Knowledge externalities and knowledge creation: the role of inventors' working relationships and mobility

Favaro, Donata and Ninka, Eniel and Turvani, Margherita

Department of Economics and Management, University of Padova, Department of Planning and Design in Complex Environment, University IUAV of Venice, Department of Planning and Design in Complex Environment, University IUAV of Venice

December 2014

Knowledge externalities and knowledge creation: the role of inventors' working relationships and mobility

Donata Favaro
(corresponding author)
Department of Economics and Management
University of Padova
Via del Santo, 33
35123 Padova
Email: donata.favaro@unipd.it
Fax: 0039 0498274211

Eniel Ninka
Department of Planning and Design in Complex Environment
University IUAV of Venice
Department of Management and Economics
European University of Tirana

Margherita Turvani
Department of Planning and Design in Complex Environment
University IUAV of Venice

Knowledge externalities and knowledge creation: the role of inventors' working relationships and mobility

Abstract

We study the transmission of tacit knowledge arising from working relationships established by inventors and its impact on firms' knowledge creation. First, we consider knowledge spillovers that originate through inventor working relationships that are not the result of collaboration agreements among patenting firms. Second, we analyse their effect on the creation of new knowledge as measured by companies' patenting activity. The study focuses on the role played by geographical proximity. The analysis was carried out on the population of firms located in the Italian region of Veneto and is based upon the original OECD REGPAT database that records all patenting applications at EPO.

Keywords: patenting activity, knowledge externalities, working relationships, mobility, geographical proximity

JEL codes: O3, J24, R1

1. Introduction

The literature on innovation generally states that the greatest source of new knowledge is R&D, along with other inputs, including human capital endowments, skilled labour and educational levels (Cohen and Klepper, 1992). However, this model of knowledge production was found to be stronger at broader levels of aggregation, such as countries or industries (Griliches, 1984), and the relationship appeared to be less robust at the disaggregated microeconomic level of the firm or its working units. In particular, the literature shows that there is no such direct relationship, at firm level, between inputs and innovation performance: in fact, while the largest firms undertake most industrial R&D, the innovative intensity of small firms exceeds that of large firms (Acs and Audretsch, 1988).

As clearly argued in the contribution by Audretsch and Feldman (2004), the dichotomy between the model of innovation on one side, holding at the macroeconomic level, and its failure at the firm level, motivated researchers to study the role played by firm externalities and knowledge spillovers in affecting innovation outputs from a firm-level perspective. A branch of literature investigating this puzzle highlighted the importance of the socio-economic context, as a factor that helps the sharing and transmission of knowledge, enhancing innovation and growth. Accordingly, innovation results from the interaction between human capital endowment and other factors, such as social capital, knowledge spillovers and networks of knowledge. These interactions create the social and structural conditions that affect the capacity to innovate: Lundvall's (1992) *systems of innovation* , Morgan's (1997) *learning regions* and Rodriguez-Pose's (2001) *social filter* are different ways of labelling the environment in which social interactions occur, laying the foundation for places to become particularly innovative territories.

Initial contributions to the analysis of knowledge spillovers concluded that knowledge externalities are spatially bounded: this was true for the branch of literature that studies the transmission of knowledge embedded in patent production, by observing patent citations. The seminal works by Jaffe and co-authors (Jaffe, 1986; Jaffe et al., 1993) laid the foundations of a literature that analyses the issue of knowledge diffusion using databases containing patent registrations and their subsequent citations. Indeed, they were the first to show that a spatial model of knowledge diffusion can be observed and that geography plays a role in the transmission of knowledge (these results referred to the USA). In recent years, Maurseth and Verspagen (2002) applied the same methodology as Jaffe and co-authors to the European regions, to investigate the determinants of knowledge flows within and between countries. They showed that technology flows across Europe are industry-specific and confined by geography, language and national borders.

These results laid the foundation for further studies discussing the role of proximity as a key for knowledge diffusion: Verspagen and Schoenmakers (2004), Fischer et al. (2006), LeSage et al. (2007) and Paci and Usai (2009). In particular, Verspagen and Schoenmakers (2004) showed that 'patent citations are located relatively near to each other', while Fischer et al. (2006) confirmed that technological proximity is an important factor for knowledge transfer: inter-regional knowledge flows occur most often between regions close to one another. LeSage et al. (2007), applying a spatial interaction model with spatially structured origin and destination effects to high-technology patent activity, found that knowledge spillovers are geographically localized. Paci and Usai (2009) confirmed these results, and they found that 'technological flows among firms and inventors are favored when they share the same language, culture, and institutional setting'. Rodriguez-Pose and Crescenzi (2008), on the other hand, found evidence of both intra- and inter-regional knowledge spillovers in the European area.[1]

The approach of Jaffe and his co-authors has been further developed in subsequent years through the study of the relationship between mobility of labour/human capital and innovation diffusion (Zucker et al., 1998; Almeida and Kogut, 1999; Agrawal et al., 2006; Breschi and Lissoni, 2009). Overall, this branch of literature emphasizes the role played by labour markets in the transmission of knowledge. Labour mobility is one of the main channels of knowledge diffusion; moreover, if

[1] A further aspect that has been studied using patents databases is the role played by co-inventorships, i.e. explicit collaborations between/among applicants. Breschi and Lissoni (2009), in their study of US inventors' patent applications to the European Patent Office, showed that, in some sectors, social chains of co-inventors are largely responsible for the localization of knowledge flows. These results are also found in Breschi and Lissoni (2005) for Italy. Maggioni et al. (2011) found that innovation systems can be industry specific.

innovation diffusion is localized it is because of the localization of labour mobility, especially human capital (Breschi and Lissoni, 2009). As a result, geography matters in the diffusion of innovation because human capital mobility follows a localized model.[2] The contribution of Agrawal et al. (2006) emphasizes that 'an important component of the knowledge associated with patented inventions may be held tacitly by skilled engineers'. Tacit knowledge and not only codified knowledge appears to be important: it belongs to researchers and engineers who take part in the patenting process, and can be transferred through their social relationships. Such knowledge can spread when the inventors interact with other people, in particular other inventors.

Recent works have focused on the effect of worker interactions/mobility on firm innovation and the overall economic performance, by looking at different perspectives, using alternative datasets and different empirical specifications. Simonen and McCann (2008, 2010), using a survey of Finnish high-technology firms, studied the effect of worker mobility on the probability of innovation, taking into account the geographical extension of mobility (same sub-region *versus* different sub-region). They showed that human capital mobility improves innovation performance if it occurs between different areas.[3] Boschma et al. (2009) recently addressed the issue of skill portfolios of mobile workers and their effect on firm economic performance. They showed how worker mobility affects firm economic performance depending on the mix of geographical proximity and competences.[4] Boschma et al. (2009) argue that 'the effects of labor mobility on firm performance depend on whether new employees are recruited from the same region or from other regions'. Eriksson (2011) confirmed these results.

[2] Agrawal et al. (2008) proposed an interesting development of the issue of proximity by distinguishing between spatial and social proximity. Using the same metropolitan statistical area and co-ethnicity as proxies for spatial and social proximity, respectively, they demonstrated that both types of proximity increase the probability of knowledge flows between individuals; however, spatial and social proximity are substitutes in their influence on knowledge flows. In addition, the marginal benefit of geographic proximity is greater for inventors who are not socially close.

[3] The geographical area of reference coincides with Finnish commuting areas and identifies Finland's local labour markets.

[4] Developing the idea of Boschma and Iammarino (2009), Boschma et al. (2009) studied Swedish firm performance at plant level – measured by growth in labour productivity between 2001 and 2003 – as a function of labour mobility, as measured by the number of highly skilled job movers. They split mobility into intra-regional and inter-regional mobility according to the local labour market (LLM) definition. Boschma et al. (2009) define LLMs according to a specific commuting-minimizing algorithm.

In this article, we follow this recent stream of literature and explore the role of inventors' working relationships and inventor mobility in mediating the sharing and transfer of tacit knowledge that, in the end, can have positive externalities and be responsible for the creation of new knowledge and a higher production of patents at the firm level. Our study's contribution to the literature on knowledge spillovers driven by workers' professional relationships is distinguished by the focus on inventors, who are the keepers of patented knowledge; a definition of mobility that allows for the transmission of any knowledge accumulated by inventors; and taking into account inventors' working relationships as possible channels of knowledge accumulation and transmission. Unlike co-inventorships and multi-firm collaborations, these relationships are not codified by any agreement between firms; they depend on the professional activity of inventors, of which the firms they work for may not be aware. Despite this, these relationships can have significant positive externalities at the firm level. Indeed, the patented output is the result of not only the number of inventors that work on the project but also the endowment of knowledge any inventor can provide. The hypothesis we make is that: the greater the number of distinct-firms' patenting projects the inventor participates in, the higher the knowledge capital he uses in any firm he works for, giving rise to knowledge spillovers and enhancing the creation of new knowledge. To the best of our knowledge, there are no other studies that measure these types of externalities and their effects on the production of new knowledge.

Thus, the main goal of this study is to evaluate whether the production of patents at the firm level is positively affected by the sharing and transfer of knowledge due to inventors' working relationships and inventor mobility. A second question then arises: is the transmission of knowledge localized, i.e. within the borders of each single territorial unit from which it originates? And further: is the temporal dynamic of workers' professional relationships relevant? In line with the related literature discussed above, we focus on the role played by geographical proximity and time in shaping the effects of the spillovers on patented innovation.

The analysis was carried out on the population of firms in Veneto, the Northeastern Italian region, and filing patents with the European Patent Office (EPO) in the pre-crisis period, 1998–2007. We chose Veneto for a number of reasons. First, it is one of the most dynamic regions in Italy; historically and up to the late 1990s, it has been characterized by its well-developed manufacturing industry composed of national and international companies. Veneto has been characterized by a good technological profile and innovativeness. In the past, it was largely concentrated in big firms, while nowadays, because of delocalization and deindustrialization processes that happened in the region along with all advanced economies, industrial development is dispersed in a wide number of small firms that are active in traditional and more technologically advanced sectors. Such dispersion shows firms clustering in specific geographical areas within the region; from the knowledge-creation perspective, the clustering seems to play a major role. Veneto is then an example of a particularly innovative territory, where economies of proximity and agglomeration have historically played a role, and intra-region labour mobility and local clustering is facilitated by cultural, linguistic and geographical conditions. It is therefore interesting to verify what Agrawal et al. (2006) pointed out: whether tacit knowledge held by researchers can be transferred beyond the borders of the place in which is generated, through their working relationships.

We maintain that Veneto represents an emblematic case and offers the possibility of investigating the role played by different forms of knowledge, in particular tacit knowledge, competencies and skills in supporting knowledge externalities and transmission, and to contribute to the international debate. The Veneto region is quite representative of a larger territorial context, the Italian Northeast, where firms appear to be innovative and characterized by frequent interactions. These characteristics of the local economies that also operate on much larger and global markets, are useful for studying the role of proximity in the transmission of knowledge. The questions are: given such a context, where interactions are significant, is tacit knowledge flowing among firms and having a positive impact on the creation of new knowledge? If a positive effect exists, is it localized or does the transmission of knowledge go beyond the borders of each single territorial unit from which it originates? Our study shows that knowledge held by inventors involved in patenting activities spreads beyond local boundaries, even in Veneto, where agglomeration and proximity still play an important role for the localization of production and service activities.

By limiting our study to this region, we were able to develop and document a precise procedure for cleaning the data that, in the near future, we may extend to the whole of Italy.

The paper is structured as follows. In Section 2 we describe the original dataset. Section 3 covers the empirical model and the constructed measures of inventor mobility and working relationships. The results are discussed in Section 4.

2. Data

We carried out the analysis on data from the OECD REGPAT database (December 2010 edition).[5] REGPAT is a regionalized database whose data have been linked to regional information at NUTS3 level using applicants' and inventors' addresses. REGPAT includes two main datasets: patent applications filed with the EPO in the period 1977–2007 and patent applications filed under the Patent Co-operation Treaty (PCT) at international phase, 1977–2008. They both contain data on applicants – individuals or firms that apply for a patent – and relative inventors – individuals who contributed to the invention. We chose to work only on the part of the database containing EPO applications because, although the number of patent applications filed under the PCT at the international level has substantially increased since 2005, compared to the number of applications filed to the EPO, the PCT archive is smaller than the EPO archive. For the Veneto region, in the period covered by REGPAT (1977–2008), 8,059 patent applications were filed under the EPO compared with 3,621 filed under the PCT. The difference in the size of the two archives could be because it is much more expensive to file under the PCT than under the EPO. For this reason, and because of the economic structure of

5 The OECD REGPAT database, December 2010, derives from two complementary sources of data: the European Patent Office's (EPO) Worldwide Statistical Patent Database (PATSTAT, September 2010) and the OECD patent database that relies on EPO's Epoline database (EPO Bibliographic Database and Abstracts – EBD), covering all publications up to November 2010. See Maraut et al. (2008) for a thorough description of the database and the regionalization procedure. These data have been used by a recent and growing body of literature on knowledge spillovers, inventors' mobility, etc. See among others Miguélez et al. (2010) and Hendy and Sissons (2011).

the context of interest, which is mainly made up of small- and medium-sized firms that pay particular attention to the costs of innovation, we preferred to use the EPO archive to be sure of including the largest possible number of firms and inventors innovating in the region. However, by choosing the EPO database, we excluded the patent applications filed with the Italian Patent Office. This choice allows us to deal with patents that, on average, are expected to have a higher commercial value, since applying to the EPO is more expensive and time-consuming than only applying to national patent offices (Hoekman et al., 2009).

REGPAT is a very rich database: every record contains information on each patent application filed by one or more applicants, resulting from the contribution of one or more inventors. Every single record can be linked to information on each applicant and inventor participating in the project. The variables include the EPO application number; the application identifier, i.e. a surrogate key identifying patent applications; the EPO patent publication number; and the priority year, i.e. the year of first filing. The priority year is the date closest to the actual date of invention; for this reason, we used it in our analysis instead of the application year. Further information is strictly related to inventors and applicants listed in each application. This information, together with patent data, allows us to identify contemporaneous and subsequent inventor collaborations, and to measure the variables of interest (discussed in the next section). In other fields, for every applicant and every inventor we gathered their identification codes[6], full names and addresses, and country and NUTS3 regions of residence.[7]

Despite the relevant improvement in data quality of recent REGPAT releases, the dataset presents serious problems for the identification of applicants and inventors. This issue, known in the literature as the ‘who is who’ problem (Trajtenberg et al., 2006), comes from two main kinds of error that affect the correct identification of persons and firms. The first type of problem comes from the erroneous or varied spelling of names of individuals, for example, Guiseppe instead of Giuseppe; Il’ya instead

6 Both identification codes are surrogate keys borrowed from the original PATSTAT database.

7 The latter piece of information is a result of the regionalization procedure carried out by OECD using the postcode or, in its absence, the town name identified in the address that appears on the first page of the patent document.

of Illya; Gian Carlo instead of Giancarlo; Jan-Douwe instead of Jan Douwe. The second type of error comes from writing the name of the applicant, usually a company, in various ways, for example, Glaxo, Glaxo Wellcome, GlaxoSmithKline, GSK. Additional problems arise in those cases where two different addresses are listed in relation to a single inventor. These cases need further investigation to decide whether there are two inventors with the same name or one inventor who has moved.

To solve these problems, we made considerable effort to clean the data and then to correctly identify the applicants and inventors through the unambiguous assignment of personal identification codes, addresses and municipality of residence and firm establishment. Appendix A gives details of the cleaning procedure.

After cleaning the data, the matrix we used for the study consisted of about 3500 inventors who, between 1998 and 2007, collaborated on over 2000 patent applicants in the Veneto region. In this period, the number of patent applications in the whole region exceeded 4700 applications.[8] At this stage, besides the original information available in the dataset, we were able to exactly identify any inventor and the applicants he collaborated with at any time, the correct inventor's address of residence in every year (street and city) and the right location of each applicant in every year (street and city). This allowed us to exactly establish inventors' patenting activity by year, applicant and city, and their new patenting activities at any time. Accordingly, going back to the applicant the inventors cooperate with, it was possible to measure the extension of each applicant's channels of knowledge transmission and sharing that occurred through inventors' working relationships. We discuss these measures in the following section.

3. The empirical model

[8] These applications are summed at the firm level by year. We ended up with around 3300 observations organized by firm and year.

3.1 The variables of interest

Our main research goal was to test whether inventors' working relationships facilitate the production, sharing and transfer of tacit knowledge among patent applicants, giving rise to applicant-level positive externalities that increase the production of patents. We studied two different measures of inventors' working relationships.[9] The first relates to applicant engagement of 'new' inventors, that is inventors who did not worked for the applicant before. This *mobility* of inventors can be the source of new knowledge for the applicant and can be the means by which knowledge is transferred, and the firm's ability to patent is increased (see the literature discussed in the Introduction). The second relates to the working relationships that inventors simultaneously have with different applicants, in the absence of patenting agreements between the applicants themselves. Our hypothesis is that these relationships, that we call *connections*, contribute to enrich the knowledge of the inventor and thus to increase the knowledge that the inventor provides to each applicant he works for. The number of inventors' *connections* for each applicant determines the extent of the externality of knowledge from which the applicant may benefit. All things being equal, we can imagine that two applicants with the same number of inventors but different numbers of inventors' *connections* may benefit from a different amount of externalities and thus have different patenting potential outputs.[10]

Let us now turn to a more detailed description of these variables. *Mobility* is designed to capture the transfer of knowledge a mobile inventor can be responsible for. This knowledge can be either strictly related to a specific patent the inventor developed in the past or more generic, i.e., related to the inventor's past research experience and the skills he accumulated in his professional history. Our goal is to catch the transfer of any type of knowledge inventors hold. For this reason, we define a mobile inventor towards applicant i at time t as an individual being already registered in the dataset in correspondence with any applicant and any patent filed in any year $t - x$ (with $x > 0$), and

[9] We do not take account of social relationships in general, although this is also an interesting topic where important results have been shown in the literature.

[10] We measure *mobility* and *connections* without counting those relationships between applicants who belong to the same group of companies, and controlling for cases of 'false' mobility due to changes of company type or name. See the Appendix and Laforgia and Lissoni (2009) for details.

participating at time t in the production of a patent of an applicant i he has not worked for before. Notice that, by construction, an inventor who collaborates with two different applicants at the same time t is not a mobile inventor. *Mobility* is then measured at firm i in year t as the sum of mobile inventors 'towards' firm i at time t.

A few measures of firm-to-firm labour mobility have already been proposed in some recent articles that study the relationship between labour mobility and knowledge flows. Agrawal et al. (2006) use data on patent production and citations that are similar to our data. In their study, *mobility* is detected when an inventor files a patent with a 'new' applicant (as in our case) and, afterwards, the 'old' applicant cites that patent. This measure can be used to identify the networks that inventors' mobility can give rise to. But, in our opinion, it does not capture the flow of knowledge that the mobile inventor brings to the 'new' applicant.[11] The measure of mobility we propose does not depend on any patent citation, and can be used to catch the transfer of any type of knowledge – specific or generic – that the inventor may be responsible for when he 'moves' to the 'new' applicant. Our measure is like the measures used in Simonen and McCann (2008, 2010) and Boschma et al. (2009), where the focus is on the potential effect of worker mobility to be detected at the destination firm. The difference between our study and these contributions relies on the data used and, consequently, on the types of worker under observation: in our case, only inventors.

The variable *connections* is designed to capture the extent of applicant-level knowledge externalities due to their inventors' current working activities. As previously explained, our hypothesis is that the greater the number of distinct-applicants' patenting projects the inventor participates in, the higher his knowledge and the potential externality for any applicant he works for. We measured *connections* in two different ways. For the first measure of *connections* of applicant i at time t, we added up the number of inventors who collaborate on any patent of applicant i at time t, and who participate, at the same time t, in patent applications of other applicants. We call this variable

[11] Moreover, Agrawal's measure of mobility may be endogenously determined: indeed, the inventor may move towards an applicant whose patents have already been cited by the 'original' applicant. Unfortunately, the authors do not clarify this point.

the *number of connected inventors* (of applicant *i* at time *t*). For the second measure of *connections*, we used the number of applicants to which applicant *i* is 'connected' at time *t* because of their inventors' working relationships. We call this measure the *number of applicant connections*. We normalized both measures by the number of inventors who took part in the considered patent, and excluded multiple counting. The measures also do not include those cases of inventors whose applicants co-participate in the same patent. This exclusion is motivated by the fact that we aim to capture the potential effect of the flows of knowledge that arise from the only inventor working relationships that are not related to co-operation agreements between the relative applicants.[12]

Actually, the two measures of *connections* differ in their meaning. The *number of connected inventors* of applicant *i* evaluates how many inventors who collaborate with it are part of research activities of other firms, and does not account for the extent of inventors' relationships. However, the *number of applicant connections* does account for the number of other applicants any inventor works for, and gives a more realistic measure of the potential sources of knowledge flows. This probably explains why, in the next section, the *number of applicant connections* has a more significant effect in our estimates.

Let's go through a simple example to better understand the difference between the two measures. Figure 1 shows the case of Applicant 1 that at time *t* files Patent P1 in collaboration with Inventors A, B, and C. At the same time *t*, Inventor B is collaborating with Applicant 2 on Patent P2 and Applicant 3 on Patent P3. According to the definition of the two variables, for Applicant 1, the *number of connected inventors* is equal to 1/3: one connected inventor (Inventor B) divided by the number of inventors working on Patent P1. However, the *number of applicant connections* is equal to 2/3: two applicants (Applicants 2 and 3) linked to Applicant 1 through Inventor B, divided by the number of inventors working on Patent P1.

[12] In a previous version of the paper, we proposed the empirical analysis both excluding and including this type of relationship. The results showed that when studying the effects of co-inventorships on patenting performance you should weigh the patent output by the applicant share in the patent, otherwise the estimated effect is negative.

Mobility and the two measures of *connections* (these, as alternatives) constitute our 'basic' explicative variables. Later in this section we go through the details of the geographical disaggregation of these variables and the dynamic specification the model. Before doing that, we devote a few words to the explanation of the dependent variables we use. We remind readers that our analysis aims to evaluate the effects that inventors' working relationships and inventor mobility can have, at the applicant level, in terms of a higher production of patents.

The simplest variable of interest suitable for measuring the patented activity of each firm/applicant *i* in any year *t* is the *sum of patents*: the total number of patents registered at the EPO by applicant *i* at time *t*. However, this measure of patenting activity may present some limits in representing the real patent output when the company has partnered with other companies to carry out the project of patenting. In these cases, the patent is the result of the effort of different entities and is shared by all of them, according to the quotas declared to the EPO. In these cases, a better measure of patented activity is given by the *weighted sum of patents* registered by firm *i* at time *t*, where each patent of firm *i* is weighted by the firm's participation share in the production of the patent. The *sum of patents* and the *weighted sum of patents* of each applicant *i* in year *t* are the two measures of patenting we use in the empirical analysis as dependent variables.

Table 1 summarizes the distribution of the *sum of patents*, *mobility*, and the two measures of *connections*. More than 20% of our observations record more than one patent application and almost 5% of cases registered four or more patents. Around 12% of observations record at least one mobile inventor, while *connections* are less frequent: whatever measure we use, they amount to 7% of the whole dataset. If we compare the distributions of the *number of connected inventors* and the *number of applicant connections*, however, we notice that the frequency of one connection is lower in the latter than in the former case; in contrast, the number of observations recorded for cases of two or more connected firms is higher than the number of cases of two or more connected inventors.

[Table 1]

Knowledge Externalities and Inventor Working relationships

According to the existing literature on the diffusion of knowledge discussed in the Introduction, firm proximity can play an important role in driving knowledge spillovers and externalities. To investigate the geographical model of knowledge diffusion, we separate the geographical extent of *connections* and *mobility* by looking at the place of origin and destination of every relationship. The idea is to evaluate whether the knowledge flow occurs according to the degree of geographical proximity between the origin and destination territory.

The territorial unit of reference we use is defined by the non-administrative territorial unit of the Local Labour System (LLS) as defined by the Italian Institute of Statistics in 1997 (a similar definition is used in Boschma et al. (2009) for Sweden). LLSs are constructed on commuter routes between home and work, as identified in the most recent population census. By construction, they are aggregations of municipalities that identify homogeneous labour markets and functional economic areas. As already argued in previous studies on the topic, LLSs are appropriate units for studying widespread urban areas, as they most closely correspond with economic and functional areas and local labour markets (Boschma et al., 2009).

We define *local*, *regional* and g*lobal* c*onnections* as follows: *local connections* are when the applicants involved are located within the same LLS (intra-LLS *connections*); *regional connections* are when inventors' working relationships connect applicants established in different LLSs of Veneto; *extra-regional connections* couple any applicant of Veneto to applicants established outside Veneto (either in Italy or a foreign territory). Similarly, we define *local mobility* as when inventors 'move' inside the same LLS; *regional mobility* as when inventors move between firms/applicants located in different LLSs of Veneto; and *extra-regional mobility* as when inventors 'move' to the observed applicant from firms located outside the region (either in Italy or a foreign territory).

The geographical specification of the variables of interest adds a further element to the analysis: the time needed for *connections* and *mobility* to affect patenting output. As shown in most recent contributions to the study of the effects of labour market relationships on the transfer of knowledge, relationships that occur within the same LLS may involve firms that are more 'related', in terms of production specialization and worker competencies, than firms belonging to different LLSs or different regions/countries. Thus, the transfer of knowledge may be faster when relationships are within the same LLS than between different LLSs or regions/countries. For this reason, we add the time dimension to the geographical specification and end up with a time–space specification for both *connections* and *mobility*.

Therefore, the different measures of *connections* are both evaluated at the same time (t) the patenting output is observed and estimated (variable *connections*) and at different time lags (variables *connections lag* $i - j$ *years*). When lagged, *connections* measures relationships that occurred in the past – in years $t - 1$,.., $t - i$, $t - j$ – and is meant to capture the lagged effect of knowledge diffusion. For simplicity, we grouped lagged *connections* at five-year intervals, after carrying out robustness checks for different time intervals and making sure that the dynamics were satisfied regardless of the chosen interval. The variable *mobility* already measures, by construction, the 'movement' of inventors between time t and any time $t - x$ (with $x > 0$), so, the temporal specification simply details the time interval (always at five-year intervals).

3.2 The econometric model

Our aim is to test empirically whether inventors' working relationships, measured by *connections* and *mobility*, are responsible for the creation and transmission of knowledge and, eventually, positively affect patenting. The econometric model is estimated for all firms that apply for a patent at any time t in the period 1998–2007, and the dependent variable is some measure of patenting activity, as explained in the previous sections.

Data on patent applications of firm i at any time t are typical count data. The clear discrete nature of these data and the preponderance of small values suggest that we can improve on least squares with a model that accounts for those characteristics, using the Poisson regression model.

A Poisson regression is a form of a generalized linear model where the response variable is modelled as having a Poisson distribution; random variables with non-negative integer values are modelled as Poisson distributions. A random variable Y is said to have a Poisson distribution with the parameter μ, $Y \approx P(\mu)$, if it takes integer values $y = 0, 1, 2, \ldots$ with the probability:

$$\Pr\{Y = y\} = \frac{e^{-\mu}\mu^{y}}{y!} \tag{1}$$

For $\mu > 0$, the mean and the variance of this distribution can be shown to be: $E(Y) = \mathrm{var}(Y) = \mu$. Since the mean is equal to the variance, any factor that affects one will also affect the other.

The Poisson regression model stipulates that a sample of *n* observations $y_1, y_2, \ldots, y_n$ can be treated as realizations of independent Poisson random variables, with $Y_i \approx P(\mu_i)$ and y_i taking integer values. A common transformation of the Poisson regression model is given by the log-linear Poisson model, where μ_i depends on a vector of explanatory variables x_i through a log-linear model such as:

$$\log(\mu_i) = x_i^{'}\beta \tag{2}$$

In which the regression coefficient β_j represents the expected change in the log of the mean per unit change in the predictor x_j. In other words, increasing x_j by one unit is associated with an increase of β_j in the log of the mean.

Exponentiating equation (2) we obtain a multiplicative model for the mean itself:

$$\mu_i = \exp\ \{x_i^{'}\beta\} \tag{3}$$

where the exponentiated regression coefficient $\exp\{\beta_j\}$ represents a multiplicative effect of the *j*-th predictor on the mean.

The problem with the Poisson regression model is that the assumption that the conditional mean and variance of Y, given X, are equal may be too strong. Inappropriate imposition of this restriction may produce spuriously small estimated standard errors. Moreover, the model is based on the assumption that events occur independently over time.

A way to correct these issues is to allow for unexplained randomness, by replacing equation (2) by the stochastic equation:

$$\log\left(\mu_i\right) = x_i^{'}\beta + \varepsilon_i \tag{4}$$

where the error term is assumed to be normally distributed.

Equation (4) represents a natural generalization of the Poisson regression model, where the error term can reflect a specification error such as unobserved omitted exogenous variables (Cameron and Trivedi, 1986), and allows for cross-firm heterogeneity.

The generalized Poisson regression model is very similar to the non-negative binomial model. In fact, the two models differ only for the distribution of the error term: the error is assumed to be distributed according to a normal density function for the generalized Poisson regression and according to a gamma distribution in the non-negative binomial case. Actually, the negative binomial model is a more general model than the generalized Poisson regression, because it allows for the variance to exceed the mean. However, under a specific assumption on a parameter of the gamma distribution, mean and variance converge and the non-negative binomial model becomes identical to the Poisson. In deciding on the model to use, we took account of the distribution of the dependent variables of interest (see Appendix A). It is clear that the distribution of the number of patents is not characterized by overdispersion. Indeed, its variance is lower than its mean. However, the variance becomes slightly larger than the mean when we derive the weighted sum of patents. However, as the difference between the variance and the mean was negligible, it was not appropriate to adopt the non-negative binomial model and we decided to use the generalized Poisson regression model.

Given the longitudinal dimension of our data, we estimated the model as the following:

$$\log\left(\mu_{it}\right) = cons + \alpha_i + \alpha_t + \beta * connections_{it} + \delta * mobility_{it} + \varepsilon_i \tag{5}$$

where each applicant *i* is observed in each year *t* of the period 1998–2007.

Given the panel dimension of the data, we included fixed effects at applicant level, α_i, in the equation. We measured these fixed effects by the stock of patents registered by each applicant in the decade before the period of interest, 1988–1997. When estimating the model for the weighted sum of patents, we measured the fixed effect by the weighted stock of patents (always registered in the period 1988–1997). The inclusion of the stock of patents/weighted stock of patents at applicant level also allows us to investigate the existence of persistence in patenting, although we did not investigate the sources of persistence and the mechanisms it comes from. According to the empirical evidence, pre-existing knowledge stocks are important for innovation (Roper and Hewitt-Dundas, 2008) and for patenting activities (Cefis and Orsenigo, 2001). Therefore, we expect the variable to have a positive and significant effect.

To complete the model specification, random effects at applicant level are also included to capture the role played by applicant's unobservables, together with a time trend, α_t.

Connections and *mobility* are the explanatory variables of interest already discussed in the previous section. For *connections*, we remind readers that the two different measures proposed and discussed – *number of connected inventors* and *number of applicant connections* – are used alternatively to one other in different regressions. As expected, the variable *number of applicant connections* showed a higher degree of significance. So, for reasons of space, we decided to report and discuss from here onwards only the results of the estimates where this variable is included.[13]

We carried out the initial empirical analysis using the specification without geographical disaggregation but including the lagged variables (*base model*). In a second step, we added the geographical specification for the explicative variables (local, regional, global) and estimated the *geographical model*. Estimation results for the *base model* are reported in Table 2; those concerning the *geographical model* are shown in Table 3. We estimated both models using the two dependent variables *sum of patents* and *weighted sum of patents*. Before estimating the different specifications, we carried out a correlation analysis to be sure that the explanatory variables were not causing multicollinearity problems.[14]

[13] We can provide tables of the estimation results obtained including the variable *number of connected inventors* on request.
[14] We can provide the correlation matrix on request.

4. Results

In this section we discuss the results of the estimates of the econometric model illustrated in the Section 3 that allows us to evaluate the effect of inventors' working relationships on the transmission and creation of knowledge. As previously stated, we only discuss the model in which inventors' connections are measured by the variable *number of applicant connections*. Results obtained using the variable *number of connected inventors* can be provided on request.[15]

We processed equation (5) to two different specifications: the *base model* and the *geographical model*. In the *geographical model*, explanatory variables are defined in their territorial extension. We ran both specifications using the two alternative dependent variables: the *sum of patents* and the *weighted sum of patents*. Table 2 lists the estimated coefficients, the marginal effects and the usual statistics.

[Table 2]

The general result that emerges from the different specifications of the base model is that there exists a growing trend in the creation of knowledge. Moreover, the stock of knowledge produced in the decade preceding the period of interest has a significant and positive impact on the creation of new knowledge: the higher the number of patents filed by the firm in the period 1988–1997, the higher the sum of patents produced yearly by the firm in the observed sample. This result indicates the existence of some degree of persistence in patenting activities at the firm level, confirming what previous empirical studies have shown (see for example Cefis and Orsenigo, 2001).

We now turn to the discussion of the results for the variables that measure inventor mobility and working relationships. The overall result is that both *mobility* and *connections* explain significantly

[15] We remind readers that, by construction, all measures of *connections* and *mobility* are net of relationships occurring between applicants that belong to the same business group and of 'false' mobility.

the creation of new knowledge, even in the simple *base model* where the territorial extension of the relationships is not separated. However, some caveats are necessary.

Inventor mobility enhances the creation of knowledge measured by patenting activity, but only when the period between the patent registered with the leaving applicant and that filed with the destination applicant is not greater than five years. After five years, inventor mobility does not have any significant knowledge enhancing effect at the destination firm. This result shows that mobility has knowledge productivity effects if the contribution of the incoming inventor materializes in the production of new patents in the short term. Thus, it suggests an evaluation criterion of the recruitment strategies of patenting firms. Indeed, this outcome shows that the ability of an inventor to increase the patenting activity of a firm is negatively correlated to the time lag between the moment in which the inventor filed the last patent with the origin company and the time he records the first patent with the destination company. If this period is longer than five years, the incoming inventor will not significantly contribute to the increase in the production of knowledge of the 'recruiting' firm. In short, a good recruitment is evident in the short term.

Now let us concentrate on the estimation results for inventors' patenting relationships measured by the variable *number of applicant connections*. The variable has significant and positive effects on the transfer and creation of knowledge: coefficients and marginal effects are, for the most significant, regardless of the time span that elapses between the year in which inventors' working relationships occurred and the year of observation of the company's patenting output. However, this effect appears to increase over time: relations that occurred several years earlier (11–15 years) have a much higher marginal effect on the production of patents than relations that occurred in recent years (1–5 years) or in the same year of observation of patenting. As we see in the discussion of the geographical model, this outcome is partly due to the correlation between the time and the spatial extent of the area of influence of inventors' working relations. However, some caution has to be taken in interpreting the results. In fact, the increasing effect detected as the time lag increases may also be linked to inventors' characteristics, in particular to their working experience. Connections that are more distant in time may be those of inventors with a longer working experience. Thus, the coefficient increases over time because it also captures the effect of the human capital that inventors accumulate over time.

To better summarize and understand the role played by the variables of the *base model* in explaining the production of patents, in Figure 2 we plot the line that interpolates the relationship between each variable with a statistically significant coefficient and the prediction of the dependent variable (predicted sum of patents). The graph clearly shows the positive relationship between the different variables of interest and the predicted sum of patents filed yearly at the EPO, confirming the importance of inventors' working relationships for the transfer of knowledge and the creation of new knowledge. The most significant effect is due to connections, especially in the long term.

We can move now to the discussion of the estimation results for the geographical model (Table 3).

[Table 3]

The geographical specification confirms the path dependency and the positive time trend of patenting: the coefficients of the stock of patents in the period 1988–1997, either weighted or not according to the used dependent variable, and the time trend are significant, positive and stable. In general, the estimated coefficients of the *geographical model*, when significant, are greater than those predicted in the *base model*. This suggests that, if we do not consider the geographical issue , we underestimate the productive effects of knowledge flows.

Regarding the measures of inventors' *mobility* and *connections*, the geographical model adds some interesting elements to the understanding of the mechanisms of the transfer of knowledge in the hands of inventors. For mobility, the results of the base model hold in the geographical specification: mobility significantly affects patenting only if the incoming inventor participates in some patenting output in the short term. However, the geographical specification clearly shows that inventor mobility affects the creation of knowledge at a very local level only if mobility occurs within the same LLS.

Connections, in contrast, have local, regional and extra-regional effects and time plays some role in shaping these effects. While local connections significantly affect the creation of knowledge in the very short term and in the medium to long term, regional connections only have an effect in the medium term and extra-regional connections in the long term. Thus, the geographical model seems to highlight the importance of time in determining the significant impact of working relationships that spatially go beyond the LLS.

Turning to the size of the knowledge transfer, we can see that regional and extra-regional connections have knowledge effects (marginal effects) sharply higher than the effects of local connections in the very short and short term. This seems to support the outcomes of recent empirical studies showing that human capital's working relationships improve knowledge creation and/or innovation if they occur between different areas (Simonen and McCann, 2008, 2010; Boschma et al., 2009; Eriksson, 2011).

We summarize the results of the geographical model in Figure 3. This shows the pairwise relationship between the predicted sum of patents and each variable with statistically significant effects. The figure clearly shows the positive relationship between the variables of interest and the predicted patenting activity, and how the spatial extent of inventors' relationships and time jointly play a role in explaining the transfer and creation of knowledge. Thus, the pattern of knowledge transfer is not unique. For inventors' mobility, knowledge externalities occur in the short term and within very narrow boundaries (within LLSs). Otherwise, inventors' *connections* are responsible for knowledge externalities that spread beyond the local system where knowledge is created. In the end, even in an area like Veneto, where companies – small to medium size – agglomerate and manufacturing districts still have some importance in the sharing of skills and competencies, knowledge externalities spread beyond local boundaries.

5. Conclusions

In this article, we study inventors' working relationships and inventors' mobility and their role in mediating the sharing and the transfer of tacit knowledge; we evaluate the effects of these externalities on the creation of new knowledge, which we measure by patenting activity. This work fits into the branch of literature that highlights the role played by the labour markets in the transmission of knowledge, in particular of knowledge tacitly held by individuals taking part in patenting processes.

The study adds some novelty to the literature. First, besides considering inventors' mobility, we also take into account a possible source of knowledge externality – inventors' working relationships (*connections*) – that has not been considered in any previous research. Unlike co-inventorships and multi-firm collaborations, the peculiarity of these relationships is that they are not codified by any formal agreement between firms. Inventors' working relationships only depend upon their professional activities, and firms they work for may not be aware of them. While inventors' mobility implies that inventors leave a firm to join a new one, working relationships among inventors highlight collaborations that occur simultaneously and through multiple firms. This simultaneity gives rise to knowledge externalities across firms and can positively affect the creation of knowledge. Second, we evaluate the impact of inventors' mobility and working relationships on the production of patents, e.g. the production of new knowledge, thus evaluating the productive effect of these externalities.

The article also explores the spatial extent and the dynamic pattern of the spread of knowledge and it contributes to the wide literature on knowledge externalities. To do so, we identified and measured inventors' mobility and working relationships (*connections*) considering the territorial dimension, i.e., the LLS location, of those firms involved by inventor's relationships/mobility. Thus, we can capture the local, regional and extra-regional extent of the knowledge externalities and we measure their short- and long-term effects.

We carried out the analysis on the population of firms located in Veneto that filed patents with the EPO, in the pre-crisis period 1998–2007. After cleaning the data, we measured the patenting activity and the variables of interest year by year at the firm level. We ended up with an unbalanced panel of 2018 applicants filing at least one patent with the EPO in the period 1998–2007. We estimated the patenting model using a Poisson specification and exploiting the panel dimension.

Our results confirm the role played by human capital in the transmission and creation of knowledge: specifically, inventors are responsible for positive externalities that benefit the companies they patent with. Such externalities, which arise and spread through labour relations and mobility, enhance the companies' capacity for patenting. However, in general, *connections* have a higher positive impact on patenting activity than *mobility*.

By focusing on the spatial extent of inventors' *mobility* and their *connections*, we have contributed to a better understanding of knowledge spillovers. In line with most of the literature on worker mobility and the transfer of knowledge, we found that the transfer of knowledge that occurs through inventors' mobility is localized and it significantly affects patenting only within the borders of the specific LLSs where it is taking place, and when the production of new knowledge occurs in the short term. However, we also obtained original results because knowledge externalities that occur through inventors' working relationships have local, regional and extra-regional effects. This study shows the existence of a complex pattern of knowledge relationships, where both very local and distant working relationships play a role in the transfer and creation of knowledge. Inventors' working relationships thus produce productivity effects, in terms of patenting activity, that are driven by factors related to the spatial extent of these relationships, in particular to the skill content of local and more extended relationships. Thus, not only local relationships, supported by physical proximity, but also distant relationships can be important for firm performance. We contend that distant relationships can be relevant for the transfer of those skills that are not similar to those existing in the knowledge base of the firm, but are complementary to them. Although we could not control for inventors' competences and skills, our findings support previous research findings that firm performances are affected by worker relationships, depending on the mix of geographical proximity and competences (Boschma et al., 2009; Eriksson, 2011).

These results are of particular significance in relation to the territorial context of our study, the Veneto region. The productive structure of Veneto is characterized by strong local relations, where face to face interactions are widespread and regularly occur, thus supporting a preliminary hypothesis of a localized knowledge diffusion, supported by territorial and physical proximity. However, our results emphasize that beyond the role of proximity in generating knowledge spillovers, working relationships among inventors are channelling knowledge diffusion at a much broader scale, both in term of geography and of knowledge creation dynamics.

Considering these results, we suggest that policies should promote all possible channels of knowledge exchanges, including innovation centres, thus freeing the potential of working relationships away from firms' reluctance and fears of knowledge dispossession.

References

Acs ZJ, Audretsch DB (1988) Innovation in large and small firms: an empirical analysis. *American Economic Review* 78: 678–690

Agrawal A, Cockburn I, McHale J (2006) Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships. *Journal of Economic Geography* 6: 571–579

Agrawal A, Kapur D, McHale J (2008) How do spatial and social proximity influence knowledge flows? Evidence from patent data. *Journal of Urban Economics* 64: 258–269

Almeida P, Kogut B (1999) Localization of knowledge and the mobility of engineers in regional networks. *Management Science* 45: 905–917

Audretsch DB, Feldman MP (2004) Knowledge spillovers and the geography of innovation. In: Henderson JV, Thisse JF (eds) *Handbook of Regional and Urban Economics*. Elsevier, Amsterdam

Boschma R, Eriksson R, Lindgreen 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, Iammarino S (2009) Related variety, trade linkages, and regional growth in Italy. *Economic Geography* 85: 289–311

Breschi S, Lissoni F (2005) Cross-firm inventors and social networks: localised knowledge spillovers revisited. *Annales d'Economie et de Statistique* 79/80: 189–209

Breschi S, Lissoni F (2009) Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *Journal of Economic Geography* 9: 439–468

Cameron C, Trivedi PK (1986) Econometric models based on count data: comparisons and applications of some estimators and tests. *Journal of Applied Econometrics* 1: 29–53

Cefis E, Orsenigo L (2001) The persistence of innovative activities. A cross-countries and cross sectors comparative analysis. *Research Policy* 30: 1139–1158

Cohen WM, Klepper S (1992) The tradeoff between firm size and diversity in the pursuit of technological progress. *Small Business Economics* 4(1): 1–14

Eriksson RH (2011) Localized spillovers and knowledge flows: how does proximity influence the performance of plants? *Economic Geography* 87: 127–152

Fischer MM, Scherngell T, Jansenberger E (2006) The geography of knowledge spillovers between high-technology firms in Europe: evidence from a spatial interaction modelling perspective. *Geographical Analysis* 38: 288–309

Griliches Z (1984) *R&D, Patents, and Productivity.* University of Chicago Press, Chicago

Hendy SC, Sissons CM (2011) Innovators, innovation and increasing returns to scale: Solving New Zealand's productivity paradox. *New Zealand Science Review* 68: 28–32

Hoekman J, Frenken K, Van Oort F (2009) The geography of collaborative knowledge production in Europe. *Annals of Regional Science* 43: 721–738

Jaffe AB (1986) Technological opportunity and spillovers of R&D: evidence from firms' patents, profits and market share. *American Economic Review* 76: 984–1001

Jaffe AB, Trajtenberg M, Henderson R (1993) Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics* 108: 577–598

Kim J, Lee SJ, Marschke G (2008) International knowledge flows: evidence from an inventor-firm matched data set. In: Freeman RB, Goroff DL (eds) *Science and Engineering Careers in the United States: an Analysis of Markets and Employment*. University of Chicago Press, Chicago

Laforgia F, Lissoni F (2009) What do you mean by 'Mobile'? Multi-applicant inventors in the European biotechnology. In: Malerba F, Vonortas NS (eds) *Innovation Networks in Industries*. Edward Elgar Publishing Ltd, Cheltenham

LeSage JP, Fischer MF, Scherngell T (2007) Knowledge spillovers across Europe: evidence from a Poisson spatial interaction model with spatial effects. *Papers in Regional Science* 86: 393–422

Lundvall BA (1992) *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*. Pinter Publishers, London

Maggioni MA, Uberti TE, Usai S (2011) Treating patents as relational data: knowledge transfers and spillovers across Italian provinces. *Industry and Innovation* 18: 39–67

Maraut S, Dernis H, Webb C, Spiezia V, Guellec D (2008) The OECD REGPAT database: a presentation. OECD Science, Technology and Industry WP 2008/2, OECD Publishing

Maurseth B, Verspagen B (2002) Knowledge spillovers in Europe: a patent citations analysis. *Scandinavian Journal of Economics* 104: 531–545

Miguélez E., Moreno R, Suriñach J (2010) Inventors on the move: tracing inventors' mobility and its spatial distribution. *Papers in Regional Science* 89: 251–274

Morgan K (1997) The learning region: institutions, innovation and regional renewal. *Regional Studies* 31: 491–503

Paci R, Usai S (2009) Knowledge flows across European regions. *Annals of Regional Science* 43: 669–690

Raffo J, Lhuillery S (2007) How to play the 'Names Game': patent retrieval comparing different heuristics. *Research Policy* 38: 1617–1627

Rodriguez-Pose A (2001) Is R&D investment in lagging areas of Europe worthwhile? Theory and empirical evidence. *Papers in Regional Science* 80: 275–295

Rodriguez-Pose A, Crescenzi R (2008) Research and development, spillovers, innovation systems, and the genesis of regional growth in Europe. *Regional Studies* 42: 51–67

Roper S, Hewitt-Dundas N (2008) Innovation persistence: Survey and case-study evidence. *Research Policy* 37: 149–162

Simonen J, McCann P (2008) Firm innovation: the influence of R&D cooperation and the geography of human capital inputs. *Journal of Urban Economics* 64: 146–154

Simonen J, McCann P (2010) Knowledge transfers and innovation: the role of labour markets and R&D co-operation between agents and institutions. *Papers in Regional Science* 89: 295–309

Trajtenberg M, Shiff G, Melamed R (2006) The "names game": harnessing inventors' patent data for economic research. *National Bureau of Economic Research*, WP 12479

Verspagen B, Schoenmakers PL (2004) The spatial dimension of patenting by multinational firms in Europe. *Journal of Economic Geography* 4: 23–42

Zucker LG, Darby MR, Armstrong J (1998) Geographically localised knowledge: spillovers or markets? *Economic Inquiry* 36: 65–86

Tables

Table 1. Number of observations by number of patents, inventor *mobility* and *connections*

	Observations (%)		Observations (%)
Number of patents		Number of connected inventors	
1	2,587 (78.8)	0	3,055 (93)
>1	696 (21.2)	1	175 (5.3)
Of which		2	39 (1.2)
2	415 (12.6)	3+	14 (0.4)
3	129 (3.9)		
4	65 (2)		
5+	87 (2.7)		
Mobility		Number of applicant connections	
0	2,900 (88.3)	0	3,055 (93)
1	340 (10.4)	1	153 (4.7)
2+	43 (1.3)	2	49 (1.5)
		3+	26 (0.8)
Total	3,283 (100)	Total	3,283 (100)

Table 2. Base model. Estimation results

	Dependent variable			
	Sum of $patents_{it}$		Weighted sum of $patents_{it}$	
	Coef.	Marg. effect	Coef.	Marg. Effect
Time trend	0.02***	0.03***	0.02***	0.03***
	(0.01)	(0.01)	(0.01)	(0.01)
Stock of patents (1988–1997)	0.02***	0.02***		
	(0.00)	(0.00)		
Stock of weighted number of patents (1988–1997)			0.02***	0.02***
			(0.00)	(0.00)
Mobility				
Lag 1–5 years	0.09**	0.13**	0.09*	0.11*
	(0.05)	(0.06)	(0.05)	(0.06)
Lag 6–10 years	0.03	0.03	0.05	0.06
	(0.09)	(0.12)	(0.09)	(0.12)
Lag 11–15 years	0.04	0.05	0.07	0.09
	(0.13)	(0.17)	(0.13)	(0.16)
Lag 16–20 years	−-0.24	−0.32	−0.23	−0.28
	(0.26)	(0.35)	(0.27)	(0.33)
Connections (*number of applicant connections*)				
	0.11***	0.15***	0.07*	0.09*
	(0.04)	(0.05)	(0.04)	(0.05)
Lagged connections (*number of applicant connections*)				
Lag 1–5 years	0.18**	0.25**	0.20**	0.25**
	(0.08)	(0.11)	(0.09)	(0.11)
Lag 6–10 years	0.16	0.21	0.15	0.19
	(0.11)	(0.14)	(0.11)	(0.14)
Lag 11–15 years	0.32**	0.44**	0.33**	0.42**
	(0.14)	(0.19)	(0.15)	(0.18)
Lag 16–20 years	0.27	0.36	0.22	0.28
	(0.17)	(0.23)	(0.18)	(0.23)
Constant	−37.26***		−43.45***	
	(10.85)		(11.22)	
Ln(μ)	−2.54***		−2.33***	
	(0.09)		(0.09)	
Observations	3,283		3,283	
Number of applicants	2,018		2,018	
Wald test Chi^2	120.75***		110.56***	

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

Note: The variable *number of applicant connections* is normalized by the number of inventors by firm and year.

Table 3. Geographical model. Estimation results

		Dependent variable			
		Sum of patents$_{it}$		Weighted sum of patents$_{it}$	
		Coeff.	Marg. effect	Coeff.	Marg. effect
Time trend		0.02***	0.03***	0.02***	0.03***
Stock of patents (1988–1997)		0.02***	0.02***		
Stock of weighted number of patents (1988–1997)				0.02***	0.02***
Mobility					
Local mobility	*Lag 1–5 years*	0.12*	0.16*	0.12*	0.15*
	6–10 years	−0.03	−0.04	−0.01	−0.01
	11–15 years	0.04	0.05	0.07	0.08
	16–20 years	−0.21	−0.28	−0.21	−0.26
Regional mobility	*Lag 1–5 years*	0.04	0.05	0.04	0.05
	6–10 years	0.08	0.11	0.08	0.11
	11–15 years	−0.01	−0.01	0.02	0.03
	16–20 years	−0.28	−0.33	−0.22	−0.25
Extra-regional mobility	*Lag 1–5 years*	0.14	0.19	0.12	0.15
	6–10 years	0.09	0.12	0.12	0.16
	11–15 years	0.10	0.14	0.13	0.18
	16–20 years	−0.35	−0.39	−0.32	−0.34
Connections (*number of applicant connections*)					
Local connections		0.12**	0.16**	0.08	0.10
Regional connections		0.12	0.16	0.07	0.09
Extra-regional connections		0.00	0.01	−0.02	−0.02
Lagged connections (*number of applicant connections*)					
Local connections	*Lag 1–5 years*	0.18*	0.24*	0.19*	0.24*
	6–10 years	0.17	0.23	0.18	0.23
	11–15 years	0.71***	0.96***	0.72***	0.90***
	16–20 years	0.42	0.56	0.48	0.60
Regional connections	*Lag 1–5 years*	0.47*	0.62*	0.50**	0.63**
	6–10 years	−0.21	−0.28	−0.35	−0.44
	11–15 years	−1.00	−1.34	−0.94	−1.18
	16–20 years	0.06	0.08	−0.14	−0.17
Extra-regional connections	*Lag 1–5 years*	−0.08	−0.11	−0.03	−0.04
	6–10 years	0.30	0.40	0.30	0.38
	11–15 years	0.12	0.16	0.16	0.20
	16–20 years	0.43*	0.57*	0.44*	0.56*
Constant		−37.19***		−43.47***	
Ln(μ)		−2.57***		−2.36***	
Observations		3,283		3,283	
Number of applicants		2,018		2,018	
Wald test Chi^2		139.05***		129.41***	

For reasons of space, standard errors are not reported. *** p<0.01, ** p<0.05, * p<0.1.

Note: The variable *number of applicant connections* is normalized by the number of inventors by firm and year.

Figures

Figure 1. An example of inventor–applicant working relationships

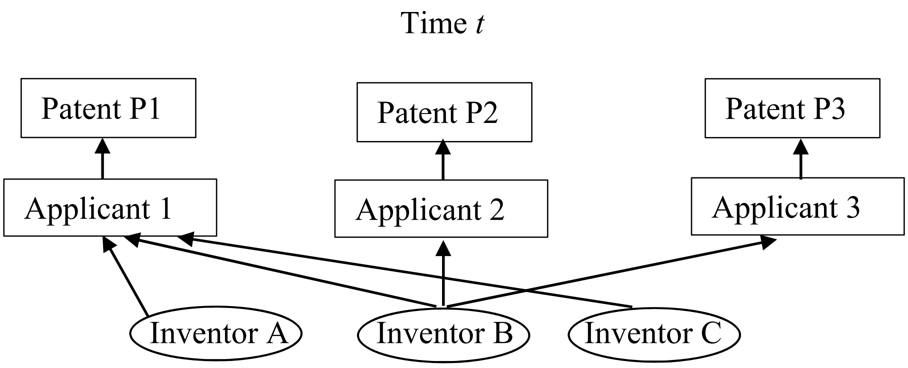
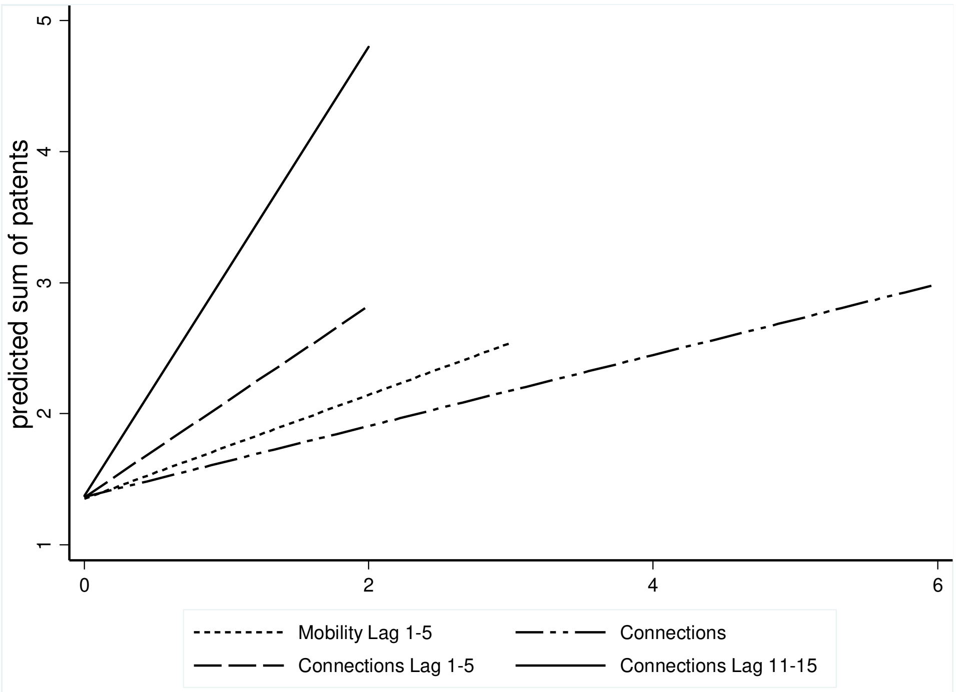
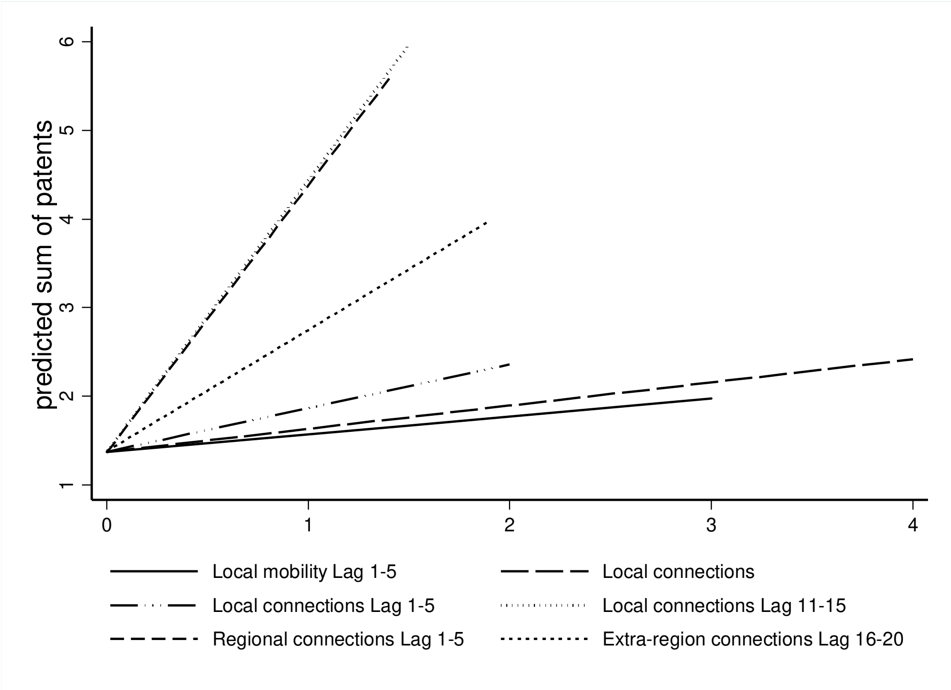


Figure 2. Base model. Linear interpolation of the *predicted sum of patents* in relation to *mobility* and *connections*.



Note: only for statistically significant covariates

Figure 3. Geographical model. Linear interpolation of the *predicted sum of patents* in relation to *mobility* and *connections*.



Note: only for statistically significant covariates

Appendix

A Descriptive statistics

	Mean	St. Dev.
Number of patents	1.442	1.415
Weighted number of patents	1.365	1.434
Stock of patents 1988–1997	1.886	8.832
Stock of weighted sum of patents 1988–1997	1.863	8.738
Mobility		
Lag 1–5 years	0.083	0.308
Lag 6 to10 years	0.029	0.173
Lag 11 to 15 years	0.013	0.118
Lag 16 to 20 years	0.004	0.070
Connections		
Number of connected inventors#	0.050	0.207
Number of applicant connections#	0.066	0.337
Lagged connections		
Number of connected inventors#		
Lag 1–5 years	0.020	0.109
Lag 6–10 years	0.011	0.085
Lag 11–15 years	0.006	0.059
Lag 16–20 years	0.004	0.055
Number of applicant connections#		
Lag 1–5 years	0.030	0.171
Lag 6–10 years	0.016	0.131
Lag 11–15 years	0.007	0.083
Lag 16–20 years	0.006	0.081

Normalized by the number of inventors by firm and year.

B The data 'cleaning' and name-matching procedure

In this section, after describing briefly the main features of the REGPAT database, we discuss the procedure we adopted to identify applicants and inventors. In REGPAT patent applications, applicants and inventors are identified by a series of surrogate keys 'borrowed' from PATSTAT.

To be able to observe inventor *mobility* and working *connections*, we needed to correctly identify applicants and inventors. The correct applicant/inventor identification procedure relies on the following three fields: name, address and country. The standardization of the names was the first step of our name-matching procedure. The fields of inventors' and applicants' names include everything in the application. Since no further standardization is carried out by either PATSTAT or REGPAT, even tiny differences like the number of spaces, commas or the use of capital letters will cause the same individual (or firm) to be treated as two distinct individuals (firms). Other sources of differences related to patent data include character variation (Baú versus Bau', Cado' versus Cado); capitalization (ROSSI versus Rossi or rossi); punctuation (Aldino Colbachini versus Aldino-Colbachini); spacing (Gianluigi versus Gian Luigi); qualifiers (Mario Rossi versus Prof. Dr. Mario Rossi); and spelling variation, including insertion (Mario versus Marrio), omission (Giannantonio versus Gianantonio), substitution (Illya versus Illia), or transposition (Giuseppe versus Guiseppe). Sources of differences might also refer to spelling variations (Pietro in Italian would be spelled Peter in English) or to the use of initials (Rossi Mario versus Rossi M.).

The address field contains all address elements for the applicant/inventor: street name and number, city, postal code. Even here, small differences in the order of appearance – street, number or number, street – or extra spaces, commas, etc. will mean that the same address is considered a distinct address because they are actually different strings.

REGPAT contains two additional fields that identify the geographic location of the inventor/applicant. The first field contains the country code, that is the ISO 2 code of the applicant's or inventor's country of residence (for example 'IT' for Italy); the second includes a regional code, which is mainly the added value of this database.[16] REGPAT has adopted the Eurostat NUTS3 regional breakdown for European countries and OECD's Territorial Level (TL3) for other countries. There are seven NUTS3 regions in Veneto that correspond to the Italian provinces. Given our regional focus, we relied on this code to select applicants located in Veneto.

As mentioned above, REGPAT applicants and inventors are identified by means of a personal identification code. This is a number given to every combination of applicant/inventor name, address and country code. Consequently, every combination of these three fields is considered unique and corresponds to a different person identifier (person ID). In the identification procedure of applicants and inventors, OECD recovers PATSTAT's surrogate person ID codes without making any additional verification or change. Unfortunately, this results in different numerical codes, quite often sequential, assigned to the same person. Therefore, any quantitative analysis on patent data needs a correct attribution of personal ID codes to both applicants and inventors. To this end, and given the dimensions of our dataset (around 14,500 records for inventors and around 8,400 records for applicants), we opted for a semi-manual procedure of data processing.[17] Although human error cannot be excluded with absolute certainty, the advantage of this approach lies in the fact that it allows for a thorough verification of the allocation of ID codes.

We treated the two subsets of data (applicants and inventors) separately and merged the information only after the completion of the data cleaning procedure. Though the procedure described below was essentially applied to both datasets, in what follows we mostly refer to the cleaning and name-matching procedure for the inventors' subset, because this set contained most of the misattributed ID codes.

[16] See Maraut et al. (2008) for the regionalization methodology.

[17] In recent years, a very promising body of research on patent data has proposed a growing number of algorithms which attempt to correctly identify inventors, starting from information on their name, address and characteristics of the invention. After making some basic assumptions, adopting ad hoc heuristics and a threshold of error, such algorithms will assign, with a degree of success, single identification codes to strings of names that are slightly different, but that reasonably belong to the same inventor. Examples of such algorithms are described in Kim et al. (2008); Miguélez et al. (2010); Raffo and Lhuillery (2007); and Trajtenberg et al. (2006).

1. As a first step, we wrote the field with the inventor's surname and name in capital letters and put them in the correct order: SURNAME, NAME.[18]
2. We dropped punctuation symbols – apostrophes, hyphens, pipes, commas, periods, dashes, numbers, qualifiers or affiliations attached to the name.
3. We trimmed extra spaces between words and corrected those names containing extra symbols like ó, ú, é.
4. We corrected the cases of misspelled names.
5. We wrote the firm's name correctly and verified for all cases where a firm changed its name after a merger or an acquisition such as, Glaxo/Wellcome/SmithKline which we assigned the generic name GLAXO SpA.

After carrying out the data cleaning procedure, we verified cases in which the same person apparently had more than one ID code. In all these cases, besides the names (which at this point matched perfectly), we controlled for the address, the town, the applicant and the priority year. Among the most frequent cases we had to deal with in the matching procedure of inventors' names were:

- same inventor name and exactly the same address;
- same inventor name and slightly different addresses: different street numbers, which may be attributable to a firm that has more than one building in the same street, or a new allocation of the street numbers due to an expansion of the residential/commercial area;
- same inventor name, different addresses, but the same applicant. Very often the inventors in the process of patent filing declare the address of their employer rather than their personal address;
- same inventor name, different street addresses in the same or in a nearby town, different patent priority years;
- same inventor name, different addresses, though in one case the inventor's address coincided with the applicant's address with whom the inventor had registered a patent in the past;
- same inventor name, different addresses, different applicants, different patent priority years.

[18] An easy task given the size of the dataset and the knowledge of Italian names and surnames.

The last one is a special case. If applicants (companies) are part of the same group or are connected somehow to each other (a change in the applicant's name; a change of company's legal status, for example, from Srl -Ltd to SpA -Joint Stock Co.; two legally different firms owned by the same person; two firms sharing exactly the same address) we may conclude that the two inventors are the same person. Otherwise, we are dealing with a dubious case that deserves further investigation as described below.

In the above-mentioned cases, we are dealing, in all probability, with the same person who might have changed his street address, the city of residence or firm. In some cases, it was not easy to understand whether it was the same person or different people with similar names. Therefore, the assignment of the ID code needed further investigation. In such dubious cases we checked whether the inventor had the same co-inventor(s) in the past or whether the IPC codes were similar – although the patents were filed by different applicants. Moreover, we carried out an internet search, using the websites of the Italian Chambers of Commerce to check whether a company had changed address, name or legal status (for example: from Srl to SpA). Furthermore, in some cases, we checked the inventor's curriculum vitae, if available online (including personal LinkedIn profiles). Finally, we assigned new codes to single applicants and inventors with different ID codes.

For to the applicants' subset, we generated two new fields containing the name of the municipality of establishment and the correct postcode. We extracted the information from each applicant's address (that in REGPAT is a not standardized text string), where we found many transcription errors that could affect the attribution of regional codes. First we standardized identical addresses that were written in different ways (for example, reporting a generic postcode in cities where there are specific postcodes; writing the street number before the street name[19]; having a place name instead of the town name). Second, we corrected every postcode, by accounting for any change in municipal boundaries, and changes made by the Italian Post Office (*Poste Italiane*). Finally, we extracted the correct

[19] The correct way of reporting the address in Italy is: street name, street number, city, postcode and provincial code – 2 capital letters in brackets.

information on the municipality and the postcode. The last step was essential in order to assign each applicant to the appropriate LLS.

Aknowledgments

We would like to thank Steven Brakman, Riccardo Crescenzi, Camilla Lenzi, Marc-Andreas Muendler, Peter Nijkamp, Roberto Patuelli, Takatoshi Tabuchi, Stefano Usai and the participants to the annual conferences of ERSA and AISRE, to the seminar hold at the Department of Economics-University of Tokyo and to the IAB conference "The Spatial Dimension of the Labour Market: the Relocation of Jobs", for helpful comments and suggestions.

The authors acknowledge funding support from CO.RI.LA (Consorzio per la Gestione del Centro di Coordinamento delle Attività di Ricerca inerenti il Sistema Lagunare di Venezia) on behalf of Venice Water Authoriry (Magistrato alle acque). Research programme: 'La Laguna di Venezia nel quadro dei cambiamenti climatici, delle misure di mitigazione e adattamento e dell'evoluzione degli usi del terriorio'.