Department of Economics and Statistics
"COGNETTI DE MARTIIS"

Working Paper Series

22/13

EUROPEAN INTEGRATION AND KNOWLEDGE FLOWS ACROSS EUROPEAN REGIONS

RICCARDO CAPPELLI and FABIO MONTOBBIO

The Department of Economics and Statistics "Cognetti de Martiis" publishes research papers authored by members and guests of the Department and of its research centers. ISSN: 2039-4004

Department of Economics and Statistics "Cognetti de Martiis"
Campus Luigi Einaudi, Lungo Dora Siena 100A, 10153 Torino (Italy)
www.unito.it/de

European Integration and Knowledge Flows across European Regions

Riccardo Cappelli[a] and Fabio Montobbio[b,c]

a) *Department of Economic Geography, Faculty of Geosciences, Utrecht University, Heidelberglaan 2, 3584 CS Utrecht, The Netherlands. Email: r.cappelli@uu.nl*

b) Department of Economics "Cognetti De Martiis", University of Turin, Lungo Dora Siena 100 A, 10153 Turin, Italy. Email: fabio.montobbio@unito.it

c) KITES, Bocconi University, Via Guglielmo Roentgen, 20136 Milan,Italy.

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

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 imlications. 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 absorptve 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 exploits a modified version of the gravity model and the results are obtained through Poisson pseudo maximum likelihood (PPML) estimates. The estimation strategy is in two steps. First, it is analysed the evolution over time of the impact of geographical distance and national border on the two measures of knowledge flows through separate cross-sectional estimates for sub-periods and panel estimates. Second, it is investigated 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 account of 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 decresed 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 bocoming 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 in more detail the literature on the diffusion of knowledge and the theoretical justification for our analysis. The third and fourth sections respectively present the gravity model used in the estimates and the methodology adopted in order to resolve some sources of bias. The fifth section presents the data used in this paper. The sixth Section presents and discusses the results of the estimates. The seventh 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 (e.g. Bottazzi and Peri, 2003). In what follow this section focuses only on those papers that have measured knowledge flows using inventor citations and inventor collaborations.[1]

Geographic localization of knowledge flows

Empirical analysis based on patent citations widely shows that the diffusion of knowledge is a geographically localized phenomenon. 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 the same 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 than the body of work on patent citations 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), using EPO data for 29 OECD countries (including 21 European countries) and a gravity model show, that the possibility of collaboration between inventors residing in two different countries decreases as the geographical distance between them (expressed as sharing the same territorial border) increases. Moreover, they show that two country

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

are more likely to collaborate if they are close in the technology space. These results are confirmed by Picci (2010). He uses a series of datasets (EPO, USPTO and other national patent office data) for 42 countries (including 14 European countries) and applies a gravity model, to show that the possibility of international collaboration is affected by geographical proximity, both physical and in the form of common territorial border, and technological proximity. Differently from Guellec and van Pottelsberghe de la Potterie (2001), he find that international collaborations are positively affected by the EU membership. At a finer geographical level, Hoeckman et al (2009), using both patent co-inventorship and publication co-authorship data for two sectors (biotechnology and semicounductors), covering 1316 NUTS3 regions of the EU27 member states, and Norway and Switzerland, provide an evidence that interregional collaboration are hampered by geographical factors (i.e. physical distance) and institutional factors (i.e. country border).

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 and there is also no consensus: some show that the diffusion of knowledge is occurring in a more localized way than in the past, while others show the opposite. Johnson et al. (2006), using USPTO data for the period 1975-1999 and using at Tobit model with geographical distance as the dependent variable, a time trend of variable of interest, and a set of control variables, 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 increase 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 makes 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, capital and people, and to the creation of supranational institutions which promote or coordinates common policies for the membership 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 considers only static effects and, therefore, does not investigate whether greater integration between countries has an effect on reducing the pre-existing barriers to knowledge flows, which would require dynamic analysis that takes account of the factors that over time contribute to the integration of countries.

Based on the hypothesis that a reduction in economic and institutional barriers may affect knowledge flows, this paper analyses the impact of the EU integration through the EU enlargement processes that occurred during the period investigated. This allows to distinguish between the effect on the three types of regions involved, i.e. new members, old members, and non-EU members. To the knowledge of the author 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

The econometric model used to analyse knowledge flows among regions of 29 European countries is a modified version of the gravity model. The gravity model is widely used in work on bilateral trade between countries (see e.g. Anderson and van Wincoop, 2003) and in the study of 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. More generally:

$$[1]\ C_{ijt} = \alpha\, P_{it}^{\beta}\, P_{jt}^{\cup}\, dist_{ij}^{\gamma}$$

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, α is a constant, P_{it} and P_{jt} are total numbers of patents (inventive mass) for the two regions, and $dist_{ij}$ is the geographical distance between the two regions.

This study identifies a citation from region j to region i as occurring when the citing patent as 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 developed by a single firm are not considered. Self-citations within firms are not considered externalities. Self-citations between inventors are also excluded because, by definition, this 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 the pairs of patents where the time lag between cited and citing patent is four years or less are considered.[2] For instance, C_{ijt} is the total number of citations contained in patents developed in region j (knowledge-receiving region) during the period t-$(t+4)$ and directed to patents developed in region i (knowledge-generating region) in period t.[3] Thus, the

[2] 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.

[3] 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 t-$(t+4)$.

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.

In order to reduce the impact of outliers, the variables for the inventive mass and distance are expressed in logarithmic form. The conditional mean of C_{ijt} can be expressed by the following equation:

$$[2] \quad E(C_{ijt}|X_{ijt}) = \mu_{ijt} = exp\big[\alpha + \beta\, ln(P_{it}) + \upsilon\, ln\big(P_{jt}\big) + \gamma\, ln\big(dist_{ij}\big) + \varrho\, Tech_{ijt} + \sigma\, Langij + \Omega\ Nationalij + \varrho\, Bordij$$

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 and that will be used in panel estimates. This gives the following equation:

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

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:

$$[4]\quad \mu_{ijt} = exp[\alpha + \beta\, ln(P_{it}) + \upsilon\, ln(P_{jt}) + \gamma_t\, ln(dist_{ij}) + \varrho\, Tech_{ijt} + \sigma\, Lang_{ij} + \Omega\, National_{ij} + \rho\, Bord_{ij} + \varrho_i + \eta_j + \Theta_t$$

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:

$$[5]\quad \mu_{ijt} = exp[\alpha + \beta\, ln(P_{it}) + \upsilon\, ln(P_{jt}) + \gamma_t\, ln(dist_{ij}) + \varrho\, Tech_{ijt} + \sigma\, Lang_{ij} + \Omega_t\, National_{ij} + \rho\, Bord_{ij} + \varrho_i + \eta_j + \Theta_t$$

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:

$$[6]\quad \mu_{ijt} = exp[\alpha + \beta\, ln(P_{it}) + \upsilon\, ln(P_{jt}) + \gamma_t\, ln(dist_{ij}) + \varrho\, Tech_{ijt} + \sigma\, Lang_{ij} + \Omega\, National_{ij} + \rho\, Bord_{ij} + \varphi\, EUboth_{ijt} + \omega\, EUone_{ijt} + \varrho_i + \eta_j + \Theta_t$$

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 one 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 one 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 one 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 one 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 one. 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 can be estimated using OLS on the log-linear version of the previous equations. However, this procedure has some problems which can lead to biased estimates. 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. These observations are treated as missing in the estimates which introduces bias in the coefficients estimated. Gravity models also have an inherent problem of heteroschedasticity, which can lead to biased estimates. To jointly address these issues a PPML estimator is aprticularly appropriate (Santos Silva and Tenreyro, 2006).[4]

The effect of European enlargement on knowledge flows is estimated using PPML estimates with regional dummies (covering both knowledge generating and knowledge receiving regions)

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

(equation [6])[5] and PPML fixed-effects (region-pairs 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 our analysis. In fact, the inclusion of pair region 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 mean that it is not possible estimate the effect for time invarying 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).[6] The pair of regions excluded by fixed-effects analysis are shaded grey in Figure 2.

5 Data

To construct the two measures of knowledge flows, i.e. patent citations and inventor collaborations, it is used the information contained in EPO patents (KITES and OECD REGPAT database). Address of inventor is used to assign a patent to the territory where it was developed.

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 the 285 regions of the aforementioned 29 European countries. In 2000, the last year of our analysis, there are 15 countries belonging to EU and 14 not EU member countries. However, the estimates consider only those regions that have at least one patent in each year of the period in question because if a region has no patents then, by definition, it cannot have a regional knowledge flow.[7] Thus, the final dataset contains patents data from 191 regions (169 regions in the EU15 countries, 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 intra-regional 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)

[5] 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 dummy regions 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.

[6] 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.

[7] 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. However, the estimation results obtained using the sample with all the regions are very similar. These results are available from the authors on request.

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. Changes in the coefficients of geographical distance and national border allow to assess whether the diffusion of knowledge is more or less circumscribed in the space than in the past. 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 patent citations (left side) and technological collaborations (right side) as percentages of the total. 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). Figure 4 shows the distribution over time of international patent citations and technological collaboration as percentages of the total (regional excluded). The international patent citations (left side) decrease over time (from 67.4% in 1981 to 58.2% in 2000), while international collaborations (right side) increase over time (from 11.9% in 1981 to 22.1% in 2000). These figures indicate 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 with inventor citations becoming more localized than in the past, and the reverse applying to inventor collaborations. These aspects will be confirmed in succeeding analysis.

- Figure 3 about here -

- Figure 4 about here -

Cross-section estimates for the whole period

The results of the estimates of equation [3] using aggregated data for the whole period analysed (1981-2000) are shown in Table 1. Table 1 presents each measure of knowledge flows in separate columns: first, in line with the extant literature, columns 1a and 1b 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 for estimates that include intraregional knowledge flows. The number of observations for the first column of the patent citations (1a) is 36,290 [(191*191)-190] and the number of observations for the first column of inventor collaborations (1b) is 18,145 [((191*191)-191)/2]. The difference in the number of observations between the patent citations (1a, 2a) and the inventor collaborations (1b, 2b) columns are due to the different characteristics of these two variables, i.e. unidirectional or bidirectional knowledge flows measure. The difference between the number of observations in the first set of columns (1a, 1b) and the second set of columns (2a, 2b) is equal to the number of regions (i.e. 191).

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 is weaker with increasing geographical distance. Moreover, adjacency (*Bord*) exacerbate the role of geographical proximity in determining knowledge flows between regions (both for citations and collaborations), i.e. 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*, irrespective of controlling for geographic proximity (*dist* and *Bord*) or technological proximity (*Tech*), can be interpreted as being due to social, institutional or other country specific reasons which lead to greater knowledge flows within than between countries. Also language matters, as the diffusion of knowledge is greater if the regions share a common language.

Another interesting result is the difference in the coefficient values for the variables for geographical (*dist* and *Bord*), social and institutional (*Lang* and *National*) proximity for both measures of knowledge flows. The coefficient values of these variables for collaboration are greater than for patent citations, meaning that technological collaboration tends to be more geographically localized than patent citation. This is consistent with inventor citations not requiring face-to-face contact. For instance, an inventor can know about the invention cited simply by reading the description contained in the patent document. To sum up, the analysis confirms the hypothesis that geographical, institutional and other country specific factors are more important for inventor citations than for inventor collaborations.

Finally, for the coefficient values of the variable *Tech*, technological proximity is more important for inventor citations than for inventor collaboration. This is consistent with the very many citations that are added by patent examiners, often aimed at limiting inventors' claims to novelty in a technological field. On the other hand, 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 new knowledge.

As a robustness check, it is estimated the above specifications including the observations for intraregional knowledge flows. The variable *region*[8] is also included to take account of 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-sections for different sub periods

The next step is analysis of the evolution over time of the coefficients of the above variables. 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. As above, the number of observations for inventor citations is 36,290 [(191*191)-191], and the number of observations for inventor collaborations is 18,145 [((191*191)-191)/2]. Table 2 presents the results of these estimates. 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

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

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).[9] 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 the 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.

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 corroborates 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 agglomaration 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 are made to check the previous results on the dynamics of the distance effect. The number of observations for patent citations is 725,800 [((191*191)-191)*20], and the number of observations for inventor collaborations is 362,900 [(((191*191)-191)/2) *20].

The results of equation [4] confirm the trends of the cross-sectional estimates. Figure 5 reports coefficient values (and the confidence interval at ± 95%) for the variable *dist.* For inventor citations the distance effect increases in abslolute value over time, while it decreases for inventor collaborations .

- Figure 5 about here -

As a further check, estimates of distance and national border effects varying over time simultaneously (equation [5]) are presented in Figure 6. It shows the results for the distance effects (graph a) and for the national border effects (graph b). National border effects decrease for inventor collaborations, while the opposite occurs for inventor citations[10]. 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 6 about here –

[9] As a robustness check, it is performed sub-period estimates excluding the citations included by the patent examiners. The results, available from the authors on request, are quite similar.

[10] As a robustness check, it is controlled for the evolution over time both of the region's internal technological specialisation and of 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.

With regard to the distance effect, the decreasing trend for the inventor collaborations is confirmed in Figure 6, 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.

Panel estimates of the impact of European integration

This paper analyses the impact of the EU integration on knowledge flows as result of the European enlargement processes. Coherent with the trade literature (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. Analysis of the effect 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. For the PPML estimates without pair fixed effects, the number of observations for inventor citations (columns 1a, 3a and 5a) is 725,800 [((191*191)-191)*20], and the number of observations for inventor collaborations (columns 1b, 3b and 5b) is 362,900 [(((191*191)-191)/2) *20]. The lower number of observations for the PPML fixed effects estimates are due to the fact that in the fixed effect estimates the observations for the pairs of region with zero variations over time of the dependent variables are dropped. Thus, the number of observations for the patent citations (columns 2a, 4a and 6a) is 540,920 (i.e. 725,800-184,880), while the number of observations for the inventor collaborations (columns 2b, 4b and 6b) 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 invarying variables.

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-pairs 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 coeffcients 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*).[11]

To sum up, European integration had a significant effect on reducing barriers to knowledge flows between new and old EU members. However, 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 the patterns of knowledge diffusion among European regions based on 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. On the one hand, inventors tend to collaborate more with other international inventors, but on the other hand, the tendency to cite national inventors increases. The evolution over time of the distance effect, which decreases only for inventor collaborations, confirms that inventor collaborations are becoming less localized, while the reverse is true for inventor citations.

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 and only for knowledge generated in new member regions (Austria, Finland and Sweden) that is more used in old EU members.

[11] 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.

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.

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., 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.

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.

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., 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 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.

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, forthcoming.

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

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.

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.

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

Variable	*Citations* (1a)		(2a)		*Collaborations* (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	
regional observations	excluded		included		excluded		included	
Log Pseudo-likelihood	-97906.8		-110892.0		-47264.1		-62547.1	
R-squared	0.955		0.926		0.908		0.983	
Number of regions	191		191		191		191	
N. observations	36290		36481		18145		18336	

Note: ***, ** and * indicate significance at 1, 5 and 10 %, respectively.

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

	Citations				*Collaborations*			
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
R-squared	0.894	0.911	0.914	0.934	0.925	0.914	0.906	0.899
Number of regions	191	191	191	191	191	191	191	191
N. observations	36290	36290	36290	36290	18145	18145	18145	18145

Note: ***, ** and * indicate significance at 1, 5 and 10 percent, respectively.

Table 3. European Integration - PPML with and without region-pairs 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 (1a)	(2a)	(3a)	(4a)	(5a)	(6a)	Collaborations (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(dist) * time dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ln(P_i) and ln(P_j)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Tech	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
National	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Bord	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Lang	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
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
region-pairs dummy	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Number of groups	36290	27046	36290	27046	36290	27046	18145	8067	18145	8067	18145	8067
Number of regions	191	191	191	191	191	191	191	191	191	191	191	191
Nunber of observations	725800	540920	725800	540920	725800	540920	362900	161340	362900	161340	362900	161340

Note: ***, ** and * indicate significance at 1, 5 and 10 %, respectively.

FIGURES

	old_j	new_{jt}	$never_j$
old_i	old_i*old_j	old_i*new_{jt}	$old_i*never_j$
new_{it}	$new_{it}*old_j$	$new_{it}*new_{jt}$	$new_{it}*never_j$
$never_i$	$never_i*old_j$	$never_i*new_{jt}$	$never_i*never_j$

old	*new*	*never*
time constant dummy variable which is set equal to one for regions of countries that are EU member since 1981. These countries are:Belgium; Denmark; Germany; Greece; France; Ireland; Italy; Luxembourg; Nederland; United Kingdom.	time varying dummy variables which is set equal to one for regions of EU entering countries for the EU annexation year and the following years. These countries are: Spain; Portugal; Austria; Finland; Sweden.	time constant dummy variable which is set equal to one for regions of counties that are not EU member. These countries are: Bulgaria; Cyprus; Czech Republic; Estonia; Hungary; Latvia; Lithuania; Malta; Norway; Poland; Romania; Slovakia; Slovenia; Switzerland.
Note: the suffixes *i*, *j* and *t* are omitted for sake of clarity.		

Figure 1. Matrix of the combinations between European regions

Initial variable	*Description*	*EU-based typology of regions*	*Final variable*
$EUboth_{ijt}$	knowledge flows between EU members	knowledge flows between "old" and "new" regions (extra-national flows)	old_i*new_{jt} $new_{it}*old_j$
		knowledge flows between "new" regions (intra-national and extra-national flows)	$(new_{it}*new_{jt})$_intra $(new_{it}*new_{jt})$_extra
		knowledge flows between "old" regions (intra-national and extra-national flows)	(old_i*old_j)_intra (old_i*old_j)_extra
$EUone_{ijt}$	knowledge flows between EU members and not EU members	knowledge flows between "old" and "never" regions (extra-national flows)	$old_i*never_j$ $never_i*old_j$
		knowledge flows between "new" and "never" regions (extra-national flows)	$new_{it}*never_j$ $never_i*new_{jt}$

Figure 2. European integration and sub group of regions

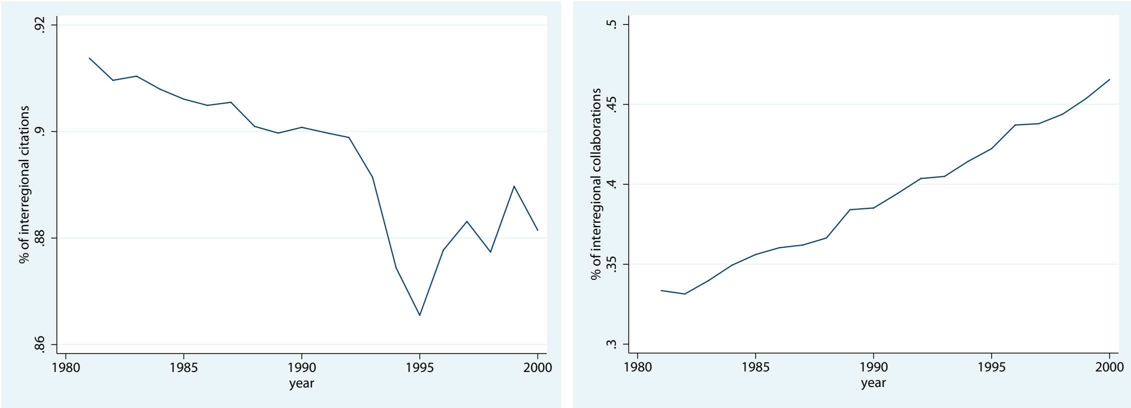


Figure 3. Interregional patent citations and collaborations in percentage on total

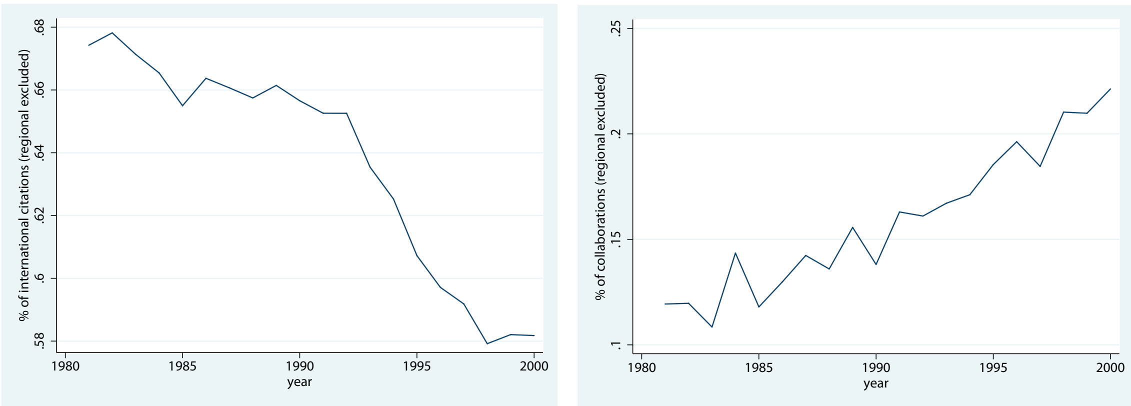


Figure 4. International patent citations and collaborations in percentage on total (regional excluded)

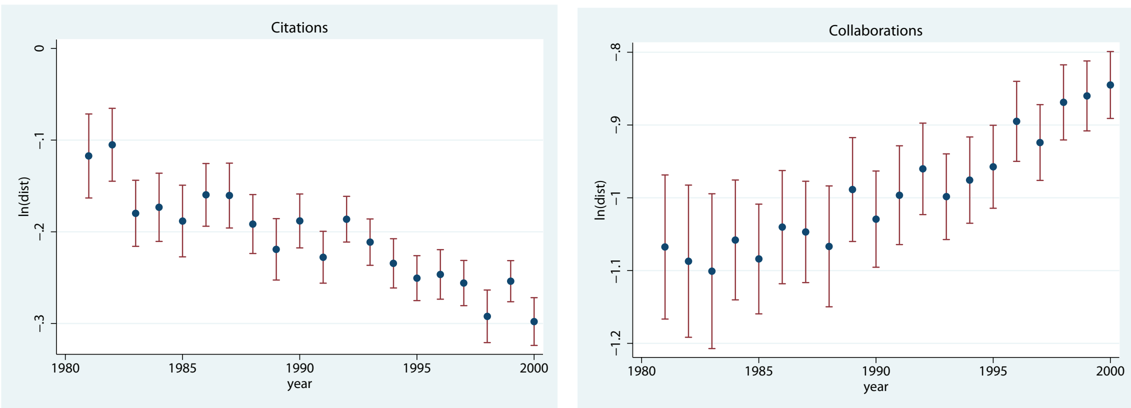
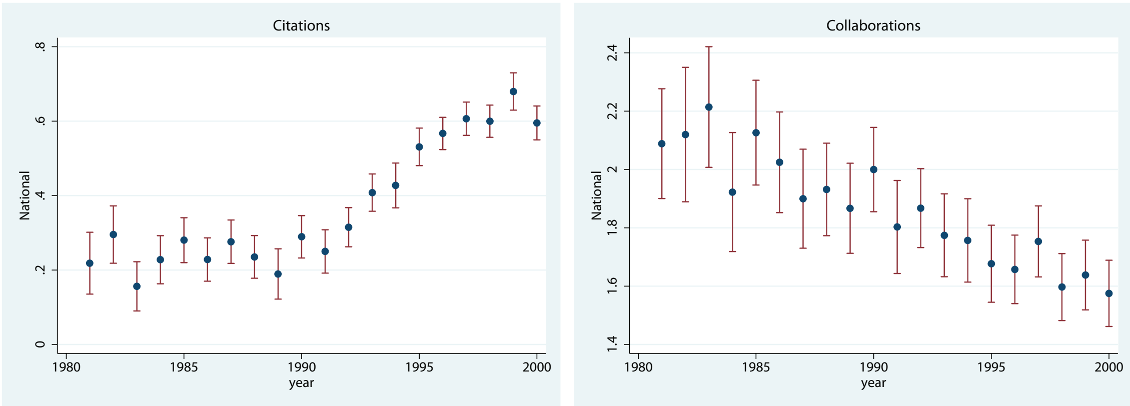


Figure 5. Coefficient estimates and 95% confidence intervals of the distance effect - Equation [4] -

a) evolution over time of the distance effect



b) evolution over time of the national border effect

Figure 6. Coefficient estimates and 95% confidence intervals of the distance and national border effect - Equation [5] -

Supplemental Material

A. The construction of the data on knowledge flows

The initial dataset was obtained by extrapolating from the KITES dataset the information on patents (EP number, priority year), inventors (name and address), applicants (name and address) and citations. The dataset contains data on patents registered in various patent offices, national or supranational, but for our analysis we consider only the patents registered at the EPO. Patents are assigned to European NUTS 2 regions (Eurostat, 2007) based on inventor's address. Patents with more than one inventor are assigned to regions based on the addresses of all inventor addresses. For instance, a patent with two inventors from two different regions is assigned to both regions.

In most cases, the connection between inventor's address and NUTS 2 region is contained in the KITES dataset. However, in order to avoid bias in the estimates, it is reduced the number of inventors without a NUTS 2 region assigned. To do this, it is merged the KITES and the OECD REGPAT datasets based on EPO publication number and inventor name in order to obtain full correspondence between datasets. This allowed to identify the NUTS 2 region for some of the inventors. For the inventors without a NUTS 2 designation, the NUTS 2 region is manually assigned based on inventor's place of residence and postcode.

At the end of the above procedures, the percentage of inventors without a NUTS 2 is approximately 0.8% of the total and is quite similar across time periods and countries.[12]

The procedures adopted for the construction of the dependent variables are described in section 3.

[12] Although we do not know the inventor's NUTS 2 region, we know the inventor's country of residence and other details of the patent developed.

B. Definition, source and descriptive statistics of the variables

Table B1 reports the definitions and sources of the variables used in the analysis.

Table B1. Definition and sources of the variables

Variable	*Definition*	*Source*
Citations *ijt*	Number of patents (at time *t*) of inventors residing in region *i* cited by patents (at time *t-(t+4)*) with at least an inventor residing in region *j*. Patents with more than one inventor residing in the same region (*i* or *j*), citations are counted only once.	KITES/OECD REGPAT
Collaborations *ijt*	Number of patents with at least an inventor residing in region *j* and an inventor residing in region *i*. Patents with more than one inventor residing in the same region (*i* or *j*), collaborations are counted only once.	KITES/OECD REGPAT
P *it*	Number of patents with at least an inventor residing in country *i*.	KITES/OECD REGPAT
P *jt*	Number of patents with at least one inventor residing in country *j*. For patent citations we consider a temporal window of four years (*t-(t+4)*).	KITES/OECD REGPAT
Tech *ijt*	Jaffe (1986) index based on 30 technology classes (OST, 2004). It is an indicator of the technological proximity between region *i* and region *j*.	KITES
National *ij*	Dummy equal to 1 if the two regions are located in the same country.	KITES/OECD REGPAT
dist *ij*	Geographical distance between two regions, calculated using the great circle distance method on the basis of the geographical coordinates of the centre point of the regions. Intra-regional distance is calculated as two thirds of the radius of the regional geographic size.	EUROSTAT/GISCO
Bord *ij*	Dummy equal to 1 if the two regions are neighbours.	Authors' elaborations
Lang *ij*	Dummy equal to 1 if the two regions have the same official language.	Authors' elaborations
region *ij*	Dummy equal to 1 for intra-regional knowledge flows (*i=j*).	KITES/OECD REGPAT
EU variables		
EUboth *ijt*	Time varying dummy equal to 1 if the two regions are from EU member states.	Authors' elaborations
EUone *ijt*	Time varying dummy equal to 1 if only one region is from EU member states.	Authors' elaborations
old*new *ijt*	Time varying dummy equal to 1 if knowledge flows from an *old* region (i.e. region of countries that were EU member before 1981) to a *new* region (i.e. region of EU entering states).	Authors' elaborations
new*old *ijt*	Time varying dummy equal to 1 if knowledge flows from a *new* region to an *old* region.	Authors' elaborations

(continued)

Table B1. (continued)

Variable	*Definition*	*Source*
(old*old)_intra *ij*	Time constant dummy equal to 1 if knowledge flows within countries that were EU member before 1981.	Authors' elaborations
(old*old)_extra *ij*	Time constant dummy equal to 1 if knowledge flows between countries that were EU member before 1981.	Authors' elaborations
old*never *ij*	Time constant dummy equal to 1 if knowledge flows from an *old* region to a never region (i.e. region of non-EU member countries).	Authors' elaborations
never*old *ij*	Time constant dummy equal to 1 if knowledge flows from a *never* region to an *old* region.	Authors' elaborations
(new*new)_intra *ijt*	Time varying dummy equal to 1 if knowledge flows within EU new member states.	Authors' elaborations
(new*new)_extra *ijt*	Time varying dummy equal to 1 if knowledge flows between EU new member states.	Authors' elaborations
new*never *ijt*	Time varying dummy equal to 1 if knowledge flows from a *new* region to a *never* region.	Authors' elaborations
never*new *ijt*	Time varying dummy equal to 1 if knowledge flows from a *never* region to a *new* region.	Authors' elaborations
(old*new)_enl86 *ijt*	Time varying dummy. The variable is *old*new*, but *new* regions include only regions of countries that joined the EU in 1986 (Spain and Portugal).	Authors' elaborations
(new*old)_enl86 *ijt*	Time varying dummy. The variable is *new*old*, but *new* regions include only regions of countries that joined the EU in 1986 (Spain and Portugal).	Authors' elaborations
(old*new)_enl95 *ijt*	Time varying dummy. The variable is *old*new*, but *new* regions include only regions of countries that joined the EU in 1995 (Austria, Finland and Sweden).	Authors' elaborations
(new*old)_enl95 *ijt*	Time varying dummy. The variable is *new*old*, but *new* regions include only regions of countries that joined the EU in 1995 (Austria, Finland and Sweden).	Authors' elaborations
(new*new)_intra_enl86 *ijt*	Time varying dummy. The variable is *(new*new)_intra*, but *new* regions include only regions of countries that entered the EU in 1986 (Spain and Portugal).	Authors' elaborations
(new*new)_intra_enl95 *ijt*	Time varying dummy. The variable is *(new*new)_intra*, but *new* regions include only regions of countries that joined the EU in the 1995 (Austria, Finland and Sweden).	Authors' elaborations

(continued)

Table B1. (continued)

Variable	*Definition*	*Source*
(new*new)_extra_enl86 *ijt*	Time varying dummy. The variable is *(new*new)_extra*, but *new* regions include only regions of countries that entered the EU in 1986 (Spain and Portugal).	Authors' elaborations
(new*new)_extra_enl95 *ijt*	Time varying dummy. The variable is (new*new)_extra, but *new* regions include only regions of countries that joined the EU in 1995 (Austria, Finland and Sweden).	Authors' elaborations
(new*never)_enl86 *ijt*	Time varying dummy. The variable is *new*never*, but *new* regions include only regions of states that joined the EU in 1986 (Spain and Portugal).	Authors' elaborations
(new*never)_enl95 *ijt*	Time varying dummy. The variable is *new*never*, but *new* regions include only regions of states that joined the EU in 1995 (Austria, Finland and Sweden).	Authors' elaborations
(never*new)_enl86 *ijt*	Time varying dummy. The variable is *never*new*, but *new* regions include only regions of states that joined the EU in 1986 (Spain and Portugal).	Authors' elaborations
(never*new)_enl95 *ijt*	Time varying dummy. The variable is *never*new*, but *new* regions include only regions of countries that joined the EU in 1995 (Austria, Finland and Sweden).	Authors' elaborations

In the Table B2 are shown the descriptive statistics of the variables used in our analysis (intra-regional observations are excluded).

Table B2. Descriptive statistics (period 1981-2000)

Variable	Citations					Collaborations				
	Obs	Mean	Std. Dev.	Min	Max	Obs	Mean	Std. Dev.	Min	Max
C_{ij}	725800	0.82	3.75	0	327	362900	0.57	5.82	0	593
P_i	725800	210.18	351.99	1	3742	362900	246.83	401.44	1	3742
P_j	725800	1190.01	1939.71	10	18784	362900	173.53	289.76	1	3742
$\log(P_i)$	725800	4.47	1.41	0	8.23	362900	4.59	1.45	0	8.23
$\log(P_j)$	725800	6.25	1.35	2.30	9.84	362900	4.35	1.35	0	8.23
Tech	725800	0.55	0.18	0	0.99	362900	0.5	0.19	0	1
National	725800	0.11	0.31	0	1	362900	0.11	0.31	0	1
dist	725800	947.27	566.52	6.56	3775.18	362900	947.27	566.52	6.56	3775.18
log(dist)	725800	6.64	0.71	1.88	8.24	362900	6.64	0.71	1.88	8.24
Bord	725800	0.02	0.15	0	1	362900	0.02	0.15	0	1
Lang	725800	0.18	0.38	0	1	362900	0.18	0.38	0	1
EUboth	725800	0.64	0.48	0	1	362900	0.64	0.48	0	1
EUone	725800	0.19	0.39	0	1	362900	0.19	0.39	0	1
old*new	725800	0.04	0.20	0	1	362900	0.09	0.28	0	1
new*old	725800	0.04	0.20	0	1					
(old*old)_intra	725800	0.10	0.30	0	1	362900	0.10	0.30	0	1
(old*old)_extra	725800	0.45	0.50	0	1	362900	0.45	0.50	0	1
old*never	725800	0.09	0.28	0	1	362900	0.17	0.38	0	1
never*old	725800	0.09	0.28	0	1					
(new*new)_intra	725800	0.00	0.04	0	1	362900	0.00	0.04	0	1

(continued)

Table B2. (continued)

Variable	Citations					Collaborations				
	Obs	Mean	Std. Dev.	Min	Max	Obs	Mean	Std. Dev.	Min	Max
(new*new)_extra	725800	0.00	0.05	0	1	362900	0.00	0.05	0	1
new*never	725800	0.01	0.08	0	1	362900	0.01	0.11	0	1
never*new	725800	0.01	0.08	0	1					
(old*new)_86	725800	0.02	0.13	0	1	362900	0.04	0.18	0	1
(new*old)_86	725800	0.02	0.13	0	1					
(old*new)_95	725800	0.03	0.16	0	1	362900	0.05	0.22	0	1
(new*old)_95	725800	0.03	0.16	0	1					
(new*new)_intra_86	725800	0.00	0.02	0	1	362900	0.00	0.02	0	1
(new*new)_intra_95	725800	0.00	0.03	0	1	362900	0.00	0.03	0	1
(new*new)_extra_86	725800	0.00	0.01	0	1	362900	0.00	0.01	0	1
(new*new)_extra_95	725800	0.00	0.05	0	1	362900	0.00	0.05	0	1
(new*never)_86	725800	0.00	0.05	0	1	362900	0.01	0.07	0	1
(new*never)_95	725800	0.00	0.06	0	1	362900	0.01	0.09	0	1
(never*new)_86	725800	0.00	0.05	0	1					
(never*new)_95	725800	0.00	0.06	0	1					

C. Cross-section estimates for the whole sample and the restricted sample

Table C1 compares the results of the estimates of equation [3] obtained for the whole[13] sample of 281 regions and the restricted sample of 191 regions (aggregate data).

Table C1. Citations and collaborations (restricted and whole sample) - PPML

	Citations				*Collaborations*			
Variable	restricted		whole		restricted		whole	
Tech	2.138	***	2.157	***	2.014	***	2.031	***
	(0.075)		(0.074)		(0.195)		(0.190)	
ln (dist)	-0.243	***	-0.247	***	-0.828	***	-0.830	***
	(0.015)		(0.015)		(0.060)		(0.058)	
Lang	0.225	***	0.230	***	0.398	***	0.434	***
	(0.023)		(0.023)		(0.096)		(0.096)	
National	0.454	***	0.449	***	1.791	***	1.793	***
	(0.024)		(0.024)		(0.128)		(0.127)	
Bord	0.152	***	0.150	***	0.733	***	0.726	***
	(0.026)		(0.026)		(0.078)		(0.076)	
region	0.351	***	0.344	***	0.374	***	0.369	***
	(0.068)		(0.068)		(0.081)		(0.079)	
constant	-2.493	***	-4.720	***	-0.051		-0.939	
	(0.199)		(0.237)		(0.401)		(0.754)	
dummy region i	Yes		Yes		Yes		Yes	
dummy region j	Yes		Yes		Yes		Yes	
regional observations	included		included		included		included	
Log Pseudo-likelihood	-110892.01		-124739.15		-62547.15		- 67897.319	
R-squared	0.926		0.927		0.983		0.983	
Number of regions	191		281		191		281	
N. observations	36481		78961		18336		39621	

Note: ***, ** and * indicate significance at 1, 5 and 10 %, respectively.

[13] We consider all regions with at least 1 EPO patent during the period analysed, which leaves 4 regions (285-281) without an EPO patent.

D. Robustness check excluding citations added by patent examiners

Table D1 compares the results of the sub-period estimates using the whole sample (inventor/applicant and examiner citations) and the restricted sample (only inventor/applicant citations). In our sample the share of patent citations attributable to the inventors/applicants are about 15%. Similar share (about 11%) is found in the sample used by Criscuolo and Verspagen (2008). If we exclude the first period were we have the 99% of observations with zero knowledge flows, the increasing trends in distance and national border effects are confirmed. In general, the coefficients obtained for distance and national border are higher than those showed in Table 2 and, thus, confirm the literature that applicant/inventor citations are more localized than EPO examiner citations (Criscuolo and Verspagen, 2008).

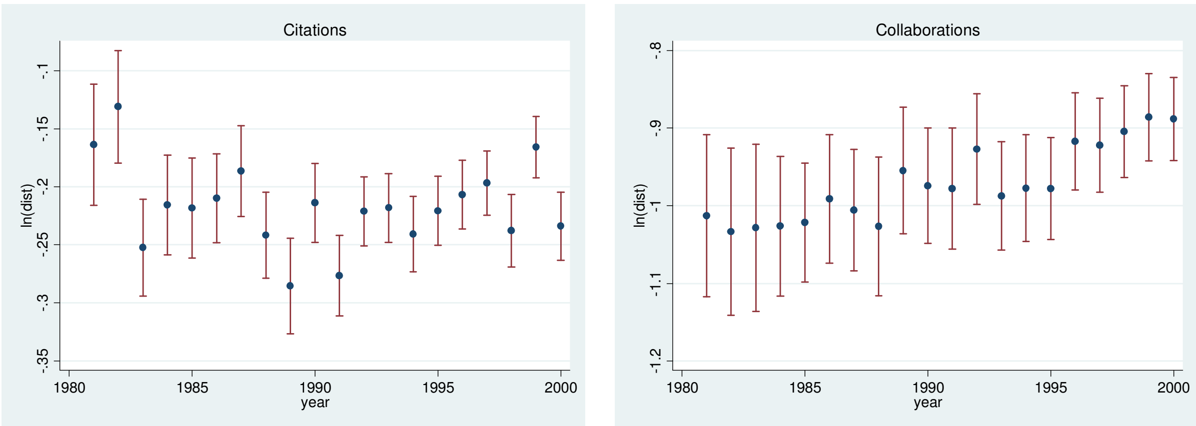
Tab D1. PPML estimates of equation [3] (sub-periods estimates) – Whole sample and restricted sample

	Whole sample (examiner and inventor/applicant citations)				*Restricted sample (inventor/applicant citations)*			
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 ***	2.262 ***	1.366 ***	2.374 ***	2.359 ***
	(0.645)	(0.064)	(0.060)	(0.055)	(0.545)	(0.303)	(0.093)	(0.082)
ln(dist)	-0.138 ***	-0.183 ***	-0.255 ***	-0.209 ***	-0.497 ***	-0.092 ***	-0.346 ***	-0.262 ***
	(0.016)	(0.016)	(0.015)	(0.014)	(0.157)	(0.077)	(0.026)	(0.024)
Lang	0.315 ***	0.194 ***	0.231 ***	0.190 ***	-0.103	0.462 ***	0.247 ***	0.204 ***
	(0.027)	(0.026)	(0.028)	(0.025)	(0.236)	(0.113)	(0.044)	(0.040)
National	0.298 ***	0.441 ***	0.389 ***	0.530 ***	0.797 ***	0.352 ***	0.457 ***	0.661 ***
	(0.027)	(0.027)	(0.032)	(0.028)	(0.254)	(0.133)	(0.046)	(0.043)
Bord	0.091 **	0.142 ***	0.165 ***	0.242 ***	0.010	0.430 ***	0.180 ***	0.298 ***
	(0.045)	(0.035)	(0.033)	(0.029)	(0.245)	(0.125)	(0.048)	(0.045)
constant	-4.798 ***	-5.096 ***	-3.741 ***	-3.262 ***	-10.592 ***	-15.330 ***	-10.003 ***	-5.137 ***
	(0.389)	(0.391)	(0.332)	(0.237)	(1.449)	(1.545)	(1.239)	(0.451)
Dummy region i	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dummy region j	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log Pseudo-likelihood	-37264.01	-48468.10	-60354.29	-71054.79	-1010.35	-4150.11	-27609.26	-35957.22
R-squared	0.894	0.911	0.914	0.934	0.212	0.335	0.821	0.853
Number of regions	191	191	191	191	191	191	191	191
N. observations	36290	36290	36290	36290	36290	36290	36290	36290

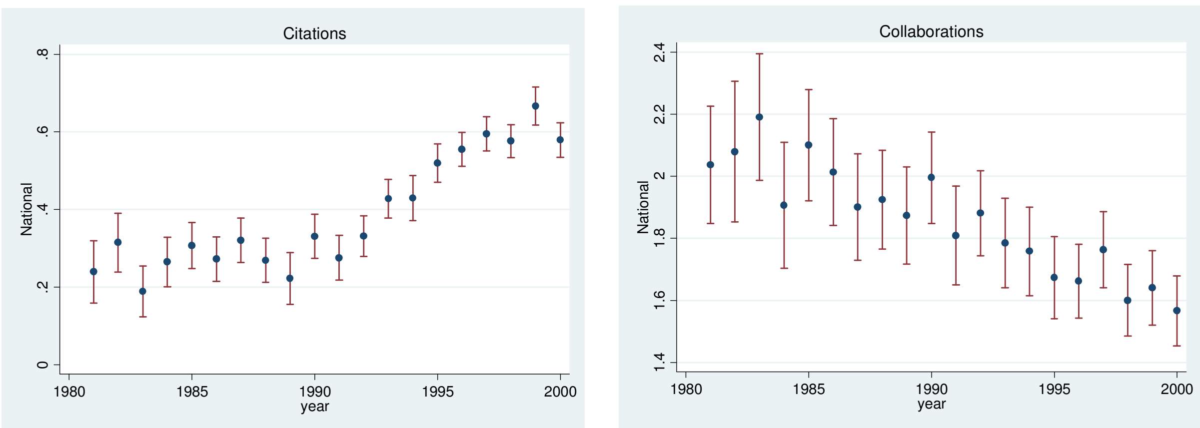
Note: ***, ** and * indicate significance at 1, 5 and 10 %, respectively.

E. Robustness check for technological specialisation

Figure E1 shows estimates of equation [5] controlling for technological specialisation. In particular an Herfindhal absolute index of internal specialisation and thirty Balassa indexes of relative specialisation are added as independent variables in equation [5].



a) evolution over time of the distance effect



b) evolution over time of the national border effect

Figure E1. Coefficient estimates and 95 % confidence intervals of the distance and national border effect - Equation [5] with the technological specialisation indexes

F. Robustness check for different time lags

For patent citations, Figures F1 and F2 show the results for distance and national border effects obtained from three different PPML estimates of equation [4] (Figure F1) and three different PPML estimates of equation [5] (Figure F2). [14] To do these estimations we created three different samples on the basis of the temporal lag between the priority years of the cited and citing patents. The temporal lags used are: 0-2 years (lag_0_2); 3-5 years (lag_3_5); 6-9 years (lag_6_9).

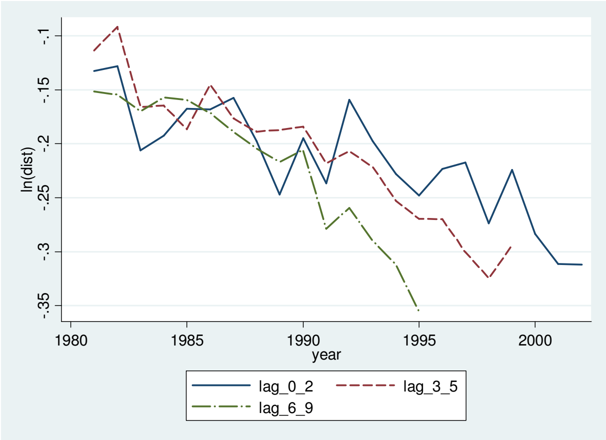


Figure F1. Citations and lag: evolution over time of the distance effect

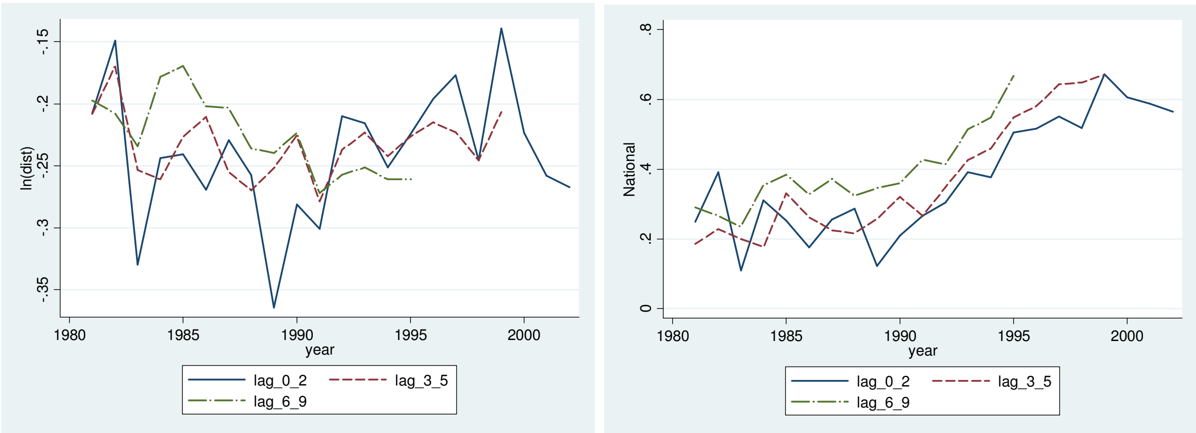


Figure F2. Citations and lag: evolution over time of the distance and national border effect

[14] All the coefficients for the distance and national border effect are significant at 1%. The results for the other variables, not shown here, are very similar to those obtained with the dataset with a time lag of 4 years.

G. List of the European regions

Table G1. List of Nuts 2 regions (Eurostat) used in our estimates

Country	Nuts 2	Name	Country	Nuts 2	Name
Austria	AT11	Burgenland (A)	Greece	GR30	Attiki
	AT12	Niederösterreich	Hungary	HU10	Közép-Magyarország
	AT13	Wien		HU21	Közép-Dunántúl
	AT21	Kärnten		HU23	Dél-Dunántúl
	AT22	Steiermark		HU32	Észak-Alföld
	AT31	Oberösterreich	Ireland	IE01	Border, Midland and Western
	AT32	Salzburg		IE02	Southern and Eastern
	AT33	Tirol	Italy	ITC1	Piemonte
	AT34	Vorarlberg		ITC3	Liguria
Belgium	BE10	Région de Bruxelles-Capitale		ITC4	Lombardia
	BE21	Prov. Antwerpen		ITD1	Provincia Autonoma Bolzano
	BE22	Prov. Limburg (B)		ITD3	Veneto
	BE23	Prov. Oost-Vlaanderen		ITD4	Friuli-Venezia Giulia
	BE24	Prov. Vlaams-Brabant		ITD5	Emilia-Romagna
	BE25	Prov. West-Vlaanderen		ITE1	Toscana
	BE31	Prov. Brabant Wallon		ITE2	Umbria
	BE32	Prov. Hainaut		ITE3	Marche
	BE33	Prov. Liège		ITE4	Lazio
	BE34	Prov. Luxembourg (B)		ITF1	Abruzzo
	BE35	Prov. Namur		ITF3	Campania
Bulgaria	BG41	Yugozapaden		ITF4	Puglia
Czech Republic	CZ01	Praha		ITG1	Sicilia

(continued)

Table G1 (continued)

Country	Nuts 2	Name	Country	Nuts 2	Name
Germany	DE11	Stuttgart	Italy	ITG2	Sardegna
	DE12	Karlsruhe	Luxembourg	LU00	Luxembourg (Grand-Duché)
	DE13	Freiburg	Nederland	NL11	Groningen
	DE14	Tübingen		NL12	Friesland (NL)
	DE21	Oberbayern		NL13	Drenthe
	DE22	Niederbayern		NL21	Overijssel
	DE23	Oberpfalz		NL22	Gelderland
	DE24	Oberfranken		NL31	Utrecht
	DE25	Mittelfranken		NL32	Noord-Holland
	DE26	Unterfranken		NL33	Zuid-Holland
	DE27	Schwaben		NL34	Zeeland
	DE30	Berlin		NL41	Noord-Brabant
	DE42	Brandenburg - Südwest		NL42	Limburg (NL)
	DE50	Bremen	Poland	PL12	Mazowieckie
	DE60	Hamburg		PL22	Slaskie
	DE71	Darmstadt	Portugal	PT17	Lisboa
	DE72	Gießen	Sweden	SE11	Stockholm
	DE73	Kassel		SE12	Östra Mellansverige
	DE80	Mecklenburg-Vorpommern		SE21	Småland med öarna
	DE91	Braunschweig		SE22	Sydsverige
	DE92	Hannover		SE23	Västsverige
	DE93	Lüneburg		SE31	Norra Mellansverige
	DE94	Weser-Ems		SE32	Mellersta Norrland
	DEA1	Düsseldorf		SE33	Övre Norrland

(continued)

Table G1 (continued)

Country	Nuts 2	Name	Country	Nuts 2	Name
Germany	DEA2	Köln	United Kingdom	UKC1	Tees Valley and Durham
	DEA3	Münster		UKC2	Northumberland and Tyne and Wear
	DEA4	Detmold		UKD1	Cumbria
	DEA5	Arnsberg		UKD2	Cheshire
	DEB1	Koblenz		UKD3	Greater Manchester
	DEB2	Trier		UKD4	Lancashire
	DEB3	Rheinhessen-Pfalz		UKD5	Merseyside
	DEC0	Saarland		UKE1	East Yorkshire and Northern Lincolnshire
	DED1	Chemnitz		UKE2	North Yorkshire
	DED2	Dresden		UKE3	South Yorkshire
	DED3	Leipzig		UKE4	West Yorkshire
	DEE0	Sachsen-Anhalt		UKF1	Derbyshire and Nottinghamshire
	DEF0	Schleswig-Holstein		UKF2	Leicestershire, Rutland and Northamptonshire
	DEG0	Thüringen		UKF3	Lincolnshire
Denmark	DK01	Hovedstaden		UKG1	Herefordshire, Worcestershire and Warwickshire
	DK02	Sjælland		UKG2	Shropshire and Staffordshire
	DK03	Syddanmark		UKG3	West Midlands
	DK04	Midtjylland		UKH1	East Anglia
	DK05	Nordjylland		UKH2	Bedfordshire and Hertfordshire
Spain	ES21	País Vasco		UKH3	Essex
	ES30	Comunidad de Madrid		UKI1	Inner London
	ES51	Cataluña		UKI2	Outer London
	ES52	Comunidad Valenciana		UKJ1	Berkshire, Buckinghamshire and Oxfordshire
	ES61	Andalucía		UKJ2	Surrey, East and West Sussex

(continued)

Table G1 (continued)

Country	Nuts 2	Name	Country	Nuts 2	Name
Finland	FI13	Itä-Suomi	United Kingdom	UKJ3	Hampshire and Isle of Wight
	FI18	Etelä-Suomi		UKJ4	Kent
	FI19	Länsi-Suomi		UKK1	Gloucestershire, Wiltshire and Bristol/Bath area
	FI1A	Pohjois-Suomi		UKK2	Dorset and Somerset
France	FR10	Île de France		UKK3	Cornwall and Isles of Scilly
	FR21	Champagne-Ardenne		UKK4	Devon
	FR22	Picardie		UKL1	West Wales and The Valleys
	FR23	Haute-Normandie		UKL2	East Wales
	FR24	Centre		UKM2	Eastern Scotland
	FR25	Basse-Normandie		UKM3	South Western Scotland
	FR26	Bourgogne		UKM5	North Eastern Scotland
	FR30	Nord - Pas-de-Calais		UKN0	Northern Ireland
	FR41	Lorraine	Switzerland	CH01	Lake Geneva region
	FR42	Alsace		CH02	Espace Mittelland
	FR43	Franche-Comté		CH03	Northwestern Switzerland
	FR51	Pays de la Loire		CH04	Zurich
	FR52	Bretagne		CH05	Eastern Switzerland
	FR53	Poitou-Charentes		CH06	Central Switzerland
	FR61	Aquitaine		CH07	Ticino
	FR62	Midi-Pyrénées	Norway	NO01	Oslo og Akershus
	FR63	Limousin		NO02	Hedmark og Oppland
	FR71	Rhône-Alpes		NO03	Sør-Østlandet
	FR72	Auvergne		NO04	Agder og Rogaland
	FR81	Languedoc-Roussillon		NO05	Vestlandet
	FR82	Provence-Alpes-Côte d'Azur		NO06	Trøndelag
				NO07	Nord-Norge