Journal of Informetrics 7 (2013) 966–971

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Do more distant collaborations have more citation impact?

Önder Nomaler[a], Koen Frenken[a], Gaston Heimeriks[b,*]

[a] School of Innovation Sciences, Eindhoven University of Technology, The Netherlands
[b] Department of Innovation Studies, Utrecht University, The Netherlands

ARTICLE INFO

Article history:
Received 17 July 2013
Received in revised form
14 September 2013
Accepted 3 October 2013

Keywords:
Citation impact
Collaborations
Distance
Country effects

ABSTRACT

Internationally co-authored papers are known to have more citation impact than nationally co-authored paper, on average. However, the question of whether there are systematic differences between pairs of collaborating countries in terms of the citation impact of their joint output, has remained unanswered. On the basis of all scientific papers published in 2000 and co-authored by two or more European countries, we show that citation impact increases with the geographical distance between the collaborating counties.

1. Introduction

International collaboration is a salient feature of present-day scientific research. Especially since the 1990s, a rapid rise occurred in internationally co-authored papers (Doré, Ojasoo, & Okubo, 1996; Georghiou, 1998; Glänzel, 2001). The increase was dramatic: the share of internationally co-authored publications doubled between 1990 and 2000 (Wagner & Leydesdorff, 2005). The number of internationally co-authored articles grew at a rate faster than traditional nationally-co-authored articles (NSB, 2002). This trend continued after 2000 (Hoekman, Frenken, & Tijssen, 2010). While there are large differences among fields in the number of international co-authorships (Heimeriks, 2013; Hoekman et al., 2010), an increase can be seen across all fields of science at more or less the same rate (Hoekman et al., 2010; Wagner & Leydesdorff, 2005).

A striking feature of internationally co-authored papers is the tendency of their citation impact to be systematically higher than that of nationally co-authored papers (Frenken, Hardeman, & Hoekman, 2009; Narin, Stevens, & Whitlow, 1991). This pattern suggests that, on average, scientists will have more impact by international partnering as opposed to national partnering. Though the citation premium for internationally co-authored papers is well known, it is unlikely that all pairs of countries equally gain from collaboration. Our question is: what explains the variation in the citations (if any) received by internationally co-authored papers? Using data for over 33,000 papers concerning all collaborations in Europe in 2000, our main result holds that citation impact increases with the geographical distance between the collaborating counties.

2. International co-authorship and its citation impact

Since the study by Narin et al. (1991) on international scientific collaboration, several studies have noted the citation premium enjoyed by internationally co-authored papers compared to nationally co-authored ones. They found that

* Corresponding author. Tel.: +31 302537802.
E-mail addresses: G.J.Heimeriks@uu.nl, gheimeriks@gmail.com (G. Heimeriks).

1751-1577/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.joi.2013.10.001

co-publications involving affiliations to several European countries were twice as heavily cited as papers reporting a single EC country affiliation. This finding has been confirmed by later studies such as the ones by Frenken, Hölzl, and De Vor (2005), Frenken, Ponds and Van Oort (2010), Persson, Glänzel, and Danell (2004), (for a review, see Frenken et al., 2009).

The difference in citation impact between nationally and internationally co-authored papers suggests that scientists have something to gain by covering distance. One type of explanation of the citation premium enjoyed by internationally co-authored papers points to the content and the quality of the underlying research: international research projects may be, on average, more creative and important than nationally co-authored ones. Recombining resources from centers located in different national systems and traditions can be expected to lead to more unique outcomes. That is, breakthrough innovations often stem from recombining ideas that previously have been remained unconnected (Fleming, 2001). Indeed, the higher expenses for international projects compared to national projects need to be legitimized by better prospects in terms of research output.

There is, however, good reason to believe that quality provides only a partial explanation for the citation premium. A second explanation, which does not necessarily exclude the first one, holds that the output of international projects diffuses more widely than nationally co-authored papers. Since research tends to be more cited in the countries where authors originate from (Pasterkamp, Rotmans, Kleijn, & Borst, 2007), one expects internationally co-authored papers to be cited more than nationally co-authored papers.

The question we pose here is a different one: rather than investigating the citation differences between nationally and internationally co-authored papers, we are interested in explaining differences in citation impact among internationally co-authored papers. We expect that the citation impact of papers will be systematically different for different pairs of collaborating countries. In particular, we expect that collaboration between more distant countries will have greater impact than collaboration between closer countries.

The reasoning underlying our hypothesis is based on the geographical nature of scientific collaboration. In international collaborations, it is known that scientists tend to collaborate with colleagues nearby (Hoekman, Frenken, & Van Oort, 2009; Hoekman et al., 2010; Maggioni & Uberti, 2009; Scherngell & Barber, 2009). Even if present-day research collaboration is supported by advanced ICTs, frequent travel to have face-to-face interaction will remain necessary. Hence, most international research projects occur between neighboring countries. As a consequence, the networks in which ideas and competencies are being shared and developed, will be much tighter between scientists nearby than farther apart. This would mean that projects recombining resources from centers located farther apart can be expected to lead to more novel and unique outcomes than projects in which nearby colleagues collaborate (Boschma, 2005). A second reason why geographical distance can lead to more citation impact relates to the diffusion of results. Distant researchers have less overlap in their personal networks than researchers located in closer vicinity (Breschi & Lissoni, 2009). Hence, results are expected to diffuse more widely.

We test this hypothesis on all internationally co-authored papers in Europe published in 2000. To probe the effect of geographical distance on the citation impact of international research collaborations, this effect needs to be carefully isolated from other impact determinants. Hence, in the following, we take into account several control variables, including the number of authors and countries involved, whether collaborating countries share language and institutions, dummies for scientific disciplines, and an interdisciplinarity indicator. Most importantly, we also employ country dummies as each country has a different baseline expectation regarding citation impact. Since the mean citation rate of countries is known to differ systematically (May, 1997; Rousseau & Rousseau, 1998), collaborations between highly cited countries will automatically result in higher citation impact. The usage of country dummies that control for these national variations allows one to isolate the hypothesized effect of geographical distance on citation impact in a precise manner.

3. Methodology

We used Elsevier's Scopus database and selected all publications from 2000 which report affiliation addresses from at least two different European countries. As we are interested in European collaborations only, we left out single authored papers as well as papers reporting any non-European addresses in addition to the European ones. This procedure resulted in a total of 33,524 papers.

The dependent variable is the total number of citations a paper received before the end of 2009. This number ranges from 0 to 1503, with a mean of 23.331 citations. Given the skewed distribution of the dependent (integer) variable, Ordinary Least Squares regression is inappropriate. Instead, dealing with citations as count data, the options were Poisson regression and Negative Binomial regression (on this, see Frenken et al., 2005; Hausman et al., 1984; Maurseth & Verspagen, 2002). Given the extreme skewness of our dependent variable (i.e., over dispersion), the Negative Binomial regression technique is the most appropriate.

Our main independent variable concerns the distance between the two collaborating countries. Many indicate the distance between two countries by computing the kilometer distance between the two capital cities (Baldwin & Taglioni, 2006). However, in many instances, capital cities are not centrally located; hence, the distances between capitals provide an imprecise measure of the distance between countries. Instead, we follow Head and Mayer (2002) and Mayer and Zignago (2011) who measure the distance between two countries as the weighted average of the bilateral distances between the biggest cities of those two countries, where the weights reflect the share of each city of the pair in the overall population of the

respective country. In this way, the distance measure between two countries does not rely on the location of their capital cities alone, but on the distribution of population across both countries.[1]

In our dataset, some papers contain addresses from more than two European countries. This means that a single paper would contain multiple distances. In such cases, we treat the single paper as multiple observations corresponding to the number of bilateral collaborations involved. To each of these observations, a single value for the distance between the two respective countries can be assigned. Then, in the regression, we weigh these observations following the logic of fractional counting. For example, for a paper with addresses from four European countries one can compute the distance between six pairs of countries. We get six observations, which are weighted by 1/6 in the regression. As a robustness check, we also ran the same regression for the subset of papers that report addresses from only two countries, which avoids the need to weigh observations individually.

As the first control variable, we look at whether a shared language (language) affects citation. Even though an increasing share of European researchers speaks English fairly well, one can assume that in the practice of collaboration, they will resort to a shared native language if possible, as this would render collaboration easier and more productive. The dummy variable language takes on the value of one if there exists at least one official language that is spoken (at least partially) in both countries involved in the collaboration (CIA World Factbook, 2013).

We are also interested in whether co-membership in the EU affects citation impact. The corresponding dummy variable (eu27) indicates whether the two collaborating countries are both part of the European Union. Here, one may reason in two ways. On the one hand, the European Commission (EC) invests in pan-European research infrastructures and training, and provides ample subsidies for collaboration. This can facilitate collaboration between researchers and hence improve the quality of their work. On the other hand, one can expect a negative sign, since more resources are available for collaboration between EU members than for other pairs of countries. Hence, for EC-funded activities, researchers may be less selective in projects and/or foreign partners, which, on the average, leads to lower citation impact.

We also include the number of authors (nr_authors) as a control variable since each additional author is expected to add to the quality of the paper. At the same time, each additional author acts in due course as yet another source of diffusion. In either case, the effect of this variable is expected to be positive (Frenken et al., 2005).

The number of countries (nr_countries) is included to single out the effect of multi-country collaborations on citation impact, so that our distance variable is indeed measuring the effect of distance as such (Guerrero Bote, Olmeda-Gómez, & de Moya-Anegón, 2013).

Finally, we take into account information on disciplines. Since citation rates differ systematically across disciplines (Bornmann & Daniel, 2008), we include a dummy for each of the 27 disciplines that are distinguished in Scopus at the 2-digit level. We also take into account the number of disciplines in which the paper's journal is classified as an interdisciplinarity variable (interdisciplinarity). Here, one expects a negative sign as interdisciplinary research is generally less recognized (Heimeriks, 2013; Porter & Rafols, 2009; Wagner et al., 2011). Admittedly, this is only a rough indication of interdisciplinarity, since the Scopus classification of science into only 27 broad disciplines ignores interdisciplinarity within these disciplines.

We will further include as a control variable country dummies, since countries differ systematically in their citation impact (May, 1997; Rousseau & Rousseau, 1998). This difference can be attributed to at least two factors. First, if there exists a domestic bias in citation, papers from larger countries may receive, on average, more citations (Schubert & Braun, 1990; Pasterkamp et al., 2007). Second, countries differ in the amount and quality of resources at their disposal with national systems profiting from economies of scale (Nelson, 1993). Hence, even if the exact country-specific processes remain unknown, one may still expect systematic differences in citation impact across countries according to size.

4. Results

Table 1 shows the descriptive statistics, and Table 2 the correlation between independent variables. As it is clear, independent variables show low levels of correlation, except (not surprisingly) for the variables nr_authors and nr_countries. No problems of multi collinearity were encountered. As a robustness check, we performed also an un weighted regression analysis on the subset of papers which were co-authored by only two countries. The results of this additional analysis is reported as Model 2. Note that the variable denoting the number of countries had to be dropped for Model 2, since this takes on the value of 2 for all papers. The comparison of Model 2 with Model 1shows that the results tend to be robust, as estimated coefficients hardly change and significance levels remain stable.

Table 3 shows the regression results, including all variables. Distance is shown to have a positive and highly significant effect on citation. The estimated coefficient indicates that for each 1000 km, citations increase by 7% in Model 1, and 9% in Model 2. Hence, we can conclude that our guiding hypothesis can be confirmed: research partners profit from being distant, that is, the more distant two collaborators are in geographical space, the more citation impact their research will have, ceteris paribus

[1] We also repeated the regression analysis presented below using simply the distance between capitals as the distance measure between countries, and we found that signs and significance levels of all coefficients did not change. In this respect, results are thus robust. This additional regression analysis is available upon request.

Table 1
Descriptive statistics.

Variable	Mean	Std. Dev.	Min.	Max.
Citations 2009	23.331	47.178	0	1503
Distance	1251	778.033	161	4633
eu27	0.714		0	1
language	0.125		0	1
nr_countries	2.900	1.617	2	13
nr_authors	6.610	7.366	2	99
Interdisciplinarity	1.959	1.148	1	8

Table 2
Correlation matrix.

	Distance	eu27	Language	nr_countries	nr_authors	Interdisciplinarity
Distance	1					
eu27	−0.421	1				
language	−0.357	−0.153	1			
nr_countries	−0.003	0.051	−0.044	1		
nr_authors	0.018	0.002	−0.015	0.552	1	
interdisciplinarity	0.006	0.009	−0.003	−0.011	−0.041	1

Table 3
Regression results.

Variable	Model 1		Model 2	
	Coefficient	z	Coefficient	z
Distance (×1000)	0.070***		0.091***	
	(0.023)	3.08	(0.027)	3.44
eu27	−0.249***		−0.306***	
	(0.080)	−3.12	(0.095)	−3.23
Language	−0.042		−0.052	
	(0.037)	−1.12	(0.043)	−1.22
nr_countries	0.085***			
	(0.010)	8.42		
nr_authors	0.029***		0.042***	
	(0.003)	10.01	(0.004)	10.11
Interdisciplinarity	0.007		0.006	
	(0.011)	0.60	(0.012)	0.50
Discipline dummies	Yes		Yes	
Country dummies	Yes		Yes	
N	33,524		28,728	
Log pseudolikelihood	−129875.18		−110454.28	

Regression results. *, ** and *** refer to significance levels of 10%, 5% and 1%, respectively. Robust standard errors between brackets.

We further observe that the number of countries involved in the publication has a large effect on citation, as expected. The number of authors also has a positive and significant effect on citation impact, albeit much lower than the number of countries. A striking result holds that the effect of the eu27 dummy is strongly negative. Collaboration between countries both belonging to the EU have lower impact with a predicted ceteris paribus average loss around 25% in Model 1, and 30% in Model 2. This supports the interpretation that more resources render researchers less selective in their collaboration projects (similarly, one can understand the negative sign of the language variable, though not significant, as reflecting that researchers who share a language may be less selective in the joint projects they undertake). It should be reminded though, that the result on the effect of EU is obtained while controlling for country dummies. Hence, from the negative sign of the eu27 dummy, one cannot conclude that the mean citation rate of papers from EU countries is lower than from other countries.

Having a closer look at the coefficient estimates for the individual country dummies, we observe a certain pattern. In Table 4 we report these coefficients as sorted by the (fractional) number of papers with which the respective country is affiliated in the dataset (all country dummies were significant at the 5% level, except for the Serbia and the Armenia dummies). What becomes clear from the table is that countries producing more papers tend to have more citations. This may reflect a domestic citation bias enjoyed by larger countries, and/or the benefits stemming from economies of scale at the country level. However, some notable deviations from this pattern exist. For example, in terms of the sheer number of papers, Russia is a very productive country but its citation impact tends to be low. And reversely, Denmark's output is only a seventh of Germany's output, yet Denmark's citation impact is the highest among all countries.

Table 4
Country dummies.

	Model 1	Model 2		Model 1	Model 2
Germany	1.962	1.956	Turkey	1.148	1.059
UK	1.986	1.974	Bulgaria	1.393	1.312
France	1.913	1.894	Slovenia	1.690	1.656
Italy	1.796	1.759	Belarus	1.414	1.250
Russia	1.113	0.972	Croatia	1.083	0.927
Netherlands	1.996	1.974	Estonia	1.634	1.649
Spain	1.805	1.773	Lithuania	1.763	1.788
Switzerland	1.873	1.827	Latvia	1.464	1.355
Sweden	1.993	1.996	Iceland	1.691	1.534
Belgium	1.853	1.862	Armenia	0.669	0.459
Poland	1.405	1.363	Georgia	1.106	1.096
Denmark	2.018	1.995	Moldova	1.099	1.029
Austria	1.939	1.934	Cyprus	1.896	1.918
Finland	1.947	1.917	Azerbaijan	1.221	1.271
Czech Rep	1.645	1.630	Luxembourg	1.531	1.5'22
Hungary	1.514	1.461	Macedonia	1.057	1.070
Greece	1.624	1.588	Albania	1.137	1.198
Ukraine	1.027	0.909	Malta	1.451	1.037
Norway	1.623	1.569	Liechtenstein	1.006	0.654
Portugal	1.716	1.689	Bosnia H.	1.521	1.564
Ireland	1.792	1.748	Monaco	1.755	1.770
Romania	1.291	1.229	Serbia	−0.512	−0.555
Slovakia	1.382	1.344	Constant	−0.800	−0.638

5. Conclusion

Research collaboration through international co-authorship has been an increasingly important phenomenon in science. Growing international collaboration is not only the result of 'big science' but also part of the globalization process in scientific research (Glänzel & Schubert, 2005; Leydesdorff & Wagner, 2009; Waltman, Tijssen, & Eck, 2011). Our results suggest that the recombination of dissemination opportunities, skills, and resources from research centers located farther away, increases a paper's citation impact. In addition, distant researchers have less overlap in their personal networks than researchers located in closer vicinity, which may lead to a wider diffusion of their work. This effect is shown to exist while controlling for the country-specific and discipline-specific influences, as well as the number of countries and authors involved in the publication. Hence, on top of the already-known citation premium associated with international collaboration, there is an additional premium to geographical distance. Future research can extend our research design to cover all countries in the world, as to validate whether the finding for Europe is a general one.

Our finding can be taken to imply that researchers and national policymakers must take careful stock of the conduct of science at the global level (Wagner & Leydesdorff, 2005). As the system expands, useful innovations and dissemination opportunities can increasingly arise somewhere else, far from the comfort zone nearby; identifying these opportunities will be an increasingly important challenge for researchers and policymakers. At the same time, the collaboration costs for long distance collaborations can be expected to be high. Hence, science policies could on the one hand promote collaborations over longer distance, while at the same time removing barriers to engage in such collaborations.

References

Baldwin, R., & Taglioni, D. (2006). *Gravity for dummies and dummies for gravity equations*. CEPR Discussion Papers 5850.

Bornmann, L., & Daniel, H.-D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, *64*, 45–80.

Boschma, R. A. (2005). Proximity and innovation. A critical assessment. *Regional Studies*, *39*, 61–74.

Breschi, S., & Lissoni, F. (2009). Mobility of skilled workers and co-invention networks: An anatomy of localized knowledge flows. *Journal of Economic Geography*, *9*, 439–468.

CIA World Factbook. (2013). http://www.cia.gov/library/publications/the-world-factbook. Accessed 21.05.13.

Doré, J.-C., Ojasoo, T., & Okubo, Y. (1996). Correspondence factorial analysis of the publication patterns of 48 countries over the period 1981–1992. *Journal of the American Society for Information Science and Technology*, *47*, 588–602.

Fleming, L. (2001). Recombinant uncertainty in technological search. *Management Science*, *47*, 117–132.

Frenken, K., Hardeman, S., & Hoekman, J. (2009). Spatial scientometrics: Towards a cumulative research program. *Journal of Informetrics*, *3*, 222–232.

Frenken, K., Hölzl, W., & De Vor, F. (2005). The citation impact of research collaborations: The case of European biotechnology and applied microbiology (1988–2002). *Journal of Engineering Management and Technology*, *22*, 9–30.

Frenken, K., Ponds, R., & Van Oort, F. G. (2010). The citation impact of research collaboration in science-based industries: A spatial-institutional analysis. *Papers in Regional Science*, *89*, 351–371.

Georghiou, L. (1998). Global cooperation in research. *Research Policy*, *27*, 611–626.

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

Glänzel, W., & Schubert, A. (2005). Analysing scientific networks through co-authorship. In *Handbook of Quantitative Science and Technology Research*. Netherlands: Springer.

Guerrero Bote, V. P., Olmeda-Gómez, C., & de Moya-Anegón, F. (2013). Quantifying the benefits of international scientific collaboration. *Journal of the American Society for Information Science and Technology*, *64*, 392–404.

Hausman, J., Hall, B. H., & Griliches, Z. (1984). Econometric models for count data with an application to the patents – R&D relationship. *Econometrica*, *52*, 909–938.
Head, K., & Mayer, T. (2002). *Illusory border effects: Distance mismeasurement inflates estimates of home bias in trade, CEPII Working Paper 2002-01, Paris*.
Heimeriks, G. (2013). Interdisciplinarity in biotechnology genomics and nanotechnology. *Science and Public Policy*, *40*, 97–112.
Hoekman, J., Frenken, K., & Tijssen, R. J. W. (2010). Research collaboration at a distance: Changing spatial patterns of research collaboration in Europe. *Research Policy*, *39*, 662–673.
Hoekman, J., Frenken, K., & Van Oort, F. (2009). The geography of collaborative knowledge production in Europe. *The Annals of Regional Science*, *43*, 721–738.
Leydesdorff, L., & Wagner, C. (2009). Is the United States losing ground in science? A global perspective on the world science system. *Scientometrics*, *78*, 23–36.
Maggioni, M. A., & Uberti, T. E. (2009). Knowledge networks across Europe: Which distance matters? *The Annals of Regional Science*, *43*, 691–720.
Maurseth, P. B., & Verspagen, B. (2002). Knowledge spillovers in Europe: A patent citations analysis. *Scandinavian Journal of Economics*, *104*, 531–545.
May, R. M. (1997). The scientific wealth of nations. *Science*, *275*(5301), 793–796.
Mayer, T., & Zignago, S. (2011). *Notes on CEPII's distances measures: The GeoDist database. CEPII Working Paper 2011-25, Paris*.
Narin, F., Stevens, K., & Whitlow, E. (1991). Scientific cooperation in Europe and the citation of multidomestically authored papers. *Scientometrics*, *21*, 313–323.
Nelson, R. R. (1993). *National innovation systems. A comparative analysis*. New York/London: Oxford University Press.
NSB. (2002). *Science and engineering indicators 2002*. Washington, DC: National Science Board., http://www.nsf.gov/statistics/seind02/.
Pasterkamp, G., Rotmans, J., Kleijn, D., & Borst, C. (2007). Citation frequency: A biased measure of research impact significantly influenced by the geographical origin of research articles. *Scientometrics*, *70*, 153–165.
Persson, O., Glänzel, W., & Danell, R. (2004). Inflationary bibliometrics values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics*, *60*, 421–432.
Porter, A. L., & Rafols, I. (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics*, *81*, 719–745.
Rousseau, S., & Rousseau, R. (1998). The scientific wealth of European nations: Taking effectiveness into account. *Scientometrics*, *42*, 75–87.
Scherngell, T., & Barber, M. (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.
Schubert, A., & Braun, T. (1990). World flash on basic research: international collaboration in the Sciences, 1981–1985. *Scientometrics*, *19*, 3–10.
Wagner, C. S., & Leydesdorff, L. (2005). Mapping the network of global science: Comparing international co-authorships from 1990 to 2000. *International Journal of Technology and Globalisation*, *1*, 185–208.
Wagner, C., Roessner, J., Bobb, K., Klein, J., Boyack, K., Keyton, J., et al. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics*, *165*, 14–26.
Waltman, L., Tijssen, R. J. W., & Eck, N. V. J. (2011). Globalisation of science in kilometres. *Journal of Informetrics*, *5*, 574–582.