European Regional Science Association Conference
August-September, 2011, Barcelona

SPATIAL MOBILITY AND LOCATION CHOICES OF HIGHLY SKILLED WORKERS[1]

Very preliminary version
Do not quote without permission from the authors

First draft: February 2010
This draft: March 2011

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
The aim of the present paper is to shed light on the determinants of geographical mobility of skilled individuals across the European regions. The most talented workers, e.g. inventors, move for a number of reasons, contributing in this manner to the geographical diffusion of knowledge as well as to reshape the geography of talent. Thus, geographic areas constitute nodes through which talent circulate, bringing knowledge from one place to another. By means of a gravity model, we will test whether social proximity between inventors' communities and the so-called National System of Innovation drive in- and out-flows of inventors between pairs of regions, above and beyond physical separation, as well as other pulling factors (amenities, economic conditions, and the like). 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, still, geographical proximity in driving this phenomenon. However, social relationships, as well as institutional, or technological and cultural proximities, are also playing a preponderant role in mediating the mobility patterns of inventors across the European geography.

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

---

[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 (Milan, 17th March, 2010), the AQR Lunch Seminar, at the University of Barcelona (Barcelona, 12th May 2010), the XIII Encuentro de Economía Aplicada, (Sevilla, 11th June, 2010), the Zvi Griliches Summer School on the Economics of Innovation (Barcelona, 12th – 14th July, 2010), the 50th Annual Meeting of the Western Regional Science Association (Monterey, California, 2nd March 2011), Camilla Lenzi, Francesco Lissoni and Jouke van Dijk. 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.
† **Presenting author**

JEL: C8, J61, O31, O33, R0

1. Introduction

Geographical mobility of skilled[2] workers has become a central subject matter in empirical economics in recent years, attracting the attention of both academics and policymakers (Trippl, 2009; European Commission, 2000). Indeed, policymakers have convincingly embraced this affair, 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 put forward a number of suggestions and considerations for debate aimed to 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 present paper focuses precisely on the analysis of this phenomenon as measured by regional mobility of inventors across European regions.

The importance of this phenomenon from different perspectives motivates its analysis. First, highly skilled personnel's mobility across firms and in space matters for the transmission of knowledge. This claim has become an aphorism in recent years in innovation economics, regional economics, and so forth (Döring and Schnellenbach, 2006). Indeed, "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" (Breschi et al., 2009). Thus, together with inter-regional input-output linkages, FDI, and research networks, movements of skilled workers across regions act as an important channel through which inter-regional linkages are set up and knowledge is transferred (Fratesi and Senn, 2009).

Like in Boschma et al. (2009), we are convinced that mobile skilled workers act as 'pipelines' to access external, global, and sometimes distant, sources of knowledge, maintain a constant influx of new ideas, avoid regional 'lock-in' and elude regional 'entropic death' (Camagni, 1991; Williams et al., 2004; Millard, 2005; Bathelt et al., 2004).

Likewise, geographical mobility of skilled workers constantly changes the geography of talent, and determines the agglomeration of talented individuals and human capital.[3] As it is well known from a theoretical as well as from an empirical viewpoint, the accumulation of talent and creative people

[2] The use of the term mobility in this paper is intentional and is preferred to, for instance, migration. As stated in Williams et al. (2004), it is more apposite to refer to mobility when treating with knowledge workers. However, in this paper we might refer to migration as well, to which we give the same meaning as mobility, though we acknowledge the differences between the two.

[3] In the present paper we will use the term 'talent' or 'human capital' indistinctively and interchangeably. However, some studies have highlighted differences between the two (Mellander and Florida, 2007), linking the former to creative occupations and the later to educational attainment. Some other works, though, report high correlations between them (Glaeser, 2005). In any case, our study focuses only on inventors, irrespective of their occupation or their educational attainment, which we assume high in both cases.

influences regional development mainly due to the existence of human capital externalities arising from the fact that skilled workers tend to be more productive when surrounded by their peers (Lucas, 1988; Glaeser et al., 1995; Florida, 2002a, 2002b, 2005, 2006; Moretti, 2004). Indeed, talent and human capital is found to be highly concentrated in the space. This concentration may, however, evolve in distinct and particular ways depending upon certain conditions. Hence, as already asserted elsewhere, 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" (Storper and Scott, 2009, p. 148).

Broadly speaking, highly talented individuals are more geographically mobile than the rest of the population (5% versus 2% for the case of researchers -European Commission, 2000). In our case, for instance, 11.54% of the inventors with at least two patents report more than one NUTS2 region of residence (5.08% of all inventors).[4] However, either researchers or inventors "are still not as mobile as they could be in proportion to requirements" (Op. Cit., pp. 16). In consequence, geographical mobility, among other things, is seen to keep on being one of the main pillars of the ERA aimed at reinforcing the Lisbon process beyond 2010 (European Commission, 2010).

The focus on inventors as a proxy of talented individuals might be somehow questioned, since it may be argued that they are only a proportion of skilled labour. This choice requires an explanation before proceed. Thus, it is not less true that they constitute a significant part of the talented labour force, and they are more involved in the production of innovations and, consequently, the transfer of larger quantities of knowledge when they move (Breschi and Lenzi, 2010). This is also stressed in Lenzi (2010), who states that inventors are highly qualified workers, normally engaged in research activities in the firm for which they work, and so, they can be considered a good proxy for a larger group of researchers and, broadly speaking, knowledge workers. Equally, Florida's research agenda points at scientists and engineers as a critical part of the creative class super-core.[5] From a more pragmatic viewpoint, and as far as we are concerned, inventors are the only group of knowledge workers for which systematic data for a long time period and at fine geographical and sectoral desegregated levels can be gathered for the whole Western Europe.

[4] In order to make these figures comparable, results of other studies are as follows: 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.).

[5] According to Florida (2004: 8), the core of the creative class are those "whose economic function is to create ideas, new technology and/or new creative content (…) basically composed of occupations in science and engineering, architecture and design, education, arts, music and entertainment".

Overall, we strongly believe that the identification of regional features and differences impeding or favouring this mobility is an important issue from a policy viewpoint. The related literature is, however, scant. Still, we find imperative to investigate which are the regional structural features that attract and mobilize talent. In the present inquiry, however, we want to take a spatial-network approach (Bergman and Maier, 2009). To step in this direction, we do not consider regions as single entities floating in the space, but as nodes of different types of networks through which skilled individuals (and knowledge thereby) circulate. With this idea in mind, we will make use of a gravity model of immigration (applied to the subsample of knowledge workers) to test whether a set of relational[6] variables reflecting the position of regions within a set of different types of networks may favour talent mobility across Western European regions of 17 countries.

Our curiosities in the present inquiry are therefore manifold. Our starting point feeds from the migration literature and, among the factors determining individuals' mobility, we include distance between regions as a proxy for migration costs, for a number of reasons.[7] We also feed from more recent literature (Florida, 2002; Glaeser et al., 2001; Gottlieb and Joseph, 2006; Scott, 2010) and test the role of several pulling factors such as amenities, job opportunities, or regional economic conditions in attracting talent. We then move towards our main variables under scrutiny in this study. First, following recent contributions in regional and innovation economics (Boschma, 2005; Meyer, 2001; Ter Wal and Boschma, 2009) as well as the literature on labour economics (Nakajima et al., 2010), the role played by social networks of inventors across distant communities is going to be empirically assessed. Our hypothesis states that, all things being equal, we should observe disproportionate mobility between couplets of regions whose inventors maintain repeated professional relationships. On the other side, we are also interested in elucidating the influence of the so-called National System of Innovation –understood as institutional proximity/distance between regions- as favouring mobility within countries versus cross-country movements, above and beyond physical distance. Thus, we aim to test whether the concerns of the European Commission (2000, 2010) regarding mobility of skilled individuals to build the ERA –i.e., the unbalanced influence in favour of the national institutional setting in front of a European common context- actually hold. Furthermore, differences across time are also explored by estimating our models for two different time windows, i.e. 1996-1999 and 2002-2005. In short, the present paper's

[6] According to Scott (2000, pp. 2-3), "attribute data" are that data regarded as the properties, qualities or characteristics that belong to the individuals or, in general, to the unit of analysis considered. "Relational data" are the ties and connections which relates one unit of analysis to another and cannot be reduced to the properties of the individual agent under study. Relations are then not the properties of the unit, but of *systems of units*. Additionally, from SNA we learn that nodes are the different actors or points of the network, which are connected to one another by means of edges or ties.

[7] Descriptive figures seem to support this initial empirical setting since, on average, the distance covered by inventors' movements is around 397 kilometres –approximately the distance between Paris and Zurich, which could be considered low, and is around half of the distance found for the US case in a similar study (Breschi and Lenzi, 2010). Moreover, 30.79% of movements into the regions come from their 5 nearest neighbours, and 44.33%, from their 10 nearest ones. However, confirmatory analysis must be undertaken in order to confirm or reject this extreme.

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 knowledge workers across European regions, laying an especial emphasis on the role played by social and institutional proximities. To the best of our knowledge, no paper has addressed this question before, and it will be the main contribution of the study.

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 spatial mobility of inventors throughout the continent. However, institutional and social distances are also playing a significant role. These results are robust to the inclusion of other relational variables such as technological or cultural distances, among others.

The outline of the paper is as follows: section 2 reviews some relevant previous studies about regional talent endowments' differences, inventors' mobility and skilled labour migration, tying together dispersed but related literature. Section 3 describes the empirical model and the hypotheses we propose, whilst also presents the data and several estimation issues. Section 4 shows the results and section 5 presents conclusions and certain limitations of our approach.

2. Literature review and previous empirical findings

As already noted in the introductory section, spatial differences in human capital endowments have been widely investigated. Very well known examples are those by Florida and colleagues. Thus, for instance, Florida (2002a,b) and Mellander and Florida (2007) find significant correlations between regional human capital endowments -measured either by those with a bachelor degree or above, or technical and professional workers or scientists and engineers- and different types of regional features for, respectively, the US and Sweden. Such tested and confirmed critical 'attractive' characteristics are those like social tolerance, diversity, coolness indexes, and consumer amenities and, in broad terms, lifestyle -including entertainment, nightlife, culture, and so on. On their side, Glaeser et al. (2001) convincingly argue that amenities are critical determinants of the spatial distribution of human capital, whilst Shapiro (2006) stresses that around 40% of the employment growth effect of college graduates is due to growth in quality of life. In our view, however, these approaches are not fully satisfactory to analyse talent mobility for two main reasons. First of all, because they are eminently static, i.e., they devote their attention to analyse stocks of talent, whilst the dynamic analysis in terms of talent flows is not explicitly considered (at least in part of these studies). Second of all, because they hardly ever differentiate between talent created within a given location, and talent attracted from abroad.

Of course, both the migration (Borjas, 2000; Lewer and Van der Berg, 2008) and the economic geography literatures (Tabuchi and Thisse, 2002; Crozet, 2004; Sanchis-Guarner and López-Bazo, 2006; Clark et al., 2007) have analysed cross-regional mobility of labour through the estimation of migration equations. Undoubtedly, our approach feeds from this literature, insofar as we are also studying migration movements of individuals across locations. Additionally, part of these studies has also adopted a gravity equation to estimate migration flows across countries, regions, or cities. However, the focus on inventors in the present inquiry put some ground between these approaches and ours, since these papers do not take into account specific particularities of knowledge workers which we find essential.[8]

According to Ackers' (2005) concerns, despite evidence of significant imbalances in the geography of skilled migrations flows, little attention has been paid to study this phenomenon. To the best of our knowledge, studies about inventors' -i.e. those applying for patents- spatial mobility are equally very scarce. A descriptive approach is taken in Breschi and Lenzi (2010) and Miguélez et al. (2010) to analyse the mobility patterns of inventors across, respectively, US and European regions, with a particular emphasis on the role of physical separation. However, geography may seem to be more important in Europe than in the US, which could be attributed to the lower tendency of European inventors to relocate far away in the space, the lower institutional incentives to cross national borders (in terms of career promotion or the portability of social security provisions), as well as cultural and language differences that do not exist within the US.

Despite these later studies, systematic evidence about the determinants of the geographical mobility of skilled individuals, especially to what refers to inventors, is absent in the major part of the related literature. Further, the lack of studies about the regional features mobilizing talent is even more severe. Few exceptions are those by Faggian and McCann (2006,2009), who analyze, by means of structural equation models, what influences regional human capital inflows –in the form of recent university graduates- across British regions. Their findings suggest that inflows of highly mobile graduates are influenced by the presence of universities as well as the quality of these universities, which act as catalyst to enhance regional patent production –while variables such as wages, quality of life, and job opportunities are found to be insignificant. More recently, Venhorst et al. (2010, 2011) investigate the spatial mobility of graduates across Dutch regions. According to their findings, there exist substantial net flows towards the economic centre of the Netherlands, and that the availability of large labour markets is a key factor in the location decision of Dutch graduates.

[8] A typical example would be the influence of wages. While the income gap is considered pivotal in the migration literature (Borjas, 2000; Ortega and Peri, 2009), recent empirical evidence for the case of skilled workers shows the non-significance of salaries and wages in explaining their spatial mobility (Faggian and McCann, 2006, 2009; Scott, 2010), though more research on this point is needed.

Gottlieb and Joseph (2006) study also the college-to-work migration patterns of US graduates (including PhD holders). They found small evidence for amenities as spatial mobility drivers of US graduates, whilst employment opportunities seem to play a stronger role. This later study also tests the role of physical distance between the sending and receiving regions, finding a significant and negative effect on graduates' mobility patterns. Its effect is, though, smaller for PhD holders than for other graduates, as it is pointed out in the literature (Swartchz, 1973). Finally, Scott (2010) analyzes what drives inflows of migrant US engineers into different MSAs for 13 different technological categories. From his analysis we learn that local employment opportunities have a dominant impact on the destination choices of these skilled individuals, far above most of the amenities variables considered, or even the wage levels. Our approach is closely related to this later set of works. However, as stressed before, we prefer a spatial-network perspective to test which relational variables (proximities and distances between regions) mobilize talent.

To sum up, few points arise from this review. First and foremost, mobility of talented individuals is critical for knowledge flows and the spatial configuration of human capital. Second, the study of features and factors attracting, retaining and, in general, mobilizing talent, is a pivotal concern both for academics and for policymakers. However, we notice a drawback on our specific understanding about what influences cross-regional mobility of talent from a regional perspective. In the present study we will try to fill in this gap.

3. Research design

3.1. Empirical specifications

Baseline model

In order to meet our goals, the migration literature will be our point of departure. By estimating a gravitational equation, we will be able to discriminate which relational variables between pairs of regions determine the mobility patterns of inventors across the European geography. Thus, we assume that the mobility of an inventor from location i to location j is a function of the characteristics of both location i and j (pushing and pulling regional factors) and the costs of migration, which is proxied, for a number of reasons, by the physical separation between i and j.

Physical separation is aimed to capture a series of distance related phenomena which are difficult to measure (transport costs, communication costs, cultural distance, and the like). It might also capture other non-observable attributes, such as individuals' risk aversion or different kind of sunk costs (Breschi and Lissoni, 2001), like information costs related to destination-region housing markets. Other authors (McCann et al., 2010) have also stressed that large levels of *relational capital* may prevent individuals to re-locate distant from their home location. Thus, physical separation from family and friends is, still, a substantial cost for international and inter-regional migration. For the specific case of knowledge workers, additional considerations may come about. These are related to the flourishing and maintenance of fruitful professional relationships. Skilled individuals may tend to relocate closely in the space in order to minimize mobility costs if face-to-face interactions and frequent meetings with colleagues and competitors are needed, for instance, in universities or rival firms for information and help.

Our data seem to support, certainly, the former ideas. An inspection of several figures reveals that, on average, the distance covered by inventors' movements reported between 2002 and 2005 is around 397 kilometres –approximately the driving distance between Paris and Luxembourg, which could be considered low, and is around half the distance found for the US case in another study (Breschi and Lenzi, 2010). In fact, the average distance between regions' centroids in our sample is around 3.8 times larger. Furthermore, 30.79% of movements into the regions come from their 5 nearest neighbours, and 44.33%, from their 10 nearest ones. To the extent that economic and innovation activities in Europe are strongly concentrated, we wonder whether these data are truly reflecting the importance of geography to explain inventors' mobility or whether they are simply the result of the spatial distribution of innovation and innovators. The proposed empirical strategy in the present paper will try to disentangle this issue.

Critically, one is tempted to support the thesis that skilled individuals' mobility across the space is marginally influenced by physical separation when the spatial distribution of innovation activity is controlled for. Straightforward arguments are those related to the 'death-of-distance' hypothesis, 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 decide to move (Schwartz, 1973); (iii) the process of global economic integration; and (iv) the reduction of the real costs of travel and the development of ICT technologies (McCann et al. 2010, p. 362; Cairncross, 1997; O'Brien, 1992). Indeed in Ackers (2005), commenting in this last point, is said that scientists location decisions are currently strongly influenced by the existence of cheap flights and the benefits of laptops in promoting more flexible approaches to work, enabling them to tolerate extended forms of commuting (temporary mobility). In Ackers and Gill (2008) is said that scientists' location decisions are mainly driven by the search of the best research facilities and less by geographical considerations. For Chompalov (2006), the labour market for scientists is much more internationalized and with larger mobility rates regarding the whole workforce. And finally, Dickson (2003) argues that talented individuals trained in one country or region may easily function in another location, almost more than in any other profession.

Given these former arguments, the first hypothesis to test would be the following:

H1. Geographical separation between regions influences spatial mobility of inventors only marginally.

The baseline model set up to test this first hypothesis is as follows:

$$y_{ij} = e^{\beta_0} (D_{ij})^{\beta_k} e^{\rho NC_{ij}} A_{ik}^{\gamma_{ik}} A_{jk}^{\gamma_{jk}} e^{\theta_{ik} d_{ik} + \theta_{jk} d_{jk}} \varepsilon_{ij} . \qquad (1)$$

where flows of inventors between pairs of regions i and j, y_{ij} , are explained by the geographical distance(s) between the two regions, $D_{ij} = f\left(GeoDIST_{ij}\right)$, a dummy controlling for non-contiguous regions, $e^{\alpha_1 NC_{ij}}$ [9], and e^{β_0} the constant term capturing the impact of all common factors affecting mobility.

$A_{ik}^{\gamma_{ik}} A_{jk}^{\gamma_{jk}} e^{\theta_{ik} d_{ik} + \theta_{jk} d_{jk}}$ are a number of continuous and dummy variables aimed to control (and test therein) for the spatial distribution of economic and innovation activities, as well as other pulling attributional effects of the destination regions. In particular, among the variables meant to control for the spatial distribution of the economic and innovation activities we include:

- Population (POP) in sending and receiving regions, proxing the spatial distribution of economic activity.
- The number of inventors (INV) in sending and receiving regions, proxing the spatial distribution of innovation and innovators.
- Origin and destination country-specific fixed effects.
- Origin and destination regional shares of patents for 7 technological sectors (SHARE.TECH), aimed to control for different propensities to apply for patents across technological branches.

Among the variables aimed to control for specific pulling features of the destination region we find:

- Spatial location variables:
    - o We measure how central is a region within Europe (CENTRAL) as the distance of each regions' centroid to Brussels
    - o We expect regions sharing a border with a foreign country to receive more inventors (BORDER).
- Employment driven variables:

[9] Given the size variability of the NUTS2 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 NUTS2. For this reason, all our regressions will include a dummy variable reflecting if two regions are contiguous or not –first order contiguity.

- We use the number of inventors in the receiving region (INV_d) as a proxy for the size of the host labour market for inventors, and therefore as a proxy for job opportunities.
- We also measure how R&D-friendly is a region (and may have better opportunities for inventors) by including the share of Human Resources in Science and Technology (according to both educational attainment and occupations) over active population (HRST_d).

- Amenities driven variables. We pretty much follow Scott (2010) in the definition of amenities, who in turn follows some of the most relevant literature on the topic. Note that important Florida-type variables (bohemian index, gay index, and the like) are not included due to data restrictions. We include four variables:
  - The annual average temperature (TEMP) has been widely used as a predictor of incoming flows of skilled people (Gottlieb and Joseph, 2006).
  - We expect regions with coast (COAST) to receive more talent since it might be understood as an important recreational amenity. It might also proxy for temperate weather during the whole year.
  - Population density (DENS) is also included, as it is done in other studies on the topic. Glaeser et al. (2001) argue that low density cities are strongly attractive to immigrants. One should expect then a negative influence of density on inventors' inflows. However, they also acknowledge that density has now less power as an immigration predictor than one or two decades ago. In fact, it could also be argued that dense, urban areas may have larger supply of producer and consumer amenities (Perugini and Signorelli, 2010), so a positive effect might be also observed.
  - Regional absolute population (POP) is included as well (Scott, 2010). Again, there is no a priory expectations about the sign of this variable. It has been argued that, like density, cultural amenities are more available in large metropolitan areas, but also greater job opportunities. Conversely, one would observe a negative influence if inventors had preferences for smaller, less polluted cities with lower crime rates.

In order to consider deviations from the theory, a stochastic version of the model will be estimated by introducing ε_{ij}, an error term assumed to be independent of the regressors.

Next step is to test whether more meaningful relational variables are better explanatory factors of spatial inventors' mobility. We believe that, all things being equal, other proximities between regions and inventors' communities may outperform physical distance and confer it a marginal or inexistent role. If these other more meaningful variables are not controlled for, we face the risk of biasing the geography parameter upward. Obviously, these other distances might be intrinsically spatially determined, which means that they might well overlap with physical distance itself (see Boschma and Ter Wal, 2007, for a discussion on that), but their separate effect must be investigated. If they are not included in the estimations, the geography coefficient might be overestimated. In particular, we focus our attention on the role played by social and professional relations between inventors' communities, as well as the institutional setting in which these communities are inserted –the so-called National System of Innovation- among other relational variables.

Social proximity and the National System of Innovation

It is already a well established fact from labour economics and the sociology of networks literature that social relationships constitute one of the most effective ways for successful recruitment (Meyer, 2001). Thus, a relationship between the employer and the future employee is set up through a third person known by both acting as the intermediary. This relationship is mutually beneficial because (1) this third person provides the employee with information about the job; (2) he guarantees the employer that the individual is suitable for the job; and, on top of this, (3) it improves the employer-employee match, allowing workers to self-select themselves for the most suitable firms (Nakajima et al., 2010).

Highly skilled mobility dynamics responds to the same logic (Meyer, 2001). Most positions are acquired via connections and, to some extent, knowledge workers make location decisions in the context of their professional relations and networks (Millard, 2005). Besides, as stated in Ter Wal and Boschma (2009) and Sorenson (2003), the probability that firms or research institutions connect to individuals in firms or institutions with which they maintain any kind of social connection is higher than to non-connected skilled workers. Additionally for the case of researchers, it is well known that the network of relations established by a scholar determines his/her PhD students' mobility choices (Millard, 2005; Williams et al., 2004). In short, networks serve individuals to get better information about job vacancies (Lenzi, 2010), but also to find out about entrepreneurship opportunities elsewhere. To the extent that social networks are not necessarily spatially mediated (Boschma, 2005), professional relationships between inventors may well cross regional boundaries. We would observe then social relations between individual inventors located in different regions. In this study we state that if two nodes (regions) of the network create an outstanding number of professional relations in the form of research collaborations, one would observe larger amounts of talent interexchange between them. With this background in mind, the second hypothesis to test will be as follows:

H2. Social and professional relationships between distant inventors' communities enhances spatial mobility of inventors

As already stated in the introductory section, one of the main concerns of the European Commission towards the construction of the ERA is related to the low levels of transnational mobility of skilled workers between EU countries. According to the European Commission (2007), the fragmentation of R&D systems, policies and programmes between countries remain a characteristic of the European research system, at "a huge cost to Europeans as taxpayers, consumers, and citizens" (Op. Cit.). Indeed, it is an extended clamour among researchers that their career opportunities and cross-country mobility choices prevail limited by legal and practical barriers. As a rule of thumb, most academic positions remain largely reserved for national staff, for instance, hampering talent mobility across different institutional settings. These claims apply for the whole inventors' population.

Overall, it is argued that the National System of Innovation still remains being the reference institutional framework for knowledge workers (European Commission, 2006) and the main reference for major research activities (European Commission, 2000), far above the role played by European research institutions and the ERA. Indeed in Trippl (2009) is said that the US is characterized by a homogeneous institutional set-up and a common research area, whilst European countries strongly differ in terms of systems of innovation, making skilled workers mobility across different institutional frameworks the exception rather than the rule.

In the present inquiry, we empirically test whether two regions belonging to two different institutional systems, or two different countries, negatively affects the probability to observe a move between a given pair of regions. If this was the case for the subsample of inventors within the knowledge workers, the Commission's concerns would be justified and policies aimed to smooth differences in institutional frameworks across European countries would be required in order to build the ERA. Thus, the following hypothesis is suggested:

H3. Institutional distance between regions hinders spatial mobility of inventors

Control variables

Additional control relational variables are considered in the estimation. In particular, in the present paper we consider

(i) *technological distance between regions*: An index of technological (di)similarity between pairs of regions is calculated to test to what extent technologically close regions inter-exchange more inventors than technologically distant regions (Maggioni and Uberti, 2009; Moreno et al., 2005). This may happen because, among other things, the language to communicate each other, transmit messages, and interact, may belong to epistemic communities of scientists and inventors, which may or may not share the same physical space or the same country. Therefore, the costs associated to move into a region technologically distant are higher than to move to a technologically closest region. A negative effect of technological distance on mobility is therefore expected.

(ii) *cultural similarity*: it is also reasonable to think that inventors may chose to re-locate in regions sharing the same cultural and idiomatic background as his origin-region. This is actually closely related to the institutional distance variable. However, since similarities between languages are used to compute cultural similarity, we will allow this variable to have regions very close to each other belonging to different countries, and regions within the same country without the same value of the variable. A positive and significant impact is expected for this variable.

(iii) *membership to networks of research excellence*: we would also expect that regions doing research and innovation efforts over the mean might belong to networks of research excellence that inter-exchange more talented individuals. Scientists' location decisions are mainly driven by the search of the best research facilities and less by, for instance, geographical considerations (Ackers and Gill, 2008).

All in all, we now let D_{ij} be a function of a broader set of meaningful distances between pairs of regions,

$$D_{ij} = f\left(GeoDIST_{ij}, SocPROX_{ij}, InstiDIST_{ij}, Z_{ij}\right). \qquad (2)$$

where Z_{ij} includes the set of control relational variables aside from social proximity and institutional distance.

3.2. Estimation issues

A straightforward way to estimate (1) is to linearize by applying a logarithmic transformation in both sides of the equation. However, as showed in Santos Silva and Tenreyro (2006), in the presence of heteroskedasticity (which is likely to occur in gravitational frameworks), log-linearizing equation (1) and using least squares as estimation method would lead to inappropriate estimates because of the fact that $\ln \varepsilon_{ij}$ becomes not statistically independent of the regressors, leading to inconsistent estimates of the parameters of interest –see Santos Silva and Tenreyro (Op. Cit.) and

Siliverstovs and Schumacher (2009) for a proof on this. Put differently, what they basically say is that the estimation itself in a gravity model may induce a form of heteroskedasticity of the error term, because of the log transformation of the data. OLS would be inconsistent. Equally, it could be the case that no inventors' flows occur between a given pair of regions. This would make the logarithmic transformation of these observations impossible. Clearly, dropping these observations or adding an arbitrary constant to the dependent variable would lead again to inconsistent estimates -see Burger et al. (2009). To solve these pitfalls, Santos Silva and Tenreyro (2006), among others, suggest estimating the multiplicative form of the model by Poisson pseudo-maximum likelihood. To do so, we use the fact that the conditional expectation of y_{ij} in (1) can be written as the following exponential function

$$E(y_{ij} \mid x_{ij}) = \exp\left[\ln \beta_0 + \beta_k \ln(D_{ij}) + \rho NC_{ij} + \gamma_{ik} \ln A_{ik} + \gamma_{jk} \ln A_{jk} + \theta_{ik} d_{ik} + \theta_{jk} d_{jk}\right], \quad (3)$$

where $x_{ij} = \left(1, D_{ij}, NC_{ij}, d_{ik}, d_{jk}, A_{ik}, A_{jk}\right)$. Thus, count data class of models can be used to estimate (3), avoiding in this way the logarithmic transformation of (1), which would lead to inconsistent estimates and misleading interpretation of the results.

Additional advantages of this estimation technique in our specific framework are as follows: first, 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.). 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). Additionally, count models will generate estimates of y_{ij}, and not $\ln\left(y_{ij}\right)$. So by means of count data models, we avoid the underprediction of large migration flows (Burger et al., 2009).

The most basic type of count data model is derived from the Poisson distribution that assumes that the probability to observe a move from region i to region j follows a Poisson distribution

$$P\left[y \mid \mu\right] = \frac{\exp(-\mu)\mu^{y}}{y!}, \quad (4)$$

with a conditional mean (μ) of the distribution that is a function of the independent variables.

The maximum likelihood estimator would be reached by maximising

$$\ln L(\beta) = \sum_{y_{ij}} \left[y_{ij} x'_{ij} \beta - \exp(x'_{ij} \beta) - \ln y_{ij}! \right]. \tag{5}$$

However, the Poisson distribution assumes equidispersion, that is to say, the conditional variance and mean are the same, i.e., $E(y_{ij} \mid x_{ij}) = Var(y_{ij} \mid x_{ij}) = \mu_{ij} = \exp(x'_{ij} \beta)$ -where the corresponding subsacript, ij, is added to extend the framework to the regression case. But 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 would lead to consistent but inefficient estimates (Burger et al., 2009), with standard errors biased downward (Cameron and Trivedi, 1986; Long, 1997). Conversely, the negative binomial regression is preferred. In such a model, the expected value is the same as in the Poisson $E(y_{ij} \mid x_{ij}) = \exp(x'_{ij} \beta)$, but the variance is specified as a function of both the conditional mean and a dispersion parameter (α).

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 $H_0 : \alpha = 0$, to assess whether or not the negative binomial model is preferred to the Poisson estimation.

Another important point must be raised. Although count data models are explicitly designed to handle with the presence of zeros in the dependent variable, these zeros may come from different processes, which make necessary specific estimation techniques. Thus, we may have zero movements between a given pair of regions because a lack of innovation resources and inventors. In this case, we would not observe mobility by definition. Besides, we may have zero counts because, in spite of the presence of observed inventors in a given pair of regions, we do not observe movements in a certain period of time due to the characteristics (both relational and attributional features) of the given regions. The different processes at work and variety of sources of zero events are making our dependent variable extremely zero-inflated. Thus, again, even though count data models may deal with zero events, our dependent variable has greater frequency of them than would be predicted by the Poisson or Negative binomial models (Greene, 1994). As a consequence, the data generating process adds additional mass at the zero value, resulting in higher probability of it than is consistent with Poisson or Negative Binomial distributions. In such a setting, we would like to model separately the existence of zeros because a lack of resources (or observed resources), and the existence of zeros because of the characteristics of our observations. The later process can be perfectly modelled by means of count data models, negative binomial in our case. The first source of zeros must be modelled differently. Hence, the literature has suggested the use of zero-inflated models. In such zero-inflated models the population is formed by two groups (Mullhay, 1986). One individual is in the first group with probability φ, and he is in the second group with probability $1-\varphi$. Thus, the estimation process includes two parts: first is estimated the probability to observe a move, φ, by means of a probit or logit model, which is a function of certain characteristics –a set of covariates that predict the probability to belong to the strictly-zero group; and second, the count data model is estimated for the probability of each count for the group that has non-zero probability. There is, therefore, an equation for "participation" and a model for the event count that is conditional on the outcome of the "participation" equation.

Maximum likelihood techniques are used to obtain the estimated values for which the log-likelihood reaches its maximum values. Thus, the following log-likelihood function is going to be estimated:

$$\ln L(\beta,\alpha_{ij}) = \sum_{y_{ij}=0} \ln\left[\varphi + (1-\varphi)\left(\frac{\alpha_{ij}^{-1}}{\alpha_{ij}^{-1}+\exp(x'_{ij}\beta)}\right)^{\alpha^{-1}}\right] + \\ + \sum_{y_{ij}>0}\left[\begin{array}{l} \ln(1-\varphi) + \ln\left(\dfrac{\Gamma(y_{ij}+\alpha_{ij}^{-1})}{\Gamma(y_{ij}+1)\Gamma(\alpha_{ij}^{-1})}\right) + \alpha^{-1}\ln\left(\dfrac{\alpha_{ij}^{-1}}{\alpha_{ij}^{-1}+\exp(x'_{ij}\beta)}\right) \\ + y_{ij}\ln\left(\dfrac{\exp(x'_{ij}\beta)}{\alpha_{ij}^{-1}+\exp(x'_{ij}\beta)}\right) \end{array}\right]. \tag{6}$$

The Vuong (Vuong, 1989) statistic may be employed to assess whether the zero-inflated negative binomial is preferred above its non zero-inflated counterpart.

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.

3.3. Data and variable construction

We estimate our models for a sample of European NUTS2 regions of 17 countries[10] –see Appendix 1, in two time periods -1996-1999 and 2002-2005- in order to study differences in the point

[10] We have omitted the regions of Las Canarias, Ceuta, Melilla, Madeira, Açores, Guadeloupe, Martinique, Guyane and Reunion due to their distance from continental Europe.

estimates of our parameters of interest over time. The data are aggregated through 4-year time windows to avoid extreme heterogeneity. The explanatory variables are computed for the previous time spans (1992-1995 and 1998-2001 respectively) except for the distance variables, as well as contiguity, institutional distance, cultural similarity, and country-specific fixed effects. Doing so, we expect to lessen potential endogeneity biases caused by simultaneous causal relationships between the explanatory and the dependent variables. In the last section of the paper we discuss the suitability of this approach and possible alternative solutions. Our final sample is made up of 220 regions. Our dependent variable is built full-counting the movements of inventors crossing regional borders. We therefore construct a mobility asymmetric matrix of 220 rows and 220 columns for each time window, where each of the elements of the matrix is the number of inventors moving from region *i* to region *j*. 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 (220)x(220-1)=48,180 observations. Mobility is computed through observed changes in the reported region of residence by the inventor. Admittedly, in this manner we only capture mobility if the inventor applies for a patent before and after the move, probably underestimating real mobility. Another challenge is related to the time span in which we assign each movement. We compute it here in between the origin and the destination patent, but only if lasts 5 or less years between the two.

The data for constructing the mobility matrix are taken from the REGPAT database (OECD, January 2010 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 follow 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.[11]

[11] We are completely aware about the caveats of using patent data in economic analysis. Thus, for instance, it is well known that not all inventions are patented, they do not have the same economic impact, and not all the patented inventions are commercially exploitable innovations (Griliches, 1991). Additionally, it is also known that firms patent in a large extent for strategic motives, building up a patent portfolio in order to improve their position in negotiations or its technological reputation (Verspagen and Schoenmakers, 2004). Equally, the mobility matrix built reflects, to some extent, either the innovation capacity of regions, the degree of decentralisation of innovation activity in the different national states, or the different sectoral specialisations in regions which in turn determine the regional propensity to apply for patents (pharmaceuticals and biotech firms have a higher-than-average patent propensity). For example, the high degree of economic decentralisation and innovative activity in Germany results in a high degree of mobility across regions. Meanwhile, mobility in France will be more limited because only the regions of Paris or

In spite of the fact that we only consider two time spans, the identification process has been applied to the whole sample of EPO[12] patents with inventors reporting a European address. All in all, this process –see table 1 below- ends up in a list of 2,297,196 records -1,041,080 patents, belonging to a group of 768,810 identified unique inventors. On average, therefore, 2.99 patents (full counting) per inventors. These inventors are mostly concentrated in a reduced group of regions. Conversely, 25% of the regions have 10 or less inventors, whilst 50%, 30 or less. Equally, the Gini coefficient of their distribution across regions, 0.71, is relatively high. Note also that from these unevenly distributed inventors, only 11.54% are considered mobile inventors (they report more than one NUTS2 region of residence within our period). As for the specific case of our dependent variable, we have identified 26,178 movements (10,813 and 15,365 in the first and second time periods respectively), which are extremely concentrated from a geographical perspective as well. Thus, 5.5% of the regions do not receive any inventor at all during the period 2002-2005 (9.5% for the period 1996-1999), while 19.1% (25.5%) of them do receive only 6 or less inventors. On the contrary, it is important to see that around 50% (44.5%) of the inflows (inventors moving in a given region) are concentrated in only 20 regions.

[Insert Table 1 about here]

Geographical distance between regions' centroids (GeoDIST) is computed in four different manners, running therefore variants of the same model in order to study the stability of the coefficients: Euclidean distance, great circle distance, driving distances (in kilometres) and driving time (in seconds), both calculated using Google Maps.

Social proximity is proxied using EPO co-patents across NUTS2 regions. Thus, when one patent contains inventors reporting their addresses in different regions, we assume that there exist cross-

Rhône-Alpes are truly patenting regions. We believe, however, that the econometric analysis takes account of these spatial differences. Another source of bias is also related to the use of patents to identify mobility patterns of individuals. Thus, patenting activity may not include all the possible job-to-job changes of a given knowledge worker, since it could be the case that no patents are reported in some of the workplaces. Lenzi (2010) shows that, at least for a group of Italian inventors, patent data underestimate spatial mobility for this reason. We do not think that the underestimation of mobility is an important concern if no time or spatial significant differences are expected due to this fact. Lenzi (2010) also shows, however, a possible source of spatial mobility overestimation. In few specific cases, firms provide the EPO with applicants' addresses instead of inventors' addresses for strategic purposes. This should not be an important problem when the same address is provided for the whole labour career of a given inventor –in this case, we would not observe spatial mobility unless it indeed exists. In few cases, however, both the inventor's address and the applicant's address are provided to the EPO indistinctly, overestimating spatial mobility as a consequence. In general, however, inventors tend to live as close as possible to their workplace. It turns out from Lenzi's analysis that, if we look at NUTS2 regions, only 3% of the inventors report two different addresses in their set of patents not corresponding to a real NUTS2-region change for the case of her group of Italian inventors. Given this results, again, we think that this fact do not pose a serious bias in our estimations. See Ter Wall and Boschma (2009) for a discussion on additional shortcomings of using patents in regional analysis.

[12] European Patent Office.

regional collaborations. We 'full-count' all the collaborations across regions, irrespective of the number of inventors reported in each patent. We therefore obtain a socio-matrix reflecting the collaboration intensity between pairs of regions. We then adopt a measure suggested in Ejermo and Karlsson (2006) called 'affinity'. Thus, 'social affinity', A, between region i and j is the observed number of links between i and j, l_{ij}, minus all the links starting from i, n_i, over the total number of regions, k. Formally,

$$A_{ij} = l_{ij} - (n_i / k). \tag{7}$$

In reality, though, we choose to compute a variant of this formula

$$A_{ij} = l_{ij} / n_i. \tag{8}$$

in order to avoid negative values and allow for a logarithmic transformation of the variable.[13] 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).

Patent data from EPO to calculate technological distance are taken from the REGPAT database and assigned to each of the technological sectors using the IPC[14] classification.

To proxy technological distance, we use the following index:

$$TechDIST_{ij} = 1 - t_{ij}, \tag{9}$$

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

$$t_{ij} = \frac{\sum f_{ik} f_{jk}}{\left(\sum f_{ik}^2 \sum f_{jk}^2\right)^{1/2}}. \tag{10}$$

In (10), f_{ik} stands for the share of patents of one technological class k according to the IPC classification (out of 30 technological classes in the subdivision chosen) of the region i, and f_{jk} for

[13] A small constant has been added to all the explanatory variables with at least one 0 value for the same reason.
[14] International Patent Classification.

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 are technologically similar, and values close to the unity, that are technologically distant.

We calculate cultural similarity computing an index of language similarity across regions, as in Picci (2010). According to the author, it is reasonable to expect that people whose language share common roots will also share similar cultural backgrounds. To compute such index, we follow Picci (2010) and Fearon (2003). We gather data from the Ethnologue Project (www.ethnologue.com) in order to assign a single language to every NUTS2 region. We are aware about the fact that this criterion may mask the existence of multilingual regions in the continent. However, data on the number of people speaking one or the other language, or both, is lacking. Thus, we look at each country in the Ethnologue Project website and select only the languages under the heading "National or official languages". Using the Project's maps, we assign each of the languages under this heading to each NUTS2 of every country. Thus, for instance, Spanish is assigned to all NUTS2 regions of Spain, and French, to all NUTS2 regions of France. Conversely, up to six (very similar) languages are assigned to Dutch regions. We then compute the language similarity index. This index is based on the distance between branches of the linguistic classification of languages. As an example, the linguistic classification, from largest, more inclusive, grouping to smallest of, respectively, Portuguese, Swedish, and Danish is: Indo-European, Italic, Romance, Italo-Western, Western, Gallo-Iberian, Ibero-Romance, West Iberian, Portuguese-Galician (Portuguese); Indo-European, Germanic, North East, Scandinavian, Danish-Swedish, Swedish (Swedish); Indo-European, Germanic, North East, Scandinavian, Danish-Swedish, Danish-Riksmal, Danish (Danish). We sum the number of coincident branches between each pair of languages and divide the result over the sum of branches of each of the two languages (in order to take into account that the granularity of branches may not be the same across languages). As a result, we obtain an index between 0 and 1, where 0 means complete dissimilarity and 1 means that these two languages are almost the same (in linguistic terms). Thus, for instance, the similarity index between Spanish and Portuguese is 0.889, and between Swedish and Danish is 0.769. Meanwhile, the index between Portuguese and Danish is just 0.125.

Finally, membership to networks of research excellence is computed with a dummy variable valued 1 if the two regions show a level of Human Resources in Science and Technology (data taken from Eurostat) over total active population above the mean and 0 otherwise. Notice that we only compute here HRST under the heading CORE, which means that the individuals computed have both high level of scientific and technological educational attainment, as well as a technological or scientific occupations (different from the R&D –friendly measure explained before, which was either those with scientific educational background, or those with a technical or scientific occupation, or both).

A summary of the variables included, the proxies used, and the data sources can be found in the Appendix 2. Table 2 below also includes some descriptive statistics of the variables under consideration. Note again from Table 1 that the average distance covered by the computed movements increases by around 25 kilometres from the first to the second time period. This would point at the fact that distance is becoming less important over time to explain inventors' geographical mobility, though the econometric specification should shed some light on this issue. Note also that the average distance between pairs of regions, 1,524 km –see Table 2, is around 4 times larger than the average distance covered by the inventors' movements.

[Insert Table 2 about here]

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, different models for each of the proxies used for physical separation, and both for the first and the second time spans in separate and different tables. Both the negative binomial and the logit estimations are presented. For the NB regression, since the covariates are 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 distance between regions' centroids would lead to a decrease of 1.34% 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 inventors (INV) coefficient is -0.39, it means that an increase of 1% of the number of inventors in a given region leads to a decrease of 0.39% of the probability to belong to the "strictly zero group" (Maggioni and Uberti, 2009).

Results on the hypothesis on the marginal role of geography

Table 3 presents the estimation of model (1), including only distance as the focal explanatory variable as predictor of mobility between 1996 and 1999 –aside from regional controls. The estimated coefficients are negative and strongly significant irrespective of the proxy used. These coefficients, between -1.34 and -1.54, are 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). Actually, these coefficients are in line to what is found in the migration literature in similar frameworks at the European regional level (Crozet, 2004).

[Insert Table 3 about here]

Table 4 shows the same estimated model, but for the period 2002-2005. Broadly speaking, the results are maintained over time. Distance is lower in the second period for 3 out of 4 estimations, though a chi-squared test on individual coefficients does not reject the null that the differences are not statistically significant –see Table 5. This seems to be a little bit contradictory, since one should expect that physical separation should decrease its importance over time with the increasing usage of communication technologies, as we stated in the third section. Contrarily, these results would support the thesis that as the economy becomes more technologically advanced and specialised, the need for frequent human capital interactions and face-to-face meetings remains so as to conditioning the location choices of talented individuals. Overall, these preliminary findings suggest that the first hypothesis must be, a priori, rejected, since the influence of geographical distance between regions' centroids is far from being "marginal". Bear in mind, however, that the geographical coefficient might well be biased upward if other more meaningful distances are not controlled for. In consequence, we now move on to present our preferred specification.

[Insert Table 4 and Table 5 about here]

Results on the role of social affinity and institutional distance

Table 6 shows the estimation of the unrestricted model, where social proximity, institutional distance, as well as other relational control variables, are included. In there, results for the period 1996-1999 are shown in the first 2 columns and for the period 2002-2005 in the last 2 columns –we only provide results using kilometres and time as physical separation proxy, but results remain unchanged with other variables. From these columns we conclude some main findings. First and foremost, our focal variables in the present inquiry, i.e. institutional distance and social proximity, are significant and with the expected sign (respectively, negative and positive). Thus, hypotheses 2 and 3 are confirmed. These results are robust irrespective of the geographical distance proxy used and the time span. However, as can be seen, the importance of social proximity increases over time, whilst institutional distance decreases. Table 5 shows that these differences in point estimates are significant according to the chi-square tests performed for the case of institutional distance, but not for social 'affinity'. These findings support the idea that, even though physical distance does not decrease its importance over time, the research and innovation setting for inventors is becoming more and more internationally based.

Second of all, we learn from these tables that the role conferred to physical distance decreases considerably, by less than a half as before, confirming our suspicions about the important bias introduced when other more meaningful distances and proximities across regions are not controlled for. Admittedly, both geographical and the other distances may overlap, but each feature might has a different and independent effect on mobility that must be isolated correctly.

In sum, results reject the idea that the effect of geographical separation is unimportant. Indeed, the empirical exercise undertaken so far has conferred it a critical role in explaining inventors' spatial mobility. However, the two main variables under scrutiny in the present paper also show significant parameters and expected signs in explaining the phenomenon under analysis. Besides, technological distance, cultural proximity, or networks of excellence are also significant –note, though, that belonging to research networks of excellence is only significant in the second period.

[Insert Table 6 about here]

Results on the role of attributional variables: amenities versus job opportunities

An interesting subproduct of the present paper was to test the role played by several pulling attributional features of regions. We enter the vibrant debate on the importance of amenities versus job opportunities ('Do jobs follow people or people follow jobs?') by including several variables widely used in the literature in order to test whether their respective importance in attracting talent hold when we focus our attention to a specific group of knowledge workers. Density may seem to have a negative influence on attracting inventors across regions, in line with Glaeser's et al. (2001) arguments. However, its point estimates are not significant in the second period, in line again with the thesis that this variable is less important now than few years ago. Population in the destination region was also included to account for the supply of cultural amenities. We find large point estimates (and strongly significant at 1%) only in the first period, whilst lower coefficients (significant at 10%) in the second period. So the attractiveness of large metropolitan areas seems to be important, but decreasingly over time. Warmer climates do not seem to influence inventors' location decisions, whilst access to the sea is positively and significantly related to inflows of inventors (recreational amenities related to the sea, as well as more temperate climates throughout the year).

As regards the variables meant to control for destination-region employment opportunities, we find the size of the inventors' community in the destination region positively (and strongly) correlated with our dependent variable, irrespective of the time span and the estimated model. The general flavour of these results seems to indicate then that the variable proxing the labour market for inventors outperforms other pulling factors. The estimated coefficient, though, would be biased if the variable is not completely exogenous, which is quite unlikely, since inventors data are also mastered from patent data. Meanwhile, regional R&D efforts (HRST_d) are not significant in the first period, but it becomes in the second.

Robustness checks

In this section we summarize some robustness checks we performed in order to study the stability and significance of the estimated parameters, and the results encountered so far. In the interests of brevity, we omit here the tables, but they can be provided upon request from the authors.

First of all, we have repeated the estimation but including the income gap between origin and destination regions, following the literature on migration economics. Despite the fact that we could not use all the regions and time spans when introduced, the income gap does not turn to be significant in any of the estimations. This is consistent with previous findings regarding highly-skilled workers. Thus, several authors note the importance of career development (as opposed to financial gains) in explaining migration decision of the highly skilled (Mahroum, 2000; Meyer, 2001). These results are also in line with the findings by Scott (2010) for the US case, 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". The unemployment rate in origin and destination regions has been also included, but it does not turn out to be a significant predictor of mobility.

Second of all, given the strong significance of the first-order non-contiguity variable, we include second and third-order non-contiguity variables and re-estimate the models. 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 NUTS2 region. Fortunately, neither of the included variables turns out to be significant, and the parameters for the remaining variables remain virtually unchanged.

Additionally, one could think that the variables affecting the probability to belong to the strictly-zero group may not coincide with the variables which determine the number of movements from region i to region j. In our framework, it could be argued that the relational variables (distances between regions) should not enter as regressors in the logit estimation, whilst the remaining attributional variables should do in both processes. Again, we re-estimate the model in this direction and the main results remain unchanged.

Finally, we also repeat the estimations by including several time lags of the dependent variable. By doing so, our idea is to tease out the effect of the main variables under scrutiny on the spatial mobility of inventors while controlling at the same time for unobserved heterogeneity across pairs of regions -like unobserved historical linkages, common labour market institutions, and so forth. In a sense, we meant to control for the historical inertia of a given pair of regions to inter-exchange inventors, as if it was a region-couplet fixed effect. With this idea in mind, we have included in the '2002-2005' model the dependent variable lagged either one period (movements 1998-2001) or two periods (movements 1994-1997) or three periods (movements 1990-1993). The results of these estimations show that the negative effect of institutional and, especially, physical distance is notably reduced, while the social affinity coefficient remains unchanged. However, the three variables remain strongly significant. Conversely, the variables proxying cultural proximity and belonging to research networks turn out to be insignificant. Equally, some amenities variables decreases their point estimates and become insignificant as well. Bear in mind, however, that in the presence of serial correlation, the lagged dependent variable induces biases in all the other variables, which would depend on the level of serial correlation and the time elapsed between the lagged variable and the dependent variable we want to explain. Thus, in the absence of further analysis, the interpretation of these estimations should be taken with extreme care.

5. Conclusions, implications, and limitations

Throughout the previous sections, we have tried to disentangle the effect of some pivotal regional features that influence the spatial mobility of skilled workers, i.e. inventors, across the European geography. More specifically, the present paper focuses on the role played by social and professional relationships between distant communities of inventors, as well as the insertion in common institutional settings, as drivers of this mobility. Our approach feeds from social network analysis and considers regions as not being single entities floating in the space, but as nodes of different networks through which talented individuals circulate.

In the advent of the knowledge based economy, the identification of territorial features favouring and hampering the attraction of talent is of the upmost importance. Indeed, if talented individuals move across spatial locations, knowledge will diffuse in the space. Attracting knowledge workers is meant to help having access to distant sources of knowledge through these mobile skilled people, who constitutes 'pipelines' to this distant knowledge, which is mastered and diffused locally through the local 'buzz' once it enters the region. Furthermore, it is an extended wisdom that the spatial agglomeration of human capital may also influence regional growth rates differentials. Consequently, the forces that influence the movements of people must be an important matter of concern. Hence, we find empirical exercises like the present one of critical importance. However, previous evidence is unfortunately scant. 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 NUTS2 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 reject, by and large, that physical separation do not matter at all in explaining mobility patterns of these particular skilled workers. In reality, we were expecting that due to the reasons sketched in the third section, geography should not play any pivotal role in explaining the phenomenon under study. However, far from the announcements of "the death of distance" (Cairncross, 1997) or "the end of geography" (O'Brien, 1992), physical distance plays a preponderant role in mediating inventors' mobility across regions. These results are robust to the sample choice, specification, and inclusion of controls.

It is not less true that other more meaningful distances are also significant predictors of inventors' mobility patterns, such as social/professional connections, the institutional framework, or technological and cultural similarities. However, these measures do not succeed in explaining the role of physical distance away.

We also obtained some results concerning the role of amenities and job opportunities as talent attractors, contributing to the current debate on the topic. Admittedly, our results seems to better support the influence of job opportunities rather than amenities' supply (or economic conditions), especially for the later period, though there seems to exist some room for the amenities explanation as well. We acknowledge, though, that deeper research on this point must be undertaken.

From the present exercise, two main policy implications arise. In the path towards the ERA, this paper confirms that the fragmented institutional framework among countries impede frictionless mobility across national borders. Thus, in spite of latest progresses –important and significant differences in parameters estimates between the first and second period arise, much work remains to be done to overcome this fragmentation, which remains a prevailing characteristic of the European public research base. As asserted elsewhere, mobility across borders tends to be penalised rather than rewarded (European Commission, 2007). To step in the correct direction, policies aimed at making more transparent recruitment procedures, improve the portability of social security provisions across countries, and smooth differences in taxation, must be accomplished sooner than later. A second policy implication concerns the role played by social connectedness between inventors' communities. As it has been shown elsewhere, strengthening relationships between skilled individuals located far apart is a way to diffuse knowledge and information by itself, but also to improve mobility of talent between distant communities of skilled workers, reinforcing in this way the inter-exchange of information and ideas. In this sense, specific policies in line with the Framework Programmes funded by the European Commission should be strengthened and expanded.

In the present version of the paper we have estimated cross-section regressions using lagged r.h.s. variables. Lagging variables of the r.h.s. of the models attempts to lessen endogeneity and reverse-causality problems. 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 use of patents may also lead to a selection bias when including variables such as social proximity and the number of inventors. Put simply, the probability to observe a move or to collaborate with other inventors is higher for more productive individuals –having more patents. We would observe a positive and significant association between the two even if no causal relationship exists. To deal with such an issue, we need to find suitable instruments that must satisfy two conditions: (1) they must be uncorrelated with the unobservable time-varying error term; and (2) they must be sufficiently correlated with the endogenous variables that we want to instrument. If this by no means trivial task is accomplished, a control function approach -2 stages residual inclusion estimation (Terza et al., 2008; Wooldridge, 2002)- could be applied. First results in this direction using 20-year time lags as instruments of these explanatory variables seem to show that a downward bias in the 'social' coefficient exists, whilst the other two variables remain unchanged.

References

Ackers, L. (2005) "Promoting scientific mobility and balanced growth in the European research Area", *Innovation: The European Journal of Social Science research*, 18(3);

Ackers, L.; Gill, B. (2008) Moving People and Knowledge: Scientific Mobility in an Enlarging European Union. Cheltenham, UK: Edward Elgar Publishing Limited.

Bathelt, H.; Malberg, A.; Maskell, P. (2004): "Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation" *Progress in Human Geography* 28(1), 31-56;

Bergman, E.; Maier, G. (2009) "Network central: regional positioning for innovative advantage", *Annals of Regional Science*, 43: 615-644;

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

Boschma R. (2005) Proximity and Innovation: A Critical Assessment, Regional Studies, 39.1, 61-74;

Boschma, R.A. & A.L.J. ter Wal (2007) "Knowledge networks and innovative performance in an industrial district: the case of a footwear district in the South of Italy", *Industry and Innovation* 14 (2), pp. 177-199

Boschma R., Eriksson R. and Lindgren U. (2009) How does labour mobility affect the performance of plants? The importance of relatedness and geographical proximity, Journal of Economic Geography 9, 169-90;

Breschi S. and Lissoni F. (2001) Knowledge spillovers and local innovation systems: A critical survey, Industrial and Corporate Change 10, 4, 975-1005;

Breschi S., and Lenzi C. (2010) "Spatial patterns of inventors' mobility: Evidence on US urban areas" *Papers in Regional Science,* 89(2)

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. (ed.) (1991) *Innovation Networks: Spatial Perspectives*. London: Belhaven

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.

Chompalov, I. (2006) "Birds of Passage: Brain drain from Bulgaria before and after the transition to democracy", *Sociological viewpoints*

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;

Dickson, D. (2003) "Mitigating the brain drain is a moral necessity", paper presented to the Science and Development Network, London, May.

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;

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

European Commission (2007a) "Annex to the Green Paper. A European Strategy for Sustainable, Competitive and Secure Energy" SEC(2006) 317/2

European Commission (2007b) "Green Paper. The European research area: new perspectives" COM(2007) 161

Faggian A, McCann P (2006) Human capital flows and regional knowledge assets: A simultaneous equation approach, *Oxford Economic* Papers, 58(3): 475-500;

Faggian A, McCann P (2009) Human capital, graduate migration and innovation in British Regions. *Cambridge Journal of Economics* 33: 317–333

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

Florida, R (2004) Response to Edward Glaeser's review of *The Rise of the Creative Class*, manuscript

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

Florida R (2002b) The economic geography of talent. *Annals of the Association of American Geographers* 92: 743–755

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 (2005) Review of Richard Florida's 'The rise of the creative class'. *Regional Science and Urban Economics* 35: 593–596

Glaeser E. L., Scheinkman, J.A. and Shleifer, A. (1995) "Economic growth in a cross-section of cities," Journal of Monetary Economics, 36(1), pages 117-143;

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

Gottlieb, P. D.; Joseph, G. (2006) "College-to-work migration of technology graduates and holders of doctorates within the United States", *Journal of Regional Science* 46: 627-659

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;

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;

Lenzi C. (2010) "technology mobility and job mobility: On the use of patent data for inventors' career analysis", unpublished manuscript;

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.

Lucas R. E. (1988) On the mechanics of economic development, Journal of Monetary Economics 22, 3-42;

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

McCann, P.; Poot, J.; Sanderson, L. (2010) "Migration, relationship capital and international travel: theory and evidence" *Journal of Economic Geography*, 10, pp. 361-387;

Mellander C, Florida R (2007) The creative class or human capital? Explaining regional development in Sweden. *KTH/CESIS*, Electronic Working Paper Series in Economics and Institutions of Innovation Paper No. 79

Meyer, J (2001) "Network Approach versus Brain Drain: Lessons from the Diaspora" *International Migration* Vol. 39 (5)

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;

Miguelez, E.; Moreno, R.; Suriñach, J. (2010) "Inventors on the move: Tracing inventors' mobility and its spatial distribution" *Papers in Regional Science,* 89(2)

Millard, D. (2005) 'The impact of clustering on scientific mobility', Innovation: The European Journal of Social Science Research, 18: 3, 343 — 359

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

Moretti, E. (2004) "Human capital externalities in cities," Handbook of Regional and Urban Economics, in: J. V. Henderson & J. F. Thisse (ed.), Handbook of Regional and Urban Economics, edition 1, volume 4, chapter 51, pages 2243-2291 Elsevier

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

Nakajima R.; Tamura, R.; Hanaki, N. (2010) "The effect of collaboration network on inventors' job match, productivity and tenure", *Labour Economics*, 17

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

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

Perugini, C.; Signorelli, M. (2010) "Youth labour market performance in European regions" Econ Change Restruct 43:151–185;

Ortega, F.; Peri, G. (2009) "The causes and effects of international labor mobility: Evidence from OECD countries 1980-2005" MPRA Paper No. 19183

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;

Sanchis-Guarner R. and Lopez-Bazo E. (2006) "Are Skilled Workers More Attracted to Economic Agglomerations?" ERSA conference papers

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

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

Scott, J (2000) *Social Network Analysis: A Handbook*. SAGE, London

Scott, A (2010) "Jobs or amenities? Destination choices of migrant engineers in the USA" *Papers in Regional Science*, Volume 89 Number 1

Shapiro J.M. (2006) "Smart cities: quality of life, productivity, and the growth effects of human capital" NBER, 11615

Siliverstovs, B.; Schumacher, D. (2009) "Estimating gravity equations: to log or not to log?", *Empirical Economics* 36(3), pages 645-669;

Storper, M. and Scott.A. (2009) "Rethinking human capital, creativity and urban growth" Journal of Economic Geography, pp. 1–21

Storper, M. and Venables, A. (2004) "Buzz: face-to-face contact and the urban economy" Journal of Economic Geography, 4

Sorenson, O. (2003) "Social networks and industrial geography" *Journal of Evolutionary Economics*, 13(5): 513-527;

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

Ter Wal A. L. J. and Boschma R. (2009) Applying social network analysis in economic geography: framing some key analytic issues, The Annals of Regional Science 43, 739-56;

Terza, J. V.; Basu, A.; Rathouz, P. J. (2008) "Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling", Journal of Health Economics 27(3): 531-543;

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;

Venhorst, V.; Van Dijk, J.; Van Wissen, L. (2010) "Do the best graduates leave the peripheral areas of the Netherlands?" Tijdschrift voor economische en sociale geografie, 101(5): 521–537;

Venhorst, V.; Van Dijk, J.; Van Wissen, L. (2011) "An analysis of trends in spatial mobility of Dutch higher education graduates" forthcoming in *Spatial Economic Analysis*;

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

Williams, A.M.; Balaz, V.; Wallace, C. (2004) International Labour Mobility and Uneven Regional Development in Europe: Human Capital, Knowledge and Entrepreneurship, *European Urban and Regional Studies* (11) 27

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	Expected sign
Inventors' flows	Counts of flows from home to host region	96-99 02-05	REGPAT and own calculations	
Geographic distance	Euclidean distance between UTM regional centroids		GIS	-
Geographic distance	Great circle distance		GIS	-
Geographic distance	Driving distance in km		Google Maps and SAS	-
Geographic distance	Driving distance in time		Google Maps and SAS	-
Contiguity	1: contiguity; 0 otherwise		GIS	-
Institutional distance	1: dif. country; 0 otherwise			-
Affinity	$A_{ij} = l_{ij} / n_i$	92-95 98-01	REGPAT and own calculations	+
Technological distance	$1 - \left(\frac{\sum f_{ik} f_{jk}}{\left(\sum f_{ik}^2 \sum f_{jk}^2 \right)^{1/2}} \right)$	Average 92-95 98-01	REGPAT and own calculations	-
Language similarity			Ethnologue Project	+
Excellence	1: share HRST (core) of active population over the mean in both regions; 0 otherwise	92-95 98-01	Eurostat	+
Inventors	# inventors in origin and destination regions	92-95 98-01	REGPAT and own calculations	+
Population	Population in origin and destination regions	Average 92-95 98-01	Eurostat	+
Border_d			ESPON	+
Time2Brussels_d	Seconds from the regions' centroids to Brussels		Google Maps and SAS	-
HRST_d	Human Resource in Science and Technology over active population	Average 92-95 98-01	Eurostat	+
Population Density_d	Population over area (km2)	Average 92-95 98-01	Eurostat	?
Annual average temperature_d		1994 2000	FOODSEC project, MARS units, EC-JRC (ISE)	+
Coast_d	1: if the region has a coast; 0 otherwise		ESPON	+

Table 1. Descriptive figures

Identified inventors(1975-2005)	768,810
Inventors' distribution across regions: Gini index	0.71
Movements 96-99	10,813
Movements 02-05	15,365
Total number of movements	26,178
Regions with 0 inflows	5.5% (9.5%)
Regions with 6 or less inflows	19.1% (25.5%)
Top 20 inflows' regions	50% (44.5%)
Movements from 5 nearest neighbours	30.79%
Movements from 10 nearest neighbours	44.33%
Movements from within national borders	76.18%
Mean distance covered by inventors' movements 1996-1999	
Euclidean	3.23º
Great circle	175.29
Km	374.68
Time	14,221.72
Mean distance covered by inventors' movements 2002-2005	
Euclidean	3.56º
Great circle	188.32
Km	397.46
Time	14,970.35

Table 2. Summary statistics

	Mean	St. Dev	Coef. Var.	Min.	Max.
Attributional variables					
BORDER_d	0.45	0.50	1.10	0	1
CENTRAL_d	640.32	475.46	0.74	10	2400
INV9295_o	648.25	1058.10	1.63	1	9140
INV9801_o	1040.30	1629.25	1.57	1	12766
INV9295_d	648.25	1058.10	1.63	1	9140
INV9801_d	1040.30	1629.25	1.57	1	12766
HRST9295_d	28.28	8.63	0.31	7.73	55.05
HRST9801_d	32.50	8.08	0.25	11.88	55.30
POP9295_o	1718268	1476858	0.86	25025	10800000
POP9801_o	1747665	1500628	0.86	25625	11000000
POP9295_d	1718268	1476858	0.86	25025	10800000
POP9801_d	1747665	1500628	0.86	25625	11000000
DENS9295_d	354.47	842.97	2.38	3.17	8163.25
DENS9801_d	359.07	857.72	2.39	3.14	8497.49
TEMP1994_d	10.87	3.43	0.32	-0.05	19.47
TEMP2000_d	11.84	3.10	0.26	1.67	20.17
COAST_d	0.54	0.50	0.93	0.00	1
Relational variables					
Euclidean distance	12.62	7.46	0.59	.06	44.60
Great circle distance	696.30	416.95	0.59	4.07	2416.55
Km	1524.76	910.27	0.59	8.06	5545
Time	57625.21	36297	0.62	1200	241200
Social affinity 1992-1995	0.00	0.03	6.62	0	1
Social affinity 1998-2001	0.00	0.03	5.98	0	1
Institutional distance	0.90	0.29	0.33	0	1
Cultural proximity	0.38	0.30	0.78	0	1
Tech. distance 1992-1995	0.56	0.23	0.41	0	1
Tech. distance 1998-2001	0.51	0.22	0.43	0	1
HRST core 1992-1995	0.26	0.44	1.70	0	1
HRST core 1998-2001	0.21	0.41	1.92	0	1

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

Table 3. Gravity model - inventors' mobility, 1996-1999. Dep. Var.: In- and Out-flows of inventors

	(i) euclidean		**(ii) great circle**		**(iii) km**		**(iv) time**	
	NegBin	Logit	NegBin	Logit	NegBin	Logit	NegBin	Logit
Intercept	-15.92***	4.61	-11.76***	0.65	-9.89***	1.06	-1.56	-2.68
	(2.02)	(6.50)	(2.05)	(5.08)	(2.01)	(5.35)	(1.89)	(5.48)
Contiguity	0.94***	-1.20**	0.87***	-1.07**				
	(0.09)	(0.49)	(0.09)	(0.42)				
ln(Euclidean Distance)	-1.34***	1.08***						
	(0.06)	(0.22)						
ln(Arc Distance)			-1.42***	1.05***				
			(0.06)	(0.17)				
ln(Km)					-1.45***	1.12***		
					(0.06)	(0.20)		
ln(Time)							-1.54***	1.46***
							(0.07)	(0.18)
ln(BORDER_d)	0.09	-0.17	0.07	-0.25	0.06	-0.27	0.06	-0.28*
	(0.07)	(0.26)	(0.06)	(0.17)	(0.06)	(0.17)	(0.06)	(0.17)
ln(CENTRAL_d)	-0.18	-0.75	-0.13	-1.02***	-0.10	-1.15***	-0.12	-1.32***
	(0.13)	(0.49)	(0.11)	(0.27)	(0.11)	(0.27)	(0.11)	(0.26)
ln(INV_o)	0.51***	-0.39**	0.52***	-0.39**	0.52***	-0.37**	0.49***	-0.34**
	(0.05)	(0.16)	(0.05)	(0.18)	(0.05)	(0.18)	(0.05)	(0.17)
ln(INV_d)	0.45***	-0.73***	0.46***	-0.81***	0.47***	-0.79***	0.45***	-0.82***
	(0.06)	(0.24)	(0.06)	(0.17)	(0.06)	(0.17)	(0.06)	(0.17)
ln(HRST_d)	0.34	0.26	0.31	0.40	0.30	0.38	0.39*	0.47
	(0.23)	(0.67)	(0.20)	(0.57)	(0.20)	(0.57)	(0.20)	(0.60)
ln(POP_o)	0.32***	-0.24	0.29***	-0.29	0.26***	-0.32	0.27***	-0.32
	(0.08)	(0.24)	(0.08)	(0.25)	(0.08)	(0.26)	(0.08)	(0.25)
ln(POP_d)	0.45***	0.28	0.43***	0.22	0.41***	0.16	0.41***	0.21
	(0.09)	(0.27)	(0.09)	(0.26)	(0.09)	(0.27)	(0.09)	(0.28)
ln(DENS_d)	-0.07*	0.04	-0.10***	-0.02	-0.13***	-0.03	-0.11***	0.01
	(0.04)	(0.14)	(0.04)	(0.12)	(0.04)	(0.12)	(0.04)	(0.12)
ln(TEMP_d)	-0.23	0.23	-0.22	0.11	-0.22	0.12	-0.32	0.15
	(0.25)	(0.58)	(0.23)	(0.51)	(0.23)	(0.51)	(0.25)	(0.52)

ln(COAST_d)	0.30***	-0.06	0.25***	-0.13	0.24***	-0.17	0.30***	-0.29
	(0.10)	(0.40)	(0.08)	(0.24)	(0.08)	(0.25)	(0.08)	(0.26)
ln(TECH.SHARES) (1)	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*
Country Origin/Destination Fixed Effects (2)	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*
Sample size	48,180	48,180	48,180	48,180	48,180	48,180	48,180	48,180
Nonzero observations	2,854	2,854	2,854	2,854	2,854	2,854	2,854	2,854
Log-pseudolikelihood	-10717.32		-10709.02		-10685.98		-10683.99	
LR test	11253.360		11269.961		11316.030		11320.009	
p-value	0.0000		0.0000		0.0000		0.0000	
Wald test	5384.85		5259.34		5144.14		5528.39	
p-value	0.0000		0.0000		0.0000		0.0000	
LR test of α	4544.88		4600.46		4565.45		4387.29	
p-value	0.0000		0.0000		0.0000		0.0000	
Voung statistic	12.42		12.54		12.57		12.27	
p-value	0.0000		0.0000		0.0000		0.0000	
McFadden's R2	0.344		0.345		0.346		0.346	
Adjusted McFadden's R2	0.337		0.337		0.339		0.339	
AIC	21672.64		21656.04		21609.97		21605.99	
Schwartz	22717.78		22701.18		22655.11		22651.13	

Notes: Robust standard errors are presented in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Each of the columns includes the negative binomial estimation and the first stage of the ZINB, the logit model. Overdispersion tests largely reject the null hypothesis of no overdispersion. Vuong statistics (Vuong, 1989), are also performed and reported at the bottom of each regression. The tests performed point to the need of the zero-inflated models to accommodate our estimations to the excess of zeros. '_o' and '_d' denotes origin-region and destination-region variables, respectively. **(1)** Inventors are assigned to each technological sectors according to the classification jointly elaborated by Fraunhofer Gesellschaft-ISI (Karlsruhe), Institut National de la Propriété Industrielle (INPI, Paris) and Observatoire des Sciences and des Techniques (OST, Paris). This classification aggregates all IPC codes into seven technology fields: 1. Electrical engineering; Electronics; 2. Instruments; 3. Chemicals; Materials; 4. Pharmaceuticals; Biotechnology; 5. Industrial processes; 6. Mechanical eng.; Machines; Transport; and 7. Consumer goods; Civil engineering. However, some patents may belong to different sectors (out of 7), and therefore also the inventors. In consequence, we first assign a main technological sector to each patent. In particular, we drop out all the sectors in each patent that do not represent more than the 35% of the total number of technological classes listed in the patent document. Few patents maintain doubled assignment of technological sector, though. We repeat the process inventor by inventor. Thus, we assign each inventor to each technological sector if at least he/she has 30% of the patents assigned to a given sector. Again, few inventors are doubled, because we are unable to categorically assign them to a unique sector. However, we do not expect this doubling to produce any bias in our estimation results. **(2)** Germany is treated as the reference country.

Table 4. Gravity model - inventors' mobility, 2002-2005. Dep. Var.: In- and Out-flows of inventors

	(i) euclidean		(ii) great circle		(iii) km		(iv) time	
	NegBin	Logit	NegBin	Logit	NegBin	Logit	NegBin	Logit
Intercept	-16.90***	3.68	-10.99***	0.17	-10.06***	-1.44	-3.32	-7.04*
	(1.79)	(3.21)	(2.21)	(3.42)	(2.68)	(4.10)	(2.29)	(3.74)
Contiguity	1.02***	-1.33***	0.88***	-1.32***	0.92***	-1.38***	1.00***	-1.38***
	(0.11)	(0.37)	(0.10)	(0.34)	(0.10)	(0.36)	(0.10)	(0.34)
ln(Euclidean Distance)	-1.30***	1.07***						
	(0.08)	(0.25)						
ln(Arc Distance)			-1.39***	1.05***				
			(0.06)	(0.16)				
ln(Km)					-1.43***	1.05***		
					(0.07)	(0.16)		
ln(Time)							-1.58***	1.21***
							(0.08)	(0.17)
ln(BORDER_d)	0.21***	0.37**	0.20***	0.34*	0.19***	0.32*	0.19***	0.38**
	(0.07)	(0.18)	(0.07)	(0.18)	(0.07)	(0.18)	(0.07)	(0.17)
ln(CENTRAL_d)	-0.07	-0.56**	-0.04	-0.65***	-0.01	-0.66***	0.01	-0.70***
	(0.14)	(0.26)	(0.13)	(0.25)	(0.13)	(0.25)	(0.13)	(0.25)
ln(INV_o)	0.72***	-0.60***	0.72***	-0.59***	0.70***	-0.58***	0.69***	-0.55***
	(0.04)	(0.07)	(0.04)	(0.07)	(0.04)	(0.08)	(0.04)	(0.07)
ln(INV_d)	0.67***	-0.57***	0.66***	-0.60***	0.66***	-0.60***	0.65***	-0.60***
	(0.05)	(0.09)	(0.05)	(0.10)	(0.05)	(0.10)	(0.05)	(0.11)
ln(HRST_d)	0.95**	-0.60	1.03**	-0.40	1.10**	-0.31	1.12***	-0.23
	(0.37)	(0.65)	(0.40)	(0.72)	(0.43)	(0.75)	(0.40)	(0.72)
ln(POP_o)	-0.01	-0.00	0.00	0.00	0.01	-0.00	0.01	0.01
	(0.04)	(0.09)	(0.04)	(0.09)	(0.04)	(0.09)	(0.03)	(0.09)
ln(POP_d)	0.03	-0.07	0.02	-0.08	0.03	-0.09	0.02	-0.09
	(0.03)	(0.07)	(0.03)	(0.07)	(0.03)	(0.07)	(0.03)	(0.07)
ln(DENS_d)	-0.01	0.09	-0.01	0.09	-0.04	0.09	-0.05	0.09
	(0.05)	(0.09)	(0.05)	(0.10)	(0.05)	(0.10)	(0.05)	(0.10)
ln(TEMP_d)	0.42	1.09**	0.18	0.66	0.23	0.89	0.21	0.96
	(0.32)	(0.53)	(0.36)	(0.66)	(0.38)	(0.76)	(0.36)	(0.66)

ln(COAST_d)	0.33***	-0.04	0.32***	-0.02	0.32***	-0.01	0.36***	0.03
	(0.09)	(0.18)	(0.10)	(0.21)	(0.10)	(0.24)	(0.10)	(0.23)
ln(TECH.SHARES) (1)	*yes*	*yes*	*yes*	*yes*	yes	yes	yes	yes
Country Origin/Destination Fixed Effects (2)	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*
Sample size	48,180	48,180	48,180	48,180	48,180	48,180	48,180	48,180
Nonzero observations	3,365	3,365	3,365	3,365	3,365	3,365	3,365	3,365
Log-pseudolikelihood	-12919.93		-12881.42		-12846		-12810.08	
LR test	12154.665		12231.673		12302.526		12374.360	
p-value	0.0000		0.000		0.0000		0.0000	
Wald test	4016.27		4537.18		4615.40		4613.13	
p-value	0.0000		0.0000		0.0000		0.0000	
LR test of α	1400.00		1400.00		1400.00		1400.00	
p-value	0.0000		0.0000		0.0000		0.0000	
Voung statistic	11.28		11.06		10.97		10.83	
p-value	0.0000		0.0000		0.0000		0.0000	
McFadden's R2	0.320		0.322		0.324		0.326	
Adjusted McFadden's R2	0.314		0.316		0.318		0.319	
AIC	26077.85		26000.85		25929.99		25858.16	
Schwartz	27122.99		27045.99		26975.13		26903.3	

Notes: Robust standard errors are presented in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Each of the columns includes the negative binomial estimation and the first stage of the ZINB, the logit model. Overdispersion tests largely reject the null hypothesis of no overdispersion. Vuong statistics (Vuong, 1989), are also performed and reported at the bottom of each regression. The tests performed point to the need of the zero-inflated models to accommodate our estimations to the excess of zeros. '_o' and '_d' denotes origin-region and destination-region variables, respectively. **(1)** Inventors are assigned to each technological sectors according to the classification jointly elaborated by Fraunhofer Gesellschaft-ISI (Karlsruhe), Institut National de la Propriété Industrielle (INPI, Paris) and Observatoire des Sciences and des Techniques (OST, Paris). This classification aggregates all IPC codes into seven technology fields: 1. Electrical engineering; Electronics; 2. Instruments; 3. Chemicals; Materials; 4. Pharmaceuticals; Biotechnology; 5. Industrial processes; 6. Mechanical eng.; Machines; Transport; and 7. Consumer goods; Civil engineering. However, some patents may belong to different sectors (out of 7), and therefore also the inventors. In consequence, we first assign a main technological sector to each patent. In particular, we drop out all the sectors in each patent that do not represent more than the 35% of the total number of technological classes listed in the patent document. Few patents maintain doubled assignment of technological sector, though. We repeat the process inventor by inventor. Thus, we assign each inventor to each technological sector if at least he/she has 30% of the patents assigned to a given sector. Again, few inventors are doubled, because we are unable to categorically assign them to a unique sector. However, we do not expect this doubling to produce any bias in our estimation results. **(2)** Germany is treated as the reference country.

Table 5. Test on differences in coefficients over time

	Coef. 1996-99	Coef. 2002-05	Chi-Sq.	p-value
First Models				
Euclidean distance	-1.34	-1.30	0.04	0.83
Great circle distance	-1.42	-1.39	0.63	0.43
Km	-1.45	-1.43	1.06	0.30
Time	-1.54	-1.58	2.01	0.16
Second Model with KM				
Km	-0.59	-0.62	0.32	0.57
Institutional distance	-0.65	-0.47	3.37*	0.07
Social Affinity	0.12	0.15	1.55	0.21
Second Model with TIME				
Time	-0.63	-0.68	1.12	0.29
Institutional distance	-0.65	-0.46	3.33*	0.07
Social Affinity	0.12	0.16	1.55	0.21

Table 6. Gravity model - inventors' mobility, 1996-1999 & 2002-2005. Dep. Var.: In- and Out-flows of inventors

	(iii) km 1996-1996		**(iv) time 1996-1996**		**(iii) km 2002-2005**		**(iv) time 2002-2005**	
	NegBin	Logit	NegBin	Logit	NegBin	Logit	NegBin	Logit
Intercept	-9.36***	-8.50*	-6.75***	-8.61*	-9.80***	-0.61	-6.90***	-0.67
	(1.73)	(4.52)	(1.72)	(4.55)	(1.93)	(4.07)	(2.05)	(4.16)
Contiguity	0.92***	-0.81**	0.98***	-0.84**	0.85***	-1.33***	0.90***	-1.32***
	(0.08)	(0.39)	(0.08)	(0.39)	(0.08)	(0.37)	(0.08)	(0.38)
ln(Km)	-0.59***	0.08			-0.62***	-0.03		
	(0.06)	(0.15)			(0.07)	(0.14)		
ln(Time)			-0.63***	0.07			-0.68***	0.00
			(0.07)	(0.17)			(0.08)	(0.15)
Institutional distance	-0.65***	4.93***	-0.65***	4.90***	-0.47***	4.60***	-0.46***	4.56***
	(0.11)	(0.53)	(0.11)	(0.54)	(0.10)	(0.40)	(0.10)	(0.41)
ln(Social affinity)	0.12***	-0.04	0.12***	-0.04	0.15***	0.03	0.16***	0.03
	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
ln(Technological Distance)	-0.15**	0.51**	-0.16**	0.51**	-0.15**	0.42***	-0.16***	0.40***
	(0.07)	(0.20)	(0.07)	(0.20)	(0.06)	(0.14)	(0.06)	(0.14)
ln(Cultural Distance)	0.05**	-0.23**	0.04**	-0.25**	0.05*	-0.46***	0.04	-0.47***
	(0.02)	(0.10)	(0.02)	(0.10)	(0.03)	(0.09)	(0.03)	(0.09)
Research Excellence	-0.02	0.05	-0.02	0.07	0.19**	0.03	0.19**	0.03
	(0.06)	(0.17)	(0.06)	(0.17)	(0.07)	(0.15)	(0.07)	(0.15)
ln(BORDER_d)	0.20***	0.32*	0.20***	0.33*	0.19***	0.22	0.18***	0.22
	(0.06)	(0.18)	(0.06)	(0.18)	(0.07)	(0.17)	(0.07)	(0.17)
ln(CENTRAL_d)	-0.19*	-0.14	-0.19*	-0.15	-0.19*	-0.15	-0.19*	-0.17
	(0.11)	(0.23)	(0.11)	(0.23)	(0.10)	(0.20)	(0.10)	(0.21)
ln(INV_o)	0.57***	-0.31**	0.56***	-0.32**	0.69***	-0.45***	0.68***	-0.44***
	(0.05)	(0.13)	(0.05)	(0.13)	(0.04)	(0.08)	(0.04)	(0.08)
ln(INV_d)	0.38***	-0.79***	0.37***	-0.80***	0.53***	-0.43***	0.53***	-0.43***
	(0.06)	(0.16)	(0.06)	(0.16)	(0.04)	(0.09)	(0.04)	(0.09)
ln(HRST_d)	0.22	0.54	0.22	0.57	0.60*	-0.11	0.62*	-0.07
	(0.20)	(0.50)	(0.20)	(0.50)	(0.34)	(0.68)	(0.35)	(0.68)
ln(POP_o)	0.11	-0.12	0.11	-0.13	-0.02	0.07	-0.02	0.07
	(0.08)	(0.21)	(0.08)	(0.21)	(0.03)	(0.08)	(0.03)	(0.08)

ln(POP_d)	0.27***	0.58**	0.27***	0.57**	0.06*	0.03	0.06*	0.03
	(0.09)	(0.25)	(0.09)	(0.25)	(0.03)	(0.07)	(0.03)	(0.07)
ln(DENS_d)	-0.08**	-0.06	-0.08**	-0.05	-0.05	-0.01	-0.06	-0.02
	(0.04)	(0.11)	(0.04)	(0.11)	(0.04)	(0.09)	(0.04)	(0.09)
ln(TEMP_d)	-0.02	0.42	-0.04	0.41	0.40	0.77	0.39	0.78
	(0.23)	(0.37)	(0.23)	(0.37)	(0.31)	(0.59)	(0.32)	(0.60)
ln(COAST_d)	0.17**	-0.42*	0.18**	-0.41*	0.32***	0.05	0.33***	0.06
	(0.07)	(0.23)	(0.07)	(0.23)	(0.08)	(0.20)	(0.08)	(0.20)
ln(TECH.SHARES)	*yes*	*yes*	*yes*	*yes*	yes	yes	yes	yes
Country Origin/Destination Fixed Effects	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*
Sample size	48,180	48,180	48,180	48,180	48,180	48,180	48,180	48,180
Nonzero onservations	2,854	2,854	2,854	2,854	3,365	3,365	3,365	3,365
Log-pseudolikelihood	-9919.116		-9930.268		-11990.45		-11992.6	
LR test	12849.766		12827.462		14013.625		3629.60	
p-value	0.0000		0.0000		0.0000		0.0000	
Wald test			3579.27		3632.83		3629.60	
p-value	0.0000		0.0000		0.0000		0.0000	
LR test of α	3118.76		3128.87		1200.00		1200.00	
p-value	0.0000		0.0000		0.0000		0.0000	
Voung statistic	9.34		9.34		10.69		10.67	
p-value	0.0000		0.0000		0.0000		0.0000	
McFadden's R2	0.393		0.392		0.369		0.369	
Adjusted McFadden's R2	0.385		0.385		0.362		0.362	
AIC	20096.23		20118.54		24238.89		24243.2	
Schwartz	21229.2		21251.5		25371.86		25376.17	

Notes: Robust standard errors are presented in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Each of the columns includes the negative binomial estimation and the first stage of the ZINB, the logit model. Overdispersion tests largely reject the null hypothesis of no overdispersion. Vuong statistics (Vuong, 1989), are also performed and reported at the bottom of each regression. The tests performed point to the need of the zero-inflated models to accommodate our estimations to the excess of zeros. '_o' and '_d' denotes origin-region and destination-region variables, respectively.