Ann Reg Sci (2009) 43:721–738
DOI 10.1007/s00168-008-0252-9

SPECIAL ISSUE PAPER

The geography of collaborative knowledge production in Europe

Jarno Hoekman · Koen Frenken · Frank van Oort

Published online: 31 July 2008

Abstract We analyse inter-regional research collaboration as measured by scientific publications and patents with multiple addresses, covering 1316 NUTS3 regions in 29 European countries. The estimates of gravity equations show the effects of geographical and institutional distance on research collaboration. We also find evidence for the existence of elite structures between excellence regions and between capital regions. The results suggest that current EU science policy to stimulate research collaboration is legitimate, but doubt the compatibility between EU science policy and EU cohesion policy.

JEL Classification R10 · R12

1 Introduction

Knowledge production has become a central concern for firms and policy makers alike. In particular, the transformation towards a 'European knowledge society' rendered science and technology of particular importance to ensure the competitiveness of Europe. Against the background of this process, the 'Lisbon agenda' of the European Union can be considered an attempt to reorient Europe's main rationale from one based on economic integration alone towards one based on the concept of a common

J. Hoekman (✉) · K. Frenken · F. van Oort
Urban and Regional Research Centre Utrecht (URU), Utrecht University,
PO Box 80115, 3508 TC Utrecht, The Netherlands
e-mail: j.hoekman@geo.uu.nl

F. van Oort
Netherlands Institute for Spatial Research (RPB), PO Box 30314,
2500 GH The Hague, The Netherlands

knowledge society. A major initiative in this direction has been to create a European research area (ERA).

The general idea underlying ERA is that *research activities at national and Union level must be better integrated and coordinated to make them as efficient and innovative as possible* (European Council 2000). Such an objective assumes that a European Research Area does not yet exist and that its creation requires action at several levels of spatial aggregation. Yet, studies assessing these assumptions are scarce and traditionally have focused only on the level of countries (e.g., Narin et al. 1991; Glänzel 2001; Frenken 2002). Little is known about the regional dimension of collaborative knowledge production despite its supposed relevance in the light of regional, national and European policies.

The present study assesses the extent to which European inter-regional research activities are already integrated based on scientific publications and patents with multiple addresses. Using these data, we address the role of *proximity* and of *elite structures* in collaborative knowledge production. Our main research question holds to what extent geographical and institutional proximity, as well as elite structures among excellence regions and among capital regions, explain the participation of regions in collaborative knowledge production.

Concerning proximity, the inter-regional perspective allows us to differentiate between geographical patterns in collaborative knowledge production within and between member states. By doing so, we can test to what extent *geographical distance* and *institutional distance* hamper collaboration, where geographical distance is expressed in terms of distance in kilometres between two regions and institutional distance is reflected in a dummy variable distinguishing between domestic and foreign collaborations. Geographical distance relates directly to the costs of collaboration, which increase with distance, while institutional distance relates to obstacles in collaboration due to different national institutions.

In our framework based on inter-regional collaboration, we also analyse *elite structures* that facilitate collaboration among favoured regions. More specifically, we focus on cognitive structures that explain why 'excellence regions' have a bias to network among themselves and on political structures that explain why capital cities have a bias to network among themselves.

This paper is organised as follows. In Sect. 2 we discuss general trends in research collaboration. Section 3 introduces some theoretical concepts and derives a number of hypotheses. Data and methodology are presented in Sect. 4 and the estimation results in Sect. 5. In the final section we discuss EU policies in the light of our evidence.

2 Previous research

If anything has characterized knowledge production in science and technology during the twentieth century, it is the increased collaborative nature of knowledge production (Meyer and Bhattacharya 2004). In science, co-authorships accounted for less than 10% of all publications at the start of the twentieth century, while co-authorships account for over 50% of all publications at the end of the twentieth century (Wagner-Doebler 2001).

The relevance of collaboration is evidenced by the fact that the number of citations that scientific articles receive increases with the number of contributing researchers (Katz and Martin 1997; Frenken et al. 2005). Similarly, the average number of inventors that contribute to a patent has increased over time during the past 20 years (Fleming and Frenken 2007). Both trends indicate an increased division of labour among researchers. With the universe of knowledge ever expanding, researchers need to specialise to continue contributing to state of the art knowledge production.

To encourage research collaboration, the European Union has always been concerned with funding international research projects and with removing barriers that currently hinder researchers in such projects, and its financial efforts in this direction have again been increased substantially in the seventh framework programme, which runs from 2007 to 2013 (Commission of the European Communities 2006).[1] European collaboration is expected to generate benefits in many ways. Economically, it provides opportunities to realize savings with regard to costs of training and sharing research infrastructures as well as to avoid duplication of research efforts. International collaboration is also expected to generate intellectual benefits from the cross-fertilization of ideas that previously were unconnected. Indeed, scientific articles stemming from international collaboration projects, on average, receive more citations than national collaboration projects (Narin et al. 1991; Katz and Martin 1997). The European Commission's objective to create an ERA by stimulating research collaboration is therefore legitimate as long as barriers exist that impede European researchers from engaging in research collaboration.

Studies analyzing collaborative knowledge production at the regional level have been mostly limited to particular countries only. Co-publications among regions have been analysed by Katz (1994) for the UK regions, Danell and Persson (2003) for Swedish regions, Liang and Zhu (2002) for Chinese regions, and Ponds et al. (2007) for Dutch regions. Co-inventorships among Swedish regions using patent data have been analysed by Ejermo and Karlsson (2006). At the European level, we know of only one patent study by Maggioni and Uberti (2007) who analysed the effect of geographical distance on inter-regional collaborations based on co-inventorships between NUTS2 regions for six countries. In line with studies done for particular countries, they also found that distance significantly affect the formation of inventor networks.

Our study takes three steps to improve the analysis of the geography of research collaboration. First, we have been able to cover a larger set of countries (EU27 plus Norway and Switzerland) at a lower level of spatial aggregation (NUTS3). Second, we will analyse not only the effect of geographical distance on the intensity of inter-regional collaboration, but we will also include other determinants (institutional proximity, elite structures) in the analysis. Third, since we collected both data on publications and patents we are able to differentiate between research collaboration in science and technology, respectively.

[1] The total budget of the seventh Framework amounts to EUR 50.521 million. The majority (64.1%) of the budget of the Seventh Framework is reserved for 'Cooperation'. Other important elements are labour mobility of researchers under the heading of 'People' (9.4%) and the enhancement of research and innovation infrastructures under the heading of 'Capacities' (8.1%) (Commission of the European Communities 2006).

3 Theoretical framework

The rationales for collaborative knowledge production are straightforward: actors engage in collaborations to learn from each other and to make a stronger impact on the field than could be achieved individually. Indeed, collaborations are expected to increase the quality of the research output, but at the same time the pursuit of quality is restricted by several constraints. The time and money required to engage in collaboration are substantial, which forces researchers to be highly selective in choosing a collaboration partner. Thus, the strength of interaction between any two actors, and any two regions, will be dependent on the learning opportunities involved in collaboration at the one hand, and the time and money required to participate on the other hand.

Starting with the costs involved, we can distinguish between two forms of proximity that are expected to bring down costs and thus to increase the probability of interaction (Boschma 2005). First, the costs of collaboration increase as a function of geographical distance. As a result, we hypothesise that research collaborations between geographically proximate researchers are more likely to occur. Second, the costs of research collaboration increase with institutional distance as a common institutional framework brings costs down (Gertler 1995; Edquist and Johnson 1997).[2] In the case of knowledge production, the relevant institutional arrangements (funding, labour markets, intellectual property right regimes, common language) have a strong, although not exclusive, national component. Hence, our hypothesis therefore holds that two regions that belong to the same country are institutionally nearby and more inclined to collaborate, while two regions belonging to different countries are institutionally distant and more reluctant to collaborate.

Turning to benefits of collaboration, we distinguish between benefits for elite researchers and other researchers. Elite researchers working at the cutting edge of research are more inclined to collaborate with other elite researchers, since they learn much more from fellow elite researchers than from those less advanced. A fundamental observation in this context is that elites are remarkably concentrated in certain regions. This generates advantages as evidenced by the mean rate of citations received by scientific publications (Frenken et al. 2007; Tijssen 2007). Hence, in research collaboration, regional hierarchies are likely to emerge, with regions hosting the elite researchers—which we call 'excellence regions'—networking primarily among them and much less with less advanced regions.

Second, elite structures exist between researchers in terms of access to their resources. Collaboration requires resources, and differential access to resources will impact the propensity of actors to collaborate. Resources are concentrated in large cities—predominantly capital cities—where banks and funding agencies are concentrated. Furthermore, most national research institutes are located in capital cities, and these institutes are typically over-represented in multilateral programmes supported by multi-lateral government funding. Following this reasoning, we expect that, all else

[2] Institutional proximity can also be taken to refer to relations between organizations that operate in the same societal subsystem, like inter-university relationships, or inter-firm relationships, or inter-governmental relationships. On this, see Ponds et al. (2007).

being equal, pairs of capital regions are likely to have stronger ties than pairs of any other type of regions.[3]

Summarising, we expect the inter-regional intensity of collaboration to be dependent on costs on the one hand and benefits on the other. The wish to minimise costs will lead researchers to be biased and to collaborate with geographically and institutionally proximate parties. Differential opportunities will be reflected in cognitive elite structures between excellence regions and political elite structures between capital regions.

4 Research design

Research on collaborative knowledge production has always been relying on partial indicators. Since knowledge is—by definition—intangible it cannot be measured and counted directly and unambiguously. Yet, many research collaboration efforts, have a tangible output: a text. Many of these texts reach the public domain in the form of publications in scientific journals or in the form of patents awarded by patent offices. Both publications and patents indicate a research activity of proven value. Publications in scientific journals have been peer-reviewed, which assures a certain minimum level of quality and originality. Patents are reviewed by patent examiners, who decide to grant a patent on the basis of the originality of the invention.

Scholars studying science and technology make extensive use of publications and patent data due to a number of advantages (Griliches 1990; Frenken et al. 2007):

1. Each publication and patent contains highly detailed information on content (title words and abstract), previous art (citations), researchers (name), organizations involved (institutional affiliation), and geographical location (address).
2. Systematic data collection on patents and publication goes back a long time.
3. The current 'stock' of patents and publications is large and ever growing.

However, we should also bear in mind that the use of these paper trails is not completely without limitations (Griliches 1990; Frenken et al. 2007). More specifically, we can identify three major drawbacks:

1. Research does not necessarily lead to publications or patents. Rejection by reviewers is one of the main reasons of research efforts not necessarily resulting in publications or patents. Other reasons include time/cost constraints of researchers to submit a report for publications or patenting and non-disclosure strategies by firms who value secrecy over property right.
2. Publications and patents do not necessarily contribute to our knowledge. Most publications and patents are rarely cited, if at all, which suggests that their added value to the knowledge system is small, and, regarding patents, the commercial value of patents varies widely.

[3] What is more, almost all capital regions also host the main airport in a country, providing an advantage in accessibility through air.

3. Publication and patenting rates differ systematically across scientific disciplines and technology fields, respectively. This means that inter-regional comparisons can be misleading due to the differences in technological specialization.

Despite these shortcomings we make use of both publications and patents as we consider these data appropriate given our purpose for a number of reasons. With regard to the first limitation, our research topic being the European Research Area renders the use of quantitative information almost indispensable. Alternative research methodologies, for example based on expert interviews, would be too limited in their scope. We address the second limitation by aggregating publications and patents to the regional level in order to minimise differences in quality. Furthermore, regarding publications, we distinguish between excellence regions and other regions as to control for quality differences. With regard to the third limitation, the separate analysis of various scientific disciplines and technology classes allows us to avoid making conclusions that are biased by regional differences in scientific or technological specialization.

4.1 Data

Data on publications have been retrieved from *Web of Science (WoS)*, which is a product of Thomson Scientific. Web of Science is an electronic archive of scientific publications in most science journals. Though WoS does not contain all journals and tends to be biased towards English-language journals, it is widely considered the most comprehensive and reliable source covering all the major journals in the world.

Data on patents have been obtained from the *European Patent Office (EPO)* database. Our focus on the European Research Area provides a clear rationale for the use of this database. Moreover, using patent data from the European Patent Office rather than from national patent offices ensures that we deal with patents with, on average, a high expected commercial value, since applying to the EPO is more expensive and time-consuming than applying only to national patent offices.

We retrieved the information for scientific articles published between 1988 and 2004, since access to WoS is restricted before 1988. Hence, patents have also been obtained from 1988 onwards, but we did not extend the patent data beyond 2001, because there is a sudden drop in the total number of patents after 2001 at the time we retrieved the data. This drop reflects to backlog in the administration of patents awarded.

We did not retrieve all publications and patents, but limited the analysis to two science-based technologies: biotechnology and semiconductors. [4] These technologies had a revolutionary global impact during the last two decades and have long been thematic priorities in many European, national and regional policies. Patents are selected on the basis of the IPC classes biotechnology and semiconductors. Following Verbeek et al. (2003), we subsequently selected scientific publications on the basis of journals that are often cited in the patents. For biotechnology, the relevant scientific discipline

[4] More details can be found in Frenken et al. (2007).

becomes biochemistry and molecular biology, while for semiconductors we chose electrical and electronical engineering as the relevant scientific discipline.[5]

With regard to the territorial breakdown, we decided to construct our data at the NUTS3 level covering the 27 countries of the European Union plus Norway and Switzerland. We consider the NUTS3 level of spatial aggregation to be relevant as it corresponds most closely to regional labour markets in casu 'regional innovation systems' (Cooke et al. 1998). Thus, all addresses occurring in publications and in patents have been assigned to one of the 1316 NUTS3 regions in the aforementioned 29 countries in Europe.[6]

One major advantage of using publications and patents is that the addresses of researchers are systematically recorded in these texts. We make use of this information to construct our dataset on research collaboration by selecting all publications and patents with multiple addresses in more than one NUTS3 region.[7] In our dataset this phenomenon represents an inter-regional collaboration link. The collaboration intensity between region i and j, labelled I_{ij}, is then defined by the number of times addresses from these two regions co-occur in a publication or a patent. In doing so, we obtain our matrices of inter-regional collaboration patterns which serve as the basis of our empirical analysis. We thus use 'full counting' to derive the interaction strength between two regions. For example, if a publication contains three addresses in three different regions, the interaction strength between each pair of regions is 1. Alternatively, one can use fractional counting were a co-occurrence of two regions in a publication or patent divided by the total number of interactions. For example, if a publication contains three addresses in three different regions, the interaction strength between each pair of regions is one-third. The final matrix of inter-regional interaction strength based on full counting is very similar to the final matrix obtained by fractional counting.[8]

It is important to note here that the occurrence of publications and patents with multiple addresses may refer to several underlying mechanisms. In most cases, an inter-regional link represents a collaboration between two or more researchers or institutions. Yet, it may also be the case that a single researcher appears on a publication or patent with two or more addresses. This phenomenon also counts as a collaboration and denotes that the researcher works for two or more organizations or conducted a research for one organization and subsequently moved to another

[5] Publications from Applied Physics are even more often cited than publications from electrical and electronical engineering, yet Applied Physics is rather broad as to account as a discipline.

[6] We were not able to locate the addresses within the greater urban areas of London and Manchester and as a result consolidated them into two new regions. Furthermore, we excluded some islands due to their remote locations and disproportional great geographical distances to other regions. These islands are: Guadeloupe Las Palmas (ES), Santa Cruz de Tenerife (ES), Guadeloupe (FR), Martinique (FR), Guyane (FR), Réunion (FR), Região Autónoma dos Acores (PT) and Região Autónoma da Madeira (PT). The outcome is a total number of 1316 NUTS3 regions instead of 1329.

[7] The address information in publication data refers to the address of the organization where the researcher works. In contrast, the address information in the patent data we used refers to the home addresses of the researchers involved. This difference should always be kept in mind, as it precludes any comparison between the collaboration patterns that are reflected in publications and those that are reflected in patents.

[8] Correlations between the full counting and fractional counting matrices are above 0.99.

organization. Thus, the inter-regional collaboration networks refer primarily to the main pillar of the Framework Programmes (i.e., 'Cooperation'); to some extent, however they also reflect labour mobility mechanisms, which are another pillar of Europe's research policies under the heading of 'People'.

4.2 Gravity model

We analyze the determinants of the constructed interregional-networks using a gravity model. Spatial interaction, the process whereby actors at different points in physical space make contacts, can be revealed by applying an analogical model of Isaac Newton's Theory of Universal Gravitation (Tinbergen 1962; Sen and Smith 1995; Roy and Thill 2004). In a gravity model, the gravitational force between two objects is assumed to be dependent on the mass of the objects and the distance between them. In our case this means that the interaction intensity of research collaborations in science and technology aggregated at the NUTS3 level is hypothesized to be dependent on the masses of the two regions and inversely dependent on the geographical distance between two regions. The basic gravity equation is therefore as follows:

$$I_{ij} = \alpha_1 \frac{MASS_i^{\alpha 2} MASS_j^{\alpha 3}}{DISTANCE_{ij}^{a4}} \tag{1}$$

Such a gravity model can be estimated using linear regression by taking a double log:

$$\ln I_{ij} = \ln \alpha_1 + \alpha_2 \ln MASS_i + \alpha_3 \ln MASS_j + \alpha_4 \ln DISTANCE_{ij} \tag{2}$$

with $\alpha_2 > 0$, $\alpha_3 > 0$ and $\alpha_4 < 0$.

Since we deal with count data, we cannot rely on an OLS estimation procedure. The use of alternative regression techniques is appropriate (Burger and Van Oort 2007). Probably the most common regression model applied to count data is Poisson regression, which is estimated by means of maximum likelihood estimation techniques. In this log–linear model, the observed interaction intensity between region i and j has a Poisson distribution with a conditional mean (μ) that is a function of the independent variables (Eq. 3).

$$\begin{aligned} \Pr[I_{ij}] &= \frac{\exp^{-\mu_{ij}} \mu_{ij}^{I_{ij}}}{I_{ij}!}, \text{ where in our model} \\ \mu_{ij} &= \exp(a_1 + a_2 \ln MASS_i + \alpha_3 \ln MASS_j + \alpha_4 \ln DISTANCE_{ij}) \end{aligned} \tag{3}$$

In order to correct for overdispersion (conditional variance is larger than the conditional mean) and an excessive number of zero counts in our data set (the incidence of zero counts is greater than would be expected for the Poisson distribution as most pair of regions do not collaborate with each other), we make use of the zero-inflated negative binomial regression, which can be perceived as an extension of the Poisson model.

Not correcting for the overdispersion and excess zero problem normally results in incorrect and biased estimates.

The zero-inflated negative binomial model considers the existence of two (latent) groups within the population: a group having strictly zero counts and a group having a non-zero probability of counts different than zero. Correspondingly, its estimation process consists of two parts. The first part contains a logit regression of the predictor variables on the probability that there is no interaction between two given regions at all. The second part contains a negative binomial regression on the probability of each count for the group that has a non-zero probability of count different than zero. A good technical discussion of the zero-inflated negative binomial model is provided by Long (1997).

4.3 Covariates

The gravity equation assumes that inter-regional interaction is dependent on the respective size or masses of the regions. In line with our count method for the interaction strength between regions, we use full counting for the masses and derive the total number of publications and patents, including single-authored texts. Since collaborations are by definition undirected we only once include the interaction between a pair of regions. Due to this fact the size of the coefficient of the two masses may slightly differ.[9] Note also that we added 1 to all masses in order to allow for logarithmic transformation of observations without any publications or patents.

We account for our theoretical suggestions regarding the spatial context of research collaboration by introducing a number of independent variables. In concordance with basic gravity models we add *DISTANCE*, which is calculated between the central points of regions using GIS ('as the crow flies'). The covariate *COUNTRY* is a variable capturing institutional proximity between regions, coded one if regions belong to the same country and coded zero otherwise.

As explained, elite structures are accounted for by defining *EXCELLENCE* and *CAPITAL*. In our analysis, excellence regions are defined as those belonging to the top 25 most publishing regions and the top 25 most patenting regions. Size is treated here as a proxy for quality. Regions that host top institutes will typically grow and attract the best talent, while regions with poor institutes will have trouble growing and retaining their talent. The assumption that size and quality are closely correlated is also supported by the empirical finding that the mean citation rate for scientific articles in a region increases with the number of articles produced in that region (Frenken et al. 2007; Tijssen 2007). Defining capital regions does not need further explanations, although we should mention that as a result of the low level of aggregation we selected more than one NUTS3 regions as capital regions for some countries.[10] From this, we create two dummy variables that capture the elite structures between regions. Excellence

[9] Alternatively, we may also subtract M_i and M_j to make a single new variable indicating the mass of both regions. Results of the regression models are similar and available on request.

[10] This is the case for Paris, France (5 regions) and Copenhagen, Denmark (2 regions). In all other countries we selected one NUTS3 region that corresponds to the capital city.

Table 1 Descriptive statistics of inter-regional collaborations

	N	Mean	SD	Min.	Max.
Publications biotechnology					
Inter-regional collaborations	865270	0.251	5.058	0	1671
Number of publications region *i*	865270	263.341	965.595	1	23694
Number of publications region *j*	865270	381.311	1410.671	1	23694
Inter-regional distance in km	865270	1045.050	633.322	6.448	4195.561
Patents biotechnology					
Inter-regional collaborations	865270	0.039	1.595	0	609
Number of patents region *i*	865270	30.684	93.153	1	1332
Number of patents region *j*	865270	30.006	100.960	1	1332
Inter-regional distance in km	865270	1045.050	633.322	6.448	4195.561
Publications semiconductors					
Inter-regional collaborations	865270	0.060	1.118	0	296
Number of publications region *i*	865270	74.990	260.974	1	4714
Number of publications region *j*	865270	117.535	350.286	1	4714
Inter-regional distance in km	865270	1045.050	633.322	6.448	4195.561
Patents semiconductors					
Inter-regional collaborations	865270	0.011	0.814	0	81
Number of patents region *i*	865270	16.141	73.646	1	1518
Number of patents region *j*	865270	12.274	71.714	1	1518
Inter-regional distance in km	865270	1045.050	633.322	6.448	4195.561

structures are measured by a dummy for relations between two regions of excellence, and capital structures by a dummy for relations between two capital regions.

The extended gravity equation to be estimated is thus as follows:

$$\ln I_{ij} = \alpha_1 + \alpha_2 \ln MASS_i + \alpha_3 \ln MASS_j + \alpha_4 \ln DISTANCE_{ij} + \alpha_5 COUNTRY_{ij} + \alpha_6 EXCELLENCE_{ij} + \alpha_7 CAPITAL_{ij} + \varepsilon \quad (4)$$

The zero-inflated negative binomial model allows for an estimation process in which the explanatory variable is predicted by two distinct processes. As we believe that in case of research collaboration the determinants predicting the change of collaborating do not differ from the determinants that predict the intensity, we include the same variables in both parts of the regression model. The only exception in the model is the variable *EXCELLENCE*, which we only include in the negative binomial part. The reason for this is that estimating the probability that there is no interaction at all is irrelevant in this case, as we only included regions that belong to the 25 most publishing or patenting regions.

Table 1 reports some descriptive statistics on the variables of main interest. Because our analysis addresses all possible pairs of regions, and not individual regions, the total number of observations amounts to $1/2 \times 1,316 \times 1,315 = 865,270$ observations. This also implies that the mean number of collaboration is very low as the large

majority of inter-regional pairs do not collaborate at all (hence, our choice for the zero-inflated negative binomial regression model).

5 Results

Before discussing the results of the regression analysis, we present correlation matrices in Table 2 to identify possible multicollinearity in the covariates. All correlations are well within the allowed range and can be included in the regression analysis.

Tables 3, 4, 5 and 6 present the estimates for the regression models with all four regression models showing successively a negative binomial part (NBP), a zero inflated

Table 2 Correlation matrix of covariates

		1	2	3	4	5	6
Biotechnology publications							
1	Mass origin (ln)	1.000					
2	Mass destination (ln)	0.011*	1.000				
3	Distance (ln)	0.036*	0.067*	1.000			
4	Same country	−0.025*	−0.128*	−0.616*	1.000		
5	Excellence	0.048*	0.043*	−0.000	−0.002	1.000	
6	Capital	0.050*	0.042*	0.009*	−0.009*	0.113*	1.000
Semiconductor publications							
1	Mass origin (ln)	1.000					
2	Mass destination (ln)	0.011*	1.000				
3	Distance (ln)	0.016*	0.058*	1.000			
4	Same country	0.007*	−0.125*	−0.616*	1.000		
5	Excellence	0.052*	0.043*	0.000	−0.002	1.000	
6	Capital	0.049*	0.040*	0.009*	−0.009*	0.070*	1.000
Biotechnology patents							
1	Mass origin (ln)	1.000					
2	Mass destination (ln)	0.005	1.000				
3	Distance (ln)	−0.121*	−0.121*	1.000			
4	Same country	0.051*	−0.003	−0.616*	1.000		
5	Excellence	0.050*	0.053*	−0.011*	−0.000	1.000	
6	Capital	0.040*	0.035*	0.009*	−0.009*	0.090*	1.000
Semiconductors patents							
1	Mass origin (ln)	1.000					
2	Mass destination (ln)	0.011*	1.000				
3	Distance (ln)	−0.161*	−0.165*	1.000			
4	Same country	0.144*	0.039*	−0.616*	1.000		
5	Excellence	0.057*	0.071*	−0.014*	0.005*	1.000	
6	Capital	0.025*	0.027*	0.009*	−0.009*	−0.009*	1.000

* Indicates significance at 1% level

Table 3 Zero-inflated negative binomial regression model on interaction intensity of co-publishing in biotechnology for the period 1988–2004

Parameter	Model A estimate (SE)	Model B estimate (SE)	Model C estimate (SE)
Negative binomial part			
Constant	−2.363 (0.067)*	−5.401 (0.086)*	−5.040 (0.087)*
Mass origin (ln)	0.640 (0.006)*	0.649 (0.005)*	0.621 (0.006)*
Mass destination (ln)	0.591 (0.005)*	0.636 (0.005)*	0.609 (0.005)*
Distance (ln)	−0.734 (0.009)*	−0.368 (0.010)*	−0.367 (0.010)*
Same country		1.160 (0.022)*	1.146 (0.022)*
Excellence			0.832 (0.056)*
Capital			0.475 (0.052)*
Zero-inflated part			
Constant	4.458 (0.112)*	7.366 (0.165)*	7.593 (0.162)*
Mass origin (ln)	−0.760 (0.009)*	−0.769 (0.009)*	−0.787 (0.009)*
Mass destination (ln)	−0.764 (0.009)*	−0.779 (0.009)*	−0.794 (0.009)*
Distance (ln)	0.739 (0.017)*	0.359 (0.021)*	0.362 (0.021)*
Same country		−1.395 (0.048)*	−1.394 (0.046)
Excellence			–
Capital			−0.974 (0.233)*
Fit statistics			
Overdispersion (α)	1.098 (0.017)*	0.881 (0.014)*	0.848 (0.013)*
Vuong-statistic	27.43*	27.25*	27.85*
Log Likelihood	−102711.865	−99774.550	−99545.800
Mc Fadden's Adj. R^2	0.442	0.458	0.459
AIC	0.237	0.231	0.230
N	865270	865270	865270
Nonzero observations	25589	25589	25009

* Indicates significance at 1% level

part (ZIP) and some general fit statistics.[11] The latter include tests checking whether the choice of the zero inflated negative binomial regression models is appropriate. Overall, the likelihood ratio test of overdispersion and the Vuong-statistic are significant, indicating that the zero-inflated negative binomial regression model fits our data best.

In each regression, Model A restricts the analysis to the respective mass of the regions and the geographical distance between them, while Model B adds institutional proximity (same country) and Model C the two elite structures related to excellence

[11] It is essential to keep in mind that a positive sign in the zero inflated part indicates that with a one percent positive change in the predictor, the chance of belonging to the 'strictly zero group' increases, holding all other predictors constant. Thus, the coefficients in the zero inflated part should be interpreted in reverse in comparison to the negative binomial part: a positive value in the negative binomial part has the same meaning as a negative value in the zero-inflated part and vice versa.

Table 4 Zero-inflated negative binomial regression model on interaction intensity of co-publishing in semiconductors for the period 1988–2004

Parameter	Model A estimate (SE)	Model B estimate (SE)	Model C estimate (SE)
Negative binomial part			
Constant	−2.091 (0.013)*	−4.064 (0.133)*	−3.763 (0.135)*
Mass origin (ln)	0.550 (0.010)*	0.533 (0.009)*	0.504 (0.010)*
Mass destination (ln)	0.526 (0.010)*	0.552 (0.010)*	0.525 (0.010)*
Distance (ln)	−0.565 (0.013)*	−0.301 (0.016)*	−0.300 (0.016)*
Same country		0.824 (0.036)*	0.836 (0.036)*
Excellence			0.626 (0.073)*
Capital			0.450 (0.076)*
Zero inflated part			
Constant	4.535 (0.145)*	6.999 (0.202)*	7.150 (0.201)
Mass origin (ln)	−0.844 (0.013)*	−0.851 (0.013)*	−0.866 (0.013)*
Mass destination (ln)	−0.810 (0.013)*	−0.832 (0.014)*	−0.845 (0.014)*
Distance (ln)	0.739 (0.021)*	0.423 (0.027)*	0.426 (0.026)*
Same country		−1.112 (0.059)*	−1.098 (0.058)*
Excellence			–
Capital			−1.146 (0.201)*
Fit statistics			
Overdispersion (α)	1.502 (0.038)*	1.333 (0.034)*	1.302 (0.034)*
Vuong-statistic	20.30*	20.41*	20.46*
Log likelihood	−52191.683	−51301.529	−51202.390
Mc Fadden's Adj. R^2	0.429	0.439	0.440
AIC	0.121	0.119	0.118
N	865270	865270	865270
Nonzero observations	12531	12531	12531

* Indicates significance at 1% level

and capital regions. The results in all models show that mass and geographical distance are indeed powerful predictors of research collaboration in co-publications and in co-patents. Naturally, the mass contributes positively, indicating an increase in the change and intensity of collaboration between two regions if these regions accommodate a larger number of knowledge producing actors.[12] Distance has a significant negative effect too on the chance and intensity of collaboration. Regions that are further apart collaborate less than regions that are in closer proximity.

[12] The effect of geographical distance tends to be stronger for patents than for publications possibly indicating the higher tacit content on technological knowledge compared to scientific knowledge. Yet, since address information in patent data refers to home address of inventors, while address information in publication data refers to addresses of the employer, strictly speaking, the two cannot be compared.

Table 5 Zero-inflated negative binomial regression model on interaction intensity of co-patenting in biotechnology for the period 1988–2001

Parameter	Model A estimate (SE)	Model B estimate (SE)	Model C estimate (SE)
Negative binomial part			
Constant	0.417 (0.105)*	−0.187 (0.147)	−0.180 (0.148)
Mass origin (ln)	0.411 (0.013)*	0.419 (0.013)*	0.414 (0.013)*
Mass destination (ln)	0.376 (0.013)*	0.387 (0.013)*	0.381 (0.013)*
Distance (ln)	−0.572 (0.015)*	−0.503 (0.018)*	−0.499 (0.018)*
Same country		0.275 (0.053)*	0.296 (0.053)
Excellence			0.046 (0.115)
Capital			0.453 (0.153)*
Zero inflated part			
Constant	−0.859 (0.124)*	3.360 (0.172)*	3.292 (0.172)*
Mass origin (ln)	−0.740 (0.013)*	−0.769 (0.014)*	−0.765 (0.014)*
Mass destination (ln)	−0.678 (0.013)*	−0.771 (0.014)*	−0.769 (0.014)*
Distance (ln)	1.458 (0.022)*	0.951 (0.025)*	0.961 (0.025)*
Same country		−1.750 (0.055)*	−1.743 (0.054)*
Excellence			–
Capital			−1.389 (0.219)*
Fit statistics			
Overdispersion (α)	2.022 (0.082)*	1.880 (0.072)*	1.865 (0.071)*
Vuong-statistic	22.22*	19.08*	12.12*
Log likelihood	−31659.830	−30738.290	−11751.202
Mc Fadden's Adj. R^2	0.369	0.387	0.408
AIC	0.073	0.071	0.027
N	865270	865270	865270
Nonzero observations	6078	6078	6078

* Indicates significance at 1%level

Institutional proximity as captured by the dummy variable *COUNTRY* is added in model B. The variable is significant in three of the four models and it has the expected positive sign, indicating that two regions belonging to the same country collaborate more frequently than two regions from different countries. Comparison of the results of Model A and Model B also reveals that the inclusion of the *COUNTRY* variable diminishes the estimate of the *DISTANCE* variable, as there is considerable correlation between geographical distance and belonging to the same country. However, though its influence diminishes, geographical distance remains significant in all cases, indicating an independent effect of both geographical distance and institutional distance.

In the final model (Model C), we add the two elite variables to denote possible elite structures in research collaboration. Taken as a whole, these models are more accurate predictors of the determinants of research collaboration, indicated by the better fit expressed in the log likelihood, AIC and adjusted R^2. However, outcomes

Table 6 Zero-inflated negative binomial regression model on interaction intensity of co-patenting in semiconductors for the period 1988–2001

Parameter	Model A estimate(SE)	Model B estimate(SE)	Model C estimate(SE)
Negative binomial part			
Constant	0.206 (0.163)*	0.567 (0.224)*	0.596 (0.228)*
Mass origin (ln)	0.424 (0.020)*	0.427 (0.020)*	0.421 (0.021)*
Mass destination (ln)	0.452 (0.023)*	0.448 (0.022)*	0.443 (0.023)*
Distance (ln)	−0.585 (0.027)*	−0.614 (0.030)*	−0.612 (0.031)*
Same country		−0.233 (0.113)	−0.239 (0.114)
Excellence			0.131 (0.166)
Capital			−0.627 (0.734)
Zero inflated part			
Constant	−2.243 (0.163)*	1.069 (0.240)*	1.076 (0.241)*
Mass origin (ln)	−0.616 (0.021)*	−0.590 (0.021)*	−0.593 (0.021)*
Mass destination (ln)	−0.533 (0.021)*	−0.597 (0.021)*	−0.599 (0.021)*
Distance (ln)	1.560 (0.034)*	1.180 (0.037)*	1.186 (0.037)*
Same country		−1.654 (0.094)*	−1.672 (0.095)*
Excellence			–
Capital			−2.305 (0.837)*
Fit statistics			
Overdispersion (α)	1.690 (0.125)*	1.647 (0.120)*	1.635 (0.120)*
Vuong-statistic	13.52*	12.24	12.12*
Log Likelihood	−11996.461	−11757.489	−11751.202
Mc Fadden's Adj. R^2	0.396	0.408	0.408
AIC	0.028	0.027	0.027
N	865270	865270	865270
Nonzero observations	2196	2196	2196

* Indicates significance at 1% level

for publications differ from the outcomes for patents. In the publication system the coefficients of collaborations between excellence regions and capital regions are all positive and significant.[13] For the patenting system, we only find a bias between capital regions for biotechnology.[14]

[13] This finding is in line with a recent study by Tijssen (2007) who found that regions with higher quality of research (indicated by the mean citation rate) have a higher propensity to collaborate internationally.

[14] A possible explanation for the absence of an elite structure in collaborative patenting can be based on the differences between science system and the innovation system. In science, knowledge production is more a collective endeavour, while in patenting the major incentive for markets. This could explain why technology researchers in excellence regions show no particular bias to collaborate with other researchers in excellence regions.

6 Discussion

In this study we adopted a gravity framework to analyse inter-regional collaboration based on scientific publications and patents with multiple addresses. More specifically, we addressed the role of proximity and elite structures in collaborative knowledge production. The results for 1316 European regions indeed showed that these two determinants affect the formation of inter-regional collaborative networks. By doing so, we confirmed the role of geographical proximity as found by other studies, yet extended our understanding of other barriers to collaborate including national borders and elite structures stemming from cognitive and political structures.

Our results bear significant implications within a European policy context.[15] The outcomes with regard to the importance of proximity indicate that the European Union is far from having created an area in which '*research efforts at national and union level are integrated*'. In such a research area the choice for a collaboration partners should be based solely on scholarly ground, while we found that this choice is significantly impeded by geographical barriers. Hence, there is a clear need to further harmonise the national research systems, including the alignment of labour market regulations, diploma systems and property rights. The current spatial heterogeneity explains why most researchers are still heavily biased towards domestic collaboration, even though European collaboration could offer more opportunities in many cases. As there is evidence that the effect of geographical proximity exists independently of national borders, the process of integration within member states is incomplete too. This implies that the research policy efforts to promote international collaboration under the heading of the seventh framework programme should be complemented with efforts of member states to integrate their own national research systems.

Next to the significance of proximity in collaborative knowledge production, we also found evidence for elite structures in which regions that host quality scholars or financial resources are more inclined to network among themselves. This finding is not incompatible with the definition of ERA, as promoting elite structures is part of the agenda. With the recent emphasis in the Seventh Framework Programme on frontier research, both by individual researchers and in collaboration networks, the gap between these regions is expected to increase rather than decrease in the future.

Thus, our results suggest that within the European context facilitating research collaboration per se will not necessary contribute to increasing cohesion at the regional level. Rather, ERA policy will remove barriers related to geography thereby fostering integration and reinforcing the centralization of knowledge flows among already well-connected excellence regions and capital regions. Reading from the commission's recent green paper on ERA (Commission of the European Communities 2007) such an outcome should be considered as intended. Yet, if the objective of the EU is to implement an inclusive policy that promotes active participation of peripheral locations in the European Research Area, it should be more specific in their policies. Stimulating linkages between elite regions and peripheral regions is such an inclusive instrument. In this way, less well connected regions profit from access to knowledge in the elite

[15] For a more detailed policy discussion, see Frenken et al. (2007).

regions. At least, for peripheral locations such a strategy seems more effective than local research policies even if the two strategies are not mutually exclusive.

Acknowledgments We thank the Netherlands Institute for Spatial Research (RPB) for financial support, Hans van Amsterdam, Stephaan DeClerck and Joep van Vliet for their research assistance, and Martijn Burger, Harry Garretsen, Gaston Heimeriks, Mario Maggioni, Roderik Ponds, Gusta Renes, Jan Schuur, Erika Uberti, Eleftheria Vasileiadou and two anonymous referees for helpful comments. All errors remain ours.

References

Boschma RA (2005) Proximity and innovation. A critical assessment. Reg Stud 39(1):61–74

Burger MJ, Van Oort FG (2007) On the specification of the gravity model of trade: zero's, excess zero's and quasi Poisson estimation, working paper, Erasmus University Rotterdam, The Netherlands

Commission of the European Communities (2006) Amended proposal for a decision of the parliament and the council concerning the 7th Framework Programme of the European Community for research, technological development and demonstration activities (2007–2013), COM(2006), 28 June 2006, 364 p

Commission of the European Communities (2007) Green paper 'The European Research Area: New Perspectives', {SEC(2007) 412}, COM(2007)161 final, Brussels, 4 April 2007

Cooke P, Uranga MG, Extebarria G (1998) Regional innovation systems: an evolutionary perspective. Environ Plann A 30(9):1563–1584

Danell R, Persson O (2003) Regional R&D activities and interactions in the Swedish Triple Helix. Scientometrics 58(2):205–218

Edquist C, Johnson B (1997) Institutions and organisations in systems of innovations. In: Edquist C (ed) Systems of innovation. Technologies Institutions and Organizations. Pinter, London, pp 41–63

Ejermo O, Karlsson C (2006) Interregional inventor networks as studied by patent coinventorships. Res Policy 35(3):412–430

European Council (2000) Presidency conclusions, Lisbon European Council, 23–24 March 2000

Fleming L, Frenken K (2007) The evolution of inventor networks in the Silicon Valley and Boston regions. Adv Comp Sys 10(1):53–71

Frenken K (2002) A new indicator of European integration and an application to collaboration in scientific research. Econ Syst Res 14(4):345–361

Frenken K, Hoekman J, Van Oort F (2007) Towards a European Research Area. RPB/NAi Publishers, The Hague/Rotterdam

Frenken K, Hölzl W, de Vor F (2005) The citation impact of research collaborations: the case of European biotechnology & applied microbiology (1988–2002). J Eng Technol Manage 22(1–2):9–30

Gertler MS (1995) 'Being there': proximity, organization, and culture in the development and adoption of advanced manufacturing technologies. Econ Geogr 71(1):1–26

Glänzel W (2001) National characteristics in international scientific co-authorship relations. Scientometrics 51(1):69–115

Griliches Z (1990) Patent statistics as economic indicators: a survey. J Econ Literat 28:1661–1707

Katz JS (1994) Geographical proximity and scientific collaboration. Scientometrics 31(1):31–43

Katz JS, Martin BR (1997) What is research collaboration. Res Policy 26(1):1–18

Liang LM, Zhu L (2002) Major factors affecting China's inter-regional research collaboration: regional scientific productivity and geographical proximity. Scientometrics 55(2):287–316

Long JS (1997) Regression models for categorical and limited dependent variables. Advanced quantitative techniques in the social sciences, vol 7. Sage publications, Thousand Oaks

Maggioni MA, Uberti TE (2007) International networks of knowledge flows: an econometric analysis. In: Frenken K (ed) Applied evolutionary economics and economic geography. Edward Elgar, Cheltenham, pp 230–255

Meyer M, Bhattacharya S (2004) Commonalities and differences between scholarly and technical collaboration. An exploration of co-invention and co-authorship analyses. Scientometrics 61(3):443–456
Narin F, Stevens K, Whitlow ES (1991) Scientific cooperation in Europe and the citation of multinationally authored papers. Scientometrics 21(3):313–323
Ponds R, Van Oort FG, Frenken K (2007) The geographical and institutional proximity of scientific collaboration networks. Pap Reg Sci 86(3):423–443
Roy JR, Thill J (2004) Spatial interaction modelling. Pap Reg Sci 83(1–2):339–361
Sen A, Smith TE (1995) Gravity modelling of spatial interaction behaviour. Springer, Berlin
Tijssen RJW (2007) Research cooperation within an expanding European Union. Geographical patterns and recent trends in international, domestic and inter-regional co-publications, Working paper CWTS. Leiden University, The Netherlands
Tinbergen J (1962) Shaping the world economy, suggestions for an international economic policy. The 20th century fund, New York
Verbeek A, Debackere K, Luwel M (2003) Science cited in patents: a geographic "flow" analysis of bibliographic citation patterns in patents. Scientometrics 58(2):241–263
Wagner-Doebler R (2001) Continuity and discontinuity of collaboration behaviour since 1800—from a bibliometric point of view. Scientometrics 52(3):503–517