

# RELATEDNESS, EXTERNAL LINKAGES AND REGIONAL INNOVATION IN EUROPE

Ernest Miguelez

GREThA, University of Bordeaux & AQR-IREA, University of Barcelona

[ernest.miguelez@u-bordeaux.fr](mailto:ernest.miguelez@u-bordeaux.fr)

Rosina Moreno

AQR-IREA, University of Barcelona

[rmoreno@ub.edu](mailto:rmoreno@ub.edu) (corresponding author)

**This is a pre-print of an article published in *Regional Studies*. The definitive publisher-authenticated version is Miguelez, E., Moreno, R., *Regional Studies* (2018). <http://dx.doi.org/10.1080/00343404.2017.1360478>**

## Abstract

This paper analyses if the generation of new knowledge benefits from the combination of similar or dissimilar pieces of existing technologies, in terms of their technological content (related versus unrelated variety), for the case of European regions. Specifically, it analyses the relevance of variety in the case of local knowledge as well as in the case of the knowledge coming from other regions. At the local level, it shows that, while related variety is conducive to regional innovation, unrelated variety does play a role too when it comes to radical innovations. Conversely, it also shows that external knowledge flows have a higher impact, the higher the similarity between these flows and the extant local knowledge base.

**Key words:** related variety, unrelated variety, patents, patent citations, knowledge

production function, European regions

**JEL:** O18, O31, O33, R11

## 1. INTRODUCTION

It is now an established fact in the literature that the combination and recombination of previously unconnected ideas lead to new knowledge production, subsequent technological innovations, and ensuing economic growth and well-being (Aghion and Howitt, 1992; Jones, 1995; Weitzman, 1998). Following a well-settled tradition in evolutionary economic geography, this paper argues that not all types of formerly existing knowledge are equally and successfully combined, and the results of such processes depend on the kind of knowledge put in contact in terms of their technological content – that is to say, the degree of knowledge relatedness (Boschma et al., 2009).

The innovation and economic geography literatures have long tried to understand whether firms located in agglomerations mainly learn from other local firms in the same industry or from other local firms in a range of other industries (Glaeser et al., 1992). The former dates back to Marshall's (1920) contributions on the benefits arising from spatial concentration. The latter relates to Jane Jacobs' contributions on cities, externalities and innovation (Jacobs, 1969; see also Glaeser et al., 1992). From her work we learn that a diversified economy brings benefits to local firms because it generates new knowledge and innovation steaming from the cross-fertilization of ideas across different industries. Following Frenken et al. (2007) and a large number of studies after them, we argue that Jacob's concept of diversification needs to be more thoroughly elaborated, by differentiating between diversification of related industries and diversification of unrelated industries – or related versus unrelated variety. Regions hosting related industries, with different but connected knowledge bases, can engage in recombinant

innovation. On the contrary, the combination of unrelated technologies is more difficult to succeed into the production of new ideas.

An issue that has been generally under-investigated by this particular literature is the role played by knowledge linkages across the space in introducing variety into regions. While most of the related literature is usually silent on the role of linkages across regions, thus implicitly assuming that innovation production draws mainly from geographically localized knowledge sources (Audretsch and Feldman, 2004), some scholars have recently posited that, at some point, co-located agents may start to combine and recombine local knowledge that eventually becomes redundant and less valuable. As a result, processes of negative lock-in may begin to occur (Boschma, 2005; David, 1993). Conversely, firms looking for external sources of knowledge may find that the knowledge they require is available beyond the boundaries of the region where the firm is located (Bergman and Maier, 2009; Bathelt et al., 2004). In the ongoing age of globalization characterized by predominantly open economies, it is naïve to assume that agents in regions source their knowledge inputs only from their immediate vicinity. In this scenario, this paper argues that not only being connected to the outside world matters, but also the degree of diversity between the external knowledge that is brought into the region and the existing knowledge base is important (Boschma et al., 2014).

In order to fill this gap, the paper estimates a knowledge production function (KPF) for the case of European regions, trying to ascertain what type of knowledge recombination – related or unrelated – is more conducive to regional innovation. Different from previous studies, the paper takes into account the geographical breadth of such knowledge. That is, it does not only analyze the relevance of local variety, but also how the external-to-

the-region knowledge flows fits into the local knowledge base. While adding the external dimension is crucial, this concern has been generally neglected by the related variety literature, and contributions introducing a “more geographical wisdom in the study of regional diversification” are still scarce (Boschma, 2016; Content and Frenken, 2016). Only Boschma and Iammarino (2009), for Italian regional growth, and Tavassoli and Carbonara (2014), for Swedish regional innovation, have concluded that it is not enough being connected to the outside world, but different, yet related, connections provide real learning opportunities and boost economic outcomes.

As a second objective, this study incorporates the idea that the combination of related technologies is not always a necessary condition for regional diversification, as unrelated diversification may occur too. The paper uses regional innovation as outcome variable, which allows us regressing not only innovation quantity, but also innovation quality (i.e., breakthrough innovations) on related variety, unrelated variety and connectedness with other regions.<sup>1</sup> This allows us to test whether breakthrough innovations draw more on unrelated and distant pieces of knowledge, as ideas with high impact tend to stem from knowledge cross-fertilization and the combination of unrelated technologies (Fleming, 2001; Saviotti and Frenken, 2008). Again, very few systematic evidence exists in this respect, being Tavassoli and Carbonara (2014) and Castaldi et al. (2015) the exceptions, for the case of, respectively, Swedish regions and US states.

In sum, this paper draws on these ideas and studies the relevance of the degree of relatedness among previous existing pieces of knowledge for the generation of new ideas, while differentiating between relatedness in the local technological structure of regions

(related vs unrelated variety) and relatedness between the internal knowledge base and the extra-regional sources of knowledge.

Using regional innovation intensity as outcome variable – patents per capita, contrary to large part of the related literature, which focuses on economic growth or employment, is an important departure from the majority of studies, for several reasons: first, while most studies conclude that related variety facilitates knowledge spillovers, which are conducive to innovation, this specific relationship is barely tested, but rather, implicitly assumed to exist in the link between the regional structure of employment or exports with economic growth. However, recent studies suggest that growth effects of related variety may be specific to knowledge-intensive industries only (Content and Frenken, 2016). In consequence, we focus here solely on innovation production and on patent intensive sectors. Moreover, we compute variety indexes using the technological classification provided in patent documents. In particular, we exploit technology information using the International Patent Classification (IPC) codes contained in patent applications to the European Patent Office (EPO) to build the diversity indexes, establishing a more direct link between regional diversification and its underlying technological nature. Second, and more importantly, our study is one of the few investigating cross-regional linkages and related variety, for which trade data has mostly been used to depict linkages across regions (Boschma and Iammarino, 2009; Tavassoli and Carbonara, 2014). Our focus on innovative sectors allows us using citations to patents as a cleaner and more direct measure of knowledge flows across the space. Patent citations directly point to the prior knowledge to which the current innovations draw upon, and therefore represent a good proxy for cross-regional linkages and knowledge flows (Jaffe and Trajtenberg, 1999; Schoenmakers and Duysters, 2010). Finally, using innovation as outcome variable allows

us exploiting heterogeneity in patent quality and its relationship with related and unrelated diversification, as mentioned earlier.

This paper makes use of a large sample of European regions (255 NUTS2 regions) belonging to 25 countries, which, to our knowledge, correspond to the largest coverage in Europe of studies of this kind. Moreover, the study utilizes data for several years, allowing us to introduce time and region fixed-effects (FE) to control for a large number of unobservables.

The outline of the paper is as follows. Section 2 reviews the related literature. Section 3 sets the empirical analysis and describes the data. We give the main results in section 4 and finally section 5 concludes.

## **2. LITERATURE REVIEW**

It is widely accepted in the literature that innovation is a process of accumulation and recombination of previously existing ideas (Weitzman, 1998). A key point is, however, if any potential combination of existing knowledge is equally successful, or only the connection of different, but related, pieces of knowledge is most effective (Frenken et al., 2007). Besides, it is established that innovation production draws mainly from geographically localized knowledge sources (Audretsch and Feldman, 2004). However, scholars have also signalled that the combination of local knowledge may eventually become redundant (Arthur, 1989; David, 1985), leading firms to look for external sources of ideas. This section discusses theoretical and empirical contributions on the different

role of related and unrelated variety on regional outcomes, both within the region (section 2.1) and across geographical areas (section 2.2).

### **2.1 Related and unrelated variety at the regional level**

Much research on the geography of innovation and regional development has addressed the question of whether specialization or diversity boosts local innovation. Proponents of the former argue that firms tend to learn from other firms in the same industry, and therefore specialization facilitates knowledge spillovers and subsequent growth. Meanwhile, advocates of the latter contend that diverse economies facilitate barter of different pieces of knowledge across industries, which are more prone to produce innovations and economic prosperity – despite implying higher communication costs between agents. The concept of diversity is complex and subtle, as first signalled by Frenken et al. (2007). These authors pose the central question of whether it is related or unrelated diversity which is most relevant for growth. Related diversity, or variety, facilitates local knowledge spillovers across industries at a lower cost. This is because the cognitive distance across these industries is not too large so that complementarities exist among them in terms of shared competences and capabilities, which enable effective connections as well as sharing knowledge and information. Conversely, unrelated variety may slow down the diffusion of ideas, given that they draw on very different and completely disconnected knowledge bases making it more uncertain and costly to engage in recombinant innovation, thereby hampering the production of new local innovation.

Frenken's et al (2007) pioneering study shows how related variety impacts regional economic growth in the Netherlands. Results are confirmed by studies in other countries:

Bishop and Gripaos (2010) for Great Britain, Boschma and Iammarino (2009) and Quatraro (2010) for Italy, Hartog et al. (2012) for Finland and Boschma et al. (2012) for Spain. The role of unrelated variety is more controversial: whereas Bishop and Gripaos (2010) find that unrelated variety affects employment growth in a larger set of industries than related variety, Boschma et al. (2012) and Hartog et al. (2012) do not find any growth effect. Meanwhile, Frenken et al. (2007) find that unrelated variety dampens unemployment growth, which the authors interpret as evidence of unrelated industries spreading risks of potential negative shocks – known as the portfolio effect of variety.<sup>2</sup>

Despite the emphasis put on earlier studies on related variety as knowledge spillovers facilitator, implicitly, these studies assume that variety and employment or economic growth are linked to each other via innovation. Little work has been done, however, on directly examining the impact of technological variety on innovation performance. To our knowledge, only Tavassoli and Carbonara (2014) and Castaldi et al. (2015) analyse the role of related and unrelated variety on regional innovation output, for the Swedish and the United States (US) cases, respectively. Their findings suggest that when it comes to variety of knowledge within regions or US states, unrelated variety does not affect regional innovation output in general, whereas the impact is robust and positive for related variety.<sup>3</sup>

To reiterate, as Frenken et al. (2007) put it, related variety “improves the opportunities to interact, copy, modify, and recombine ideas, practices and technologies across industries giving rise to Jacobs externalities” (p. 59). Therefore, in search for recombination, agents focus mainly on the technological pieces in which they have prior experience (related variety), since this previous expertise allows them to understand better the nature of the



new knowledge. As a consequence, when a region presents a diversity of related technologies, connections are more effectively established given that related technologies are more easily recombined. Therefore, we expect related variety to be crucial in the generation of regional innovation.

In spite of the previous discussion, scholars have argued that truly important innovations may stem from the combination of previously unrelated technologies (Saviotti and Frenken, 2008). This is so because, when combining more different capabilities, despite implying higher costs and risks, it can result in the production of radical breakthroughs, i.e., innovation with a high technological and economic impact (R. Boschma, 2016). As Fleming (2001) puts it, knowledge producers that experiment with new and unusual components and combinations may arrive to less useful innovations on average, but with large variability, which results in turn in both failure and breakthrough inventions. If successful, unrelated pieces of knowledge become related in the form of a new invention that paves the way to future technological developments and further innovation, leading to “new operational principles, functionalities and applications” (Castaldi et al., 2015; p. 770) . In consequence, we expect unrelated variety to be key in the generation of more radical innovations.

## **2.2 Relatedness and extraregional linkages**

An important debate within the geography of innovation literature that has emerged recently is the role of external knowledge in the process of regional knowledge creation. Indeed, the widely accepted assumption that agents usually source their innovations from their immediate vicinity might have limited our understanding of the ways in which

knowledge flows across the space and the way in which innovations are generated (Coe and Bunnell, 2003). Thus, it has been highlighted the increasing importance of agents' needs to access extra-local knowledge pools to overcome potential situations of regional 'lock-in' (Boschma, 2005; Camagni, 1991; Grabher, 1993; David, 1993). Even local unrelated activities may become related when they are successfully combined, eventually becoming redundant too (Boschma, 2016; Desrochers and Leppälä, 2011). Thus, recent empirical works have extensively documented the influence of extra-local knowledge sources on firms' and regions' innovative performance and knowledge acquisition (Owen-Smith and Powell, 2004; Moreno et al., 2005; Gittelman, 2007; Gertler and Levitte, 2005; Rosenkopf and Almeida, 2003; Zhou and Li, 2012; Bottazzi and Peri, 2003; Miguelez and Moreno, 2013).

Yet, not only being connected to the outside world matters, but also the degree of relatedness between the external knowledge that is brought into the region and the existing knowledge base (Boschma et al., 2014). While the external dimension is crucial to understand regional growth, it has been generally neglected by the related variety literature (Boschma, 2016; Content and Frenken, 2016), with only few exceptions (Boschma and Iammarino, 2009; Tavassoli and Carbonara, 2014). This paper argues that in the ongoing globalized world characterized by predominantly open economies, it is naïve to assume that agents in regions source their knowledge inputs only from their local environment. Regions lacking certain capabilities could still diversify if they leverage knowledge inputs coming from external sources and allow the different unrelated sectors to find their way to interact with related sectors located beyond their regional borders.

The scarce extant empirical evidence on the role of relatedness of extra-regional knowledge flows has approached the issue using regional trade data –either imports or exports (Boschma and Iammarino, 2009, for Italian regional employment growth; and Tavassoli and Carbonara, 2014, for Swedish regional innovation). Their findings suggest that it is not enough being connected to the outside world, but different, yet related, connections provide real learning opportunities and boost economic outcomes.<sup>4</sup> When the external knowledge basically integrates prior art from the same technologies from within the region, it can be easily absorbed but the new knowledge will not add much to the existing local one. On the contrary, when the external knowledge brings technologies different from the local ones, it will be more difficult to understand but once it is integrated, the chances that they lead to successful outcomes are higher. All in all, in analogy to section 2.1, we expect extraregional knowledge inflows to be most effective when they are different, but related, to the local knowledge base.

### **3. EMPIRICAL ANALYSIS**

#### **3.1 Empirical model**

We test our hypotheses under a KPF framework at the regional level. Our point of departure is the simplest specification of this model:

$$Y_{it} = f(RD_{it}, Z_{it}), \quad (1)$$

where  $Y$  is the innovative output of a given region, which depends on regional R&D expenditures (RD) as well as  $Z$ , a number of time-variant controls that account for

specific features of the region  $i$  at time  $t$ . Among them, we include measures of variety and relatedness, as explained in the following subsections. Note that regional differences in size are accounted for by dividing the dependent and explanatory variables by total population. All in all, the following model is suggested:

$$\ln Ypc_{it} = \beta \cdot \ln RDpc_{it} + Z_{it} + \delta_i + \delta_t + \varepsilon_{it}, \quad (2)$$

where  $\ln Ypc_{it}$  is the log-transformation of the annual number of patent applications per million inhabitants in region  $i$  and year  $t$ ,  $\ln RDpc_{it}$  is the log-transformation of R&D expenditures per capita in region  $i$  and year  $t$ , and  $Z$  are a number of focal variables – as explained below – and controls. For the latter, we include a proxy for human capital, measured as the share of human resources devoted to science and technology (HRST), as well as two variables accounting for differences in the economic structure of regions: the share of manufacturing employment (ShareInd), and the share of employment in high-technology manufacturing and knowledge-intensive, high-technology services (High-tech Empl). In addition,  $\delta_i$  and  $\delta_t$  stand for, respectively, regional FE and time FE. In order to consider deviations from the theory, a well-behaved error term is also introduced,  $\varepsilon_{it}$ .

Our empirical model (the regional KPF) draws mainly from a large number of contributions in regional science and innovation economics trying to understand the role played by regional innovative efforts (R&D) and the technological structure of regions on regional innovative output. We are aware that our reduced-form model does not account for all possible determinants of regional innovation intensity. Thus, several

studies have extended the regional KPF to include a larger number of potential non-technology determinants of regional innovation outputs. For instance, one interesting avenue of research is the role of institutions and social capital on innovation, and more importantly, how they influence regional variety's role in fostering regional innovation (see Boschma, 2016, for a claim to do research in this direction). However, this lies beyond the primary focus of the present analysis. Yet, contrary to still the large majority of empirical studies using the regional KPF, we control for region FEs, and therefore account for all time-invariant features of regions that may influence the regional production of innovations (with institutions or social capital variables, which evolve slowly over time, being partially controlled for through these FEs).

### **3.2 Related and unrelated variety**

We start our analysis with a simple model that does not account for the influence of non-local capabilities – which will be introduced progressively (see section 3.3). Our first inquiry concerns the impact of knowledge diversification on regional patenting activity. In line with previous papers, as a proxy for diversified knowledge we measure variety as well as related and unrelated variety with entropy measures (Frenken et al., 2007). We borrow from Castaldi et al. (2015) the use of the technological classification of patents in order to construct the measures of regional knowledge variety. Our entropy indicators are computed using information retrieved from applications to the EPO. In particular, we use the IPC system, which provides a hierarchical system of codes for the classification of patents according to the different areas of technology to which they pertain – directly assigned by the patent office, the EPO in this case. These codes are grouped into eight sections, which are the highest level of hierarchy of the classification. Each section is

divided into three-digit classes and four-digit subclasses. The current version of the IPC classification contains 635 technological subclasses.<sup>5</sup> Scholars have reorganized these technological subclasses in meaningful fields and broad fields of technology, similar to the grouping of products or economic activities into sectors (such as the Standard International Trade Classification used in trade or International Standard Industrial Classification of All Economic Activities). The aim of this grouping is to allow time and cross-country comparisons of innovation activities, and it is based on minimizing technological heterogeneity within technology fields and broad fields. Here we use the classification built by Schmoch (2008), which grouped subclasses into 35 technology fields (35-field), which are further grouped into 5 broad fields (5-field), namely: Electrical engineering, Instruments, Chemistry, Mechanical engineering, and Other fields.<sup>6</sup>

Using the IPC codes and Schmoch's (2008) classification of technological fields, the variety variable measures the degree of knowledge diversification through the computation of an entropy measure at the four-digit level (subclasses), where  $p_j$  is the share of the four-digit sector  $j$ :

$$Variety = \sum_{j=1}^J p_j \log_2 \left( \frac{1}{p_j} \right) \quad (3)$$

The value of this index will be higher in regions characterized by a high diversified sectoral composition in its knowledge base.

We break down this measure in two different indicators. Following Frenken et al. (2007), if all four-digit subclasses  $j$  fall under a 35-field technology  $S_g$ , where  $g=1, \dots, G$ , it is possible to derive the 35-field shares,  $P_g$ , by summing the four-digit shares  $p_j$

$$P_g = \sum_{j \in S_g} p_j \quad (4)$$

Related variety is then measured by the weighted sum of the entropy at the four-digit within each 35-field technology:

$$RV = \sum_{g=1}^G P_g H_g \quad (5)$$

where:

$$H_g = \sum_{j \in S_g} \frac{p_j}{P_g} \log_2 \left( \frac{1}{p_j/P_g} \right) \quad (6)$$

Equation (6) measures the diversity of a region's portfolio at the most fine disaggregation. Thus, it assumes that sectors that belong to the same 35-field technology are technologically related to each other and, as a consequence, can learn from each other through knowledge spillovers.

Unrelated variety is proxied by the entropy of the 5-field distribution. Formally, being  $K$  the total number of 5-field sectors ( $k=1, \dots, K$ ), the unrelated variety index is given by

$$UV = \sum_{k=1}^K p_k \log_2 \left( \frac{1}{p_k} \right) \quad (7)$$

Thus, equation (7) measures the extent to which a region is diversified in very different types of activities. This measure assumes that technologies that do not share the same broad field (5-field) are unrelated to each other. Theoretically, high levels of this variable are associated to less knowledge spillovers.

The indices of related and unrelated variety are not opposites. One region can have both a high related variety (diversified into many specific subclasses in each field) and a high unrelated variety (diversified into unrelated broad 5-field technologies). In fact, they tend to correlate positively (Frenken et al., 2007; Boschma et al., 2012), although it is not always the case. In addition, given the decomposable nature of the entropy measure, variety calculated at different digit levels can be included in a regression analysis without necessarily generating collinearity.

Following with the empirical model sketched above, we include now the indices proxying for related and unrelated variety in the  $Z$  vector including controls that account for specific features of the region,

$$Z_{it} = g(RV_{it}, UV_{it}), \quad (8)$$

which once inserted into the main equation yields to:



$$\ln Ypc_{it} = \beta \cdot \ln RDpc_{it-1} + \lambda_1 HRST_{it-1} + \lambda_2 ShareInd_{it-1} + \lambda_2 High-techEmpl_{it-1} + \omega_1 RV_{it-1} + \omega_2 UV_{it-1} + \delta_i + \delta_t + \varepsilon_{it} \quad (9)$$

Note that we introduce the subscript t-1 to all explanatory variables in order to indicate that they have been time lagged one period to lessen endogeneity concerns due to system feedbacks. Section 3.4 includes further details regarding the construction of all the variables used in the present analysis.

### 3.3 Relatedness and external interactions

Here we extend our baseline model to account for the role of non-local knowledge sources in the process of regional knowledge creation. Although some studies, at the level of European regions, have consistently shown the importance of cross-regional interactions to the process of regional innovation (Maggioni and Uberti, 2009; Ponds et al., 2010), little attention has been paid to which kind of external interactions may be more beneficial. We conjecture that, even if new variety may enter a region thanks to the interactions with other regions – in the form of, e.g. trade linkages, FDI, research collaboration or labour mobility, extra-regional knowledge flows should be related, but not too similar, to the technological base of a region in order to positively impact the region's outcomes.

We directly look at the actual knowledge flows through the use of patent citations as a proxy for these flows. Patent citations point directly to prior art on which the patent is based (Trajtenberg, 1990) and, consequently, represent a “paper trail” worthwhile for the analysis of knowledge diffusion (Jaffe et al., 1993). Since Jaffe's et al. pioneering paper,

patent citations have been considered to be useful to depict knowledge linkages between inventions, inventors and applicants along time, geographical space and technological fields, among other dimensions (Hall et al., 2005; Jaffe and Trajtenberg, 1999; Schoenmakers and Duysters, 2010). In our case, since patents record the residence of the inventors, they are an exceptional source for studying knowledge flows across regions.

To build our variables, we use citations made by inventors resident in the focal region to EPO applications of inventors living outside the region. In particular, we look at backward citations listed in patents produced in a given region and collect the cited patents (alongside their technology codes) with all inventors living outside the region. Even though the use of patent citations does not come without limitations – e.g., some citations are added by the examiner, and not the applicant (Alcacer and Gittelman, 2006), they have been widely used in innovation economics as a proxy for knowledge flows (Criscuolo and Verspagen, 2008; Jaffe et al., 1993; Jaffe and Trajtenberg, 1999). Moreover, as citations relate cited patents with citing ones, they include detailed descriptions of technological characteristics and classification into technical domains (Popp et al., 2011) allowing the computation of the necessary indexes.

We use an indicator of *RELATEDNESS* to account for knowledge inflows that are related, but are not the same, to the actual knowledge base of the region. This indicator is built in a similar fashion to Boschma and Iammarino (2009):

$$RELATEDNESS = \sum_j CIT_4^M(j) PAT_4(j) \quad (10)$$

where  $CIT_4^M(j)$  is the entropy measure obtained with data for extra-regional backward citations in four-digit technologies (subclasses) other than  $j$ , but within the same 35-field technology, and  $PAT_4(j)$  is the relative size of the four-digit patent technology  $j$  in the total regional patenting. The idea is that for each four-digit patent technology in a region (e.g., technology C07G), we measure the entropy of the citations to patents from the other four-digit subclasses (e.g., C07K, C12M, C12N, C12P, C12Q, C12R, and C12S) pertaining to the same 35-field sector (e.g., the biotechnology field), excluding the focal four-digit subclass itself (i.e., subclass C07G).

In order to complement the analysis, and again in line with Boschma and Iammarino (2009), we also use an index to determine the similarity between the external knowledge entering a region and its existing knowledge base (*SIMILARITY*). In our case it is computed as the sum of the products of the absolute sizes of the four-digit subclass patents ( $PAT_4(j)$ ), as a proxy of the knowledge stock in a region, and the four-digit subclass extra-regional patents the former have cited ( $CIT_4(j)$ ):

$$SIMILARITY = \log \sum_j PAT_4(j) CIT_4(j) \quad (11)$$

This measure gets a maximum when the region is specialized in just one technology and this technology coincides with the extra-regional patents cited. The lowest values are obtained when the more diverse the region is in its patent portfolio as well as in the extra-regional patents it cites, and at the same time the less similar both profiles are. When a region gets knowledge from other regions, but such knowledge comes from the same technologies that are present in the region, the knowledge base of the economy will be

able to absorb it but it will not add much to the existing knowledge. Therefore, we expect SIMILARITY to have little or null effect on regional innovation.

With these two indices (RELATEDNESS and SIMILARITY) we aim to measure how close the knowledge that flows into a region is to the current regional knowledge stock of a given region, in order to infer the role of such relatedness in the creation of new knowledge.

### **3.4 Data**

We use a sample of 255 NUTS2 European regions of 25 countries – EU-27 (except Cyprus and Malta, as well as Denmark and Greece, for which we have very little information at the NUTS2 level) plus Norway and Switzerland, to estimate a regional KPF from 1999 to 2007. Our dependent variable, innovation output, is measured by patent applications, a variable widely used in the literature to proxy innovation outcomes. As widely documented, this proxy presents serious caveats since not all inventions are patented, nor do they all have the same economic impact, as they are not all commercially exploitable (Griliches, 1991). In spite of these shortcomings, patent data have been considered useful for proxying inventiveness as they present minimal standards of novelty, originality and potential profits, and as such are a good proxy for economically profitable ideas (Bottazzi and Peri, 2003). We retrieve patent data at the regional level from the OECD REGPAT database – July 2013 edition (Maraut et al., 2008). When patents have been produced by inventors resident in different NUTS2, they have been fractionally assigned to the different regions, according to the number of inventors out of all inventors listed in a patent living there – fractional counting.

We slightly modify our dependent variable in order to account not only for the quantity of patents produced, but also for their quality – as explained in previous sections. As largely argued in the related literature, the number of forward citations received presumably conveys information about the relevance of patents, thus providing a way of assessing the enormous heterogeneity in the value of patents (Hall et al., 2005). This extreme is confirmed by several studies that have found strong correlations between the number of forward citations received and the economic value of patents (Trajtenberg, 1990; Harhoff et al., 1999; Lanjouw and Schankerman, 2004). We therefore use citations as an imperfect, but widely used, proxy for patent quality and weight the number of patents by the number of citations the patent has received in subsequent patent documents.<sup>7</sup>

As for the explanatory variables, R&D expenditures data (both private and public expenditures in regions) were collected from Eurostat and some National Statistical Offices. Data to measure ShareInd and High-tech Empl were collected also from Eurostat. As for the level of human capital of regions, which likely determines the regions' capacity to transform technological inputs into outputs, we use the variable HRST, which, according to Eurostat, include all tertiary educated workers employed in science and technology occupations (over all workers in the region).<sup>8</sup>

As mentioned above, variety indexes are constructed using the information of IPC codes listed in patent documents (again from the OECD REGPAT database – July 2013 edition). Again, based on the available data, there are 635 four-digit patent classes, 35 technological fields and 5 broad fields. Knowledge flows are proxied through patent

citations as explained in section 3.3. We use unit-record data retrieved from EPO patents to construct the patent citation variables (OECD Citations database, July 2013 edition; see Webb et al., 2005). All the patent data used to build the focal explanatory variables are retrieved for moving time windows of five years.

Table 1 provides summary statistics of the variables used in the present analysis whereas the correlation matrix of explanatory variables is given in Table A.2 of the online appendix.<sup>9</sup> We observe high correlations between some variables, although most of them do not jointly appear in the same regressions. For the remaining, table A.5 of the online appendix shows additional regressions in which we remove some of the problematic variables, to ensure that our results and conclusions hold.

[Insert Table 1 about here]

Further, figure A.1 in the online appendix depicts the distribution of our variables (dependent and explanatory) in maps – as averages of the whole period. Interestingly, even if some of these variables seem to follow the same concentration pattern in core regions of Europe, some others seem to be more spread across the space.

## **4. RESULTS**

### **4.1 Local variety and innovation**

We estimate an unbalanced panel model of 9 periods (1999-2007). Table 2 provides the two-way FE estimates for the regional KPF model, including all the controls listed in

section 3. Columns (i) to (iii) use as dependent variable the logarithmic transformation of the number of patents per million inhabitants.

[Insert Table 2 about here]

In all the cases, the Hausman test rejects the null hypothesis that individual effects are uncorrelated with the independent variables, so the FE model is preferred to the expense of the random-effects – results available on request. In general, the KPF holds in the European regional case for the period under consideration. The elasticity of patents with respect to R&D expenditures presents significant values (0.13-0.22), which is in line with the value obtained in the literature (Jaffe, 1989; Bottazzi and Peri, 2003).

With respect to the variety index, results indicate that the variety in knowledge stocks of regions is indeed positively and significantly related to regions' innovation output, similar to the results for the role of variety on employment and productivity (Boschma and Iammarino, 2009). Interestingly, once variety is split into related and unrelated, only related variety is significant. This result indicates that the higher the number of related technologies in a region, the larger the knowledge spillovers and, as a consequence, the more the learning opportunities across them (Frenken et al., 2007). That is, learning opportunities generated by a variety of technologies within the region are relevant when such technologies are related, which ultimately will generate more knowledge externalities across them. Meanwhile, if the knowledge flows across technologies are far away from each other (unrelated variety), it will be more difficult to assemble them and produce new ideas and innovation.

Columns (iv) and (v) of Table 2 look at patent quality, as explained in section 3. All our results and conclusions with respect to columns (ii) and (iii) hold, except for the case of unrelated variety, that increases considerably its point estimate and becomes now highly significant. It seems therefore that when the combination of unrelated technologies is attained, not only general innovation is obtained (as suggested by the positive and significant parameter for related variety) but also knowledge of presumably high value and economic impact can be achieved, which accords with our expectations.

Overall, our results are qualitatively comparable to recent studies that have looked at related variety and innovation in regions – i.e., Castaldi et al. (2015) and Tavassoli and Carbonara (2014), even though we do not share with them neither the regions analyzed (US states and Swedish functional regions, respectively) nor their estimation method (negative binomial models).<sup>10</sup> For comparison purposes, table A.3 in the online appendix presents negative binomial estimates of our preferred models. This implies taking the number of patents as dependent variable instead of the number of patents per capita, and then having R&D as a regressor and not R&D per capita. We also add population as a control to account for size effects (in columns (iii) to (vi)). Results are comparable to our OLS estimates – although Castaldi et al. (2015) do not find evidence of a positive effect of related variety on breakthrough innovations, as we do. Columns (v) and (vi) present random effects estimations to make our paper fully comparable to Tavassoli and Carbonara (2014). As expected, some of the coefficients become larger, making them closer to the ones found by the mentioned authors.<sup>11</sup>

Next, as argued in the introductory and theory sections, it is critical to account for extra-local knowledge sources on regions' innovative performance, as well as the degree of



relatedness between the external knowledge that is brought into the region and the existing knowledge base (Boschma et al., 2014). This issue is important from a methodological viewpoint too, as estimates in earlier regressions could be biased if the external dimension is not accounted for. We discuss this in turn.

#### **4.2 Technological relatedness and external linkages**

This section looks at the role of external-to-the-region inflows of knowledge. To do so, we introduce a variable accounting for external flows of knowledge which are different, but related, to the local knowledge base (RELATEDNESS), using data on patent citations to build it. For completeness, we also build a variable proxying for the amount of incoming knowledge flows that remain within the same technology (SIMILARITY).

Table 3 shows the results when the RELATEDNESS and the SIMILARITY indices are included to explicitly consider to what extent the knowledge that flows from other regions is related to the knowledge stock of the host region. The remaining explanatory variables are those of Table 2. Reassuringly, as observed in column (i), the majority of coefficients do not change to a large extent, which indicates that the omission of the external dimension in Table 2 was not biasing our results concerning the role of related and unrelated variety.

From Table 3, column (i), we also learn that, contrary to our initial assumptions, RELATEDNESS does not significantly correlate with regional innovation. Thus, it seems that knowledge inflows that are different, but related, to the local knowledge base, do not create useful interconnections that can end up producing any significant innovation

outcome. In turn, and against our expectations, the higher the SIMILARITY between the technological composition of the local knowledge and that of the cross-regional knowledge flows, the higher the impact on the regions' innovative output. In other words, if the knowledge that flows into a region comes from technologies in which the region already patents, there seems to be plenty of opportunities for using such knowledge in a creative way. As in Boschma et al. (2009), we interpret these results as evidence that the knowledge coming from other regions already convey a certain degree of novelty as compared to the local knowledge base, which is not embodied in the technological classification used in the present paper.

[Insert Table 3 about here]

Interestingly, when the patents are weighted by their quality (column ii), the coefficient accompanying the RELATEDNESS index increases considerably and becomes statistically significant, suggesting that an extra-regional knowledge that is complementary, but not similar, to the existing knowledge base in the region will particularly boost interactive learning that can bring out breakthrough innovations. We conclude, therefore, that in order to generate average innovations, it is necessary to have a certain level of technological similarity so as to have the opportunity to learn and absorb across technologies coming from different regions. Whereas for the generation of more radical innovations, related, but not the same, incoming knowledge flows are also critical.

### **4.3 Robustness analysis**

Several robustness analyses are presented in the online appendix. In table A.4 we test the theoretical statements discussed earlier through the use of a more general dependent variable on regional economic performance, such as the annual growth rate of GDP per capita. Despite the fact that GDP growth does not reflect a direct measure of innovation, its use avoids potential criticisms derived from the use of patent data to build both the dependent and independent variables, as we did in previous sections. Data on regional GDP per capita is retrieved from Eurostat, and the dependent variable is computed as the log of the ratio between per capita GDP at time  $t_1$  and per capita GDP at  $t_0$ . Moreover, regressions include the log of per capita GDP at  $t_0$  as an additional control, as done in much of the growth literature.

Results reported in columns (i) and (ii) concerning related and unrelated variety are in line with much of the related literature for specific countries (Frenken et al., 2007, for the Netherlands; Boschma and Iammarino, 2009, for Italy; Bishop and Gripiaios, 2010, for Great Britain; Quatraro, 2010, for Italy; Hartog et al., 2012, for Finland and Boschma et al., 2012, for Spain) even if in our regressions, variety indicators are computed using technology fields from patent applications, instead of employment by economic activities. The results reported show the significant impact of variety, both in related and unrelated technologies. This evidence supports the hypothesis that economic growth benefits from diversification in technologies too. Note that in previous tables we found that unrelated variety only impacts innovation if weighted by their value using forward citations – breakthrough innovations. Interestingly, both related and unrelated variety strongly influence regional economic growth, which we attribute to the strong link between economic growth and breakthrough innovations, as witnessed by the recent report of the World Intellectual Property Organization (Wipo, 2015). Results concerning incoming

knowledge flows and regional economic growth (column (iii)) are also consistent with the previous results presented in Table 3.

Reassuringly, we have shown that our results are not driven by mechanical correlation between dependent and independent variables, given that the use of an alternative measure not directly retrieved from patent documents, such as per capita GDP growth, does support our key findings.

Finally, as commented in the data section with respect to the high correlation between R&D expenditures and some of our focal variables, we turn now to analyse the robustness of our results to potential collinearity problems. In Table A5 of the online appendix, we observe that after eliminating R&D expenditures from our models, the results are virtually unchanged. The same is true when the related and unrelated variety variables are suppressed from the equations (columns (iv) and (v)). This corroborates that potential collinearity problems do not exert any influence in the obtained results on the impact of variety and external relatedness on regional knowledge production.

## **5. CONCLUSION**

This paper has investigated the role of variety on regional innovation, for a sample of 255 NUTS2 European regions of 25 countries, from 1999 to 2007. In particular, it has looked at the differential role played by various degrees of relatedness, across different spatial scales, on regional patenting and on citations-weighted regional patenting.

According to our results, diversity of knowledge, or variety, is critical for regional innovation. However, only knowledge flowing from different but related technologies (related variety) will generate new knowledge that incrementally constructs on established cognitive structures across related technologies – in line with the vast majority of previous studies. Notwithstanding these results, an interesting conclusion arises from our empirical approach when the patenting activity is weighted by the quality of such patents through the forward citations received – as an attempt to give more importance to breakthrough innovations. In this case, the more diversified across unrelated technologies is a region, the higher is the output in terms of high-quality innovations. Thus, evidence supports the idea that general innovation benefits from diversification in related technologies whereas more radical innovation also benefits from variety in unrelated technologies.

In addition, since knowledge can also be brought into a region from “outside”, we assess whether the degree of relatedness between incoming knowledge and the local knowledge base influences regional innovation performance. As it is usually done in the related literature, knowledge flows are proxied through the use of backward patent citations. Our results show that extra-regional incoming knowledge flows have a higher impact, the higher the similarity between these knowledge flows and the extant local knowledge base, which goes somewhat against our initial expectations. While this is true for the generation of average innovations, again differences emerge when accounting for the impact of the innovations produced: for radical innovations, the technological contents of the extra-regional linkages do not necessarily need to be very similar to the local technological base, but a certain degree of relatedness seems to be sufficient. This degree of relatedness assures certain cognitive proximity between agents located at a geographical distance,

while at the same time brings in the necessary variety to offer the building blocks for technological revolutions.

Regional diversification and relatedness are hot-button issues nowadays, not only for academics, but also for policymakers. These concepts have become especially relevant recently, as many European regions are still being hit by the economic crisis, which requires promoting new industries and economic activities (Boschma and Gianelle, 2013). These academic concepts go hand in hand with the Smart Specialization Strategy policy. Smart Specialization aims to focus policy support to key industries and economic activities already building in current national and regional strengths, thus avoiding to pick sectors that do not match the actual and potential technological capabilities of regions (Boschma and Gianelle, 2013). The concept of relatedness is thus the appropriate academic tool for smart specialization policies, advocating for the promotion of economic activities related, but different, to the actual technological structure of regions (McCann and Ortega-Argilés, 2013). Notably, our results on the positive effects of unrelated variety as well as the role of similar versus related knowledge inflows from outside the region have important policy implications in the framework of EU's smart specialization strategy, and must be accounted for.

Future research should thoroughly look at the effect of regional unrelated variety on breakthrough innovations. On the one hand, it could be interesting to analyse if breakthrough innovations – i.e., those at the upper-tail of the citations distribution – in a region actually combine technology classes that are unrelated, defined through co-occurrence analysis (see Boschma et al., 2015, as an example of this type of analysis), but present in the region concerned. On the other hand, it is plausible to think that the impact

of technological unrelated variety on the generation of breakthrough innovations can be stronger in the long run since the combination and recombination of previously unrelated technologies may imply some time to be fulfilled. Thus, it would be interesting to analyse the time profile of the impact of related and unrelated variety on the probability to produce breakthroughs.

### **Acknowledgements**

The authors would like to share their appreciation for the helpful comments received from Ron Boschma, Tom Broekel, Jackie Krafft, Francesco Quatraro and Pier Saviotti, as well as the participants of the Summer School “Knowledge Dynamics, Industrial Evolution, Economic Development” KID2015 in Nice and in the 3<sup>rd</sup> Geography of Innovation Conference in Toulouse. The authors also acknowledge financial support from the project “Redes de colaboración tecnológica e innovación. Determinantes y efectos sobre la competitividad de las empresas españolas” funded by the Fundación BBVA Ayudas a Proyectos de Investigación 2014. Ayudas Fundación BBVA. Ernest Miguelez acknowledges financial support from the Regional Council of Aquitaine (Chaire d’Accueil en Economie de l’Innovation, to Francesco Lissoni). Rosina Moreno acknowledges financial support provided by the Ministerio de Economía y Competitividad for the project entitled “Innovation and territory. Technological collaboration, international competitiveness and location factors”, ECO2014-59493-R, as well as to the Program ICREA Acadèmia. However, any mistakes or omissions remain the authors’ alone.

### **REFERENCES**

- Aghion, P., Howitt, P., 1992. A Model of Growth Through Creative Destruction. *Econometrica* 60, 323–351. doi:10.2307/2951599
- Alcacer, J., Gittelman, M., 2006. Patent citations as a measure of knowledge flows: The influence of examiner citations. *The Review of Economics and Statistics* 88, 774–779.
- Audretsch, D.B., Feldman, M.P., 2004. Knowledge spillovers and the geography of innovation. *Handbook of regional and urban economics* 4, 2713–2739.
- Bathelt, H., Malmberg, A., Maskell, P., 2004. Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. *Progress in human geography* 28, 31–56.
- Bergman, E.M., Maier, G., 2009. Network central: regional positioning for innovative advantage. *The Annals of Regional Science* 43, 615–644.
- Binz, C., Truffer, B., Coenen, L., 2014. Why space matters in technological innovation systems—Mapping global knowledge dynamics of membrane bioreactor technology. *Research Policy* 43, 138–155.
- Bishop, P., Gripaos, P., 2010. Spatial externalities, relatedness and sector employment growth in Great Britain. *Regional Studies* 44, 443–454.
- Boschma, R., 2016. Relatedness as driver of regional diversification: a research agenda. *Regional Studies* Published online, 1–14. doi:10.1080/00343404.2016.1254767
- Boschma, R., 2016. Relatedness as driver behind regional diversification: a research agenda. *Regional Studies*. Published online.
- Boschma, R., 2005. Proximity and innovation: a critical assessment. *Regional studies* 39, 61–74.



- Boschma, R., Balland, P.-A., Kogler, D.F., 2015. 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.
- Boschma, R., Eriksson, R., Lindgren, U., 2009. How does labour mobility affect the performance of plants? The importance of relatedness and geographical proximity. *Journal of Economic Geography* 9, 169–190.
- Boschma, R., Gianelle, C., 2013. Regional branching and smart specialization policy. JRC technical reports.
- Boschma, R., Heimeriks, G., Balland, P.-A., 2014. Scientific knowledge dynamics and relatedness in biotech cities. *Research Policy* 43, 107–114.
- Boschma, R., Iammarino, S., 2009. Related Variety, Trade Linkages, and Regional Growth in Italy. *Economic Geography* 85, 289–311. doi:10.1111/j.1944-8287.2009.01034.x
- 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., Minondo, A., Navarro, M., 2012. Related variety and regional growth in Spain\*. *Papers in Regional Science* 91, 241–256.
- Bottazzi, L., Peri, G., 2003. Innovation and spillovers in regions: Evidence from European patent data. *European economic review* 47, 687–710.
- Camagni, R., 1991. *Innovation networks*. John Wiley & Sons, Inc.
- Castaldi, C., Frenken, K., Los, B., 2015. Related variety, unrelated variety and technological breakthroughs: an analysis of US state-level patenting. *Regional Studies* 49, 767–781.

- Coe, N.M., Bunnell, T.G., 2003. "Spatializing" knowledge communities: towards a conceptualization of transnational innovation networks. *Global networks* 3, 437–456.
- Colombelli, A., 2016. The impact of local knowledge bases on the creation of innovative start-ups in Italy. *Small Business Economics* 1–14.
- Content, J., Frenken, K., 2016a. Related variety and economic development: a literature review. *European Planning Studies* 24, 2097–2112.
- Content, J., Frenken, K., 2016b. Related variety and economic development: a literature review. Utrecht University, Section of Economic Geography.
- Criscuolo, P., Verspagen, B., 2008. Does it matter where patent citations come from? Inventor vs. examiner citations in European patents. *Research Policy* 37, 1892–1908.
- David, P.A., 1993. Path-dependence and predictability in dynamic systems with local network externalities: a paradigm for historical economics. *Technology and the Wealth of Nations* 208–231.
- Dawley, S., 2014. Creating new paths? Offshore wind, policy activism, and peripheral region development. *Economic Geography* 90, 91–112.
- Desrochers, P., Leppälä, S., 2011. Opening up the "Jacobs Spillovers" black box: local diversity, creativity and the processes underlying new combinations. *Journal of Economic Geography* 11, 843–863.
- Essletzbichler, J., 2015. Relatedness, industrial branching and technological cohesion in US metropolitan areas. *Regional Studies* 49, 752–766.
- Fleming, L., 2001. Recombinant uncertainty in technological search. *Management science* 47, 117–132.

- 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*.
- Frenken, K., Van Oort, F., Verburg, T., 2007. Related variety, unrelated variety and regional economic growth. *Regional studies* 41, 685–697.
- Gertler, M.S., Levitte, Y.M., 2005. Local nodes in global networks: the geography of knowledge flows in biotechnology innovation. *Industry and Innovation* 12, 487–507.
- Gittelman, M., 2007. Does geography matter for science-based firms? Epistemic communities and the geography of research and patenting in biotechnology. *Organization Science* 18, 724–741.
- Glaeser, E.L., Kallal, H.D., Scheinkman, J.A., 1992. Andrei Shleifer. *The Journal of Political Economy* 100, 1126–1152.
- Grabher, G., 1993. The weakness of strong ties: the lock-in of regional development in the Ruhr area. *The embedded firm: On the socioeconomics of industrial networks*. London: Routledge 255–277.
- Griliches, Z., 1991. The search for R&D spillovers. *National Bureau of Economic Research*.
- Guo, Q., He, C., Li, D., 2015. Entrepreneurship in China: The role of localisation and urbanisation economies. *Urban Studies* 0042098015595598.
- Hall, B.H., Jaffe, A., Trajtenberg, M., 2005. Market value and patent citations. *RAND Journal of economics* 16–38.
- Harhoff, D., Narin, F., Scherer, F.M., Vopel, K., 1999. Citation frequency and the value of patented inventions. *Review of Economics and statistics* 81, 511–515.

- Hartog, M., Boschma, R., Sotarauta, M., 2012. The impact of related variety on regional employment growth in Finland 1993–2006: high-tech versus medium/low-tech. *Industry and Innovation* 19, 459–476.
- Hidalgo, C., Klinger, B., Barabasi, A.-L., Hausmann, R., 2007. The product space and its consequences for economic growth, in: *APS Meeting Abstracts*. p. 22006.
- Isaksen, A., 2015. Industrial development in thin regions: trapped in path extension? *Journal of Economic Geography* 15, 585–600.
- Jacobs, J., 1969. *The economy of cities*. Vintage, New York.
- Jaffe, A.B., 1989. Real effects of academic research. *The American Economic Review* 957–970.
- Jaffe, A.B., Trajtenberg, M., 1999. International Knowledge Flows: Evidence From Patent Citations. *Economics of Innovation and New Technology* 8, 105–136.  
doi:10.1080/10438599900000006
- Jaffe, A.B., Trajtenberg, M., Henderson, R., 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. *the Quarterly journal of Economics* 577–598.
- Jones, C.I., 1995. R & D-based models of economic growth. *Journal of political Economy* 759–784.
- Kogler, D.F., Rigby, D.L., Tucker, I., 2013. Mapping knowledge space and technological relatedness in US cities. *European Planning Studies* 21, 1374–1391.
- Lanjouw, J.O., Schankerman, M., 2004. Patent quality and research productivity: Measuring innovation with multiple indicators\*. *The Economic Journal* 114, 441–465.
- Maggioni, M.A., Uberti, T.E., 2009. Knowledge networks across Europe: which distance matters? *The Annals of Regional Science* 43, 691–720.

- Maraut, S., Dernis, H., Webb, C., Spiezia, V., Guellec, D., 2008. The OECD REGPAT Database (OECD Science, Technology and Industry Working Papers). Organisation for Economic Co-operation and Development, Paris.
- Marshall, A., 1920. Principles of economics: an introductory volume. London: Macmillan and Co., Ltd.
- McCann, P., Ortega-Argilés, R., 2013. Transforming European regional policy: a results-driven agenda and smart specialization. *Oxford Review of Economic Policy* 29, 405–431.
- Migueluez, E., Moreno, R., 2013. Do labour mobility and technological collaborations foster geographical knowledge diffusion? The case of European regions. *Growth and Change* 44, 321–354.
- Moreno, R., Paci, R., Usai, S., 2005. Spatial spillovers and innovation activity in European regions. *Environment and Planning A* 37, 1793–1812.
- Neffke, F., Hartog, M., Boschma, R., Henning, M., 2014. Agents of structural change. The role of firms and entrepreneurs in regional diversification. Utrecht University, Section of Economic Geography.
- 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.
- Owen-Smith, J., Powell, W.W., 2004. Knowledge networks as channels and conduits: The effects of spillovers in the Boston biotechnology community. *Organization science* 15, 5–21.
- Ponds, R., Van Oort, F., Frenken, K., 2010. Innovation, spillovers and university–industry collaboration: an extended knowledge production function approach. *Journal of Economic Geography* 10, 231–255.

- Popp, D., Hascic, I., Medhi, N., 2011. Technology and the diffusion of renewable energy. *Energy Economics* 33, 648–662.
- Quatraro, F., 2010. Knowledge coherence, variety and economic 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, 1922–1937.
- Rosenkopf, L., Almeida, P., 2003. Overcoming local search through alliances and mobility. *Management science* 49, 751–766.
- Saviotti, P.P., Frenken, K., 2008. Export variety and the economic performance of countries. *Journal of Evolutionary Economics* 18, 201–218.
- Schmoch, U., 2008. Concept of a technology classification for country comparisons. Final Report to the World Intellectual Property Office (WIPO), Karlsruhe: Fraunhofer ISI.
- Schoenmakers, W., Duysters, G., 2010. The technological origins of radical inventions. *Research Policy* 39, 1051–1059.
- Tanner, A.N., 2016. The emergence of new technology-based industries: the case of fuel cells and its technological relatedness to regional knowledge bases. *J Econ Geogr* 16, 611–635. doi:10.1093/jeg/lbv011
- Tanner, A.N., 2014. Regional branching reconsidered: emergence of the fuel cell industry in European regions. *Economic Geography* 90, 403–427.
- Tavassoli, S., Carbonara, N., 2014. The role of knowledge variety and intensity for regional innovation. *Small Business Economics* 43, 493–509.
- Trajtenberg, M., 1990. A penny for your quotes: patent citations and the value of innovations. *The Rand Journal of Economics* 172–187.

- Webb, C., Dernis, H., Harhoff, D., Hoisl, K., 2005. Analysing European and International Patent Citations: A Set of EPO Patent Database Building Blocks. OECD Science, Technology and Industry Working Papers, No. 2005/09, OECD Publishing, Paris.
- Weitzman, M.L., 1998. Recombinant growth. Quarterly journal of Economics 331–360.
- Wipo, 2015. World Intellectual Property Report 2015 - Breakthrough Innovation and Economic Growth (WIPO Economics & Statistics Series). World Intellectual Property Organization - Economics and Statistics Division.
- Xiao, J., Boschma, R., Andersson, M., 2016. Industrial diversification in Europe: The differentiated role of relatedness. Utrecht University, Section of Economic Geography.
- Zhou, K.Z., Li, C.B., 2012. How knowledge affects radical innovation: Knowledge base, market knowledge acquisition, and internal knowledge sharing. Strategic Management Journal 33, 1090–1102.

**Table 1. Summary statistics**

<b>Variable</b>	<b>Obs.</b>	<b>Mean</b>	<b>Std.Dev</b>	<b>Min.</b>	<b>Max.</b>
PAT pc	2,219	111.84	131.52	0	1,017.78
Weighted PAT pc	2,219	264.42	324.30	0	2,575.42
Variety	2,219	5.85	1.50	0	7.78
Related Variety	2,219	1.78	0.77	0	3.20
Unrelated Variety	2,219	1.96	0.35	0	2.31
Similarity	2,219	6.33	3.36	0	13.68
Relatedness	2,219	0.03	0.03	0	0.43
R&Dpc	2,219	0.40	0.41	0.00	2.88
HRST	2,219	14.12	4.73	3.90	34.40
ShareInd	2,219	19.21	6.74	5.21	38.55
High-tech Empl.	2,219	4.26	1.80	0.70	12.80
GDP pc	1,827	21,253.04	8,802.56	3,400	84,600

**Note:** Variables in this table are expressed without taking the logarithmic transformation.



**Table 2. Related/unrelated variety and regional innovation**

	(i) Patents pc	(ii) Patents pc	(iii) Patents pc	(iv) Quality- weighted	(v) Quality- weighted
ln(R&D pc)	0.223*** (0.0598)	0.163*** (0.0536)	0.172*** (0.0558)	0.144* (0.0770)	0.160** (0.0778)
HRST	0.0159** (0.00689)	0.00868 (0.00677)	0.00845 (0.00649)	0.00881 (0.00833)	0.00875 (0.00814)
Variety		0.105*** (0.0306)		0.160*** (0.0369)	
Rel. variety			0.229*** (0.0647)		0.267*** (0.0777)
Unrel. variety			0.0919 (0.0693)		0.227*** (0.0814)
ShareInd		0.0421*** (0.00925)	0.0434*** (0.00855)	0.0635*** (0.0114)	0.0630*** (0.0109)
High-tech Empl.		0.0322*** (0.0116)	0.0315*** (0.0115)	0.0405*** (0.0152)	0.0417*** (0.0154)
Constant	3.934*** (0.147)	2.316*** (0.282)	2.332*** (0.292)	2.443*** (0.369)	2.490*** (0.372)
Observations	2,219	2,219	2,219	2,219	2,219
Number of regions	255	255	255	255	255
Region FE	yes	yes	yes	yes	yes
Time FE	yes	yes	yes	yes	yes
Overall-R2	0.733	0.603	0.627	0.444	0.472

Notes: Robust standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.



**Table 3. Relatedness and external linkages**

	(i) Patents pc	(ii) Quality- weighted
ln(R&D pc)	0.140*** (0.0536)	0.128* (0.0764)
HRST	0.00560 (0.00584)	0.00553 (0.00741)
Rel. variety	0.183*** (0.0619)	0.218*** (0.0765)
Unrel. variety	0.0937 (0.0652)	0.230*** (0.0771)
Relatedness	0.489 (0.348)	0.941** (0.418)
Similarity	0.0705*** (0.0150)	0.0743*** (0.0182)
ShareInd	0.0383*** (0.00804)	0.0575*** (0.0108)
High-tech Empl.	0.0247** (0.0108)	0.0343** (0.0147)
Constant	2.075*** (0.270)	2.220*** (0.349)
Observations	2,219	2,219
Number of regions	255	255
Region FE	yes	yes
Time FE	yes	yes
Overall -R2	0.748	0.610

**Notes:** Robust standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

<sup>1</sup> This paper uses interchangeably breakthrough innovations, radical breakthroughs or radical innovations. They all try to convey the idea that not all inventions have the same technological and economic impact, and therefore this innovation quality heterogeneity needs to be taken into account. In the empirical part of this study, this heterogeneity is accounted for by weighting the number of patents produced in regions by the forward citations each of them receives.

<sup>2</sup> A complementary perspective is offered by the branching literature – after Hidalgo et al. (2007) at the country level, which looks at whether variety enhances regional diversification – that is to say, renewal and broadening of an economy’s industrial base (Xiao et al., 2016). Indeed, as Frenken and Boschma (2007) suggest, regions tend to diversify into economic activities related to the existing portfolio of local industries. Therefore, this idea of regional branching into related manufacturing industries is especially useful for understanding how new economic growth paths may be linked to preexisting industrial structures in a region (Tanner, 2014). Evidence on how regions diversify over time is now large too and include the case of Swedish regions (Neffke et al., 2011), Spanish regions (Boschma et al., 2013) and US metropolitan areas (Boschma et al., 2015; Essletzbichler, 2015; Kogler et al., 2013) yet showing at the same time that the process of technological transition is relatively slow (Rigby, 2015).

<sup>3</sup> Other studies have also looked at the role of variety on patents (Kogler et al., 2013; Rigby, 2015; Tanner, 2016), scientific publications (Boschma et al., 2014) or new firm formation (Guo et al., 2015; Colombelli, 2016).

---

<sup>4</sup> Recent case-study work has called attention to the relevance of external linkages for creating knowledge diversification. For instance, Binz et al. (2014) look at the membrane bioreactor technology and show that networks transcending national borders are of great importance for innovation processes – and therefore deserve more attention in theoretical and empirical work. More systematic evidence is presented in Neffke et al. (2014), who argue that the unrelated diversification needed for structural change is mostly created via non-local firms and entrepreneurs, according to the evidence they obtain using Swedish matched employer-employee data.

<sup>5</sup> Subclasses are further divided into groups and subgroups, so each IPC code can contain up to 10 digits.

<sup>6</sup> See the online appendix for the list of the 35 fields and the 5 broad fields.

<sup>7</sup> To compute this variable, we simply multiply the patents by the number of forward citations they received, and add up by region and year. In order to avoid eliminating a patent in case it has not received any forward citation, what we do is multiplying the number of patents by the number of citations plus 1, that is, Patents \* (Citations+1).

<sup>8</sup> We have experimented with alternative measures of human capital, such as the share of tertiary educated inhabitants (data from Eurostat), but the coefficient associated to this variable tends to be smaller and largely not significant. Results are available on request. This result confirms the intuition that only those more directly involved in knowledge and innovation activities are likely to determine the regions' capacity to innovate.

<sup>9</sup> In the empirical analyses, because of the existence of zero patents in some cases, a small constant, 1, is added before the logarithmic transformation.

<sup>10</sup> Tavassoli and Carbonara (2014) estimate a panel negative binomial model employing data for the 81 Swedish functional regions (local labour market) over the period 2002-2007 and provide robust evidence that related variety of knowledge plays a superior role than unrelated variety. Castaldi et al (2015), using patent data for US states in the period 1977–1999, provide evidence that innovation in general benefits from diversification in related technologies whereas states with higher unrelated variety would outperform states with lower unrelated variety in producing breakthrough innovations.

<sup>11</sup> Our empirical model (the regional knowledge production function) draws mainly from a large number of contributions in regional science and innovation economics trying to understand the role played by regional innovative efforts (R&D) and the technological structure of regions on regional innovative output. Other approaches have extended the regional KPF to include a large number of potential non-technology determinants of regional innovation outputs. We face a trade-off here, between accuracy of our empirical model (we want all the potential controls to be there) and completeness (we want to analyse a large number of regions and years). For instance, one interesting hypothesis to test would be the role of institutions and social capital on innovation, and more importantly, how they influence regional variety's role in fostering regional innovation (see Boschma, 2016, for a claim to do research in this direction). However, institutions and social capital variables are usually available for fewer regions, or at NUTS1 level, or for short periods of time (normally, they are not available on a yearly basis). Given that this is not the primary focus of our analysis, we have chosen to go for a large sample of regions and years to the expense of not adding these type of variables. Yet, contrary to still the large majority of empirical studies using the regional KPF, we control for region fixed-effects, and therefore account for all time-invariant features of regions that may influence the regional production of innovations (with institutions or social capital variables, which evolve slowly over time, being partially controlled for through these fixed effects).