Productivity in innovation: the role of inventor connections and mobility

Favaro, Donata and Ninka, Eniel and Turvani, Margherita

20 May 2012

Productivity in Innovation: The Role of Inventor Connections and Mobility[♣]

Donata FAVARO[*]
Department of Economics M. Fanno, University of Padova, Italy
Affiliations: CHILD, Torino, Italy.; TESS, Venice, Italy.

Eniel NINKA
University IUAV and TESS, Venice, Italy

Margherita TURVANI
University IUAV and TESS, Venice, Italy

Abstract

We study the transmission of knowledge arising from working relationships established by inventors, and its impact on firms' innovation production. The study's contribution to the literature is twofold. First, we consider those relationships that originate through inventor connections ("multi-applicant" inventors) and inventor mobility. Second, we analyse their effect on companies' innovation production. The study focuses on the role played by geographical proximity, and the dynamic effects of knowledge flows. The geographical question is dealt with on a detailed level, by measuring knowledge spillovers observed within the same Local Labour System (LLS), between different LLSs of the region and, finally, with extra-regional LLSs. Dynamics are captured by measuring inventor mobility and connections occurring up to 20 years before patent filing.

The analysis is carried out on the Italian region of Veneto and is based upon the original OECD REGPAT database of patent applications filed at the European Patent Office. The manual procedure we used to clean the data allows us to resolve some issues raised in the literature. Our results show that the impact of working relationships on innovation production depends on both geography and dynamics. Therefore, we can not conclude that productivity effects of knowledge flows occurring through the labour market are localized. However, we can conclude that working relationships have sizable productivity effects on innovation, either in the short or in the long run, depending on the geographical distance.

Keywords: labour mobility, inventor connections, knowledge diffusion, innovation, geographical proximity

JEL codes: O3, J24, J61, R12

[♣] We wish to thank Takatoshi Tabuchi and the participants to the Regional Economics Workshop at Tokyo University and to the ERSA Conference (Barcelona, 2011) for their comments and suggestions. This study was supported by CO.RI.LA, Venice, www.corila.it.

[*] Corresponding author: Favaro Donata, Department of Economics M. Fanno, via del Santo 33, 35123 Padova, Italy, donata.favaro@unipd.it.

1. Introduction

Theoretical analysis and empirical research have demonstrated the fundamental role played by human capital in the innovation sector. This is true not only in its individual character – in the individual contribution of researchers and inventors to knowledge – but also in its collective and dynamic dimension. Interpersonal relationships ease the diffusion and transfer of knowledge by affecting the innovative capacity of individuals, companies or systems (Lundvall, 1992; Morgan, 1997; Rodriguez-Pose, 2001). Knowledge can be enhanced through human capital networks; some of these are observable and derive from formal agreements, while others are less explicit and originate informally. In the first case, knowledge is shared and enhanced through explicit forms of collaboration between researchers and/or companies; in the second case, knowledge is spread mainly through the mobility of workers in the labour market, their cooperation with different companies at the same time, and non-work social relations.

In this article we deal with informal channels of knowledge transmission that arise from working relationships established by inventors, and study their impact on firm innovation production. The study's contribution to the literature is twofold. First, we consider those relationships not formally settled in co-participations in the same patent (co-inventorships)[1]. These relationships, although informal, can be responsible for the creation of networks of knowledge sharing and transfer (Almeida and Kogut, 1999; Power and Lundmark, 2004). Specifically, we consider spillovers that are born through multi-firm inventor collaborations and inventor mobility, and analyze separately their innovation effects, detailing their dynamic and geographical effects. The idea behind our analysis is that inventors accumulate knowledge when participating to the production of a patent, and this knowledge can flow to other firms with which inventors are collaborating with at the same time (*inventor*

[1] However, the empirical analysis is carried out with an additional definition of the variable where co-inventorships are included in the measure of inventor-firm connections. This allows us to show that, at least for our region of interest, co-inventorships indeed are not a means for increasing the productivity of innovation.

connections). A fortiori, the flow of knowledge occurs when inventors move from one firm to another (*inventor mobility*). Our perspective is different from that used in the literature that studies the flow of knowledge through patent citations and labour mobility. In that case, the diffusion of knowledge occurs when a patent is cited. In our case, knowledge flows when the inventor collaborates to the production of patents filed by different firms. Moreover, unlike the literature that studies co-inventorships, our first interest is on inter-firm connections established through inventors working on patents registered by different applicants. Second, we study the productive effects of these flows of knowledge, which means their effects in terms of innovation production. The question we try to answer is whether multi-applicant inventor work relationships and mobility give rise to knowledge flows that positively affect the production of innovation.

A further contribution of this article is given by the analysis of the geographical pattern of the above mentioned relationships and their dynamic effects. The territorial unit of reference is defined by the smallest non-administrative territorial unit at our disposal: the Local Labour System (LLS) as defined by Istat in 1997. 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, these are appropriate units to use for this type of study of widespread urban areas, as they most closely correspond with economic and functional areas and local labour markets (Boschma et al., 2009). Then, to verify the localization of knowledge sharing and transfer, we disentangle among intra-LLS, inter-LLS and extra-regional interactions. The underlying idea is to evaluate whether the effect of the flow of knowledge on productivity – that occurs through the labour market – is different according to the degree of geographical proximity of origin and destination of worker mobility and of the poles of inter-firm inventor connections.

However, time also plays an important role in determining the productive effect of knowledge flows and can interact with proximity/distance. Then, we measure all relationships (by geographical proximity) occurring in the 20 years before the sample interval (see Section 4 for a detailed description).

The analysis is carried out on the population of firms located in Veneto, a North-eastern Italian region, and filing patents with the European Patent Office (EPO) in the pre-crisis period, 1998-2007. The analysis was restricted to the region of Veneto, for a number of reasons. First, Veneto is one of the most dynamic regions in Italy. It has been historically characterized by the weight and importance of its manufacturing industry that, up to the late 90s, consisted of big national and international companies, mostly concentrated in the industrial area around Venice. Veneto used to be characterized by a good technological profile and innovativeness. Nowadays, as a result of delocalization and deindustrialization processes common to all advanced economies, there is a widespread industrial development with a high proportion of small firms, producing in both traditional and more technological sectors. In this latter case, firms appear to be aggregated in specific areas of the region. Therefore, knowledge seems to be concentrated. A second reason for the choice of Veneto is related to the scientific interest demonstrated for similar economic systems, consisting of complex industrial structures – known as districts or clusters – made of small but highly dynamic and competitive firms even on international level. Veneto represents an emblematic example of these systems that gave rise to an international debate on the role played by different forms of knowledge, competencies and skills in supporting growth; in particular, tacit knowledge, proximity and networking. Production of patents and the endowment of human capital are two of the many forms of knowledge production and diffusion. Last, but not least, studying the only Veneto allowed us to perfect the procedure for cleaning the data that, in the near future, we are going to extend to the all Italy.

Data are extracted from the OECD REGPAT database (released 2010) that contains regionalized data on all patent applications filed with the EPO from 1977 to 2009[2]. The original data have been cleaned by manually and, then, reprocessed to obtain our measures of the relationships. The manual procedure of cleaning has allowed us to exactly identify applicants and inventors and, then, to construct mobility net of any cases of "false" mobility.

The long period of time at our disposal allows us to study the production of innovation across the decade 1998-2007 in function of the variables of interest observed over a period going back as far as 20 years. The econometric model of innovation output (total number of patents and, in alternative, weighted sum of patents by firm/applicant and by year) is specified as a Poisson model as a function of the various measures of inter-firm informal relationships; we exploit the panel data by adding random and fixed effects.

The paper is structured as follows. In Section 2 we briefly review the literature related to knowledge transfer and labour mobility. Section 3 describes the original dataset. The empirical model and the constructed measures of inventor mobility and connections are described in Section 4. The results are discussed in Section 5.

2. Brief review of the reference literature

The theoretical models of endogenous growth explain the fundamental role played by human capital – education and learning – in overcoming the neoclassical hypothesis of the decreasing marginal productivity of capital and in determining long-term balanced growth paths (Romer, 1986; Romer, 1990). In particular, Romer (1990) shows how long-run growth strictly depends on the size of human capital employed in the research and development (R&D) sector, which is responsible for the production of new products. It is the process of innovation that allows growth, and a necessary condition for this to happen, is that the level of human capital in the

[2] The data for 2008-2009 are excluded from the analysis due to their incompleteness.

economy is above than a certain threshold. In the absence of an appropriate number of researchers, the economy is not able to develop a R&D sector capable of innovating and to escaping the neoclassical long-run equilibrium where per capita income does not grow. Innovation and growth both need significant human capital.

The theoretical predictions of the endogenous growth models have often been tested empirically with a regional and local focus. These analyses highlight the importance of the socio-economic context, as a factor in sharing and transmitting knowledge, enhancing innovation and growth. Innovation results from interaction between human capital endowment and other factors, such as social capital, knowledge spillovers, and networks of knowledge. Social and structural conditions affect the capacity to innovate: the s*ystems of innovation* of Lundvall (1992), the *learning regions* of Morgan (1997) and the *social filter* of Rodriguez-Pose (2001) are different ways to labelling places in which social interactions occur that allow them to become particularly innovative territories.

Knowledge can move both in time and space, across the boundaries of areas or regions with pro-innovative social and structural conditions; this may occur through labour mobility, professional relationships and, even, informal job networks, such as extra-professional relationships. Rodríguez-Pose and Crescenzi (2008)[3] found evidence both of intra and inter-regional knowledge spillovers in the European area. The intensity of interregional flows, however, depends on several factors: the stage of technological development of the receiving country, average firm size, financial infrastructure, and the similarity of their economic systems. R&D plays an important positive role, while geographical distance between regions negatively affects inter-regional flows; furthermore, innovation systems can be industry specific (Maggioni et al. 2011[4]).

[3] Rodríguez-Pose and Crescenzi (2008) use regional data for the EU-25 countries.
[4] Inter-regional flows are identified by the different localization of inventors and applicants collaborating in the same patent applications.

In recent years, a considerable literature that studies the transmission of knowledge by observing patents citations, has developed. The seminal works by Jaffe and his co-authors (Jaffe, 1986 and Jaffe et al., 1993) laid the foundations of a literature that analyses the issue of knowledge diffusion through the use of 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 United States). Applying the same methodology to European regions, Maurseth and Verspagen (2002) investigated the determinants of knowledge flows within and between countries and showed that technology flows across Europe are both industry-specific and confined by geography, language and national borders. The issue of proximity is addressed also by Verspagen and Schoenmakers (2004), Fisher 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 Fisher et al. (2006) confirmed that technological proximity is an important factor for knowledge transfer: interregional knowledge flows occur most often between regions close to one another. LeSage et al. (2007), applied a spatial interaction model with spatially structured origin and destination effects to high-technology patent activity, and found that knowledge spillovers are geographically localised. "Knowledge spillovers most often occur between origin-destination regions that belong to the same country and are in geographical proximity. On the contrary, "national borders have a negative impact on knowledge flows, and this effect is very substantial". Paci and Usai (2009) confirmed these results, and found that "technological flows among firms and inventors are favoured when they share the same language, culture, and institutional setting".

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; Breschi and Lissoni, 2009). This literature generally has defined a mobile inventor as a researcher who has worked for several patent applicants, being found in at least one other patent. Then, mobility is detected when a patent is cited by an applicant whose inventor is listed as a collaborator of the applicant whose patent is cited. The results emphasized the considerations of Jaffe et al. (1993), and stressed the role played by the labour markets in the transmission of knowledge. Labour mobility is one of the main channels of knowledge diffusion; moreover, innovation diffusion is localized 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. Agrawal et al. (2008) proposed an interesting development of the issue of proximity by distinguishing between spatial and social proximity. Using 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.

A second aspect that has been studied by using this type of dataset is the role played by co-inventorships. Breschi and Lissoni (2009), in their study on 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) in the Italian context.

Our study is part of this literature that focuses on the relationship between (local) labour markets and knowledge transmission. In particular, we evaluate how multi-firm researcher collaboration and mobility enhance firm innovation performance. Some recent works have focused on this issue, from different points of view, using alternative datasets, and different empirical specifications. Simonen and McCann (2008, 2010), using a survey of Finnish high-technology firms, calculated the probability of innovation as a function of R&D expenditure, inter-firm R&D collaboration and worker mobility, disentangling the geographical extension of these effects (same sub-region *versus* different sub-region). Inter-firm R&D cooperation is detected using a dummy variable equal to one if the firm collaborates with other firms in a project; mobility is measured by firms' recent acquisitions. The authors showed that localization of knowledge flows works through inter-firm collaboration. On the other hand, human capital mobility improves innovation performance if it occurs between different areas[5].

The issue of skill portfolios of mobile workers and their effect on firm economic performance was recently addressed in Boschma et al. (2009) who showed how worker mobility affects firm economic performance depending on the mix of geographical proximity and competences. By developing the idea of Boschma and Iammarino (2009), accounting for the effects of labour mobility, 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. Mobility is split in intra-regional and inter-regional mobility according to local labour market definition (LLM)[6]. Boschma et al. (2009) argued that "the effects of labour mobility on firm performance can only be accounted for after differentiating between types of labour inflows, in this case depending on whether new employees are recruited from the same region or from other regions". Then, they showed that the productivity impact of labour mobility strictly depends on the type of labour inflows, i.e. the types of skills that flow into the plant, and on whether new employees are recruited from the same LLM or from other LLMs. The analysis of Boshma et al. (2009) showed that geography matters when assessing the effects of different

[5] The geographical area of reference coincides with Finnish commuting areas and identifies Finland's local labour markets.
[6] Like in our study, Boschma et al. (2009) define Local Labour Markets according to a specific commuting-minimizing algorithm.

types of labour mobility. Inflows of unrelated skills positively affect plant performance when workers are recruited in the same region; on the other hand, labour mobility across regions only has a positive effect on productivity growth when incoming workers are skilled-related. These results were confirmed in Eriksson (2011).

A recent contribution to the analysis of the economic performance (however specified) of firms that benefit from knowledge spillovers was proposed in Lychagin et al. (2010), a study on the effect of knowledge spillovers and proximity – geographical, technological and market proximity – on firm's sales[7]. Proximity of spillovers and localization are studied with respect to inventors' addresses rather than headquarters, and allowing for one-year productivity lag. The authors found that geography matters: intraregional spillovers are significant, sizeable, and economically important, even after accounting for technological and product-market spillovers. Distance matters, and knowledge spillovers are very local.

3. Data

The analysis has been carried out on data from the OECD REGPAT database (December 2010 edition)[8]. REGPAT is a regionalised database whose data have been linked to regional information at NUTS3 level by means of applicants' and inventors' addresses. Two main datasets are included in REGPAT: patent applications filed with the European Patent Office (EPO) in the period 1977-2007 and patent applications filed under the Patent Co-operation Treaty (PCT) at international phase, 1977-2008. Both of them contain data on applicants – individuals or firms that apply for a patent – and inventors – individuals contributing to the

---

[7] The study is carried out merging data from the U.S. Patent and Trademark Office (USPTO) and from the U.S. Compustat 1980-2000.

8 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).

invention. Although the number of patent applications filed under the PCT at international phase have substantially increased from 2005 onwards compared to the number of applications filed to the EPO, and the global number of patent applications filed under the PCT in years 2005-2007 is higher than the number of EPO applications, we chose to work only on the part of the database containing EPO applications, for a couple of reasons. First, the PCT archive is smaller than EPO: for the Veneto region, in the period covered by REGPAT (1977-2008), patent applications filed with the EPO were 8059 against 3621 filed under the PCT. The difference in the size of the two archives could be explained by the much more expensive procedure to file under the PCT than under EPO. For that reason, and due to the economic structure of the context of interest, mainly made up of small- and medium-sized firms which pay particular attention to the costs of innovation, we preferred to use the EPO archive to be sure to include the largest possible number of firms and inventors innovating in the region. By choosing the EPO database, on the other hand, 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 and the result of the contribution of one or more inventors. Then, every single record can be linked to information on each applicant and inventor participating in the project. Among the variables, we find: the EPO application number; the application identifier, i.e. a surrogate key identifying patent applications; the EPO patent publication number; the technological classes of each patent (IPC codes); 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, it will be used 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 inventor connections and mobility. In other fields, for every applicant and every inventor we gathered: their identification codes[9]; full names and addresses; country and NUTS3 regions of residence[10].

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), basically comes from two main kinds of errors which affect the correct identification of persons and firms. The first type of problem consists in the erroneous or in varied spelling of names of individuals: Guiseppe instead of Giuseppe; Il'ya instead of Illya; Gian Carlo instead of Giancarlo; Jan-Douwe instead of Jan Douwe. The second type of error consists in writing the name of the applicant, usually a company, by using various terms: Glaxo, Glaxo Wellcome, GlaxoSmithKline, GSK. Additional problems arise in those cases where two different addresses are listed in relation to a single inventor. In these cases further investigation is needed in order to distinguish between homonymy and mobility. To solve these problems, considerable effort was made to clean the data and, then, to correctly identify the applicants and inventors through the unambiguous assignment of personal identification codes. In addition, the allocation of NUTS3 codes were checked, by controlling and correcting addresses. The details of the cleaning procedure are discussed in Appendix A.

After cleaning the data, the matrix we used for the study consists of about 3500 inventors who, between years 1998 and 2007, collaborated with on over 2000 patent applicants in the

---

9 Both identification codes are surrogate keys borrowed from the original PATSTAT Database.
10 The latter piece of information is a result of the regionalization procedure carried out by OECD utilizing the postcode, or in its absence, the town name identified in the address that appears on the first page of the patent document.

Veneto region. In this period the number of patent applications in the whole region exceeded 4700 applications[11].

4. The empirical model

Our aim is to assess whether the flows of knowledge, occurring through inventor multi-firm collaborations and inventor mobility, affect the production of innovation, and whether the effect follows a geographical pattern, and a dynamic scheme. To measure these effects we estimate an econometric model of the number of patents filed in years 1998-2007[12]. The dataset includes all patent applicants that filed at least one patent in any of the years considered. It has two dimensions: applicant and date. The dependent variables we use are, alternatively, the *number of patents* and the *weighted sum of patents* filed in one year by a firm. The applicant participation share in the production of the patent is used to weigh each patent.

The channels of knowledge transmission we are interested in arise from inter-firm connections established through the labour market. On the one hand, we consider inter-firm connections that originate through the contribution of researchers to patents filed by different firms. These links are mostly implicit forms of knowledge sharing. This variable will be called simply *Connections*, and will be measured in two ways discussed below. Secondly, we consider the transfer of knowledge occurring through inventor mobility. This variable will be called simply *Mobility*. A detailed explanation of the construction of variables is given below.

These channels of knowledge transmission are checked for geographical proximity and "productivity lag". With regards to proximity, we disentangle the geographical extent of labour connections and mobility by looking at the place of "origin" and "destination" of the interaction, using Local Labour System (LLS) as the unit of reference (the same definition is

[11] These applications are summed at the level of firm by year. We end up with around 3300 observations organized by firm and year.
12 We work with an unbalanced panel.

used in Boschma et al., 2009). Then, we define a) intra-LLS connections, if the applicant/firm is connected to other applicants located in the same LLS; b) inter-LLS connections, if the applicant/firm is connected to other applicants located in a different LLS; c) extra-region connections, if the applicant/firm is connected to other applicants located outside the region. Similarly, we label it intra-LLS labour mobility if inventors move inside the same LLS; inter-LLS mobility if inventor mobility occurs between firms located in different LLSs of the Veneto region; extra-region mobility if inventors come from firms in regions other than Veneto.

As for the productivity lag, we try to add new evidence on the dynamic effect of knowledge transmission, by checking for connections and labour mobility occurring up to 20 years before the patent filing. Therefore, we are able to answer two different questions: a) what is the time interval in which incoming inventors affect the innovation output in the destination firm b) how long does it take for informal connections to produce knowledge spillovers that affect innovation production. The meaning of the dynamic effect is different for the two variables, and will be explained below.

Before going into the details of the variable construction, we briefly discuss the econometric model adopted. Data on patent applications are typical count data. The clearly discrete nature of the dependent variable 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 $y = 0,1,2,...$ with the probability:

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

For μ > 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 $y_1, y_2,..., y_n$ 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 specification we estimate is more detailed than the model (1)-(3) because we exploit the longitudinal dimension of the data and add fixed and random effects at firm/applicant level.

The simplest model we estimate is specified as follows:

$$\log(\mu_{it}) = cons + \alpha_i + \alpha_t + \beta * CONNECTIONS_{it} + \delta * MOBILITY_{it} + \nu_i \quad (4)$$

where i identifies the patent applicant and t the year of application, for $t \in [1998, 2007]$; ν_i is the applicant random effect with gamma distribution; α_i is the fixed effect and α_t the time trend.

As previously outlined, the variables *Connections* and *Mobility* measure different channels of knowledge transmission that arise from working relationships established by inventors. *Connections* measures inter-firm knowledge linkages occurring through inventor collaboration with more than one firm. *Mobility* captures those connections that occur when a worker moves from a firm to another. Both *Connections* and *Mobility* are constructed by eliminating those relationships between firms that belong to the same group and controlling for cases of "false" mobility due to changes of company type or name (see Laforgia and Lissoni, 2009)[13].

Connections has been proxied in two different ways. A first measure of *Connections* of applicant i at time t is constructed by summing the number of inventors working for that applicant at time t – i.e. applying for a patent at that time – and participating, at the same time t, in patent applications of other applicants. This measure captures the size of inter-firm human capital sharing that is likely to be responsible for knowledge sharing/transfers. We will call this variable *number of connected inventors*. We propose a second measure of *Connections* given by the number of applicants to which applicant i is "connected" to at time t through shared inventors. This measure will be called *number of applicant connections*. Both measures are constructed through the identification of multiple-collaborations of inventors and excluding multiple counting. However, they differ in their intrinsic meaning. The *number of connected inventors* focuses on the extension of the linkages headed by inventors, stressing the role played by the source of knowledge flows. The *number of applicant connections*, on

[13] See Appendix A for details on the cleaning process.

the other hand, shifts the attention from the source of the externality to the extent of the network. This variable should allow us to evaluate more properly each applicant connection and the potential for knowledge spillovers.

In light of the main aim of the study, we initially excluded co-inventorships, i.e. formal multi-firm participations to a single patent in the definition of *Connections*. In so doing we were able to measure inter-applicant connections that occur only through inventor working collaborations. Later we became interested in evaluating the role played by co-inventorships as well; indeed, the productive structure of the Veneto region, made up of small- and medium-sized firms, it is particularly interesting to study the role of formal inter-firm collaboration in enhancing innovation production. So, we defined *Connections* in three ways: a) excluding co-inventorships (*net connections*); b) including co-inventorships (*gross connections*); c) including co-inventorships and weighting each inventor link by the corresponding firm's share in the patent (*gross* connect*ions weighted by patent share*). These alternative definitions, together with the very disaggregated definition of the territory (sub-regional functional areas), allowed us to highlight some interesting results that differ from what has generally been found at regional level concerning the role played by co-inventorships in the transmission of knowledge. Inter-firm *Connections* are observed in any year up to 20 years before the patent application. In the tables of results they will be labelled simply *Connections* current year, *Connections* lag 1-5 years, *Connections* 6-10 years, and so on.

A mobile inventor at time t is defined as an individual being already registered in the dataset in correspondence to a patent filed in any previous year by any applicant, and participating, at time t, in the production of a patent of a different applicant i with which she/he has not shared any of his last ten patents. Then, *Mobility* at firm i at time t is constructed as the sum of mobile inventors (or incoming inventors) at time t. By construction, our definition of mobility excludes from the count those inventors who had already collaborated with the firm in a patent application.

It has to be underlined that our definition of mobility is slightly different from that defined by means of patent citations. We could say that we adopt a more general measure. Indeed, when using patent citations, knowledge transmission is detected when the incoming inventor participates in a patent that cites one of his previous innovations produced in collaboration with another firm. This means that, when the incoming inventor participates to a patent that does not cite any of his previous innovations, no knowledge transfer at all is detected. In our case, the flow of knowledge is detected when we observe an inventor who produces an innovation with a new applicant he had not worked with before. As a matter of fact, this measure of mobility captures the flow of knowledge embedded in the human capital of a mobile inventor and is not related only to a specific patent. So it allows us to detect knowledge flows that occur through inventors who move between firms which do not produce strictly-related innovations.

By construction, *Mobility* is observed when there is at least a one-year interval between the dates of patent filing at the leaving and destination firms. This prevents that *Mobility* overlaps *Connections*. In fact, if we observe an inventor filing two patents with two different firms in the same year, this is recorded as a connection and not as mobility. So, *Mobility* is measured by taking account of the interval that elapses between the dates of patent registrations with the origin and destination firms (simply called time lag).

In Table 1 we summarize the distribution of patents, inventor mobility, and connections (net of co-inventorships) of the dataset used. More than 20% of our observations record more than one patent application, and almost 5% of cases found four or more patents. Around 12% of observations record at least one incoming inventor, while connections are less frequent. Inter-firm connections, whatever measure we apply, 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; on the contrary, 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. One explanation for these results is that some of the connections that occur through a single inventor correspond to more than two applicant connections. This means that the inventor works for more than two applicants at the same time.

[Table 1]

5. Results

The initial empirical analysis has been carried out using the specifications without geographical disaggregation. We will call this the *Base model*. Then, we added the territorial disaggregation and estimated the *Geographical model*. Estimation results related to the *base model* are reported in Table 2; those concerning the *geographical model* are shown in Table 3[14]. Both models have been estimated using the two measures of *net Connections*, and the dependent variables *number of patents* and *weighted sum of patents*. Results obtained using the *number of connected inventors* variable are summarized in part a) of every table. Those obtained using the *number of applicant connections* are reported in part b) of each table. In Appendix C we show the estimates we derived using *gross connections*, only for the case of the *number of applicant connections* variable [15]. We will go through the discussion of the estimation results for the specification with *gross Connections* later in this section.

[14] The correlations between the explicative variables of both models are very small.
[15] We do not list the results for the model estimated with the *number of connected inventors* because it replicates, in general, the results obtained with the *number of applicant connections* we discuss. The tables can be provided upon request.

All specifications include a time trend and the stock of patents filed in the decade before the period of interest. However, when using the *weighted sum of patents* as dependent variable, the stock of patents filed in the decade before is constructed weighting each patent by the firm's participation share to it.

The *base model* estimates appear substantially stable, whatever the measure of *net Connections* being used, and show a positive trend in the production of innovation. Moreover, it is clear that firms that applied for a higher number of patents in the overall period 1988-1997, produce more patents in the years of interest.

[Table 2]

Inventor mobility enhances the production of innovation, but only in cases in which the period of time between the patent produced with the leaving applicant and that filed with the destination applicant is not greater than five years. This suggests that mobility can have productivity effects on innovation if the contribution of the incoming inventor is realized rapidly (in the short run) in the production of a new patent. After five years, inventor mobility does not have any significantly demonstrable productivity effect.

As expected, the estimated coefficients of the two measures of *Connections* have slightly different significance. The result is not surprising. Indeed, the *number of applicant connections* assigns a larger number of connections than the *number of connected inventors* if, given the same number of connected inventors, these co-operate with a greater number of applicants (net of duplication). Then, the *number of applicant connections* captures more correctly the potential of inter-firm flows of knowledge. Estimation results appear to confirm this idea. The *number of applicant connections* is significant both in the very short run (at time *t*), and in the short run (one to five years), and in the medium run (11-15 years). The *number of connected inventors*, on the contrary, is significant only in the medium-long run. Then, focusing on the coefficients of the *number of applicant connections*, reported in part b) of the table, it is clear that "multi-applicant" inventors give rise to flows of knowledge that enhance the production of patents of the involved applicants. Moreover, the effect increases over time, and is highest for the most remote *connections*.

The results for the *base model* are robust to the geographical specification of *Mobility* and *Connections* (see Table 3). The stock of patents produced in the decade prior to the sample starting year and the trend over time remained significant and stable. Moreover, the two measures of *Connections* confirm their different capacities to evaluate the relevance of inventor labour market relationships for the transmission of knowledge and the enhancement of innovation production. As for the *base model*, the significance of *Connections* is generally higher when using the *number of applicant connections* rather than the *number of connected inventors*.

[Table 3]

The pattern of the dynamic effects of knowledge flows on the production of new patents are confirmed, in general, in the *geographical model* as well. Inventor mobility positively affects firm production of innovation if incoming inventors contribute to a new patent within 5 years of the last patent registered with the leaving applicant. Therefore, *Mobility* is significant only in the short-run. In contrast, *Connections* have both short, medium and long run effects, and the last are stronger than the very short run effects. However, geographic distance plays some role in explaining the relevance of the flows of knowledge, and their consequences in terms of innovation production.

Looking at Table 3 part b), we see that the estimated coefficients of the geographical specification, when significant, are greater than those predicted in the *base model*. This suggests that, if the geographical issue is not taken into account, productive effects of knowledge flows are underestimated.

Knowledge flows occurring through inventor mobility and connections appear, at first sight, strongly localized. Productivity effects are, in general, significant and positive when these working relationships happen within the same local labour market, e.g. the same LLS. This has been verified for both *Mobility* and *Connections* in the very short and short run (up to the 5 year lag) and in the medium run (lag 11-15 years). Intra-LLS *Connections* show, in the medium term, the strongest productive effect on innovation. On the other hand, we find a positive and significant impact even for connections of applicants located in different LLSs of the same region or in different regions. In the first case, the effect on innovation is relatively short (within 5 years) and much greater than that produced by intra-LLS *Connections* (double, in terms of marginal effects). In the case of *Connections* with extra-regional LLSs, the effect is much delayed in time – between 16 and 20 years – and not particularly high in marginal terms, when compared with the marginal effect in the medium run of intra-LLS *Connections*.

To summarize, inventor working relationships with more than one applicant give rise to knowledge flows that positively affect the production of innovation, especially in the long run, and when these relationships are extremely localized. However, if we focus on the short-run, the highest productive effect can be achieved through the relationships occurring between different labour markets of the same region. Extra-region *Connections*, on the other hand, have an influence only in the long run, and the marginal effect is not particularly relevant.

Given the estimation results, we derived the predicted value of the number of patents for each model. To better summarize the relationship between patent production and our variables of interest, in Figure 1 we represent the linear interpolation between the predicted values of patents and variables *Mobility* and *Connections*. Part a) of the Figure refers to the *base model*; part b) to the *geographical model*.

[Figure 1]

Looking at the *base model*, we observe that time plays an important role in transferring knowledge through labour market relationships and in transforming that knowledge into increased production of innovation. Productivity effects increase with the time lag between the time at which we observe the working relationships and the time at which we observe a new patent application. Even in the geographical model, time remains a key element for enabling the translation of the flow of knowledge into new innovative output. This, however, applies in particular when the flow of knowledge is extremely localized and occurs within the boundaries of the LLS. Indeed, the relation between the predicted number of patents and inter-LLS *Connections* is extremely high in the short run (1-5 years); by this time, inter-LLS *Connections* are able to produce the same innovative output that intra-LLS *Connections* achieve in the long run.

As mentioned in the introduction, the analysis discussed up to now has been repeated using a measure of *Connections* including inter-applicant formal relationships documented by co-participation in patents. More exactly, *gross connections* include *net connections* plus those relationships that derive from inventor working relationships with more applicants that co-participate in the invention of a patent[16]. The idea behind this extension of the study is to evaluate the role played by co-inventorhips in enhancing the production of innovation, in a territory that consists, mainly, of small and medium firms. In this context, employers may be urged to carry out innovation activities in collaboration with other firms to share the costs of

[16] Remember, by construction, all measures of *Connections* and *Mobility* are net of relationships occurring between applicants that belong to the same business group and "false" mobility.

innovation. If this were true, co-inventorships would not necessarily have an effect of increasing the production of innovation. On the contrary, they may not affect the product of innovation.

Estimation results using *gross connections* measured by the *number of applicant connections* are reported in Appendix C[17]. Because *gross connections* include co-inventorships, we also used a measure in which each connection is weighted by the respective applicant participation share in the patent (variable *gross connections weighted by patent share*). In Column 1 we list the estimates for the *number of patents* using *gross connections*; in Column 2 the *weighted sum of patents* is regressed on gross connections; in Column 3 the *weighted sum of patents* is estimated on *gross connections weighted by patent share*.

Our results show that, if we include co-inventorships in the measure of *Connections* and do not weight each connection for its applicant participation share in the patent, the effect of *Connections* on innovation is, in general, either insignificant or negative. This may confirm our hypothesis, and suggest that co-inventorships, in the Veneto region, do not affect the production of innovation. However, results change, and *gross connections* become positive and significant, with some coefficients close to those estimated using *net connections*, when each connection is weighted by the respective applicant participation share in the patent (*gross connections weighted by patent share*) and the dependent variable is the *weighted sum of patents* (Column 3). This result seems to suggest that, to properly estimate the effect of *Connections* on innovation production, when including co-inventorships, it is necessary to weigh each connection by its applicant participation share in the patent. If this precaution is not taken, *Connections* are overvalued; then, productive effects are insignificant or underestimated. This result, although related to a specific economic territory, suggests that, in

[17] Estimates have also been carried out using the *number of connected inventors*. Similarly to what happens for the specification with *net Connections*, the *number of connected inventors* is less significant than the *number of applicant connections*.

other situations, the proper evaluation of the productive effect of co-inventorships could be undermined by incorrect measurement.

6. Conclusions

This study arises from our interest in analysing the effect of knowledge flows that occur through inventor labour market relationships – established by "multi-applicant inventors" – and inventor mobility. The article contributes to the literature on the empirical study of the transfer of knowledge by addressing several issues. First of all, we focus on the productive effect of the flow of knowledge, and study its impact on the production of innovation. Secondly, we evaluate the dynamic and geographical effect of these relationships, by measuring inventor connections and mobility up to twenty years before a patent registration with a applicant, and observing their geographical distribution. In this respect, we distinguish between relationships and mobility that are very localized – occurring within the same local labour market –, "regionalized" – occurring between different local labour markets in the same region, and not localized – arising from inter-regional interactions.

The analysis is carried out on the population of firms located in Veneto, a region in North-eastern Italy, and filing patents with the European Patent Office (EPO) in the pre-crisis period, 1998-2007. The data have been extracted from the OECD REGPAT database, and cleaned manually. This allowed us to construct the variables of interest excluding cases of "false" mobility. The empirical model is specified as a panel Poisson model.

Our study makes an innovative contribution to the literature on the transmission of knowledge by showing that inventor working relationships and mobility can enhance the production of innovation. This effect is particularly high if inventor relationships and mobility are very localized (they occur within the same local labour market), confirming, in general, what has been argued previously in the literature on the transmission of knowledge. However, we show that the study of the geographical distribution of the flows of knowledge, and the evaluation of their impact on innovation, need dynamic specification. The geographical focus, together with dynamics, allows us to detect significant effects in the long run of inter-regional relationships. These results, although related to the Veneto region, provide a general contribution to the literature that studies the productive effects of knowledge spillovers.

Specifically, we find that the effect of inventor mobility on the production of innovation is very localized and short-term: incoming inventors have productivity effects on innovation if they move from the same local labour market and if their contribution is realized rapidly in the production of a new patent. On the contrary, the effect of inventor connections on innovation varies, with geographic distance playing some role in explaining the significance and the relevance of the flows of knowledge and their dynamic impact on innovation.

The relationships that guarantee a productive effect on innovation, both in the very-short run and with a lag of a few or several years (up to 15), are the connections that occur within the same local labour market. Moreover, the effect increases with time. However, also inventor connections between different local labour markets of the same region have significant effects in the short run, and these effects are much higher than those achieved, in the same time interval, by localized connections. On the contrary, inter-regional connections have productive effects only in the long run. However, their productive effects are as great as connections between intra-region local labour markets.

To summarize, the effect of knowledge flows on innovation, caused by “multi-applicant” inventors, depends on both geographical distance and time. In general, we can say that the greater the distance, the longer the time needed to achieve productive results. However, in the short run, both very localized connections and connections between intra-region local labour markets have significant effects. In the last case, the size of the effect is much higher than in the case of very localized connections.

We suggest interpreting these results in terms of “technological proximity” (Boschma and Iammarino, 2009; Boschma et al., 2009). Labour inflows and inter-applicant connections are very effective in terms of innovation production when they “connect” applicants from different local labour markets of the same region, which we can imagine as being related but not as similar as the applicants belonging to the same local labour market. On the other hand, extra-regional relations are effective only in the long run and their impact is not particularly high compared to that of intra-regional relationships, because applicants that belong to different regions are “technologically” too distant: they are neither similar nor related.

A final remark on the estimates including co-inventorships needed to be made. We show that co-inventorships do not aid the output of innovation. This suggests that co-inventorships could be a means of cost sharing in innovation, rather than a channel for the transmission of knowledge and the increase of the innovation output. This result, although related to the economic context of the Veneto region, may be found in other economic systems characterized by small- and medium-sized firms and could be a focus for future studies.

References

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.

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.

Eriksson R.H. (2011) Localized spillovers and knowledge flows: how does proximity influence the performance of plants? *Economic Geography*, 87: 127-152.

Fischer, M.M., 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.

Hendy, S.C., Sissons, C.M. (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 A.B. (1986) Technological opportunity and spillovers of R&D: evidence from firms’ patents, profits and market share, *American Economic Review*, 76: 984–1001.

Jaffe, A.B., 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, S.J., Marschke, G. (2008) International knowledge flows: evidence from an inventor-firm matched data set. In Freeman, R.B., Goroff, D.L. (Eds) *Science and Engineering Careers in the United States: An Analysis of Markets and Employment*, pp. 321-348. Chicago: University of Chicago Press.

Laforgia, F., Lissoni, F. (2009) What do you mean by 'Mobile'? Multi-applicant inventors in the European biotechnology. In Malerba F., Vonortas N.S. (Eds) *Innovation Networks in Industries*, pp. 157-176. Cheltenham: Edward Elgar Publishing Ltd.

LeSage, J.P., Fischer, M.F., 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, B.A. (1992) *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*. London: Pinter Publishers.

Lychagin, S., Pinkse, J., Slade, M.S., Van Reenen, J. (2010) Spillovers in space: does geography matter? Discussion Paper 7928, London: Centre for Economic Policy Research.

Maggioni, M.A., Uberti, T.E., 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 Working Paper 2008/2, OECD Publishing. doi:10.1787/241437144144

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.

Power, D., Lundmark, M. (2004) Working through knowledge pools: labour market dynamics, the transference of knowledge and ideas, and industrial clusters, *Urban Studies*, 41: 1025-1044.

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.

Romer, P.M. (1986) Increasing returns and long-run growth, *Journal of Political Economy*, 94: 1002-1037.

Romer, P.M. (1990) Endogenous technological change, *Journal of Political Economy*, 98: 71-102.

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. Working Paper 12479, National Bureau of Economic Research.

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

Zucker, L.G., Darby, M.R., Armstrong, J. (1998) Geographically localised knowledge: spillovers or markets? *Economic Inquiry*, 36: 65-86.

Figures and 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)

**Net of co-inventorships*

Table 2. Base model. Estimation results

a) *Connections* measured by the *net number of connected inventors*

	Number of patents		Weighted sum of patents	
	Coeff.	Marg. effect	Coeff.	Marg. effect
Time trend	0.02***	0.03***	0.02***	0.03***
	(0.01)	(0.01)	(0.01)	(0.01)
Stock of patents (period 1988 – 1997)	0.02***	0.02***		
	(0.00)	(0.00)		
Stock of weighted number of patents (period 1988 – 97)			0.02***	0.02***
			(0.00)	(0.00)
Mobility				
Lag 1-5 years	0.10**	0.13**	0.09*	0.12*
	(0.05)	(0.06)	(0.05)	(0.06)
Lag 6-10	0.01	0.01	0.02	0.03
	(0.09)	(0.12)	(0.09)	(0.11)
Lag 11-15	0.03	0.04	0.07	0.08
	(0.13)	(0.17)	(0.13)	(0.16)
Lag 16-20	-0.25	-0.33	-0.23	-0.29
	(0.26)	(0.35)	(0.27)	(0.34)
Connections - current				
Number of connected inventors#	0.06	0.08	-0.01	-0.02
	(0.07)	(0.10)	(0.08)	(0.10)
Connections lags				
Number of connected inventors#				
Lag 1-5 years	0.14	0.19	0.19	0.23
	(0.14)	(0.19)	(0.15)	(0.18)
Lag 6-10	0.22	0.29	0.21	0.27
	(0.17)	(0.22)	(0.17)	(0.22)
Lag 11-15	0.42*	0.56*	0.43**	0.54**
	(0.21)	(0.28)	(0.22)	(0.27)
Lag 16-20	0.32	0.43	0.23	0.29
	(0.26)	(0.34)	(0.28)	(0.35)
Constant	-39.38***		-45.49***	
	(10.86)		(11.22)	
Ln(μ)	-2.52***		-2.32***	
	(0.09)		(0.09)	
Observations	3,283		3,283	
Number of Applicants	2,018		2,018	
Wald test Chi^2	103.74***		99.65***	

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Variable normalized by the number of inventors by firm and year.

b) *Connections* measured by the *net number of applicant connections*

Dependent variable	Number of patents		Weighted sum of patents	
	Coeff.	Marg. effect	Coeff.	Marg. Effect
Time trend	0.02***	0.03***	0.02***	0.03***
	(0.01)	(0.01)	(0.01)	(0.01)
Stock of patents (period 1988 – 97)	0.02***	0.02***		
	(0.00)	(0.00)		
Stock of weighted number of patents (period 1988 – 97)			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	0.02	0.02	0.03	0.04
	(0.09)	(0.12)	(0.09)	(0.11)
Lag 11-15	0.04	0.05	0.07	0.09
	(0.13)	(0.17)	(0.13)	(0.16)
Lag 16-20	-0.24	-0.32	-0.23	-0.28
	(0.26)	(0.35)	(0.27)	(0.33)
Connections – current				
Number of applicant connections#	0.11***	0.15***	0.07*	0.09*
	(0.04)	(0.05)	(0.04)	(0.05)
Connections - lags				
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	0.16	0.21	0.15	0.19
	(0.11)	(0.14)	(0.11)	(0.14)
Lag 11-15	0.32**	0.43**	0.33**	0.42**
	(0.14)	(0.19)	(0.15)	(0.18)
Lag 16-20	0.27	0.36	0.22	0.28
	(0.17)	(0.23)	(0.18)	(0.23)
Constant	-37.44***		-43.67***	
	(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	120.49***		110.31***	

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Variable normalized by the number of inventors by firm and year.

Table 3. Geographical model. Estimation results

a) *Connections* measured by the *net number of connected inventors*

Dependent variable	Number of patents		Weighted sum of patents	
	Coeff.	Marg. effect	Coeff.	Marg. effect
Time trend	0.02***	0.03***	0.02***	0.03***
Stock of patents (period 1988 – 97)	0.02***	0.02***		
Stock of weighted number of patents (period 1988 – 97)			0.02***	0.02***
Mobility				
Intra-LLS, Lag 1-5 years	0.12**	0.16**	0.12*	0.15*
Lag 6-10	-0.04	-0.05	-0.02	-0.03
Lag 11-15	0.02	0.03	0.06	0.08
Lag 16-20	-0.21	-0.29	-0.21	-0.27
Inter-LLS, Lag 1-5 years	0.03	0.04	0.04	0.05
Lag 6-10	0.05	0.06	0.04	0.05
Lag 11-15	-0.00	-0.01	0.03	0.04
Lag 16-20	-0.29	-0.34	-0.23	-0.26
Extra-region, Lag 1-5 years	0.14	0.19	0.12	0.15
Lag 6-10	0.09	0.12	0.12	0.16
Lag 11-15	0.09	0.13	0.12	0.17
Lag 16-20	-0.35	-0.40	-0.33	-0.35
Connections – current				
Number of connected inventors#				
Intra-LLS	0.06	0.08	-0.01	-0.01
Inter-LLS	0.16	0.21	0.04	0.06
Extra-region	-0.08	-0.10	-0.16	-0.20
Connections – lags				
Number of connected inventors#				
Intra-LLS, Lag 1-5 years	0.18	0.25	0.21	0.26
Lag 6-10	0.30	0.41	0.33	0.41
Lag 11-15	0.74***	1.00***	0.77***	0.96***
Lag 16-20	0.62	0.84	0.61	0.77
Inter-LLS, Lag 1-5 years	0.25	0.34	0.32	0.40
Lag 6-10	-0.12	-0.17	-0.28	-0.35
Lag 11-15	-0.73	-0.98	-0.68	-0.86
Lag 16-20	0.12	0.16	-0.07	-0.09
Extra-region, Lag 1-5 years	-0.27	-0.36	-0.19	-0.24
Lag 6-10	0.27	0.36	0.30	0.38
Lag 11-15	0.13	0.18	0.23	0.29
Lag 16-20	0.95	1.27	1.01	1.27
Constant	-39.14***		-45.22***	
Ln(μ)	-2.54***		-2.34***	
Observations	3,283		3,283	
Number of Applicants	2,018		2,018	
Wald test Chi^2	115.27***		111.34***	

For reasons of space standard errors are not reported. *** p<0.01, ** p<0.05, * p<0.1.
Variable normalized by the number of inventors by firm and year.

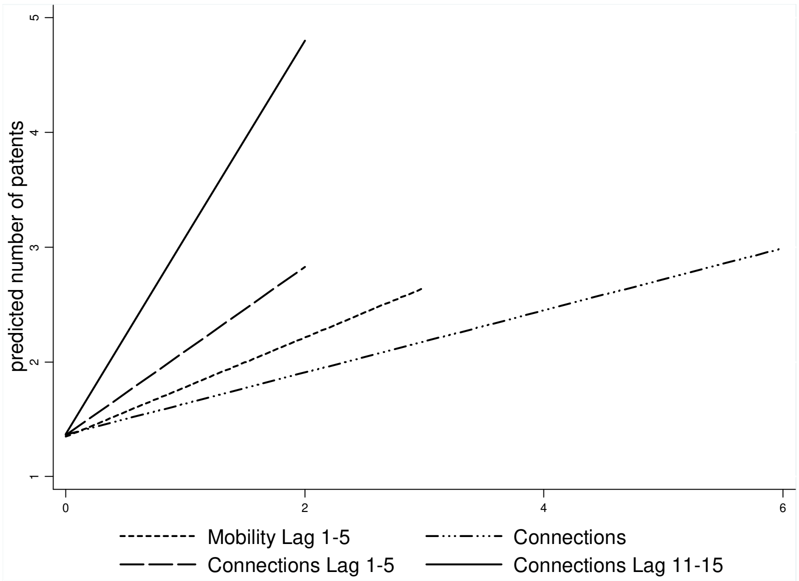
b) *Connections* measured by the *net number of applicant connections*

Dependent variable	Number of patents		Weighted sum of patents	
	Coeff.	Marg. effect	Coeff.	Marg. effect
Time trend	0.02***	0.03***	0.02***	0.03***
Stock of Patents (period 1988 – 97)	0.02***	0.02***		
Stock of weighted number of patents (period 1988 – 97)			0.02***	0.02***
Mobility				
Intra-LLS, Lag 1-5 years	0.11*	0.15*	0.11*	0.14*
Lag 6-10	-0.03	-0.04	-0.02	-0.02
Lag 11-15	0.03	0.04	0.07	0.08
Lag 16-20	-0.21	-0.28	-0.21	-0.26
Inter-LLS, Lag 1-5 years	0.04	0.05	0.04	0.05
Lag 6-10	0.05	0.07	0.04	0.06
Lag 11-15	-0.01	-0.01	0.02	0.02
Lag 16-20	-0.28	-0.33	-0.22	-0.25
Extra-region, Lag 1-5 years	0.14	0.19	0.13	0.16
Lag 6-10	0.09	0.13	0.12	0.16
Lag 11-15	0.09	0.13	0.13	0.17
Lag 16-20	-0.35	-0.39	-0.32	-0.34
Connections – current				
Number of applicant connections#				
Intra-LLS	0.12**	0.16**	0.08	0.10
Inter-LLS	0.12	0.16	0.07	0.09
Extra-region	0.00	0.01	-0.02	-0.02
Connections – lags				
Number of applicant connections#				
Intra-LLS, Lag 1-5 years	0.18*	0.24*	0.19*	0.24*
Lag 6-10	0.17	0.23	0.18	0.23
Lag 11-15	0.71***	0.95***	0.71***	0.89***
Lag 16-20	0.42	0.57	0.48	0.60
Inter-LLS, Lag 1-5 years	0.46*	0.62*	0.50**	0.63**
Lag 6-10	-0.20	-0.27	-0.34	-0.43
Lag 11-15	-0.98	-1.31	-0.93	-1.16
Lag 16-20	0.06	0.08	-0.14	-0.17
Extra-region, Lag 1-5 years	-0.08	-0.11	-0.03	-0.04
Lag 6-10	0.30	0.40	0.30	0.38
Lag 11-15	0.12	0.16	0.16	0.20
Lag 16-20	0.43*	0.57*	0.44*	0.56*
Constant	-37.37***		-43.66***	
Ln(μ)	-2.57***		-2.36***	
Observations	3,283		3,283	
Number of Applicants	2,018		2,018	
Wald test Chi^2	138.51***		128.98***	

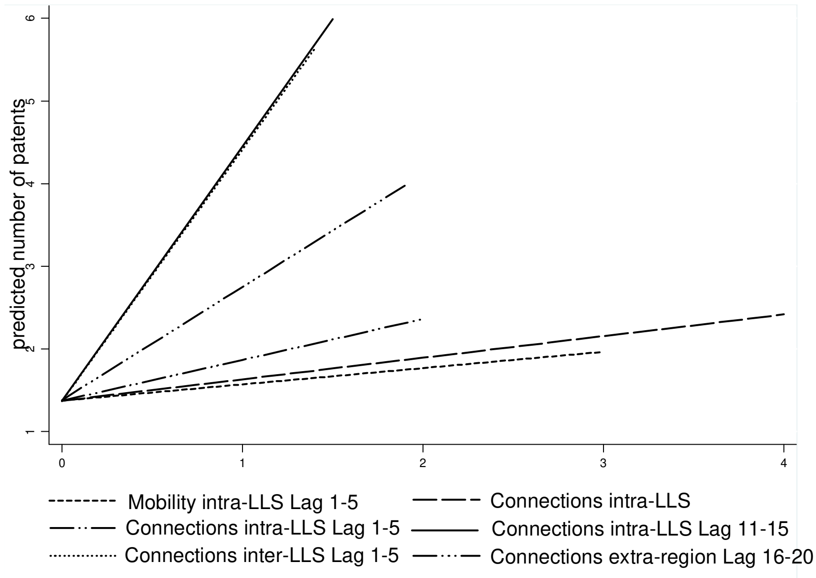
For reasons of space standard errors are not reported. *** p<0.01, ** p<0.05, * p<0.1.
Variable normalized by the number of inventors by firm and year.

Figure 1. Linear interpolation between the predicted number of patents, *Mobility* and *Connections*.

a) Base model



b) Geographical model



Appendix

A 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.

In order 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 either by PATSTAT or by 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). There are several other sources of differences related to patent data that might refer to 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), or qualifiers (Mario Rossi versus Prof. Dr. Mario Rossi), 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 in this case, small differences in the order of appearance –

street, number or number, street – or extra spaces, commas, etc. will make the same address to be considered a distinct addresses (they actually are different strings).

REGPAT contains two additional fields that identify 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[18]. 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 in order to select applicants located in Veneto.

As mentioned above, REGPAT applicants and inventors are identified by means of a personal identification code. Basically, it is a number given to every combination of Applicant/Inventor name, address and country code. As a consequence, 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 both to 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[19]. Although human error cannot be

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

[19] 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).

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 will mostly refer to the cleaning and name-matching procedure for the inventors' subset, given the fact that this set contained most of the ID codes misattributed.

As a first step (i), we wrote the field with the inventor's surname and name in capital letters and put them in the correct order: SURNAME, NAME[20]; (ii) we dropped punctuation symbols – apostrophes, hyphens, pipes, commas, periods, dashes, numbers, qualifiers or affiliations attached to the name; (iii) we trimmed extra spaces between words and corrected those names containing extra symbols like ó, ú, é; (iv) we corrected the cases of misspelled names; (v) we wrote the firms' name correctly and verified for all cases where a firm changed its name after a merger or an acquisition such as, Glaxo/Wellcome/SmithKline to which was assigned the generic name GLAXO SpA.

After having carried 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:

- 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;

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

- 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 inventors' address coincided with the applicants' address with whom the inventor has registered a patent in the past;
- same inventor name, different addresses, different applicants, different patent priority years.

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 it was a case of homonymy. 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 changed address, name or legal status (for example: from Srl to SpA). Furthermore, in some cases, we checked the inventors' curriculum vitae, if available online (including personal LinkedIn profiles). Finally, in cases of single applicants and inventors with different ID codes we assigned them a new codes.

Referring to the applicants' subset, we generated two new fields containing the name of the municipality of establishment and the correct postcode. The information was extracted from each applicant's address (that in REGPAT is a not standardised text string), where we found many transcription errors that can affect the attribution of regional codes. Then, we first standardised identical addresses which 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[21]; having a placename instead of the town name, etc.). 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 information on the municipality and the postcode. The last step was essential in order to assign each applicant to the appropriate Local Labour System.

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

B Descriptive statistics

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

Variable normalized by the number of inventors by firm and year.

C Estimates with gross *Connections* measured by *number of applicant connections*

	(1)	(2)	(3)
Dependent variable	Number of patents	Weighted sum of patents	Weighted sum of patents
Time trend	0.02***	0.02***	0.02***
Stock of Patents (period 1988 – 97)	0.02***		
Stock of weighted sum of patents (period 1988 – 97)		0.02***	0.02***
Mobility			
Intra-LLS, Lag 1- 5 years	0.12**	0.13**	0.12*
Lag 6-10	-0.04	-0.03	-0.03
Lag 11-15	0.03	0.04	0.05
Lag 16-20	-0.21	-0.23	-0.21
Inter-LLS, Lag 1- 5 years	0.05	0.04	0.04
Lag 6-10	0.05	0.03	0.04
Lag 11-15	-0.01	0.02	0.03
Lag 16-20	-0.29	-0.26	-0.24
Extra-region, Lag 1- 5 years	0.14	0.14	0.14
Lag 6-10	0.08	0.11	0.13
Lag 11-15	0.08	0.10	0.13
Lag 16-20	-0.35	-0.34	-0.33
Gross Connections - current			
Number of applicant connections#			
Intra-LLS	0.02	-0.15***	
Inter-LLS	0.06	-0.14	
Extra-region	-0.11	-0.71***	
Gross Connections -lags			
Number of applicant connections#			
Intra-LLS, Lag 1-5 years	0.08	0.03	
Lag 6-10	0.18*	0.19*	
Lag 11-15	0.65***	0.64***	
Lag 16-20	0.52	0.75*	
Inter-LLS, Lag 1-5 years	0.32*	0.35*	
Lag 6-10	-0.29	-0.35	
Lag 11-15	-0.62	-0.73	
Lag 16-20	0.24	0.15	
Extra-region, Lag 1-5 years	-0.04	-0.13	
Lag 6-10	0.01	0.00	
Lag 11-15	0.11	0.12	
Lag 16-20	0.32	0.33	
Gross Connections weighted by patent share - current			
Number of applicant connections#			
Intra-LLS			-0.01
Inter-LLS			0.02
Extra-region			-0.67***
Gross Connections weighted by patent share – lags			

Intra-LLS, Lag 1-5 years			0.12
Lag 6-10			0.18
Lag 11-15			0.70***
Lag 16-20			0.63
Inter-LLS, Lag 1-5 years			0.45*
Lag 6-10			-0.39
Lag 11-15			-0.92
Lag 16-20			0.21
Extra-region, Lag 1-5 years			-0.07
Lag 6-10			0.22
Lag 11-15			0.15
Lag 16-20			0.40
Constant	-38.02***	-43.27***	-43.80***
Ln(μ)	-2.55***	-2.39***	-2.37***
Observations	3,283	3,283	3,283
Number of Applicants	2,018	2,018	2,018

Variable normalized by the number of inventors by firm and year.