Institut de Recerca en Economia Aplicada Regional i Pública
Research Institute of Applied Economics

Document de Treball 2012/04, 64 pàg.
Working Paper 2012/04, 64 pag.

Grup de Recerca Anàlisi Quantitativa Regional
Regional Quantitative Analysis Research Group

Document de Treball 2012/03, 64 pàg.
Working Paper 2012/03, 64 pag.

"What attracts knowledge workers? The role of space, social connections, institutions, jobs and amenities"

Ernest Miguélez and Rosina Moreno

Research Institute of Applied Economics

Regional Quantitative Analysis Research Group

Institut de Recerca en Economia Aplicada Regional i Públic
Research Institute of Applied Economics

WEBSITE: www.ub-irea.com • CONTACT: irea@ub.edu

Grup de Recerca Anàlisi Quantitativa Regional
Regional Quantitative Analysis Research Group

WEBSITE: www.ub.edu/aqr/ • CONTACT: aqr@ub.edu

Universitat de Barcelona
Av. Diagonal, 690 • 08034 Barcelona

WHAT ATTRACTS KNOWLEDGE WORKERS? THE ROLE OF SPACE, SOCIAL CONNECTIONS, INSTITUTIONS, JOBS AND AMENITIES

Ernest Miguélez
Economics and Statistics Division, World Intellectual Property Organization,
34, Chemin des Colombettes, CH-1121 Geneva 20, Switzerland
& AQR-IREA, University of Barcelona. E-mail: ernestmiguelez@gmail.com

Rosina Moreno
AQR-IREA. Department of Econometrics, Statistics and Spanish Economy. University of Barcelona, Av. Diagonal 690, 08034 Barcelona, Spain. E-mail: rmoreno@ub.edu

Abstract
The aim of the present paper is to identify the determinants of the geographical mobility of skilled individuals, such as inventors, across European regions. Their mobility contributes to the geographical diffusion of knowledge and reshapes the geography of talent. We test whether geography, amenities, job opportunities and social proximity between inventors' communities, and the so-called National System of Innovation, drive in- and out-flows of inventors between pairs of regions. We use a control function approach to address the endogenous nature of social proximity, and zero-inflated negative binomial models to accommodate our estimations to the count nature of the dependent variable and the high number of zeros it contains. Our results highlight the importance of physical proximity in driving the mobility patterns of inventors. However, job opportunities, social and institutional relations, and technological and cultural proximity also play key roles in mediating this phenomenon.

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

JEL: C8, J61, O31, O33, R0

1. Introduction

The geographical mobility[1] of skilled workers has become a key issue in economics in recent years, attracting the attention of both academics and policymakers (Trippl, 2009; European Commission, 2000). Indeed, policymakers have actively endorsed the phenomenon: the mobility of researchers, scientists and, in general, highly skilled personnel has become one of the main pillars of the efforts to create the European Research Area (ERA) launched by the Lisbon Agenda back in 2000. Hence, in order to foster the establishment of the ERA, the European Commission has encouraged, among others, the promotion of "greater mobility of researchers" and "improving the attraction of Europe for researchers from the rest of the world" (op. cit., pp. 8). The present paper analyses precisely this phenomenon, by measuring the mobility of inventors across European regions.

The issue is important for a variety of reasons. First, human capital[2] endowments are often said to influence differentials in economic prosperity across space. Highly skilled workers are the engine of innovation (Dahl and Sorenson, 2010) as well as major sources of knowledge externalities (Lucas, 1988; Glaeser et al., 1995; Moretti, 2004). Second, when skilled workers move from place to place, their knowledge and skills move as well. "[K]nowledge 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, pp. 367).

In the analysis of what attracts and mobilizes talent, we find a regional focus essential. However, few studies have approached the issue from this angle. Here, we use a gravity model of immigration (applied to the subsample of knowledge workers) to test whether

[1] 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 occasionally use the term migration as well, to which we give the same meaning as mobility, though we acknowledge the differences between the two.
[2] In this paper we use the terms 'human capital' and 'talent' indistinctively and interchangeably. Some studies have highlighted differences between the two (Mellander and Florida, 2007), linking the latter to creative occupations and the former to educational attainment. In other work, however, high correlations are reported between the two parameters (Glaeser, 2005). In any case, our study focuses only on inventors, irrespective of their occupation or their educational attainment (which we assume to be high in both cases).

a set of regional ‘attribute’ and ‘relational’[3] variables influence talent mobility across regions in 17 Western European countries. We aim to estimate disproportionate levels of high-skilled individuals’ flows, above and beyond a baseline level that would be expected from the spatial distribution of overall invention activity.

We hope that the present study will provide the answers to a range of questions. Our starting point feeds from the migration literature and aims to assess whether the migration costs associated to physical distance play any significant role. We also draw on more recent literature and test the role of several pulling factors such as amenities, job opportunities, and regional economic conditions in attracting talent. Thirdly, we acknowledge that inventors are a highly specific type of migrant. In consequence, we extend the standard analytical framework to the analysis of other more meaningful distances across locations that may determine the spatial location choices of mobile knowledge workers. Among a set of relational characteristics, the main variables under scrutiny in this study will be (1) the spread of inventors’ social networks to distant inventors’ communities and (2) the influence of the *National System of Innovation* (i.e., institutional proximity) in favouring mobility within countries versus cross-country movements, above and beyond physical distance. We also show that, broadly speaking, within-firm mobility does not have a strong influence on our results. We acknowledge the endogenous nature of cross-regional social networks of inventors. We base our identification strategy on the use of geographical/spatial variables to instrument social proximity and 2-stage residual inclusion (2SRI) estimation procedures, and show that endogeneity does not pos a serious concern. We take a static comparative approach by estimating our models in two separate time periods, 1996-1999 and 2002-2005. As regards the econometrics, we rely on zero-inflated negative binomial models to test our hypotheses.

This paper therefore contributes to the literature in four distinct ways: (1) in broad terms, it analyses the determinants of spatial mobility patterns of knowledge workers, which have not been addressed in depth before; more specifically, (2) it studies the

[3] According to Scott (2000, pp. 2-3), “attribute data” are the 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 relate one unit of analysis to another and cannot be reduced to the properties of the individual agent under study. Relations, then, are not the properties of the unit, but of *systems of units*.

influence of physical distance from former work colleagues in the destination choices of spatially mobile inventors; (3) it tests the role played by amenities versus job opportunities in attracting talent to the regions; and (4) it assesses the role played by more meaningful, observable linkages across distant regions. Our findings indicate the importance of physical separation in mediating the spatial mobility of inventors throughout the continent. However, institutional and social distances also play a significant role. These results are robust to the inclusion of other relational variables such as technological or cultural distances.

The outline of the paper is as follows: section 2 reviews some relevant previous studies on the differences in regional talent endowment, inventors' mobility, and skilled labour migration. Section 3 describes the empirical model and our research design, while also presenting the data and several estimation issues. Section 4 shows the results, and section 5 presents the conclusions and discusses certain limitations of our approach.

2. Literature review and previous empirical findings

Spatial differences in human capital endowments have been investigated in detail. The studies by Florida and colleagues are particularly well known. Analysing the US and Sweden respectively, Florida (2002a,b) and Mellander and Florida (2007) find significant correlations between regional talent endowments and various types of regional features, like social tolerance, diversity, coolness indexes, lifestyle indicators, and consumer amenities. Glaeser et al. (2001) argue convincingly that amenities are critical determinants of the spatial distribution of human capital. Shapiro (2006) notes that around 40% of the employment growth of college graduates is due to growth in quality of life. Yet, in our view, these approaches to the analysis of talent mobility have two main drawbacks. First, they are eminently static, i.e., they analyse stocks of talent, but do not explicitly consider flows of talent. Second, they rarely differentiate between talent created within the region and talent attracted from outside.

Both the migration (Borjas, 2000; Lewer and Van der Berg, 2008) and the economic geography literature (Tabuchi and Thisse, 2002; Crozet, 2004) have analysed cross-regional mobility of labour through the estimation of migration equations. Clearly, our approach draws on this literature, insofar as we also study the migration movements of individuals across locations. However, our focus on knowledge workers' practices to some extent distinguishes our approach from the traditional ones.

In sum, most of the related literature does not present systematic evidence of the determinants of the geographical mobility patterns of skilled individuals, especially with regard to inventors. Some studies have analysed the spatial mobility and location choices of recent college graduates. Faggian and McCann (2006, 2009) use structural equations models to explore the causes of regional human capital inflows 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 a 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. (2011) investigate the spatial mobility of graduates across Dutch regions, finding that the availability of large labour markets is a key factor in their location decisions. Gottlieb and Joseph (2006) also study the college-to-work migration patterns of US graduates and PhD holders. They find little evidence for amenities as spatial mobility drivers; employment opportunities seem to play a stronger role.

To our knowledge, few studies have focused on the determinants of location choices of highly skilled individuals. Scott (2010) analyses what drives inflows of migrant US engineers into different MSAs for 13 different technological categories, and finds that local employment opportunities have a major impact on the destination choices of these skilled individuals, far above amenities or even wages. Studying Danish scientists, engineers and entrepreneurs, Dahl and Sorenson (2009, 2010) report that distance to family, friends, former classmates, and so on are stronger motivations than the influence of potential income in their spatial location choices. This brief collection of results seems at odds with the generally accepted argument that high skilled individuals are less affected by physical distance in their location decisions (Ackers and Gill, 2008; Schwartz, 1973).

Briefly, a few points arise from this review. First and foremost, the main debate focuses on the influence of economic and job opportunities versus amenities in attracting talent, though no consensus has been reached so far. Second, the literature has also stressed the strong influence of unobservable linkages to the origin region, like family, friends, or colleagues. The present paper is closely related to these two branches of studies and contributes to this debate. However, little is known about what influences the location choices of highly-qualified knowledge workers from a regional perspective, since the impact of more meaningful linkages across locations has not been addressed. The present inquiry will try to fill this gap.

Our focus on inventors as a proxy of talented individuals may appear controversial, since it could be argued that they are only a proportion of skilled labour. Indeed, their numbers are small, but in general they have a critical economic significance (Calmfors et al., 2003): they are deeply involved in the production of innovations and, as a result they transfer larger quantities of knowledge when they move (Breschi and Lenzi, 2010). Indeed, scientists and engineers are central elements of Florida's (2004) super-core creative class.[4]

3. Research design

3.1. Empirical specifications

Baseline equation

The hypotheses sketched above have been tested in a regional migration framework (Biagi et al., 2011; Faggian and Royuela, 2010; Lewer and Van der Berg, 2008; Wall, 2001). This well-known setting is based on an individual's utility-maximizing framework, where the decision to move is influenced by the comparison between expected utilities of the origin and destination locations. The utility of a given location i for the n^{th} individual is a function of the region's economic features, including its job opportunities and its supply of amenities (all these factors are included under the label ' V ' in equation 1). An inventor will decide to move if and only if the expected utility of the destination region is greater than the expected utility of the origin region plus the costs of moving, both monetary and non-monetary. More formally,

[4] According to Florida (2004: 8), the core of the creative class comprises 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".

$$E[u_j^n(V_j^n)] - c(D_{ij}^n) > E[u_i^n(V_i^n)]. \quad (1)$$

As is customary in the related literature, the costs of migrating across regions, $c(D_{ij}^n)$, are proxied by the geographical separation between i and j. It aims to take on board several distance-related phenomena that are difficult to measure empirically, such as the sunk costs of re-location and aversion to risk of unemployment, the influence of family and friends from the origin region or, more importantly, inventors' preferences to re-locate close to their former colleagues and workmates if face-to-face interactions, information exchange and technical help are required. Several related studies note, however, that knowledge workers represent a highly specific type of skilled individuals whose location decisions are not greatly affected by physical separation (Ackers and Gill, 2008; Schwartz, 1973). This issue is critical from the perspective of particular regions, especially in the case of peripheral lagging regions whose strategy for attaining a critical level of human capital endowments is decisively based on the attraction of talent to catch up with the technological frontier of European core regions. Thus, ascertaining the specific role of geographical distance to explain mobility patterns of inventors across Europe, above and beyond the spatial distribution of innovation and economic activities, is one of the main objectives of the present paper.

Next, when equation (1) is met, the variable y_{ij}^n is set to 1, and 0 otherwise. By aggregating all individual decisions by pairs of regions, we end up specifying a general gravity model of regional immigration in the form of

$$y_{ij} = e^{\beta_0}(D_{ij})^{\beta_k} e^{\rho C_{ij}} \prod_{k=1}^{K} A_{ik}^{\gamma_{ik}} \prod_{k=1}^{K} A_{jk}^{\gamma_{jk}} \prod_{r=1}^{R} e^{\theta_{ir} d_{ir}} \prod_{r=1}^{R} e^{\theta_{jr} d_{jr}} \varepsilon_{ij} \quad (2)$$

where y_{ij} is the sum of individual location choices of inventors moving from region i to region j, and is a multiplicative function of a number of covariates. Among them we include a dummy controlling for contiguous regions, $e^{\rho C_{ij}}$, and a constant term capturing the impact of all common factors affecting mobility, e^{β_0}. In (2), $\prod_{k=1}^{K} A_{ik}^{\gamma_{ik}} \prod_{k=1}^{K} A_{jk}^{\gamma_{jk}} \prod_{r=1}^{R} e^{\theta_{ir} d_{ir}} \prod_{r=1}^{R} e^{\theta_{jr} d_{jr}}$ k are continuous variables and r are dummy variables designed to control for the spatial distribution of economic and innovation activities in both sending and receiving regions, as well as other pulling effects of the destination region, such as amenities (both natural and non-natural) or job opportunities.

The variables chosen to control for the spatial distribution of the economic and innovation activities are:

- Population (POP) in sending and receiving regions, proxying the spatial distribution of economic activity.
- The number of inventors (INV) in sending and receiving regions, proxying the spatial distribution of innovation and innovators.
- Origin and destination country-specific fixed effects.
- Share of patents for seven technological sectors (SHARE.TECH), in both sending and receiving regions, designed to control for differences in patent application propensities across technological branches.
- Distance from Brussels of the centroid of the destination region (CENTRAL_d).
- Dummy variable valued 1 if the destination region shares a physical border with a foreign country, and 0 otherwise (BORDER_d).

The effect of the number of inventors in the origin region on mobility is ambiguous. In principle, the larger this number, the higher the probability of observing cross-regional movers. However, inventors find more job opportunities in larger markets, and thus have less need to find them elsewhere: so the larger the number of inventors, the lower the probability of detecting a move.

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

(i) Job opportunities:

— 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.

— A healthy R&D environment in the destination region is expected to provide job and research opportunities for knowledge workers. The share of the active population that either successfully completed tertiary level education or are employed in a ‘Science and Technology’ occupation (HRST_d) is included as a general proxy for human capital, private R&D investment, the presence of universities and research centres, and the presence of technology-oriented venture capital firms.

(ii) Amenities:[5]

- Warmer winters, proxied by the average temperature in January (TEMP), as a predictor of incoming flows of skilled people (Gottlieb and Joseph, 2006).
- Access to coast (COAST), an important recreational amenity. It might also proxy for temperate weather during the whole year.
- Regional population density (DENS). Glaeser et al. (2001) argue that low density areas are highly attractive to immigrants. One should expect, then, a negative influence of density on inventors' inflows. However, these authors also acknowledge that density now has less power as an immigration predictor than ten or twenty years ago. In fact, it could also be argued that dense, urban areas may have a larger supply of producer and consumer amenities (Perugini and Signorelli, 2010), so a positive effect might also be observed.
- Total regional population (POP) is included (Scott, 2010). The sign of this variable is ambiguous. On the one hand, it has been argued that, like density, the availability of cultural amenities is greater in regions containing large cities and metropolitan areas. Conversely, the influence is negative if inventors have a preference for smaller, less polluted metropolitan areas 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.

So we have sketched a benchmark framework to define the factors that influence talent mobility across Europe. However, other more economically meaningful proximities across regions may play a role in explaining spatial mobility, above and beyond geographical distance -raising its point estimate if not controlled for.

Social proximity, the National System of Innovation and other relational variables

It is widely agreed in labour economics that social relationships are among the most effective ways of attaining successful recruitment (Meyer, 2001). The relationship

[5] We basically follow Scott (2010) in our definition of amenities.

between the employer and the future employee is set up through a third person known by both, acting as the intermediary. This 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). The dynamics of highly skilled mobility responds to the same logic (Meyer, 2001). Most positions are acquired via connections and, to some degree, knowledge workers make location decisions in the context of their professional relations and networks (Millard, 2005). To the extent that social networks are not necessarily spatially mediated, professional relationships between inventors may well cross regional boundaries. In this study we state that if two regions establish a large number of professional relations in the form of research collaborations, one would expect to see higher levels of inventor mobility between them. We label this *social proximity*.

Next, as we noted in the introductory section, one of the main concerns of the European Commission regarding the construction of the ERA is the low level of transnational mobility of skilled workers between EU countries. European R&D systems, policies, and programmes are characterized by fragmentation between countries – a situation that contrasts strikingly with the US – at "a huge cost to Europeans as taxpayers, consumers, and citizens" (European Commission, 2007, pp. 6). Indeed, it is a frequent claim among scientists and technology experts that their career opportunities and cross-country mobility choices are limited by legal and practical barriers. Generally speaking, most academic positions remain reserved for national staff, which restricts talent mobility across different institutional settings. Overall, it is argued that the *National System of Innovation* remains the institutional framework of reference for knowledge workers (European Commission, 2006) and the main reference point for major research activities (European Commission, 2000). Here, we empirically test whether the fact that two regions belong to two different institutional systems, or two different countries, negatively affects the probability of observing movement of highly-skilled professionals between them. If this were the case, the Commission's concerns would be justified, and new policies aimed to smooth differences across institutional frameworks would be required.

Additional control relational variables are considered in the estimation. Specifically,

(i) *technological distance*: included in order to test to what extent cognitive proximity (a shared, related, and complementary knowledge base) explains mobility across physically distant epistemic communities. We expect to find a negative effect of technological distance on mobility.
(ii) *cultural proximity*: inventors may choose to re-locate in regions sharing the same cultural background and language as their origin-region, in order to minimize migration costs. A positive and significant impact is expected for this variable.
(iii) *membership to elites of research excellence*: we also expect regions with above average efforts in research and innovation to belong to elite structures of research excellence (Hoekman et al., 2008) prone to exchange more talented individuals.

In sum, we now let D_{ij} be a vector of a broader set of meaningful distances between pairs of regions,

$$D_{ij} = f\begin{pmatrix} \text{GeographicDist.}_{ij}, \text{SocialProx.}_{ij}, \text{InstitutionalDist.}_{ij} \\ \text{TechnologicalDist.}_{ij}, \text{CulturalProx.}_{ij}, \text{ResearchExcellence}_{ij} \end{pmatrix}. \tag{3}$$

Inter-firm vs. intra-firm spatial mobility

A critical issue in our study is the role played by spatial movements of inventors within the same firm or group of firms. Quite often, for strategic purposes, firms have branches in separate locations. Individuals in firms usually obtain valuable knowledge from colleagues at their firm, including those in other locations. Spatial mobility of employees within firms' boundaries is one of the main ways through which knowledge spreads. This is not a trivial issue, since, according to our definition of labour mobility, 44.13% of the movements in the 2002-2005 period occurred within firms (8,585 movements in absolute terms).[6] Clearly, this phenomenon also implies knowldge diffusion and changes in the spatial configuration of talent. However, its implications from a regional point of view may be rather different. It is therefore important to ensure

[6] Figures computed using our data, as we will explain later on.

that our main hypotheses hold when we remove movements that do not correspond to real labour mobility.

3.2. Estimation issues

A logarithmic transformation of (2) and OLS techniques would be a straightforward estimation method. Santos Silva and Tenreyro (2006) show, however, that this standard procedure in a gravity model may induce a form of heteroskedasticity of the error term because of the log transformation of the data, and OLS would be inconsistent. Equally, it could be that there are no inventors' flows between a given pair of regions, making the logarithmic transformation of these observations impossible. Clearly, dropping these observations or adding an arbitrary constant to the dependent variable would again lead to inconsistent estimates (Burger et al., 2009). Santos Silva and Tenreyro (2006) 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 (2) can be written as the following exponential function

$$E(y_{ij} \mid x_{ij}) = \exp\begin{bmatrix} \ln\beta_0 + \beta_k \ln(D_{ij}) + \rho C_{ij} + \\ + \sum_{k=1}^{K}\gamma_{ik}\ln A_{ik} + \sum_{k=1}^{K}\gamma_{jk}\ln A_{jk} + \sum_{r=1}^{R}\theta_{ir}d_{ir} + \sum_{r=1}^{R}\theta_{jr}d_{jr} \end{bmatrix}, \quad (4)$$

where $x_{ij} = (1, D_{ij}, C_{ij}, A_{ik}, A_{jk}, d_{ir}, d_{jk})$. Thus, count data models can be used to estimate (4), avoiding in this way the logarithmic transformation of (2). Additionally, the response variable is a discrete one with a distribution that places the probability mass at non-negative integer values only, with data concentrated in a few small discrete values skewed to the left and intrinsically heteroskedastic, with variance increasing with the mean (Cameron and Trivedi, 1998). Again, count data models are more suitable in this framework.

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

$$P\left[y_{ij} \mid \mu_{ij}\right] = \frac{\exp(-\mu_{ij})\mu_{ij}^{y_{ij}}}{y_{ij}!}, \tag{5}$$

with a conditional mean (μ) of the distribution that is a function of the independent variables. The maximum likelihood estimator would be achieved by maximizing

$$\ln L(\beta) = \sum_{y_{ij}} \left[y_{ij} x'_{ij} \beta - \exp(x'_{ij} \beta) - \ln y_{ij}!\right]. \tag{6}$$

However, the Poisson distribution assumes equidispersion; that is to say, the conditional variance equals the conditional mean, i.e., $E(y_{ij} \mid x_{ij}) = Var(y_{ij} \mid x_{ij}) = \mu_{ij} = \exp(x'_{ij} \beta)$. But the conditional variance often exceeds the conditional mean (Burger et al., 2009; Long, 1997), which is a clear symptom of overdispersion. 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 may lead to consistent but inefficient estimates (Burger et al., 2009), with standard errors biased downward (Cameron and Trivedi, 1998; Long, 1997). Therefore, the negative binomial regression is preferred. In this 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 should be noted at this stage. Although count data models are explicitly designed to deal with the presence of zeros in the dependent variable, these zeros may come from different data generating processes. As a consequence, our dependent variable may have a greater frequency of zeros than would be predicted by the Poisson or negative binomial models (Greene, 1994). Specific estimation techniques are therefore required, such as the use of zero-inflated models. In these zero-inflated models the population is formed by two groups (Mullhay, 1986). One individual is in the first group with probability φ, and in the second group with probability $1-\varphi$. Thus, the estimation process includes two parts: first the probability of observing mobility from i to j, φ, is estimated by means of a probit or logit model, which is a function of certain characteristics – a set of covariates that predict the probability of belonging 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. The full model is specified:

$$\begin{cases} P(Y_{ij}=0)=\varphi_{ij}+(1-\varphi_{ij})\left(\dfrac{\alpha_{ij}^{-1}}{\alpha_{ij}^{-1}+\mu_{ij}}\right)^{\alpha_{ij}^{-1}} \\ P(Y_{ij}=y_{ij}>0)=(1-\varphi_{ij})\dfrac{\Gamma(y_{ij}+\alpha_{ij}^{-1})}{\Gamma(\alpha_{ij}^{-1})\Gamma(y_{ij}+1)}\left(\dfrac{\alpha_{ij}^{-1}}{\alpha_{ij}^{-1}+\mu_{ij}}\right)^{\alpha_{ij}^{-1}}\left(\dfrac{\mu_{ij}}{\alpha_{ij}^{-1}+\mu_{ij}}\right)^{y_{ij}} \end{cases}. \tag{7}$$

Thus, the log-likelihood function to be maximized is:

$$\begin{aligned} \ln L(\beta,\alpha_{ij}) = {} & \sum_{y_{ij}=0}\ln\left[\varphi_{ij}+(1-\varphi_{ij})\left(\frac{\alpha_{ij}^{-1}}{\alpha_{ij}^{-1}+\exp(x'_{ij}\beta)}\right)^{\alpha_{ij}^{-1}}\right]+ \\ & +\sum_{y_{ij}>0}\left[\begin{aligned} & \ln(1-\varphi_{ij})+\ln\left(\frac{\Gamma(y_{ij}+\alpha_{ij}^{-1})}{\Gamma(y_{ij}+1)\Gamma(\alpha_{ij}^{-1})}\right)+\alpha_{ij}^{-1}\ln\left(\frac{\alpha_{ij}^{-1}}{\alpha_{ij}^{-1}+\exp(x'_{ij}\beta)}\right) \\ & +y_{ij}\ln\left(\frac{\exp(x'_{ij}\beta)}{\alpha_{ij}^{-1}+\exp(x'_{ij}\beta)}\right) \end{aligned}\right]. \end{aligned} \tag{8}$$

The Vuong (Vuong, 1989) statistic can be used to assess whether the zero-inflated negative binomial is preferred to its non zero-inflated counterpart. In principle, there is no formal restriction to including the same regressors both in the binary and the negative binomial process, aside from possible theoretical considerations.

3.3. Data, variables construction, and descriptive figures

Dependent variable

We estimate our models for a sample of 220 European NUTS2 regions of 17 countries[7] (see Appendix 1) in two time periods – 1996-1999 and 2002-2005 – in order to study differences in point 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). In doing so, we expect to lessen potential endogeneity biases caused by system feedbacks. In the last section of the paper, we discuss the suitability of this approach and possible alternative solutions. Our dependent variable is built by full-counting the movements of inventors crossing regional borders. We therefore construct a mobility asymmetrical matrix of 220 rows and 220 columns for each time window, where each of the elements in the matrix is the number of inventors moving from region *i* to region *j*. If an inventor moves more than once, or if she returns to her former region, we compute these movements as separate and independent. Since by definition movements from region *i* to region *i* do not exist, we end up with a dependent variable reflecting flows between pairs of regions – (220)x(220-1)=48,180 observations. Mobility is computed through the changes observed in the region of residence reported by the inventor in patent documents from the European Patent Office (EPO). Of course, in this way we only capture mobility if the inventor applies for a patent before and after the move, and so we probably underestimate real mobility. We compute each movement between the origin and the destination patents, but only if there is a maximum time lapse of five years between them.

The data needed to build the matrix are taken from the REGPAT database (OECD, January 2010 edition). In spite of the vast amount of information contained in patent documents, there is no single ID for each individual inventor. To be able to trace the mobility history of inventors, we need to identify them individually by their name and surname, as well as via other useful information contained in the patent document. The method chosen to identify the inventors is of the utmost importance in studies of this nature. Here, we follow Miguélez and Gómez-Miguélez (2010), who, in line with a

[7] We have omitted the regions of Las Canarias, Ceuta, Melilla, Madeira, Açores, Guadeloupe, Martinique, Guyane and Reunion due to their distance from continental Europe. We do not expect this omission to alter our results significantly.

growing number of researchers in the field, use different customary heuristics for singling out individual inventors using patent documents.[8]

For the whole 1975-2005 period, 768,810 individual inventors were identified. Table 1 reports some notable figures. The spatial distribution of these inventors across regions is very uneven – the Gini coefficient, 0.71, is relatively high. Note also that of these unevenly distributed inventors, only 11.54% are considered mobile (i.e., they report more than one NUTS2 region of residence within our period).[9]

[Insert Table 1 about here]

As for the specific case of our dependent variable, we identified 26,178 movements (10,813 in the first period and 15,365 in the second), which are also highly concentrated from a geographical perspective: 5.5% of the regions did not receive any inventors at all during the 2002-2005 period (9.5% for the 1996-1999 period), while 19.1% (25.5%) of them received only six or fewer. On the other hand, around 50% (44.5%) of the inflows (inventors moving into a given region) were concentrated in only 20 regions.[10]

On average, the distance covered by inventors' movements reported between 2002 and 2005 was around 397 kilometres – approximately the driving distance between Paris and Luxembourg. This figure is relatively low, and is around half the distance found in another study for the US (Breschi and Lenzi, 2010). Furthermore, 30.79% of movements into the regions come from their five nearest neighbours, and 44.33% from their ten nearest ones. Note again from Table 1 that the average distance covered by the

[8] We are fully aware of the dangers of using patent data in economic analysis. The criticisms by Griliches (1991) are well known. Others have stressed that firms to a large extent build up a patent portfolio for strategic reasons, in order to improve their position in negotiations or their technological reputation (Verspagen and Schoenmakers, 2004). Likewise, it has been shown that patent data may underestimate or overestimate real mobility (Lenzi, 2010), and other studies have raised more general criticisms concerning the use of patents for regional analysis (Ter Wall and Boschma, 2009). We do not think that these shortfalls influence differences across regions, and so they do not pose a serious bias in our estimates.

[9] For comparative purposes, the 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 MSAs. Trajtenberg and Shiff (2008) find that 19.8% of software inventors from the USPTO report more than one geographical location, while 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.).

[10] Noord-Brabant (NL), Île de France (FR), Koeln (DE), Surrey, East and West Sussex (UK), Oberbayern (DE), Karlsruhe (DE), Darmstadt (DE), Stuttgart (DE), Dusseldorf (DE), Rheinhessen-Pfalz (DE), Rhone-Alpes (FR), Mittelfranken (DE), Tubingen (DE), Bretagne (FR), Freiburg (DE), Berlin (DE), Etelae-Suomi (FI), Wien (AT), East Anglia (UK), and Hamburg (DE).

movements computed increases by around 25 kilometres between the first and the second time periods. This suggests that, over time, distance is becoming less important as an explanation of inventors' geographical mobility, though the econometric specification should shed some light on this issue.

Using maps, Figure 1 depicts the patterns of innovator mobility in the two time-windows. The lines connect the regions' centroids when at least one inventor has moved from one region to another. It does not matter how many inventors have moved from i to j, since the thickness of the line does not take this into account. Figures 1.3 and 1.4 take the intensity of the pairwise mobility and depict as linked only those pairs of regions with five total movements or more (in at least one of the directions). As can be seen from all pictures, most movements involve central regions, which in turn are the most innovative. The result of this is that the majority of movements involve relatively short distances.

[Insert Figure 1 about here]

To illustrate this point further, we plot the kernel density estimations of the distribution of the distance covered by inventors' movements in the two periods (in km). The distribution of movements is extremely skewed to the left, i.e. the distance covered tends to be low. Note that, surprisingly, differences across the two periods are unappreciable – although probably the interval between the two periods is too short to reveal important changes.

[Insert Figure 2 about here]

In sum, physical distance seems to be pivotal in explaining the mobility patterns of inventors across space. However, these figures may be due to the skewed distribution of innovation across space. In order to explore this possibility in more detail, figure 3 examines whether inventors move from highly productive regions (in terms of patenting activity) to other highly productive regions, or whether movements from low to high-productive regions (or vice versa) dominate. For the two periods under study, the figures depict histograms of the number of movements as a function of the difference between the patent intensities of the regions involved. Although not strictly symmetrical (especially in the second period), these figures show that the majority of the movements occur between regions with similar levels of innovative activity (high-high, low-low). Very similar findings are reported in Azoulay et al. (2011) for US scientists.

[Insert Figure 3 about here]

Bearing this concentration of innovation and economic activities in mind, we wonder whether these figures are an artefact of this distribution or whether they truly reflect inventors' preferences for short-distance movements. Therefore, our aim in the present paper is to determine (1) whether, after controlling for the fact that the spatial distribution of innovators is not random throughout space, migration costs associated to physical separation influence the mobility patterns of these skilled workers; and (2) whether other variables may explain this phenomenon, after controlling for physical distance as well.

Explanatory variables

With respect to the explanatory variables, the geographical distance between regions' centroids (GeoDIST) is computed in different ways, running variants of the same model in order to study the robustness of the coefficients: driving distances (in kilometres) and driving time (in seconds), both calculated using Google Maps. Robustness checks include Euclidean and great circle distances as well.

Institutional distance is proxied with a dummy variable valued 1 if the pair of regions do not belong to the same country and 0 otherwise (as in Ponds et al., 2007 and Hoekman et al., 2008). Social proximity is proxied using EPO co-patents across NUTS2 regions (REGPAT database). Thus, when one patent contains inventors who report their addresses in different regions, we assume that there is cross-regional collaboration. We 'full-count' all the collaborations across regions, irrespective of the number of inventors reported in each patent. We thus 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'. 'Social affinity' between regions i and j, A_{ij}, 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, J. Formally,

$$A_{ij} = l_{ij} - (n_i / J). \tag{9}$$

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

$$A_{ij} = l_{ij} / n_i . \tag{10}$$

in order to avoid negative values and to allow the logarithmic transformation of the variable. [11]

The patent data from EPO needed to calculate technological distance are taken from the REGPAT database and assigned to each of the technological sectors using the IPC[12] classification system. To proxy technological distance, we use the following index:

$$TechDis._{ij} = 1 - t_{ij}, \tag{11}$$

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

$$t_{ij} = \frac{\sum f_{ih} f_{jh}}{\left(\sum f_{ih}^2 \sum f_{jh}^2\right)^{1/2}} . \tag{12}$$

In (12), f_{ih} stands for the share of patents of one technological class h according to the IPC classification (out of 30 technological classes in the subdivision chosen) of region i, and f_{jh} for the share of patents of one technological class h of region j. Thus, values of the index close to zero indicate that two regions are technologically similar, and values close to unity indicate that they are technologically distant (see Jaffe, 1986).

[11] A small constant has been added to all the explanatory variables with at least one 0 value for the same reason.

[12] International Patent Classification.

As in Picci (2010), we calculate cultural proximity by computing an index of language similarity across regions. According to the author, it is reasonable to expect that people whose languages share common roots will also share similar cultural backgrounds. To compute this index, we gather data from the Ethnologue Project (www.ethnologue.com) in order to assign a single language to every NUTS2 region. 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, an index based on the distance between branches in the classification of languages.[13] We sum the number of branches that coincide between each pair of languages and divide the result by the sum of branches of each of the two languages (in order to take into account the fact 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. For instance, the similarity index between Spanish and Portuguese is 0.889, and between Swedish and Danish is 0.769, whereas the index between Portuguese and Danish is just 0.125.

Finally, membership to elite structures of research excellence is computed with a dummy variable valued 1 if the proportion of individuals in the total active population who successfully completed a tertiary education degree and who are currently employed as professionals or technicians in a 'Science and Technology' occupation is above the mean in the two regions, and 0 otherwise (Human Resources in Science and Technology data are retrieved from Eurostat databases).

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

---

[13] For example, the linguistic classification of Portuguese, Swedish, and Danish, from the largest, most inclusive grouping to the smallest, 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).

variables under consideration. Note that the average distance between pairs of regions, 1,524 km, is around four times larger than the average distance covered by the inventors' movements.

[Insert Table 2 about here]

Spatial-labour mobility

Two alternative matrices for constructing spatial-labour mobility dependent variables are also built. Even though the identification of geographical mobility is reasonably easy in most of the cases, the identification of strict labour mobility may be difficult (see Laforgia and Lissoni, 2006). To narrow our definition of spatial labour mobility, we look at the patents that surround each spatial movement. Previously, we gathered firm and group information from the KITeS-Bocconi University databases, and matched them with our REGPAT datasets. If at least one firm or group of firms coincides in both the origin and the destination patent-region, we remove that movement from our dependent variable. As a result, we obtain two matrices reflecting spatial mobility between firms and spatial mobility between groups of firms, which can be used to build two additional dependent variables. Note that our definition of labour mobility is very strict and probably underestimates real mobility. We prefer to be, however, conservative; in fact, using other more relaxed definitions of spatial labour mobility the results (provided upon request) did not change substantially.

4. Results

In this section we summarize the main results obtained with 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 (driving distance in kilometres and driving distance in time), and for both time spans. Both the negative binomial and the logit estimations were estimated. 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, as shown in Table 3, a 1% increase in the distance between regions' centroids would lead to a 1.45% decrease in the probability of observing 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 a 1% increase in the number of inventors in a given region leads to a 0.39% decrease in the probability of belonging to the "strictly zero group" (Maggioni and Uberti, 2009) – that is, the probability of zero bilateral mobility. For the sake of brevity, the present section shows only the negative binomial estimations. The remaining results are displayed in tables A.3.1 to A.3.4 in the Appendix 3 section, where other proxies for physical distance are also shown.

Physical distance

Columns (i) and (ii) in Table 3 present the estimation of equation (2), including distance as the only focal relational variable - 1996-1999 period, aside from other regional controls. The estimated coefficients are negative and strongly significant, irrespective of the proxy used. These coefficients, between -1.45 and -1.54, are larger than we initially expected. In reality, the elasticity is very close to what we find in similar frameworks for trade data (see Disdier and Head, 2008, for a meta-analysis of this topic) or co-patenting data (Maggioni and Uberti, 2009), and considerably higher than what we find for citations data (Peri, 2005). Actually, these coefficients are in line with the migration literature (see Crozet, 2004, for an analysis at the European regional level).

[Insert Table 3 about here]

Columns (iii) and (iv) in Table 3 show the same estimated model, but for the period 2002-2005. Broadly speaking, the results are maintained over time. A chi-squared test of individual coefficients does not reject the null hypothesis that the differences between the two periods are not statistically significant (test provided upon request). This seems to be slightly contradictory, since one would expect the importance of physical separation to decrease over time with the increasing use of communication technologies, for instance.

Overall, these preliminary findings suggest that the distance from family and friends and, especially, from former work colleagues, is pivotal in explaining the spatial location choices of migrant inventors, and also that its importance does not seem to decrease as the economy becomes more technologically advanced and specialized (although, again, there is only a lapse of six years between the two periods of time under consideration). Bear in mind, however, that the geographical coefficient may well be biased upward if other more meaningful distances are not controlled for. We hope to shed further light on this issue in the following subsection.

Social proximity, institutional distance and other relational variables

Table 4 shows the estimation of the unrestricted model, which includes social proximity and institutional distance as well as other relational control variables. The table shows the results for the 1996-1999 period in the first two columns and for the 2002-2005 period in the last two columns. Again, we only provide results using kilometres and time as physical separation proxies, but the results remain unchanged with other distance variables (see Appendix 3). From these columns we extract the following findings. First and foremost, our focal variables in the present inquiry, i.e., institutional distance and social proximity, are significant and have the expected sign (negative and positive respectively). 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, while that of institutional distance decreases. These differences in point estimates over time are shown to be significant by the chi-square tests performed for the case of institutional distance, but not for social proximity (tests provided upon request). In brief, even though physical distance does not decrease in importance over time, the general innovation framework seems to become progressively more internationally based.

Second, these tables show that once other proximities across regions are controlled for, the role conferred on physical distance decreases considerably, by more than half, confirming our suspicions that a sizeable bias is introduced if they are neglected. Certainly, geographical and other distances may partially overlap, but each feature may have a different, independent effect on mobility that must be isolated correctly. Finally, technological distance, cultural proximity, and networks of excellence are also significant – note, though, that belonging to elites of research excellence is only significant in the second period.

In sum, the empirical exercise conducted so far assigns a critical role to geographical separation in explaining inventors' spatial mobility and location choices. However, the two main variables under scrutiny in the present paper also show significant values and the expected signs in explaining the phenomenon under analysis.

[Insert Table 4 about here]

Attribute variables: amenities versus job opportunities

We also enter the ongoing debate on the importance of amenities versus job opportunities by including several variables that are widely used in the literature in order to test whether their role in attracting talent is also witnessed in this specific group of knowledge workers. For instance, density in the destination region (DENS_d) seems to have a negative influence on attracting inventors, corroborating the arguments of Glaeser et al. (2001). However, its point estimates are not significant in the second period, in line again with the thesis that this variable is less important today than it was a few years ago. Population in the destination region (POP_d) 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, and lower coefficients (significant at 10%) in the second period. So the attractiveness of large metropolitan areas seems to be important, but it decreases over time. As regards natural amenities, warmer climates (TEMP_d) have only a slight influence on inventors' location decisions in the first period, but a strong influence in the second. Meanwhile, access to the sea (COAST_d) is positively and significantly related to inflows of inventors.

Among the variables designed to control for destination-region job opportunities, we find that the size of the inventors' community in the destination region (INV_d) is positively (and strongly) correlated with our dependent variable, irrespective of the time span and the estimated model. Meanwhile, regional R&D efforts (HRST_d) also seem to matter, especially in the second period.

Thus, despite the roughness of the proxies used, both amenities and job opportunities seem to play a certain role in attracting talent. However, the size of the destination-region labour market for inventors has a stronger influence than other pulling factors, making this variable the most decisive attraction characteristic.

Inter-firm spatial mobility

Table 5 reproduces the estimation of columns (iii) and (iv) in table 4, but considers only strict labour mobility as a dependent variable. The relational variables included remain significant and with the expected sign – cultural proximity increases its point estimate and becomes significant at 5%. However, the results for the case of the attribute variables in the destination region are slightly ambiguous. Human Resources in Science and Technology in the destination region decreases its coefficient and increases its standard error, becoming insignificant. This is also true for Population in the destination region – columns (iii) and (iv). Other changes are not worth reporting and, in general, the main conclusions continue to apply when strict labour mobility is considered; therefore, it does not have a great effect on our results.

[Insert Table 5 about here]

Causality

A critical concern in any empirical analysis is endogeneity, which produces biased and inconsistent estimates. By lagging r.h.s. variables, we can reduce the endogeneity problems due to simultaneity (a future event cannot 'cause' a past event). We believe that endogeneity no longer poses a serious problem for the majority of the variables included in the model. We acknowledge the remaining endogeneity concerns with respect to social proximity. Both inventors' spatial mobility and cross-regional co-patents are rare phenomena which depend heavily on patent data and patenting inventors' practices. Thus, even if lagged r.h.s. variables are included, unobserved heterogeneity may introduce endogeneity problems, such as the tendency of a given technological sector or firm to patent above the average. If this is the case, social proximity would not be completely exogenous, and biased estimates would arise.

We adopt several approaches to address this issue. First, we repeat estimation (iv) from table 4, but include several time lags of the dependent variable as additional explanatory variables. By doing so, we aim 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 – unobserved historical linkages, common labour market institutions, and so forth. In a sense, we mean to control for the historical inertia of a given pair of regions to exchange inventors, as if it were a region-pair fixed effect.[14] In short, time lags account for "*historical* factors that cause *current* differences in the dependent variable that are difficult to account for in other ways" (Wooldridge, 2002, pp. 289). With this idea in mind, in the 'unrestricted 2002-2005' model we include the dependent variable lagged either one period (movements 1998-2001), two periods (movements 1994-1997) or three periods (movements 1990-1993). The results of these estimations – columns (i) to (iii) in table 6 – show that the negative effect of institutional and, especially, physical distance is notably reduced, while the social proximity coefficient remains virtually unchanged. However, the three variables remain strongly significant. At the same time, other variables decrease 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 toward negligible values, which depend on the level of serial correlation and the time elapsed between the lagged variable and the dependent variable we want to explain (Achen, 2001). Therefore, these estimations should be interpreted with extreme care.

An alternative approach is to find suitable instruments for the social proximity variable. They must be (1) uncorrelated to the unobservable time-varying error term; and (2) sufficiently correlated to the endogenous variables that we want to instrument. In other words, the instrument must be completely exogenous and must be relevant. This is by no means a trivial task. We have a list of potential spatial/geographical candidates as instruments, i.e. origin- and destination-region fixed effects, as well as other variables such as whether the two regions belong to the same NUTS1 region, whether origin and destination regions host the country's capital city, the log of the average area in squared kilometres of the two regions, whether they belong to contiguous countries, whether

[14] For gravity models of trade, for instance, Eichengreen and Irwin (1998) and Anderson et al. (2004) argue that historical hysteresis between pairs of countries as regards bilateral trade should be accounted for by including time lags of the dependent variable in the r.h.s. of the equation, especially in the absence of fixed effects. For the case of gravity models of immigration, Anjomani and Hariri (1992), Kazakevitch (1996), or Fry (1999) argue that lagged migration variables in the r.h.s. of the equation may help to control for unobserved causes of migration.

they belong to the core regions of Europe[15], and the sum of their distance to Brussels, in logs.

We then apply the 2-stage residual inclusion (2SRI) estimator (Terza et al., 2008) or control function approach (Wooldridge, 2002). As has been shown, the 2SRI is consistent in non-linear models while 2-stage prediction substitution (2SPS) estimators are not – conversely, they are fully consistent in linear models, as in the well-known case of the 2-stage least squares (2SLS). The reason for this is the non-additive nature of either the observable or the unobservable confounders (see Terza et al., 2008). In practice, therefore, we regress the instruments on our social proximity variable in the first stage, conditional upon the other exogenous variables of the original model (except all the attribute variables, whose effect is picked up by the fixed effects), and recover the predicted residuals of this estimation, to plug them into our original model (without excluding the social proximity variable) - inference based on bootstrapping over all two-step procedure, 1,000 iterations. The result of this process is shown in the last column in table 6. Note that the partial R^2 of the first stage is 0.513 and the value of the F-tests statistic, 23.82, is well above 10, which is usually considered a good threshold, and so the instruments cannot be judged as weak (see table A.3.7 in the Appendix for the results of the first stage OLS regression). The positive coefficient of the control term included tells us that the latent factor captured by the instruments is positively correlated with cross-regional mobility. Hence, endogeneity seems to cause a small upward bias in the social proximity coefficient in our previous estimates. Note, however, that the bias is small and the control term is not significant, so the main conclusions of the analysis undertaken so far hold.

[Insert Table 6 about here]

Robustness checks

In this section we summarize some robustness checks performed to study the stability and significance of the estimated parameters, and the results encountered so far. For the

[15] Core regions are defined as regions whose centroid lies within a pentagon formed by a straight line linking Milan, Munich, Hamburg, London, and Paris.

sake of brevity, we omit the tables in the main section, but they can be found in Appendix 3.

These tables show the estimation of our main models, including Euclidean and great circle distances. Unlike the tables in the main part of the text, they also include the logit estimations. Columns (i) and (ii) in table A.3.8. repeat the estimation but, following the literature on migration economics, they include the average income in the manufacturing sector of the destination region and the income gap between origin and destination regions. Despite the fact that we could not use all the regions and time spans due to missing data, these variables did not turn out to be significant in any of the estimations. This is consistent with previous findings regarding high-skilled workers. Scott (2010, pp. 59) argues that “engineers (may be) relatively insensitive to wage and salary differences across geographic space in relation to potential employment opportunities”. Column (iii) includes the number of citations per capita the destination region receives, in order to reflect the attractiveness of more productive (and thus more cited) regions. No notable changes are reported. Third, given the strong significance of the first-order contiguity variable, we include second and third-order contiguity variables and re-estimate the models (column (iv)). Fortunately, none of the variables included turns out to be significant, and the parameters for the remaining variables remain virtually unchanged.

In columns (i) and (ii) in table A.3.9 we play around with the variables predicting the probability of belonging to the strictly zero group, that is, the logit estimation. Column (i) excludes the relational variables from the logit estimation, whilst in column (ii) we omit the attribute ones. As can be seen, the main results remain unchanged, with few exceptions.

In column (iii) in Table A.3.9 we remove the ‘mass’ variables (inventors in origin and destination regions), remove the share of patents in each technological sector, and include the absolute number of inventors in the origin and destination regions split into seven technological sectors. For the most part, the results are maintained. Logically, technological distance is no longer significant. Column (iv) re-estimates the model but excludes from the identification process all the inventors with Soundex codes of name plus surname with more than 50 records (see Miguélez and Gómez-Miguélez, 2011), because it seems that the algorithm works better with Soundex codes with fewer records. Again, the main results are maintained.

5. Conclusions and implications

Throughout the above sections, we have tried to disentangle the effect of some key regional features on the spatial mobility patterns of skilled workers, i.e., inventors, across the European geography. With the advent of the knowledge-based economy, identifying territorial features that favour or hinder the attraction of talent is of the utmost importance. The ability to attract knowledge workers increases access to distant sources of knowledge; they act as 'pipelines' to distant pools of ideas, which are mastered and diffused locally through the local 'buzz' once they enter the region. Furthermore, it is widely agreed that the spatial agglomeration of human capital may also influence regional growth rate differentials. Consequently, the map of human capital is constantly reshaped by labour migration, and so it is important to investigate "the forces that influence 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). We consider empirical exercises like the present one to be of critical importance. However, little evidence on the issue is currently available. In this 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 highlight a number of factors likely to affect inter-regional mobility, and test them in the empirical section.

Our empirical analysis shows that physical separation from the inventors' former workplace is a critical predictor of their spatial movements, even after controlling for the spatial distribution of innovation and economic activities. In fact, we expected this variable to play a more secondary role. However, in spite of the announcements of "the death of distance" (Cairncross, 1997), we find physical space to be pivotal in mediating inventors' mobility across regions. These results are robust to the sample choice, specification, and inclusion of controls. To the extent that inventors are carriers of knowledge, these results may partially help to explain the well-known findings reported by Jaffe et al. (1993) on the localization of knowledge flows (Breschi and Lissoni, 2009).

Oher 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 results for the role of amenities and job opportunities as talent attractors. Our results suggest that job opportunities have a greater influence, especially in the later period, though amenities also appear to play a role as well (for a recent discussion on the topic, see Biagi et al., 2011). We acknowledge that further research on this point is required.

The implications of these results are not particularly encouraging. We interpret the large and strongly significant geography coefficient as follows: when knowledge workers decide to move, they place a high value on locating close to their former colleagues, from whom they receive constant inflows of information about job and business opportunities, technical solutions, and, in general, knowledge spillovers (similar conclusions are found in Dahl and Sorenson, 2010, p.44). Second, on the way towards the ERA, this paper confirms that the fragmentation of the institutional framework between countries impedes frictionless mobility across national borders. Despite recent progress – there are significant differences in parameters estimates between the first and second period – much work remains to be done to overcome this fragmentation, which remains a prevailing characteristic of the European research base. Policies aimed at making recruitment procedures more transparent, improving the portability of social security provisions across countries, and reducing differences in taxation must be implemented sooner rather than later. In sum, our results suggest that there is little scope for policy action at regional level to attract talent and to decrease technological gaps and income inequalities across regions unless a given region is located within a reasonable physical distance of 'highly-talented' regions. In this situation, regions should ideally devote greater efforts to helping their own populations to raise their level of human capital rather than trying to attract skilled individuals from other places, in order to attain a critical mass of talent within the region.

One final remark is in order. In spite of the generally negative tone of the results, promising findings are the decreasing role of institutional distance over time, and the significant influence of formal and professional relationships across distant inventors' communities. Thus, from a regional perspective, joining international and inter-regional networks of research collaboration is beneficial for two main reasons: first, because of the direct knowledge acquired via research collaborations, and second, because of their effect in smoothing out frictions that may impede the free mobility of talent across Europe.

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 also received many helpful comments from the participants at 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, (Seville, 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), the 51st European Regional Science Association Conference (Barcelona, 1st September 2011), the 36th Simposio de la Asociación Española de Economía (Málaga, 16th December 2011), Camilla Lenzi, Francesco Lissoni, Johannes Rode and Jouke van Dijk. We also acknowledge financial support from the Ministerio de Ciencia e Innovación, ECO2008-05314 and ECO2011-30260-C03-03, 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 mistakes or omissions remain ours.

References

Achen, C. H. (2001) “Why Lagged Dependent Variables Can Suppress the Explanatory Power of Other Independent Variables,” presented at the Annual Meeting of the Political Methodology Section of the American Political Science Association, UCLA, July 20-22, 2000.

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

Anderson, M.A.; Ferrantino, M.J.; Schaefer, K.C. (2004) Monte Carlos Appraisals of Gravity Model Specifications, Office of Economics Working Paper No. 2004-05-A, US International Trade Commission

Anderson, J. E. and van Wincoop, E. (2003) “Gravity with Gravitas: A Solution to the Border Puzzle” American Economic Review, 93(1): 170-192;

Anjonami, A.; Hariri, V. (1992) “Migration stock and the issue of competing and complementary flows in United States interstate migration” Journal of Population Economics 5(2): 87-100

Azoulay, P., Graff Zivin, J. S., and Sampat, B. N. (2011): “The Diffusion of Scientifc Knowledge Across Time and Space: Evidence from Professional Transitions for the Superstars of Medicine”, Manuscript Prepared for the “Rate and Direction of Inventive Activity NBER 50th Anniversary Conference”.

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;

Biagi, B., Faggian, A. and McCann, P. (2011) “Long and short distance migration in Italy: the role of economic, social and environmental characteristics” Spatial Economic Analysis, 6(1), 111

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 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.

Calmfors, L.; Giancarlo Corsetti, John Flemming, Seppo Mikko Sakari Honkapohja, John Kay, Willi Leibfritz, Gilles Saint-Paul, Hans-Werner Sinn and Xavier Vives (2003) “Should We Worry about the Brain Drain?” EEAG Report on the European Economy from CESifo Group Munich

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

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;

Dahl, M. S. and Sorenson, O. (2009) “The embedded entrepreneur” *European Management Review*, Vol. 6, pp. 172-181

Dahl, M. S. and Sorenson, O. (2010) "The migration of technical workers" Journal of Urban Economics, vol. 67(1), pages 33-45

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;

Dumont, J.C. and G. Lemaître (2005) “Counting Immigrants and Expatriates in OECD Countries: A New Perspective.” Social, Employment, and Migration Working Papers 25. Paris: OECD Directorate for Employment, Labour, and Social Affairs.

Eichengreen, B. and Irwin, D. (1998) “The role of history in bilateral trade flows”, in J. Frankel (ed.) The Regionalisation of the World Economy. Chicago: University of Chicago Press, pp. 33-57.

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 (2006) “Annex to the Green Paper. A European Strategy for Sustainable, Competitive and Secure Energy” SEC(2006) 317/2

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

European Commission (2010) “A vision for strengthening world-class research infrastructures in the ERA”, Report of the Expert Group on Research Infrastructures

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

Faggian, A. & Royuela, V. (2010) “Migration flows and quality of life in a metropolitan area: the case of Barcelona, Spain” Applied Research in Quality of Life, 5(3), 241_259

Feenstra, R. C. (2004) Advanced International Trade: Theory and Evidence, Princeton, Princeton University Press.

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.

Fry, J.; T. Fry; M. Peter (1999) “Inter-Regional Migration in Australia: An Applied Economic Analysis”. Australasian Journal of Regional Studies, 5: 111-130
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;
Griliches, Z. (1991) "Patent Statistics as Economic Indicators: A Survey," NBER Working Papers 3301
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;
Jaffe, A.B. (1986) Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value. *American Economic Review* 76(5): 984-1001.
Jaffe, A.B., Trajtenberg, M., and Henderson, R. (1993) “Geographic localization of knowledge spillovers as evidenced by patent citations” Quarterly Journal of Economics 63, 577-598;
Kazakevitch, G. (1996) Regional Labour Market Disparities and International Migration in Australia. Zurich: European Reg
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;
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 Gómez-Miguélez I. G. (2010) Singling out individual inventors from patent data, IREA Working Paper 2011/05;
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.
Ortega, F.; Peri, G. (2009) “The causes and effects of international labor mobility: Evidence from OECD countries 1980-2005” MPRA Paper No. 19183
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;
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;
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
Schwartz A. (1973) “Interpreting the Effect of Distance on Migration”, Journal of Political Economy, 81 (5) pp. 1153-1169;
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. (2011) “An analysis of trends in spatial mobility of Dutch higher education graduates” forthcoming in Spatial Economic Analysis;
Verspagen, B. and Schoenmakers, W. (2004) "The spatial dimension of patenting by multinational firms in europe" Journal of Economic Geography, vol. 4(1), pages 23-42

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

Wall, H. J. (2001) “Voting with your feet in the United Kingdom: using cross-migration rates to estimate relative living standards” Papers in Regional Science, 80, 1–23.

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

Wooldridge, J. (2002) Econometric Analysis of Cross Section and Panel Data, MIT Press.

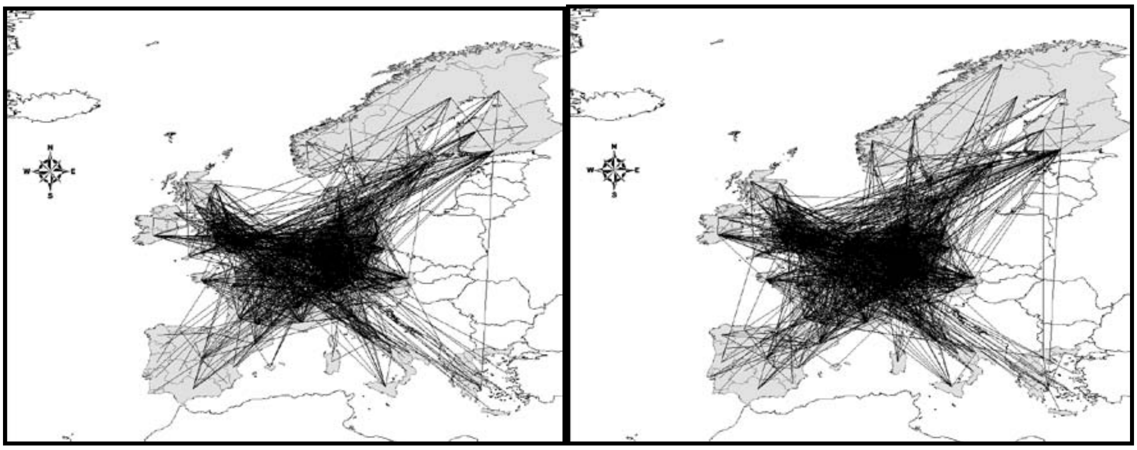
Table 1. Descriptive figures

Inventors identified (1975-2005)	768,810
Share of mobile inventors (1975-2005)[(1)]	11.54%
Inventors' distribution across regions: Gini index (1975-2005)	0.71
Movements	15,365 (10,813)
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 inflow 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%
Average distance covered by inventors' movements	
Euclidean	3.56º (3.23º)
Great circle	188.32 (175.29)
Km	397.46 (374.68)
Time (seconds)	14,970.35 (14,221.72)

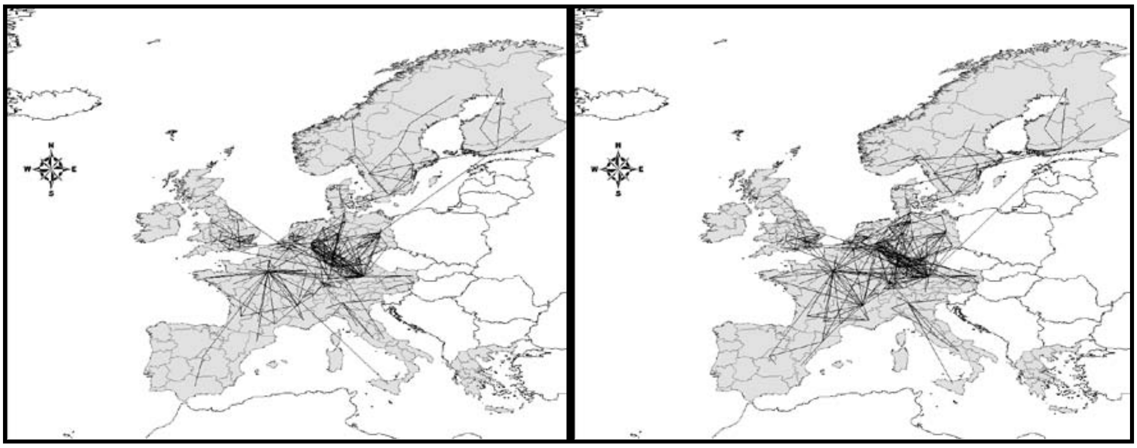
Notes: Values for the period 2002-2005, when applicable. In parentheses, 1996-1999. **(1)** Mobile inventors are those reporting more than one NUTS2 region of residence throughout the whole period.

Figure 1. Movements connecting regions' centroids

1.1. Movements 1996-1999 **1.2. Movements 2002-2005**



1.3. Movements 1996-1999 **1.4. Movements 2002-2005**



Notes: In figures 1.3 and 1.4, the threshold is set at five movements (in at least one of the directions).

Figure 2. Distribution of the spatial extent of individual movements, in km.

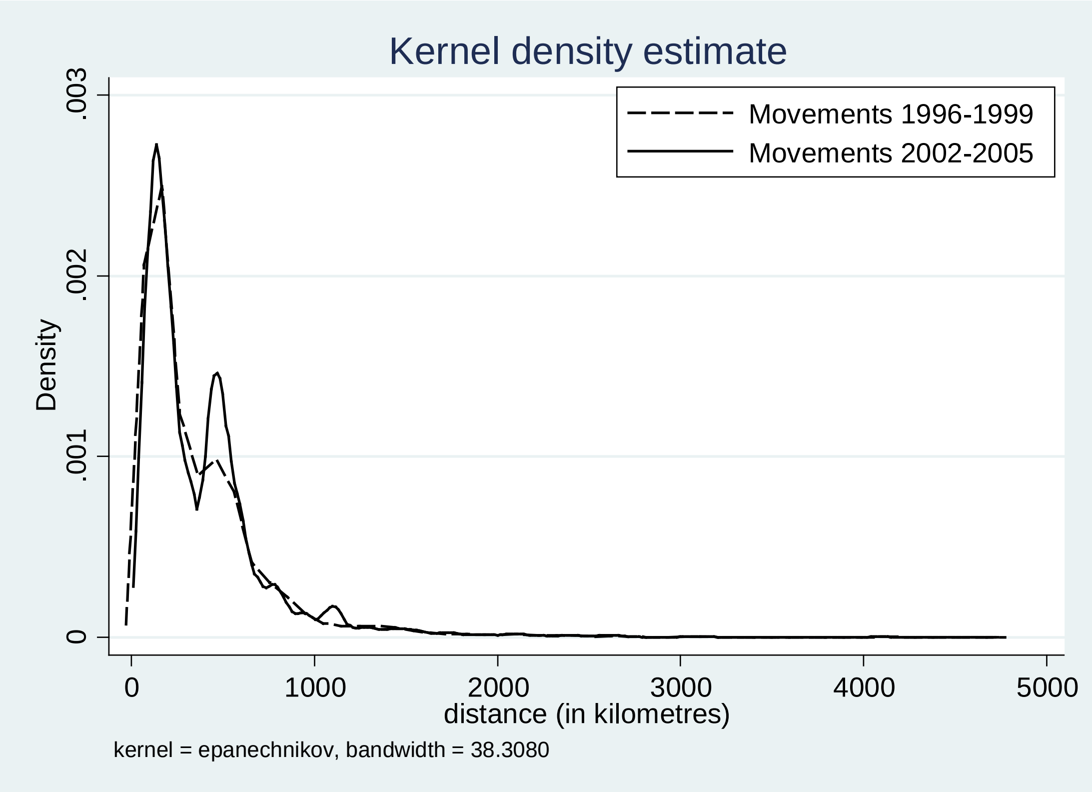


Figure 3. Patent intensity similarity between origin and destination regions

3.1. Movements 1996-1999 **3.2. Movements 2002-2005**

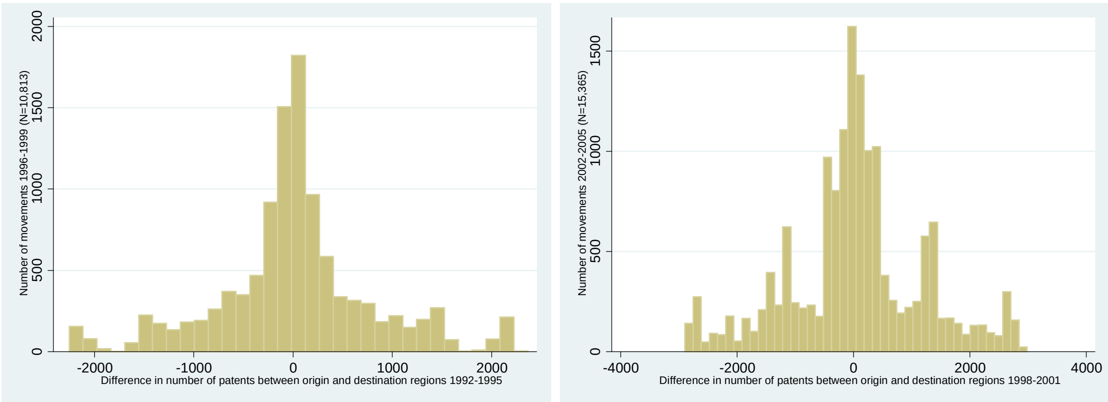


Table 2. Summary statistics

	Mean	St. Dev	Coef. Var.	Min.	Max.
Total movements 1996-1999	0.22	2.06	9.17	0	84
Total movements 2002-2005	0.32	5.09	15.96	0	467
Attributional variables					
BORDER_d	0.45	0.50	1.10	0	1
CENTRAL_d	640.32	475.46	0.74	10	2,400
INV9295_o	648.25	1,058.10	1.63	1	9,140
INV9801_o	1,040.30	1,629.25	1.57	1	12,766
INV9295_d	648.25	1,058.10	1.63	1	9,140
INV9801_d	1,040.30	1,629.25	1.57	1	12,766
HRST9295_d	28.28	8.63	0.31	7.73	55.05
HRST9801_d	32.50	8.07	0.25	11.88	55.30
POP9295_o	1,718,268	1,476,858	0.86	25,025	10,800,000
POP9801_o	1,747,665	1,500,628	0.86	25,625	11,000,000
POP9295_d	1718268	1,476,858	0.86	25,025	10,800,000
POP9801_d	1,747,665	1,500,628	0.86	25,625	11,000,000
DENS9295_d	354.47	842.97	2.38	3.17	8,163.25
DENS9801_d	359.07	857.72	2.39	3.14	8,497.49
TEMP9295_d	36.91	6.82	0.19	16.97	56.75
TEMP9801_d	40.83	6.47	0.16	20.66	57.38
COAST_d	0.54	0.50	0.93	0.00	1
Relational variables					
Contiguity	0.02	0.14	6.98	0	1
Euclidean distance	12.62	7.46	0.59	.06	44.60
Great circle distance	696.30	416.95	0.59	4.07	2,416.55
Km	1,524.76	910.27	0.59	8.06	5,545
Time	57,625.21	36,297	0.62	1,200	241,200
Social proximity 1992-1995	0.00	0.03	6.62	0	1
Social proximity 1998-2001	0.01	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: Data are not log transformed. See appendix 2 for the names of the variables. '_o' and '_d' stand for origin-region and destination-region variables respectively.

Table 3. Gravity model, ZINB estimations. Periods 1996-1999 & 2002-2005. Dependent variable: cross-regional pair-wise mobility of inventors.

	(i) km 96_99	(ii) time 96_99	(iii) km 02_05	(iv) time 02_05
Intercept	-10.62***	-5.40**	-13.21***	-6.34**
	(2.29)	(2.66)	(4.39)	(3.01)
Contiguity	0.92***	1.01***	0.92***	1.00***
	(0.09)	(0.09)	(0.10)	(0.10)
ln(Km)	-1.40***		-1.43***	
	(0.06)		(0.06)	
ln(Time)		-1.57***		-1.58***
		(0.07)		(0.07)
BORDER_d	0.11	0.08	0.21***	0.21***
	(0.07)	(0.07)	(0.07)	(0.07)
ln(CENTRAL_d)	-0.09	-0.07	0.04	0.06
	(0.11)	(0.11)	(0.14)	(0.13)
ln(INV_o)	0.51***	0.50***	0.70***	0.69***
	(0.05)	(0.05)	(0.04)	(0.04)
ln(INV_d)	0.48***	0.48***	0.67***	0.67***
	(0.06)	(0.06)	(0.05)	(0.05)
ln(HRST_d)	0.39*	0.35*	1.15**	1.13***
	(0.20)	(0.21)	(0.48)	(0.39)
ln(POP_o)	0.27***	0.25***	0.01	0.00
	(0.08)	(0.08)	(0.04)	(0.03)
ln(POP_d)	0.37***	0.37***	0.03	0.02
	(0.09)	(0.09)	(0.03)	(0.03)
ln(DENS_d)	-0.12***	-0.14***	-0.04	-0.06
	(0.04)	(0.04)	(0.05)	(0.06)
ln(TEMP_d)	0.36	0.35	0.69	0.80
	(0.44)	(0.45)	(0.62)	(0.56)
COAST_d	0.23***	0.23***	0.28**	0.32***
	(0.08)	(0.08)	(0.11)	(0.10)
ln(TECH.SHARES) [(1)]	yes	yes	yes	yes
Country Fixed Effects [(2)]	yes	yes	yes	yes
Sample size	48,180	48,180	48,180	48,180
Nonzero observations	2,854	2,854	3,365	3,365
Log-pseudolikelihood	-10,706.02	-10,666.66	-12,846.29	-12,809.8
LR test of α	4,509.92	4,410.46	1,400	1,300
p-value	0.0000	0.0000	0.0000	0.0000
Vuong statistic	12.54	12.46	10.97	10.83
p-value	0.0000	0.0000	0.0000	0.0000
Adjusted McFadden's R2	0.338	0.340	0.318	0.319

Notes: Robust standard errors are presented in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1, + p<0.12. Only the negative binomial estimation is shown here. 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' stand for origin-region and destination-region variables, respectively. **(1)** Inventors are assigned to each technological sector according to the classification produced jointly y 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. Inventors are assigned to sectors according to the majority of the IPC codes of their patent portfolio. These control variables are included in all the estimations unless otherwise stated. **(2)** The UK is treated as the reference country.

Table 4. Gravity model, ZINB estimations. Periods 1996-1999 & 2002-2005. Dependent variable: cross-regional pair-wise mobility of inventors.

	(i) km 96_99	(ii) time 96_99	(iii) km 02_05	(iv) time 02_05
Intercept	-12.80***	-9.73***	-14.34***	-11.47***
	(2.19)	(2.21)	(3.24)	(3.24)
Contiguity	0.92***	0.99***	0.85***	0.90***
	(0.08)	(0.08)	(0.08)	(0.08)
ln(Km)	-0.60***		-0.62***	
	(0.06)		(0.07)	
ln(Time)		-0.63***		-0.68***
		(0.07)		(0.08)
Institutional distance	-0.65***	-0.64***	-0.47***	-0.46***
	(0.11)	(0.11)	(0.10)	(0.10)
ln(Social Proximity)	0.12***	0.13***	0.16***	0.16***
	(0.02)	(0.02)	(0.02)	(0.02)
ln(Technological Distance)	-0.16**	-0.16**	-0.15**	-0.16***
	(0.07)	(0.07)	(0.06)	(0.06)
ln(Cultural Proximity)	0.05**	0.04**	0.05*	0.05*
	(0.02)	(0.02)	(0.03)	(0.03)
Research Excellence	-0.03	-0.02	0.17**	0.17**
	(0.06)	(0.06)	(0.07)	(0.07)
BORDER_d	0.23***	0.23***	0.24***	0.23***
	(0.07)	(0.07)	(0.08)	(0.08)
ln(CENTRAL_d)	-0.13	-0.14	-0.11	-0.11
	(0.11)	(0.11)	(0.11)	(0.11)
ln(INV_o)	0.56***	0.56***	0.69***	0.68***
	(0.05)	(0.05)	(0.04)	(0.04)
ln(INV_d)	0.41***	0.40***	0.55***	0.55***
	(0.06)	(0.06)	(0.04)	(0.04)
ln(HRST_d)	0.23	0.23	0.63*	0.65*
	(0.20)	(0.20)	(0.34)	(0.34)
ln(POP_o)	0.12	0.10	-0.02	-0.02
	(0.07)	(0.08)	(0.03)	(0.03)
ln(POP_d)	0.24**	0.23**	0.07*	0.07*
	(0.09)	(0.09)	(0.03)	(0.03)
ln(DENS_d)	-0.09**	-0.09**	-0.06	-0.06^{+}
	(0.04)	(0.04)	(0.04)	(0.04)
ln(TEMP_d)	0.70^{+}	0.69^{+}	1.23**	1.25**
	(0.43)	(0.44)	(0.60)	(0.59)
COAST_d	0.13*	0.13*	0.26***	0.27***
	(0.08)	(0.08)	(0.08)	(0.08)
ln(TECH.SHARES)	yes	yes	yes	yes
Country Fixed Effects [(1)]	yes	yes	yes	yes
Sample size	48,180	48,180	48,180	48,180
Nonzero observations	2,854	2,854	3,365	3,365
Log-pseudolikelihood	-9,917.573	-9,928.545	-11,986.95	-11,988.82
LR test of α	3,116.03	3,128.56	1,200	1,200
p-value	0.0000	0.0000	0.0000	0.0000
Vuong statistic	9.31	9.35	10.72	10.70
p-value	0.0000	0.0000	0.0000	0.0000
Adjusted McFadden's R2	0.385	0.385	0.362	0.362

Notes: Robust standard errors are presented in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1, + p<0.12. Only the negative binomial estimation is shown here. 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' stand for origin-region and destination-region variables respectively. **(1)** The UK is treated as the reference country.

Table 5. Gravity model, ZINB estimations. Period 2002-2005. Dependent variable: cross-regional pair-wise mobility of inventors - labour mobility only.

	(i) firm mobility km 02_05	(ii) firm mobility time 02_05	(iii) group mobility km 02_05	(iv) group mobility time 02_05
Intercept	-16.97***	-14.18***	-18.33***	-15.30***
	(2.63)	(2.70)	(3.17)	(3.29)
Contiguity	0.77***	0.82***	0.81***	0.86***
	(0.09)	(0.08)	(0.09)	(0.09)
ln(Km)	-0.61***		-0.61***	
	(0.08)		(0.08)	
ln(Time)		-0.67***		-0.67***
		(0.08)		(0.09)
Institutional distance	-0.48***	-0.47***	-0.38***	-0.37***
	(0.11)	(0.11)	(0.12)	(0.12)
ln(Social Proximity)	0.15***	0.15***	0.15***	0.16***
	(0.02)	(0.02)	(0.03)	(0.03)
ln(Technological Distance)	-0.17***	-0.18***	-0.20***	-0.21***
	(0.06)	(0.06)	(0.06)	(0.06)
ln(Cultural Proximity)	0.08***	0.07***	0.06***	0.06**
	(0.02)	(0.02)	(0.02)	(0.02)
Research Excellence	0.18**	0.18**	0.25***	0.24***
	(0.08)	(0.08)	(0.08)	(0.08)
BORDER_d	0.32***	0.32***	0.32***	0.32***
	(0.07)	(0.07)	(0.08)	(0.08)
ln(CENTRAL_d)	-0.02	-0.01	-0.03	-0.03
	(0.10)	(0.11)	(0.11)	(0.11)
ln(INV_o)	0.71***	0.71***	0.66***	0.65***
	(0.04)	(0.04)	(0.04)	(0.04)
ln(INV_d)	0.59***	0.58***	0.58***	0.58***
	(0.05)	(0.05)	(0.06)	(0.06)
ln(HRST_d)	0.56^{+}	0.59^{+}	0.56	0.59
	(0.36)	(0.36)	(0.42)	(0.42)
ln(POP_o)	-0.02	-0.03	-0.02	-0.02
	(0.03)	(0.03)	(0.04)	(0.04)
ln(POP_d)	0.08**	0.08**	0.06	0.05
	(0.03)	(0.03)	(0.04)	(0.04)
ln(DENS_d)	-0.08*	-0.08*	-0.06	-0.06
	(0.04)	(0.05)	(0.05)	(0.05)
ln(TEMP_d)	1.70***	1.71***	1.80***	1.76***
	(0.48)	(0.51)	(0.58)	(0.59)
ln(COAST_d)	0.19**	0.20**	0.18*	0.20**
	(0.08)	(0.08)	(0.10)	(0.10)
ln(TECH.SHARES)	yes	yes	yes	yes
Country Fixed Effects **(1)**	yes	yes	yes	yes
Sample size	48,180	48,180	48,180	48,180
Nonzero observations	2,812	2,812	2,391	2,391
Log-pseudolikelihood	-9,930.056	-9,932.441	-8,552.124	-8,553.164
LR test of α	6,699.87	6,670.30	5,353.87	5,323.15
p-value	0.000	0.0000	0.0000	0.0000
Vuong statistic	9.55	9.57	8.83	8.85
p-value	0.000	0.0000	0.0000	0.0000
Adjusted McFadden's R2	0.368	0.368	0.372	0.372

Notes: Robust standard errors are presented in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1, + p<0.12. Only the negative binomial estimation is shown here. 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' stand for origin-region and destination-region variables, respectively. **(1)** The UK is treated as the reference country.

Table 6. Gravity model, ZINB estimations. Period 2002-2005. Dependent variable: cross-regional pair-wise mobility of inventors - endogeneity.

	(i) time 1st lag depvar	(ii) time 2n lag depvar	(iii) time 3rd lag depvar	(iv) time 2SRI
Intercept	-8.25***	-8.26***	-12.18***	-12.62***
	(2.18)	(2.34)	(3.22)	(4.53)
Contiguity	0.60***	0.52***	0.87***	0.94***
	(0.08)	(0.08)	(0.08)	(0.15)
Lag Dependent var. 98-01	0.05***			
	(0.01)			
Lag Dependent var. 94-97		0.07***		
		(0.01)		
Lag Dependent var. 90-93			0.00	
			(0.00)	
ln(Time)	-0.38***	-0.54***	-0.68***	-0.70***
	(0.06)	(0.08)	(0.08)	(0.13)
Institutional distance	-0.30***	-0.35***	-0.44***	-0.56*
	(0.10)	(0.09)	(0.10)	(0.31)
ln(Social Proximity)	0.15***	0.14***	0.15***	0.13*
	(0.02)	(0.02)	(0.02)	(0.07)
ln(Technological Distance)	-0.11*	-0.10*	-0.16**	-0.19*
	(0.06)	(0.06)	(0.06)	(0.10)
ln(Cultural Proximity)	0.02	0.01	0.04**	0.06*
	(0.02)	(0.02)	(0.02)	(0.03)
Research Excellence	0.10	0.02	0.16**	0.18**
	(0.08)	(0.07)	(0.07)	(0.09)
BORDER_d	0.18***	0.13**	0.25***	0.25***
	(0.06)	(0.07)	(0.08)	(0.09)
ln(CENTRAL_d)	-0.12	-0.11	-0.10	-0.10
	(0.10)	(0.10)	(0.11)	(0.13)
ln(INV_o)	0.47***	0.50***	0.63***	0.69***
	(0.04)	(0.04)	(0.05)	(0.11)
ln(INV_d)	0.33***	0.35***	0.52***	0.57***
	(0.05)	(0.05)	(0.05)	(0.12)
ln(HRST_d)	0.54	0.94***	0.78**	0.72**
	(0.37)	(0.32)	(0.38)	(0.37)
ln(POP_o)	-0.02	-0.06*	-0.03	-0.02
	(0.03)	(0.03)	(0.03)	(0.04)
ln(POP_d)	0.07**	0.03	0.07*	0.07*
	(0.03)	(0.03)	(0.03)	(0.04)
ln(DENS_d)	-0.01	-0.06	-0.07+	-0.07*
	(0.04)	(0.04)	(0.04)	(0.04)
ln(TEMP_d)	0.87**	0.95**	1.39***	1.30**
	(0.38)	(0.39)	(0.49)	(0.57)
COAST_d	0.22***	0.21***	0.29***	0.26***
	(0.07)	(0.07)	(0.08)	(0.09)
Control term				0.03
				(0.08)
ln(TECH.SHARES)	yes	yes	yes	yes
Country Fixed Effects [(1)]	yes	yes	yes	yes
Sample size	48,180	48,180	48,180	48,180
Nonzero observations	3,365	3,365	3,365	2,391
Partial R2 first stage				0.513
F-stat first stage				23.82
Adjusted McFadden's R2	0.384	0.379	0.365	0.362

Notes: Robust standard errors are presented in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1, + p<0.12. Only the negative binomial estimation is shown here. '_o' and '_d' stand for origin-region and destination-region variables, respectively. **(1)** The UK is treated as the reference country. Standard errors in (iv) are calculated via bootstrapping with 1000 iterations.

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), Luxembourg (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, and PATSTAT-KITeS	
Geographical distance	Euclidean distance between UTM regional centroids		GIS	-
Geographical distance	Great circle distance		GIS	-
Geographical distance	Driving distance in km		Google Maps and SAS	-
Geographical distance	Driving distance in time (seconds)		Google Maps and SAS	-
Contiguity	1: contiguity; 0 otherwise		GIS	-
Institutional distance	1: dif. country; 0 otherwise			-
Social proximity	$A_{ij} = l_{ij} / n_i$	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	Border with a foreign country		ESPON	+
Time2Brussels_d	Time (in seconds) from the regions' centroids to Brussels		Google Maps and SAS	-
HRST_d	Human Resource in Science and Technology (core) over active population	Average 92-95 98-01	Eurostat	+
Population Density_d	Population over area (km2)	Average 92-95 98-01	Eurostat	?
Average temperature_d	Average temperature in January (degress Fahrenheit)	Average 92-95 98-01	FOODSEC project, MARS units, EC-JRC	+
Coast_d	1: if the region has a coast; 0 otherwise		ESPON	+

Notes: '_o' and '_d' stand for origin-region and destination-region variables, respectively.

Appendix 3: Complementary results and robustness checks

Table A. 3.1. Gravity model, ZINB estimations. Periods 1996-1999. Dependent variable: cross-regional pair-wise mobility of inventors.

	(i) Euclidean		**(ii) Great circle**		**(iii) km**		**(iv) time**	
	NegBin	Logit	NegBin	Logit	NegBin	Logit	NegBin	Logit
Intercept	-19.33***	-0.73	-14.90***	-2.03	-10.62***	6.62	-5.40**	-5.56
	(2.46)	(6.38)	(2.53)	(6.04)	(2.29)	(6.39)	(2.66)	(6.39)
Contiguity	0.92***	-1.33***	0.88***	-1.04**	0.92***	-0.98**	1.01***	-1.01**
	(0.10)	(0.40)	(0.09)	(0.42)	(0.09)	(0.43)	(0.09)	(0.47)
ln(Euclidean Distance)	-1.36***	0.94***						
	(0.07)	(0.21)						
ln(Arc Distance)			-1.41***	1.07***				
			(0.06)	(0.17)				
ln(Km)					-1.40***	1.25***		
					(0.06)	(0.18)		
ln(Time)							-1.57***	1.39***
							(0.07)	(0.21)
BORDER_d	0.17**	0.07	0.11^{+}	-0.19	0.11	-0.25	0.08	-0.20
	(0.07)	(0.18)	(0.07)	(0.18)	(0.07)	(0.18)	(0.07)	(0.18)
ln(CENTRAL_d)	-0.08	-0.34	-0.08	-0.96***	-0.09	-1.17***	-0.07	-1.21***
	(0.13)	(0.35)	(0.11)	(0.27)	(0.11)	(0.26)	(0.11)	(0.27)
ln(INV_o)	0.51***	-0.39***	0.52***	-0.39**	0.51***	-0.38**	0.50***	-0.31*
	(0.05)	(0.14)	(0.05)	(0.17)	(0.05)	(0.17)	(0.05)	(0.18)
ln(INV_d)	0.51***	-0.54***	0.49***	-0.78***	0.48***	-0.79***	0.48***	-0.74***
	(0.06)	(0.17)	(0.06)	(0.17)	(0.06)	(0.17)	(0.06)	(0.17)
ln(HRST_d)	0.38*	0.51	0.32	0.43	0.39*	0.42	0.35*	0.47
	(0.20)	(0.61)	(0.20)	(0.57)	(0.20)	(0.63)	(0.21)	(0.62)
ln(POP_o)	0.30***	-0.26	0.28***	-0.30	0.27***	-0.27	0.25***	-0.41
	(0.08)	(0.24)	(0.08)	(0.25)	(0.08)	(0.26)	(0.08)	(0.27)
ln(POP_d)	0.40***	0.15	0.39***	0.18	0.37***	0.20	0.37***	0.07
	(0.09)	(0.26)	(0.09)	(0.27)	(0.09)	(0.29)	(0.09)	(0.29)
ln(DENS_d)	-0.08**	0.08	-0.11***	-0.03	-0.12***	0.02	-0.14***	-0.04
	(0.04)	(0.12)	(0.04)	(0.12)	(0.04)	(0.12)	(0.04)	(0.12)
ln(TEMP_d)	0.55	1.57	0.46	0.92	0.36	0.95	0.35	0.83
	(0.46)	(1.15)	(0.44)	(1.07)	(0.44)	(1.06)	(0.45)	(1.05)

COAST_d	0.31***	0.09	0.19**	-0.21	0.23***	-0.34	0.23***	-0.26
	(0.08)	(0.26)	(0.08)	(0.25)	(0.08)	(0.27)	(0.08)	(0.28)
ln(TECH.SHARES) (1)	yes	yes	yes	yes	yes	yes	yes	yes
Country 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	-10,715.65		-10,709.25		-10706.02		-10,666.66	
LR test	11,256.700		11,269.494		11,275.954			
p-value	0.0000		0.0000		0.0000		0.0000	
Wald test	5,600.92		5,321.96		5,524.86		4,992.81	
p-value	0.0000		0.0000		0.0000		0.0000	
LR test of α	4,528.81		4,585.39		4,509.92		4,410.46	
p-value	0.0000		0.0000		0.0000		0.0000	
Vuong statistic	12.02		12.52		12.54		12.46	
p-value	0.0000		0.0000		0.0000		0.0000	
McFadden’s R2	0.344		0.345		0.345		0.347	
Adjusted McFadden’s R2	0.337		0.337		0.338		0.340	
AIC	21,669.3		21,656.5		21,650.04		21,571.31	
Schwartz	22,714.44		22,701.64		22,695.18		22,616.46	

Notes: Robust standard errors are presented in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1, + p<0.12. Only the negative binomial estimation is shown here. 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’ stand for origin-region and destination-region variables, respectively. **(1)** Inventors are assigned to each technological sector 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. Inventors are assigned to sectors according to the majority of the IPC codes of their patent portfolio. These control variables are included in all the estimations unless otherwise stated. **(2)** The UK is treated as the reference country.

Table A. 3.2. Gravity model, ZINB estimations. Periods 2002-2005. Dependent variable: cross-regional pair-wise mobility of inventors.

	(i) Euclidean		**(ii) Great circle**		**(iii) km**		**(iv) time**	
	NegBin	Logit	NegBin	Logit	NegBin	Logit	NegBin	Logit
Intercept	-22.72***	-8.58*	-13.45***	0.62	-13.21***	-3.17	-6.34**	-8.66
	(2.42)	(4.84)	(3.01)	(5.20)	(4.39)	(10.18)	(3.01)	(5.40)
Contiguity	0.98***	-1.38***	0.88***	-1.33***	0.92***	-1.36***	1.00***	-1.39***
	(0.10)	(0.31)	(0.10)	(0.34)	(0.10)	(0.40)	(0.10)	(0.34)
ln(Euclidean Distance)	-1.31***	1.02***						
	(0.06)	(0.13)						
ln(Arc Distance)			-1.39***	1.04***				
			(0.06)	(0.16)				
ln(Km)					-1.43***	1.05***		
					(0.06)	(0.18)		
ln(Time)							-1.58***	1.20***
							(0.07)	(0.16)
BORDER_d	0.29***	0.48**	0.22***	0.33*	0.21***	0.33*	0.21***	0.38**
	(0.08)	(0.19)	(0.07)	(0.18)	(0.07)	(0.18)	(0.07)	(0.18)
ln(CENTRAL_d)	0.10	-0.17	0.00	-0.60**	0.04	-0.59**	0.06	-0.63**
	(0.12)	(0.26)	(0.13)	(0.25)	(0.14)	(0.26)	(0.13)	(0.25)
ln(INV_o)	0.72***	-0.51***	0.72***	-0.59***	0.70***	-0.58***	0.69***	-0.56***
	(0.04)	(0.07)	(0.04)	(0.07)	(0.04)	(0.07)	(0.04)	(0.07)
ln(INV_d)	0.78***	-0.33***	0.67***	-0.59***	0.67***	-0.58***	0.67***	-0.58***
	(0.05)	(0.10)	(0.05)	(0.10)	(0.05)	(0.10)	(0.05)	(0.11)
ln(HRST_d)	0.89**	-1.05	1.04***	-0.45	1.15**	-0.28	1.13***	-0.29
	(0.35)	(0.88)	(0.40)	(0.71)	(0.48)	(0.99)	(0.39)	(0.70)
ln(POP_o)	-0.04	-0.03	-0.00	-0.00	0.01	-0.01	0.00	0.01
	(0.04)	(0.10)	(0.04)	(0.09)	(0.04)	(0.10)	(0.03)	(0.09)
ln(POP_d)	0.09**	0.12	0.02	-0.08	0.03	-0.09	0.02	-0.09
	(0.04)	(0.08)	(0.03)	(0.07)	(0.03)	(0.07)	(0.03)	(0.08)
ln(DENS_d)	-0.07^{+}	-0.06	-0.02	0.10	-0.04	0.11	-0.06	0.11
	(0.04)	(0.11)	(0.05)	(0.09)	(0.05)	(0.10)	(0.06)	(0.11)
ln(TEMP_d)	1.52***	3.53***	0.60	0.66	0.69	1.15	0.80	1.25
	(0.44)	(1.05)	(0.57)	(1.10)	(0.62)	(1.45)	(0.56)	(1.07)
COAST_d	0.37***	0.17	0.29***	-0.02	0.28**	-0.04	0.32***	0.01

	(0.08)	(0.19)	(0.10)	(0.23)	(0.11)	(0.27)	(0.10)	(0.24)
ln(TECH.SHARES) [(1)]	yes	yes	yes	yes	yes	Yes	Yes	yes
Country 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	-12,882.33		-12,881.26		-12,846.29		-12,809.8	
LR test	12,229.862		12,232.001		12,301.944		12,374.924	
p-value	0.0000		0.000		0.0000		0.0000	
Wald test	4,607.51		4,491.59		4,568.44		4,591.74	
p-value	0.0000		0.0000		0.0000		0.0000	
LR test of α	1,400		1,400		1,400		1,300	
p-value	0.0000		0.0000		0.0000		0.0000	
Vuong statistic	11.48		11.06		10.97		10.83	
p-value	0.0000		0.0000		0.0000		0.0000	
McFadden's R2	0.322		0.322		0.324		0.326	
Adjusted McFadden's R2	0.316		0.316		0.318		0.319	
AIC	26,002.66		26,000.52		25,930.57		25,857.6	
Schwartz	27,047.8		27,045.66		26,975.72		26,902.74	

Notes: Robust standard errors are presented in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1, + p<0.12. Only the negative binomial estimation is shown here. 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' stand for origin-region and destination-region variables, respectively. **(1)** Inventors are assigned to each technological sector 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. Inventors are assigned to sectors according to the majority of the IPC codes of their patent portfolio. These control variables are included in all the estimations unless otherwise stated. **(2)** The UK is treated as the reference country.

Table A. 3.3. Gravity model, ZINB estimations. Periods 1996-1999. Dependent variable: cross-regional pair-wise mobility of inventors.

	(i) Euclidean		**(ii) Great circle**		**(iii) km**		**(iv) time**	
	NegBin	Logit	NegBin	Logit	NegBin	Logit	NegBin	Logit
Intercept	-14.01***	-7.78	-13.59***	-14.91**	-12.80***	-14.94**	-9.73***	-15.16**
	(2.32)	(7.00)	(2.19)	(5.85)	(2.19)	(5.82)	(2.21)	(6.10)
Contiguity	0.94***	-0.53	0.90***	-0.75*	0.92***	-0.82**	0.99***	-0.88**
	(0.08)	(0.41)	(0.08)	(0.38)	(0.08)	(0.39)	(0.08)	(0.39)
ln(Euclidean Distance)	-0.54***	0.26*						
	(0.06)	(0.15)						
ln(Arc Distance)			-0.59***	0.04				
			(0.06)	(0.15)				
ln(Km)					-0.60***	0.02		
					(0.06)	(0.15)		
ln(Time)							-0.63***	0.06
							(0.07)	(0.17)
Institutional distance	-0.61***	4.54***	-0.66***	4.84***	-0.65***	4.82***	-0.64***	5.00***
	(0.11)	(1.33)	(0.11)	(0.49)	(0.11)	(0.49)	(0.11)	(0.54)
ln(Social Proximity)	0.13***	-0.01	0.12***	-0.04*	0.12***	-0.04*	0.13***	-0.04
	(0.01)	(0.03)	(0.01)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
ln(Technological Distance)	-0.05	0.93***	-0.15**	0.51**	-0.16**	0.49**	-0.16**	0.49**
	(0.07)	(0.23)	(0.07)	(0.20)	(0.07)	(0.20)	(0.07)	(0.20)
ln(Cultural Proximity)	0.05**	-0.43***	0.05***	-0.19*	0.05**	-0.20*	0.04**	-0.24**
	(0.02)	(0.13)	(0.02)	(0.11)	(0.02)	(0.11)	(0.02)	(0.10)
Research Excellence	0.03	0.23	-0.02	0.09	-0.03	0.08	-0.02	0.09
	(0.06)	(0.18)	(0.06)	(0.17)	(0.06)	(0.17)	(0.06)	(0.17)
BORDER_d	0.29***	0.51***	0.24***	0.39**	0.23***	0.38**	0.23***	0.41**
	(0.07)	(0.19)	(0.07)	(0.18)	(0.07)	(0.18)	(0.07)	(0.19)
ln(CENTRAL_d)	-0.13	0.02	-0.14	-0.02	-0.13	-0.01	-0.14	-0.04
	(0.11)	(0.24)	(0.11)	(0.23)	(0.11)	(0.23)	(0.11)	(0.24)
ln(INV_o)	0.58***	-0.34**	0.57***	-0.31**	0.56***	-0.32**	0.56***	-0.31**
	(0.05)	(0.15)	(0.05)	(0.13)	(0.05)	(0.13)	(0.05)	(0.13)
ln(INV_d)	0.41***	-0.65***	0.41***	-0.71***	0.41***	-0.70***	0.40***	-0.73***
	(0.06)	(0.18)	(0.06)	(0.16)	(0.06)	(0.16)	(0.06)	(0.17)
ln(HRST_d)	0.30	0.76	0.22	0.54	0.23	0.54	0.23	0.58
	(0.20)	(0.54)	(0.20)	(0.49)	(0.20)	(0.49)	(0.20)	(0.51)
ln(POP_o)	0.14*	-0.06	0.13*	-0.10	0.12	-0.09	0.10	-0.14

	(0.07)	(0.21)	(0.07)	(0.20)	(0.07)	(0.20)	(0.08)	(0.21)
ln(POP_d)	0.21**	0.32	0.25***	0.47*	0.24**	0.46*	0.23**	0.48*
	(0.09)	(0.28)	(0.09)	(0.26)	(0.09)	(0.25)	(0.09)	(0.26)
ln(DENS_d)	-0.04	0.12	-0.08**	-0.05	-0.09**	-0.06	-0.09**	-0.06
	(0.04)	(0.12)	(0.04)	(0.10)	(0.04)	(0.10)	(0.04)	(0.11)
ln(TEMP_d)	0.76*	1.86	0.71*	1.90*	0.70^{+}	1.93*	0.69^{+}	1.92*
	(0.45)	(1.35)	(0.43)	(1.13)	(0.43)	(1.12)	(0.44)	(1.16)
COAST_d	0.17**	-0.32	0.13*	-0.46**	0.13*	-0.46**	0.13*	-0.51**
	(0.08)	(0.25)	(0.08)	(0.23)	(0.08)	(0.23)	(0.08)	(0.24)
ln(TECH.SHARES)	yes	yes	Yes	yes	yes	Yes	yes	yes
Country Fixed Effects [(1)]	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	-9,920.899		-9,915.472		-9,917.573		-9,928.545	
LR test	12,846.199		12,857.054		12,852.851		12,830.908	
p-value	0.0000		0.0000		0.0000		0.0000	
Wald test	4,643.54		3,629.38		3,585.09		3,590.69	
p-value	0.0000		0.0000		0.0000		0.0000	
LR test of α	3,104.78		3,117.13		3,116.03		3,128.56	
p-value	0.0000		0.0000		0.0000		0.0000	
Vuong statistic	9.49		9.28		9.31		9.35	
p-value	0.0000		0.0000		0.0000		0.0000	
McFadden's R2	0.393		0.393		0.393		0.393	
Adjusted McFadden's R2	0.385		0.385		0.385		0.385	
AIC	20,,099.8		20,088.94		20,093.15		20,115.09	
Schwartz	21232.77		21,221.91		21,226.11		21,248.06	

Notes: Robust standard errors are presented in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1, + p<0.12. 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' stand for origin-region and destination-region variables, respectively. **(1)** The UK is treated as the reference country.

Table A. 3.4. Gravity model, ZINB estimations. Periods 2002-2005. Dependent variable: cross-regional pair-wise mobility of inventors.

	(i) Euclidean		**(ii) Great circle**		**(iii) km**		**(iv) time**	
	NegBin	Logit	NegBin	Logit	NegBin	Logit	NegBin	Logit
Intercept	-17.55***	-9.57	-14.88***	-8.25	-14.34***	-8.68	-11.47***	-8.84
	(3.38)	(9.19)	(3.17)	(8.68)	(3.24)	(9.11)	(3.24)	(8.86)
Contiguity	0.92***	-1.21***	0.84***	-1.33***	0.85***	-1.37***	0.90***	-1.37***
	(0.09)	(0.38)	(0.08)	(0.37)	(0.08)	(0.38)	(0.08)	(0.39)
ln(Euclidean Distance)	-0.49***	0.15						
	(0.08)	(0.15)						
ln(Arc Distance)			-0.59***	-0.03				
			(0.07)	(0.13)				
ln(Km)					-0.62***	-0.04		
					(0.07)	(0.14)		
ln(Time)							-0.68***	-0.02
							(0.08)	(0.15)
Institutional distance	-0.49***	4.31***	-0.48***	4.60***	-0.47***	4.58***	-0.46***	4.55***
	(0.11)	(0.38)	(0.10)	(0.40)	(0.10)	(0.40)	(0.10)	(0.41)
ln(Social Proximity)	0.17***	0.05	0.16***	0.03	0.16***	0.03	0.16***	0.03
	(0.02)	(0.04)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
ln(Technological Distance)	-0.13**	0.44***	-0.14**	0.44***	-0.15**	0.43***	-0.16***	0.41***
	(0.06)	(0.14)	(0.06)	(0.14)	(0.06)	(0.14)	(0.06)	(0.14)
ln(Cultural Proximity)	0.06**	-0.45***	0.06**	-0.45***	0.05*	-0.46***	0.05*	-0.47***
	(0.03)	(0.11)	(0.03)	(0.10)	(0.03)	(0.10)	(0.03)	(0.10)
Research Excellence	0.19***	0.02	0.18**	0.02	0.17**	0.01	0.17**	0.01
	(0.07)	(0.16)	(0.07)	(0.15)	(0.07)	(0.16)	(0.07)	(0.16)
BORDER_d	0.26***	0.40	0.24***	0.32	0.24***	0.32	0.23***	0.32
	(0.09)	(0.25)	(0.08)	(0.22)	(0.08)	(0.23)	(0.08)	(0.23)
ln(CENTRAL_d)	-0.15	-0.05	-0.12	-0.01	-0.11	-0.02	-0.11	-0.04
	(0.11)	(0.24)	(0.11)	(0.23)	(0.11)	(0.24)	(0.11)	(0.24)
ln(INV_o)	0.70***	-0.45***	0.70***	-0.45***	0.69***	-0.45***	0.68***	-0.45***
	(0.04)	(0.08)	(0.04)	(0.08)	(0.04)	(0.08)	(0.04)	(0.08)
ln(INV_d)	0.54***	-0.46***	0.55***	-0.41***	0.55***	-0.41***	0.55***	-0.41***
	(0.05)	(0.13)	(0.04)	(0.09)	(0.04)	(0.09)	(0.04)	(0.09)
ln(HRST_d)	0.59*	-0.04	0.61*	-0.07	0.63*	-0.01	0.65*	0.03
	(0.35)	(0.72)	(0.34)	(0.71)	(0.34)	(0.74)	(0.34)	(0.74)
ln(POP_o)	-0.03	0.07	-0.02	0.07	-0.02	0.08	-0.02	0.07

	(0.03)	(0.08)	(0.03)	(0.08)	(0.03)	(0.08)	(0.03)	(0.08)
ln(POP_d)	0.07*	0.04	0.07*	0.03	0.07*	0.03	0.07*	0.03
	(0.04)	(0.08)	(0.03)	(0.08)	(0.03)	(0.08)	(0.03)	(0.08)
ln(DENS_d)	-0.04	-0.00	-0.04	-0.00	-0.06	-0.01	-0.06^{+}	-0.02
	(0.04)	(0.09)	(0.04)	(0.09)	(0.04)	(0.09)	(0.04)	(0.09)
ln(TEMP_d)	1.41**	2.57	1.23**	2.09	1.23**	2.18	1.25**	2.21
	(0.65)	(1.77)	(0.59)	(1.63)	(0.60)	(1.71)	(0.59)	(1.68)
COAST_d	0.24***	-0.10	0.26***	-0.03	0.26***	-0.03	0.27***	-0.01
	(0.09)	(0.23)	(0.07)	(0.19)	(0.08)	(0.19)	(0.08)	(0.19)
ln(TECH.SHARES)	yes	yes	yes	yes	yes	Yes	yes	yes
Country Fixed Effects **(1)**	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	-12,007.38		-11,993.06		-11,986.95		-11,988.82	
LR test	13,979.767		14,008.409		14,020.622		14,016.886	
p-value	0.0000		0.0000		0.0000		0.0000	
Wald test	3,296.87		3,605.62		3,620.31		3,620.92	
p-value	0.0000		0.0000		0.0000		0.0000	
LR test of α	1,200		1,200		1,200		1,200	
p-value	0.0000		0.0000		0.0000		0.0000	
Vuong statistic	10.89		10.77		10.72		10.70	
p-value	0.0000		0.0000		0.0000		0.0000	
McFadden’s R2	0.368		0.369		0.369		0.369	
Adjusted McFadden’s R2	0.361		0.362		0.362		0.362	
AIC	24,272.75		24,244.11		24,231.9		24,235.63	
Schwartz	25,405.72		25,377.08		25,364.86		25,368.6	

Notes: Robust standard errors are presented in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1, + p<0.12. 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’ stand for origin-region and destination-region variables, respectively. **(1)** The UK is treated as the reference country.

Table A. 3.5. Gravity model, ZINB estimations. Periods 2002-2005. Dependent variable: cross-regional pair-wise mobility of inventors – only labour mobility.

	(i) km firm mob.		**(ii) time firm mob.**		**(iii) km group mob.**		**(iv) time group mob.**	
	NegBin	Logit	NegBin	Logit	NegBin	Logit	NegBin	Logit
Intercept	-16.97***	-12.38*	-14.18***	-12.45*	-18.33***	-16.15**	-15.30***	-16.13**
	(2.63)	(7.30)	(2.70)	(7.52)	(3.17)	(7.35)	(3.29)	(7.52)
Contiguity	0.77***	-1.40***	0.82***	-1.38***	0.81***	-1.12***	0.86***	-1.10***
	(0.09)	(0.42)	(0.08)	(0.52)	(0.09)	(0.36)	(0.09)	(0.36)
ln(km)	-0.61***	-0.04			-0.61***	0.05		
	(0.08)	(0.17)			(0.08)	(0.17)		
ln(Time)			-0.67***	-0.02			-0.67***	0.08
			(0.08)	(0.18)			(0.09)	(0.19)
Institutional distance	-0.48***	4.19***	-0.47***	4.18***	-0.38***	3.56***	-0.37***	3.53***
	(0.11)	(0.51)	(0.11)	(0.50)	(0.12)	(0.43)	(0.12)	(0.44)
ln(Social Proximity)	0.15***	0.03	0.15***	0.03	0.15***	0.04	0.16***	0.04
	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)
ln(Technological Distance)	-0.17***	0.53***	-0.18***	0.51***	-0.20***	0.58***	-0.21***	0.55***
	(0.06)	(0.17)	(0.06)	(0.17)	(0.06)	(0.17)	(0.06)	(0.17)
ln(Cultural Proximity)	0.08***	-0.38***	0.07***	-0.39***	0.06***	-0.43***	0.06**	-0.43***
	(0.02)	(0.10)	(0.02)	(0.10)	(0.02)	(0.12)	(0.02)	(0.11)
Research Excellence	0.18**	0.00	0.18**	0.01	0.25***	0.09	0.24***	0.09
	(0.08)	(0.18)	(0.08)	(0.18)	(0.08)	(0.19)	(0.08)	(0.19)
BORDER_d	0.32***	0.47**	0.32***	0.47**	0.32***	0.63***	0.32***	0.63***
	(0.07)	(0.20)	(0.07)	(0.20)	(0.08)	(0.22)	(0.08)	(0.22)
ln(CENTRAL_d)	-0.02	-0.01	-0.01	-0.03	-0.03	0.06	-0.03	0.03
	(0.10)	(0.23)	(0.11)	(0.23)	(0.11)	(0.25)	(0.11)	(0.26)
ln(INV_o)	0.71***	-0.35***	0.71***	-0.35***	0.66***	-0.44***	0.65***	-0.43***
	(0.04)	(0.08)	(0.04)	(0.09)	(0.04)	(0.09)	(0.04)	(0.09)
ln(INV_d)	0.59***	-0.39***	0.58***	-0.39***	0.58***	-0.44***	0.58***	-0.44***
	(0.05)	(0.12)	(0.05)	(0.13)	(0.06)	(0.12)	(0.06)	(0.12)
ln(HRST_d)	0.56^{+}	0.43	0.59^{+}	0.50	0.56	0.67	0.59	0.71
	(0.36)	(0.89)	(0.36)	(0.90)	(0.42)	(0.92)	(0.42)	(0.92)
ln(POP_o)	-0.02	0.03	-0.03	0.03	-0.02	0.04	-0.02	0.04
	(0.03)	(0.08)	(0.03)	(0.08)	(0.04)	(0.09)	(0.04)	(0.09)
ln(POP_d)	0.08**	0.05	0.08**	0.05	0.06	-0.02	0.05	-0.02
	(0.03)	(0.08)	(0.03)	(0.08)	(0.04)	(0.08)	(0.04)	(0.08)

ln(DENS_d)	-0.08*	-0.17	-0.08*	-0.17	-0.06	-0.09	-0.06	-0.09
	(0.04)	(0.11)	(0.05)	(0.12)	(0.05)	(0.11)	(0.05)	(0.11)
ln(TEMP_d)	1.70***	3.06**	1.71***	3.04**	1.80***	3.92***	1.76***	3.78***
	(0.48)	(1.43)	(0.51)	(1.42)	(0.58)	(1.38)	(0.59)	(1.39)
COAST_d	0.19**	-0.16	0.20**	-0.15	0.18*	-0.25	0.20**	-0.23
	(0.08)	(0.24)	(0.08)	(0.24)	(0.10)	(0.26)	(0.10)	(0.26)
ln(TECH.SHARES)	yes	yes	yes	yes	yes	Yes	yes	yes
Country Fixed Effects (1)	yes	yes	yes	yes	yes	Yes	yes	yes
Sample size	48,180		48,180		48,180		48,180	
Nonzero observations	2,812		2,812		2,391		2,391	
Log-pseudolikelihood	-9,930.056		-9,932.441		-8,552.124		-8,553.164	
LR test	11,979.563		11,974.793		10,555.194		10,553.114	
p-value	0.000		0.000		0.0000		0.0000	
Wald test	3,516.15		3,509.39		2,887.87		2,880.84	
p-value	0.000		0.000		0.0000		0.0000	
LR test of α	6,699.87		6,670.30		5,353.87		5,323.15	
p-value	0.000		0.0000		0.0000		0.0000	
Vuong statistic	9.55		9.57		8.83		8.85	
p-value	0.000		0.0000		0.0000		0.0000	
McFadden's R2	0.376		0.376		0.382		0.382	
Adjusted McFadden's R2	0.368		0.368		0.372		0.372	
AIC	20,118.11		20,122.88		17,362.25		17,364.33	
Schwartz	21,251.08		21,255.85		18,495.22		18,497.3	

Notes: Robust standard errors are presented in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1, + p<0.12. 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' stand for origin-region and destination-region variables, respectively. **(1)** The UK is treated as the reference country.

Table A.3.6. Gravity model, ZINB estimations. Periods 2002-2005. Dependent variable: cross-regional pair-wise mobility of inventors – endogeneity.

	(i) time 1st lag depvar		(ii) time 2n lag depvar		(iii) time 3rd lag depvar		(iv) time (no ZINB)	(v) time 2SRI
	NegBin	Logit	NegBin	Logit	NegBin	Logit	NegBin	NegBin
Intercept	-8.25***	-1.75	-8.26***	-0.10	-12.18***	-7.92	-12.62***	-11.44
	(2.18)	(4.01)	(2.34)	(4.48)	(3.22)	(6.90)	(4.53)	(10.72)
Contiguity	0.60***	-0.99***	0.52***	-1.32***	0.87***	-0.88**	0.94***	-1.15**
	(0.08)	(0.33)	(0.08)	(0.40)	(0.08)	(0.35)	(0.15)	(0.49)
Lag Dependent var. 1998-2001	0.05***	-1.37***						
	(0.01)	(0.12)						
Lag Dependent var. 1994-1997			0.07***	-1.19***				
			(0.01)	(0.16)				
Lag Dependent var. 1990-1993					0.00	-1.66***		
					(0.00)	(0.29)		
ln(Time)	-0.38***	0.26**	-0.54***	0.07	-0.68***	-0.02	-0.70***	-0.09
	(0.06)	(0.12)	(0.08)	(0.14)	(0.08)	(0.15)	(0.13)	(0.20)
Institutional distance	-0.30***	2.39***	-0.35***	2.75***	-0.44***	3.16***	-0.56*	3.90***
	(0.10)	(0.26)	(0.09)	(0.30)	(0.10)	(0.36)	(0.31)	(1.02)
ln(Social Proximity)	0.15***	0.05*	0.14***	0.02	0.15***	0.02	0.13*	-0.07
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.07)	(0.13)
ln(Technological Distance)	-0.11*	0.26**	-0.10*	0.35***	-0.16**	0.35**	-0.19*	0.29
	(0.06)	(0.13)	(0.06)	(0.13)	(0.06)	(0.14)	(0.10)	(0.25)
ln(Cultural Proximity)	0.02	-0.42***	0.01	-0.43***	0.04**	-0.42***	0.06*	-0.42***
	(0.02)	(0.08)	(0.02)	(0.09)	(0.02)	(0.09)	(0.03)	(0.12)
Research Excellence	0.10	-0.10	0.02	-0.18	0.16**	-0.04	0.18**	0.04
	(0.08)	(0.13)	(0.07)	(0.13)	(0.07)	(0.15)	(0.09)	(0.22)
BORDER_d	0.18***	0.22*	0.13**	0.13	0.25***	0.36**	0.25***	-0.40***
	(0.06)	(0.13)	(0.07)	(0.14)	(0.08)	(0.17)	(0.09)	(0.14)
ln(CENTRAL_d)	-0.12	-0.16	-0.11	-0.23	-0.10	-0.12	-0.10	-0.35**
	(0.10)	(0.18)	(0.10)	(0.19)	(0.11)	(0.23)	(0.13)	(0.17)
ln(INV_o)	0.47***	-0.51***	0.50***	-0.52***	0.63***	-0.46***	0.69***	0.08
	(0.04)	(0.06)	(0.04)	(0.07)	(0.05)	(0.07)	(0.11)	(0.10)
ln(INV_d)	0.33***	-0.50***	0.35***	-0.52***	0.52***	-0.44***	0.57***	0.04
	(0.05)	(0.08)	(0.05)	(0.09)	(0.05)	(0.10)	(0.12)	(0.09)
ln(HRST_d)	0.54	-0.01	0.94***	0.24	0.78**	0.40	0.72**	0.26

	(0.37)	(0.57)	(0.32)	(0.58)	(0.38)	(0.71)	(0.37)	(0.83)
ln(POP_o)	-0.02	-0.01	-0.06*	-0.05	-0.03	0.04	-0.02	0.38*
	(0.03)	(0.07)	(0.03)	(0.07)	(0.03)	(0.08)	(0.04)	(0.21)
ln(POP_d)	0.07**	0.04	0.03	-0.01	0.07*	0.03	0.07*	-0.05
	(0.03)	(0.06)	(0.03)	(0.06)	(0.03)	(0.07)	(0.04)	(0.31)
ln(DENS_d)	-0.01	0.04	-0.06	-0.04	-0.07+	-0.03	-0.07*	-0.04
	(0.04)	(0.08)	(0.04)	(0.08)	(0.04)	(0.09)	(0.04)	(0.08)
ln(TEMP_d)	0.87**	1.30*	0.95**	1.51*	1.39***	2.30**	1.30**	2.39
	(0.38)	(0.73)	(0.39)	(0.78)	(0.49)	(1.11)	(0.57)	(1.54)
COAST_d	0.22***	-0.00	0.21***	-0.01	0.29***	0.08	0.26***	-0.06
	(0.07)	(0.15)	(0.07)	(0.17)	(0.08)	(0.19)	(0.09)	(0.26)
Control term							0.03	0.10
							(0.08)	(0.12)
ln(TECH.SHARES)	yes	yes	yes	yes	yes	Yes	yes	yes
Country Fixed Effects (1)	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
Partial R2 first stage							0.513	
F-stat first stage							23.82	
Log-pseudolikelihood	-1,1566.19		-11,662.16		-11,926.66		-11,989.831	-11.988.816
LR test	14,862.139		14,670.196		14,141.203		14,104.856	14,016.887
p-value	0.0000		0.0000		0.0000		0.0000	0.0000
Wald test	2,633.67		2,291.37		2,861.55		-	-
p-value	0.0000		0.0000		0.0000		-	-
LR test of α	7,500.30		9,023.60		1,200		1,400	1,400
p-value	0.0000		0.0000		0.0000		0.0000	0.0000
Vuong statistic	14.91		12.05		11.12		10.54	10.71
p-value	0.0000		0.0000		0.0000		0.0000	0.0000
McFadden's R2	0.391		0.386		0.372		0.369	0.369
Adjusted McFadden's R2	0.384		0.379		0.365		0.362	0.362
AIC	23,394.38		23,586.32		24,115.32		24,077.66	24,075.63
Schwartz	24,544.91		24,736.86		25,265.85		24,508.02	24,505.98

Notes: Robust standard errors are presented in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1, + p<0.12. 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' stand for origin-region and destination-region variables, respectively. **(1)** The UK is treated as the reference country.

Table A.3.7. Endogeneity. First stage regression, OLS. Period 2002-2005. Dependent variable: cross-regional co-patents [ln(Social Proximity)]

	(i)
Intercept	-3.85***
	(0.91)
Contiguity	1.82***
	(0.11)
ln(Time)	-0.73***
	(0.05)
Institutional distance	-4.25***
	(0.07)
ln(Technological Distance)	-1.23***
	(0.05)
ln(Cultural Proximity)	0.24***
	(0.02)
Research Excellence	0.43***
	(0.05)
Same NUTS1	-0.20
	(0.13)
Both regions host the country capital	0.79***
	(0.16)
Both regions belong to the European CORE	0.39***
	(0.06)
ln(average area both regions)	0.09
	(0.06)
ln(sum km to Brussels)	-0.02
	(0.11)
The countries of both regions are contiguous	0.35***
	(0.04)
Origin fixed effects	yes
F-test Joint significance	17.55***
p-value	0.000
Destination fixed effects	yes
F-test Joint significance	33.02***
p-value	0.000
Observations	48,180
Partial R-squared	0.513
First stage F-stat	23.82
p-value	0.000

Notes: Robust standard errors are presented in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A.3.8. Gravity model, ZINB estimations. Periods 2002-2005. Dependent variable: cross-regional mobility – robustness checks.

	(i) time + Income_d		**(ii) time + Income Gap**		**(iii) time + citations_d**		**(iv) time + 2&3 cont**	
	NegBin	Logit	NegBin	Logit	NegBin	Logit	NegBin	Logit
Intercept	-12.37***	-12.87	-11.66***	-8.43	-11.76***	-9.01	-12.39***	-9.71
	(3.19)	(8.23)	(2.94)	(7.92)	(3.92)	(8.75)	(3.33)	(8.81)
Contiguity 1st order	0.89***	-1.32***	0.88***	-1.38***	0.89***	-1.34**	0.96***	-1.18**
	(0.08)	(0.38)	(0.08)	(0.38)	(0.09)	(0.58)	(0.14)	(0.49)
ln(Time)	-0.68***	-0.03	-0.69***	-0.06	-0.69***	-0.03	-0.64***	0.02
	(0.08)	(0.16)	(0.08)	(0.15)	(0.09)	(0.16)	(0.09)	(0.19)
Institutional distance	-0.44***	4.49***	-0.42***	4.69***	-0.44***	4.48***	-0.47***	4.43***
	(0.10)	(0.41)	(0.10)	(0.42)	(0.11)	(0.44)	(0.10)	(0.37)
ln(Social Proximity)	0.16***	0.03	0.16***	0.03	0.16***	0.03	0.15***	0.03
	(0.02)	(0.04)	(0.02)	(0.03)	(0.04)	(0.06)	(0.02)	(0.04)
ln(Technological Distance)	-0.16**	0.41***	-0.14**	0.39***	-0.15*	0.42***	-0.16***	0.41***
	(0.06)	(0.14)	(0.06)	(0.15)	(0.08)	(0.14)	(0.06)	(0.15)
ln(Cultural Proximity)	0.05*	-0.44***	0.05**	-0.47***	0.05+	-0.45***	0.05+	-0.47***
	(0.03)	(0.10)	(0.02)	(0.10)	(0.03)	(0.09)	(0.03)	(0.09)
Research Excellence	0.17**	0.03	0.17**	0.04	0.17**	0.01	0.18**	0.04
	(0.07)	(0.16)	(0.07)	(0.16)	(0.07)	(0.18)	(0.07)	(0.16)
ln(Manuf. Income_d)	-0.02	0.37						
	(0.13)	(0.27)						
ln(Manuf. Income Gap)			-0.00***	-0.00***				
			(0.00)	(0.00)				
ln(Citations_d per capita)					-0.05	-0.19		
					(0.15)	(0.13)		
Contiguity 2nd order							0.10	-0.17
							(0.10)	(0.30)
Contiguity 3rd order							-0.05	0.33
							(0.08)	(0.24)
BORDER_d	0.25***	0.38*	0.23***	0.31	0.25***	0.40*	0.24***	0.35
	(0.08)	(0.21)	(0.08)	(0.20)	(0.08)	(0.21)	(0.08)	(0.23)
ln(CENTRAL_d)	-0.08	-0.10	-0.09	-0.01	-0.11	-0.09	-0.10	-0.03
	(0.11)	(0.23)	(0.11)	(0.22)	(0.11)	(0.23)	(0.11)	(0.24)
ln(INV_o)	0.67***	-0.47***	0.69***	-0.46***	0.67***	-0.48***	0.67***	-0.48***

	(0.05)	(0.08)	(0.04)	(0.08)	(0.14)	(0.09)	(0.05)	(0.08)
ln(INV_d)	0.56***	-0.73***	0.54***	-0.41***	0.60***	-0.22	0.55***	-0.41***
	(0.12)	(0.22)	(0.04)	(0.09)	(0.16)	(0.17)	(0.05)	(0.09)
ln(HRST_d)	0.66*	0.50	0.68**	-0.15	0.60*	-0.08	0.64*	-0.03
	(0.41)	(0.79)	(0.34)	(0.73)	(0.36)	(0.75)	(0.34)	(0.73)
ln(POP_o)	-0.02	0.09	-0.02	0.10	-0.02	0.08	-0.02	0.08
	(0.03)	(0.08)	(0.03)	(0.08)	(0.05)	(0.08)	(0.03)	(0.08)
ln(POP_d)	0.07**	0.05	0.05*	0.01	0.02	-0.14	0.07*	0.03
	(0.03)	(0.08)	(0.03)	(0.08)	(0.15)	(0.15)	(0.03)	(0.08)
ln(DENS_d)	-0.06	-0.11	-0.06	-0.01	-0.07*	-0.05	-0.06	-0.01
	(0.04)	(0.10)	(0.04)	(0.09)	(0.04)	(0.10)	(0.04)	(0.09)
ln(TEMP_d)	1.37**	2.66*	1.25**	2.22	1.37**	2.54	1.27**	2.29
	(0.58)	(1.51)	(0.53)	(1.46)	(0.64)	(1.77)	(0.61)	(1.72)
COAST_d	0.27***	-0.09	0.27***	-0.01	0.26***	-0.07	0.28***	-0.01
	(0.08)	(0.20)	(0.08)	(0.20)	(0.08)	(0.19)	(0.08)	(0.20)
ln(TECH.SHARES)	yes	yes	yes	yes	yes	yes	yes	yes
Country Fixed Effects	yes	yes	yes	yes	yes	yes	yes	yes
Sample size	47,523	47,523	46,872	46,872	48,180	48,180	48,180	48,180
Nonzero observations	3,344	3,344	3,323	3,323	3,365	3,365	3,365	3,365
Log-pseudolikelihood	-11,883.93		-11,788.74		-11,989.65		-11,983.64	
LR test	13,920.619		13,,849.585		14,015.226		14,027.239	
p-value	0.0000		0.0000		0.0000		0.0000	
Wald test	3,351.53		3566.72		3,435.51		3,665.77	
p-value	0.0000		0.0000		0.0000		0.0000	
LR test of α	1,200		1,200		1,200		1,200	
p-value	0.0000		0.0000		0.0000		0.0000	
Vuong statistic	10.38		10.52		10.61		10.51	
p-value	0.0000		0.0000		0.0000		0.0000	
McFadden's R2	0.369		0.370		0.369		0.369	
Adjusted McFadden's R2	0.362		0.363		0.362		0.362	
AIC	24,029.86		23,839.49		24,241.29		24,233.28	
Schwartz	25,178.6		24,986.41		25,391.83		25,401.38	

Notes: Robust standard errors are presented in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1, + p<0.12. 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' stand for origin-region and destination-region variables, respectively. **(1)** The UK is treated as the reference country.

Table A.3.9. Gravity model, ZINB estimations. Periods 2002-2005. Dependent variable: cross-regional mobility – robustness checks.

	(i) time + non-rel.		**(ii) time + non-attrib.**		**(iii) time + mass 7 sectors**		**(iv) time + Soundex with less than 50 records**	
	NegBin	Logit	NegBin	Logit	NegBin	Logit	NegBin	Logit
Intercept	-9.42***	-0.89	-9.57***	-22.66***	-3.96	-21.75***	-11.74***	-13.48*
	(2.45)	(12.09)	(1.98)	(1.10)	(2.94)	(7.48)	(2.94)	(7.13)
Contiguity 1st order	0.95***		1.02***	-0.77**	0.92***	-1.40***	0.86***	-1.28***
	(0.09)		(0.09)	(0.37)	(0.08)	(0.43)	(0.08)	(0.38)
ln(Time)	-0.73***		-0.61***	0.39***	-0.69***	0.04	-0.70***	0.04
	(0.07)		(0.08)	(0.12)	(0.08)	(0.15)	(0.09)	(0.16)
Institutional distance	-1.17***		-0.62***	19.72***	-0.43***	17.74***	-0.37***	4.34***
	(0.10)		(0.10)	(0.12)	(0.10)	(0.81)	(0.12)	(0.51)
ln(Social Proximity)	0.16***		0.15***	-0.02	0.15***	0.02	0.16***	0.02
	(0.01)		(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.04)
ln(Technological Distance)	-0.17***		-0.01	0.74***	-0.05	0.37***	-0.14**	0.32**
	(0.06)		(0.06)	(0.14)	(0.06)	(0.14)	(0.06)	(0.15)
ln(Cultural Proximity)	0.09*		0.08***	0.07***	0.06**	-0.44***	0.05**	-0.29
	(0.05)		(0.03)	(0.03)	(0.03)	(0.10)	(0.02)	(0.22)
Research Excellence	0.24***		0.28***	0.05	0.19***	0.02	0.19**	0.02
	(0.07)		(0.08)	(0.12)	(0.07)	(0.15)	(0.08)	(0.16)
BORDER_d	0.19***	-0.07	0.17***		0.25***	0.36*	0.17**	0.27
	(0.07)	(0.44)	(0.06)		(0.08)	(0.21)	(0.07)	(0.18)
ln(CENTRAL_d)	-0.22**	-0.62	-0.11		-0.20*	-0.06	-0.07	0.17
	(0.11)	(0.57)	(0.09)		(0.11)	(0.24)	(0.11)	(0.22)
ln(INV_o)	0.75***	-0.47***	0.77***		-	-	0.66***	-0.45***
	(0.03)	(0.14)	(0.03)		-	-	(0.07)	(0.09)
ln(INV_d)	0.60***	-0.49***	0.63***		-	-	0.54***	-0.32***
	(0.04)	(0.17)	(0.04)		-	-	(0.05)	(0.10)
ln(HRST_d)	0.72**	-0.26	0.73**		0.55*	-0.09	0.69**	0.30
	(0.34)	(1.36)	(0.32)		(0.32)	(0.68)	(0.34)	(0.69)
ln(POP_o)	-0.04	-0.15	-0.03		-0.01	0.07	-0.02	0.03
	(0.03)	(0.18)	(0.03)		(0.03)	(0.07)	(0.04)	(0.08)
ln(POP_d)	0.07**	-0.11	0.07**		0.06**	0.02	0.06*	0.07
	(0.03)	(0.15)	(0.03)		(0.03)	(0.07)	(0.03)	(0.08)
ln(DENS_d)	-0.04	0.07	-0.06^{+}		-0.08*	-0.06	-0.08*	-0.07

	(0.05)	(0.23)	(0.04)		(0.04)	(0.09)	(0.04)	(0.09)
ln(TEMP_d)	0.88*	1.87	0.84**		0.89*	1.92	1.24**	2.56*
	(0.46)	(2.19)	(0.33)		(0.55)	(1.51)	(0.49)	(1.43)
COAST_d	0.29***	-0.29	0.31***		0.32***	-0.05	0.27***	-0.15
	(0.08)	(0.54)	(0.06)		(0.08)	(0.19)	(0.08)	(0.20)
ln(INV7sectors_o)	no	no	no	no	yes	yes	no	no
ln(INV7sectors_d)	no	no	no	no	yes	yes	no	no
ln(TECH.SHARES)	yes	yes	yes	yes	no	no	yes	yes
Country 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 observations	3,365	3,365	3,365	3,365	3,365	3,365	2,917	2,917
Log-pseudolikelihood	-12,216.83		-12,153.57		-12,014.73		-10,321.22	
LR test	13,560.866		13,687.381		13,965.064		12,375.641	
p-value	0.0000		0.0000		0.0000		0.0000	
Wald test	7,580.87		-		3,614.33		3,183.04	
p-value	0.0000		-		0.0000		0.0000	
LR test of α	1,300		1,300		1,200		7,032.68	
p-value	0.0000		0.0000		0.0000		0.0000	
Vuong statistic	6.70		7.77		10.56		9.82	
p-value	0.0000		0.0000		0.0000		0.0000	
McFadden's R2	0.357		0.360		0.368		0.375	
Adjusted McFadden's R2	0.350		0.356		0.361		0.367	
AIC	24,677.65		24,443.14		24,279.46		20,900.45	
Schwartz	25,749.14		25,040.36		25,377.29		22,033.42	

Notes: Robust standard errors are presented in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1, + p<0.12. 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' stand for origin-region and destination-region variables, respectively. **(1)** The UK is treated as the reference country.

Institut de Recerca en Economia Aplicada Regional i Públic
Research Institute of Applied Economics

WEBSITE: www.ub-irea.com • **CONTACT:** irea@ub.edu

Grup de Recerca Anàlisi Quantitativa Regional
Regional Quantitative Analysis Research Group

WEBSITE: www.ub.edu/aqr/ • **CONTACT:** aqr@ub.edu

Universitat de Barcelona
Av. Diagonal, 690 • 08034 Barcelona