XXIII Applied Economics Meeting. Sevilla, 10th-11th June 2010

A GRAVITY APPROACH TO CROSS-REGIONAL MOBILITY OF INVENTORS: EVIDENCE FROM EUROPE[1]

Very preliminary version
Do not quote without permission from the authors

First draft: February 2010
This draft: March 2010

Ernest Miguélez & Rosina Moreno

AQR-IREA. Department of Econometrics, Statistics and Spanish Economy. University of Barcelona, Av. Diagonal 690, 08034 Barcelona, Spain. Phone: 934021412; Fax: 934021821.
emiguelez@ub.edu, rmoreno@ub.edu

Abstract
We investigate the aggregated determinants of inventors' inter-regional mobility between European NUTS3 regions. To do so, we estimate a gravity model where counts of in- and out-flows of inventors between pairs of regions are explained by geographical distance and regional characteristics. However, economic, social, technological, and institutional distances between regions may also play a significant role. As for the econometrics is concerned, in order to accommodate our estimations to the count nature of our dependent variable and the high number of zeros in it, zero inflated negative binomial models are used. Our first results point out to the importance of geographical distance, as well as technological, social, and institutional distance, in mediating the mobility patterns of inventors across the European geography.

Key words: inventors' mobility, gravity model, zero-inflated negative binomial, European regions

JEL: C8, J61, O31, O33, R0

[1] **Acknowledgements:** Part of this work was carried out while Ernest Miguélez was visiting the 'Knowledge, Internationalization and Technology Studies' (KITeS) Research Group at Bocconi University (Milan, Italy). The use of KITeS' facilities is gratefully acknowledged. The authors would also like to share their appreciation of helpful comments received from participants to the Brown Bag Discussion Meetings, at Bocconi University (17th March, 2010), Camilla Lenzi, and Francesco Lissoni. We also acknowledge financial support from the Ministerio de Ciencia e Innovación, ECO2008-05314 and Ernest Miguélez, from the Ministerio de Educación, AP2007-00792 and the European Science Foundation, for the activity entitled 'Academic Patenting in Europe'. However, any mistake or omission remains ours.

1. Introduction

Knowledge diffusion is essential for knowledge creation and growth (Grossman and Helpman, 1991; Romer, 1986, 1990). Highly skilled personnel's mobility across firms and in space matters for the transmission of knowledge. These two claims have become nearly an aphorism in recent years in the field of innovation economics. However, the importance of mobility as a knowledge transmission mechanism goes back to Marshall's (1920) contributions. In Europe, policymakers have also embraced this conviction, and mobility of researchers, scientists and, in general, highly skilled personnel, became one of the main pillars of the creation of the European Research Area (ERA) launched by the Lisbon Agenda, back in the 2000. Thus, the European Commission (2000) put forward a number of suggestions and considerations for debate aimed at the creation of such an Area. Amongst them, "greater mobility of researchers" and "improving the attraction of Europe for researchers from the rest of the world" were pivotal (Op. Cit.).

The literature on the economics of innovation and knowledge diffusion points out at mobility of human capital as one of the main mechanism through which knowledge is transferred and spread (Döring and Schnellenbach, 2006). Mobility of professionals and skilled labour is found to be huge within the local labour market (Camagni and Capello, 2009), and largely explains the diffusion of knowledge within it and subsequent regional innovative activities (Miguélez and Moreno, 2010; Riggi and Maggioni, 2009). When this mobility also entails crossing an urban or regional border, geographical knowledge flows are at work. The present paper focuses precisely on the analysis of geographical knowledge flows as measured by cross-regional mobility of inventors across European regions. Although large part of literature has focused its attention on job-to-job mobility of inventors, their geographical mobility is also important for its implications for regional economic performance, and issues related to brain drain, brain gain, and brain circulation. Even though location changes are not as frequent as job-changes *per* se, together with inter-regional input-output linkages, FDI, and research networks, the movement of skilled workers across regions acts as an important channel through which inter-regional linkages are set up and knowledge is transferred (Fratesi and Senn, 2009). Likewise, their mobility also entails positive externalities since their new colleagues might benefit from a continued exposition to the inventor's knowledge, and such, social benefits of mobility exceed the direct returns of it (Trajtenberg and Shiff, 2008). Thus, quoting Breschi et al. (2009), "knowledge always travels along with people who master it. If those people move away from where they originally learnt, researched, and delivered their inventions, knowledge will diffuse in space. Otherwise, access to it will remain constrained in bounded locations".

All in all, however, mobility of researchers in general, and of inventors in particular, is higher than of the rest of the population[2] (European Commission, 2000). As pointed out in Storper and Scott (2009), the map of human capital is constantly reshaped by labour migration, and therefore it is important to investigate "the forces that influences the movements of people, that contribute to changes in the geographical distribution of human capital, and that hence might play a role in local economic growth" (Op. Cit. p. 148). Therefore, the identification of features and differences impeding or favouring this mobility is an important issue from a policy viewpoint in order to design measures aimed at (1) providing regions with the endowments able to attract talented individuals, and (2), reducing the differences (distances) that are hindering mobility between regions, in order to improve and increase the spread of people, and therefore knowledge, throughout the European geography.

In short, the present paper's intention is to shed some light in one aspect which has not deserved enough attention in the literature, i.e., what is driving the geographical mobility patterns of inventors across European regions. Put differently, this paper will try to answer what hinders and favours the spread of inventors' mobility, above and beyond the location of economic and innovative activities, uniformly throughout the whole continent. To step in that direction, a gravity equation framework is used, where both relational (distances) and attributional (regional features) variables are included. To the best of our knowledge, no paper has addressed this question from a territorial perspective before, and it will be the main contribution of the study.

In the globalisation era, the so-called 'end of geography' (O'Brien, 1992) or 'death of distance' (Cairncross, 1997), due to the existence of telecommunication technologies, communication highways, broadband connections, and, in general, lowering transportation costs, may induce us to think on a minored, residual importance of physical space on driving the mobility patterns of inventors across Europe. So, first of all, in the present paper our intention is to test whether or not physical distance plays a preponderant role in driving this mobility and, therefore, implicitly test the hypothesis of the 'death of distance' versus the localisation of knowledge flows –see section 3.

Further, we will suggest a couple of additional hypotheses. First, we would like to assess the importance of distance once other regional characteristics, aside from innovation, related to income level and consumption amenities, are controlled for. Whether these features are also spatially concentrated, the coefficient of distance will be biased upward if they are not taken

[2] According to the European Commission (2000), on average, the mobility of researchers is around 5% of their active population, whilst that of other professional groups is around 2%. For a group of US inventors, Breschi and Lissoni (2009) found that only 28.4% of all cross-firm inventors (9.2% of all inventors) are mobile across MSA's. On its side, Trajtenberg and Shiff (2008) find that 19.8% of software inventors from the USPTO report more than one geographical location, whilst 13.9% of Israeli inventors report more than one district of residence, and 6.8% of the inventors move in and/or out of the country (Op. Cit.).

appropriately into account. Second of all, we would like to test if other 'type of distances' between pairs of regions, which might be overlapping with physical distance *per se*, are actually driving our results. In particular, we are interested in elucidating the role played by institutional, technological, industrial, and social distance in mediating knowledge flows driven by cross-regional mobility of inventors in Europe. Summarising, we would like to test (i) whether geography plays any significant role, even when controlling for the spatial distribution of economic and innovation activities; (ii) whether regional/atributional features taken from innovation economics literature (regional inventors' productivity), from immigration economics literature (income gap between pairs of regions), and from the 'creative class-consumer city' branch of economic geography literature (amenities) influence this mobility pattern; and (iii) whether other more meaningful distance, which could overlap with physical distance, are playing any significant role, in line which what is said in part of the innovation and regional economics literature.

As for the econometrics is concerned, the count nature of our dependent variable (counts of flows of inventors) lead us to the utilisation of count data models, which might well be corrected for overdispersion and excess of zeros by applying zero-inflated negative binomial models. Our preliminary findings seem to indicate the (still) strong importance of geography in mediating knowledge flows driven by geographical mobility of inventors throughout the continent. However, technological, institutional, and social distances are also playing a significant role and, more and more, becoming as important as geography and space in explaining the transfer of knowledge across locations.

The outline of the paper is as follows: section 2 reviews some relevant previous studies about inventors' mobility and the gravity equation as a tool for the analysis of knowledge flows, tieing together dispersed but related literature. Section 3 describes the model and the hypotheses we propose, whilst also presents the data and several econometric issues. Section 4 includes some results and section 5 presents conclusions and certain limitations of our approach.

2. Background

A body of literature on inventors' mobility has sprung up in recent years. Thus, for instance, earlier studies have examined how the labour mobility of inventors acts as a key mechanism in the diffusion of knowledge (Rosenkopf and Almeida, 2003; Breschi and Lissoni, 2009; Agrawal et al., 2006), and have stressed the role of mobility insofar as it increases the hiring firm's use of a hired inventor's prior knowledge (Almeida and Kogut, 1999; Rosenkopf and Almeida, 2003; Corredoria and Rosenkopf, 2006; Song et al., 2003).

Likewise, another stream has paid more attention on the determinants of inventors' mobility. Most of this literature lay an emphasis on the effect of individual's productivity on the chances to change job or to move from the academy to the industry. However, results are not conclusive so far, since some works find a positive relationship (Zucker et al., 2002; Lenzi, 2009), whilst others a negative one (Hoisl, 2007; Trajtenberg and Shiff, 2008; Shalem and Trajtenberg, 2008). Other results found negative effects of age and experience on mobility (Crespi et al., 2007; Lenzi, 2009), while monetary incentives increase the probability to observe a move (Hoisl, 2007). Besides, network effects and regional characteristics –regional income, industrial structure- are found to positively influence mobility (Lenzi, 2009).

In the present paper we suggest a shift-focus on the study of inventors' mobility determinants. We will bring out our attention entirely on the aggregated determinants of mobility instead of the individual's characteristics. This is a significant shift-focus, because, in spite of the rareness of the phenomenon under study (Breschi et al., 2009), we are convinced about the importance of highly skilled inventors, scientists, and engineers, and their attraction and retention, for the performance of regions (Trippl, 2009). Thus, by using aggregated data, we will be able to identify which regional economic factors may influence geographical mobility of inventors.

In order to test our hypotheses already suggested in the introductory section and thoroughly discussed in the next section, we apply a gravity equation framework. This framework has been already used in few studies for the analysis of knowledge flows. The pioneering work can be considered the one by Peri (2005) estimating a sort of interaction models for a set of OECD NUTS1 regions using citation data. The main findings of his work are that geographical distance explains a substantial part in the variation of inter-regional flows of citations. However, the estimated elasticity of distance is found to be smaller than the gravity models of trade, pointing to the importance of telecommunication technologies for transferring knowledge. These findings are also found in Paci and Usai (2008), who estimate a gravity equation for patent citations flows for the European NUTS2 regions. Again, geographical distance, alongside other geographical proxies, is found to be significant in mediating knowledge flows measured by patent citations. In a similar vein, Fisher et al. (2009), Fisher and Griffith (2008), and LeSage et al. (2007) use patent citation data to examine knowledge flows across European regions but, as a novelty, their analysis is mainly focused on the presence and influence of spatial autocorrelation in spatial interaction models such as the gravity equations, and how they can be solved.

Though patent citations have been used as a measure for knowledge flows and as evidence from the localisation of knowledge from the work by Jaffe et al. (1993), this view has received numerous criticisms due to methodological concerns (Ejermo and Karlsson, 2006; Thompson and Fox-Kean, 2005) and due to the fact that it has been shown that networks of co-inventors (Breschi and Lissoni, 2009; Singh, 2005) and labour mobility (Breschi and Lissoni, 2009; Zucker et al., 1998; Almedia and Kogut, 1999) are largely responsible for these two phenomena. Thereby, several studies have used the gravity approach to study the extent of these flows measured by co-inventorshipness in patents. The idea is that the relationship which is established in a co-invention implies the transfer of a large part of knowledge from both sides. This line of thinking comes from the empirical evidence encountered as networks of research collaboration to be a mechanism for knowledge flows (Singh, 2005). Thus, co-patenting behaviour is also believed to be geographically mediated (even more than citations) due to the fact that, in general, collaboration tasks requires face-to-face interactions and frequent meetings (whilst the knowledge that a citation bring about could be easily transmitted by means of telecommunications). In this sense, several studies have appeared in recent years using a gravity equation to explain co-inventorship data at the regional level –like that of Ejermo and Karlsson (2006) for Swedish regions, Ponds et al. (2007) for Dutch regions, Maggioni and Uberti (2007, 2009) for European NUTS2 regions, or Hoekman et al. (2009) for European NUTS3 regions. Again, a common conclusion arises: the importance of geographical distance in mediating co-patenting behaviour across regions. More recently Picci (2009) and Montobbio et al. (2010) are examples of the gravity approach applied to co-autorshipness of patents at the international level. To our knowledge, nobody has used the gravity equation approach to assess the patterns of knowledge flows driven by inventors' mobility across locations.

We cannot overlook a large body of literature coming from labour and migration economics (Borjas, 2000) and economic geography (Tabuchi and Thisse, 2002). These and related studies have focused on the estimation of migration equations, where several home and host regional or country characteristics are responsible for the flows of migrants across locations (Crozet, 2004; Sanchis-Guarner and López-Bazo, 2006; Clark et al., 2007; Lewer and Van der Berg, 2008; Ortega and Peri, 2009; Blackburn et al., 2010). Once more, geographical distance, which has been taken as a proxy for migration costs, is found to be pivotal in determining migration flows across locations. Undoubtedly, our approach feeds from this literature, insofar as we are also studying migration movements of individuals across locations. However, the focus on inventors in the present inquiry put some ground between these approaches and ours.

To sum up, the reviewed literature point to two main issues. First of all, it points at the importance of mobility for transferring knowledge across firms and also across locations. Second of all, and more important for our purpose, it stresses the preponderant role of geography in mediating both knowledge flows and migration movements of labour. However, at the same time it reflects a drawback on our specific understanding about what influences cross-regional knowledge flows driven by mobility of inventors from a regional perspective. In the present study, we will try to fill in this gap.

3. Research design

The Model

The gravity equation, taken from Newton's "Law of Universal Gravitation", has been widely applied to estimate trade flows across locations, as well as other social interactions like FDI, migration movements, or knowledge flows measured by patent citations or co-patents. Our study steps in that way by using flows of inventors (those who have applied for patents) between pairs of regions. On its simplest version, the gravity equation states that the "gravitational" force between two objects is directly proportional to the "masses" of the two objects and inversely proportional to the (physical) distance between them (Burger et al., 2009). Such a model can be written as:

$$M_{ij} = \beta_0 \cdot \frac{I_i^{\beta_1} \cdot I_j^{\beta_2}}{DIST_{ij}^{\beta_3}}, \qquad (1)$$

where M_{ij} in our framework measures the number of people (counts) moving out from region i into region j. Taking logs in the former expression,

$$\ln M_{ij} = \ln \beta_0 + \beta_1 \ln I_i + \beta_2 \ln I_j - \beta_3 \ln DIST_{ij}. \qquad (2)$$

I_i and I_j stand for the "masses" of people potentially movers –inventors in our case- while $\ln \beta_0$ will be the constant term capturing the impact of all common factors across regions affecting mobility.

Geography

As listed in McCann et al. (2010, p. 362), geographical mobility patterns has undoubtedly come into a more complex drawing in the globalisation era, due to a number of reasons such as (i) migrants better informed about opportunities elsewhere; (ii) the reduction of institutional barriers, especially in Europe since the Maastricht treaty –which is particularly true for highly skilled individuals who, in turn, seem to be less sensitive to distance when they decided to move (Schwartz, 1973); (iii) because of a process of global economic integration; and (iv) due to a reduction of the real costs of travel. Once more, the 'death-of-distance-hypothesis' supporters (Cairncross, 1997; O'Brien, 1992) usually go one step beyond arguing that, due to air travels, fax, Internet, and so on, the explanatory power of classical variables affecting factor mobility have sharply declined (Clarck, 2002). All in all, therefore, the naïve expectation would suggest that physical distance within Europe is not influencing geographical mobility anymore.

Whether the former statements hold, one should observe a random pattern of mobility across regions only driven by employer-employee matches that, at the same time, are determined by the spatial distribution of production and innovation activities, and the employment opportunities derived from this distribution. Therefore, geographical distance should not play any significant role once this spatial distribution is controlled for.

Against this view, however, several authors have stressed the relative immobility of labour (Malmber, 2003), and especially to what refers to inventors (Breschi and Lissoni, 2009). Thus, in line with immigration economics, physical distance is seen to play a preponderant role in the re-location decisions of highly skilled individuals, such inventors. According to a large body of literature (see, recently, Ibrahim et al., 2009), inventors may tend to relocate closer in the space, in order to (i), minimized mobility costs if face-to-face interactions and frequent meetings with colleagues and competitors are needed, and (ii), maintain previously fruitful labour relationships. Thus, the hypothetical existence of *localised knowledge spillovers* would favour the co-location of individuals close to each other in the space in order to access and transmit relevant knowledge. Further, both professional as well as personal ties could be behind low observed mobility of skilled individuals (Fischer et al., 2000) due to the existence of sunk costs related to the loss of accumulated place-specific human capital when moving far away. This *relationship capital*[3] loss (McCann et al., 2010) has prevented individuals to re-locate distant from their home location. Critically, however, Breschi and Lissoni (2009, p. 15) have equally stressed that "workers that embody relevant knowledge may tend to move 'locally', for a number of reasons", which, according to them, have nothing to do with the need to meet with their colleagues in universities or in rival companies for information and help.[4] Given these former arguments, the first hypothesis to test would be the following:

---

[3] Physical separation from family and friends, who are part of this so-called *relationship capital* (Dollahite and Rommel, 1993), remains a huge cost for international and inter-regional migration, in spite of the aforementioned changes related to the globalisation era (McCann et al., 2010).

[4] An interesting debate in regional economics and innovation economics has emerged in recent years on the existence and nature of localised knowledge spillovers and the existence or not of a relationship between them and the agglomeration and concentration of the economic activity, research infrastructures and, consequently, the agglomeration of inventors. The reasons behind the concentration of innovation activity and the existence of localised knowledge spillovers debate (and therefore the reasons behind the influence of distance in mediating inventors' mobility) is an interesting question which, however, goes beyond the scope of the present study (for a critical survey, see Breschi and Lissoni, 2001).

H1. Geographical distance influences the mobility patterns of inventors across regions.

Thus, we would expect, all things being equal, that physical distance between regions is still playing a critical role on mediating knowledge flows driven by mobility of inventors. To test this hypothesis, equation (1) and (2) can be easily extended to the following empirical model,

$$
\begin{aligned}
M_{ij} &= \ln\beta_0 + \beta_1 \ln I_i + \beta_2 \ln I_j + \beta_3 \ln EcoSIZE_i + \beta_4 \ln EcoSIZE_j + \\
&+ \beta_5 \ln DIST_{ij} + \beta_6 NonCONT_{ij} + \omega_{n-1} CO\_DUMM_i + \omega_{n-1} CO\_DUMM_j + \varepsilon_{ij}\,.
\end{aligned}
\tag{3}
$$

In equation (3), geographical distance between regions (DIST) is the main explanatory variable under scrutiny.[5] As for the controls, $EcoSIZE_i$ and $EcoSIZE_j$, and I_i and I_j will account for, respectively, the spatial distribution of economic activity (total population in home and host regions) and innovation activity (number of inventors –employment opportunities therefore), in order to not bias the 'distance' coefficient upward. We will repeat the estimations using GDP and area as 'size' variables, running therefore variants of the same model in order to study the stability of the remaining coefficients –see the 'robustness checks' section. Moreover, an unobservable stochastic component, ε_{ij}, will capture non-observable features in the empirical model –like the personal situation of a given inventor, or her perception of the home and host region characteristics.

Given the size variability of the NUTS3 regions in terms of area, we would like to separate the net effect of distance to movements that may well be occurring within the cities and large metropolitan areas containing more than one NUTS3. For this reason, all our regressions will include a dummy variable reflecting if two regions are contiguous or not –first order contiguity. Further, in order to take fully account of any country-specific non-observable feature that might be affecting the expulsion or attraction of inventors –research prestige, country-specific wage premium, and the like, country-specific fixed-effects of the home and host regions should also be included.

Regional features

Next, from different streams of literature, we would like to test whether certain regional features may act as pushing and pulling forces behind inter-regional inventors' mobility. From the

[5] Note that in equation (3), the dependent variable is not expressed in logarithmic form anymore. Given the count nature of the dependent variable, we were forced to use a log-linear specification instead of a log-log one. We will come back to this point at the end of section 3.

migration literature we learn the importance of monetary rewards differences between locations in driving migration choices (see Ortega and Peri, 2009; and Blackburn, 2010, for very recent contributions). Hence, the income gap ($IncGAP_{ij}$) between pairs of regions (measured by the ratio[6] between GDP per capita of the destination and origin regions) is included, and a positive and significant coefficient is expected.

Besides, the innovation economics literature, and specifically the most related to inventors' individual labour mobility (Lenzi, 2009; Hoisl, 2009; Shalem and Trajtenberg, 2008), has examined the causal relationship between productivity and mobility. Though a consensus has not been reached, and in spite of the different frameworks used in these studies and ours, we include the aggregated average productivity of inventors in our model. On the one hand, we hypothesize that higher levels of productivity in the origin region ($\Pr oductivity_i$) will induce other firms from outside to look for the most talented individuals of that location, so influencing positively the level of out-flows in every couplet of regions. Likewise, higher levels of productivity in the destination region ($\Pr oductivity_j$) will also attract talented individuals, who look for better job matching and career opportunities, so positively influencing inventors' flows from the home to the host region. Home and host regional productivities were separately included –and not the gap between the two- to explore the different effect of each one.

Finally, certain substreams of the economic geography literature, such as the 'creative class' concept by Florida and co-authors (Florida, 2002) and the 'consumer city' tradition by Glaeser and co-authors (Glaeser et al., 2001), have put forward the importance of amenities acting as magnetic factors for highly-skilled labour. As have been already asserted (Eger, 2003), the young, talented individuals wanted by all firms like to live in locations with a "healthy and vibrant cultural life" (Op. Cit., p. 13). Thus, regions with higher supply of cinemas and theatres, bars and restaurants, or simply with a higher degree of multi-culturality and multi-ethnicity may attract skilled individuals to live in. We will proxy this idea using population density ($PopDENS_j$) of the host region. One may expect that in dense, metropolitan areas, amenities' supply should be larger than in rural, less populated areas.[7] Thus,

[6] Other papers (Ortega and Peri, 2009) have used the linear difference between host and home regions' income –or related proxy, though in our framework this would discard the negative observations after the logarithmic transformation, so we decided against the linear difference in favour of the ratio.

[7] Against this widespread view, Storper and Scott (2009) and Scott (2010) argue that the role of amenities (being environmental or cultural ones) has been exaggerated and, even worse, has lead to wrong policy implications and recommendations. On the contrary, they argue, employment opportunities are also behind migration behaviours of the creative class.

H2. Certain region-specific features like their income level, their productivity level, or the supply of amenities, might influence the inventors' location decision.

All in all, model (3) can be enlarged as follows:

$$\begin{aligned} M_{ij} &= \ln\beta_0 + \beta_1 \ln I_i + \beta_2 \ln I_j + \beta_3 \ln EcoSIZE_i + \beta_4 \ln EcoSIZE_j + \\ &+ \beta_5 \ln DIST_{ij} + \beta_6 NonCONT_{ij} + \beta_7 IncGap_{ij} + \beta_8 \Pr oductivity_i + \\ &+ \beta_9 \Pr oductivity_j + \beta_{10} PopDens_j + \omega_{n-1} CO\_DUMM_i + \omega_{n-1} CO\_DUMM_j + \varepsilon_{ij} \end{aligned} \tag{4}$$

Other distances

Even though geographical distance might matter, for a number of reasons sketched above, one could think that other economically meaningful distances might also play an important role. Thus, other relational variables between pairs of regions are also hindering inventors' mobility across space. These relational features are intrinsically spatially correlated, which means that they might well be correlated with physical distance itself, but their separate effect must be investigated. Obviously, the existence of such other distances is facilitated by geographical proximity itself. However, if they are not included in the estimations, the geography coefficient might be biased upward and, even worse, wrong policy implications could be derived. Thus, through the following hypotheses we ask ourselves:

H3a. Are other types of distance, like institutional, technological, industrial, or social distances, relevant in mediating knowledge flows driven by inventors' mobility?

H3b. What is left for the geographical explanation when other distances are taken into account?

Inventors, and knowledge, might flow easily between similar regions, and with more difficulties between different, structurally distant regions. In this final step, we are interested in elucidating the influence of different concepts of distance. We do that in order to test both, their influence in mediating knowledge flows and in order to know whether inventors'-mobility knowledge flows are not geographically mediated anymore when economically meaningful differences between regions are taken into consideration.

Institutional distance: The costs of moving into a region where the institutional framework - especially concerning the system of innovation and the R&D infrastructure- is completely different, are higher than to move to closer regions, sharing the institutional framework (Hoekman et al., 2008; Burger et al., 2009). According to the European Commission (2000), the mobility of researchers is still too much nationally bounded, because of institutional barriers for crossing national boundaries, as well as cultural and idiomatic obstacles.

Structural distance: It is also straightforward to assume that regions with similar economic structures will inter-exchange more knowledge (Maggioni and Uberti, 2009; Moreno et al., 2005) and therefore a higher degree of mobility between them might well exist. Again, the costs of moving and finding a job in a region with a structure very different to the origin region are obviously higher than with a similar region. We will introduce therefore two concepts of structural distance, that is,

- *Technological distance:* An index of technological similarity between pairs of regions is calculated to test to what extent technologically close regions inter-exchange more inventors than technologically distant regions.
- *Industrial distance:* We also calculate a similar index than before, but looking at the industrial division of economic activities (NACE[8]) instead of their technological division.

Social connectedness: Finally, the likelihood to move from one region to another may well also depend upon how well connected are these two regions in terms of social (labour) relationships. Thus, we think that mobility is more likely to occur when inventors from one region maintain repeated work relationships with inventors of the other regions, because, amongst other things, they can get better information about job vacancies (Lenzi, 2009) or entrepreneurship opportunities.

All in all, the following empirical model is going to be estimated:

$$\begin{aligned} M_{ij} &= \ln\beta_0 + \beta_1 \ln I_i + \beta_2 \ln I_j + \beta_3 \ln EcoSIZE_j + \beta_4 \ln EcoSIZE_{ij} + \beta_5 \ln DIST_{ij} + \\ &+ \beta_6 NonCONT_{ij} + \beta_7 IncGap_{ij} + \beta_8 \Pr oductivity_i + \beta_9 \Pr oductivity_j + \\ &+ \beta_{10} PopDens_j + \beta_{11} InstiDIST_{ij} + \beta_{12} Tech\_DIST_{ij} + \beta_{13} Indus\_DIS_{ij} + \\ &+ \beta_{14} SOC\_DIS_{ij} + \omega_{n-1} CO\_DUMM_i + \omega_{n-1} CO\_DUMM_j + \varepsilon_{ij} \end{aligned} \tag{5}$$

Data

We estimate our models for a sample of European NUTS3 regions of 17 countries –see Appendix 1, during the period between 1998 and 2002.[9] Our final sample is made up of 1079

[8] Statistical Classification of Economic Activities, where NACE corresponds to the French acronym for *"Nomenclature Statistique des Activités Économiques"*.

regions. Our dependent variable is built full-counting the movements of inventors crossing regional borders. We therefore construct a mobility asymmetric matrix of 1079 rows and 1079 columns, where each of the elements of the matrix is the number of inventors moving from region *i* to region *j* within our period of analysis. If an inventor moves more than once or she returns to her former region, we count them as separate and independent movements. Since movements from region *i* to region *i* do not exist by definition, we end up with a dependent variable reflecting the fluxes between pairs of regions of (1079)x(1079-1)=1,163,162 observations.

The data for constructing the mobility matrix are taken from the REGPAT database (OECD, January 2009 edition). In spite of the vast amount of information contained in patent documents, a single ID for each inventor and anyone else involved is missing. However, in order to draw the mobility history of inventors, we need to identify them individually by name and surname, as well as via other useful information contained in the patent document. The method chosen for identifying the inventors is therefore of the utmost importance in studies of this nature. Thus, here, we have followed the methodology proposed by Miguélez and Miguélez (2010), who, in line with a growing number of researchers in the field, suggest several algorithms for singling out individual inventors using patent documents. In the present study, this procedure has been used for a subsample of inventors whose patent applications have been made from one of 17 countries during our period of analysis. We only apply the algorithms to EPO[10] patents filed under the Patent Cooperation Treaty (PCT) as well. This condition was adopted for two main reasons: first, because patents filed during an international phase (the PCT) can be assumed to be more technologically and economically exploitable, since applying for EPO patents under PCT procedures is more expensive and time-consuming. Therefore, the knowledge embedded in these patents is presumably of greater worth for innovative activities. The second reason is related to time constraints, because of the fact that the chosen methodology still requires a considerable amount of manual work to ensure minimum levels of reliability. All this process ends up in a sample of 323,339 records, from which we identified 142,915 unique inventors.

Data for the "mass" variables, the total number of inventors (at home and host regions), are also taken from the PCT patents of the REGPAT database, and the procedures described above to identify individual inventors. Population data for the SIZE controls are taken from Eurostat, and calculated as an average of the whole period. We also calculate the income gap (the GDP per capita ratio between pairs of regions) using Espon data, and regional productivity is calculated as

[9] Due to data constraints, however, we have considered Denmark as a single region, the island of Sardinia as a whole NUTS 2 (instead of NUTS 3) and the German *Land* of *Sachsen-Anhalt* as a single NUTS 1 region, and we have omitted the regions of Las Palmas de Gran Canaria, Tenerife, Ceuta, Melilla, Madeira, Açores, Guadeloupe, Martinique, Guyane and Reunion due to their distance from continental Europe.
[10] European Patent Office.

the number of patents from the European Patent Office (EPO) between 1998 and 2002 taken also from the REGPAT database over the number of inventors identified in each region. Area and population to calculate population density are also taken from Eurostat.

Geographical distance between regions (DIST) is calculated using Euclidean distances between the UTM (Universal Transverse Mercator) coordinates of the centroids of each region. Institutional distance is proxied with a dummy variable valued 1 if the couplet of regions does not pertain to the same country and 0 otherwise (as in Ponds et al., 2007 and Hoekman et al., 2008). At the national level, other proxies related to institutional issues must be used (Burger et al., 2009). However, institutional indices with different dimensions are lacking at regional level – especially at NUTS3 level. Patent data to calculate technological and structural distances are also EPO patents taken from the REGPAT database and assigned to each of the technological sectors (and economic sectors) using the IPC[11] classification (and its corresponding NACE classification).[12] Data to construct a collaboration matrix are only taken from PCT patents. To proxy this social connectedness we define,

$$SocCON_{ij} = \frac{s_{ij}}{(PAT_i + PAT_j)}, \tag{6}$$

where s_{ij} are each of the components of an $n-by-n$ matrix, S, which reflects the collaboration intensity of two regions. When one patent contains inventors reporting their addresses in different regions, we assume that there exist cross-regional collaborations between

[11] International Patent Classification.

[12] To proxy technological distance, we use the following index:

$$TechDIST_{ij} = 1 - t_{ij}, \tag{f.1}$$

Where t_{ij} is the uncentered correlation between regional vectors of technological classes in the form of:

$$t_{ij} = \frac{\sum f_{ik} f_{jk}}{\left(\sum f_{ik}^2 \sum f_{jk}^2\right)^{1/2}}. \tag{f.2}$$

In (f.2), f_{ik} stands for the share of patents of one technological class k according to the IPC classification (out of 27 technological classes in the subdivision chosen) of the region i, and f_{jk} for the share of patents of one technological class k of the region j. Thus, values of the index close to zero would indicate that a given pair of regions is technologically similar, and values close to the unity, that are technologically distant. Similarly, the industrial-distance index is calculated in the same vein, but according to the NACE classification of patents (out of 21 industrial sectors in the subdividion chosen).

regions. We 'full-count' all the collaborations across regions, irrespective of the number of inventors reported in each patent.

A summary of the variables included, the proxies used, and the data sources can be found in the Appendix 2. Table 1 below also includes some descriptive statistics of the variables under consideration.

[Insert Table 1 about here]

Table 2 presents the correlation matrix. Two pairs of variables show correlations between 0.6 and 0.7, but, widely speaking, correlation among independent variables is sufficiently small and does not pose a serious problem.

[Insert Table 2 about here]

Econometric issues

Due to a number of characteristics of our dependent variable, several estimation particularities prevent us against the use of usual linear regression models, which would lead us to inconsistent, inefficient, and biased estimates if used. In what follows, we present some of these features and the way in which we have chosen to deal with them. First and foremost, the response variable, counts of flows from the home to the host region, is a discrete one with a distribution that places the probability mass at nonnegative integer values only (Cameron and Trivedi, 1998). In cases like ours (see Cameron and Trivedi, 1998, 2005, for numerous examples of count variables), data are concentrated in few small discrete values skewed to the left and intrinsically heteroskedastic with variance increasing with the mean (Op. Cit.). Second of all, it contains a number of zeros. Any attempt to normalize the variable through, for instance, a logarithmic transformation would return a large number of missing values. Adding a constant to the zero values to make them positive in order to allow for a logarithmic transformation, or just to remove the zero observations, would be misleading, and would lead us to biased and inconsistent estimates (Burger et al., 2009). In short, the use of linear regression models for count outcomes such the one of the present framework may lead to inefficient, inconsistent, and biased estimates (Long, 1997).

In such a setting, the literature has suggested the use of count data models. The most basic type of count data model, the Poisson regression, assumes that, in analogy to our case, the probability to observe a move from region i to region j follows a Poisson distribution

$$\Pr\left[M_{ij}\right]=\frac{\exp(-\mu_{ij})\mu_{ij}^{M_{ij}}}{M_{ij}}, \qquad (7)$$

with a conditional mean (μ) of the distribution that is a function of the independent variables $\mu_{ij}=\exp\left(\beta_0+\beta_1 x_i+\beta_2 x_j+\beta_3 x_{ij}\right)$. This generates estimates of M_{ij}, and not $\ln\left(M_{ij}\right)$. In this way, we would avoid the underprediction of large migration flows (Burger et al., 2009).

Although more appropriate to analyse count events than linear models, the Poisson distribution assumes equidispersion, that is to say, the conditional variance being equal to the conditional mean of the outcome. However, the conditional variance often exceeds the conditional mean (Burger et al., 2009; Long, 1997), which is a clear symptom of overdispersion. Intuitively, the presence of overdispersion in count data models has similar consequences as the presence of heteroskedasticity in linear models (Cameron and Trivedi, 2005). To be precise, overdispersion appears due to the presence of individual unobserved heterogeneity in the data generating process, which is not captured by the Poisson distribution. As a result, the Poisson regression will lead to consistent but inefficient estimates (Burger et al., 2009), with standard errors biased downward (Cameron and Trivedi, 1986; Long, 1997). To deal with this issue, the negative binomial regression is preferred -which can be estimated, for instance, by means of maximum likelihood procedures. In such a modelization, the expected value is the same as in the Poisson model, but the variance is specified as a function of both the conditional mean and a dispersion parameter (α). This dispersion parameter allows us to incorporate the unobserved heterogeneity into the conditional mean (Long, 1997).

Another important issue which has not deserved enough attention in the related literature is the excess of zeros. This problem occurs when the number of zeros in the dependent variable has greater frequency than would otherwise be predicted by the Poisson or Negative binomial models (Greene, 1994). In such a setting, some of the zeros in the data are produced by a different process than the remaining zeros (Burger et al., 2009). In our framework, for instance, it would be important to differentiate between the lack of migration movements because of a lack of resources; and the lack of movements because of the effects of the covariates included (which the usual count models would predict). To deal with such an issue, several strategies can be adopted, like the Heckman selection model or zero inflated count data models. In such zero-inflated models, as suggested by Mullhay (1986), the population is formed by two groups. One individual is in the first group with probability φ, and he is in the second group with probability $1-\varphi$. Formally (Cameron and Trivedi, 1998):

$$\Pr\left[y_{ij} = 0\right] = \varphi_{ij} + (1-\varphi_{ij})\exp(-\mu_{ij}) , \tag{8}$$

and

$$\Pr\left[y_{ij} = r\right] = (1-\varphi_{ij})\frac{\exp(-\mu_{ij})\mu_{ij}^{r}}{r!} , \; r = 1,2,... \tag{9}$$

where φ_{ij} is the proportion of observations with a strictly zero count. In the second group, we observe zero counts now, although they can potentially be positive counts in another moment in time based on the characteristics of the pair of regions (both attributional and relational characteristics).

Summing up, in a first step the probability of being in one or another group is estimated with a logit or a probit model. Subsequently, a negative binomial model is used to estimate the number of events in the second group. Overall, the predicted probability of a zero count in a zero-inflated negative binomial model would be driven by

$$\Pr = (y = 0|x) = \hat{\varphi} + (1-\hat{\varphi})\left(\frac{\hat{\alpha}^{-1}}{\hat{\alpha}^{-1} + \hat{\mu}}\right)^{\hat{\alpha}^{-1}} , \tag{10}$$

and for a positive count,

$$\Pr = (y|x) = (1-\hat{\varphi}) + \frac{\Gamma(y+\hat{\alpha}^{-1})}{y!\Gamma(\hat{\alpha}^{-1})}\left(\frac{\hat{\alpha}^{-1}}{\hat{\alpha}^{-1} + \hat{\mu}}\right)^{\hat{\alpha}^{-1}}\left(\frac{\hat{\mu}}{\hat{\alpha}^{-1} + \hat{\mu}}\right)^{y} , \tag{11}$$

In principle, there is no formal restriction to include the same regressors both in the binary and the negative binomial process –aside from possible theoretical considerations.

4. Results

In this section we summarize the main results encountered throughout the estimation of the models suggested in section 3. We have estimated, step by step, the different models in different columns. Both the negative binomial and the logit estimations are presented. For the NB regression, since the covariates are (most of them) expressed in logarithmic form, the estimated coefficients can be interpreted as elasticities (Cameron and Trivedi, 1998; Long, 1997). Thus, for instance, an increase of 1% of the number of inventors in the home region would lead to an increase of 0.94% of the probability to observe a move from the home to the host region, holding all other variables constant. The interpretation of the logit coefficients is different: if the distance coefficient is 0.60, it means that an increase of 1% of the distance between a couplet of regions leads to an increase of 0.60% of the probability to belong to the "strictly zero group" (Maggioni and Uberti, 2009).

Column (i) presents the estimation of model 3. As expected, the "mass" variables (number of inventors in the home and the host regions) are positive and significant. Distance and non-contiguity are negative and strongly significant, so we corroborate the hypothesis that geography matters in driving the patterns of labour mobility of inventors across the European geography. Its coefficient, around 0.86, is larger than we would expect at the beginning. In reality, the elasticity is very close to what is found in similar frameworks for trade data (see Disdier and Head, 2008, for a meta-analysis on this topic) or co-patenting data (Maggioni and Uberti, 2009), and largely higher than what is found for citation data (Peri, 2005). In fact, these coefficients are in line to what is found in the migration literature in similar frameworks at the European regional level (Crozet, 2004; Guarner-Sanchis and López-Bazo, 2006).

As for the size variable, negative and significant signs, especially for the home region size variables, are found. We explain this result as follows: large, populated regions tend to serve job opportunities for inventors by themselves. While the "inventors" variable captures the potential inventors that move, the size variables take into account this effect of the local labour market. As stressed by Guellec and van Pottelsberghe de la Potteire (2001), small regions tend to be more opened, including to what refers to in- and out-flows of inventors, while contrary, larger regions, do not. In any case, their quantitative effect is small.

[Insert Table 3 about here]

Column (ii) shows the results of the estimation of model 4. From this column we learn that the atributional features suggested in the theoretical section as driving inter-exchanges of inventors do not seem to matter at all (or influences mobility contrarily to what we were expecting). Thus, for instance, productivity might seem to negatively affect mobility. As stressed elsewhere (Hoisl, 2009), this might be explained due to the fact that (i), firms have huge incentives to retain their most valuable human capital; and (ii), it could be the case that some highly productive individuals (hosted in highly productive regions) may have non-compete agreements with their employers, which hinders their mobility elsewhere. This point is also stressed by Fleming et al. (2007) for the US and Lissoni (2000) for the case of Northern Italy.

On the other side, the income gap, although positive, does not seem to influence mobility. Though the point estimate has the expected sign, the coefficients are not significant at standard levels of confidence. This result is also found in Scott (2010) for US data –in a rather different empirical framework though, who argues that it could be the case "that engineers are relatively insensitive to wage differences across geographic space in relation to potential employment opportunities". Besides, the amenities proxy (population density) is not significant either. Again, this is in line with what is found in Scott (2010). According to him (see also Storper and Scott, 2009, arguments), people are mostly attracted by employment opportunities. Hence, most immigrants are unlikely to "be able to move in significant numbers from one location to another unless relevant employment opportunities are actually or potentially available", so employment-generating policies are better recommended rather than amenity-magnetic actions.[13] Similar results are found in Hansen and Niedomsyl (2009) using Swedish data.

Column (iii) displays the estimates of model 5. Institutional, technological, and industrial distances, as well as social connectedness, are included as regressors. From this column we conclude some main findings. First and foremost, all four variables included are significant and with the expected sign. Overall, these results support the idea that other distances (and proximities) are driving inventors' mobility across European regions as well, above and beyond geographical separation. Thus, knowledge, and those who carry it, seems to flow easily between similar regions (similar results are found in Maggioni and Uberti, 2009, using co-patents). Of special interest is the strong effect of institutional proximity, proxied here by country dummies, which shows the largest coefficient amongst the different distances considered. In this case, similar results are found in Hoekman et al. (2009) for the case of co-patenting activity. Second of all, it is also worth to be highlighted the downward shift in the geographical distance coefficient (which is now less than a half than in columns (i) and (ii)). Again, it is obvious that both geographical and the other distances overlap, but each feature might has a different and independent effect on mobility that must be isolated correctly.

Robustness checks

In the final part of this section, we perform few robustness checks by including several modifications in the estimations presented so far, in orther to study the stability and significance of the parameters. First of all, given the strong significance of the first-order non-contiguity variable,

[13] Note that one could also argue that this is because of the fact that the subset of regions belonging to the second group of observations in the ZINB procedure (those not belonging to the strictly zero group) share a set of structural and economic characteristics that made them similar in terms of income and level of amenities, making these variables not drivers of mobility flows of inventors.

we include second and third-order non-contiguity variables and re-estimate model 5. We do that in order to not attribute to distance any effect derived from the existence of large metropolitan, urban areas covering more than one NUTS3 region. Results of these estimations are shown in table 4 below, column (i). Fortunately, neither of the included variables turns out to be significant, and the parameters for the remaining variables remain virtually unchanged.

[Insert Table 4 about here]

Another important issue we would like to address is the importance of the National System of Innovation. Given the influence of the Institutional Distance variable in the estimations performed so far, we would like to assess to what extent country effects are driven the relational attributes between pairs of regions. With this idea in mind, we have repeated the estimation of model 5 in column (ii) of table 4, but only including inter-regional movements between regions belonging to different countries in Europe. From this estimation we can conclude that most of the findings encountered so far remain when only cross-country mobility is considered. However, few aspects deserve a more detailed discussion. First, the physical distance coefficient is slightly increased. Contrarily, industrial and, especially, technological distances are somehow decreased – making them non-significant at standard confidence levels. Meanwhile, the social connectedness variable increases between 3 and 4 times in these estimations. It is therefore clear to us that social contacts stand out from the other relational variables when only international mobility is considered. Thus, it seems that when an inventor has to cross a national border, and adapt herself to a completely different environment, previous social relationships plays a preponderant role. It is also interesting to note that the inventors' productivity variables are now positive both of them, and the host region productivity strongly significant. Thus, the ideas considered in the theoretical section about the attractiveness of highly productive places for inventors from abroad seems to only hold when national borders must be crossed. However, more research on this issue must be undertaken before going further on.

Other robustness checks are also performed, though for the sake of brevity, results are not displayed here. Thus, for instance, we perform all the estimations by removing irrelevant variables. What we have in mind is to avoid these non-significant variables to account spuriously for variation in the dependent variable. Fortunately, any significant change stands out –results provided from the authors upon request. Equally, probit estimations –instead of logit- are also performed for the first part of the ZINB and again, changes must not be reported. Finally, as said before, we have also repeated all the estimations using GDP and area as "size" variables. Any significant change does not stand out.

5. Conclusions, implications, and limitations

The literature of geographical knowledge flows and spillovers usually points at skilled people's mobility as cross-regional knowledge flow mechanism. However, the features driving this phenomenon are rarely investigated. In the present inquiry, we have tried to fill in this gap by estimating a gravity model to analyse the mobility patterns of inventors across European NUTS3 regions. In the theoretical discussion, we have highlighted a number of factors likely to affect inter-regional mobility, which have been tested in the empirical section.

Quite surprisingly, our results confirm, by and large, that one of the most important territorial determinants of inventors' (and knowledge) inter-exchange between pairs of regions is still physical distance –this effect is very robust to sample choice (when intra-country movements are excluded), specification (negative binomial, logit, probit), and inclusion of controls (atributional and relational variables). Given the lowering transportation costs, the abolition of national frontiers within Europe, and the type of people we are analysing (less sensitive to geographical distance in their decision to move) we expected a relatively lower importance of geography, although this was not the case. Far from the announcements of "the death of distance" (Cairncross, 1997) or "the end of geography" (O'Brien, 1992), physical distance is still playing a preponderant role in mediating knowledge flows driven by inventors' mobility across regions. However, other differences (or similarities) across regions, like their industrial or technological specialisation, or their social (labour) connectedness, are critical as well. Similarly, institutional distance (measured as pertaining or not to the same country) is driving heavily the phenomenon under study. Thus, in the path towards economic convergence and the creation and strengthen of the European Research Area, it is straightforward to conclude that a more active policy to smooth the frictions obstructing mobility, and to facilitate cross-regional and especially cross-national one is required.

An important issue that should be put to the forefront is the non-significance (or the unexpected negative sign) of some of the attributional variables not related to 'size' (those of the second hypothesis). This result is quite surprising, as well as very important. It seems from the empirical evidence that individual preference-seeking forces do not seem to play a significant role. Being this the case, policy actions based on Glaeser's and Florida's arguments can lead to contradictory results unless regions do not belong to the network of locations among which highly-killed labour flows in their search for employment opportunities. In any case, further research in this direction should be undertaken in order to confirm or reject this extreme.

In what follows, we will discuss some of the limitations of our approach that, rather than undermine the analysis undertaken so far, will serve as the basis for improvements of the current study. First and foremost, the quality of the data could be enhanced. Better procedures to improve the identification of inventors will surely lead to more consistent estimates on the relationships analysed in the present paper. In this sense, the extension to all EPO patents and not only to those EPO patents applying also to the PCT procedures, will increase the number of inventors identified and therefore the number of non-zero events of our dependent variable.

The second notoriously drawback is related to simultaneity and endogeneity issues. In the first version of this study, we have estimated a cross-section regression using the same time span (1998-2002) for both the dependent and independent variables. However, causality could go the other way around at least for the variables inventors, population, income gap, productivity, population density, and specially technological, industrial, and social distance. The most popular way to deal with such an issue is to lag one (or more) period the explanatory variables. Lagging variables of the r.h.s. of the models attempts to lessen endogeneity and reverse-causality problems, so implicitly assuming that weak exogeneity is enough to get consistent results. However, our dependent variable feeds from patenting activity, which time-lags might well be influencing the time lags of our independent variables, and therefore consistency will be affected. The search for suitable instruments and the use of GMM techniques to deal with this issue will be accomplished in a near future. Related to the issue of endogeneity and causality is the potential use of longitudinal data. If the data treatment can be extended, not only to all the EPO patents, but also to additional time-spans, the time dimension could be exploited (by including time-dummies or a time-trend).

References

AGRAWAL A., COCKBURN I. and MCHALE J. (2006) Gone but not forgotten: labour flows, knowledge spillovers, and enduring social capital, Journal of Economic Geography 6, 571-91;

ALMEIDA P. and KOGUT B. (1999) Localisation of knowledge and the mobility of engineers in regional networks, Management Science 45, 905-17;

BORJAS G. (2000) "Economics of Migration", *International Encyclopedia of the Social and Behavioral Sciences*, Section No. 3.4, Article No. 38;

BRESCHI S., LENZI C., LISSONI F. and VEZZULLI A. (2009) "The geography of knowledge spillovers: the role of inventors' mobility across firms and in space", in BOSCHMA R. and MARTIN R. (eds.), *Handbook of Evolutionary Economic Geography*, Edward Elgar.

BRESCHI S. and LISSONI F. (2009) "Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows", *Journal of Economic Geography* 9, 4, 439-68;

BURGER M., VAN OORT F. and LINDERS G. (2009) "On the Specification of the Gravity Model of Trade: Zeros, Excess Zeros and Zero-inflated Estimation", *Spatial Economic Analysis* 4(2), pp. 167-190;

CAIRNCROSS, F. (1997) *The Death of Distance: How Communications Revolution Will Change Our Lives*. London: Orion Business Books.

CAMAGNI, R. AND CAPELLO, R. (2009) "Knowledge-base economy and knowledge creation: The role of space", in FRATESI, U. and SENN, L. (Eds.) *Growth and Innovation of Competitive Regions: The Role of Internal and External Connections*, pp. 145-166. Springer-Verlag, Berlin.

CAMERON A. C. and TRIVEDI P. K. (2005) *Microeconometrics. Methods and applications*. Cambridge University Press, New York.

CAMERON A. C. and TRIVEDI P. K. (1998) *Regression Analysis of Count Data*. Econometric Society Monograph No.30, Cambridge University Press.

CORREDOIRA R. A. and ROSENKOPF L. (2006) Learning from those who left: the reverse transfer of knowledge through mobility ties, Management Department Working Paper

CRESPI G., GEUNA A. and NESTA L. (2007) The mobility of university inventors in Europe, Journal of Technology Transfer 32, 3, 195-215;

CROZET M. (2004) "Do migrants follow market potentials? An estimation of a new economic geography model", *Journal of Economic Geography*, 4(4), pp. 439-458;

DISDIER, A.-C. and HEAD, K. (2008) "The puzzling persistence of the distance effect on bilateral trade" Review of Economics and Statistics 90(1), pp. 37-48;

DÖRING, T. and SCHNELLENBACH, J. (2006) What do we know about geographical knowledge spillovers and regional growth?: A survey of the literature, Regional Studies 40.3, 375-395;

EGER J. M. (2003) The creative community. The California Institute for Smart Communities, San Diego, California;

EJERMO O. and KARLSSON C. (2006) "Interregional inventor networks as studied by patent coinventorships", *Research Policy* 35, pp. 412-430;

EUROPEAN COMMISSION (2000) “Towards an European research area”, COM (2000) 6

FLORIDA R. (2005) *The Flight of the Creative Class: The New Global Competition for Talent.* Harper Collins, London.

FLORIDA R. (2002) *The Rise of the Creative Class: And How It's Transforming Work, Leisure, Community and Everyday Life*. Basic Books.

FISCHER M. and GRIFFITH D. A. (2008) "Modeling Spatial Autocorrelation In Spatial Interaction Data: An Application To Patent Citation Data In The European Union" *Journal of Regional Science*, 48(5), p. 969-989;

FISCHER M., SCHERNGELL T. and JANSENBERGER E. (2009) "Geographic localisation of knowledge spillovers: evidence from high-tech patent citations in Europe" *The Annals of Regional Science*, 43(4), p. 839-858;

FRATESI, U. and SENN, L. (2009) “Regional Growth, Connections and Economic Modelling: An Introduction”, in FRATESI, U. and SENN, L. (Eds.) *Growth and Innovation of Competitive Regions: The Role of Internal and External Connections*, pp. 3-28. Springer-Verlag, Berlin.

GLAESER E. L., KOLKO J. and SAIZ A. (2001) "Consumer city," *Journal of Economic Geography* 1(1), pp. 27-50;

GUELLEC D. and VAN POTTELSBERGHE DE LA POTTERIE B. (2001) “R&D and productivity growth. Panel data analysis of 16 OECD countries” *OECD Science, Technology and Industry Working Papers*, 2001/3, OECD publishing;

GREENE W. H. (1994) “Accounting for excess zeros and sample selection in Poisson and Negative Binomial regression models”, Working Paper No. 94-10. New York: Stern School of Business, New York University, Department of Economics;

GROSSMAN. G. M. and E. HELPMAN (1991) *Innovation and Growth in the Global Economy.* MIT Press, Cambridge, MA.

HOEKMAN J., FRENKEN K. and VAN OORT F. (2009) The geography of collaborative knowledge production in Europe, The Annals of Regional Science 43, 3, 721-38;

HOISL K. (2009) Tracing mobile inventors: The causality between inventor mobility and inventor productivity, Research Policy 36, 5, 615-36;

JAFFE A.B., TRAJTENBERG M. and HENDERSON R. (1993) “Geographic localisation of knowledge spillovers as evidenced by patent citations”, *Quarterly Journal of Economics* 108: 577-598;

LENZI C. (2009) “Patterns and determinants of skilled workers’ mobility: evidence from a survey of Italian inventors”, *Economics of Innovation and New Technology* 18, 2, 161-79;

LESAGE J. P., FISCHER M. and SCHERNGELL, T. (2007) "Knowledge spillovers across Europe: Evidence from a Poisson spatial interaction model with spatial effects" *Papers in Regional Science*, 86(3), p. 393-421;

LEWER J. J. and VAN DER BERG H. (2008) “A gravity model of immigration”, *Economics Letters* 99, pp. 164-167;

LONG J. S. (1997) *Regression Models for Categorial and Limited Dependent Variables.* Thousand Oaks, CA, Sage.

MAGGIONI M. A. and UBERTI T.E. (2007) "International networks of knowledge flows" in FRENKEN K. (ed) *Applied evolutionary economics and economic geography.* Edward Elgar, Cheltenham

MAGGIONI M. A. and UBERTI T. E. (2009) Knowledge networks across Europe: which distance matters?, The Annals of Regional Science 43, 3, 691-720;

MARSHALL, A. (1961) [1920] *Principles of Economics.* Macmillan and co, London.

MIGUÉLEZ E. and MORENO R. (2010) "Research networks and inventors' mobility as drivers of innovation: Evidence from Europe", IREA Working Papers 201001;

MIGUÉLEZ E. and MIGUÉLEZ I. G. (2010) Singling out individual inventors from patent data, unpublished manuscript, http://ub.academia.edu/documents/0058/3908/Miguelez_Miguelez__2010__-_Singling_out_individual_inventors_from_patent_data.pdf;

MONTOBBIO F., PRIMI A. and STERZI V. (2010) "Do IPRs Foster International Technological Cooperation?", unpublished manuscript.

MORENO R., PACI R. and USAI S. (2005) Spatial spillovers and innovation activity in European regions, Environment and Planning A 37, 10, 1793-812;

MULLHAY J. (1986) "Specification and testing of some modified count data models", *Journal of Econometrics* 33, pp. 341-365;

O'BRIEN, R. (1992) *Global Financial Integration: The End of Geography.* New York: Council on Foreign Relations Press.

PACI, R. and USAI, S. (2008) "Knowledge flows across European regions", *The Annals of Regional Science* 43(3), 669-690;

PERI G. (2005) Determinants of Knowledge Flows and Their Effect on Innovation, The Review of Economics and Statistics 8, 2, 308-22;

PICCI, L. (2009) "The Internationalization of Inventive Activity: A Gravity Model using Patent Data", unpublished manuscript

PONDS R., VAN OORT F. and FRENKEN K. (2007) The geographical and institutional proximity of research collaboration, Papers in Regional Science 86, 3, 423-43;

RIGGI, M. R. AND MAGGIONI, M. A. (2009) "Regional growth and the co-evolution of clusters: The role of labour flows", in FRATESI, U. and SENN, L. (Eds.) *Growth and Innovation of Competitive Regions: The Role of Internal and External Connections*, pp. 245-268. Springer-Verlag, Berlin.

ROMER P. M. (1986) Increasing Returns and Long-run Growth, Journal of Political Economy 94, 5, 1002-37;

ROMER P. M. (1990) Endogenous Technological Change; Journal of Political Economy 98, 5, 71-102;

ROSENKOPF L. and ALMEIDA P. (2003) Overcoming Local Search Through Allainces and Mobility, Management Science 49, 6, 751-66;

SANCHIS-GUARNER R. and LOPEZ-BAZO E. (2006) "Are Skilled Workers More Attracted to Economic Agglomerations?" ERSA conference papers

SAXENIAN A. (1994) Regional Advantage. Culture and Competition in Silicon Valley and Route 128. Harvard University Press, Cambridge, MA.

SCHWARTZ A. (1973) "Interpreting the Effect of Distance on Migration", *Journal of Political Economy*, 81 (5) pp. 1153-1169;

SCOTT, A. J. (2010) Jobs or amenities? Destination choices of migrant engineers in the USA, *Papers in Regional Science* 89, 1(3), pp. 43-63;

SINGH J. (2005) Collaborative Networks as Determinants of Knowledge Diffusion Patterns, Management Science 51, 5, 756-70;

SHALEM R. and TRAJTENBERG M. (2008) Software patents, inventors and mobility, unpublished Working Paper

STORPER, M. and SCOTT, A. J. (2009) Rethinking human capital, creativity and urban growth, *Journal of Economic Geography* 9(2), Pp. 147-167;

SONG J., ALMEIDA P. and WU G. (2003) Learning-by-hiring: When is mobility more to facilitate interfirm knowledge transfer?, Management Science 49, 4, 351-65;

TABUCHI T. and THISSE J.-F. (2002) "Taste heterogeneity, labour mobility and economic geography", CEPR Discussion Paper No. 3114;

TRAJTENBERG M. and SHIFF G. (2008) "Identification and mobility of Israeli patenting inventors", The Pinhas Sapir Center for Development, Tel Aviv University, DP No. 5-2008;

TRIPPL, M. (2009) "Scientific mobility, international knowledge circulation and regional development", unpublished manuscript;

TRIPPL M. and MAIER G. (2007) Knowledge spillover agents and regional development, SRE-Discussion 2007/01

THOMPSON P. and FOX-KEAN M. (2005) "Patent citations and the geography of knowledge spillovers: a reassessment", *American Economic Review* 95, pp. 450–460;

VUONG, Q. H. (1989) "Likelihood ratio tests for model selection and non-nested hypothesis", *Econometrica* 57, pp. 307-333;

ZUCKER LG, DARBY MR, and BREWER MB (1998) "Intellectual human capital and the birth of U.S. biotechnology enterprises", *American Economic Review* 88(1): 209-306;

ZUCKER LG, DARBY MR, TORERO M (2002) Labor Mobility from Academe to Commerce. *Journal of Labor Economics* 20(3): 629-660;

Appendices

Appendix 1: List of countries

Austria –AT-, Belgium –BE-, Switzerland –CH-, Germany –DE-, Denmark –DK-, Spain –ES-, Finland –FI-, France –FR-, Greece –GR-, Ireland –IE-, Italy –IT-, Luxemburg –LU, the Netherlands –NL-, Norway –NO-, Portugal –PT-, Sweden –SE-, United Kingdom –UK-.

Appendix 2: Variables to be included

Variable	Proxy	Time span	Source
Inventors' flows	Counts of flows from home to host region	1998-2002	REGPAT and own calculations
Mass	Number of inventors	1998-2002	REGPAT and own calculations
Size	Population	1998-2002	Eurostat
Geographic distance	Euclidean distance between UTM regional centroids		GIS
Non-contiguity	1: non-cont; 0 otherwise		GIS
Income gap	GDP PPP p.c. ori. / GDP PPP p.c. dest.	2000	ESPON
Productivity	Num. Patents per inventor	1998-2002	REGPAT and own calculations
Population density	Population / Area	1998-2002	Eurostat
Institutional distance	1: dif. country; 0 otherwise		
Technological distance	$1-\left(\frac{\sum f_{ik} f_{jk}}{\left(\sum f_{ik}^2 \sum f_{jk}^2\right)^{1/2}}\right)$	1998-2002	REGPAT and own calculations
Industrial distance	$1-\left(\frac{\sum g_{il} g_{jl}}{\left(\sum g_{il}^2 \sum g_{jl}^2\right)^{1/2}}\right)$	1998-2002	REGPAT and own calculations
Social connectedness	$\frac{s_{ij}}{(PAT_i + PAT_j)}$	1998-2002	REGPAT and own calculations

Table 1. Summary statistics

	Observations	Mean	St. Dev	Coef. Var.	Min.	Max.
Inventors $_O$	1,163,162	197.17	410.61	2.08	1.00	5443.00
Inventors $_D$	1,163,162	197.17	410.61	2.08	1.00	5443.00
Popul. $_O$	1,163,162	357.87	475.16	1.33	15.00	5338.00
Popul. $_D$	1,163,162	357.87	475.16	1.33	15.00	5338.00
GDPppp $_O$	1,163,162	21339.86	8112.77	0.38	8092.50	120637.70
GDPppp $_D$	1,163,162	21339.86	8112.77	0.38	8092.50	120637.70
Area $_O$	1,163,162	3305.82	6815.05	2.06	34.90	106011.50
Area $_D$	1,163,162	3305.82	6815.05	2.06	34.90	106011.50
EcoGap	1,163,162	1.11	0.55	0.50	0.07	14.91
Product. $_O$	1,163,162	2.02	5.47	2.71	0.00	16.06
Product. $_D$	1,163,162	2.02	5.47	2.71	0.00	16.06
Density. $_D$	1,163,162	488.67	1066.28	0.61	1.50	20228.30
GeoDist.	1,163,162	10.97	7.13	0.65	0.02	47.60
NonCont.	1,163,162	1.00	0.07	0.07	0.00	1.00
InstiDist.	1,163,162	0.81	0.39	0.50	0.00	1.00
TechDist.	1,163,162	0.42	0.21	0.50	0.00	1.00
IndusDist.	1,163,162	0.34	0.23	0.69	0.00	1.00
SocDist.	1,163,162	0.04	0.11	2.84	0.00	0.78

Notes: Summary statistics are calculated using the raw variables before any logarithmic transformation.

Table 2. Correlation matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Inventors(o)	1.000													
	-													
2. Inventors(d)	-0.001	1.000												
	0.317	-												
3. Popul.(o)	0.575	-0.001	1.000											
	0.000	*0.565*	-											
4. Popul.(d)	-0.001	0.575	-0.001	1.000										
	0.565	*0.000*	*0.317*	-										
5. EcoGap	0.364	-0.364	0.159	-0.159	1.000									
	0.000	*0.000*	*0.000*	*0.000*	-									
6. Product.(o)	0.302	0.000	0.020	0.000	0.240	1.000								
	0.000	*0.774*	*0.000*	*0.984*	*0.000*	-								
7. Product.(d)	0.000	0.286	0.000	0.010	-0.232	-0.001	1.000							
	0.778	*0.000*	*0.985*	*0.000*	*0.000*	*0.328*	-							
8. PopDens	0.000	0.417	0.000	0.256	-0.320	0.000	0.276	1.000						
	0.676	*0.000*	*0.798*	*0.000*	*0.000*	*0.777*	*0.000*	-						
9. GeoDis	-0.188	-0.188	0.085	0.085	0.000	-0.278	-0.272	-0.219	1.000					
	0.000	*0.000*	*0.000*	*0.000*	*1.000*	*0.000*	*0.000*	*0.000*	-					
10. NonCont	-0.008	-0.008	-0.007	-0.007	0.000	-0.005	-0.005	0.006	0.240	1.000				
	0.000	*0.000*	*0.000*	*0.000*	*0.996*	*0.000*	*0.000*	*0.000*	*0.000*	-				
11. InstiDis	-0.049	-0.049	0.134	0.134	0.000	-0.118	-0.115	-0.128	0.636	0.129	1.000			
	0.000	*0.000*	*0.000*	*0.000*	*1.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	-			
12. TechnoDis	-0.245	-0.245	-0.146	-0.146	0.000	-0.129	-0.127	-0.134	0.207	0.059	0.078	1.000		
	0.000	*0.000*	*0.000*	*0.000*	*0.831*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	-		
13. IndusDis	-0.181	-0.181	-0.113	-0.114	0.000	-0.121	-0.117	-0.088	0.166	0.047	0.049	0.687	1.000	
	0.000	*0.000*	*0.000*	*0.000*	*0.857*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	-	
14. SocProx	0.027	0.027	0.008	0.008	0.000	0.009	0.008	0.017	-0.157	-0.228	-0.127	-0.068	-0.052	1.000
	0.000	*0.000*	*0.000*	*0.000*	*0.972*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	-

Notes: The correlation matrix has been calculated having log-transformed the variables, except for the case of Institutional Distance and Non-contiguity. In italics, the p-values of the significance below each correlation are shown.

Table 3. Robust zero-inflated negative binomial estimation

	(i)		(ii)		(iii)	
	NegBin	Logit	NegBin	Logit	NegBin	Logit
ln(InventorsOrigin)	0.94***	-0.47***	0.92***	-0.49***	0.75***	-0.61***
	(0.04)	(0.05)	(0.07)	(0.06)	(0.03)	(0.04)
ln(InventorsDestination)	0.90***	-0.47***	0.90***	-0.46***	0.73***	-0.59***
	(0.04)	(0.05)	(0.06)	(0.06)	(0.03)	(0.04)
ln(POP – origin)	-0.07***	0.02	-0.07**	-0.01	-0.11***	0.03
	(0.03)	(0.07)	(0.03)	(0.07)	(0.03)	(0.05)
ln(POP – destin)	-0.06*	0.02	-0.06*	0.01	-0.10***	0.02
	(0.03)	(0.07)	(0.03)	(0.09)	(0.03)	(0.05)
ln(GeoDistance)	-0.86***	0.60***	-0.86***	0.64***	-0.40***	0.29***
	(0.04)	(0.09)	(0.07)	(0.15)	(0.03)	(0.04)
Non-CONTIGUITY	-1.00***	3.26***	-0.92***	3.31***	-0.59***	3.61***
	(0.08)	(0.60)	(0.11)	(0.60)	(0.06)	(1.05)
ln(IncomeGap)			0.03	0.01	0.04	0.03
			(0.04)	(0.09)	(0.04)	(0.06)
ln(ProductivityOrigin)			-0.14***	0.11	-0.09*	0.02
			(0.05)	(0.14)	(0.05)	(0.07)
ln(ProductivityDestin)			-0.13**	0.10	-0.06	0.11
			(0.06)	(0.12)	(0.05)	(0.07)
ln(PopDens)			-0.00	-0.04	0.01	-0.00
			(0.02)	(0.04)	(0.02)	(0.03)
InstiDistance (Dummy)					-0.77***	1.52***
					(0.07)	(0.10)
ln(TECH_DIST)					-0.16***	0.06
					(0.03)	(0.04)
ln(IndustrialDistance)					-0.04*	0.05
					(0.02)	(0.03)
ln(SocialProximity)					0.23***	-1.07***
					(0.05)	(0.25)
Intercept	-10.14***	1.03	-9.21***	1.03	-7.34***	3.09***
	(0.51)	(0.81)	(0.75)	(0.98)	(0.27)	(1.10)
Country Origin Fixed Effects	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*
Country Destination Fixed Effects	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*
Sample size	1,163,162	1,163,162	1,163,162	1,163,162	1,163,162	1,163,162
Nonzero onservations	12591		12591		12591	
Log-pseudolikelihood	-54325.6		-54275.5		-52444.33	
Overdispersion $(\ln\alpha)$	0.48***		0.43***		-0.28**	
	(0.06)		(0.10)		(0.11)	
Voung statistic	18.89		17.45		14.60	
p-value	0.000		0.000		0.000	
AIC	108809.2		108725.1		105078.7	
Schwartz	109754.6		109763.4		106212.5	

Notes: Robust standard errors are presented in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Models 3, 4, and 5 of the theoretical section are estimated in columns (i), (ii), and (iii) respectively. Each of the columns includes the first stage of the ZINB, the logit model, and the negative binomial estimation. Overdispersion tests largely reject the null hypothesis of no overdispersion. When the dispersion parameter, α , is zero, the negative binomial model reduces to the Poisson model. Therefore a likelihood ratio test on α can be computed, where $H_0 : \alpha = 0$. Vuong statistics (Vuong, 1989), are also performed and reported at the bottom of each regression. In a nutshell, the Vuong test examines if significant evidence for excessive zeros exist and therefore whether zero-inflated models should be used. The tests performed point to the need of the zero-inflated models to accommodate our estimations to the excess of zeros.

Table 4. Robust zero-inflated negative binomial estimation

	(i)		(ii)	
	NegBin	Logit	NegBin	Logit
ln(InventorsOrigin)	0.76***	-0.62***	0.71***	-0.75***
	(0.03)	(0.04)	(0.07)	(0.07)
ln(InventorsDestination)	0.73***	-0.60***	0.65***	-0.75***
	(0.03)	(0.04)	(0.07)	(0.08)
ln(POP – origin)	-0.12***	0.04	-0.25***	-0.05
	(0.03)	(0.05)	(0.06)	(0.08)
ln(POP – destin)	-0.11***	0.04	-0.14*	0.11
	(0.03)	(0.05)	(0.08)	(0.09)
ln(GeoDistance)	-0.40***	0.09	-0.48***	0.08
	(0.03)	(0.06)	(0.09)	(0.11)
Non-CONTIGUITY1	-0.57***	5.08*	0.28	7.19*
	(0.09)	(3.07)	(0.31)	(3.68)
Non-CONTIGUITY2	0.05	1.17***		
	(0.07)	(0.18)		
Non-CONTIGUITY3	0.01	0.43***		
	(0.09)	(0.15)		
ln(IncomeGap)	0.04	0.01	0.03	0.03
	(0.04)	(0.06)	(0.08)	(0.10)
ln(ProductivityOrigin)	-0.09*	-0.00	0.07	0.10
	(0.05)	(0.07)	(0.12)	(0.14)
ln(ProductivityDestin)	-0.06	0.10	0.21**	0.41***
	(0.05)	(0.07)	(0.10)	(0.12)
ln(PopDens)	0.02	-0.01	0.05	-0.01
	(0.02)	(0.03)	(0.04)	(0.05)
InstiDistance (Dummy)	-0.79***	1.54***		
	(0.07)	(0.10)		
ln(TECH_DIST)	-0.16***	0.06	-0.06	0.19**
	(0.03)	(0.04)	(0.06)	(0.08)
ln(IndustrialDistance)	-0.04*	0.05	-0.05	0.03
	(0.02)	(0.03)	(0.05)	(0.06)
ln(SocialProximity)	0.25***	-0.86***	0.90***	0.03
	(0.04)	(0.22)	(0.27)	(0.29)
Intercept	-7.47***	0.48	-8.19***	2.24
	(0.34)	(3.20)	(0.77)	(3.58)
Country Origin Fixed Effects	*yes*	*yes*	*yes*	*yes*
Country Destination Fixed Effects	*yes*	*yes*	*yes*	*yes*
Sample size	1,163,162	1,163,162	909,252	909,252
Nonzero onservations	12591		4,537	
Log-pseudolikelihood	-57285.81		-23166.98	
Overdispersion $(\ln\alpha)$	-0.26**		0.27***	
	(0.11)		(0.07)	
Voung statistic	15.32		7.31	
p-value	0.000		0.000	
AIC	104956.1		44511.25	
Schwartz	106137.7		45601.25	

Notes: Robust standard errors are presented in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. The model 5 is now estimated in both columns (i) and (ii). Each of the columns includes the first stage of the ZINB, the logit model, and the negative binomial estimation. Overdispersion tests largely reject the null hypothesis of no overdispersion. When the dispersion parameter, α, is zero, the negative binomial model reduces to the Poisson model. Therefore a likelihood ratio test on α can be computed, where $H_0 : \alpha = 0$. Vuong statistics (Vuong, 1989), are also performed and reported at the bottom of each regression. In a nutshell, the Vuong test examines if significant evidence for excessive zeros exist and therefore whether zero-inflated models should be used. The tests performed point to the need of the zero-inflated models to accommodate our estimations to the excess of zeros.