The final published version of this paper is available at:
http://www.springerlink.com/index/10.1007/s11192-010-0305-6

Please, quote as: Acosta, M., Coronado, D., Ferrándiz, E., y León, M. D. 2011. Factors affecting inter-regional academic scientific collaboration within Europe: the role of economic distance. *Scientometrics*, 87(1): 63-74.

Factors affecting inter-regional academic scientific collaboration within Europe: The role of economic distance

Manuel Acosta
Daniel Coronado
Esther Ferrándiz
Mª Dolores León
UNIVERSITY OF CADIZ

Facultad de Ciencias Económicas y Empresariales, Universidad de Cádiz.
c/ Duque de Nájera, 8
11002-CADIZ
SPAIN

Corresponding author:

Daniel Coronado
e-mail: daniel.coronado@uca.es
Tf: 34-956015415
Fax: 34-956015388

Abstract: This paper offers some insights into scientific collaboration (SC) at the regional level by drawing upon two lines of inquiry. The first involves examining the spatial patterns of university SC across the EU-15 (all countries belonging to the European Union between 1995 and 2004). The second consists of extending the current empirical analysis on regional SC collaboration by including the economic distance between regions in the model along with other variables suggested by the extant literature. The methodology relies on co-publications as a proxy for academic collaboration, and in order to test the relevance of economic distance for the intensity of collaboration between regions, we put forward a gravity equation. The descriptive results show that there are significant differences in the production of academic scientific papers between less-favoured regions and core regions. However, the intensity of collaboration is similar in both types of regions. Our econometric findings suggest that differences in scientific resources (as measured by R&D expenditure) between regions are relevant in explaining academic scientific collaborations, while distance in the level of development (as measured by per capita GDP) does not appear to play any significant role. Nevertheless, other variables in the analysis, including geographical distance, specialization and cultural factors, do yield significant estimated coefficients, and this is consistent with the previous literature on regional SC.

Key words: economic distance, academic scientific collaboration, gravity equation, co-authorship

MSC classifications: 62P20, 62J02, 97K40
JEL classifications: R11, O33, I23

Introduction

In this paper, we provide some insight into the factors affecting SC using two lines of inquiry. The first involves examining the university SC patterns across Nomenclature of Territorial Units for Statistics (NUTS) II regions in the EU-15 (all countries belonging to the European Union between 1995 and 2004). The second line of inquiry consists of extending the empirical literature by introducing economic aspects, such as the role of economic distance (regional differences in the levels of development and higher education research and development (R&D) expenditures), along with other variables suggested in the literature. This paper then addresses a problem in the small literature on this topic, whereby there is some focus on the effects of different kinds of distances (geographical, institutional and cultural), but the effects of economic factors at this scale remain largely unknown, especially in Europe.

The dataset used to measure regional scientific production and collaboration consists of a regionalized sample of 994,938 scientific papers by authors affiliated with European universities from 1998 to 2004. The data were obtained from the Thomson ISI (Information Sciences Institute) database and include papers from all scientific fields (except the social sciences and humanities) for over 500 European universities at the NUTS II level of regional aggregation. The empirical analysis involves two parts. First, we carry out a brief descriptive study of the co-publication patterns across regions. Second, we put forward an econometric framework to test the relevance of the economic factors, along with other variables, on the intensity of collaboration. The paper is organized as follows. Section 2 reviews the relevant literature. Section 3 describes the methodology. Section 4 details the data. Section 5 provides the results. The main conclusions and policy implications are given in the final section.

Empirical background

As stated by the scarce literature on the collaborative patterns of research at the regional level (Liang and Zhu, 2002; Okubo and Zitt; 2004; Ponds et al., 2007; Boshoff, 2010; and Hoekman et al., 2009, 2010), geographical distance, similarities in the specialization patterns of SC in regions and cultural proximity are significant factors that positively affect SC between regions. However, no previous research has considered the role of economic factors on SC at a regional scale. We need to turn to the international literature on collaboration to find suitable references. In this regard, the centre–periphery discussion at the international level (Schott, 1998; Schubert and Sooryamoorthy, 2010) provides some clues on whether differences in economic development between areas may determine the patterns of scientific collaboration. According to this hypothesis, peripheral countries are willing to collaborate in order to gain access to resources, while core (centre) countries collaborate for the purpose of complementarities. For example, using interviews, Hwang (2008) concluded that the main aim for Korean scientists and engineers in international collaboration was to obtain advanced knowledge and technologies from core scientists in exchange for funding core knowledge production. Sonnenwald (2007) described some examples of collaboration between Africans and non-Africans; Africans granted access to local communities and non-Africans provided free treatment, lab equipment and training. Also between China and Taiwan; Taiwan provided the experienced, mid-career scientists that China lacked because of the Cultural Revolution and China provided a large number of younger scientists to increase the size of Taiwan’s scientific community. Despite this promising work, empirical research including specific economic indicators is limited and the evidence is weak. In light of this discussion, some scepticism on the centre–periphery hypothesis remains. For example, Wagner and Leydesdorff (2005) pointed out that the centre–periphery model does not explain the dynamic through which scientific centres both collaborate and compete with one another for partners at the international level. Some doubt also remains on the role of the level of economic resources devoted to R&D on international collaboration. In fact, among the scant empirical evidence, Kim (2005) surprisingly concluded a negative relationship between R&D expenditure and international collaboration in Korea.

Methodology

In order to test the role of economic distance along with other variables on regional academic scientific collaboration, we put forward a gravity equation. The gravity model has been increasingly applied in a number of studies of the regional scientific collaboration (Ponds et al., 2007; Scherngell and Barber, 2009; Hoekman et al., 2010). Typically, a simple version of the gravity equation takes the form:

$$Fij = \alpha_0 Y_i^{\alpha_1} Y_j^{\alpha_2} D_{ij}^{\alpha_3} \qquad \text{(Eq. 1)}$$

where F (scientific collaboration) is a function of characteristics of the origin (Y_i), characteristics of the destination (Y_j), and some measurement of distance between both areas (D_{ij}). To account for deviations from the theory, stochastic versions of the equation are used in empirical studies by adding an independent stochastic error term. Given the count nature of our data and the large number of zero observations in the sample, we estimate the gravity equation using a negative zero-inflated binomial (ZINB) model in which the collaboration count between regions is specified as an exponential mean regression model. The zeros for the several regions (observations) not including collaboration potentially arise from two sources. The first distinguishes those regions with no potential for scientific collaboration (for example, one or both do not have any publications during the initial period of the sample). The second source stems from those regions with the capacity to collaborate, but which did not present any collaboration in the observed period. The model takes the form:

$$\Pr(F_{ij}) = \begin{cases} \lambda_{ij} + (1-\lambda_{ij})h(F_{ij} = 0, \theta \mid X) & for\ \ F_{ij} = 0 \\ (1-\lambda_{ij})h(F_{ij}, \theta \mid X) & for\ \ F_{ij} = 1,2,... \end{cases} \qquad \text{(Eq. 2)}$$

where $h(F_{ij}, \theta \mid X)$ is the negative binomial density with mean $\exp(X, \beta)$, dispersion parameter α, and $\theta = (\beta' \alpha)'$. Here, λ is a zero-inflation parameter representing the proportion of observations with a strictly zero count $(0 < \lambda < 1)$ as determined by a logit model on all (or several) observed explanatory variables: $\lambda_{ij} = \exp(X\varphi)/1 - \exp(X\varphi)$.

The dependent variable F_{ij} = $\mathbf{Asc_{ijt}}$ represents the counts of academic scientific collaboration between region *i* and region *j* for the period *t*. We use co-publications to measure collaboration among regions, as it is a well-established indicator with a long tradition in scientific collaboration studies at both the individual and international levels of analysis (for reviews, see Melin and Persson (1996), Katz and Martin (1997) and Laudel (2002)).

X is a vector including the following independent variables as suggested by the literature:

$\mathbf{Pub_{it0}}$ is the number of academic scientific publications in region *i* for a period t_0 before the collaboration takes place.

$\mathbf{Pub_{jt0}}$ is the number of academic scientific publications in region *j* for a period t_0 before the collaboration takes place.

$\mathbf{Gdist_{ij}}$ is the geographical distance in kilometres between the capitals of regions *i* and *j*.

$\mathbf{Cont_{ij}}$ is a dummy variable that takes a value of 1 if *i* and *j* are contiguous regions, 0 otherwise. Note that while this is an easy way to measure regional proximity, it does have several drawbacks. Of these, the main limitation is that we are unable to capture differences in the size of the regions, the number of bordering regions and the regional concentration of higher education institutions using this simple dummy variable.

Country$_{\mathbf{ij}}$ is a dummy variable that takes a value of 1 when regions *i* and *j* are in the same country, 0 otherwise. This is because regions in the same country usually share a similar culture, language, policies, etc. This variable helps capture these factors.

Spec$_{\mathbf{ijt0}}$ is the proximity in scientific specialization between regions *i* and *j* in period t_0. This variable is measured using an index similar to that proposed by Peri (2005) of the correlation coefficient between the 12-field composition of scientific papers in regions *i* and *j*.

In order to contrast the centre–periphery hypothesis in collaboration patterns across regions, we also include in our model two variables to account for the effects of economic distance between pairs of regions:

Edist$_{\mathbf{ijt0}}$: economic inequality between regions *i* and *j* is proxied with the absolute difference in per capita income for the period t_0.

RDdist$_{\mathbf{ijt0}}$: the difference in academic economic resources between regions *i* and *j* is captured with the absolute difference in per capita higher education R&D expenditure for the period t_0.

To prevent endogeneity, the explanatory variables refer to the initial period t_0, that is, before collaboration takes place. This is because although the theoretical gravity equation establishes that the number of collaborations between a pair of regions depends on the "mass" of publications in each region, reverse causality is also possible, given the effect that collaboration may exert on scientific productivity. Details on the estimation procedure of the ZINB model can be found in Long (1997) and Cameron and Trivedi (1998, 2009).

Data

The empirical data used in this study comprises a set of research articles published in scientific journals indexed by the Science Citation Index Expanded (SCI). As is well known, the SCI is a bibliographical database produced by the Information Sciences Institute (ISI), which is in turn a part of Thomson Reuters' Web of Science. The main advantage of ISI citation indexes is that they provide a complete list of all authors and their affiliations. There are also some known limitations of this database. For example, it does not include all journals, and the ISI journal list is biased towards journals published in English. At the regional level, collaboration takes place when a paper is co-authored by researchers affiliated with universities located in different regions.

The procedure to account for collaboration between pairs of NUTS II regions in the EU-15 followed these four steps: i) Data on academic publications containing at least one author affiliated with a university from an EU-15 country for 1998–2004 were retrieved from the SCI. We included several search terms to help identify higher education institutions in both English and other languages (*fachhochschule, yliopisto, ecole, institut nacional polytehcnique, politécnico, scuola, hogskola,* etc.*).* This search resulted in 994,938 publications.

ii) The second step involved regionalization at the NUTS II level of aggregation of the academic publications obtained in Step 1 (213 regions[1]). We first identified the NUTS II associated with each university using the list provided by the members of the

[1] Number of regions in the EU-15 according to Regulation (EC) No. 1059/2003 of the European Parliament and the Council of 26 May 2003 on the establishment of a common classification of territorial units for statistics (NUTS) (excluding extra-regio).

European Indicators, Cyberspace and the Science-Technology-Economy System (EICSTES). For those universities not included in the EICSTES list, we searched for the address on each university's website. Following the full-count process (assigning the entire publication to those regions that collaborated in its production), we obtained 1,206,644 publications and 387,545 inter-regional collaborations. iii) The third step involves classification by scientific field. We grouped the ISI categories into 12 broad scientific disciplines using the Third European Report on S&T indicators[2]. iv) The fourth step provided the collaboration matrix between regions. The data on scientific collaboration was placed into a (213 × 213) symmetrical matrix containing all co-publications between regions. Each cell then includes the number of scientific co-publications between region *i* and region *j*, and therefore excludes domestic collaboration (academic scientific collaboration between researchers in the same region). Consequently, there are potentially (213 × 212) ÷ 2 =22,548 collaboration links (observations) in the EU-15 at the NUTS II scale of analysis.

A summary of the main statistics from the full sample is reported in Table 1. Note that the number of academic publications increased by 22.83% from 1998 to 2004, while the number of regional collaborations increased by 51.28% over the same period.

[Table 1. About here]

The indexes in Table 1 reveal that both the production of scientific knowledge and the patterns of scientific collaboration present a high level of concentration in a few regions. As shown in Table 1, the Gini coefficient for publications takes a value of 0.61 for the

[2] The classification was established by the Centre for Science and Technology Studies (CWTS) at Leiden University (see Tijssen and van Leeuwen, 2003). For categories not included in the CWTS 2003 classification, we used an updated (but unpublished) classification kindly provided by the CWTS.

initial year (1998) and 0.59 for the latest year (2004) in the sample. The value of the Gini coefficient is slightly lower for regional collaborations. Further, as shown, the trend in the Gini coefficients for both publications and collaborations is slightly downward over the period 1998 to 2004. The remaining concentration indexes in Table 1 lead to the same conclusion. For example, the value of the C5 index takes a value of about 12 for collaborations, suggesting that just five regions account for 12% of papers co-authored with academics in other regions. Likewise, the value of the C10 index is 21, indicating that 10 regions provide 21% of co-authored papers.

Table 2 details the number of academic papers and the number of papers in collaboration by type of NUTS region. Drawing on this Table, it is clear that 29% of the less-developed EU-15 regions contributed to only 15.7% of the EU-15 published papers. On average, the capacity for publication of a region in this group is about 45% of the capacity of a developed region in the core group. The disparities are rather stronger when we consider a classification of regions based on higher education R&D expenditure. However, despite this apparently unbalanced picture of the generation of academic papers, the intensity of collaboration is similar in both groups of regions. The main question to respond to in the next section is to what extent economic distance is an obstacle to collaboration between regions with different levels of development or university R&D resources.

Results

The empirical equation of the gravity model was run using cross-sectional data where the dependent variable, **Asc**, includes the counts of academic scientific collaboration between EU-15 regions from 1998 to 2004. The explanatory variables capturing the *mass* of publications, specialization and economic distances refer to the initial year, 1998. Descriptive statistics are presented in Table 3.

[Table 3. About here]

We estimated three models (Table 4). Model I includes the number of publications (in logs) in each region, measures of distance and the dummy variable capturing cultural factors. Model II contains the variable capturing the similarities in scientific specialization between regions, along with the explanatory variables in Model I. We first estimated these two models in order to determine whether the effects of the factors affecting academic collaboration between regions provide similar behaviour to that in the empirical literature. Model III adds two new variables capturing the differences between regions in terms of economic development and resources. Note that the number of observations is different for each model. To start with, Model I was estimated with all possible observations between the pairs of regions. In Model II, some observations were excluded on the basis of the variable **Spec**, because it does not make sense to obtain the coefficient of correlation between scientific specializations in regions *i* and *j* when one or both have no publications. Finally, Model III was obtained with fewer observations because of missing data on higher education R&D expenditure for some regions.

[Table 4. About here]

To test the reliability of these estimates, we follow a top-down procedure, and first estimate the ZINB equations and then other count data models, including Zero-Inflated Poisson, Negative Binomial and Poisson (results not presented). For the purposes of comparison, we applied the usual statistics of over-dispersion and the LR test in Vuong (1989) (see Table 4). In all cases, the ZINB models were preferred. Based on this table, the main findings suggest that:

1. The coefficients of the variables capturing the differences in the levels of development and economic resources devoted to R&D in universities (Model 3) both display a negative sign, suggesting that the greater the economic difference between two regions, the fewer the number of collaborations. However, the economic distance measured as the absolute differences in per capita income does not affect collaborative behaviour, while the absolute differences in the level of economic resources devoted to university R&D are highly significant. Together, these results indicate that regions tend to collaborate with other regions independently of their level of economic development. They collaborate with regions with similar characteristics in terms of the level of resources devoted to R&D[3]. These results are inconsistent with the SC centre–periphery hypothesis because, as explained earlier and according to the SC centre–periphery hypothesis, we expect a positive relationship between SC and economic distance. Although our data do not provide reasons for this result, a tentative explanation is that the economic distance between our pairs of regions is probably not sufficiently wide, as it is in other contexts where this hypothesis holds (as in, say, Boshoff's (2010) study of SC between African and non-African countries). In addition, the significant negative relationship between SC and R&D distance in European regions is not entirely unexpected, as the greater the amount of resources, the greater the opportunities to attend international conferences and to engage in collaboration. Moreover, core regions may not find complementarities with less-developed regions (those with scarce resources devoted to R&D).

[3] Note that the level of development and university R&D expenditure (or their differences) do not have to be necessarily related. For example, some regions may have a high level of development because of tertiary activities (such as tourism) that have little to do with university R&D expenditure.

2. The other variables included in the models present signs and coefficient as expected and consistent with previous empirical literature:

a) The coefficients of the variables capturing the number of publications for each pair of regions are both significant and have positive signs. This is a natural result because the mass of publications usually implies more researchers in each region and therefore more opportunities for collaboration.

b) The coefficients of the two variables of distance are both significant, and also have their expected signs. Accordingly, geographical distance and contiguity are both relevant variables in explaining academic scientific collaboration between regions. The negative sign of the first variable indicates that collaboration decreases with distance, while the positive sign of the second variable shows that bordered regions explain their scientific collaborative behaviour. The main argument explaining this result relies on the fact that collaboration usually requires the mobility of researchers; that is, coordination, knowledge sharing and feedback sometimes require face-to-face contact.

c) The variable capturing the correlation between the levels of scientific specialization between regions displays a positive sign, suggesting that proximity in scientific specialization is significant in explaining the number of collaborations between regions.

d) Finally, the binary variable capturing collaboration between regions in the same country is also relevant. This suggests that cultural similarities and other characteristics, such as a common language or policies, help explain scientific collaboration between regions.

Note that these findings hold for all three models. This means that a reduction in the number of observations used in estimating the models produces some change in the estimated coefficients, but not in their levels of significance.

Conclusions

This paper attempts to identify the spatial distribution of academic scientific collaboration patterns across European regions, and is mainly aimed at evaluating the role of economic differences between regions. A preliminary descriptive analysis suggests a growing trend in collaboration between regions, increasing from 28.35% of co-authored publications in 1998 to 34.92% in 2004. The data also displays a high level of concentration of SC in a few regions, with little change over the period 1998–2004. The separation of regions according to different levels of economic development indicates that an Objective 1 region (one with a GDP per capita less than 75% of the EU-15 mean) produced on average less than half (45%) the papers of a more economically advanced region. However, both groups of regions display a similar rate of publications involving collaboration with other regions.

Another important question we responded to in the empirical analysis was the extent to which economic distance is an obstacle to collaboration between regions with different levels of development and/or university R&D resources. For this purpose, we estimated a gravity equation using empirical ZINB models for the period 1998–2004. The results lead to the following conclusions:

- The centre–periphery hypothesis applying to SC suggests that researchers in research-lagging countries are willing to collaborate with those in core countries in order to gain access to resources, while researchers in core countries collaborate by seeking complementarities. According to our analysis, this hypothesis does not hold at the regional level in the EU-15. From a policy viewpoint, this finding suggests that if collaboration becomes a priority, economic distance (in terms of per capita R&D expenditures) needs to be reduced in order to successfully attain the fulfilment of a European Research Area.

- Other results in the gravity model indicate that there are also other variables explaining SC between European regions. In particular, we found that the number of publications in the initial year, geographical distance and border contiguity, similarities in scientific specialization between the two regions, and the sharing of similar languages, cultures and policies, also help explain SC. Results concerning the relevance of these variables are similar to those obtained in previous work.

Finally, the focus of this paper was to analyze the relationship between SC and economic distance, along with other variables, but we have offered only a few clues on the reasons for our outcomes. Further research is necessary to explain, for example, the variables capturing economic distance in European regions, particularly whether transport costs can explain the negative effect of geographical distance on SC.

Acknowledgements

The authors would like to thank Martin Feldkircher (Austrian National Bank) for providing the spatial weight contiguity matrix. They are also grateful to Raffaele Paci and Barbara Dettori from the Centro Ricerche Economiche Nord Sud (CRENOS) at the University of Cagliari for their assistance in the construction of the distance matrix and for providing the coordinates of the centre regions, and Robert Tijssen of the Centre for Science and Technology Studies (CWTS) at Leiden University for providing the updated classification. An early version of this paper was presented at the 12th European Network on Industrial Policy (EUNIP) International Conference held in Reus, Spain, from 9–11 June 2010. We thank the conference participants, particularly James Wilson, for their helpful comments and suggestions. The authors are also very grateful to the reviewers for constructive and insightful comments. This work was supported by the Consejería de Innovación, Ciencia y Empresa (Junta de Andalucía-Spain) [Grants P06-SEJ-02087 and P08-SEJ-3981].

References

Boshoff, N. (2010). South–South Research Collaboration of Countries in the Southern African Development Community (SADC). *Scientometrics,* 84, 481–503.

Cameron, C., Trivedi, P.K. (1998). *Regression Analysis of Count Data.* Cambridge: Cambridge University Press.

Cameron, C., Trivedi, P.K. (2009). *Microeconometrics using Stata*. College Station Texas: Stata Press.

Hoekman, J., Frenken, K., Oort, F. (2009). The Geography of Collaborative Knowledge Production in Europe. *Annals of Regional Science*, 43, 721–738.

Hoekman, J., Frenken K., Tijssen, R.J.W. (2010). Research Collaboration at a Distance: Changing Spatial Patterns of Scientific Collaboration within Europe. *Research Policy,* doi:10.1016/j.respol.2010.01.012.

Hwang, K. (2008). International Collaboration in Multilayered Center–Periphery in the Globalization of Science and Technology. *Science Technology Human Values*, 33, 101–133.

Katz, J.S., Martin, B.R. (1997). What is Research Collaboration? *Research Policy*, 26, 1–18.

Kim, M.-J. (2005). Korean science and international collaboration, 1995–2000. *Scientometrics*, 63, 321–339.

Laudel, G. (2002). What Do We Measure by Co-authorships? *Research Evaluation,* 11, 3–15.

Liang, L., Zhu, L. (2002). Major Factors Affecting China’s Inter-regional Research Collaboration: Regional Scientific Productivity and Geographical Proximity. *Scientometrics*, 55, 287–316.

Long, J.S. (1997). *Regression Models for Categorical and Limited Dependent Variables,* Sage, Thousand Oaks.

Melin, G., Persson, O. (1996). Studying Research Collaboration using Co-authorships. *Scientometrics*, 36, 363–377.

Okubo, Y., Zitt, M. (2004). Searching for Research Integration across Europe: A Closer Look at International and Inter-regional Collaboration in France. *Science and Public Policy,* 31, 213–226.

Peri, G. (2005). Determinants of Knowledge Flows and their Effect on Innovation. *Review of Economics and Statistics,* 87, 308–322.

Ponds, R., van Oort, F., Frenken, K. (2007). The Geographical and Institutional Proximity of Research Collaboration. *Papers in Regional Science,* 86, 423–443.

Scherngell, T., Barber, M.J. (2009). Spatial Interaction Modelling of Cross-region R&D Collaborations: Empirical Evidence from the 5th EU Framework Programme. *Papers in Regional Science*, 88, 531–546.

Schott, T. (1998). Ties between Center and Periphery in the Scientific World-System: Accumulation of Rewards, Dominance and Self-Reliance in the Center. *Journal of World-Systems Research*, 4, 112–144.

Schubert, T., Sooryamoorthy, R. (2010). Can the Centre–Periphery Model Explain Patterns of International Scientific Collaboration among Threshold and Industrialised Countries? The Case of South Africa and Germany. *Scientometrics*, 83, 181–203.

Sonnenwald, D.H. (2007). Scientific Collaboration: A Synthesis of Challenges and Strategies, in: Cronin, B. (Ed.), *Annual Review of Information Science and Technology*, vol. 41. Medford NJ: Information Today, Inc.

Tijssen, R.J.W, van Leeuwen, T.N., (2003). Bibliometric Analyses of World Science, Extended Technical Annex to Chapter 5 of the Third European Report on Science & Technology Indicators.

Vuong, Q.H. (1989), Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica*, 57, 307–33.

Wagner, C.S., Leydesdorff, L. (2005). Mapping the Network of Global Science: Comparing International Co-authorship from 1990 to 2000. *International Journal of Technology and Globalization*, 1, 185–208.

Table 1. Evolution and regional concentration indexes of academic scientific publications and collaborations 1998-2004

		1998	1999	2000	2001	2002	2003	2004	1998–2004
Publications	N1	157,446	164,492	166,669	170,603	174,266	179,770	193,398	1,206,644
	Mean	739.19	772.27	782.49	800.96	818.16	844.00	907.98	5,664.99
	Max.	5,794	5,950	5,887	6,162	6,186	6,401	6,701	43,081
	Min.	0	0	0	0	0	0	0	0
	Std Dev.	937.41	972.52	976.27	995.60	1,013.52	1,046.23	1,100.44	7,024.09
	C. Var.(1)	1.26	1.25	1.24	1.24	1.23	1.23	1.21	1.23
	Gini coeff. (2)	0.61	0.60	0.60	0.60	0.59	0.59	0.58	0.59
	C5(3)	13.38	13.38	13.16	13.31	13.12	13.32	12.91	13.18
	C10(4)	23.04	22.88	22.82	22.65	22.41	22.64	22.13	22.61
	C25(5)	44.86	44.69	44.13	44.05	44.10	44.06	43.37	44.02
Collaborations	N2	44,643	48,588	51,459	55,169	58,203	61,947	67,536	387,545
	(N2÷N1)×100	28.35	29.54	30.87	32.34	33.40	34.46	34.92	32.12
	Mean	209.59	228.11	241.59	259.00	273.25	290.83	317.07	1,819.46
	Max.	1,374	1,566	1,665	1,821	1,882	2,020	2,180	12,508
	Min.	0	0	0	0	0	0	0	0
	Std Dev.	251.90	272.54	287.61	304.44	323.30	346.52	369.17	2,149.98
	C. Var.(1)	1.20	1.19	1.19	1.17	1.18	1.19	1.16	1.18
	Gini coeff. (2)	0.59	0.59	0.58	0.58	0.58	0.58	0.57	0.56
	C5(3)	12.35	12.31	12.30	12.30	12.29	12.38	11.98	12.22
	C10(4)	21.94	21.74	21.77	21.37	21.55	21.66	21.02	21.53
	C25(5)	42.78	42.66	42.50	41.64	42.19	42.54	41.84	42.22

(1) Coefficient of variation = Std Dev. ÷ Mean; (2) The Gini coefficient ranges between 0 and 1; the larger the value the higher the level of regional concentration in publications or collaborations. (3)(4)(5) Concentration indexes of publications for the top 5, 10 and 25 regions with the largest number of scientific papers, respectively.

Table 2. Regional production of academic papers and collaborative papers by type of NUTS region (*)

Groups of regions according to their level of development					Groups of regions according their level of higher education R&D expenditure				
		1998	2004	1998–2004 % increase			1998	2004	1998–2004 % increase
Regions with less than 75% of the EU-15 average GDP per capita	A. No. Papers	20,996	30,463	45.09	Regions with less than 75% of the EU-15 average R&D per capita	A. No. Papers	12,064	15,855	31.42
	Mean	338.65	491.34			Mean	236.55	317.10	
	Std. Dev	434.08	589.93			Std Dev.	319.20	430.61	
	B. No. Coll.	6,108	10,548	72.69		B. No. Coll.	3,443	5,905	71.51
	Mean	98.52	170.13			Mean	67.51	118.10	
	Std Dev.	122.34	197.46			Std Dev.	78.00	150.39	
	(B/A)*100	29.09	34.63	19.02		(B/A)*100	28.54	37.24	30.50
Regions with more than 75% of the EU-15 average GDP per capita	A. No. Papers	136,450	162,935	19.41	Regions with more than 75% of the EU-15 average R&D per capita	A. No. Papers	86,905	102,274	17.68
	Mean	903.64	1,079.04			Mean	1,259.49	1,461.06	
	Std Dev.	1,035.29	1,211.95			Std Dev.	1,054.00	1,214.21	
	B. No. Coll.	38,535	56,988	47.89		B. No. Coll.	24,635	34,998	42.07
	Mean	255.20	377.40			Mean	357.03	507.22	
	Std Dev.	276.41	405.18			Std Dev.	278.48	398.17	
	(B/A)*100	28.24	34.98	23.85		(B/A)*100	28.35	34.22	20.72

(*) The group of less-developed regions comprises 62 NUTS regions, where the number of NUTs regions with more than 75% of the EU-15 average GDP per capita is 151 (213 in total). Because of the lack of data, the number of regions with less and more than 75% of EU-15 average university R&D per capita falls to 120, with 51 NUTS regions in the first group and 69 in the second group.

Table 3. Descriptive statistics

Variable	Mean	Std Dev.	Min.	Max.
Asc_{ij}	23.9821	73.0597	0	1,487
$Ln(Pub_i)$	6.1460	1.6060	0	8.6645
$Ln\ (Pub_j)$	6.2067	1.5884	0	8.6645
$Gdist_{ij}$	10.9701	7.6767	0.4228	60.0071
Country	0.1855	0.3887	0	1
$Cont_{ij}$	0.0341	0.1815	0	1
$Spec_{ij}$	0.4618	0.3639	–0.5502	0.9905
$Edist_{ij}$	7.9900	6.1366	0.0023	36.7493
$RDdist_{ij}$	0.2890	0.2893	0	1.4652

Table 4. Results of zero-inflated negative binomial regressions

	MODEL 1 (ZINB)			MODEL 2 (ZINB)			MODEL 3 (ZINB)		
	Coeff.	Std Err.		Coeff.	Std Err.		Coeff.	Std Err.	
Constant	–6.9972	0.1044	***	–6.7297	.1041	***	–8.905	0.1635	***
Pub_i	0.7614	0.0103	***	0.7083	.0104	***	0.8113	0.0149	***
Pub_j	0.6310	0.0093	***	0.5865	.0096	***	0.7976	0.0155	***
$Gdist_{ij}$	–0.0152	0.0013	***	–0.0131	.0013	***	–0.1798	0.0027	***
Country	1.8186	0.0303	***	1.7509	.0303	***	1.9096	0.0352	***
$Cont_{ij}$	0.8971	0.0615	***	0.9138	.0607	***	0.8846	0.0600	***
$Spec_{ij}$				0.6405	.0389	***	0.5934	0.0483	***
$Edist_{ij}$							–0.0044	0.0027	
$RDdist_{ij}$							–0.1716	0.0513	***
Inflated (logit)									
Constant	8.3158	0.3389	***	8.4406	.3647	***	9.7186	0.7727	***
Pub_i	–0.8210	0.0371	***	–0.8416	.03973	***	–0.9331	0.0801	***
Pub_j	–0.8603	0.0289	***	–0.8071	.0338	***	–0.9498	0.0812	***
$Gdist_{ij}$	–0.0049	0.0054		–0.0086	.0055		–0.0187	0.0153	
Country	–1.8372	0.1961	***	–1.9299	.2105	***	–3.4769	0.6475	***
$Cont_{ij}$	–2.2896	0.6269	***	–2.2542	1.061	***	–0.7423	1.0202	***
$Spec_{ij}$				–0.6781	.1731	***	0.2543	0.3212	***
$Edist_{ij}$							0.0189	0.0167	
$RDdist_{ij}$							0.2187	0.3470	
LnAlpha	–0.1089		***	–0.1441		***	–0.6509		***
Alpha	0.8967			0.8682			0.5215		
LR-test	12,627.72		***	12,863.10		***	6,899.73		***
Likelihood-ratio test alpha	1.2e+05		***	1.2e+05		***	3.7e+04		***
Vuong test	7.75		***	8.37		***	4.24		***
No. obs.	22,578			16,110			5,978		

* p < 0.10, ** p < 0.05, *** p < 0.01