Are knowledge flows all alike?
Evidence from European regions

Francesco Quatraro[a,c] and Stefano Usai[b]

a) University of Nice Sophia Antipolis, GREDEG-CNRS, 250, rue Albert Einstein, 06560, Valbonne (France)
b) University of Cagliari, CRENoS, Via S. Ignazio, 78, Cagliari (Italy)
c) BRICK, Collegio Carlo Alberto, Moncalieri (Torino, Italy)

ABSTRACT. The paper investigates the impact of distance, contiguity and technological proximity on cross-regional knowledge flows, by comparing the evidence concerning co-inventorship, applicant-inventor relationships and citation flows. We find evidence of significant differences across these diverse kinds of knowledge flows for what concerns the role of distance, and the moderating role of contiguity and technological proximity. Moreover, we show that border effects may prove crucial in a twofold sense. On the one hand we show that contiguity between regions belonging to two different countries still plays a moderating role, although weaker as compared to that of within-country contiguity. On the other hand, regions sharing a frontier with a foreign country are more likely to exchange knowledge with this foreign country than other regions which are far away from the border.

Keywords: Knowledge Flows, Border regions, Patents, regional competitiveness, Europe, Gravity

JEL Classification Codes: R11, O33

Acknowledgments. Preliminary versions of this paper have been presented at the conference of the Italian branch of the Regional Science Association (AISRE) held in Palermo (September 2013); the Geography of Innovation conference held in Utrecht (January 2014); the RSA conference held in London (November 2013) and at a seminar held at the University of Padua (November 2013). We thank the participants for comments and remarks. Francesco Quatraro acknowledges the funding of European Commission, within the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement No. 266959 (PICK-ME).

1 Introduction

The creation and the diffusion of technology are crucial pre-requisites for economic growth (Romer, 1986). Both phenomena have an important geographical content in that their dynamics depend on local increasing returns and on local knowledge spillovers. Arthur (1989) and Krugman, (1991) provide convincing theoretical arguments to explain the multifaceted nature of local economies which make the generation of technological knowledge a polarized activity across space. At the same time, Grossman and Helpman (1991) explain that knowledge has both a tacit and a codified nature, and, as a result, a public good component which may work in different ways across territories.

Since the seminal work by Jaffe et al. (1993), an increasing body of empirical literature has focused on the analysis of knowledge flows and spillovers, mainly drawing upon data about patent citations. Despite the wide range of empirical works, the debate about the localization of knowledge spillovers is still far from finding an exhaustive conclusion. A common criticism is that citations may prove to be a 'noisy' indicator of knowledge spillovers (Jaffe et al., 1998), since they do not always imply an actual flow of knowledge from cited to citing inventor. Indeed, Thomson and Fox Kean (2005) show that the results obtained by Jaffe et al. (1993) are due to an imperfect matching of patent data, which is likely to produce a biased evidence concerning the geographical clustering of citations. Following this result, Thompson (2006) proposes an alternative citing-cited patent matching scheme, showing that citations still appear to be localized both within and across international borders. More recently, Belenzon and Schankerman (2013) study the geography of university knowledge spillovers, confirming that citations to patents are localized and sensitive to border effects whilst citations to publications are not. Criscuolo and Verspagen (2008) extend the debate and the analysis to the European case and find that geographical distance is a factor that strongly diminishes the probability of knowledge flows. This probability is found to be influenced also by cognitive distance and time.

Patent citations, however, are not the only available counterexample to Krugman's (1991, p. 53) observation that "knowledge flows . . . leave no paper trail by which they may be measured and tracked." As a matter of fact, Jaffe et al. (1998), in light of a quantitative and qualitative analysis, conclude that geographic spillovers are underestimated by patent citations and point to the necessity to go beyond this indicator. This research avenue has been recently explored in some contributions (Picci, 2010, Maggioni et al. 2011, Capelli and Montobbio, 2013) which investigate other patent related indicators, such as collaborations among inventors and relationships among patent inventors and applicants. Giuri and Mariani (2013) follow a different route by collecting direct information from the patent inventors themselves on their use of knowledge spillovers to produce inventions. All these contributions find that knowledge spillovers have a rather important geographical component but also that space is not the only proximity dimension at stake. Moreover, there is an indication that knowledge flows can be differentiated according to the medium used for their transfer.

Our paper intends to contribute exactly on these final suggestions by investigating the differential effects of proximity across different types of knowledge flows. We, therefore, extend the analysis of knowledge spillovers so as to consider cooperative relationships among inventors and their relationship with formal patent applicants (most often firms), besides citations as proxies of cross-regional knowledge flows. The paper's contribution to the field is three-fold. First, we compare three indicators of knowledge flows across regions in Europe in the last decade, *i*.e. citations, applicant-inventor links and co-inventorships, in order to ascertain if knowledge flows are all alike in terms of their dependence on geographical distance and contiguity. Secondly, we provide evidence of the moderating role of technological proximity on the effect of physical distance. Thirdly, we investigate the differential patterns of inter-national vs intra-national flows and knowledge exchanges among core and peripheral regions.

Our results indicate that these indicators show different responses to proximity, citations being less dependent on physical contiguity than co-inventorships. On the contrary, when one considers the role of technological proximity, citations appear to be more sensitive than co-inventorship. The applicant-inventor relationship always appears as an intermediate phenomenon. These different patterns can be explained by noticing that co-inventorships concern mainly the exchange of tacit knowledge, while citations are more likely to involve the flow of codified knowledge.

The paper is structured as follows. In section 2 we present and discuss our theoretical and empirical background. Section 3 describes the dataset, the variables and the methodology. In section 4 we present the results of econometric estimations, while in the final section we conclude with some policy implications.

2 Knowledge Flows, Proximity and Border Effects

According to the conventional Marshallian tradition (Meade, 1952; Viner, 1932), knowledge spillovers are qualified as 'untraded' interdependencies among firms. Knowledge generated by a given firm is an unpaid factor (*i*.e. an externality) that enters the production and innovation processes of other firms by means of accidental effects of co-location and spontaneous learning. Knowledge spills over and engenders positive externalities essentially due to its non-exclusive and non-rival use (Arrow, 1962).

Systemic approaches to innovation activities depict the generation of technological knowledge as an outcome of a collective undertaking strongly influenced by the availability of external sources of knowledge and by the way in which interactions are organized and carried out (Allen, 1983; von Hippel, 1988; Lundvall, 1992; Nelson, 1993). Internal and external knowledge inputs are so complementary that too low levels of each of them can hinder the entire knowledge production process (Antonelli, 1999). The intentional participation of firms to organized knowledge exchanges favours the acquisition of knowledge sourced externally in other firms and institutions (Dicken and Malmberg, 2001; Nicholas, 2009).

The collective and interactive dimension of technological knowledge raises the issue of proximity of innovating agents (Foray, 2004). A wide body of literature has shown that knowledge spillovers tend to be geographically clustered, and firms are likely to base their location choices on the opportunities of taking advantages of the positive feedbacks associated to co-location with other innovative actors (Audretsch and Feldman, 1996; Baptista and Swann, 1998). The spatial concentration applies, above all, when informal rather than formal cooperation ties are at work (Audretsch and Stephan, 1996). Feldman (1994a and 1994b) argues that co-location mitigates the inherent uncertainty of innovative activity: proximity enhances the ability of firms to exchange ideas and be aware of important incipient knowledge. Social and institutional ties, localized accumulation of labor, capital and R&D are the main requirements for knowledge spillovers and spontaneous learning from external sources to take place[1], and to exert an unconditional positive effect on output and productivity growth (Dekle, 2002; Dumais, Ellison and Glaeser, 2002; Rosenthal and Strange, 2003)[2].

In this context, the distinction between tacit and codified knowledge is especially relevant. Definitions of tacit knowledge often recall the well-known Polanyi's quotation according to which people know more than they can tell. In this sense, tacit knowledge is

[1] On this point, we should remember the distinction between unintended and intended spillovers (Maggioni et al., 2007): in the latter case, knowledge may flow among agents on a voluntary basis thanks to formal or informal agreements. Moreover, such exchanges can be either market or non-market mediated (Breschi and Lissoni, 2001) and in the former case take the form of pecuniary externalities (Antonelli et al., 2011?)

[2] Acknowledging that knowledge spillovers are important sources of increasing returns which tend to be geographically clustered, does not provide any assessment of the mechanisms by which externalities show up. The literature usually distinguishes between Marshall-Arrow-Romer (MAR), Jacobs and Porter externalities. Digging into this theoretical issue goes beyond the purposes of the present work. The reader can find exhaustive review in Frenken et al. (2007), Audretsch and Feldman (2004) and Basile and Usai (2014)

highly idiosyncratic and difficult to communicate. On the contrary, codified knowledge, thanks to a shared codebook that allows for coding and decoding, is better transmittable and understandable by people knowing the codebook. However, knowledge is not created codified. Codification is indeed the outcome of a process triggered by intentional efforts of innovating agents. In this perspective codified and tacit knowledge are not to be considered as discrete states, but rather as two extreme poles of a continuum (Saviotti, 1998; Cowan, David and Foray, 2000). An implication of the distinction between codified and tacit knowledge is that the marginal cost of transmitting codified knowledge across geographic space has been rendered more invariant with respect to geographical distance by the revolution in telecommunications. On the contrary, Von Hipple (1994) explains that most of economic agents' tacit knowledge is 'sticky'; *i*.e. highly contextual and uncertain and concludes that it is best transmitted via face-to-face interaction and through frequent and repeated contact (Steinmuller, 2000).

While the Economics of knowledge literature stresses the bearing of the distinction between tacit and codified knowledge upon the sensitivity of knowledge flows to distance, the New Economic Geography approach (NEG henceforth) emphasizes the difference between core and peripheral regions. According to this stream of literature, a reduction in trade costs leads to catastrophic agglomeration (Krugman, 1991). In this direction, trade liberalization could affect the core-periphery configuration insofar as it allows a decrease in trade costs between border regions (Krugman and Livas, 1996; Monfort and Nicolini, 2000; Paluzie, 2001). In a context such as the European one, characterized by gradual enlargement over time, border regions could take advantage of the possibility to build systematic exchanges with neighbor regions in close countries at lower costs. Although the evidence is not conclusive, Lafourcade and Paluzie (2011) show that, actually, border regions of core areas have obtained trade advantages from the integration process, as compared to other border regions. Moreover, border regions show better performances in cross-country trade exchanges than interior regions. However, the issue of cross-country patterns of exchanges is important not only as far as the flows of goods are concerned. Knowledge flows can, as well, be characterized by differential patterns in border and in interior regions. Border regions of core areas should show better performance than interior regions, especially when the exchange of tacit knowledge is at stake.

The last stream of literature which supports our analytical framework has been started by the so called French School of Proximity which claims that geographical proximity is neither necessary nor sufficient for knowledge spillovers and that a separate role for a-spatial links among economic entities is possible (see Carrincazeaux and Coris, 2011). Such links have been classified by Boschma (2005) into five dimensions of proximity across agents: geographical, institutional, technological (or cognitive), social (or relational) and organizational. Several recent works have proved the relative importance of a-spatial dimensions on either economic performance (Basile et al., 2012) or on innovative activity (Marrocu et al., 2013).

In view of the arguments elaborated so far, we can now spell out our working hypotheses as it follows.

Hypothesis 1. Knowledge flows are affected by multi-faceted proximity. However, knowledge flows are not all alike and the diverse kinds of proximity have, consequently, differential impacts. Citations and co-inventorship may be thought as standing at two poles of a continuum marked by codified and tacit knowledge respectively. In this direction, co-inventorship is expected to be more sensitive to geographical proximity than citations, while the latter are expected to be more sensitive to technological proximity than the former.

Hypothesis 2. Being near an international border implies international contiguity which creates a better environment for knowledge exchanges with other regions in nearby countries. We therefore expect that inner regions, in countries which share a border with other countries, are less prone to exploit knowledge flows than border regions.

3 Data, Methodology and Variables

3.1 The dataset

In order to obtain information on citation patterns, co-inventorship and applicant-inventor relationships, we use data extracted from the OECD REGPAT Database and the OECD Citations Database (January 2012). The former database presents patent data that have been linked to regions utilizing the addresses of the applicants and inventors. Two main dataset are covered by REGPAT: patent applications filed to the European Patent Office (EPO) and patent applications filed under the Patent Cooperation Treaty (PCT) at international phase. The OECD Citations database provides information on patent citations found in patent applications filed directly to the EPO or via the PCT. The geographical coverage relates to 276 NUTS2 regions located in 29 European countries (the EU-27 countries plus Norway and Switzerland)[3]. The reference period is the priority year: since it corresponds to the first filing worldwide and it is considered the closest date to the invention.

The REGPAT database is used in order to build the inter-regional matrices on co-inventorships and applicant-inventor links, while this database has to be combined with the Citations database in order to build the matrix on citation flows made and received by each region. Patent applications of citing and cited patents are, as a matter of fact, linked to regions on the basis of inventors' address thanks to the information provided in the REGPAT

[3]Data on patents in the OECD REGPAT database (Maraut et al., 2008), provides information on inventive activity and its multiple dimensions (e.g. geographical location, technical and institutional origin, individuals and networks).

database. In case of multiple inventors, a proportional share is assigned to each region and, as a result, cells are not going to be made of integers.

It is important to emphasize that the majority of citations at EPO comes from patent examiners during their searches rather than from patent applicants and inventors (Criscuolo and Verspagen 2008). Nonetheless, since we aggregate citations to proxy knowledge interactions among regions rather than inventors' contacts, this issue becomes less crucial (Breschi and Lissoni 2006). In other words, we believe that, even though examiners play an essential role in the citation process at EPO, it is reasonable to assume that professionals in R&D laboratories know existing patents (that is public knowledge) in their fields.

As for collaborations in inventive activity we consider all those cases where patents have more than one inventor and they reside in different regions in Europe. For each patent, we first link each inventor's region to all the other regions of the same patent. To every pair of regions is then assigned a weight which is inversely proportional to the total number of pairs created for each patent. The final matrix is made of the sum of weights for all the regions pairs for all the patents considered.

Finally, as far as the relationship among applicants and inventors of the same patent is concerned, we consider those patent applications where at least one applicant and/or at least one inventor reside in different regions in Europe. In this case, patents are linked to regions by utilizing the addresses of the applicants and inventors. In case of multiple applicants and/or inventors, a proportional share is attributed. More detail on the construction of these two latter matrices can be found in Maggioni et al. (2011).

It is worth noting that the citation and the applicant-inventor matrices are bi-directional, that is the flow is different when we consider the region *i* as the origin and region *j* as the destination or vice versa. On the contrary, with the co-inventorship matrix flows are not bi-directional and therefore there is no difference between origin and destination regions.

This matrix is, in other words, symmetric. This implies that in the former two cases the number of observations is 75900 (276*275), whilst in the latter case we have to halve this number to obtain 37950.

Table 1 shows the countries included in the analysis, as well as some key figures on patent activities and collaboration patterns. There are a few aspects which are worth noting. As for patent intensity, that is the amount of patents per employee, the country with the highest values are Switzerland and Germany, whilst those with the lowest values are Romania, Portugal and Bulgaria. The ranking based on patent intensity is quite similar to the one which can be obtained by looking at knowledge links measured by our three indicators. Nonetheless, it is worth noting that least innovative countries have a relatively high intensity of cooperation due to the fact that their internal innovative background in terms of patents, inventors or innovative firms is rather empty.

>>> INSERT TABLE 1 ABOUT HERE <<<

3.2 The Methodology

In order to investigate the effects of the multidimensional aspects of proximity on different kinds of knowledge flows we implemented a traditional gravity model taking the following form:

$$K_{ijt} = \alpha X_{it-s}^{\beta} X_{jt-s}^{\gamma} d_{ij}^{-\delta} \tag{1}$$

The gravity model is widely used in work on bilateral trade between countries (see e.g. Anderson and van Wincoop, 2003; Costantini and Crespi, 2008; Lafourcade and Paluzie, 2013) as well as in the study of knowledge flows (see e.g.: Maurseth and Verspagen, 2002; Paci and Usai, 2009; Picci (2010), Maggioni et al. (2011); Montobbio and Sterzi, 2013; Cappelli and Montobbio, 2013).

In our setting the dependent variable, K_{ijt}, represents knowledge flows between region *i* and region *j* at time *t*. They are measured in three ways: citation flows, co-inventorship relations and inventor-applicant links, respectively. The variables X_{it-s} and X_{jt-s} represent the mass of the two regions in a previous period (*t-s*), which may have affected knowledge exchanges at time *t*. It is worth stressing that we do not focus exclusively on the inventive mass of the two regions as in other previous works. We rather identify other potential sources of attraction for knowledge flows. Therefore, besides the obvious measures of the stock of patent applications, we consider regional R&D expenditures, the employment level[4], population density and country dummies to consider other potential institutional or structural factors which are in common to regions within the same nation. Finally d_{ij} stands for the distance between regions *i* and *j*, and it is going to be differentiated in several ways to take into account different dimensions of proximity, both geographical and technological.

By taking logs of equation (1) we obtain the following specification of the full empirical model:

$$\log(K_{ijt}) = \log(\alpha) + \beta \log(X_{it-s}) + \gamma \log(X_{jt-s}) - \delta \log(d_{ij}) + \varphi_I + \varphi_J + \varepsilon_{ij} \qquad (2)$$

In this specification α, β, γ and δ are the vectors of the coefficients of interest, φ_I and φ_J are country dummies for each region of the pair, while ε_{ij} is the error term. We focus mainly on α and δ. The former represents, according to Buch et al. (2004), a crucial comparative element for assessing the impact of distance across time[5] and along different contexts. The latter is a measure of how important are knowledge flows among regions which are far away relative to those of close-by regions.

The model in Equation (2) can be estimated by using different econometric techniques. Most previous papers implement OLS estimates of the coefficients and we follow

[4] We also run estimations considering skilled labour force instead of employment levels with very similar results.
[5] Buch et al. (2004) prove that changes in distance costs are largely reflected in the constant term rather than in the distance coefficients. They show that a proportional fall in distance costs is consistent with constant distance coefficients.

this approach. In what follows we discuss in detail the variables used in the analysis and then show the results of the econometric estimations.

3.3 The Variables

3.3.1 Endogenous variables

In order to investigate the effects of the multifaceted dimensions of proximity on the different kinds of knowledge flows we use three different dependent variables: a) $Ln(coinv_{i,j})$ is the natural logarithm of co-inventorship collaboration between regions *i* and *j*, identified when, in a patent developed by more than one inventor, at least one co-inventor is resident in region *i* and at least one co-inventor is resident in region *j*; b) $Ln(appinv_{i,j})$ is the natural logarithm of applicant-inventor link between regions *i* and *j*, identified whenever a patent has (at least) one inventor in region *i* and one applicant (which is usually a firm) resident in another region *j*[6]; c) $Ln(cit_{i,j})$ is the natural logarithm citations link between regions *i* and *j*, which occurs when the citing patent has at least one inventor residing in the region *j* and the cited patent has at least one inventor residing in the region *i*.

These three different indicators can be thought as mapping onto different kinds of knowledge defined according to the tacit/codified distinction. As it is shown in Figure 1, one can imagine tacit and codified knowledge as two separate poles of a continuum. In this frame, citation links better proxy the flow of codified knowledge between two regions, while co-inventorship is mostly related to the exchange of tacit knowledge. The link applicant-inventor can instead be seen as a sort of intermediate collaboration form. Actually, applicants are usually companies[7], and the kind of link established between an applicant and an inventor is much more similar to an employer-employee relationship than to collaboration. However, a

[6] In some patents the applicant can be the inventor him/herself. This does not create any problem in this context, as in these cases inventor and applicant appear to belong to the same region and therefore they are not counted.
7 The case in which the applicant id is the same as the inventor id is not taken into account by definition.

successful innovation process requires not only skilled inventors, but also qualified employers able to screen and monitor inventors' activities. The sharing of some codified knowledge is therefore crucial. At the same time, the interactions between applicants and inventors are also sensitive to tacitness insofar as the invention leading to the patent application emerges as a specific and idiosyncratic outcome.

>>> INSERT FIGURE 1 ABOUT HERE <<<

Finally, the dependent variable is the log of average values in the period 2002-2004 of the three types of knowledge flows detailed above. All explanatory variables but the dummies are, on the contrary, calculated in a previous period, that is 1999-2001, in order to partially avoid potential endogeneity problems. By way of robustness check, we also run estimations with different lag specifications. We regress in particular the log of average values of knowledge flows in the period 2005-2007 against the average values of explanatory variables in the period 1999-2001 and against the average values of explanatory variables in the period 2002-2004.

3.3.2 Explanatory variables

The variety of dimensions related to proximity have been measured by a number of indicators. First of all, geographical distance ($geodist_{i,j}$) is measured by logarithm of the row-normalized distance between regions i and j. Secondly, we build up a contiguity matrix ($cont_{ij}$) between regions i and j. We further decomposed the contiguity measure so as to appreciate the difference between contiguity of regions belonging to the same country ($wtbrd_{ij}$) and contiguity of regions belonging to different countries ($crossbrd_{ij}$) (see Figure 2 for a synthesis). Finally, we follow Lafourcade and Paluzie (2011) and analyze whether border regions (usually peripheral regions) are better off than inner regions (usually core regions) in exchanging knowledge with neighbour countries. To do so, we calculate one more dummy variable, i.e. $inner_{ij}$ which is equal to 1 if regions i and j are not contiguous but belong to two contiguous countries, and 0 otherwise.

>>> INSERT FIGURE 2 ABOUT HERE <<<

Table 2 allows to grasp the magnitude of the former two distinct phenomena and to observe some very interesting facts. Applicant-inventor links (which are reported in the first four columns) in Germany, for example are mainly intra-national (83%), and consist of contiguous German regions for a significant quota (36%). Across-border links in Germany are, therefore, only 17% in contrast with the opposite case of Ireland where we find the highest quota of international links, equal to 88%. When we consider citations, in the middle of the table, we find that Germany is a much more international player with a quota of intra-national citations of 56% and of international ones of 44%. Amongst the most innovative countries, the one which shows the highest propensity to cross-citations with other countries is Switzerland with a quota of 83% (of which only 5% between contiguous regions). Other very open countries are those ones with a negligible number of patents, such as Romania and Bulgaria. These countries appear to be rather internationalized also with respect to co-inventorships, with quotas of 100%. It is worth noting that Germany again shows mainly nationwide networks (85% of inventors cooperations are within borders), whilst Switzerland is quite open with an equal distribution of intra and inter-national co-inventorships. In an intermediate position we find other important innovative countries, such as Sweden and Finland with a quota of international co-inventorships of 40% and 32%, respectively.

>>> INSERT TABLE 2 ABUT HERE <<<

As far as the other dimensions of proximity are concerned, we focus only on technological and institutional proximity and we dismiss organisational and social proximity for a twofold reasons. On the one hand, previous works (Maurseth and Verspagen, 200x and Paci and Usai, 2009,) have reported that the former two aspects are relatively important in determining flows of knowledge across regions. On the other hand, the latter two dimensions, social and organizational, are very difficult to capture at the regional level (Marrocu et al 2013).

$Techprox_{i,j}$ is the technological proximity between regions i and j. It draws upon Jaffe's cosine index (Jaffe, 1986 and 1989) and is based on the technological classes (technologies henceforth) to which patents are assigned[8].

First of all we computed a measure of proximity amongst all observed technologies in our sample of patents, *i*.e. the cosine index. Let P_{lk} = 1 if the patent k is assigned the technology l [l= 1, …, n], and 0 otherwise. The total number of patents assigned to technology $l$is $O_l = \sum_k P_{lk}$. Similarly, the total number of patents assigned to technology $m$is $O_m = \sum_k P_{mk}$. We can, thus, indicate the number of patents that are classified in both technological fields l and m as: $V_{lj} = \sum_k P_{lk} P_{mk}$. By applying this count of joint occurrences to all possible pairs of classification codes, we obtain a square symmetrical matrix of co-occurrences whose generic cell V_{lm} reports the number of patent documents classified in both technological fields l and m.

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

[8] See Strumsky et al., 2012, for a critical assessment of opportunities and shortcomings related to the use of technological classes in empirical analyses.

$$S_{lm} = \frac{\sum_{m=1}^{n} V_{lz} V_{mz}}{\sqrt{\sum_{m=1}^{n} V_{lz}^2}\sqrt{\sum_{m=1}^{n} V_{mz}^2}} \tag{3}$$

The idea behind the calculation of this index is that two technologies l and m are similar to the extent that they co-occur with a third technology *z*. Such measure is symmetric with respect to the direction linking technological classes, and it does not depend on the absolute size of technological field. The cosine index provides a measure of the similarity between two technological fields in terms of their mutual relationships with all the other fields. S_{lm} is the greater the more two technologies *l* and *m* co-occur with the same technologies. It is equal to one for pairs of technological fields with identical distribution of co-occurrences with all the other technological fields, while it goes to zero if vectors V_{lz} and V_{mz} are orthogonal (Breschi et al., 2003)[9].

Once the technology proximity index has been calculated, we can use it to measure the technological proximity amongst any pair of regions. Let $R_{l,i} = 1$ if the technology l is observed in region *i*, 0 otherwise. Similarly, let $R_{m,j}= 1$ if technology m is observed in region *j*, 0 otherwise. The technological proximity is obtained as follows:

$$TechProx_{i,j} = \frac{\sum_m \sum_l R_{l,i} R_{m,j} S_{l,m}}{N} \tag{4}$$

Where N is the number of technological classes observed in the two regions. The technological proximity amongst regions is defined as the weighted average of the proximity amongst the technologies observed in the two regions.

Institutional proximity is usually measured in a much simpler way (see Marrocu et al, 2013): a dummy which is equal to unity when region *i* and *j* belong to the same country or zero viceversa. In other words, the sharing of the same legal framework and common culture

9For Engelsman and van Raan (1991), this approach produces meaningful results particularly at a 'macro' level, *i*.e. for mapping the entire domain of technology. An alternative approach to calculating technological proximity can be found in Sorenson and Singh (2007).

is a proxy for institutional proximity. Such common background is bound to affect transaction costs and make knowledge exchange easier and less costly.

In line with gravity models, we also consider a number of phenomena which are meant to account for the masses of the two regions *i* and *j*. First of all, we include the population density (*dens*) of sampled regions, calculated as the ratio between the number of inhabitants and area (land use). Since we are focusing on knowledge flows, regions' attraction degree may depend on the local availability of human capital (*loghk*), which is the natural logarithm of people with tertiary education attainment. In the same vein we also include the natural logarithm of regional R&D expenditure (*logrdexp*) and of regional patent stock (*kcap*), which is the stock of patents calculated by applying the permanent inventory method to patent applications.

In Table A1 in appendix we provide a synthetic account of the variables used in the econometric estimations, as well as of the time period over which they have been calculated and the different data sources. We also report the descriptive statistics concerning both the endogenous and the explanatory variables.

>>> INSERT TABLE 3 ABOUT HERE <<<

Table 3 shows instead the Spearman correlation coefficients amongst the variables included in the empirical analysis.

4 Econometric results

In order to analyze the effects of the different dimensions of proximity, we have estimated a log-linear transformation of the gravity equation, as in equation (2). Firstly, we present the whole set of results for each type of knowledge flow estimation in table 5, 6 and 7. In each table we report a set of five models which starts with a basic estimation of the gravity model with only geographical distance and the controls' set for regional characteristics. The other models follow (in columns 2 to 4) with sequential complications of the explanatory set. Model 5 is our preferred model. Secondly, table 8 offers a summary of results where the computation of standardised coefficients allow a full and detailed comparison of the impact of the main explanatory factors of knowledge flows across the three different typologies.

Table 4 reports the estimations of knowledge flows measured by citation.

>>> INSERT TABLE 4 ABOUT HERE <<<

The first column shows the estimation of the baseline model, and the geographical distance coefficient is negative and significant, as expected. In column (2) we introduce contiguity in the empirical model as it is usually done in gravity models and we find a positive (and significant) sign associated with this indicator. This is in line with our expectations but it is worth noting that this inclusion makes the coefficient of geographical distance decrease (in absolute value) of about 23%. Column (3) reports the results after the inclusion of technological (*techprox*) and institutional proximity (*instprox*), which have both the expected positive and significant coefficient. Note that this inclusion lowers the impact of both geographical distance and contiguity, which are nevertheless still significant. Column (4) reports the results of the estimations where contiguity is disentangled in international and intra-national bordering. Both indicators of adjacency have positive and significant coefficients, although the impact of within-border contiguity is higher than that of cross-border one. Finally, in column 6 we complete our estimation by adding the dummy *inner*, which is positive but not significant which implies that being at the border or not does not affect the patterns of cross country citation flows.

Table 5 shows the same pattern of results of the estimations in which the link applicant-inventor is the dependent variable. Results are analogous: the coefficient on distance is negative and significant as expected (column (1)) and the inclusion of contiguity to the baseline model (column (2)) makes the coefficient of distance decrease appreciably (-

30%). Column (3) includes technological and institutional proximity, both showing a positive and significant coefficient. It is worth noting that in this case the coefficient of institutional proximity (*instprox*) is three times the one of technological proximity. Column (4) shows the estimations including the within-border and cross-border contiguity where, as expected, the coefficient of the former is far higher than that of the latter. In column (5) we include the variable *inner*, which is now negative and significant. Since applicant-inventor relationships involve more tacit exchanges than citations, being or not a border region matters: cross-border regions are better off in international knowledge flows than inner regions.

>>> INSERT TABLE 5 ABOUT HERE <<<

Finally, Table 6 shows the results of estimations aiming at assessing the impact of proximities on the last type of knowledge flow: co-inventorships. As for the previous estimations, column (1) reports the baseline specification in which only geographical distance is taken into account. The coefficient is negative and significant, as expected and its effect is halved when contiguity is included in column (2)

.

>>> INSERT TABLE 6 ABOUT HERE <<<

In column (3) we introduce technological and institutional proximity which have positive and significant coefficients with the latter having a prevailing impact with respect to the former. In column (4) we dig into the differences between cross-border and within-border contiguity, by obtaining results consistent with the previous estimates, *i*.e. suggesting that the latter has a higher impact than the former. We finally include the dummy *inner* in column (6), which is also in this case negative and significant, suggesting that border regions are better off than inner regions when international co-inventorship links, with their tacit content, are at work.

Table 7 reports results by showing standardised coefficients, which allow for direct comparison of the impact of our main determinants on knowledge flows across estimations with different dependent variables. We replicate only the final model, that is our preferred estimation of columns (6) of the tables above, to assess the differential impact of geographical distance, contiguity (in its three different qualifications), technological and institutional proximities. We do not show the coefficients for the control variables which are not the main focus of our analysis.

>>> INSERT TABLE 7 ABOUT HERE <<<

Results comply with our conceptual framework and therefore with our expectations. In particular, we find that in the first model the standardized coefficient of geographical distance is quite similar for the three types of flows. Results are different when contiguity is considered. Results on contiguity-related dummies show that the within-border and the cross-border contiguity yield more significant impacts on co-inventorship (column 3) than applicant-inventor links (column 2) and citation flows (column 1). A similar result is found for the *inner* variable which has no significant impact on citations, while it has a higher impact on co-inventorship than on applicant-inventor links. All in all, these results suggest that the higher the tacit content of knowledge flows, the more sensitive they are to contiguity.

The situation is instead reversed when we look at the coefficient of technological proximity, which is very similar for co-inventorship and applicant-inventor links, and higher for citation flows. The effect related to 'epistemic communities' makes therefore citation flows more sensitive to cognitive similarity. Finally, institutional proximity yields the highest impact on co-inventorship and the lowest one on citation flows. The difference between these two knowledge flows is more marked for what concerns institutional proximity than geographical distance.

4.1 Robustness check

The results discussed in the previous section show clear-cut patterns as far as the differential impact of multidimensional proximity on diverse kinds of knowledge flows are concerned. An interesting issue concerns the robustness of this evidence to different lag specifications. The result are shown in Table 8.

INSERT TABLE 8 ABOUT HERE

In columns (1) to (3) of Table 8 we report the standardized coefficients obtained by running the estimations of the determinants of knowledge flows, these latter being calculated as an average over the period 2005-2007. The coefficients for geographical distance change only marginally: now applicant-inventor links clearly show the lowest coefficient, followed by citations and then co-inventorship. Technological proximity is basically not affected, and the ranking across the different kinds of knowledge flows does not change. The same applies to institutional proximity. As far as the contiguity-related dummies are concerned, we still find that the coefficients for co-inventorship are higher than those for applicant-inventor links, and in turn than those of citations. It is worth noting that for each dummy and each kind of knowledge flow the coefficients are slightly higher than the estimations carried out on the dependent variable calculated over the period 2002-2004. This applies also to the *inner* dummy, which is not significant for citations, but it is higher for co-inventorship than applicant-inventor links.

These estimations provide evidence of the determinants of knowledge flows when a longer lag is admitted between dependent and exogenous variables. The last three columns of Table 8 reports the results obtained by regressing knowledge flows over the period 2005-2007 against exogenous variables calculated over the period 2002-2004, i.e. by reproducing the same lag structure as the baseline estimations. The results are overall in line with those of the previous estimations, suggesting that both envisaged relationships between the variables and differences amongst the diverse knowledge flows are fairly robust to different specifications.

5 Conclusions and Policy Implications

Knowledge flows are not all alike. This is the answer to our main question, based on an empirical test which has assessed the functioning of three types of knowledge flows: citations, applicant-inventor links and co-inventorships. More specifically, the estimation of a set of gravity models show that knowledge flows are affected by contiguity and proximities to different extents. We prove that, depending on the content of tacitness entailed in the knowledge flow, physical distance and more precisely contiguity may play a very different role. The highest impact of contiguity (both within and across countries) is registered, as a matter of fact, for co-inventorship collaborations, that is those flows which are essentially based on tacit knowledge, cooperation and trust and are facilitated by face to face contacts. Consequently, facial contacts, and therefore contiguity, are less important for applicant-inventors links and are the least important for citations flows, since they are less dependent on personal contacts.

Sharing the same institutional context has also a diverse impact on knowledge flows as it is more important for collaborations among inventors and for applicant-inventors relationships whilst it is relatively less important for citations links. The rationale for this result is that in the former two cases a common institutional framework reduces the uncertainty and makes exchanges among economic agents less risky. The effect of contiguity and of institutional closeness are associated when we discriminate between contiguous regions within the same country and those which share an international border. As expected, being contiguous within national borders implies a stronger impact with respect to the case of contiguity across borders.

Knowledge flows which happen thanks to citation links are, on the contrary, more influenced by technological relatedness than the other two knowledge flows. This confirms that some elements of knowledge flow more easily within epistemic communities which share codified knowledge thanks to some rules for knowledge diffusion and they convey messages to whatever distance and independently from contiguity (see Breschi and Lissoni, 2001).

Finally, international border regions are shown to have an advantage with respect to other regions within the same country which are not on the border. This implies that bordering regions can emerge as more central thanks to their cross-border nature, and this effect may counteract at least partially the diseconomies due to peripherality.

Since knowledge flows are diverse and based on different behaviours and relationships among actors, policies aimed at knowledge diffusion have to take such differences into account. In other words, policies should balance their action by considering all potential moderating factors of the geographical distance which is not the only dimension at work.

References

Allen, R.C., (1983). "Collective invention," *Journal of Economic Behavior & Organization*, 4(1), 1-24,

Anderson J. E. and van Wincoop E., (2004) Trade Costs, *Journal of Economic Literature*, 42, 691–751.

Antonelli C., Patrucco P.P. and Quatraro F. (2011) Productivity Growth and Pecuniary Knowledge Externalities: An Empirical Analysis of Agglomeration Economies in European Regions, *Economic Geography*, 87, 23-50.

Antonelli, C. (1999). *The microdynamics of technological change*. London: Routledge.

Arrow K.J., (1962) The Economic Implications of Learning by Doing, *Review of Economic Studies*, 29, pp. 155-173.

Arthur W.B. (1989) Competing Technologies, Increasing Returns, and Lock-In by Historical Events, *Economic Journal*, 99, No. 394., pp. 116-131.

Audretsch D.B. and Feldman M.P., R&D spillovers and the geography of innovation and production, *American Economic Review*, (1996) 86, 630–640.

Audretsch D.B. and Feldman M.P. (2004) Knowledge Spillovers and the Geography of Innovation, in Henderson J.V. and J.F. Thisse (eds.) *Handbook of Urban and Regional Economics*, 4, pp.2713-2739. Elsevier, Amsterdam

Audretsch, D.B. and Stephan P., (1996), Company-scientist locational links: the case of biotechnology, *American Economic Review* 86(4): 641-652

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

Basile R. and Usai S., (2012). "Analysis of regional endogenous growth," *Working Paper CRENoS* 201211.

Basile R., Capello R. and Caragliu A. (2012) Technological interdependence and regional growth in Europe: Proximity and synergy in knowledge spillovers, *Papers in Regional Science*, 91, 697-722.

Belenzon S. and Schankerman M. (2013), Spreading the word: Geography, Policy and Knowledge Spillovers, *The Review of Economics and Statistics*, July 2013, 95(3): 884–903

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

Breschi, S. and Lissoni, F. (2001) Knowledge Spillovers and Local Innovation Systems: A Critical Survey, *Industrial and Corporate Change*, 10, 975-1005.

Buch, C.M., Kleinert, J. and Toubal, F., (2004). The distance puzzle: on the interpretation of the distance coefficient in gravity equations, *Economics Letters*, 83(3), 293-298

Cappelli, R. and Montobbio, F., (2013). "European Integration and Knowledge Flows across European Regions," *Department Cognetti de Martiis Working Papers* 201322,

Carrincazeaux C. and Coris M., (2011) Proximity and Innovation, in Cooke P., Asheim B.T. and Boschma R. (eds) *Handbook of Regional Innovation and Growth*. Cheltenham: Edward Elgar.

Costantini, V. and Crespi, F., (2008). "Environmental regulation and the export dynamics of energy technologies," *Ecological Economics*, 66(2-3), 447-460,

Cowan, R., David P. A. and Foray, D., (2000). "The Explicit Economics of Knowledge Codification and Tacitness," *Industrial and Corporate Change*, 9(2), 211-53.

Criscuolo, P. and Verspagen, B., 2008. "Does it matter where patent citations come from? Inventor vs. examiner citations in European patents," *Research Policy*, 37(10), 1892-1908.

Dekle R., (2002), "Industrial Concentration and Regional Growth: Evidence From The Prefectures," *The Review of Economics and Statistics*, 84(2), 310-315,

Dicken, P. and Malmberg, A. (2001) Firms in territories: a relational perspective, *Economic Geography*

Dumais G., Ellison G. and Glaeser E.L., (2002). "Geographic Concentration As A Dynamic Process," *The Review of Economics and Statistics*,. 84(2), 193-204,.

Engelsman, E. C., and Van Raan, A. F. J. (1991). *Mapping Technology. A first exploration of knowledge diffusion amongst fields of technology*. The Hague: Ministry of Economic Affairs.

Feldman, M.P. (1994a). *The Geography of Innovation,* Kluwer Academic Publishers, Boston.

Feldman, M.P. (1994b), "Knowledge complementarity and innovation, Small Business Economics 6(3): 363-372.

Foray, D. (2004). *The economics of knowledge*. Cambridge, Mass.: MIT Press

Frenken K., Van Oort F and Verburg T., (2007). "Related Variety, Unrelated Variety and Regional Economic Growth," *Regional Studies*, 41(5), 685-697

Giuri P. and Mariani M., (2013). "When Distance Disappears: Inventors, Education, and the Locus of Knowledge Spillovers," *The Review of Economics and Statistics*, 95(2), 449-463,

Grossman G.M and Helpman E., (1990) Trade, Innovation, and Growth, *American Economic Review*, 80, 86-91.

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

Jaffe A.B. (1986) Technological Opportunity and Spillovers of R&D: evidence from Firms' Patents, Profits and Market Value, *American Economic Review*, 76, 984-1001.

Jaffe A.B., Trajtenberg M. and Henderson R. (1993) Geographic localization of knowledge spillovers as evidenced by patient citations, *Quarterly Journal of Economics*, 108, 577-598

Jaffe, A.B, Fogarty M.S and Banks B.A., (1998) "Evidence from Patents and Patent Citations on the Impact of NASA and Other Federal Labs on Commercial Innovation," Journal of Industrial Economics, 46(2), 183-205,.

Krugman, P. and Elizondo R.L., (1996). "Trade policy and the Third World metropolis," Journal of Development Economics, 49(1), 137-150,.

Krugman, P. (1991), Increasing Returns and Economic Geography, *Journal of Political Economy*, 99, 483-99

Lafourcade M. and Paluzie E., (2011). "European Integration, Foreign Direct Investment (FDI), and the Geography of French Trade," *Regional Studies*, 45(4), 419-439.

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

Maggioni M.A., Nosvelli, M. and Uberti, T.E. (2007) Space versus networks in the geography of innovation: a European analysis, *Papers in Regional Science*, 86, 471-493.

Maggioni M.A., Uberti T.E. and Usai S. (2011) Treating patents as relational data: knowledge transfers and spillovers across Italian provinces, *Industry & Innovation*, 18, 39-67.

Maraut S., Dernis H., Webb C., Spiezia V and Guellec D., (2008). "The OECD REGPAT Database: A Presentation," *OECD STI Working Papers* 2008/2

Marrocu E., Paci R. and Usai S. (2013) Proximity, Networking and Knowledge Production in Europe: what lessons for innovation policy? *Technological Forecasting and Social Change,* 80 (2013), pp. 1484-1498,

Maurseth, P. B. and Verspagen, B., (2002). " Knowledge Spillovers in Europe: A Patent Citations Analysis," *Scandinavian Journal of Economics*, 104(4), 531-45,.

Meade J., (1952), *A Geometry of International Trade*, London, George Allen & Unwin, 1952.

Monfort, P. and Nicolini, R., (2000). "Regional Convergence and International Integration," *Journal of Urban Economics*, 48(2), 286-306,

Montobbio F. and Sterzi V. (2013) The globalization of technology in emerging markets: a gravity model on the determinants of international patent collaborations, forthcoming in *World Development*.

Nelson, R.R., (1993) *National Innovation Systems: A Comparative Analysis*. Oxford University Press

Nicholas, T. (2009). Spatial diversity in invention: evidence from the early R&D labs. *Journal of Economic Geography* 9: 1-31.

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

Paci R., Marrocu E. and Usai S. (2014) The complementary effects of proximity dimensions on knowledge spillovers, forthcoming in *Spatial Economic Analysis*.

Paluzie E., (2001). "Trade policy and regional inequalities," *Papers in Regional Science*, 80(1), 67-85.

Picci L. (2010) The internationalization of inventive activity: A gravity model using patent data, *Research Policy*, 39, 1070-1081.

Romer, P.M. (1986) Increasing Returns and Long-run Growth, *Journal of Political Economy*, 94, 1002-1037.

Rosenthal S.S. and Strange W.C:, (2003). "Geography, Industrial Organization, and Agglomeration," *The Review of Economics and Statistics*, 85(2), 377-393.

Santos Silva J.M.C. and Tenreyro S., (2006), The Log of Gravity, *The Review of Economics and Statistics*, 88(4): 641–658

Saviotti, P.P, (1998). "On the dynamics of appropriability, of tacit and of codified knowledge," *Research Policy*, 26(7-8), 843-856.

Sorenson O. and Singh J., (2007). "Science, Social Networks and Spillovers," Industry and Innovation, 14(2), 219-238.

Steinmueller, W E., (2000). "Will New Information and Communication Technologies Improve the 'Codification' of Knowledge?," *Industrial and Corporate Change*, 9(2), 361-76.

Strumsky, D., Lobo, J., Van der Leeuw, S. (2011) "Using Patent Technology Codes to Study Technological Change" *Economics of Innovation and New Technology*, 1-20

Thompson P., (2006). "Patent Citations and the Geography of Knowledge Spillovers: Evidence from Inventor- and Examiner-added Citations," *The Review of Economics and Statistics*, 88(2), 383-388,

Thompson P. and Fox-Kean M., (2005). "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment," *American Economic Review* 95(1), 450-460,

Viner, J. (1931). Costs curves and supply curve. *Zeitschrift für Nationalokonomie* 3: 23-46.

Von Hippel E. (1988), *The Sources of Innovation*, Oxford University Press

Von Hippel E. (1994) "Sticky Information" and the Locus of Problem Solving: Implications for Innovation, *Management Science*, 40, 429-439

Table 1–Patenting activity in the sampled countries (2001-2004)

Country	Patents	Collaborative Patents	Applicant-Inventors	Citation flows	Employees	Patent Intensity
	(a)	(b)	(c)	(d)	(e)	(1000*a/e)
AT	4038	162	425	8502	11250	359
BE	4104	248	627	8635	12279	334
BG	53	2	4	45	8498	6
CH	8391	384	1895	28764	11887	706
CZ	312	24	45	85	14123	22
DE	66111	2857	8892	175312	107665	614
DK	3017	98	287	3847	8169	369
ES	3079	68	136	3120	51897	59
FI	3883	86	555	5142	7102	547
FR	23504	645	2874	56923	74837	314
GR	225	8	9	154	12763	18
HU	404	20	28	603	11693	35
IE	681	18	147	659	5453	125
IT	13074	279	800	26486	66288	197
NL	10477	293	1779	17399	24395	430
NO	1112	50	152	1154	6800	164
PL	222	10	16	121	41193	5
PT	164	7	13	122	15378	11
RO	48	0	3	22	27849	2
SE	6222	160	995	10893	12928	481
SK	75	10	9	15	6453	12
UK	16166	672	2144	40702	84251	192

Table 2 – Knowledge flows in the sampled countries: within-country vs. cross-border patterns (percentage).

	Applicants in region *i* and inventor in any other region				Patents in region *i* citing a patent in any other region				Patents with inventors in region *i* and in any other region			
Country	Same country	of which contiguous	Cross-country	of which contiguous	Same country	of which contiguous	Cross-country	of which contiguous	Same country	of which contigous	Cross-country	of which contiguous
AT	40.1%	26.8%	59.9%	2.9%	9.5%	5.4%	90.5%	3.8%	48.2%	33.1%	51.8%	7.5%
BE	46.6%	27.6%	53.4%	1.1%	18.9%	12.1%	81.1%	1.4%	60.2%	42.7%	39.8%	4.2%
BG	38.5%	2.4%	61.5%	0.0%	0.1%	0.0%	99.9%	0.0%	0.0%	0.0%	100.0%	0.0%
CH	31.1%	25.3%	68.9%	14.2%	16.8%	12.4%	83.2%	5.1%	49.2%	41.0%	50.8%	20.1%
CZ	46.1%	31.4%	53.9%	0.0%	4.4%	1.8%	95.6%	0.2%	36.0%	21.3%	64.0%	2.4%
DE	82.7%	36.4%	17.3%	0.8%	56.4%	15.5%	43.6%	0.8%	85.1%	52.5%	14.9%	2.4%
DK	46.3%	24.8%	53.7%	0.2%	6.8%	4.3%	93.2%	0.2%	52.5%	38.2%	47.5%	1.2%
ES	51.7%	10.8%	48.3%	0.4%	6.5%	1.9%	93.5%	0.4%	28.7%	9.6%	71.3%	0.4%
FI	60.8%	47.8%	39.2%	0.0%	7.1%	5.5%	92.9%	0.0%	68.0%	56.1%	32.0%	0.0%
FR	73.5%	17.0%	26.5%	0.4%	24.5%	6.5%	75.5%	1.3%	68.7%	25.6%	31.3%	4.5%
GR	79.8%	23.3%	20.2%	0.0%	2.6%	0.6%	97.4%	0.0%	20.0%	1.4%	80.0%	0.0%
HU	50.6%	39.2%	49.4%	0.0%	2.4%	1.7%	97.6%	0.0%	47.5%	32.1%	52.5%	0.0%
IE	11.6%	11.6%	88.4%	0.1%	1.4%	1.4%	98.6%	0.1%	40.9%	40.9%	59.1%	0.0%
IT	81.1%	41.5%	18.9%	0.3%	22.6%	12.6%	77.4%	1.4%	66.0%	36.4%	34.0%	2.2%
NL	31.2%	22.9%	68.8%	1.6%	14.4%	9.2%	85.6%	2.1%	63.7%	45.9%	36.3%	1.9%
NO	48.3%	25.1%	51.7%	0.8%	5.4%	3.6%	94.6%	1.4%	50.9%	29.1%	49.1%	2.2%
PL	56.5%	12.4%	43.5%	0.0%	0.4%	0.1%	99.6%	0.0%	25.2%	7.9%	74.8%	0.0%
PT	38.7%	33.1%	61.3%	0.0%	3.3%	2.1%	96.7%	0.4%	30.6%	26.2%	69.4%	0.0%
RO	60.2%	16.0%	39.8%	39.8%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	100.0%	0.0%
SE	43.1%	16.5%	56.9%	0.2%	9.1%	4.8%	90.9%	0.2%	60.0%	35.0%	40.0%	0.1%
SK	26.4%	18.5%	73.6%	0.0%	1.6%	0.0%	98.4%	6.5%	16.7%	4.3%	83.3%	6.9%
UK	79.8%	27.7%	20.2%	0.0%	21.4%	8.0%	78.6%	0.0%	71.6%	34.5%	28.4%	0.0%

Table 3 - Spearman Correlation Coefficients

	lncit	lnappinv	lncoinv	geodist	techprox	samec	a_dens	a_HK	a_kcap	a_RDexp
lncit	1.0000									
lnappinv	0.4743*	1.0000								
lncoinv	0.5203*	0.6606*	1.0000							
geodist	-0.3649*	-0.3402*	-0.3850*	1.0000						
techprox	-0.3426*	-0.2239*	-0.2540*	0.1762*	1.0000					
instprox	0.2145*	0.4144*	0.4381*	-0.3901*	-0.0674*	1.0000				
a_dens	0.2773*	0.2472*	0.2475*	-0.2481*	-0.2267*	0.1022*	1.0000			
a_HK	0.2683*	0.2277*	0.2407*	-0.0606*	-0.2315*	0.0705*	0.5005*	1.0000		
a_kcap	0.4392*	0.3508*	0.3453*	-0.2623*	-0.3862*	0.0673*	0.5897*	0.5977*	1.0000	
a_RDexp	0.3934*	0.3303*	0.3359*	-0.2046*	-0.3541*	0.0774*	0.5857*	0.7212*	0.9053*	1.0000

Note : * $p < 0.05$.

Table 4 – Econometric Results. Dependent Variables ln(citations)

	(1)	(2)	(3)	(4)	(5)
geodist	-0.1049***	-0.0849***	-0.0529***	-0.0537***	-0.0536***
	(0.0029)	(0.0025)	(0.0029)	(0.0029)	(0.0032)
contig		0.2136***	0.1685***		
		(0.0260)	(0.0258)		
techprox			0.2165***	0.2165***	0.2165***
			(0.0179)	(0.0179)	(0.0179)
instprox			0.1606***	0.1556***	0.1558***
			(0.0111)	(0.0113)	(0.0112)
crsbrd				0.1155***	0.1157***
				(0.0343)	(0.0344)
wtnbrd				0.1875***	0.1876***
				(0.0325)	(0.0325)
inner					0.0003
					(0.0051)
a_dens	-0.0243***	-0.0228***	-0.0229***	-0.0229***	-0.0229***
	(0.0014)	(0.0014)	(0.0014)	(0.0014)	(0.0014)
a_HK	-0.0019	-0.0042	-0.0008	-0.0007	-0.0007
	(0.0028)	(0.0028)	(0.0029)	(0.0029)	(0.0029)
a_kcap	0.0724***	0.0732***	0.0790***	0.0790***	0.0790***
	(0.0020)	(0.0020)	(0.0021)	(0.0021)	(0.0021)
a_RDexp	-0.0011	-0.0012	-0.0036*	-0.0037*	-0.0037*
	(0.0019)	(0.0019)	(0.0020)	(0.0020)	(0.0020)
b_dens	-0.0186***	-0.0171***	-0.0171***	-0.0172***	-0.0172***
	(0.0014)	(0.0014)	(0.0014)	(0.0014)	(0.0014)
b_HK	-0.0149***	-0.0171***	-0.0132***	-0.0131***	-0.0131***
	(0.0028)	(0.0028)	(0.0030)	(0.0030)	(0.0029)
b_kcap	0.0768***	0.0776***	0.0836***	0.0836***	0.0836***
	(0.0020)	(0.0020)	(0.0020)	(0.0020)	(0.0020)
b_RDexp	0.0058***	0.0057***	0.0029	0.0029	0.0029
	(0.0020)	(0.0019)	(0.0020)	(0.0020)	(0.0020)
_cons	1.0414***	0.7644***	0.1230***	0.1360***	0.1347**
	(0.0445)	(0.0398)	(0.0473)	(0.0469)	(0.0524)
N	75900	75900	74256	74256	74256
R^2	0.369	0.374	0.390	0.390	0.390
adj. R^2	0.368	0.373	0.390	0.390	0.390
AIC	11220.3217	10560.4357	9819.7130	9805.5100	9807.5018
BIC	11829.9750	11179.3262	10455.5669	10450.5791	10461.7862

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5 – Econometric Results. Dependent Variables ln(AppInv)

	(1)	(2)	(3)	(4)	(5)
geodist	-0.1755***	-0.1231***	-0.0399***	-0.0437***	-0.0471***
	(0.0046)	(0.0034)	(0.0031)	(0.0031)	(0.0033)
contig		0.5609***	0.4427***		
		(0.0377)	(0.0364)		
techprox			0.1373***	0.1372***	0.1371***
			(0.0165)	(0.0164)	(0.0164)
instprox			0.4226***	0.4002***	0.3937***
			(0.0134)	(0.0136)	(0.0134)
crsbrd				0.2051***	0.1958***
				(0.0405)	(0.0404)
wtnbrd				0.5280***	0.5247***
				(0.0465)	(0.0464)
inner					-0.0109**
					(0.0042)
a_dens	0.0070***	0.0109***	0.0126***	0.0124***	0.0123***
	(0.0019)	(0.0019)	(0.0018)	(0.0018)	(0.0018)
a_HK	-0.0044	-0.0104***	-0.0121***	-0.0117***	-0.0113***
	(0.0032)	(0.0031)	(0.0031)	(0.0031)	(0.0031)
a_kcap	0.0423***	0.0444***	0.0528***	0.0528***	0.0525***
	(0.0021)	(0.0020)	(0.0021)	(0.0021)	(0.0021)
a_RDexp	0.0096***	0.0094***	0.0057***	0.0055***	0.0055***
	(0.0021)	(0.0020)	(0.0021)	(0.0021)	(0.0021)
b_dens	-0.0056***	-0.0017	-0.0003	-0.0005	-0.0006
	(0.0016)	(0.0016)	(0.0016)	(0.0016)	(0.0016)
b_HK	0.0151***	0.0091***	0.0054*	0.0059*	0.0062*
	(0.0033)	(0.0033)	(0.0032)	(0.0032)	(0.0032)
b_kcap	0.0202***	0.0224***	0.0298***	0.0297***	0.0295***
	(0.0020)	(0.0020)	(0.0020)	(0.0020)	(0.0020)
b_RDexp	0.0089***	0.0087***	0.0068***	0.0065***	0.0065***
	(0.0023)	(0.0022)	(0.0023)	(0.0023)	(0.0023)
_cons	2.0959***	1.3688***	0.1222**	0.1806***	0.2280***
	(0.0660)	(0.0525)	(0.0567)	(0.0560)	(0.0585)
N	75900	75900	74256	74256	74256
R^2	0.246	0.282	0.346	0.348	0.349
adj. R^2	0.245	0.281	0.345	0.348	0.348
AIC	28569.3958	24867.8594	18907.7047	18621.3303	18613.7624
BIC	29179.0491	25486.7499	19543.5586	19266.3995	19268.0469

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6 – Econometric Results. Dependent Variables ln(Coinv)

	(1)	(2)	(3)	(4)	(5)
geodist	-0.1433***	-0.0772***	-0.0318***	-0.0348***	-0.0457***
	(0.0048)	(0.0029)	(0.0026)	(0.0026)	(0.0028)
contig		0.6618***	0.5582***		
		(0.0384)	(0.0381)		
techprox			0.0995***	0.0994***	0.0995***
			(0.0138)	(0.0137)	(0.0137)
instprox			0.3702***	0.3475***	0.3180***
			(0.0138)	(0.0132)	(0.0133)
crsbrd				0.3379***	0.3041***
				(0.0569)	(0.0560)
wtnbrd				0.6395***	0.6320***
				(0.0470)	(0.0466)
inner					-0.0394***
					(0.0039)
a_dens	0.0007	0.0051***	0.0051***	0.0050***	0.0049***
	(0.0018)	(0.0017)	(0.0017)	(0.0016)	(0.0016)
a_HK	-0.0068**	-0.0135***	-0.0075**	-0.0072**	-0.0065**
	(0.0031)	(0.0029)	(0.0029)	(0.0029)	(0.0029)
a_kcap	0.0215***	0.0244***	0.0240***	0.0242***	0.0237***
	(0.0017)	(0.0017)	(0.0017)	(0.0017)	(0.0017)
a_RDexp	0.0167***	0.0151***	0.0130***	0.0125***	0.0127***
	(0.0025)	(0.0023)	(0.0023)	(0.0023)	(0.0023)
b_dens	-0.0107***	-0.0057***	-0.0055***	-0.0058***	-0.0062***
	(0.0016)	(0.0015)	(0.0015)	(0.0015)	(0.0015)
b_HK	0.0155***	0.0073**	-0.0014	-0.0010	0.0004
	(0.0035)	(0.0033)	(0.0033)	(0.0033)	(0.0033)
b_kcap	0.0192***	0.0220***	0.0324***	0.0322***	0.0313***
	(0.0022)	(0.0021)	(0.0021)	(0.0021)	(0.0021)
b_RDexp	0.0046**	0.0056***	0.0050**	0.0049**	0.0049**
	(0.0021)	(0.0020)	(0.0020)	(0.0020)	(0.0020)
_cons	2.1746***	1.0965***	0.0739	0.1039	0.2656**
	(0.1202)	(0.1074)	(0.1112)	(0.1106)	(0.1111)
N	37950	37950	37128	37128	37128
R^2	0.320	0.404	0.466	0.470	0.472
adj. R^2	0.319	0.403	0.465	0.469	0.471
AIC	-10024.2835	-15018.8554	-18024.7038	-18285.5194	-18395.1052
BIC	-9460.3778	-14446.4057	-17436.6770	-17688.9705	-17790.0342

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7 – Standardized Coefficients

	Baseline specification		
	Ln(Cit)	Ln(AppInv)	Ln(Coinv)
geodist	-0.114***	-0.097***	-0.124***
techprox	0.056***	0.035***	0.033***
instprox	0.118***	0.291***	0.308***
crsbrd	0.022***	0.036***	0.074***
wtnbrd	0.063***	0.172***	0.271***
inner	0.000	-0.012**	-0.056***
N	74256	74256	37128

Standardized beta coefficients;
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8 – Robustness check. Comparison amongst different lag specifications.

	Dep.Var. Av. Val. 2005-2007			Dep. Var.: Av. Val. 2005-2007 – Explanatory vars: Av. Val. 2002-2004		
	Ln(Cit)	Ln(AppInv)	Ln(Coinv)	Ln(Cit)	Ln(AppInv)	Ln(Coinv)
geodist	-0.115***	-0.076***	-0.103***	-0.113***	-0.073***	-0.102***
techprox	0.050***	0.036***	0.034***	0.050***	0.037***	0.034***
instprox	0.123***	0.278***	0.306***	0.124***	0.279***	0.306***
crsbrd	0.029***	0.049***	0.064***	0.030***	0.049***	0.064***
wtnbrd	0.069***	0.197***	0.292***	0.070***	0.197***	0.292***
inner	0.001	-0.020***	-0.074***	0.002	-0.019***	-0.074***
N	74256	74256	37128	74256	74256	37128

Standardized beta coefficients;
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 1 – Relationship between type of knowledge flows and knowledge tacitness degree

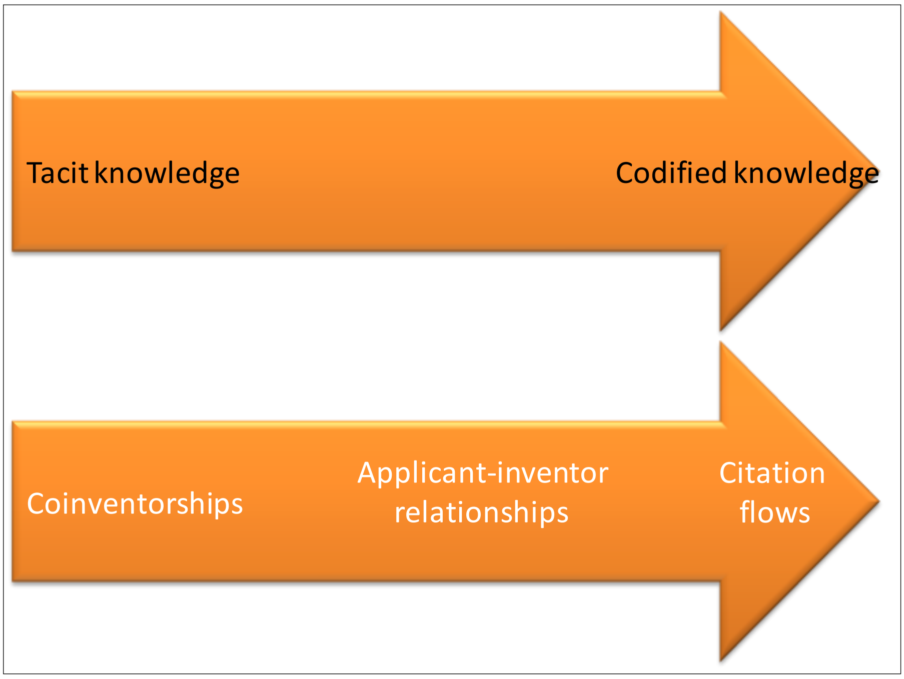

Figure 2 - Relevant Patterns of Knowledge Flows

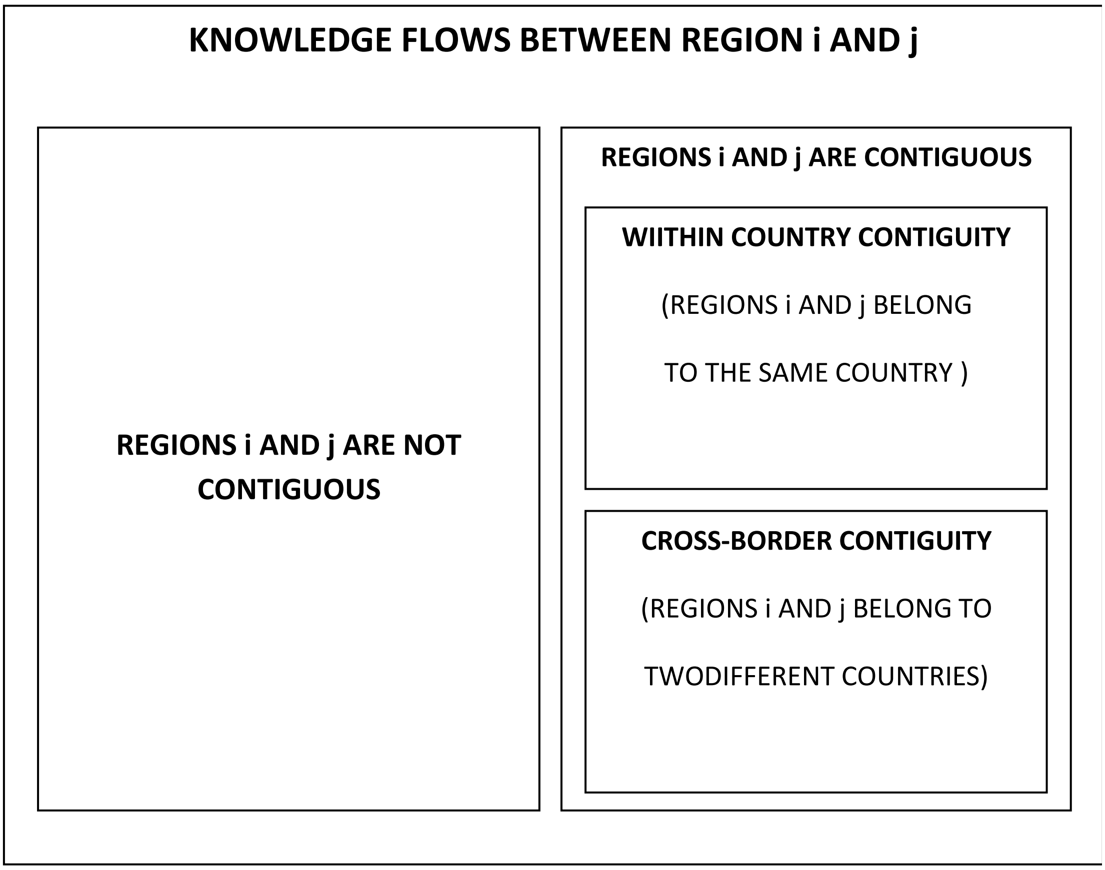

Table A1 – Definition of Variables and descriptive statistics

Variable	Definition	Source	N	Min	max	mean	sd	skewness	kurtosis
	Knowledge flows								
lncit	Natural logarithm of patent citations between region i and j (average value 2002-2004)	OECD-Regpat Database (Jan 2012)	75900	0,000	4,387	0,115	0,328	4,561	30,041
lnappinv	Natural logarithm of patents with applicant from region i and inventor from region j (average value 2002-2004)	OECD-Regpat Database (Jan 2012)	75900	0,000	6,103	0,087	0,336	6,350	57,657
lncoinv	Natural logarithm of patents with inventors in the region i and in the region j (average value 2002-2004)	OECD-Regpat Database (Jan 2012)	37950	0,000	4,933	0,070	0,257	7,050	72,675
	Proximities								
geodist	Distance (in kilometers)		75900	9,790	15,336	13,930	0,704	-1,036	4,374
techprox	Technological proximity between regions i and j, calculated on the basis of Jaffe's cosine index.	OECD-Regpat Database (Jan 2012)	74256	0,322	1,000	0,626	0,086	1,049	4,811
instprox	Dummy variable equal to 1 if regions i and j belong to the same country		75900	0,000	1,000	0,066	0,249	3,480	13,113
	Controls								
dens	Ratio between population and area (land use)	Cambridge Econometrics	75900	1,415	9,071	4,934	1,178	0,324	4,370
loghk	Natural logarithm of people with tertiary education attainment (average value 1999-2001)	Cambridge Econometrics	75900	1,352	7,549	4,742	0,914	-0,525	4,004
kcap	Natural logarithm of regional knowledge stock (average value 1999-2001)	OECD-Regpat Database (Jan 2012)	75900	0,000	10,813	5,427	2,445	-0,444	2,463
logrdexp	Natural logarithm of R&D expenditure (average value 1999-2001)	Cambridge Econometrics	75900	0,810	9,524	5,326	1,736	-0,445	2,624