* in year *t* (PAT). All the controls are one-year lagged and include the logs of region population (POP), gross regional R&D expenditures (GERD) and the intensity of employment in high tech sectors (HiTechEmpl). Moreover, the lagged values of the Kr0/Kr1 ratio and the technological complexity index (TCI) are used. The Kr0/Kr1 ratio provides a rarity-weighted measure of variety of the technology knowledge base of a specific region. An higher value of this indicator implies that the patent portfolio of the region is specialised in more technological domains that are also more rare among all the analysed regions.

The TCI and the Kr0/Kr1 ratio have been computed using a period of 2 years with a moving average approach, i.e. the value of the TCI for region *i* in year *t* was based on the processing of data on patent filings in region *i* and all other regions in the sample, during the years *t*-1 and *t*-2. Since the adoption of a wider time window for the definition of Kr0, Kr1 and TCI could account for a higher inertial effect, we have also run models using a moving average of 4 years and similar results to those presented below were obtained. Standard summary statistics of the variables are presented in Table 2.

[TABLE 2 AROUND HERE]

The processes observed at the regional level can be affected by significant cross-regional interdependencies mediated by geographical distances. Hence, in our econometric analysis we correct for spatial autocorrelation. In particular, we have applied to our data a weighting matrix based on the inverse of the geographical distances between all the analysed regions and we have adopted Spatial Durbin Models (SDM)[8], in which the TCI and the Kr0/Kr1 variables were weighted through the region distance matrix. Since this class of models requires the use of balanced panel we have preliminary conducted an imputation for those control variables that

[8] The spatial Durbin models were estimated with the *xsmle* routine for STATA 12. We have also run the baseline models using standard OLS panel model with fixed effects. Results are confirmed. Related tables are available from the authors upon request.

had some missing observations[9]. The effect of the knowledge base indicators on the innovation output at the region level was first estimated separately and then jointly. In the empirical analysis we further investigate the presence of different patterns in the relationship between knowledge base composition and the innovation performance between high-performing regions and non high-performing regions according to the Regional Competitiveness Index presented in the previous section. In order to this do this, we interact a dummy variable for the high-performing regions (HIGH_REG), with the knowledge composition indicators (Kr0/Kr1 and TCI). Such approach allows us to maintain the spatial data structure while observing the presence of significant differentials between the two subgroups of regions.

[TABLE 3 HERE]

The results presented in Table 3 indicate that, after accounting for potential spatial autocorrelation and aggregated R&D inputs, the indicators of weighted variety and of technological complexity show a significant correlation with the subsequent level of patenting performance. The results of the Durbin spatial models highlight the actual presence of spatial autocorrelation, thus suggesting the appropriateness of this modelling approach. Results confirm the robustness of the proposed patent-based indicators. Interestingly, the splitting of regions in Model IV of Table 3 indicate that the patenting output of high-performing regions has a higher correlation with the complexity measure TCI, while the opposite is found for the indicator Kr0/Kr1. This evidence seems to provide ground for a deeper understanding of the meaning and implications of the knowledge base indicators. The TCI indicator displays a greater effect for high-performing regions that have possibly already achieved an average specialisation pattern focussed on more rare technologies. In this perspective, the indicator Kr0/Kr1 seem to capture mostly the process of technological reconfiguration of lagging regions, while the TCI is more effective in representing

[9] Missing data refer only to the aggregated region-level control variables (GERD and Hi-Tech Employment). We have used for this purpose the *mi-impute* routine of STATA 12 with 45 imputations. The related OLS model for the imputation includes time and region dummies. The indicators RVI and FMI (Fraction Missing Information) indicate the validity of the imputation process.

the properties of the dense network or interactions among technological capabilities, that in turn is correlated to a higher innovation output among more advanced regions.

6. Conclusions

The results of our empirical analysis suggest that the characterization of regional knowledge bases through the HH indicators provides interesting information to understanding its composition and the source of Jacob's externalities.

The analyses suggest that the Technological Complexity Index positively contributes to knowledge generation performance, supporting the view that Jacobs knowledge externalities stemming from the composition of the knowledge base are a complementary input for knowledge generation, next to R&D expenditures and the size of the local knowledge base.

Jacobs knowledge externalities stemming from the qualified composition of the local knowledge base do exert strong pecuniary effects that contribute the recombinant generation of knowledge: not only the quantity of external knowledge matters, but also its quality in terms of relative rarity of its components. Moreover, we found evidence that the TCI indicator displays a greater effect for high-performing regions that have possibly already achieved an average specialization pattern focussed on more rare technologies. In this perspective, the TCI appears to be particularly effective in representing the properties of the dense network or interactions among technological capabilities, that in turn is correlated to a higher innovation output among more advanced regions.

The overall evidence points to the relevance of the composition of the portfolio of technologies at the region level to support further innovation development. This result appears to be consistent with the idea behind the current European policy approach envisaged by the *smart specialisation strategy*. The observed patterns seem to indicate at the European level a process of diffusion and diversification of technological capabilities. However, it is among regions with more advanced economic systems, where the command of key enabling technological competencies is higher, that we observe an higher correlation between the complexity of the knowledge base and innovation performance. The increased technological interdependencies in new products and the hybridisation of related scientific domains calls for a policy design targeted at bringing together diverse and qualified knowledge pools rather than pursuing competitive advantage through the specialisation into a single technological domain or unqualified diversification.

References

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

Adams, J. (2006), Learning internal research and spillovers, *Economics of Innovation and New Technology* 15: 5-36.

Aghion, P., Howitt, P.W. (1998), *Endogenous Growth Theory*, MIT Press, Cambridge.

Annoni, P., Dijkstra, L. (2013), *EU Regional Competitiveness Index*, European Commission Report EUR 26060 EN.

Antonelli, C., David, P.A. (eds.) (2016), *The Economics of Knowledge and Knowledge Driven Economy*, Routledge, London.

Antonelli, C., Colombelli, A. (2015), External and internal knowledge in the knowledge generation function, *Industry and Innovation* 22(4): 273-298.

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

Archibugi, D., Pianta, M. (1992), *The Technological Specialization of Advanced Countries*, Kluwer Academic Publishers & The Commission of the European Communities, Dodrecht/Boston/London.

Arthur, W.B. (2009), *The Nature of Technology*, The Free Press, New York.

Asheim, B., Coenen, L. (2005), Knowledge bases and regional innovation systems: Comparing Nordic clusters, *Research Policy* 34, 1173-1190.

Baptista, R., Swann, P. (1998), Do firms in clusters innovate more? *Research Policy*, *27*(5): 525-540.

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

Boschma, R., & Iammarino, S. (2009). Related variety, trade linkages, and regional growth in Italy, *Economic Geography*, *85*(3), 289-311.

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

Boschma, R., Balland, P.A., Kogle, D.F. (2014), Relatedness and technological change in cities: The rise and fall of technological knowledge in US metropolitan areas from 1981 to 2010, *Industrial and Corporate Change* 24: 223–250

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.

Cooke, P., Uranga, M. G., & Etxebarria, G. (1997). Regional innovation systems: Institutional and organisational dimensions. *Research Policy 26*(4): 475-491.

Duguet E., MacGarvie, M. (2005), How well do patent citation measures flows of technology? Evidence from French innovation surveys, *Economics of Innovation and New Technology* 14: 375–93.

Frenken, K., Boschma, R. A. (2007), A theoretical framework for evolutionary economic geography: Industrial dynamics and urban growth as a branching process, *Journal of Economic Geography* 7:635–49.

Frenken, K., Van Oort, F., & Verburg, T. (2007), Related variety, unrelated variety and regional economic growth. *Regional Studies 41*(5): 685-697.

Galliano, D., Magrini, M.B., and Triboulet, P. (2014), Marshall's versus Jacobs' Externalities in Firm Innovation Performance: The Case of French Industry, *Regional Studies*, (ahead-of-print), 1-19.

Gehringer, A. (2011a), Pecuniary knowledge externalities and innovation: Intersectoral linkages and their effects beyond technological spillovers, *Economics of Innovation and New Technology* 20: 495-515.

Gehringer, A. (2011b), Pecuniary knowledge externalities across European countries – Are there leading sectors? *Industry and Innovation* 18: 415-436.

Gehringer, A. (2012), A new sectoral taxonomy based on pecuniary knowledge externalities: Knowledge interactions in a vertically integrated system, *Economic System Research* 24: 35-55.

Glaeser, E., Kallal, H., Scheinkman, J., Shleifer, A. (1992), Growth of cities. *Journal of Political Economy* 100 (6): 1126-1152.

Grossman, G.M., Helpman, E. (1991), The quality ladders of the theory of growth, *Review of Economic Studies* 58, 43-61.

Han, Y-J., Park, Y. (2006), Patent network analysis of inter-industrial knowledge flows: The case of Korea between traditional and emerging industries, *World Patent Information* 28, 235–247

Hidalgo, C.A., Hausmann, R. (2008), A network view of economic development, *Developing Alternative* 12, 5-10.

Hidalgo, C.A., Hausmann, R. (2009), The building block of economic complexity, *Proceedings of the National Academy of Sciences of the United States of America* 106: 10570-75.

Imbs, J., Wacziarg, R. (2003), Stages of diversification, *American Economic Review* 93 (1): 63-86.
Jacobs, J. (1969), *The Economy of Cities,* Jonathan Cape, London.
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.
Leoncini R., Maggioni M., Montresor, S. (1996), Intersectoral innovation flow and national technological systems: Network analysis for comparing Italy and Germany, *Research Policy* 25: 415–30.
Mina, A., Ramlogan, R., Tampubolon, G., Metcalfe, J.S. (2007), Mapping evolutionary trajectories: Applications to the growth and transformation of medical knowledge, *Research Policy* 36: 591–806.
Neffke, F., Henning, M., Boschma, R. (2011), How do regions diversify over time? Industry relatedness and the development of new growth paths in regions, *Economic Geography* 87: 237-265.
Nesta, L., Saviotti, P.P. (2005), Coherence of the knowledge base and the firm's innovative performance: evidence from the U.S. pharmaceutical industry, *Journal of Industrial Economics* 53: 123-142.
Nesta L., Saviotti P.P. (2006), Firm knowledge and market value in biotechnology, *Industrial and Corporate Change* 15: 625-652.
Noteboom B. (2000), *Learning and Innovation in Organizations and Economies*, Oxford University Press, Oxford.
Pakes, A., Griliches (1984), Patents and R&D at the firm level: A first look, in Griliches, Z. (ed.) *R & D, Patents, and Productivity*, University of Chicago Press for the NBER, Chicago, pp. 55-72.
Pianta, M., Meliciani, V. (1996), Technological specialization and economic performance in OECD countries, *Technology Analysis & Strategic Management* 8(2): 157-174.
Quatraro, F. (2010), Knowledge coherence variety and productivity growth: Manufacturing evidence from Italian regions, *Research Policy* 39: 1289-1302.
Rigby, D.L. (2015), Technological relatedness and knowledge space: Entry and exit of US cities from patent classes, *Regional Studies* 49 (11): 1922-1937.
Saviotti, P.P., Frenken K. (2008), Export variety and the economic performances of countries, *Journal of Evolutionary Economics* 18, 201-218.
Sternitzke, C., Bartkowski, A., Schramm, R. (2008), Visualizing patent statistics by means of social network analysis tools, *World Patent Information* 30:115–131.
Usai, S. (2011), The geography of inventive activity in OECD regions, *Regional Studies* 45: 711-731.
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.

FIGURES

Figure 1 – Kr0 vs Kr1 by region. H, M, and L indicate high-, medium-, and low-performance regions, respectively. Axes represent the average values of the indicators; the dashed line is a linear regression.

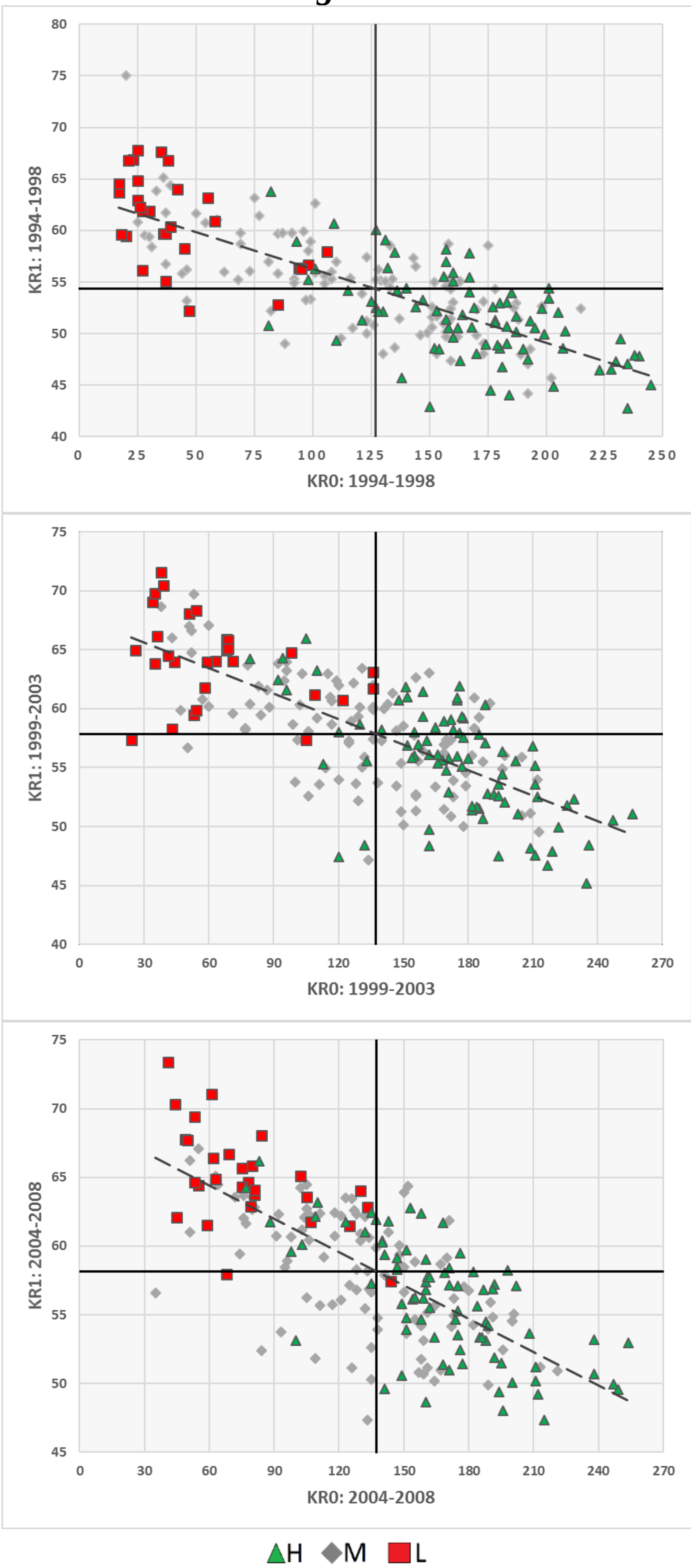


Figure 2 – Level of Kr0 and its variation (ΔKr0) between 1994-1998 and 2004-2008. H, M, and L indicate high-, medium-, and low-performance regions, respectively. Axes represent the average values of the indicators.

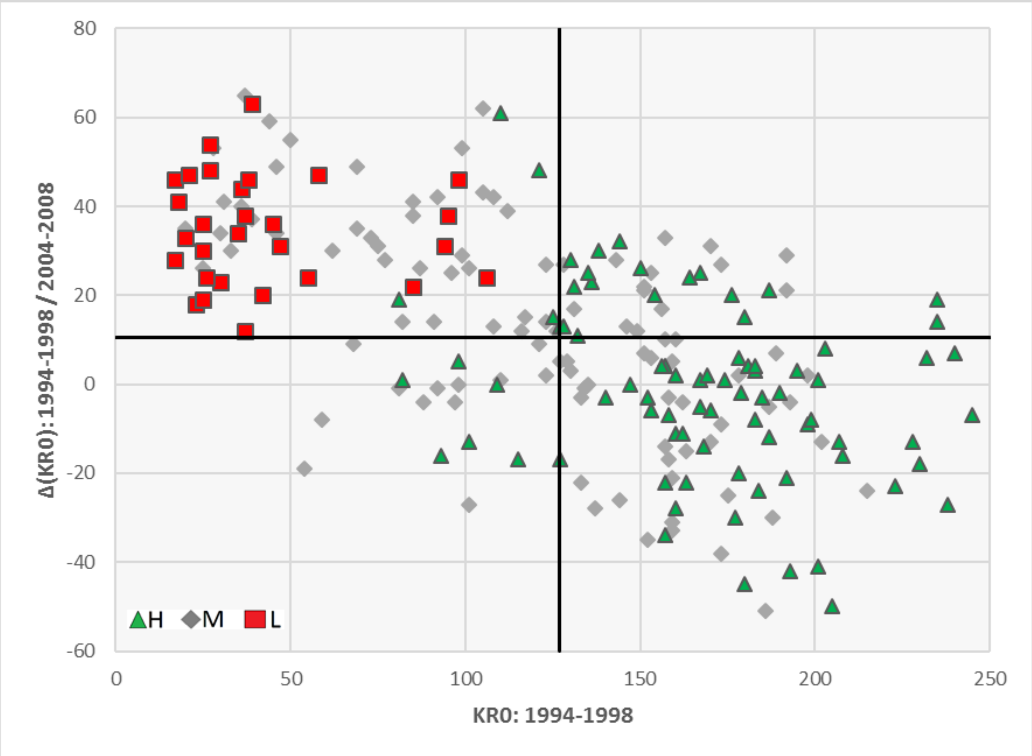


Figure 3 – Level of Kr1 and its variation (ΔKr1) between 1994-1998 and 2004-2008. H, M, and L indicate high-, medium-, and low-performance regions, respectively. Axes represent the average values of the indicators.

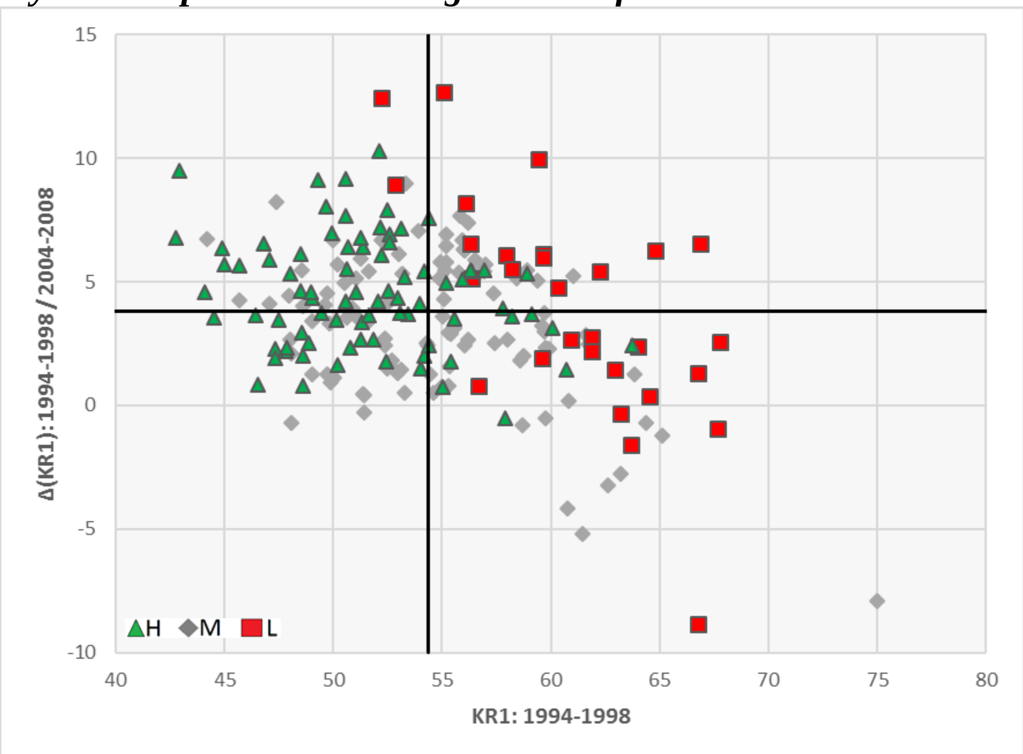
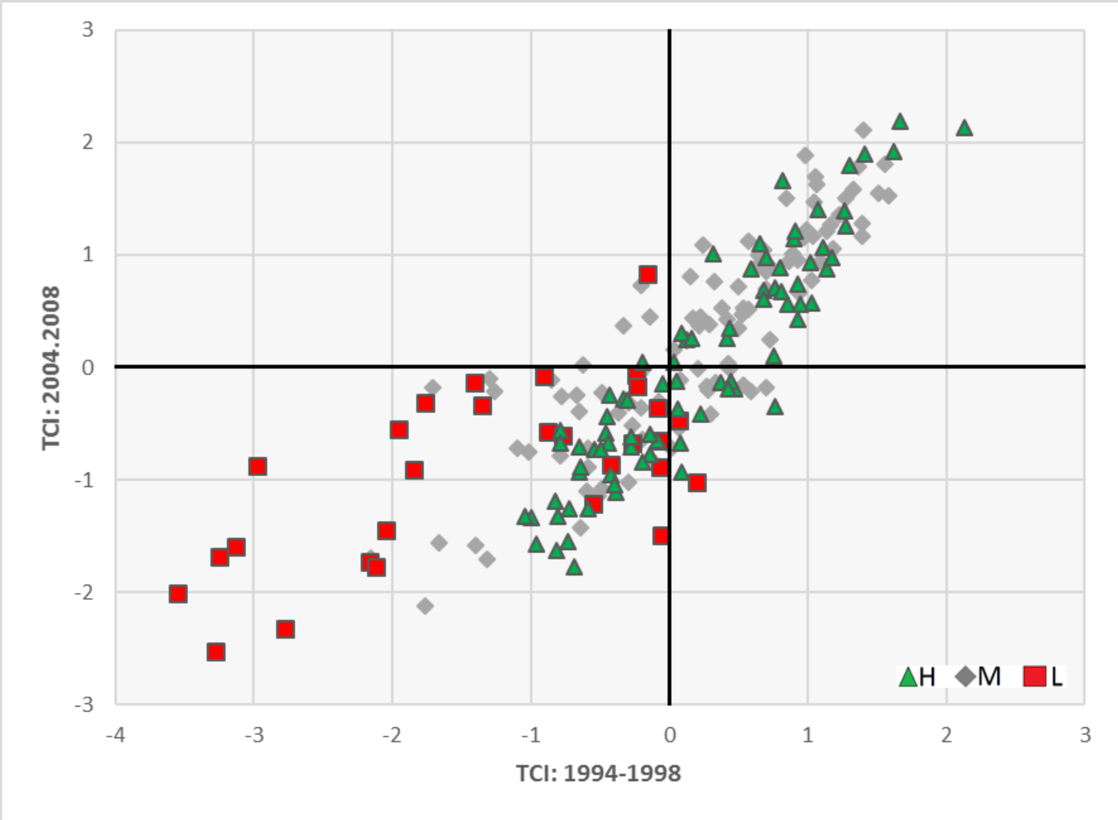


Figure 4 – TCI in 1994-1998 vs. TCI in 2004-2008. H, M, and L indicate high-, medium-, and low-performance regions, respectively. Axes represent the average values of the indicators.



TABLES

Table 1 - The search for Jacobs externalities

	LOW RARITY	HIGH RARITY
LOW VARIETY	<u>POOR SPECIALIZATION:</u> A LIMITED BUNDLE WITH HIGH LEVELS OF SPECIALIZATION IN UBIQUITOUS TECHNOLOGIES: LOW LEVELS OF JACOBS KNOWLEDGE EXTERNALITIES	<u>HYPERSPECIALIZATION:</u> LOW LEVELS OF VARIETY LIMIT RECOMBINATION BUT THE COMMAND OF RARE KNOWLEDGE INPUTS MAY MAKE IT HIGHLY PRODUCTIVE: HIGH VARIANCE IN JACOBS KNOWLEDGE EXTERNALITIES
HIGH VARIETY	<u>UNQUALIFIED VARIETY:</u> DISPERSION ACROSS LARGELY COMMON TECHNOLOGICAL ACTIVITIES: LOW LEVELS OF JACOBS KNOWLEDGE EXTERNALITIES	<u>QUALIFIED VARIETY:</u> A LARGE BUNDLE OF DIVERSE TECHNOLOGICAL ACTIVITIES SOME OF WHICH ARE RARE: HIGH LEVELS OF JACOBS KNOWLEDGE EXTERNALITIES

Table 2 – Variables definition and summary statistics

Variable	Definition	Mean	Std.	Min	Max
PAT	Log of the number of patent applications by region *i* in year *t*	5.413	1.487	1.386	8.948
Pop	Log of the population of region *i* in year *t*	7.370	0.698	5.486	9.359
GERD	Log of the gross regional R&D expenditures in region *i* and year *t*	5.978	1.297	2.073	9.583
HiTechEmpl	Log of employment in high tech sectors in region *i* and year *t*	6.505	0.698	4.482	8.543
HIGH_REG	A dummy variable that takes the value of one if region i belongs to the top tertile of the distribution of the analysed regions on the Regional Competitiveness Index	0.3		0	1
Kr0/Kr1	Ubiquity-weighted indicator of the technological diversification in region *i* in year *t*	2.254	1.351	0.057	6.344
TCI	Technology Complexity Index in region *i* and year *t*	0.000	0.998	-4.559	2.181

Table 3 – Spatial Durbin Model. Dependent variable: log of the number of new patent applications by region i in year t.

Models	I	II	III	IV
GERD (t-1)	0.156***	0.128***	0.152***	0.131***
	(0.024)	(0.024)	(0.024)	(0.024)
HiTechEmpl (t-1)	0.045	0.223	0.043	0.043
	(0.128)	(0.141)	(0.128)	(0.131)
Pop (t-1)	-0.002	-0.684**	0.043	-0.099
	(0.283)	(0.303)	(0.282)	(0.279)
Kr0/Kr1 (t-1)	0.318***		0.313***	0.432***
	(0.017)		(0.017)	(0.023)
TCI (t-1)		0.044***	0.029***	0.020**
		(0.010)	(0.009)	(0.010)
Kr0/Kr1 (t-1) * HIGH_REG				-0.265***
				(0.034)
TCI (t-1) * HIGH_REG				0.053*
				(0.029)
Weight - Kr0/Kr1 (t-1)	1.012***		0.896***	0.981***
	(0.103)		(0.109)	(0.111)
Weight – TCI (t-1)		0.863***	0.259***	0.251**
		(0.096)	(0.096)	(0.125)
Spatial rho	0.801***	0.921***	0.800***	0.761***
	(0.030)	(0.017)	(0.029)	(0.033)
Variance	0.038***	0.0457***	0.038***	0.037***
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	2782	2782	2782	2782
Avg RVI	0.220	0.192	0.178	0.131
Lar FMI	0.399	0.388	0.404	0.404
F test	268.93***	279.74***	222.75***	171.15***