European Integration and Knowledge Flows across European Regions

Abstract

Using data on inventor citations and inventor collaborations, this article analyses changes in geographical patterns of knowledge flows between European regions during the period 1981-2000. It shows that inventor collaborations become less geographically localized, while inventor citations become more localized. The European integration process has a significant effect on reducing barriers to knowledge flows between new and old EU members. For inventor citations, this effect relates only to the EU enlargement of 1995 and is confined to knowledge flows from Austria, Finland and Sweden to old EU members.

Keywords: Knowledge flows; European integration; Regional gravity model
JEL Classification: O31; R12; R15

Word count (excludind references and tables): 9387

Word count (including references and tables): 11795

1 Introduction

Catching up between different regions crucially depends upon the diffusion of technology. However, a lot of empirical evidence suggests that the diffusion of technology is constrained by geography and national borders. In fact, there are communication and learning costs because knowledge is rarely a public good, and, on the contrary, a relevant portion of it is tacit (or costly to be codified). In turn, it is widely accepted that innovation activity is characterized by substantial agglomeration effects and, possibly, increasing returns at the regional level (Grossman and Helpman, 1991; Keller, 2004; Krugman, 1991).

Understanding precisely the constraints to knowledge diffusion - and the right balance between centrifugal and centripetal forces - has major policy implications. In Europe it is very important to ask whether innovation policy is reinforcing these agglomeration effects and whether dissemination can favour economic convergence across European regions. Recently, the EU, in its Europe 2020 strategy (European Commission, 2010), underlines the importance of knowledge flows for achieving a “smart” economy. One of the goals of this strategy is to promote the diffusion of knowledge among member countries in order to develop an integrated European Research Area (ERA). As a consequence, it becomes relevant to understand whether a continuous process of reducing barriers that divide countries and a growing implementation of EU innovation policies is able to promote the diffusion of knowledge.

This paper studies whether knowledge diffuses in Europe using two different indicators of knowledge flows: patent citations and inventor collaborations. It compares two types of indicators to capture different characteristics of knowledge. Since tacit knowledge is costly to transfer and requires absorptive capacity, knowledge flows are facilitated by interpersonal links and face-to-face contacts (e.g. Keller and Yeaple, 2009; Montobbio and Sterzi, 2013). This diffusion mechanism is captured using inventor collaborations. This paper compares the geographical patterns of collaborations with the geographical scope of patent citations that measure the codified component of the knowledge.

The existing literature on patent citations in Europe (Maurseth and Verspagen, 2002; Paci and Usai, 2009) already shows that there are significant barriers preventing knowledge from flowing freely across regional and national borders in Europe. This work builds on these papers and ask whether Europe is becoming more integrated in the field of knowledge, i.e. whether the diffusion of knowledge is becoming less constrained by geography and national borders. Over time, decreased transport costs, technological advances and diffusion of ICT and the greater (commercial and political) integration among countries have eased the exchange of knowledge over long distances (see for example the “death of distance” argument in Cairncross, 1997). In Europe this effect could be stronger assuming that the integration process contributes to decrease communication and transport costs facilitating the international diffusion of knowledge.

So this paper is focused on effect of the EU enlargement. Empirical evidence shows that the EU integration has affected trade and factors flows (see, e.g., Carrère, 2006; Brenton et al., 1999; Bauer and Zimmerman, 1999). This paper goes further, and considers the hypothesis that an EU enlargement process facilites the exchange of knowledge between the regions already member of the EU before the enlargement and the EU entering regions. This paper covers the processes of European integration from 1981 to 2000, when two processes of enlargement expanded the EU from 10 to 15 members.

This paper uses EPO data to analyse the evolution over time of knowledge flows between 191 NUTS2 regions of 21 European countries, including both EU and not EU member states, for the period 1981-2000. It estimates a modified version of the gravity model using a Poisson pseudo maximum likelihood (PPML). The estimation strategy is in two steps. First, the paper analyses the evolution over time of the impact of geographical distance and national border on the two measures of knowledge flows through separate cross-sections (for different sub-periods) and panel estimates. Second, it investigates the existence of changes in the diffusion of knowledge due to the process of

European integration. In this case, fixed-effects estimates are performed to take into account possible heterogeneity bias due to the presence of unobserved factors.

The results show that in the case of tacit knowledge flows (inventor collaborations) there is a decreased geographical localization, while in the case of codified knowledge flows (inventor citations) there is an increased process of localiziation: geographical distance and national borders are becoming more important. The results show also that the European integration process has a significant effect on reducing barriers to knowledge flows. In particular, this paper finds that, after the enlargement, knowledge diffusion is increased between regions in new and old EU members (for both types of knowledge). However, for inventor citations, this effect relates only to the EU enlargement of 1995 and is confined to knowledge flows from Austria, Finland and Sweden to old EU members.

The paper is organized as follows. The second section discusses the literature on the diffusion of knowledge and the theoretical justification for our analysis. The third and fourth sections respectively explain the gravity model and the methodology adopted. The fifth section presents the data. The sixth section presents and discusses the results. The last section offers some final considerations.

2 Background to the study

The literature on knowledge flows in Europe shows that the diffusion of knowledge in Europe is geographically localized (see, e.g., Mauseth and Verspagen, 2002; Bottazzi and Peri, 2003). This section focuses only on those papers that have measured knowledge flows using inventor citations and inventor collaborations.[1]

Geographic localization of knowledge flows

The pioneering work of Jaffe et al. (1993) provides evidence of the geographical patterns of knowledge flows for the US. Using USPTO data and a matching procedure that takes account of the existing geographic concentration of patent activity, they find that a patent is more likely to be cited by other patents originating in the same country, state or metropolitan statistical area. The first empirical evidence for Europe was provided by Maurseth and Verspagen (2002). They use EPO data for 112 regions of 14 European countries and gravity model estimates to show that the likelihood of citations between two patents developed in two different regions is negatively affected by the presence of an international border and by the geographical distance between them. Moreover, they provide empirical evidence of the importance to control for technological proximity in knowledge flows analysis. Similarities in technological specialisation facilitate knowledge flows across regions and estimates that do not control for regional specialisation are biased because physical proximity might capture the effects of technological proximity. Several studies follow and arrive at similar conclusions (see, e.g., Fischer et al, 2009; Paci and Usai, 2009): in Europe knowledge flows are strongly localised; barriers include physical proximity and other forms of proximity, especially institutional (such as country border) and technological proximity.

The empirical literature on the determinants of knowledge flows using inventor collaboration data is more recent and is mostly at country level (Guellec and van Pottelsberghe de la Potterie, 2001; Picci, 2010; Montobbio and Sterzi, 2013). Guellec and van Pottelsberghe de la Potterie (2001) show that sharing the same territorial border affects collaborations between inventors residing in two different countries. Moreover, they show that two countries are more likely to collaborate if they are close in the technology space. These results are confirmed by Picci (2010). In addition, he finds that international collaborations are positively affected by the EU membership. Also Hoeckman et al (2009), using both patent co-inventorship and publication co-authorship data for two sectors (biotechnology and semicounductors), provide evidence that interregional collaborations are hampered by physical distance and institutional factors (i.e. country border).

Finally, Montobbio and Sterzi (2013) show that technological proximity and sharing a common language are key drivers of international technological collaborations between emerging and advanced countries. They show also that the effect of geographical distance and longitude depends on the type of firms considered and the effect is stronger for companies whose ownership is in the emerging country.

Evolution over time of proximity factors

Despite the growing interest in the spatial diffusion of knowledge, few studies analyse the evolution over time of the impact of physical distance and other proximity factors on knowledge flows. Some authors show that in recent years the diffusion of knowledge is more localized than in the past, while others studies show the opposite. Johnson et al. (2006), using USPTO data for the period 1975-1999, show that the average distance between the citing and the cited patents increases by almost seven miles per year and the average distance between coinventors also increases by four miles per year. Griffith et al. (2007) analyse the changes over time of the propensity for inventor citations to be national using USPTO data for the period 1975-1995 for 5 countries (US, Japan, France, Germany, UK) and two groups of countries (EU countries and Rest of the World). They apply a duration model that looks at the "speed" of the patents of different countries to cite the same patent, and show that the national border effect decreased during the period investigated. Sonn and Storper (2008), using USPTO data for the period 1975-1997 for the US and matching procedures (e.g. Jaffe et al., 1993), find that inventor citations became more localized at country, state and metropolitan levels.

Paci and Usai (2009) use EPO patent citations data for the regions in 17 European countries and make use of gravity model estimates. They construct two cohorts of citing patents, of patents granted in 1990 and in 1998. For each cohort they consider backward citations (i.e. citations to previous patents), for 1978-1990 for the first cohort and for 1978-1998 for the second cohort. They run two separate estimates, one for each cohort, and compare the results for the impact of geographical distance and national borders on interregional knowledge flows. Their results show that the geographical distance effect has increased, while the impact of national border has decreased.

The present analysis extends that by Paci and Usai (2009) by comparing the impact of physical distance and country border for 20 periods during 1981-2000. In addition, it uses forward citations (i.e. citations from later patents) and considers inventor collaborations as further measure of knowledge flows. This study also analyses the effect of EU integration on knowledge flows. Since the Treaty of Rome in 1957, the EU integration is an ongoing process in which several policy measures are adopted to reduce the barriers to the free movement of products, services, capital and people, and to the creation of supra-national institutions which promote or coordinate common policies for the membership countries. In addition, the free movement of knowledge, the so called "fifth freedom", represents the fundamental objective for the development of the ERA; for this reason several policies, among which an important tool is the Framework Programme, [2] are implemented to fostering collaborative R&D partnerships between European countries. At the same time, a series of enlargements have been extended the area, i.e. the number of countries, interested by the EU's institutions and rules. Hence, the entering EU countries take advantage of a greater economic and institutional proximity with the pre-existing EU countries.

The effects of the EU enlargement processes for the countries involved are mostly analysed in studies on trade flows (Bussière et al, 2008; Gil et al, 2008), but work on knowledge flows ignores the effect of the enlargement dimension of the EU. Some studies analyse the difference between the diffusion of knowledge across EU members and the diffusion of knowledge across not EU members (see, e.g., Picci, 2010). However, these studies consider only static effects and, therefore, do not investigate whether greater integration between countries has an effect on reducing the pre-existing barriers to knowledge flows.

Based on the existing empirical evidence, we test whether a reduction in economic and institutional barriers affect knowledge flows. Consequently, this paper analyses the impact of the EU integration (i.e. the EU enlargement processes) on knowledge flows in three types of regions, i.e. new members, old members, and non-EU members. To the knowledge of the authors of this paper, this analysis is the first attempt to test the dynamic impact of EU enlargements on the diffusion of knowledge.

3 The empirical model

This paper analyses knowledge flows among European regions using a modified version of the gravity model. The gravity model is widely used to study bilateral trade between countries (see, e.g., Anderson and van Wincoop, 2003) and is also useful to estimate knowledge flows (see, e.g., Maurseth and Verspagen, 2002; Montobbio and Sterzi, 2013). In its basic form, the model predicts that the diffusion of knowledge between two regions is directly proportional to the inventive mass of the regions and inversely proportional to the geographic distance between regions. Tipically, other factors facilitating or hindering interregional knowledge flows are added at the basic model. More generally, most of the gravity models used in literature can be represented in the following equation for the conditional mean of the knowledge flows:

$$[1]\ E(C_{ijt} \mid X_{ijt}) = e^{\alpha} P_{it}^{\beta} P_{jt}^{\upsilon} dist_{ij}^{\gamma} e^{V'_{ijt}\tau}$$

where C_{ijt} is the variable capturing knowledge flows (in our case measured by number of citations or collaborations) between regions i and j at period t and $X_{ijt} = (1, P_{it}, P_{jt}, dist_{ij}, V'_{ijt})$ represents the vector of the explanatory variables. P_{it} and P_{jt} are total numbers of patents (inventive mass) for the two regions, $dist_{ij}$ is the geographical distance between the two regions and V'_{ijt} represents the vector of other "link" indicators between the two regions.

This study identifies a citation from region j to region i as occurring 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. In the case of patents with more than one inventor residing in the same region (i or j), citations are counted only once.

Since the interest is focused on "pure knowledge spillovers" (Griliches, 1979), citations between two patents within a single firm are not considered. Self-citations within firms are not considered externalities. Self-citations between inventors are also excluded because, by definition, these cannot be considered an exchange of knowledge between individuals.

Inherent in the use of patent citations is a truncation bias problem (see, e.g., Bacchiocchi and Montobbio, 2010), due to the fact that only a limited period of the legal life of the patent is observed. This problem is greater for recent patent cohorts. This citation lag is a source of bias in evaluation of the changes in distance or country border effects because the diffusion of knowledge could follow paths that are influenced by time. For instance, it is possible that the "new" knowledge flows, in the first periods, more easily at the local level than beyond.

In order to overcome this problem, only citations where the time lag between cited and citing patents is four years or less are considered.[3] For instance, C_{ijt} is the total number of citations contained in patents developed in region j (knowledge-receiving region) during the period from *t to (t+4)* and directed to patents developed in region i (knowledge-generating region) in period *t*.[4] Thus, the sample consists of a set of cited patents for the period 1981-2000, and a set of citing patents for the period 1981-2004.

The second measure of knowledge flows used in this paper is technological collaborations. This study identifies a collaboration between the region i and the region j if, 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. Similar to the case of patent citations, if a patent has more than one inventor resident in the same region (i or j) the collaboration is counted only once. For inventor collaborations, there is obviously no truncation problem.

In the empirical studies, variables are added to the basic model in order to take account of regional differences in terms of technological specialization, social and institutional differences between regions and other factors that may enhance the localization effect determined by physical distance. In this paper, the following control variables are included:

- Technological proximity ($Tech_{ijt}$): this variable controls for the sectoral distribution of patents within the two regions because geographical proximity effect could be influenced by the technological specialization of regions. Following the literature (see, e.g., Peri, 2005; Montobbio and Sterzi, 2013), this variable is measured by the Jaffe (1986) index, i.e. the uncentred correlation between the vectors expressing the distribution of the patents in 30 technology classes (OST, 2004) for the region i and the region j, that is: $Tech_{ijt} = P_{it}P_{jt}'/[(P_{it}P_{it}')(P_{jt}P_{jt}')]^{1/2}$. This variable takes values between 0 (when the vectors are orthogonal) and 1 (when the vectors are identical).

- Common language ($Lang_{ij}$): this variable controls for the language spoken in the two regions. A common language facilitates interpersonal relationships and, thus, facilitates the diffusion of knowledge between regions. This variable is a dummy that takes the value 1 if the two regions have the same language.

- Common border ($Bord_{ij}$): this variable controls for whether the regions are neighbours. It determines whether adjacent regions engage in greater exchange of knowledge. It is a dummy that takes the value 1 if the two regions have a common border.

- Country border ($National_{ij}$): this variable controls for whether two regions are located in the same country and takes account of institutional, social and other features specific to a nation. These features can facilitate the exchange of knowledge among inventors located in the same nation. The variable is represented by a dummy that takes the value 1 if the two regions belong to the same country.

The empirical models from equation [1] can be expressed in the following alternative way (Santos Silva and Tenreyro, 2006):

$$[2] \quad E(C_{ijt} \mid X_{ijt}) = exp\big(\alpha + \beta\, ln(P_{it}) + \mho\, ln\big(P_{jt}\big) + \gamma\, ln\big(dist_{ij}\big) + V'_{ijt}\,\tau\big)$$

The first step of the analysis is the cross-sectional estimate using aggregated data for the whole period and for different sub periods in order to evaluate changes over time in the estimated parameters. Region specific effects, both for region i and region j (denoted ϱ_i and η_j), are included to take into account regional-specific unobservable effects and to correct for cross-sectional bias (Anderson and van Wincoop, 2003; Baldwin and Taglioni, 2006). The inclusion of these fixed effects capture all the variables that are region specific (e.g. the inventive mass) and that will be used in panel estimates. Using the standard approach that $C_{ij} = E(C_{ij} \mid X_{ijt}\,)\, \varepsilon_{ij}$, where ε_{ij} is a multiplicative error term, this gives the following equation:

$$[3] \quad C_{ij} = exp\big[\alpha + \gamma\, ln\big(dist_{ij}\big) + \varrho\, Tech_{ij} + \sigma\, Lang_{ij} + \Omega\, National_{ij} + \varphi\, Bord_{ij} + \varrho_i + \eta_j\big]\, \varepsilon_{ij}$$

Further estimations are conducted to check the previous results on the dynamics of the distance effect. In particular, these estimates are performed on a panel dataset obtained by pooling annual data. Time dummies (denoted Θ_t) capture all-region-pairs-common time varying effects affecting knowledge flows, which are not captured by the other explanatory variables. Moreover, the variable for the distance is interacted with time dummies in order to allow the coefficient of distance to shift yearly. This gives the following equation:

[4] $C_{ijt} = exp\left[\alpha + \beta\, ln(P_{it}) + \mho\, ln(P_{jt}) + \gamma_t\; ln(dist_{ij}) + \varrho\, Tech_{ijt} + \sigma\, Lang_{ij} + \Omega\, National_{ij} + \varphi\, Bord_{ij} + \varrho_i + \eta_j + \Theta_t\right] \varepsilon_{ij}$

Since changes in the distance effect can also capture changes in the country border effect, a further check is made allowing the coefficient of country border to vary over time. This gives:

[5] $C_{ijt} = exp\left[\alpha + \beta\, ln(P_{it}) + \mho\, ln(P_{jt}) + \gamma_t\; ln(dist_{ij}) + \varrho\, Tech_{ijt} + \sigma\, Lang_{ij} + \Omega_t\, National_{ij} + \varphi\, Bord_{ij} + \varrho_i + \eta_j + \Theta_t\right] \varepsilon_{ij}$

One of the aims of this analysis is to examine the effect of the European integration process on interregional knowledge flows. The time period covered by this analysis, 1981 to 2000, includes two enlargement processes. The first is in 1986, following the entry of Spain and Portugal to the EU, and the second is in 1995, following the entry of Austria, Finland and Sweden. In the bilateral trade literature (see, e.g., Gil et al., 2008), the impact of European integration is estimated using a dummy variable added to the basic gravity model in order to capture deviations from the volumes of trade predicted by the model. This paper follows the same methodology and makes use of a time varying dummy variable ($EUboth_{ijt}$) which is set equal to 1 if both region *i* and region *j* are members of the EU at time *t*. In order to take account of a possible effect on knowledge flows towards non-EU members, it is added a time varying dummy ($EUone_{ijt}$) which is set equal to 1 when only one region (*i* or *j*) is a member of the EU at time *t*. These two dummies are time varying variables as there are regions of countries that are not EU members in 1981, the first year of analysis, but are EU members in 2000, the last year of analysis. Therefore, regions of countries that join the EU during the period 1981-2000 are considered not EU members until the year of entrance in the EU and EU members after that.

The following equation is used to estimate the effect of European enlargement on the knowledge flows between regions:

[6] $C_{ijt} = exp\left[\alpha + \beta\, ln(P_{it}) + \mho\, ln(P_{jt}) + \gamma_t\; ln(dist_{ij}) + \varrho\, Tech_{ijt} + \sigma\, Lang_{ij} + \Omega\, National_{ij} + \varphi\, Bord_{ij} + \varphi\, EUboth_{ijt} + \omega\, EUone_{ijt} + \varrho_i + \eta_j + \Theta_t\right] \varepsilon_{ij}$

As further step, a set of dummy variables are created to capture the differences between regions based on membership of the EU: old_i (old_j) is a time constant dummy variable which is set equal to 1 for regions *i* (*j*) of countries that are EU member since 1981; $never_i$ ($never_j$) is a time constant dummy variable which is set equal to 1 for regions *i* (*j*) of countries that are not EU member, i.e. did not enter the EU during the period 1981-2000; new_{it} (new_{jt}) is a time varying dummy variables which is set equal to 1 for regions *i* (*j*) of the EU entering countries, i.e. countries that join the EU in the period 1981-2000, for the EU integration year and the following years. The interaction between these indicators generates a new set of variables (see Figure 1) that define each pair of regions included in the sample on the basis of their EU membership.

- Figure 1 about here -

The variable $EUboth_{ijt}$ is equal to 1 when old_i*old_j, old_i*new_{it}, new_{it}*old_i *or* new_{it}*new_{jt} are equal to 1. $EUone_{ijt}$ is equal to 1 when old_i*$never_j$, $never_i$*old_j, new_{it}*$never_j$ *or* $never_i$*new_{jt} are equal to 1. This allows to identify whether the aggregate effect of EU membership ($EUboth_{ijt}$ and $EUone_{ijt}$) hides different behaviors in the different subgroups. Since the dataset is at regional level, it is also possible to distinguish between the effects of European integration on the diffusion of knowledge within and between countries by breaking down the above variables on the basis of a shared national border. Therefore, the suffixes *intra* and *extra* are used to distinguish between intra-national (*intra*) and extra-national (*extra*) knowledge flows for the variables old_i*old_j, and new_{it}*new_{jt}. The other variables, by definition, regard only extra-national knowledge flows. Thus, it is adopted a further specification in which the variables $EUone_{ijt}$ and $EUboth_{ijt}$ are replaced with their subgroups (see Figure 2).

- Figure 2 about here -

Finally, since it is possible to distinguish between the two phases of EU enlargement (1986 and 1995), it is tested whether the effect of EU integration is different in the two periods and, consequently, in the two different groups of nations. Therefore, the suffixes *enl86* and *enl95* are used to distinguish between regions that enter the EU in 1986 (*enl86*) and regions that enter the EU in 1995 (*enl95*) for the variables old_i*new_{jt}, $new_{it}*old_j$, $(new_{it}*new_{jt})\_intra$, $(new_{it}*new_{jt})\_extra$, $new_{it}*never_j$ *and* $never_i*new_{jt}$. The other variables regard only knowledge flows between regions that do not change their status of EU member during the period 1981-2000.

Note that the two measures of knowledge flows used in the analysis have some characteristics that need to be taken into account in determining the specification to be used in the estimates. In particular, patent citations capture the diffusion of knowledge from patent inventors to other inventors who developed a patent in a subsequent period. Thus, patent citations measures unidirectional flows between inventors or regions. Collaborations captures the interchange of knowledge between inventors for the generation of a new patent. Thus, inventor collaborations measure bidirectional flows between inventors or regions. This distinction means that in evaluating the impact of European integration on pairs of regions using patent citations rather than inventor collaborations, it is possible to disentangle the effects on the knowledge generating region and on the knowledge receiving region. For instance, for the pairs of "old" and "new" regions, the diffusion of knowledge from "old" to "new" regions (old_i*new_{jt}) can be identified separately from the knowledge flows from "old" to "new" regions ($new_{it}*old_j$). In the case of inventor collaborations there are only the bidirectional flows between "old" and "new" regions, thus, there is only one variable (old_i*new_{jt}).

4 Methodology

The gravity models in equations [3] to [6] can be estimated using different estimators. Following a procedure widely used in the literature on international trade, the gravity model could be estimated using OLS on the log-linear version of the previous equations. However, this procedure leads to biased estimates (Santos Silva and Tenreyro, 2006).[5] First, there are pairs of regions that do not have any interchange of knowledge (either citations and/or collaborations), which means a zero value of the dependent variable. Treating these observations as missing in the estimates would introduce a bias in the estimated coefficients. Gravity models also have an inherent problem of heteroschedasticity, which again lead to biased estimates. To jointly address these issues a PPML estimator is particularly appropriate (Santos Silva and Tenreyro, 2006).[6]

The effect of European enlargement on knowledge flows is therefore estimated using PPML estimates with regional dummies (covering both knowledge generating and knowledge receiving regions) (equation [6])[7] and PPML fixed-effects (i.e. with region-pair dummies). The latter are statistically more robust than the former because they control for unobserved region-pair heterogeneity (Cheng and Wall, 2005), which can explain the amount of bilateral knowledge flows and, additionally, the probability that two regions are in the same European agreement. However, this procedure has the disadvantage that does not allow to estimate the impact of the European integration for pair of regions whose EU member status does not change during the period covered by this analysis. In fact, the inclusion of region-pair dummies implies that only information on time variation in the variables is used to estimate their coefficient values, while information on cross-sectional variations is excluded.

This means that it is not possible estimate the effect for time invaryng variables such as those used to represent the pairs of regions that do not involve at least one *new* region. Thus, the fixed-effects models allow estimates of the European integration effects for only six pairs of regions that involve at least one country that became a new member of the EU. This might be seen as a limitation, but is not because this paper tests the effect on knowledge flows of greater integration among countries, and this effect is captured by looking at the exchange of knowledge between new EU member regions and other regions (EU members or not).[8] Pairs of regions excluded by fixed-effects analysis are shaded grey in Figure 2.

5 Data

The two measures of knowledge flows, i.e. patent citations and inventor collaborations, are built using the information contained in EPO patents (KITES and OECD REGPAT database). Patents are assigned to a region using the addresses of the inventors.

The analysis of knowledge flows for the period 1981-2000 is performed at the level of NUTS2 regions (EUROSTAT, 2007). The initial dataset contains data on patents with at least one inventor residing in one of 29 European countries.[9] However, the estimates consider only those regions that have at least one patent in each year in the sample because if a region has no patents then, by definition, it cannot have a regional knowledge flow. [10] Thus, the final dataset contains patents data from 191 regions (169 regions in the EU15 countries and 22 regions in the remaining countries). As discussed above, using inventor citations it is possible to measure unidirectional knowledge flows from one region to another; inventor collaborations measure only bidirectional flows between two regions. Thus, the final dataset contains 729,620 observations [191 regions* 191 regions *20 years] for patent citations and 366,720 observations [(((191^2-191)/2) +191)*20] for inventor collaborations.

The geographical distance between two regions is calculated using the great circle distance method on the basis of the geographical coordinates of the centre point of the regions (Maurseth and Verspagen, 2002). In considering knowledge flows within regions, the intraregional distance is calculated as two thirds of the radius of the regional geographic size, which is presumed to be circular in shape (Hoeckman et al., 2010). As mentioned above, to construct the variable related to technological proximity (*Tech*), this paper uses the 30 technological classes from the OST (2004) classification. Finally, the variable that controls for the language (*Lang*) is built on the basis of the regional official languages.

6 Results

This section presents and compares the results of the estimates for the two measures of knowledge flows. It starts with some descriptive statistics and shows the results of the estimates of equation [3] for the whole period. Then, the full sample is splitted into sub periods and separate estimates are provided for each sub period in order to assess the evolution over time of the impact of geographical factors. This section shows also a set of robustness checks for the previous results on estimates carried out using panel data (equations [4] and [5]). Finally the impact of European integration on the diffusion of knowledge (equation [6]) is considered.

Descriptive statistics

Figure 3 shows the distribution over time of interregional and international patent citations (panel a)) and technological collaborations (panel b)) as percentages of the total (regional excluded for the international). Interregional patent citations have decreased over time (from 91.4% in 1981 to 88.1% in 2000), while interregional collaborations have increased over time (from 33.5% in 1981 to 46.6% in 2000). The international patent citations decrease over time (from 67.4% in 1981 to 58.2% in 2000), while international collaborations increase over time (from 11.9% in 1981 to

22.1% in 2000). This figure points at two aspects of the diffusion of knowledge between regions. One the one hand, inventor collaborations, throughout the period examined, are more localized than inventor citations. On the other hand, these two measures of knowledge flows exhibit different time trends: inventor citations seems to become more localized, while the opposite occurs to inventor collaborations. This evidence is tested by the econometric exercise described in the next paragraphs.

- Figure 3 about here -

Cross-section estimates for the whole period

Table 1 shows the results of the estimates of equation [3] using aggregated data for the whole period (1981-2000). It presents each measure of knowledge flows in a separate column: first columns (1a) for the inventor citations and (1b) for the inventor collaborations show the results of the estimates that do not consider intraregional knowledge flows (i.e. excluding observations for which *i=j*), while, as a robustness check, columns (2a) and (2b) show the results including intraregional knowledge flows.[11]

In Columns (1a) and (1b) all the coefficients are statistically significant and their signs are consistent as expected. The distance (*dist*) effect is negative for both citations and collaborations. Therefore, the diffusion of knowledge between European regions decreases with geographical distance. Moreover, the positive and significant effect of *Bord* underlines the role of geographical proximity in determining knowledge flows between regions (both for citations and collaborations) showing that knowledge flows are higher for geographically contiguous regions.

For the variables *National* and *Lang* the coefficients are positive for both measures of knowledge flows. The diffusion of knowledge is higher for two regions in the same country. The significance of the variable *National*, controlling for geographic proximity (*dist* and *Bord*) or technological proximity (*Tech*), draws the attention to the social, institutional or other country specific reasons which lead to greater knowledge flows within countries. Also language matters, as the diffusion of knowledge is greater if the regions share a common language.

It is important to emphasise the difference between the two measures of knowledge flows in the estimated coefficients for geographical (*dist* and *Bord*), social and institutional (*Lang* and *National*) proximity. These values are greater in the case of technological collaborations. This means that technological collaborations tend to be more geographically localized than patent citations. This is consistent with inventor citations not requiring face-to-face contact. For instance, an inventor can learn about the invention cited simply by reading the description contained in the patent document. In general, the analysis confirms the hypothesis that geographical, institutional and other country specific factors are more important for inventor collaborations than for inventor citations.

Finally technological proximity (*Tech*) is more important for inventor citations than for inventor collaborations. Technological complementarities are an important incentive for inventors to collaborate. While absorptive capacity and, thus, a degree of technological proximity are necessary for effective knowledge exchange between inventors, technological complementarities and, thus, a degree of technological distance, allow inventors to learn about new knowledge.

As a robustness check, the above specifications are estimated including the observations for intraregional knowledge flows. The variable *region*[12] is also included to take into account the possible existence of regional barriers to knowledge flows. The results of these estimates (columns (2a) and (2b)) show a significant and positive effect of *region* on both measures of knowledge flows. This means that knowledge flows are more likely within regions and, thus, there are regional barriers that contribute to the geographically localized diffusion of knowledge.

- Table 1 about here -

Cross-section for different sub periods

The next step is the analysis of the evolution of the coefficients over time. The dataset is divided into four sub periods (i.e. 1981-1985, 1986-1990, 1991-1995 and 1996-2000) and four separate cross-sectional analyses (equation [3]) are performed, i.e. one for each sub period. Table 2 presents the results of these estimates. Columns (1a) to (4a) show the results for the inventor citations and columns (1b) to (4b) the results for the inventor collaborations. In general, the estimates confirm that geographical and the other forms of proximity hinder the diffusion of knowledge among European regions, and the evolution over time is different for patent citations and collaborations. The distance effect increases over time (from -0.14 to -0.21) for citations, but decreases for collaborations (from -1.05 to -0.88). So, the reduction in the distance effect is found only for collaborations. At the same time, the national border effect increases for patent citations (from 0.30 to 0.53) and decreases for technological collaborations (from 1.86 to 1.71).[13] The coefficient of *Bord* increases for patent citations (from 0.09 to 0.24), but slightly decreases for the collaborations (from 0.69 to 0.68). Based on these results, it can be stated that over time interregional collaborations among European inventors are affected less and less by geographical proximity and territorial borders, while the opposite effect occurs for patent citations.

Buch et al. (2004) address the issue of the changes in the distance coeffcients when different cross-section estimates for different time periods are compared. In trade literature a key question is how transport and information costs affect the level of trade. They show that when the change in transport costs is proportional across different distances, the distance coefficient does not change and the level effect is captured by the constant term. In Table 2 the constant terms increase both for collaborations and citations. This is probably due to decreased communication costs and increased access to information. At the same time the change in the distance coefficients can be exactly interpreted as follows: in the case of citations an increase (in absolute value) of the distance effect means that citations between regions that are far away decrease relative to citations between regions that are geographically closer. In the case of collaborations a decrease in absolute value of the *dist* coefficient means that collaborations between regions that are far away become more important relative to collaborations between regions that are geographically closer.

In addition, the effect of technological proximity (*Tech*) increases over time for patent citations (from 2.09 to 2.28), and decreases for inventor collaborations (from 1.89 to 1.65). Finally, the importance of sharing common language (*Lang*) decreases for both measures of knowledge flows, i.e. from 0.30 to 0.17 for inventor citations and from 0.69 to 0.42 for inventor collaborations.

In sum, these results corroborate the hypothesis of decreased importance of spatial proximity as a determinant of interregional knowledge flows only for technological collaborations. Results on citations, on the contrary, suggest that the presence of agglomeration forces in line with the "missing globalization puzzle" observed in the trade literature (Bhavnani et al., 2002).

- Table 2 about here -

Panel estimates of the distance effect

Pooled cross-section estimates using a panel dataset obtained by pooling annual data confirm the previous results on the dynamics of the distance effect. The results of equation [4] confirm the trends of the cross-sectional estimates. Figure 4 reports coefficient values (and the confidence interval at ± 95%) for the variable *dist.*[14] For inventor citations the distance effect increases in absolute value over time, while it decreases for inventor collaborations.

- Figure 4 about here -

As a further check, estimates of distance and national border effects varying over time simultaneously (equation [5]) are presented in Figure 5.[15] It shows the results for the distance effects (panel a)) and for the national border effects (panel b)). National border effects decrease for inventor collaborations, while the opposite occurs for inventor citations[16]. These findings confirm that citations and collaborations follow two different trends in which the former become more geographically localized, and the latter become less localized.

- Figure 5 about here –

With regard to the distance effect, the decreasing trend for the inventor collaborations is confirmed in Figure 5, and the trend for inventor citations follows a U-shaped curve. This means that if it is not controlled for the national border effect over time, the distance effect captures mainly the increased tendency for EU inventors to cite national patents. Thus, the increased localized diffusion of patent citations is due mainly to the increased home bias effect (see also Bacchiocchi and Montobbio, 2010).

Summing up, this paper corroborates the idea that tacit knowledge flows (driven by collaborations) are much more constrained by communication and transports costs than codified knowledge flows (tracked by patent citations). According to the estimates (Figures 4 and 5), the elasticity of the former to geographical distance is in the range between -0.9 and -1 while the elasticity of the latter is around -0.2. However, the impact of geographical distance and national borders decrease over time for tacit knowledge flows. On the contrary, codified knowledge flows between regions that are physically distant decrease relatively to the ones between regions that are geographically closer.

One of the main factor that could explain why interregional collaborations are less affected by geographical distance and national borders is that the process od European integration and the related policies have facilitated the contact between different inventors. This is directly tested in the next session. In addition, many other factors could have affected the openness of the European regions and have facilitated collaborations. The increased internationalisation of product and innovation activities could be driven by foregn direct investments and international mergers and aquisitions. Different motivations explain these cross-national activities in particular in Europe (Archibugi and Iammarino, 1999; Guellec and van Pottelsberghe de la Potterie, 2001): the necessity to adapt a product to a local and *common* market; the need to monitoring the technology developments in other European regions; the willingness to develop complementary technologies in which the other European regions have a comparative advantage. These international activities - facilitated by the increased integration of the European markets - stimulate contacts and exchanges of informations between distant inventors. Moreover, international cooperations could be fostered by the increasing specialisation of high skilled workforce, and by the increasing complexity in the innovation process in many industries. Finally, the first three Framework Programmes may have played a role in fostering intra-EU collaborations.

The second issue is why patent citations are more affected over time by geographical distance and national borders, and therefore are increasingly localized. Indeed, the evidence provided in this paper is that the probability to have a localised citation has increased over time. This is not in contrast with the previous evidence and can be interpreted in the following way. The increased internationalisation of innovation in Europe could be associated with a process of increased specialisation and localisation. Internationalisation and the process of specialisation are compatible with a greater openess of regions (Niosi and Bellon, 1994). Krugman (1991), comparing employment data for US in 1977 and Europe in 1985, argues that the relatively low industrial specialisation in Europe is due to the presence of national barriers to the movement of goods and of factors of production, and he anticipates the idea that a greater integration of Europe involves a greater geographical concentration of similar and related sectors. Several studies, that extend the data coverage including post-1985 data (see, e.g., Midelfart-Knarvik et al., 2002; Combes and Overman, 2004), provide evidences that European countries are becoming, althought slowly, more specialised. The increasing localisation of production brings closer people with similar or related skills making easier the comunication and absorption of production techniques and of ideas at local level. This could have also increased some typical local externalities generated by spin-out and the related provision of localised services.

Panel estimates of the impact of European integration

This paper analyses also the impact of the EU integration on knowledge flows as result of the European enlargement processes. Coherent with the trade literature (see, e.g., Carrère, 2006; Gil et al., 2008), a set of dummies is used to identify the impact of European integration on interregional knowledge flows. The analysis of the effects of the European integration process on interregional knowledge flows is conducted using equation [6], either with or without region-pair fixed effects. Table 3 presents the results.[17] Columns (1a) to (6a) show the results for the inventor citations and columns (1b) to (6b) the results for the inventor collaborations. PPML estimates are shown in columns (1a), (1b), (3a), (3b), (5a) and (5b), while PPML fixed-effects estimates are shown in columns (2a), (2b), (4a), (4b), (6a) and (6b).

PPML estimates (columns (1a) and (1b)) show that European integration increases knowledge flows between EU regions (*EUboth*), for both citations and collaborations. In addition, the EU integration process reduces knowledge flows between EU regions and non-EU regions (*EUone*).

PPML fixed-effects estimates (columns (2a) and (2b)) control for region-pair effects in order to obtain unbiased estimates of the integration effect (Cheng and Wall, 2005; Carrère, 2006). For the *EUboth* dummy, the coefficient is positive for both measures of knowledge flows, but significant only for patent citations. Thus, it seems that there is an EU integration effect only in the case of citations. The *EUone* dummy is insignificant for both measures, thus, there are no effects on third countries of EU integration.

For the different groups of regions in *EUboth* (columns (4a) and (4b)) the picture of European integration effects is more detailed. For collaborations, Table 3 shows a positive and significant effect on collaboration between *old* and *new* regions (*old*new*). Thus, European integration increases international collaborations between EU regions, but it has no effect on knowledge flows between new EU members (*(new*new)_intra* and *(new*new)_extra*). Finally, there are no effects on knowledge flows between new and non-EU members (*new*never)*.

With regards to patent citations, it is observed a positive and significant effect between *old* and *new* regions in relation to *old* regions citing the patents of *new* regions (*new*old*), but a negative and insignificant effect for *new* regions citing the patents of *old* regions (*old*new*). Thus, EU integration increases international knowledge flows only from new EU members to old EU members. Also, there is a positive and significant effect on international knowledge flows between *new* regions (*(new*new)_extra*) and a negative and significant effect on national knowledge flows between *new* regions (*(new*new)_intra*). Thus, EU integration increases international knowledge flows while decreasing national flows between new EU members. Finally, the EU integration has no effects on knowledge flows between new and not EU members (*new*never* and *never*new*).

Table 3 (columns (6a) and (6b)) shows also the estimated coefficients separating the first EU enlargement in 1986 with Spain and Portugal and the second one in 1995 with Austria, Finland and Sweden. For collaborations, there is a positive and significant effect confirmed between *old* and *new* regions with each EU enlargement (*(old*new)_enl86 and (old*new)_enl95*). For patent citations, the aggregate effects of European integration are based only on the second EU enlargement (with the exception of *(new*new)_extra_enl86*).[18]

To sum up, European integration had a significant effect on reducing barriers to knowledge flows between new and old EU members. It appears that these countries benefit from the EU governed increased free in the movements of goods, services, capital and people, and of the science, innovation and technology policies implemented by the EU like the EUREKA project and the first three Framework Programmes. In fact, this conjecture is in line with the work of Balland (2011) and Scherngell and Barber (2009) showing networks of collaborations created in more recent Framework Programmes. Finally, note that, for patent citations, this effect relates only to the second

EU enlargement and is confined to knowledge flows from new (Austria, Finland and Sweden) to old EU members.

- Table 3 about here –

7 Conclusion

This paper analyses the evolution over time of knowledge diffusion among European regions using patent citations and technological collaborations. The results show that knowledge flows are geographically localized for both measures and that the impacts of geographical and country specific factors are higher for inventor collaborations than for inventor citations. The results show also that, although national borders are still important barriers to the diffusion of knowledge, their impacts on the two measures of knowledge flows follow different time trends. In particular, the national border effect decreases for technological collaborations and increases for patent citations. Inventors tend to collaborate more with other international inventors and, at the same time, they tend to cite more inventors that are closer geographically or localized in the same country. The evolution over time of the distance effect confirms that inventor collaborations are becoming less localized, while the reverse is true for inventor citations. The diverging trends between inventor citations and collaborations can be explained by the contemporaneus increased specialisation and internationalisation of production and technological activities. On one side, a more localised concentration of inventors working in similar and related industries increases the probability to cite other local inventors. On the other side, a more international engagment of firms provides the contact information needed to foster cross-national border collaborations. In addition, the increased complexity in the innovation process spurs international collaboration since it is difficult to find in the same territory expertise in the required fields.

This paper also analyses whether European integration has an impact on reducing the economic and institutional barriers to knowledge flows. It shows that European integration favors international collaborations between entering EU members and existing EU members. For patent citations, it seems that European integration positively affects the diffusion of knowledge only in the case of the second EU enlargement, i.e. when the most technologically advanced countries are involved, and only for knowledge generated in new member regions (Austria, Finland and Sweden) that is more used by old EU members. In general, the results support the hypothesis that EU integration policies contribute to enhance knowledge flows across European regions, which is a key objective of the ERA.

However, more work is necessary to understand exactly which are the economic forces that drive the patterns of knowledge diffusion described in this paper. In addition, it would be relevant to explore more in detail the consequences on regional growth and employment of the specific patterns of knowledge diffusion. Deep concerns have been expressed about growing disparities between EU regions. This paper shows that European integration helps to drive down possible obstacles to knowledge flows but at the same time results on patent citations suggest also a process of localization and specialization. As emphasised by the recent literature on smart specialization it is important that economic benefits in some regions due to increased returns and agglomeration economies spill over to other regions. However, in some sectors innovation and R&D are more productive. This means that a given amount of R&D creates a higher return in terms of productivity and growth. As a consequence, more attention has to be paid to the quality and type of specialization and possible consequences on adverse specialization.

Aknowledgments

We thank two anonymus referees for valuable comments and suggestions on a previous version of this paper. All the usual disclaimers apply.

[1] Other measures of knowledge flows are interfirm networks (Balland et al, 2012; Morrison, 2008) and co-authorships in scientific journal (Ponds et al., 2007).

[2] For a detailed discussion about EU science, innovation and technology policies see Caloghirou et al. (2002) and Roediger-Schluga and Barber (2006). In addition, a series of studies analysed collaborations in the context of the 5th and 6th Framework Programme showing that geographical factors are important determinants of cross-regional collaborations (Balland, 2011; Scherngell and Barber, 2009).

[3] In the analysis are also considered different time lags, but the results obtained are quite similar. These results are available from the authors on request.

[4] Moreover, the inventive mass in all the equations with patent citations as dependent variable is adapted in order to take into account of these temporal windows of four years. Thus, the term P_{jt} becomes the total number of patents developed during the period from t to $(t+4)$.

[5] The lack of the explicit consideration of multilateral interactions linking all the regions is a potential source of biases due to spatial auto-correlation possibly generated by spatial spillovers. This analysis controls for that using region fixed effects. An alternative way to control for cross-sectional interdependencies is to use spatial econometric techniques. As robustness check, further estimates are made adopting the spatial filtering approach (Griffith, 2003; Fischer and Griffith, 2008). In particular, a new set of variables are constructed as linear combination of selected eigenvectors obtained from a modified contiguity matrix (Fischer and Griffith, 2008). The eigenvectors are selected on the basis of the value of the Moran index (MI) associated to their respective eigenvalues. These are the eigenvector whose MI value divided by the maximum MI value is at least 0.25. Altogether, 42 eigenvectors for the knowledge generating regions and 42 eigenvectors for the knowledge receiving regions are added at the gravity models. The results, available from the authors on request, are very similar. Finally, even recognizing that some results could still be affected by spatial autocorrelation, it can be safely assumed that comparison over time of the estimated coefficients remain largely unaffected (see below Figure 4 and 5). It is explicitly recognized that these procedures do not remove the potential biases in the coefficient values stemming from the omitted consideration of multilateral interactions as determinant of the bilateral knowledge flows. However, control variables like adjacency and language should reduce a lot of these potential biases.

[6] Santos Silva and Tenreyro (2011) show that PPML estimator performs well also when the sample has a large proportion of zeros and when the conditional variance is not proportional to the conditional mean.

[7] The time constant region dummies allow to take account of the cross-section correlation between the omitted variables and the included variables, but do not control for the time-series correlation. Time-varying region dummies should be used to remove the time-series correlation (Baldwin and Taglioni, 2006), but the large number of regions and years investigated makes this calculation difficult.

[8] Also, with regard to *EUboth* and *EUone* variables, PPML models (equation [6]) estimating the effect of being part of the EU, while PPML fixed-effects models estimating the effect of joining the EU because information on time invariant pairs of regions ($(old_i*old_j)\_intra$, $(old_i*old_j)\_extra$, $old_i*never_j$ and $never_i*old_j$) are not used.

[9] The 29 European countries considered are: Belgium; Denmark; Germany; Greece; France; Ireland; Italy; Luxembourg; The Netherlands; United Kingdom; Spain; Portugal; Austria; Finland; Sweden; Bulgaria; Cyprus; Czech Republic; Estonia; Hungary; Latvia; Lithuania; Malta; Norway; Poland; Romania; Slovakia; Slovenia; Switzerland.

[10] As a result of this procedure, all the regions belonging to 8 countries (Cyprus, Estonia, Latvia, Lithuania, Malta, Romania, Slovakia and Slovenia) and some regions of the other 21 countries are discarded. Considering the aggregated data for the whole period investigated (1981-2000), the regions excluded account for about the 0.68% of the total patents, for about the 1.11% of the total inventor citations and for about the 0,9% of the total inventor collaborations. However, the estimation results obtained using the sample with all the regions are very similar. These results are available from the authors on request.

[11] The number of observations in the cross-section estimates, both for the full sample and for the different sub periods, is 36,290 [(191*191)-190] for the inventor citations and 18,145 [((191*191)-191)/2] for the inventor collaborations. This difference in the number of observations is due to the different characteristics of these two variables, i.e. unidirectional or bidirectional knowledge flows measures. Finally, the difference between the number of observations for the estimates in columns *a* and *b* is equal to the number of regions (i.e. 191).

[12] This dummy variable is set equal to 1 when knowledge flows occur within a region ($i=j$).

[13] In our sample 85% of the total number of citations are added by the patent examiners. Recent literature suggests that examiners' citations display different properties (for EPO data see Criscuolo and Verspagen, 2008). If citations are added by the examiners, their use as an indicator of knowledge flows is more questionable. The usual assumption is also that examiner citations are less localized than inventor citations. As a robustness check, models are re-estimated excluding the citations from patent examiners. Confirming the results in Bacchiocchi and Montobbio (2010), the results, available from the authors on request, are not different from the ones displayed in Table 2.

[14] The variables P_i and P_j are significantly different from zero for both measures of knowledge flows. Coefficient values are about 1 for inventor citations and about 0.7 for inventor collaborations. The coefficients of the other control variables (not reported here) are all significant and have similar values as the above estimates.

[15] See note 14.

[16] As a robustness check, this paper controls for the evolution over time of both the region's internal technological specialisation and the region's relative technological specialisation across sectors with respect to the rest of the European regions. Further estimates of equation [5] are performed by adding in each period t, both for region i and for region j, an Herfindhal absolute index of internal specialisation and thirty Balassa indexes (one for each OST class) of relative specialisation (Malerba and Montobbio, 2003). These time varying indexes are constructed using disaggregated annual data on number of patents by regions and by technological classes. The results, available from the authors on request, are very similar.

[17] In the fixed effects estimates the observations for the pairs of regions with zero variations over time of the dependent variables are dropped. So, as a consequence, the number of observations is lower. Thus, the number of observations for the patent citations is 540,920 (i.e. 725,800-184,880), while the number of observations for the inventor collaborations is 161,340 (i.e. 362,900-201,560). Finally, the difference in the number of variables between PPML and PPML fixed-effects estimates is due to the fact that the latter do not allow estimation of time invariant variables.

[18] The high coefficient values of *(new_{it}*new_{jt})_extra_enl86* are due to the initial low levels of citations/collaborations before 1986 and the relatively high increase after 1986.

References

Anderson J.E. and van Wincoop E. (2003) Gravity with gravitas: a solution to the border puzzle, American Economic Review, 93,170-192.

Archibugi D. and Iammarino S. (1999) The policy implications of the globalisation of innovation, Research policy, 28, 317-336.

Arellano M. and Honoré B.E. (2001) Panel Data Models. Some Recent Development, in Heckman J.J. and Leamer E.E. (Eds) Handbook of Econometrics, Elsevier, Amsterdam.

Bacchiocchi E. and Montobbio F. (2010) International knowledge diffusion and home-bias effect: do USPTO and EPO patent citations tell the same story?, Scandinavian Journal of Economics, 112, 441-470.

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, 293-298.

Baldwin R.E. and Taglioni D. (2006) Gravity for dummies and dummies for gravity equations, NBER Working Paper No. 12516, National Bureau of Economic Research.

Balland, P. A. (2011). Proximity and the evolution of collaboration networks: Evidence from research and development projects within the global navigation satellite system (GNSS) industry. Regional Studies, 46(6), 741–756.

Balland P.A., De Vaan M. and Boschma R.A. (2012) The dynamics of interfirm networks along the industry life cycle: the case of the global video game industry, 1987–2007, Journal of Economic Geography, forthcoming.

Bauer T. and Zimmermann K.F. (1999) Assessment of possible migration pressure and its labour market impact following EU enlargement to Central and Eastern Europe, IZA Research Reports No. 3, Institute for the Study of Labor, Bonn.

Bhavnani R., Coe D.T., Subramanian A. and Tamirisa N.T. (2002) The missing globalization puzzle, IMF Working Papers No. 02/171, International Monetary Fund, Washington D.C..

Bottazzi L. and Peri G. (2003) Innovation and Spillovers in Regions: Evidence from European Patent Data, European Economic Review, 47, 687-710.

Brenton P., Di Mauro F. and Lucke M. (1999) Economic integration and FDI: an empirical analysis of foreign investment in the EU and in Central and Eastern Europe, Empirica, 26, 95-121.

Bussière M., Fidrmuc J. and Schnatz B. (2008) EU enlargement and trade integration: lessons from a gravity model, Review of Development Economics, 12, 562-576.

Cairncross F. (1997) The death of distance. How the communication revolution will change our lives, Harvard Business School Press, Boston.

Caloghirou Y., Vonortas N.S. and Ioannides S. (2002) Science and technology policies towards research joint ventures, Science and Public Policy, 29, 82-94.

Carrère C. (2006) Revisiting the effects of regional trade agreements on trade flows with proper specification of the gravity model, European Economic Review, 50, 223-247.

Cheng I. and Wall H. (2005) Controlling for heterogeneity in gravity models of trade and integration, Federal Reserve Bank of St. Louis Review, 87, 49-63.

Combes P.P. and Overman H. (2004) The Spatial Distribution of Economic Activities in the European Union, in Henderson J.V. and Thisse J.F. (Eds) Handbook of Urban And Regional Economics, Elsevier, North-Holland.

Criscuolo, P., and Verspagen, B. (2008), Does it Matter where Patent Citations Come From? Inventor vs. Examiner Citations in European Patents, Research Policy 37, 1892– 1908.

European commission (2010) Europe 2020 - A strategy for smart, sustainable and inclusive growth, COM(2010) 2020, Brussels.

Eurostat (2007) Regions in the European Union. Nomenclature of territorial units for statistics, European Commission, Brussels.

Fischer M.M. and Griffith, D.A. (2008) Modeling spatial autocorrelation in spatial interaction data: an application to patent citation data in the European Union, Journal of Regional Science, 48, 969-989.

Fischer M., Scherngell T. and Jansenberger E. (2009) Geographic localisation of knowledge spillovers: evidence from high-tech patent citations in Europe, The Annals of Regional Science, 43, 839-858.

Gil S., Llorca R. and Martínez-Serrano J.A. (2008) Assessing the enlargement and deepening of the European Union, The World Economy, 31, 1253-1272.

Griffith D.A. (2003) Spatial Autocorrelation and Spatial Filtering: Gaining Understanding through Theory and Scientific Visualization, Springer, Berlin, Heidelberg, and New York.

Griffith R., Sokbae L. and van Reenen J. (2007) Is distance dying at last? Falling home bias in fixed effects models of patent citations, NBER Working Papers No. 13338, National Bureau of Economic Research, Cambridge, MA.

Griliches Z. (1979) Issues in assessing the contribution of research and development to productivity growth, Bell Journal of Economics, 10, 92-116.

Grossman G.M. and Helpman E. (1991) Innovation and Growth in the Global Economy, Cambridge, MIT Press, Cambridge, MA.

Guellec D. and van Pottelsberghe de la Potterie B. (2001) The internationalisation of technology analysed with patent data, Research Policy, 30, 1253-1266.

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

Hoekman J., Frenken K. and Tijssen R.J.W. (2010) Research collaboration at a distance: changing spatial patterns of scientific collaboration within Europe, Research Policy, 39, 662-673.

Jaffe A.B., Trajtenberg M. and Henderson R. (1993) Geographic localization of knowledge spillovers as evidenced by patent citations, The Quarterly Journal of Economics, 108, 577-98.

Johnson D.K.N., Siripong A. and Brown A.S. (2006) The demise of distance? The declining role of physical proximity for knowledge transmission, Growth and Change, 37, 19-33.

Keller W. (2004) International technology diffusion. Journal of Economic Literature, 42, 752-782.

Keller W. and Yeaple S.R. (2009) Gravity in the Weightless Economy, NBER Working Paper No. 15509, National Bureau of Economic Research, Cambridge, MA.

Krugman P. (1991) Geography and trade, MIT Press, Cambridge, MA.

Malerba F. and Montobbio F. (2003) Exploring factors affecting international technological specialization: the role of knowledge flows and the structure of innovative activity, Journal of Evolutionary Economics, 13, 411-434.

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

Midelfart-Knarvik K.H., Overman H.G., Redding S.J. and Venables A.J. (2002) Integration and industrial specialisation in the European Union, Revue Economique, 53, 469-481.

Montobbio F. and Sterzi V. (2013) The globalization of technology in emerging markets: a gravity model on the determinants of international patent collaborations, World Development, 44, 281-299.

Morrison A. (2008) Gatekeepers of knowledge within industrial districts: who they are, how they interact, Regional Studies, 42, 817–835.

Niosi J. and Bellon B. (1994) The global interdependence of national innovation sytems: evidence, limits, and implications, Technology in Society, 16, 73–197.

OST (2004) Indicateurs de sciences et de technologies, Rapport de l'Observatoire des Sciences et des Techniques, Economica, Paris.

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

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

Ponds R., Van Oort F.G. and Frenken K. (2007) The geographical and institutional proximity of research collaboration, Papers in Regional Science, 86, 423-443.

Roediger-Schluga T and Barber M.J. (2006) The structure of R&D collaboration networks in the European Framework Programmes. UNU-MERIT working paper series No. 2006-36, The United Nations University - Maastricht Economic and Social Research Institute on Innovation and Technology, Maastricht.

Santos Silva J.M.C. and Tenreyro S. (2006) The log of gravity, The Review of Economics and Statistics, 88, 641-658.

Santos Silva J.M.C. and Tenreyro S. (2011) Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator, Economics Letters, 112, 220-222.

Scherngell, T., and Barber, M. (2009). Spatial interaction modelling of crossregion R&D collaborations: Empirical evidence from the 5th EU framework programme. Papers in Regional Science, 88, 531–546.

Sonn J.W. and Storper M. (2008) The increasing importance of geographical proximity in knowledge production: an analysis of US patent citations, 1975-1997, Environment and Planning, 40, 1020-1039.

TABLES

Table 1. Determinants of knowledge flows (aggregated data for the period 1981-2000)- PPML

	Citations				*Collaborations*			
Variable	(1a)		(2a)		(1b)		(2b)	
Tech	2.221	***	2.138	***	1.615	***	2.014	***
	(0.047)		(0.075)		(0.213)		(0.195)	
ln(dist)	-0.215	***	-0.243	***	-0.939	***	-0.828	***
	(0.011)		(0.015)		(0.057)		(0.060)	
Lang	0.226	***	0.225	***	0.505	***	0.398	***
	(0.020)		(0.023)		(0.084)		(0.096)	
National	0.452	***	0.454	***	1.763	***	1.791	***
	(0.023)		(0.024)		(0.111)		(0.128)	
Bord	0.180	***	0.152	***	0.705	***	0.733	***
	(0.025)		(0.026)		(0.084)		(0.078)	
Region			0.351	***			0.374	***
			(0.068)				(0.081)	
Constant	-2.731	***	-2.493	***	3.968	***	-0.051	
	(0.179)		(0.199)		(0.424)		(0.401)	
dummy region i	Yes		Yes		Yes		Yes	
dummy region j	Yes		Yes		Yes		Yes	
Intraregional observations	excluded		included		excluded		included	
Log Pseudo-likelihood	-97906.8		-110892.0		-47264.1		-62547.1	
N. observations	36290		36481		18145		18336	

Note: The dummy region *i* (*j*) is equal to 1 for region *i* (*j*) and 0 elsewhere. For the sake of clarity the coefficient values of region dummies, both for region *i* and region *j*, are not reported; ***, ** and * indicate significance at 1, 5 and 10 %, respectively; robust standard errors are in parentheses.

Table 2. Determinants of interregional knowledge flows (sub-periods estimates) - PPML -

	Citations								*Collaborations*							
	(1a)		(2a)		(3a)		(4a)		(1b)		(2b)		(3b)		(4b)	
Variable	1981-1985		1986-1990		1991-1995		1996-2000		1981-1985		1986-1990		1991-1995		1996-2000	
Tech	2.092	***	2.093	***	2.286	***	2.283	***	1.890	***	1.683	***	1.527	***	1.652	***
	(0.645)		(0.064)		(0.060)		(0.055)		(0.153)		(0.226)		(0.234)		(0.196)	
ln(dist)	-0.138	***	-0.183	***	-0.255	***	-0.209	***	-1.062	***	-0.971	***	-0.994	***	-0.883	***
	(0.016)		(0.016)		(0.015)		(0.014)		(0.078)		(0.073)		(0.066)		(0.051)	
Lang	0.315	***	0.194	***	0.231	***	0.190	***	0.696	***	0.650	***	0.514	***	0.424	***
	(0.027)		(0.026)		(0.028)		(0.025)		(0.121)		(0.119)		(0.097)		(0.077)	
National	0.298	***	0.441	***	0.389	***	0.530	***	1.871	***	1.819	***	1.791	***	1.710	***
	(0.027)		(0.027)		(0.032)		(0.028)		(0.153)		(0.139)		(0.130)		(0.102)	
Bord	0.091	**	0.142	***	0.165	***	0.242	***	0.687	***	0.759	***	0.676	***	0.684	***
	(0.045)		(0.035)		(0.033)		(0.029)		(0.082)		(0.075)		(0.070)		(0.061)	
constant	-4.798	***	-5.096	***	-3.741	***	-3.262	***	1.394		1.261	**	2.617	***	3.227	***
	(0.389)		(0.391)		(0.332)		(0.237)		(0.984)		(0.616)		(0.585)		(0.387)	
Dummy region i	Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Dummy region j	Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Log Pseudo-Lik.	-37264.0		-48468.1		-60354.3		-71054.7		-9794.5		-15133.4		-19469.6		-30896.9	
N. observations	36290		36290		36290		36290		18145		18145		18145		18145	

Note: The dummy region *i* (*j*) is equal to 1 for region *i* (*j*) and 0 elsewhere. For the sake of clarity the coefficient values of region dummies, both for region *i* and region *j*, are not reported; ***, ** and * indicate significance at 1, 5 and 10 percent, respectively; robust standard errors are in parentheses. Intraregional observations are excluded.

Table 3. European Integration - PPML with and without region-pair fixed effects -

Variable	*Citations*						*Collaborations*					
	(1a)	(2a)	(3a)	(4a)	(5a)	(6a)	(1b)	(2b)	(3b)	(4b)	(5b)	(6b)
EUboth	0.148***	0.085***					0.071**	0.058				
	(0.014)	(0.019)					(0.029)	(0.039)				
EUone	-0.121***	0.033					-0.236***	0.040				
	(0.016)	(0.039)					(0.044)	(0.090)				
old*new			0.006	-0.014					0.263***	0.266***		
			(0.018)	(0.022)					(0.040)	(0.047)		
new*old			0.261***	0.220***								
			(0.019)	(0.026)								
(old*old)_intra			-0.239***		-0.224***				0.868***		0.953***	
			(0.053)		(0.055)				(0.110)		(0.117)	
(old*old)_extra			0.160***		0.169***				0.358***		0.434***	
			(0.032)		(0.034)				(0.075)		(0.080)	
(new*new)_intra			-0.279***	-0.200***					-0.024	-0.099		
			(0.065)	(0.069)					(0.041)	(0.067)		
(new*new)_extra			0.388***	0.193***					-0.424***	-0.087		
			(0.046)	(0.039)					(0.142)	(0.152)		
old*never			-0.002		-0.001				-0.317***		-0.323***	
			(0.037)		(0.037)				(0.053)		(0.055)	
never*old			0.081*		0.086**							
			(0.037)		(0.037)							
new*never			-0.034	-0.006					-0.226***	0.048		
			(0.047)	(0.059)					(0.074)	(0.090)		
never*new			-0.034	0.078								
			(0.049)	(0.051)								

(continued)

Table 3. (continued)

Variable	Citations						Collaborations					
	(1a)	(2a)	(3a)	(4a)	(5a)	(6a)	(1b)	(2b)	(3b)	(4b)	(5b)	(6b)
(old*new)_enl86					0.162 **	0.115					0.880 ***	0.551 ***
					(0.059)	(0.086)					(0.146)	(0.214)
(new*old)_enl86					0.153 **	-0.010						
					(0.069)	(0.090)						
(new*new)_intra_enl86					0.420 ***	1.073					0.712 **	0.989
					(0.155)	(1.029)					(0.289)	(0.703)
(new*new)_extra_enl86					1.695 ***	11.765 ***					1.032	10.732 ***
					(0.453)	(0.708)					(0.677)	(0.744)
(new*never)_enl86					-0.248 **	-0.320					-1.051 ***	-0.718
					(0.111)	(0.268)					(0.219)	(0.440)
(never*new)_enl86					-0.334 ***	-0.166						
					(0.094)	(0.265)						
(old*new)_enl95					-0.004	-0.017					0.236 ***	0.260 ***
					(0.018)	(0.023)					(0.042)	(0.048)
(new*old)_enl95					0.270 ***	0.224 ***						
					(0.019)	(0.026)						
(new*new)_intra_enl95					-0.306 ***	-0.202 ***					-0.031	-0.100
					(0.066)	(0.069)					(0.041)	(0.067)
(new*new)_extra_enl95					0.378 ***	0.193 ***					-0.467 ***	-0.088
					(0.047)	(0.039)					(0.145)	(0.153)
(new*never)_enl95					-0.019	-0.000					-0.102	0.058
					(0.049)	(0.059)					(0.075)	(0.091)
(never*new)_enl95					0.042	0.084						
					(0.052)	(0.052)						

(continued)

Table 3. (continued)

Variable	Citations						Collaborations					
	(1a)	(2a)	(3a)	(4a)	(5a)	(6a)	(1b)	(2b)	(3b)	(4b)	(5b)	(6b)
ln(P_i)	1.074 ***	1.047 ***	1.06 ***	1.034 ***	1.061 ***	1.035 ***	0.667 ***	0.683 ***	0.674 ***	0.690 ***	0.673 ***	0.689 ***
	(0.013)	(0.020)	(0.013)	(0.019)	(0.013)	(0.019)	(0.029)	(0.034)	(0.030)	(0.034)	(0.030)	(0.034)
ln(P_j)	1.081 ***	1.208 ***	1.099 ***	1.043 ***	1.098 ***	1.042 ***	0.699 ***	0.705 ***	0.700 ***	0.706 ***	0.700 ***	0.706 ***
	(0.015)	(0.025)	(0.016)	(0.024)	(0.015)	(0.024)	(0.025)	(0.029)	(0.025)	(0.029)	(0.025)	(0.029)
Tech	1.924 ***	0.951 ***	1.912 ***	0.958 ***	1.915 ***	0.96 ***	1.465 ***	1.028 ***	1.482 ***	1.024 ***	1.482 ***	1.024 ***
	(0.021)	(0.042)	(0.021)	(0.041)	(0.021)	(0.041)	(0.049)	(0.063)	(0.049)	(0.063)	(0.049)	(0.063)
National	0.377 ***		0.771 ***		0.767 ***		1.688 ***		1.214 ***		1.211 ***	
	(0.011)		(0.038)		(0.038)		(0.039)		(0.068)		(0.069)	
Bord	0.194 ***		0.205 ***		0.205 ***		0.702 ***		0.688 ***		0.689 ***	
	(0.012)		(0.012)		(0.011)		(0.020)		(0.020)		(0.020)	
Lang	0.272 ***		0.275 ***		0.274 ***		0.529 ***		0.542 ***		0.536 ***	
	(0.010)		(0.010)		(0.010)		(0.030)		(0.030)		(0.030)	
constant	-12.670 ***		-12.849 ***		-12.843 ***		-3.095 ***		-2.560 ***		-2.522 ***	
	(0.211)		(0.212)		(0.212)		(0.322)		(0.327)		(0.328)	
ln(dist) * year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
dummy region i	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
dummy region j	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
dummy region-pair ij	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Log Pseudo-Lik.	-525645.16	-426874.40	-525162.64	-426701.37	-525090.49	-426692.74	-146369.44	-99111.85	-146208.30	-99068.46	-146131.89	-99065.37
Number of groups	36290	27046	36290	27046	36290	27046	18145	8067	18145	8067	18145	8067
Number of observations	725800	540920	725800	540920	725800	540920	362900	161340	362900	161340	362900	161340

Note: The variables *National*, *Bord* and *Lang* are not included in PPML fixed-effects estimates because are time invariant. PPML fixed-effects estimates do not include the intercept α (Arellano and Honoré, 2001). The dummy region *i* (*j*) is equal to 1 for region *i* (*j*) and 0 elsewhere; the dummy region-pair *ij* is equal to 1 for region-pair *ij* and 0 elsewhere. For the sake of clarity the coefficient values of region dummies, of region-pair dummies, of year dummies and of interaction terms between geographical distance and year dummies are not reported; ***, ** and * indicate significance at 1, 5 and 10 %, respectively; robust standard errors are in parentheses. Intraregional observations are excluded.

FIGURE CAPTIONS

Figure 1. Matrix of the combinations between European regions

Figure 2. European integration and sub group of regions

Figure 3. Interregional and international patent citations and collaborations

Figure 4. Coefficient estimates and 95% confidence intervals of distance effect -Equation [4] -

Figure 5. Coefficient estimates and 95% confidence intervals of distance and national border effects -Equation [5] -