

# **The Internationalization of Inventive Activity: A Gravity Model Using Patent Data**

**Lucio Picci<sup>1</sup>**

First draft. Reports preliminary results. 15 September 2008

## **Abstract**

This paper discusses the extent and the determinants of the internationalization of European inventive activity, between 1990 and 2004, using an innovative method to treat the information contained in the European Patent Office's Patstat database.

The observed level of internationalization of inventive activities, while being rather low, has steadily increased over time. The amount of collaboration between actors residing in different countries is assessed by means of a "gravity model", as it is familiar in the literature in international trade. The amount of bilateral collaboration is positively affected by the presence of a common language and of a common border, and by the common participation in the European Union. Participation in the Euro Zone is also found to have an effect.

International collaboration is negatively affected by distance, with estimated elasticities that are significantly smaller than the ones that characterize international trade. Contrary to the rumors about the "death of distance", this effect has become stronger in recent years.

---

<sup>1</sup> European Commission - DG JRC, Institute for Prospective Technological Studies, C/ Inca Garcilaso, s/n - E-41092 Sevilla, and Department of Economics, University of Bologna, Strada Maggiore 45, 40125 Bologna. E-mail: [lucio.picci@ec.europa.eu](mailto:lucio.picci@ec.europa.eu). I thank Marc Bogdanowicz and Andrea De Panizza for useful discussions on the topics discussed in this document. The opinions expressed in this paper do not necessarily coincide with those of the European Commission.

## 1. Introduction

Almost all contemporary products are the result of some form of international collaboration and trade. A cell phone is a bundle of many components that, almost invariably, are produced in different countries. Raw materials are traded to such an extent, that anything containing plastic or metal components is also very likely to be the result of international trade. As for services, the current lore on international outsourcing of back-office routines indicates that though internationalization may not be preponderant, it is certainly increasing.

Innovative activities, on the other hand, are largely the result of efforts which take place in single countries. Almost twenty years ago, Patel and Pavitt (1991), considering the patenting activities of a sample of big US firms, observed that "in most cases, the(ir) technological activities [...] are concentrated in their home country", and concluded that "despite being a critical resource in the global competition and performance of both companies and countries [...] the production of technology remains far from globalised." Certainly, since then globalization has also made inroads into this domain, thanks to robust increases in international collaboration of various types, to the off-shoring of many R&D labs, and to advances in the protection and trade of intellectual property rights. However, as it will be seen, it is still true that innovative activities are predominantly national in scope.

This paper aims to identify the determinants of the intensity of international collaboration in inventive activities, and what they tell us about why the degree of internationalization of innovative activities is still relatively low. To address these issues, several researchers have exploited in various ways the information contained in patent data (e.g. Patel and Pavitt, 1991; Patel and Vega, 1991, and Bas and Serra, 2002). The research presented here also uses patents data. Unlike in most previous studies, however, the analysis is not carried out on the patent portfolio of (multinational) firms. Instead than to firms, patents are attributed to countries, by exploiting the fact that patents data provide separate information on the nationality of the inventors and of the applicants, where the former are individuals, and the latter may be either individuals from organized entities of various types. If a given patent involves only people and organizations residing in the same country, it is defined as "national". If there is at least one inventor or one applicant resident in a different country from the others, then that patent is "international". Moreover, the presence of an important functional distinction between inventors and applicants, allows the construction of different measures of internationalization, and the contrast between these can lead to interesting interpretations of the results.

Patent data, as a measure of inventive output, have virtues and shortcomings (Smith, 2005, and Griliches, 1990). Many innovations, particularly of production processes, do not result in any patent applications, and firms often prefer to protect their inventions by keeping them secret, rather than by asking for the protection afforded by patents. However, this limitation of patent statistics is less important when the focus is on international innovation, because the propensity to patent is bound to be higher, given that trade secrets are more difficult to keep in situations where the innovators reside in several countries and may belong to distinct organizations.

While not all innovations are patented, the opposite is also true - that is, not all patented inventions produce innovations. Patents may have very different values, and for each superstar patent, which introduces a very relevant and successful product or process, there are countless others which have limited or no applications. Possibly a more serious problem is the fact that patenting activity is increasingly just one of many strategies that oligopolistic firms have at their disposal. "Defensive patenting", in particular, has become a way of accumulating a sizeable patent portfolio to be used as a bargaining chip. Indeed, at times the collections of patents owned are better summarized as ammunition for the lawyers, than blueprints for the engineers. While the worst excesses of such "patent inflation" are probably confined to the United States, the issue is obviously a very serious one (see Jaffe and Lerner, 2004 for an assessment and a critique; see van Zeebroek et al, 2008 for a quantitative assessment of the problem at the European Patent Office (from now, EPO).

One way of putting smaller weightings on those patents that are judged to be of low quality is to only consider the applications presented at particularly important patent offices. One possibility that has become fairly popular is to consider "triadic patents", meaning all patents (or patent applications) filed at least at the EPO, the United States Patent and Trademark Office and the Japan Patent Office (see Dernis and Khan, 2004). Since a triple filing is quite expensive, it may be expected that applicants choose to incur the related costs only when they believe that their invention deserves it.<sup>2</sup> However, this approach does not help

---

<sup>2</sup> According to this metric, a patent receives a weighting of one if it has been filed to these three patent offices, and zero otherwise. Triadic patents have a second rationale, that of controlling for the "home bias effect" – the higher propensity to use one's national patent office, compared to patent offices abroad. This method also has the shortcoming of not being applicable to the most recent filings, considering that, after a first filing, the applicant has 12 months of "priority time" to approach another patent office for a subsequent filing, so that the delay with respect to the earliest priority year (the year in which the application was filed for the first time, irrespective of where) may be considerable.

filter out strategic patenting activities, and in fact may achieve the opposite, since strategic patenting is more likely to affect the patenting offices covering the widest markets.

As a consequence, those approaches that filter out low quality patents often also overlook a lot of inventive activity. In particular, low quality patents, even when they have little or no practical value, indicate the presence of some kind of an inventive effort. Low quality patented inventions, even when they do not cause genuine advancement of a technological frontier, may witness a preparedness to get closer to it. The presence of low quality patents may nonetheless be an indication of the presence of a significant absorptive capability (see Cohen and Levinthal, 1989 on the double nature of R&D – part innovation, part learning). For this reason, a methodology that computes patent statistics inclusively, rather than selectively<sup>3</sup> has been adopted.

A gravity model has been chosen to study the determinants of the intensity of collaboration between pairs of countries. The gravity model is a popular tool for explaining bilateral trade flows, where it has been given several theoretical rationalizations, starting with Anderson (1979). It describes trade between two countries as increasing in their economic size, and as decreasing in their distance. It has had a remarkable degree of success, to the point where it is considered "one of the more successful empirical models in economics" (Frankel and Rose, 2002). Over the years, various incarnations of the model have estimated the effect of possible determinants of bilateral trade. Disdier and Head (2008) performed a meta-analysis on 103 papers, and reported a mean elasticity of bilateral trade with respect to distance of 0.9, "indicating that, on average, bilateral trade is nearly inversely proportionate to distance". They also show that distance effects "decreased slightly between 1870 and 1950 and then began to rise". Moreover, other variables have generally been found to positively determine bilateral trade, such as the presence of common borders, a common language, a common market, and a common currency area (for the latter, see Rose, 2000, and Baldwin, 2006). Distance effects have also been found in the trade of services (Ceglowski, 2006, and Kimura and Lee, 2006), trade through the Internet (Berthelon and Freund, 2008), knowledge flows through patents' citations (Peri, 2005), immigration flows (for a recent example, see Lewer and Van den Berg, 2007), and in fields outside economics, such as international relations and biology.

---

<sup>3</sup> True, most low-quality patents are likely not to be characterized by any international collaboration. However, not considering them leads to an overestimate of the degree of internationalization of inventive activity, and may also cause a sample selection bias when carrying out statistical inference.

The approach followed in this research appears to have as the only one antecedent in the literature a paper by Guellec and van Pottelsberghe de la Potterie (2001). However, apart from the obvious fact that with respect to their work, one more decade's worth of data is available for analysis, this research differs in many respects from that of these two authors. One dimension in which this paper innovates with respect to the previous literature is in its treatment of the information contained in the Patstat database (European Patent Office, 2008). Here, Patstat is used to develop patent statistics that, as it will be discussed, reflect the inventive prowess of the 27 countries in the European Union. However, to do so effectively, many practical issues have to be addressed, as the next Section explains. In Section 3, three alternative measures of internationalization are discussed. In Section 4, the empirical model is illustrated and the results that it produces discussed.

## **2. The data**

The source of the data is the Patstat database, based on the EPO's "master bibliographic database DocDB". It covers data from more than 80 countries and contains a wealth of information on individual filings, including the identity and country of residence of applicants and inventors, citations, priority claims, etc. (European Patent Office, 2008b). An important and innovative characteristic of Patstat is that it allows us to track patent applications that claim the right to priority (that is, to distinguish between multiple applications for the same inventions in several patent offices).

Patent applications, not granted patents, are considered. Focusing on applications instead of granted patents, besides being a common practice in the literature, has the important advantage of allowing an analysis of more recent data, considering that several years typically elapse between the filing and the granting of a patent. Henceforth, for the sake of brevity, 'patents' will, in fact, refer to 'patent applications'. Moreover, successive filings of the same invention are discarded, so as to consider only "priority" applications. Excluding all applications that follow from a prior application allows to assess only those filings that reflect true inventive activity that is centered in the geographic area of interest.

In order to assign patents to countries, two alternative criteria may be chosen: Either according to the nationality of the applicant(s), or of the inventor(s). The former defines the "applicant criterion", and the latter the "inventor criterion". Quite often, an application has more than one inventor or applicant, and at times they come from different countries. In this

case, assignments of patents to countries are carried out resorting to fractional counts. The methods is standard in the computation of patent statistics, but here it is worth to describe it with a simple example, because it reverberates to the definition of measures of internationalization<sup>4</sup>.

Assume that there are only three countries, United States (US), France (FR), and Germany (DE), that in a given year have produced a total of  $N=3$  patents. Column I in Table 1 declares the nationality of the different inventors and applicants that contributed to these three inventions.

### Table 1 about here

Column II of Table 1 reports, for each of the 3 patents,  $Inv_{i,p}$ , the fraction attributed to each country  $i$  according to the inventor criterion. It must hold that, for each patent application, the sum of all the country's contribution according to the inventor criterion has to be equal to 1: for each patent,  $Inv_{US,p} + Inv_{DE,p} + Inv_{FR,p} = 1$ , where the first subscript indicates the country, and the second the patent. These sums are indicated in Column IV. The contribution of each country, according to the inventor criterion, is given by the sum on the contribution of that country to each individual patent. It is reported in the last but one line of Table 1, and is equal to 1 for the US, 0.75 for Germany, and 1.25 for France.

Analogous computations apply to the contribution of each country to each one of the three patents according to the applicant criterion, reported in Column III. The last row reports the fractional number of patents for each country. According to the applicant criterion, the contribution of the three countries to the three patents is equal to 2 for the United States, and one half each for Germany and for France. In a more generic case, with countries indexed by  $i = 1, \dots, N$  and patents by  $p = 1, \dots, P$ , the assignment of patents to country  $I$  may be expressed as follows (while omitting a time index for simplicity):

$$(1) \quad Inv_i = \sum_{p=1}^P Inv_{ip}$$

---

<sup>4</sup> In most European cases, adopting one criterion or the other does not lead to very different results (see Picci, 2008b). Unfortunately, quite often Patstat does not report the country of inventor or applicant. Picci (2008a) contains an explanation of the procedure adopted to solve most of these occurrences, based on the consideration of successive transfers of priority filings.

where  $Inv_{ip}$  is the fraction of inventors from country  $i$  within the total number of inventors contributing to invention  $p$ . In Table 1, Column II reports these quantities for our example. Similarly, the applicant criterion is defined as follows:

$$(1') \quad App_i = \sum_{p=1}^P App_{ip}$$

where  $App_{ip}$  is the fraction of applicants from country  $i$  within the total number of applicants contributing to invention  $p$ .

In most studies on patents, information on a single patent offices are used. Given our focus on the internationalization of patents, this approach would not be viable, because the presence of the "home bias effect" would potentially bias any study bases on a single, even if very important, national patent office. For this reason, applications filed in any one of the patents office of a member state of the European Union (in its current configuration of 27 States<sup>5</sup>), or at the EPO, are considered. Guellec and van Pottelsberghe de la Potterie (2001) focus their attention to filings to the European Patent Office only, a strategy that rests on its continental scope. However, such a choice also is questionable, for two reasons. First, many European patents never reach the EPO. This can be seen, for a single year, in Table 2, providing a comparison of patent counts for EU27, and for a few selected countries, using two alternative methods.

### **Table 2 about here**

The first two lines report data published by OECD (2007) and by Eurostat (2008), both looking at the number of filings at the European Patent office, regardless of whether these filings have a priority right, or are successive filings of inventions originally filed elsewhere. The third line reports patent counts using the methodology adopted here, that is, by looking at all priority filings presented at any European national patent office, or at the EPO, both using the inventor and the applicant criterion. There are important differences: For example, the share of German patents over the total is equal to 47% for Oecd and for Eurostat, and to 42% with the present approach. The differences that are observed are driven by two effects: on the one hand, considering the filings at many patent offices leads to higher numbers than the ones of OECD and Eurostat. On the other hand, an exclusive focus on priority filings has the

---

<sup>5</sup> Cyprus and Malta do not present any patenting activity and as such are excluded from the analysis.

opposite effect. The first of these two effects typically dominates for European countries, and the second one for countries outside of Europe, whose inventors and applicants more often first file at their national patent office.

The second reason why it is wrong to focus only on filings to the EPO, at least when the objective is to carry out international comparisons, derives from the fact that the decision to file to the EPO is influenced by factors that are country-dependent. In particular, agents from countries that are smaller, whose national patent office has higher filing fees, and that have been member of the EPO for a longer term, have a higher propensity to use the EPO (De Rassenfosse and van Pottelsberghe de la Potterie, 2007)<sup>6</sup>.

### **3. Measures of internationalization**

Three measures of internationalization are computed using the available information on the country of residence of inventors and applicants. They are illustrated using the same example that was used to demonstrate how to attribute patents to countries. Internationalization may be detected in more than one way. First, the extent to which a patent application is produced by inventors, or by applicants, coming from distinct countries, may be measured. In Table 1, for example, all three patents involve a collaboration among inventors of at least two countries. This is defined to be an instance of "inventor internationalization". Moreover, the first two patents, but not the third one, have applicants from different countries. This case is defined to be of "applicant internationalization". Please note that the presence of one type of internationalization does not imply the other. For example, patent n. 3 in the example is characterized by inventor, but not by applicant, internationalization. One further type of internationalization occurs when at least one inventor and one applicant are residents of different countries. This is true for all three patents. This is defined to be an instance of "inventor-applicant internationalization".

Whenever a patent presents either "inventor" or "applicant" internationalization (or both), it follows that it also has a degree of "inventor-applicant internationalization" – if, say, at least two inventors come from different countries (inventor internationalization), then it is also true that there will be at least one inventor and an applicant that are from different

---

<sup>6</sup> Obviously, using data from the EPO only (or from one or few national patent offices) was also a matter of convenience and, until some time ago, almost unavoidable. The point is that now, with the maturity of databases such as Patstat, researchers can do better.

countries (inventor-applicant internationalization). The opposite is not true. There could be inventor-applicant internationalization without either inventor nor applicant internationalization – for example, when a patent has a single inventor from one country, and a single applicant from another country.

The first of these measures, indicating inventor internationalization, is considered first. For each patent, the strength of the relation between country  $i$  and country  $j$  inventor is expressed as the product of the attribution of that patent to the two countries:

$$(2) \quad Inv_{ijp} = Inv_{ip} \cdot Inv_{jp}$$

The aggregate strength of the relation between the inventors of two countries is defined as the sum of the above, over all patents:

$$3) \quad Inv_{ij} = \sum_{p=1}^P Inv_{ijp}$$

Using (1), it is easy to see that

$$(4) \quad \sum_{j=1}^N Inv_{ij} = Inv_i$$

Below are reported the values for all the combinations of the three patents of Table 1 where for clarity, instead then the indexes  $i$  and  $j$ , the acronyms of the countries are employed.

$$Inv_{US,US} = 0.5 \cdot 0.5 + 0 \cdot 0 + 0.5 \cdot 0.5 = 0.5$$

$$Inv_{US,DE} = 0.25 \cdot 0.5 + 0 \cdot 0 + 0 \cdot 0 = 0.125$$

$$Inv_{US,FR} = 0.25 \cdot 0.5 + 0 \cdot 0 + 0.5 \cdot 0.5 = 0.375$$

As predicted by (4),

$$Inv_{US,US} + Inv_{US,DE} + Inv_{US,FR} = 0.5 + 0.125 + 0.375 = 1 = Inv_{US}$$

Analogous formulas hold for the applicant criterion:

$$(2') \quad App_{ijp} = App_{ip} \cdot App_{jp}$$

$$(3') \quad App_{ij} = \sum_{p=1}^P App_{ijp}$$

$$(4') \quad \sum_{j=1}^N App_{ij} = App_i$$

Equation (4) and (4') indicate that the measures in (3) and (3') effectively decompose the attribution of patents to countries. Note that  $Inv_{ij} = Inv_{ji}$  and  $App_{ij} = App_{ji}$  (the order of the countries is irrelevant).

To construct a measure for "inventor-applicant internationalization", the strength of collaboration between inventors in country  $i$  and applicants in country  $j$ , for a single patent  $p$ , is defined to be:

$$(5) \quad Invapp_{ijp} = Inv_{ijp} \cdot App_{ijp}$$

Adding over patents provides a measure of the strength of the overall collaboration between country  $i$  inventors and country  $j$  applicants:

$$(6) \quad Invapp_{ij} = \sum_{p=1}^P Inv_{ijp} \cdot App_{ijp}$$

These measure aggregates to the patent attributed to a country either according to the inventor or to the applicant criterion:

$$(7) \quad \sum_{j=1}^N Invapp_{ij} = Inv_i$$

$$(7') \quad \sum_{i=1}^N Invapp_{ij} = App_j$$

To exemplify, the measure of "inventor-applicant internationalization" for the United States and for Germany is computed below, using the example of Table 1. The acronyms of the countries are used instead of the indexes  $i$  and  $j$ . The pairwise absolute measures of inventor-applicant internationalization (eq. 6), for the United States, are:

$$InvApp_{US,US} = 0.5 \cdot 0.5 + 0 \cdot 0.5 + 0.5 \cdot 1 = 0.75$$

$$InvApp_{US,DE} = 0.5 \cdot 0.5 + 0 \cdot 0 + 0.5 \cdot 0 = 0.25$$

$$InvApp_{US,FR} = 0.5 \cdot 0 + 0 \cdot 0.5 + 0.5 \cdot 0 = 0$$

Note that  $InvApp_{ij}$  is generally different from  $InvApp_{ji}$ , and also that (7) and (7') indicate that the measure of inventor-applicant innovation allows to split the fractional counts of patents. Here is one case taken from the example of Table 1, where the first of the three term in the sum is the “national” component of US patents:

$$InvApp_{US,US} + InvApp_{US,DE} + InvApp_{US,FR} = 0.75 + 0.25 + 0 = 1 = Inv_{US}$$

The quantities defined in (3), (3') and (6) are the three measures of internationalization of innovative activities that will be modelled in the next Section.

In order to provide a first description of the degree of internationalization, the measures of internationalization are usefully expressed in relative terms, as a share of the total number of patents. It is straightforward to construct relative measures of (3) and (3'):

$$Inv_{ij|i} = Inv_{ij} / Inv_i \quad \text{and} \quad App_{ij|i} = App_{ij} / App_i$$

The conditional measures of inventor-applicant internationalization (Eq. (8) and (8')) are in fact two, depending on whether the normalization is carried out with respect to the inventors of country  $i$ , or to the applicants of country  $j$ :

$$Invapp_{ij|i} = Invapp_{ij} / Inv_i$$

$$Invapp_{ij|j} = Invapp_{ij} / App_j$$

where  $\sum_{j=1}^N Invapp_{ij|i} = 1$  and  $\sum_{i=1}^N Invapp_{ij|j} = 1$ .

Figure 1 shows, in percentage points, the different measures of internationalization. Two main conclusions emerge. First, all measures indicate an increase of internationalization in inventive activity in Europe over the fifteen years under consideration. Secondly, Figure 1 indicates the presence of a vast difference among the four alternative metrics, with measures of inventor-applicant internationalization being well above the others.

Applicant internationalization is very low, always below 1%, indicating that only in a very small minority of cases applicants from different countries jointly file an application. This measure, in fact, is likely to be a lower bound for the degree of international collaboration

among applicants, that may decide to form jointly owned firms in only one country, and whose activity that would not be observable under the measure of applicant internationalization. In these cases, however, it is possible that researchers from both countries would team-up in the jointly owned entity, so that their patenting activities would show up as inventor, or inventor-applicant, internationalization when the inventor is from a country different from the one of the new entity. The measure of inventor internationalization at the end of the period is slightly above the 3% mark.

The highest degree of internationalization regards the investment-application measures, with an interesting difference between the two. The first of the two (indicated in the Figure as InvApp) refers to patents that have European inventor(s) and extra-European applicants, and AppInv is the opposite. The fact that the first is higher than the second indicates a relatively high weight of extra-European applicants in European inventive activity, higher than the weight of European applicant elsewhere. The typical case accounted for by the InvApp measure, is the one of the extra-European multinational that owns a lab in Europe, and that files its patents through the headquarter located in the home country.

Our measures of relative internationalization have similarities with those of Guellec and van Pottelsberghe de la Potterie (2001), who adopt three measures that they call SHAI, SHIA, and SHII. The first one is similar to our  $Invapp_{ij|i} = Invapp_{ij} / Inv_i$ , the second to  $Invapp_{ij|j} = Invapp_{ij} / App_j$ , and the third to  $Inv_{ij|i} = Inv_{ij} / Inv_i$ . Our  $App_{ij|i} = App_{ij} / App_i$  has no analogue in their paper. There are however several differences in the way that the measure are constructed, the main one perhaps being that here fractional counts of patents are considered, so as to count as "more international" those patents where international collaboration is more pronounced. The measures adopted by Guellec and van Pottelsberghe de la Potterie (2001), on the other hand, do not make this distinction, and consider alike all patents where there is at least "some" international collaboration of a given type.

Having clarified the differences, it is still useful to compare the degree of internationalization that Guellec and van Pottelsberghe de la Potterie report with what emerges from the present analysis. When looking at the EPO, they find, for example for Germany and for the years 1993-1995, values of their measures SHAI, SHIA and SHII of respectively 6.3%, 4.5% and 6.6%. In the present case, the more directly comparable measure are equal, for year 1994, to 2.0%, 1.9% and 0.6%. The difference is probably attributable mainly to the fact that in those years the EPO afforded, particularly in those early years, a very biased sample of European inventive activity, as it was discussed in the previous section.

#### 4. A gravity model of international inventive activity

The basic model to be estimated is the following:

$$\ln(INT_{ijt}) = \beta_0 + \beta_1 \ln(A_{it}) + \beta_2 \ln(A_{jt}) + \beta_3 \ln(dist) + \lambda L_{ijt} + \beta_4 D_i + \beta_5 D_j + \beta_6 DT_t + \varepsilon_{ijt}$$

where  $\ln$  is the natural log,  $INT_{ijt}$  is one of the three bilateral measures of internationalization of inventive activity,  $A$  is the "mass" of each country, that will be proxied with a measure of their inventive activity,  $dist$  is the distance between the capital cities of pairs of countries, computed using the great circle formula.  $L$  is a vector of other conditioning variables. It includes dummies for the presence of common borders and of a common language<sup>7</sup> (presenting no variation in time), for inclusion in the European Union and in the European Monetary Union (presenting variation in time for several of the countries considered). In  $L$  two more time-varying variables are included. One is a multiplicative term of logged distance and time, to test for the presence of time varying effects of distance. The other is an indicator of pairwise "inventive proximity". It is computed as the pairwise correlation between two vectors, each one composed, in a given year, by the number of patent applications by the country in one of six technology classes<sup>8</sup>. The two  $D$  variables are country specific fixed effects, one for each country, and the variable  $DT$  is meant to capture any time shock affecting all bilateral relations.

---

<sup>7</sup> The language dummy takes fractional values for multilingual countries. For example, it is equal to one half between Belgium and France (the presence of a small German speaking minority in Belgium is ignored), and one third for the pairs of Switzerland with Germany, France and Italy.

<sup>8</sup> These are the eight top "sections" (A through H) of the International Patent Classification (IPC) taxonomy. See <http://www.wipo.int/classifications/ipc/en/> (last visited on 9 September 2008). The correlations are computed for each year on the same dataset used for all the exercises carried out in the paper. Fractional counting is adopted to address all the cases when more than one IPC category is assigned to a patent.

The model specification is quite flexible, including all fixed effects that is possible to incorporate, short of estimating a Fixed-Effects panel model<sup>9</sup>. Flexibility is also added by pooling observations from different years, instead of summing them over periods of a few years as it sometime done in the literature. The presence of a time dummy variable controls for the presence of time shocks. The model is estimated using OLS, both including the whole set of pairwise measures of internationalization, and also limited to the pairwise relations that only include countries that are currently members of the European Union. This is meant to address the asymmetric treatment among European and non-European countries, the latter being affected by a home effect bias, following the fact that "only" European patent offices are considered<sup>10</sup>. In most cases, however, both samples provide qualitatively similar results.

One common problem in the estimation of gravity model derives from the presence of many cases when pairs of countries do not entertain any relation – be that of trade, or of inventive activity. Given that the model is expressed in logs, all these cases generate missing values, and a "truncation", of the dependent variable. To address this problem, a two-step procedure introduced by Heckman, popularly known as the Heckit estimator (Heckman 1979, and Wooldridge, 2002, pg 564) is also used. When interpreting results, in the few cases when the OLS and the Heckit estimators suggest conflicting conclusion, the Heckit estimates will be privileged.

---

<sup>9</sup> A Fixed Effect model, in its Least Square Dummy Variable representation, would simply obtain by including dummies indexed  $i$  and  $j$ , that is, one for each statistical unit of the model. In this case, the parameters would be identified uniquely by the time variation of the data. All time-invariant (such as distance, and most likely candidates for inclusion into the L vector) would be swept away by the fixed effects, so that the model would be of very little empirical interest.

<sup>10</sup> Other patent offices could be included in principle in the construction of the data, up to the whole set currently covered by Patstat. However, consider the following. First, the degree of complexity of the computations as they are is already quite high. To produce the output of this paper, starting from the output of 60 SQL queries (four for each year considered, producing a total of more than 13 million records), a set of 19 Stata ".do files" run sequentially, with a total execution time of approximately 30 hours using a small workstation. Second, the use of Patstat still requires much care, because at times the data present idiosyncrasies that require careful examination. A line had to be drawn somewhere.

Table 3 shows the estimates of the gravity model on the  $Inv_{ij}$  measure of internationalization<sup>11</sup>. With the exception of the OLS estimate of the limited sample, a significant negative effect of distance is found, with elasticities that, in the Heckit case, are equal to 0.31 for the whole sample, and to 0.24 for the reduced one. These are important effects, although much smaller than the ones that are usually found for international trade. In all cases the coefficient of the interaction term of distance with time are negative and significantly estimated, indicating that the effect of distance has increased in time.

### Table 3 about here

A common border and a common language have strong effects in all specifications, as does participation in the European Union. Interestingly, participation in the Euro Zone also seems to have a positive effect, at least in the estimates that only include European countries. In this case, as in the following ones, our index of technological similarity positively affects international collaboration.

The results of our estimates on the  $App_{ij}$  measure of internationalization, shown in Table 4, indicate an ambiguous effect of distance, significantly negative in the Heckit estimates on the wider group of countries, and positive, although not significant, in the Heckit estimate on the EU25 estimate. The cross product of distance with time is also negative and precisely estimated in the EU25 case. A lower role for distance is also witnessed by the fact that the estimated coefficient on the border dummy, while negative and significant, is of smaller magnitude than before. In interpreting this result, please note that  $App_{ij}$  is the lowest of the measure of internationalization (see Figure 1), being defined by rare collaboration between applicants of different countries. This also may explain why in this case the differences between the OLS and the Heckit estimates are more important than in the other cases: the lower is the ratio of non-zero observations, the more likely it is that OLS estimates will be prone to a sample selection bias.

---

<sup>11</sup> In this and in all following cases, the first step of the Heckit estimator, a Probit describing the selection into the sample, used the same list of dependent variables of the gravity model, minus the measures of inventive activity, plus GDP based on purchasing power parity, expressed at current dollars, taken from IMF's World Economic Outlook Database (<http://www.imf.org/external/ns/cs.aspx?id=28>, last accessed on 12 Sept 2008).

#### **Table 4 about here**

The effect of participation in the European Union is positive and marginally insignificant in the Heckit estimate of the EU25 group of countries. Participation in the Euro Zone, once again, is found to positively affect internationalization.

The estimates on the  $InvApp_{ij}$  measure of internationalization are in Table 5. Distance is again found to negatively affect international collaboration, as is having a common border and sharing a common language. Participation in the European Union also has a significantly positive effect on this type of international collaboration. Of particular interest is to observe that the estimate of the coefficient of the Euro Zone variable is negative and, in the Heckit estimates, significant in the full sample case, and not far from being significant in the other.

#### **Table 5 about here**

Table 5 shows the estimates of the gravity model on the  $InvApp_{ij}$  measure of internationalization. A tentative explanation for this finding is that the multinational firms of countries within the Euro Zone find a less urgent need to set up laboratories in other countries that also belong to the currency union, and as such can choose the location for their labs privileging other criteria of choice. In other words, it is possible to imagine that beyond a certain level of functional or institutional integration, one important reason (political, etc) to set up a R&D lab abroad vanishes, so that multinational firms can choose their location with more attention towards other countries.

It is useful to consider this issue within the broader debate on the presence of a so-called “Rose effect”, that is, of a significant and positive impact of currency unions on international trade (Rose, 2001; Baldwin, 2006). The results seem to indicate that a currency union may have a subtle influence on the particular manifestation of international interactions here considered. It provokes an increase of some type of collaboration, and possibly a decrease of others.

## 5. Discussion and conclusions

In this paper evidence has been presented on the extent and causes of internationalization of innovative activities, with a focus on the European economies, using a dataset on patent applications containing data from the EPO and for all the national patent offices of the European Union.

It emerged that the degree of internationalization of innovative activities, while steadily increasing in time, is still relatively limited. Such a lasting "lack of globalization" may come as a surprise, at least if observed under the light of the amazing intricacy of the relations that govern today's international division of labor. An understanding of this finding comes from the identification of the factors that determine the level of internationalization observed.

Sharing a common language positively affect the degree of collaboration, as does being a member of the European Union. It is not possible to disentangle the relative role of the different ingredients that together define the European Union. Obvious candidates would be the presence of a common market, the presence of a common regulatory framework, and of innovation policy instruments, both encourage and force collaboration among actors residing in different EU countries<sup>12</sup>.

From the estimation results, it emerges inconclusive evidence that being part of the Euro Zone also affects the extent of internationalization. In the case of inventor and applicant internationalization, the influence is estimated to be positive, although not very strong. This evidence adds to a literature that over the last years has tried to estimate the effect of currency unions on international trade, and it suggests that a currency unions may have a positive effect on international interactions that go beyond trade. On the inventor-applicant measure of internationalization, however, there is some evidence on the presence of a *negative* effect of the Euro Zone, an apparent puzzle that received a tentative explanation.

Distance unambiguously negatively affect internationalization, with estimated elasticities as high as 0.4, with an effect that tends to decrease in time. As other recent studies do in different concepts, from this researches also it emerges that the rumors on the death of distance have been greatly exaggerate. On why is this so, a possible explanation is the following. It could well be that what really matters is not physical distance, but what may be seen as a measure of a loosely defined "cultural" distance. Doing research together implies the

---

<sup>12</sup> An example of a European policy instrument that forces the collaboration of innovators from different countries of the European Union are the framework programmes, now in their seventh round.

functioning of countless tacit agreements, many of them unenforceable, certainly comprising research tasks proper, but encompassing many accompanying activities – building of infrastructure, sunk investments of various type, down to the protection of the produced intellectual asset.

Arguably, the further away are two cultures, the more difficult it is for persons to make tacit agreements work, the more difficult it is to collaborate. The physical distance variable could then proxy for such an omitted cultural distance variable. The idea could explain the increasing role of distance that emerges from the estimates of the gravity model. In a situation where many impediments to international collaboration have decreased over the last decades, arguably cultural distances have remained roughly constant, so that their *relative* importance is seen to have grown over time.

## References

- Anderson, James (1979) The Theoretical Foundation for the Gravity Equation, *American Economic Review* 69, 106-116.
- Baldwin, Richard (2006), The Euro's Trade Effect (2006), Proceedings of June 2005 workshop on what effects is EMU having on the Euro Area and its member countries?, European Central Bank
- Berthelon, Matias and Freund, Caroline (2008) On the conservation of distance in international trade, *Journal of International Economics* 75, 310-320
- Blum, Bernardo S. and Goldfarb, Avi (2006), *Journal of International Economics*, 70, pp. 384-405
- Carlsson, Bo (2006) Internationalization of innovation systems: A survey of the literature. *Research Policy* 35, pp. 56-67
- Ceglowki, Janet (2006) Does Gravity Matter in a Service Economy?, mimeo
- Cohen, Wesley M. and Levinthal, Daniel A. (1989), Innovation and learning: The two faces of R&D, *The Economic Journal*, Vol. 99, N. 397, pp. 569-596.
- De Rassenfosse, Gaétan, and van Pottlesberghe de la Potterie, Bruno (2007) Per un pugno di dollari: A first look at the price elasticity of patents, *Oxford Review of Economic Policy*, 23(4), 588-604
- Dernis, Helene and Khan, Mosahid (2004), Triadic patent families methodology, OECD, Directorate for Science, Technology and Industry, DSTI/DOC(2004)2
- Disdier, Anne-Célia, and Head, Keith (2008), The Puzzling Persistence of the Distance Effect on Bilateral Trade, *The Review of Economic and Statistics*, 90(1): 37-48
- European Patent Office (2008a), PATSTAT database, April 2008 Edition
- European Patent Office (2008b), Data catalog for the EPO Worldwide Patent Statistical Database, April 2008 Edition
- Eurostat (2008) Patent applications to the EPO in the ICT sector 1993 to 2003, *Statistics in Focus*, 22/2008
- Frankel, Jeffrey, and Rose, Andrew (2002), An Estimate of the Effect of Common Currencies on Trade and Income Vol. 117, No. 2, Pages 437-466
- Granstrand, Ö and S. Sjölander (1992) Internationalization and diversification of Multi-technology Corporations. In Ö. Granstrand, L. Hakanson and S. Sjölander (eds) *Technology Management and International Business: Internationalization of R&D and Technology*. Chicester: John Willey and Sons.
- Griliches, Zvi. 1990. "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature*, Vol. 28, pp. 1661-1707
- Guellec, Dominique, and van Pottlesberghe de la Potterie, Bruno (2001), The internationalization of technology analysed with patent data, *Research Policy* 30, pp. 1253-1266
- Heckman, J. J. (1979) Sample Selection Bias as a Specification Error", *Econometrica* 47, 153-161.
- Jaffe, Adam B. and Lerner, Josh (2004), *Innovation and Its Discontents: How Our Broken Patent System is Endangering Innovation and Progress, and What to Do About It*, Princeton University Press, Princeton, PA.
- Kimura, Fukunari and Lee, Hyun-Hoon (2006) The Gravity Equation in International Trade in Services, mimeo
- Le Bas, Christian and Sierra, Christophe (2002) 'Location versus home advantages' in R&D activities: some further results on multinationals' location strategies, *Research Policy* 31, 589-609.

- Lewer, Joshua J, and Van den Berg, Hendrik (2007), A gravity model of immigration, *Economic Letters* 99, 164-167.
- OECD (2008), Compendium of Patent Statistics 2007, available online: <http://www.oecd.org/dataoecd/5/19/37569377.pdf>
- Patel, Pari and Pavitt, Keith (1991), Large firms in the production of the world technology: An important case of "non-globalization", *Journal of International Business Studies* 22 (1), pg. 1-21
- Patel, Pari And M. Vega (1999) Patterns of internationalisation of corporate technology: location versus home country advantages. *Research Policy* 28 (2/3), 145-155
- Peri, Giovanni (2005) Determinants of Knowledge Flows and Their Effect on Innovation, *The Review of Economics and Statistics*, 87(2): 308-322
- Picci, Lucio (2008) A methodology to develop patents statistics on the European inventive activity using the Patstat database, IPTS-JRC, mimeo
- Rose, Andrew (2000), One Money, One Market: Estimating the Effects of Common Currencies on Trade," *Economic Policy*, Vol. 15 (April), pp. 7-45
- Smith, Keith (2005), Measuring Innovation, in Fagerberg, Jan; Mowery, David C. and Nelson, Richard R. (ed.), *The Oxford Handbook of Innovation*, Oxford University Press, Oxford, U.K.
- van Zeebroeck, N; N. Stevnsborg, B. van Pottelsberghe de la Potterie, D. Guellec and E. Archontopoulos (2008) Patent inflation in Europe, *World Patent Information*, Vol 30, n. 1, pp. 43-52

## Tables

**Table 1. Fractional counts of three patents**

I	II			III			IV	
	$Inv_{US,p}$	$Inv_{DE,p}$	$Inv_{FR}$	$App_{US,p}$	$App_{DE,p}$	$App_{FR,p}$	$\sum_{i=1}^N Inv_{ip}$	$\sum_{i=1}^N App_{ip}$
Patent 1 Inv: DE, FR, US, US App: DE, US	0.5	0.25	0.25	0.5	0.5	0	1	1
Patent 2 Inv: DE, DE, FR, FR App: FR, US	0	0.5	0.5	0.5	0	0.5	1	1
Patent 3 Inv: FR, US App: US, US	0.5	0	0.5	1	0	0	1	0
$Inv_i = \sum_{p=1}^P Inv_{ip}$	<b>1</b>	<b>0.75</b>	<b>1.25</b>					
$App_i = \sum_{p=1}^P App_{ip}$				<b>2</b>	<b>0.5</b>	<b>0.5</b>		

**Table 2. Comparison with Eurostat and OECD statistics. Number of applications, year 2003**

	EU27	DE	FR	UK	IT	SE	FI	NL	ES	US	JP	CN	TW
Source:													
<b>Eurostat</b>	50785	21469	7759	5264	4269	1939	1245	3386	920	30830	20665	813	-
<b>OECD</b>	50973	21496	7807	5297	4297	1979	1253	3388	924	31124	20717	807	513
<b>Our data:</b>													
<b>Inventor</b>	76202	36018	13430	9826	1634	1767	1402	2385	2030	829	783	139	807
<b>Applicant</b>	74443	35648	13265	9289	1419	1961	1498	2318	1952	2036	991	83	826

Sources: OECD (2007), Eurostat (2008) and analysis of the Patstat database (April 2008 release) (See Section 2 of this paper for a description). Patent counts are rounded to nearest integer.

**Table 3. Regression results: InvInv**

Dep variable: $\ln(Inv_{ij})$	All countries		EU25 only	
	OLS	Heckit	OLS	Heckit
<b>ln(agginv_x)</b>	.2012 0.000	.3180 0.000	.0800 0.155	.1800 0.005
<b>ln(agginv_y)</b>	.2012 0.000	.3180 0.000	.0800 0.155	.1800 0.005
<b>ln(dist)</b>	-.2300 0.000	-.3148 0.000	-.0952 0.169	-.2419 0.000
<b>ln(dist)*time</b>	-.0053 0.090	-.0077 0.016	-.0031 0.000	-.0265 0.000
<b>border</b>	.8906 0.000	.9557 0.000	.6368 0.000	.7026 0.000
<b>language</b>	1.0424 0.000	1.1681 0.000	1.4030 0.000	1.5748 0.000
<b>EU Union</b>	.2741 0.000	.2535 0.000	.4906 0.000	.4552 0.000
<b>Euro Zone</b>	.06181 0.305	.0151 0.800	.2193 0.001	.1548 0.015
<b>Techproximity</b>	.1477 0.061	.1579 0.046	.3415 0.001	.3689 0.000
n	4276	4276	2726	2726
R <sup>2</sup>	0.6534		0.6792	
$\lambda$	.5470		.5155	

**Table 4. Regression results: AppApp**

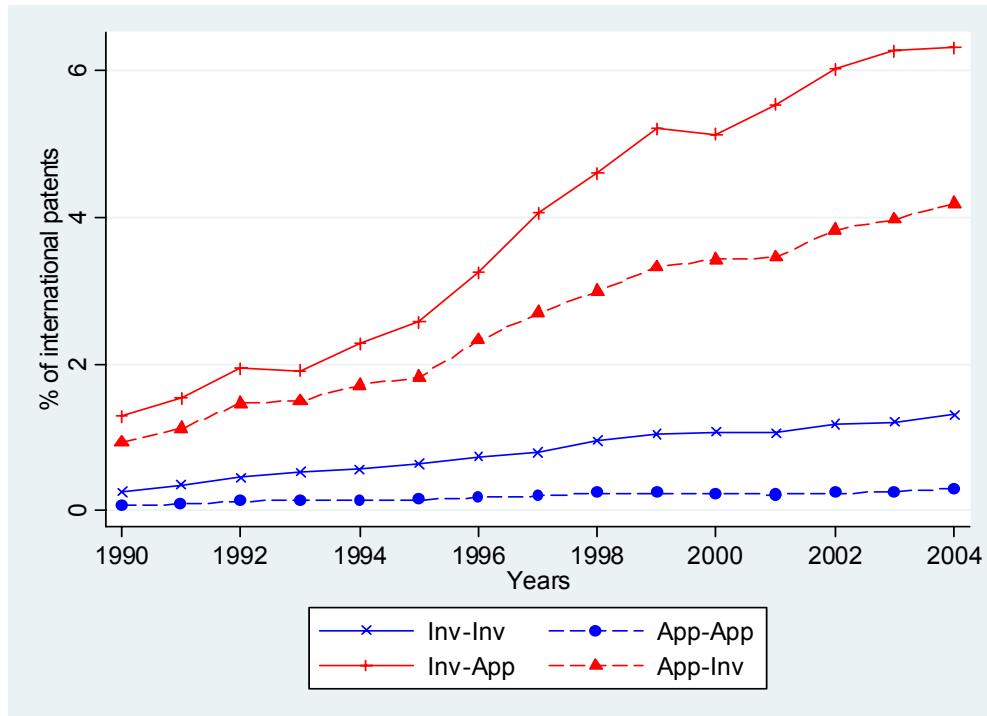
Dep variable: $\ln(App_{ij})$	All countries		EU25 only	
	OLS	Heckit	OLS	Heckit
<b>ln(aggapp_x)</b>	.1057 0.102	.2502 0.000	.0322 0.689	.1203 0.150
<b>ln(aggapp_y)</b>	.1057 0.102	.2502 0.000	.0322 0.689	.1203 0.150
<b>ln(dist)</b>	-.0123 0.822	-.2238 0.000	.3127 0.000	.0528 0.542
<b>ln(dist)*time</b>	.0048 0.231	.0047 0.223	-.0259 0.001	-.0254 0.001
<b>border</b>	.4763 0.000	.6871 0.000	.4536 0.000	.6077 0.000
<b>language</b>	.9898 0.000	1.164 0.000	1.0911 0.000	1.3676 0.000
<b>EU Union</b>	-.0129 0.868	-.0687 0.337	.1860 0.059	.1343 0.142
<b>Euro Zone</b>	.1083 0.150	.1571 0.029	.3497 0.000	.4306 0.000
<b>Techproximity</b>	.1216 0.293	.2304 0.050	.2821 0.061	.3290 0.028
n	2146	2146	1328	1328
R <sup>2</sup>	0.543		0.5805	
$\lambda$		.8229		.7297

**Table 5. Regression results: InvApp**

Dep variable: $\ln(\text{InvApp}_{ij})$	All countries		EU25 only	
	OLS	Heckit	OLS	Heckit
<b>ln(agginv_x)</b>	.0785 0.116	.2394 0.000	.0768 0.796	.2110 0.008
<b>ln(aggapp_y)</b>	.1672 0.007	.2851 0.000	.3057 0.001	.4135 0.000
<b>ln(dist)</b>	-.2594 0.000	-.4674 0.000	-.0401 0.653	-.2935 0.001
<b>ln(dist)*time</b>	-.0071 0.077	-.0074 0.064	-.0321 0.000	-.0279 0.000
<b>border</b>	.5880 0.000	.6830 0.000	.4782 0.000	.5490 0.000
<b>language</b>	1.0808 0.000	1.2212 0.000	1.2022 0.000	1.4158 0.000
<b>EU Union</b>	.0797 0.259	.0702 0.304	.5757 0.000	.5060 0.000
<b>Euro Zone</b>	-.1410 0.061	-.1740 0.019	-.0589 0.467	-.1160 0.150
<b>Techproximity</b>	.4493 0.000	.4561 0.000	.6945 0.000	.7562 0.000
n	3958	3958	2448	2448
R <sup>2</sup>	0.6157		0.6334	
$\lambda$	.7831		.6797	

## Figures

**Figure 1. International patent, EU27, national averages**



Note:

$$\text{Inv-Inv: } Inv_{ij|i} = Inv_{ij} / Inv_i$$

$$\text{App-App: } App_{ij|i} = App_{ij} / App_i$$

$$\text{Inv-App: } Invapp_{ij|i} = Invapp_{ij} / Inv_i$$

$$\text{App-Inv: } Invapp_{ij|j} = Invapp_{ij} / App_j$$

Source of the data: Analysis of the Patstat database (April 2008 release) (See Section 2 of this paper for a description).