NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Stanislav Avdeev

INTERNATIONAL COLLABORATION IN HIGHER EDUCATION RESEARCH: A GRAVITY MODEL APPROACH

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: EDUCATION
WP BRP 54/EDU/2019

International collaboration in higher education research: A gravity model approach *

Stanislav Avdeev †

Higher School of Economics

April 20, 2021

Abstract

Although geographical distance has become less relevant in co–authorship for monodisciplinary fields such as economics, mathematics, and physics, little is known about international collaboration in multidisciplinary fields such as higher education. This paper studies collaboration patterns in higher education research using the Scopus database with the application of the gravity model. The results show that the intensity of collaboration is negatively associated with geographical distance and positively associated with linguistic commonality but these findings differ significantly between various world regions. European scholars appear to give preference to linguistically proximate partners over geographical neighbours. Although English is the lingua franca in science, language is not a significant factor for the formation of collaboration for North American and Asian researchers. These findings have policy implications for fostering multidisciplinary research in international partnerships.

Keywords: economics of science, gravity model, higher education, international collaboration, spatial econometrics

JEL Codes: C21, I23, O30

*I gratefully acknowledge valuable comments and suggestions from Bernardo Pincheira, Dmitrii Shchetinin, Maria Yudkevich, colleagues at the Center for Institutional Studies, conference and workshop participants in Amsterdam, Jyväskylä, Kassel, Moscow, and Saint Petersburg, and two anonymous referees. Any errors are those of the author. The paper was prepared within the framework of the Higher School of Economics Basic Research Programme and funded by the Russian Academic Excellence Project "5–100". This is a preprint of an article published in the Scientometrics. The final version is available online: https://doi.org/10.1007/s11192-021-04008-8

†Center for Institutional Studies, Higher School of Economics, Moscow, Russia; e–mail: stnavdeev@gmail.com. ORCID: https://orcid.org/0000-0002-3130-1418

1 Introduction

Geographical distance has become less relevant while linguistic ties are still a strong factor in collaboration between countries (Adams, 2013; Wagner et al., 2017). Countries located in close physical proximity profit from each other through knowledge spillovers (Almeida et al., 2009). When authors speak the same language, they are more likely to collaborate. However, these factors are found to be relevant for monodisciplinary research fields such as economics, mathematics, and physics (Newman, 2004). Increased multidisciplinary research requires the involvement of researchers from different disciplinary backgrounds with various patterns of collaboration (Leahey, 2016; Gates et al., 2019). An example of a multidisciplinary field is higher education which is defined by its object of research: scholars from different fields come to research higher education with their methods and theories (Tight, 2004; Altbach, 2014; Yokoyama, 2016). The field of higher education is influenced by the norms and traditions of other fields, particularly economics, psychology, and sociology. Nevertheless, higher education researchers remain unreflective about the structure of their collaboration (Yokoyama, 2016).

Waltman et al. (2011) found that social scientists are less likely to collaborate with people from different countries. There are substantial differences in collaboration patterns of higher education researchers who come from different fields (usually social sciences). Scholars' choices of international collaboration partners are influenced by the norms, habits, and routines of specific disciplines (Henriksen, 2016). Economists are more likely to cooperate with international scholars than sociologists and psychologists (Rosenblat and Mobius, 2004; Leahey, 2016; Kliegl and Bates, 2011; Kuld and O'Hagan, 2018). Almeida et al. (2009) found that countries located in close geographical proximity to each other show similar specialisation patterns: policy–oriented papers involve more researchers from different countries, whereas individual scholars focus more on student–specific topics (Altbach, 2014).

Collaboration patterns are affected not only by discipline but by country of origin, e.g. European, American, and Asian scholars have different collaboration patterns. For instance, European researchers benefit from the EU and the variety of funding agencies where research programs often require cross–country teams (Leydesdorff and Wagner, 2008; Hoekman et al., 2009; Wagner et al., 2015). Asian researchers prefer to choose scholars from other Asian countries due to the similar-

ities of higher education systems and academic traditions. Likewise, if scholars in Asia are cited by authors within the region, they do not need recognition from European and North American scholars (Glänzel, 2001). Geographical partners might be preferred not only due to their proximity but due to the similarities and benefits of collaboration with neighbouring partners.

This paper studies collaboration patterns in the multidisciplinary field of higher education across different world regions. The question of international collaboration is critical in science since the national level of organisation is challenged by researchers focusing on the globalisation process (Frenken et al., 2009). To measure the research collaboration of different countries, the most robust approach is to use spatial econometrics techniques, analysing the institutional affiliations contained in the Scopus bibliometric data (Frenken et al., 2009). This paper contributes to the current discussion of the importance of international collaboration in science, paying special attention to the growing public interest in multidisciplinary research.

2 Literature Review

2.1 International collaboration and the role of geographical proximity and linguistic commonality

Katz and Martin (1997) noted that co–authorship of scientific papers is only one of the possible outcomes of research collaboration, and it is only one of the forms in which collaboration can be expressed. Nevertheless, co–authorship has been adopted as a proxy for measuring research collaboration, i.e. this is a classical outcome of research activities. Therefore, in this study international research collaboration is a co–authorship relation between two or more countries.

Many studies have shown the inexorable growth of international collaboration across different fields (Adams, 2013; Wagner et al., 2015; Marginson, 2020). Wagner et al. (2015) found that the proportion of internationally co–authored papers rose from 10.1% in 1990 to 19.5% in 2000 and 24.6% in 2011. After the year 2000, the worldwide collaboration rate rose in all disciplines, including hard sciences and social sciences (Wagner and Leydesdorff, 2005; Henriksen, 2016; Wagner et al., 2017). Adams (2013) found that the rise in the total annual output for each country is due to international collaboration. Almost all countries have similar patterns in

the growing proportion of articles that have international co–authors. However, the world average figures of international collaboration do not explain patterns of variation in international co–authorship on a smaller scale: by regions and disciplines. This study focuses not only on the world scale but distinguishes regional patterns of international collaboration in the multidisciplinary field of higher education.

To find out how co–authors work together, previous papers explored some characteristics of such collaboration. For instance, geographical proximity and linguistic commonality play an important role in fostering high intensity of collaboration. Researchers are biased towards international collaboration with partners who are proximate with respect to geography and language (Hoekman et al., 2010). Waltman et al. (2011) revealed that research has globalised in recent decades: the average collaboration distance per publication has increased from 334 km in 1980 to 1553 km in 2009. Increasingly, funding agencies and public policies have encouraged collaboration by prioritising research in partnership. The establishment of the European Research Area has stimulated international collaboration (Leydesdorff and Wagner, 2008; Hoekman et al., 2009; Wagner et al., 2015). Hoekman et al. (2010), using spatial econometrics, found a diminishing effect of geographical proximity on co–publishing, with territorial borders becoming less relevant. Overall, the research design of spatial econometrics papers is more elaborated than previous descriptive papers on international collaboration.

When authors speak the same language, they are more likely to collaborate with each other rather than with researchers with different characteristics (Hoekman et al., 2009). Adams (2012) found that Nigeria collaborates not with its neighbours in West Africa but with co–linguists in East Africa. Hoekman et al. (2010) showed that co–publication intensity is higher within regional, national, and linguistic areas after controlling for the size of regions and their research specialisation profiles. Although researchers speaking a common language are not necessarily located within a single geographical area, I expect co–authors, who speak a shared language, to have a higher intensity of international collaboration.

Geographical clusters tend to play an important role in the intensity of international collaboration regardless of distance and linguistic proximity. The rapid growth of regional links in Asia and established regions, such as Europe and North America, reveals the importance of collaboration within world regions (Glänzel, 2001). Leydesdorff and Wagner (2008) revealed that the number of internationally

co–authored publications has grown linearly while the number of addresses of internationally collaborating authors grew exponentially, suggesting that the growth of networks extends to many more places around the globe. Adams (2012) showed growing research networks of countries in Asia, Latin America, and the Middle East. A dramatic growth of international collaboration between China with neighbouring countries is one example of regional development (Jung and Horta, 2013). Therefore, in spite of language differences, geographically proximate countries within a certain region are preferred. Despite the vast body of empirical evidence, the association of geographical and linguistic proximity with the intensity of international collaboration in a multidisciplinary field has been understudied.

2.2 International collaboration in multidisciplinary fields

To explore international collaboration in a multidisciplinary field, there should be a clear definition of mono– and multi–disciplinary fields. Monodisciplinary research is concerned with the study of a research topic within a *single* discipline, and with *its own methods*. Multidisciplinary research is concerned with the study of a research topic across *multiple* disciplines, and with the *transfer of methods* from one discipline to another.

Different disciplines operate under different norms and paradigms, thus, I exclude studies of international collaboration in hard sciences from the analysis due to their distinctive collaboration patterns (Henriksen, 2016). To the best of my knowledge, international collaboration in a multidisciplinary field have never been investigated with the application of the spatial econometrics.

Several studies have investigated patterns of international collaboration in social sciences: Rosenblat and Mobius (2004); Kuld and O'Hagan (2018) in economics; Leahey (2016) in sociology; Kliegl and Bates (2011) in psychology. Previous papers on international collaboration in social sciences focused primarily on a descriptive analysis of the co–occurrence of countries in joint papers. Few studies have been published about the geographical barriers of social scientists (Waltman et al., 2011). Different barriers exist when researchers in social disciplines having differing norms and traditions are collaborating. For instance, Waltman et al. (2011) explored the differences in average geographical distance per publication among the fields of science: economics and business (1939 km) and psychology (1478 km) are more

globalised fields than sociology and anthropology (1063 km) and educational sciences (969 km).

Some studies focused on the role of international collaboration in scientific production in multidisciplinary fields. Bartneck and Hu (2010) found that there are no significant differences between domestic and international collaboration in terms of citations in the Computer–Human Interaction community. They found that North America and Europe collaborate most within the Computer—Human Interaction community, particularly, the USA, the UK, and Canada co–author many papers. Wang et al. (2015) found that sport science researchers showed a strong tendency to collaborate, especially among European countries. They highlighted that the share of international collaboration in Asian countries is below 40%, and the growth rate is lower than that of these countries' overall output, while the trend is reversed in many western countries: the share is above 50% and the growth rate is higher. Correia et al. (2018), analysing the Computer Supported Cooperative Work community, demonstrated that distance is no longer a barrier it was in the past, despite the heterogeneity between some regions in their propensity to collaborate. Therefore, it is not straightforward what the collaboration patterns are between scholars from various disciplines, as in higher education research, as they could be less or more inclined to collaborate than two researchers in monodisciplinary fields.

2.3 Higher education as a multidisciplinary field

Tight (2004) defined higher education as an "interdisciplinary field of research in which multiple communities of practice operate". The field of higher education research is a fragmented community of researchers from various disciplines with different types of degrees, theoretical approaches, and methodologies. It is even not a sub–field of educational studies (Yokoyama, 2016). Lovakov and Yudkevich (forthcoming) covered the disciplinary foundations and roots of higher education research and revealed the different influences of economics, psychology, and sociology on the field. Psychology has the highest citation rate by higher education researchers, followed by sociology and economics.

Altbach (2014) said "higher education is not a scholarly or scientific discipline; it has no central and accepted methodology nor does it have a set of concerns for research and study. Rather, it is a field that uses the disciplinary insights of other

fields, mainly in the social sciences, to inform research themes that often require interdisciplinary insights". Yokoyama (2016) defined higher education as a "multi–disciplinary and loosely coupled community [which] suggest diversity in the field and in its identity rather than coherence and consolidation". Internationalisation in the form of increased student and academic mobility has broadened the scope of higher education, particularly for scholars of the economics of education, psychology, public administration, and the sociology of education (Altbach, 2014). The multidisciplinary field of higher education has been formed by interaction with other fields and disciplines, therefore, is influenced by them and their collaboration patterns.

Higher education studies differ not only because of the several disciplines that comprise it but because of its substantial geographical variation. The American higher education community is more practice–oriented, engaging in meso– and micro–level research, whereas European scholars are more policy–focused, emphasising analysis at the macro level (Yokoyama, 2016). Although very few studies have explored international collaboration in higher education research, several studies have analysed different subfields of higher education research (Kosmützky and Krücken, 2014; Kuzhabekova et al., 2015). Kosmützky and Krücken (2014) focused specifically on comparative research and found a much higher share of international collaboration in comparative research compared to non–comparative. They showed that 46% of comparative papers are internationally co–authored publications compared to about 20% of non–comparative papers. Kuzhabekova et al. (2015) found that just 11.3% of papers are authored by researchers coming from at least two countries. International comparative research appears to be the most globalised topic in the field of higher education and many international publications have been co–authored by scholars from different countries.

Overall, there is sparse literature that focuses on international collaboration in multidisciplinary fields. I take a different approach with the application of spatial econometrics and study geographical and linguistic proximity, which are related to the formation of international collaboration in higher education. I hypothesise that geographical and linguistic proximity still play an important role for international research collaboration because of the language the authors speak and the region in which they operate. By focusing on international collaboration in a multidisciplinary field of research, this study contributes to the broader literature that explores geographical proximity and linguistic commonality across various fields.

3 Methodology

3.1 Data

In this section, I describe the data and the main variables I use in the analysis, paying particular attention to the construction of the proximity measures. I take into account only scientific papers without accounting for books and other possible ways of cooperating and publishing research as they are under–represented and, therefore, further analysis would be inhibited by a lack of data. The analysis is based on publication types "article" and "review" from Scopus. Scopus has a wider coverage of papers in social sciences than other databases, so it seems meaningful to exploit this database (Mongeon, 2015).

I examine articles and reviews published in 24 journals considered key in the field of higher education found by Lovakov et al. (forthcoming). Data were retrieved from the Scopus database in January 2019. The overall data set covers 17,413 publications from the period 1978—2017 excluding papers with no affiliation information.

First, I choose papers that are internationally co–authored: papers that have been published with the cooperation of at least two different countries. International co–publications, internationally co–authored papers, and international collaboration will be used as synonyms. The determination of the country of origin of the authors is based on their institutional affiliation. In sum, there are 1,414 internationally co–authored papers.

Second, the total number of authors in each country is counted. If an author has multiple affiliations in different countries, the paper is considered as an international paper and this author is counted multiple times: scholars with multiple affiliations gain access to additional research resources or networks and form stronger ties between institutions in different countries, therefore, these authors and papers are included and analysed.

Third, all author affiliations of the selected publications are reduced to a country, giving 96 countries. For each country I find the capital and its latitude and longitude. Since I am interested in international collaboration between two countries rather than cities, I take into account only one geographical point per country (Frenken et al., 2009; Wagner et al., 2017). Then, for each pair of countries, the number of times it occurs in the selected publications as an international co–authorship is

counted. For all capitals, coordinates are obtained using the R package *geosphere*. Next, distance is calculated for all pairs as the Euclidean distance in kilometers between the capitals of two countries. For each country the official or de facto official language is identified using R package *lingtypology*. If a country has several languages, all of them are counted.

3.2 Empirical model

I follow the spatial econometrics framework of Frenken et al. (2009) in exploring the geographical patterns of international collaboration in higher education research. I start estimating a base model which is modelled by analogy to the Newton's law of universal gravitation (Tinbergen, 1962; Frenken et al., 2009). The gravity model states that the gravitational force between two entities is dependent on their masses and the distance between them. Collaboration frequency between two countries is assumed to be dependent on the number of authors and the distance between them:

$$I_{ij} = \beta_0 \frac{MASS_i^{\beta_1} MASS_j^{\beta_2}}{DISTANCE_{ij}^{-\beta_3}} \tag{1}$$

Empirically, taking natural logarithms on both sides of the gravity model and adding a random error term, Model 1 can be converted into a testable Model 2:

$$\ln I_{ijt} = \beta_0 + \beta_1 \ln MASS_{it} + \beta_2 \ln MASS_{jt} + \beta_3 \ln DISTANCE_{ij} + \varepsilon_{ijt} \tag{2}$$

It is important to take into consideration the total number of authors in a country, because collaboration intensity is highly dependent on size. To take into account linguistic proximity, I use a dummy variable. The results of the analysis are obtained from Model 3:

$$\ln I_{ijt} = \beta_0 + \beta_1 \ln MASS_{it} + \beta_2 \ln MASS_{jt} + \beta_3 \ln DISTANCE_{ij} + \beta_4 LANG_{ij} + \varepsilon_{ijt} \tag{3}$$

I_{ijt} is the total number of co–authored papers between country i and country j in year t; $MASS_{it}$, $MASS_{jt}$ is the number of authors in country i and country j in year t; $DISTANCE_{ij}$ is the Euclidean distance between capitals of country i and country j; $LANG_{ij}$ is a dummy variable equal to 1 if country i and country j have

a common official or de facto official language/languages; ε_{ijt} is the error term.

The dependent variable is the number of internationally co–authored papers between two countries, i.e. collaboration frequency. The standard model for count data (non–negative integers) is the Poisson regression model. A Poisson distribution assumes that the mean and variance are equal. The data show that the mean is less than the variance, suggesting some overdispersion. This occurs when for a random variable $Y \sim Pois(\lambda)$:

$$\mathbb{E}(Y) < Var(Y) \tag{4}$$

There may be an issue of unobserved heterogeneity in the data. Unobserved heterogeneity leads to overdispersion. The standard parametric model to account for overdispersion is a negative binomial. The negative binomial model assumes a particular form of dependence for the underlying stochastic process, with the occurrence of an event increasing the probability of further occurrences.

Zero event counts are often observed in the data, leading to a skewed distribution. A zero event count is a situation when two countries do not have any papers for a given year, which leads to inconsistency with the Poisson model. A zero–inflated count model provides another way to model excess zeros, therefore, I include this model with the negative binomial model.

In sum, I conduct the analysis using the negative binomial and zero–inflated count models to partially solve the problem of the spatial gravity model. The approach that I use provides measurements of the geographical proximity and linguistic commonality in higher education research across various regional groupings. First, I present descriptive statistics that include all regional groupings of countries: Africa, Asia, Europe, North America, Oceania, and South America. Each co–authored paper is included to aforementioned groups if one of the authors is from this region. Second, I analyse a model for all 96 countries. Third, I estimate several models that include only three top–publishing regions: Asia, Europe, and North America.

3.3 Limitations

The results of the paper should be interpreted with some caution. First, Scopus has a bias towards publications in English, suggesting there is an over-representation of papers in English. Second, the method of identifying the current location of authors is not foolproof. For example, a scholar might work temporarily in a foreign

university and choose to list their home institute. I would consider such collaboration as an international one. Third, this paper includes international co-publications between countries and not regions/cities, thus, some information about geographical proximity might be lost. Previous studies have found that the impact of these limitations is relatively small, suggesting they should not significantly bias the results of the paper (Frenken et al., 2009; Waltman et al., 2011; Wagner et al., 2017).

4 Results

4.1 Descriptive statistics

The number of internationally co–authored publications has grown exponentially, as demonstrated in Figure 1 and Figure 2, which indicate the increased interest in higher education research. The average share of international collaboration is 8.1% and it has been rising annually by 5.5% before 2000, but by 13.8% after 2000. In the following analysis, I restrict the sample to 2000–2017 since there are few publications before 2000. There are 1,262 publications from 2000 to 2017.

Figure 1: The share of international papers in 1978–2017

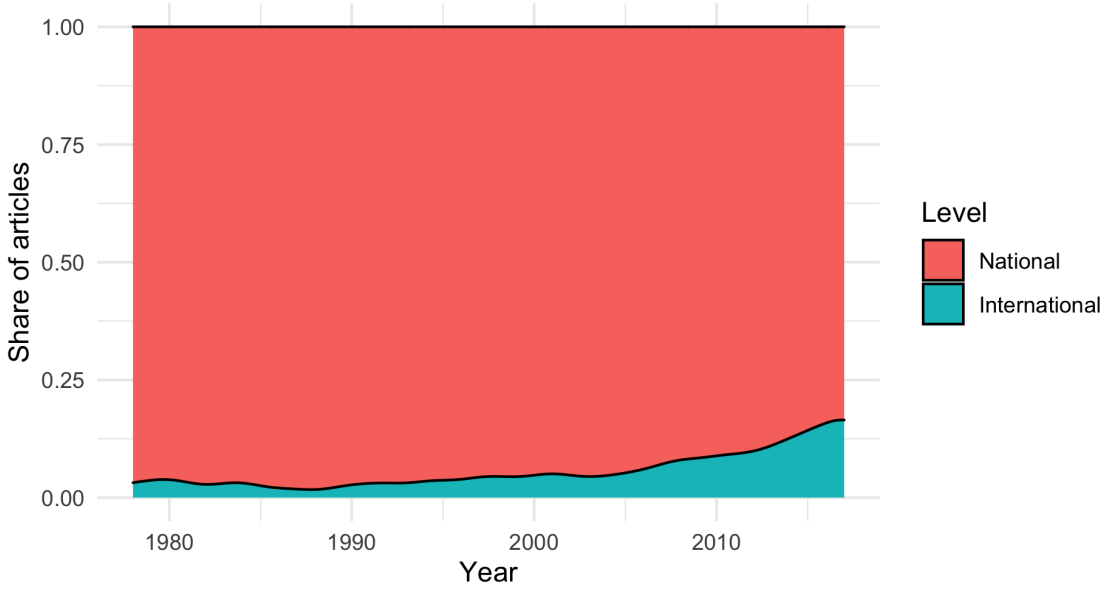


Figure 2: The number of international papers in 1978–2017

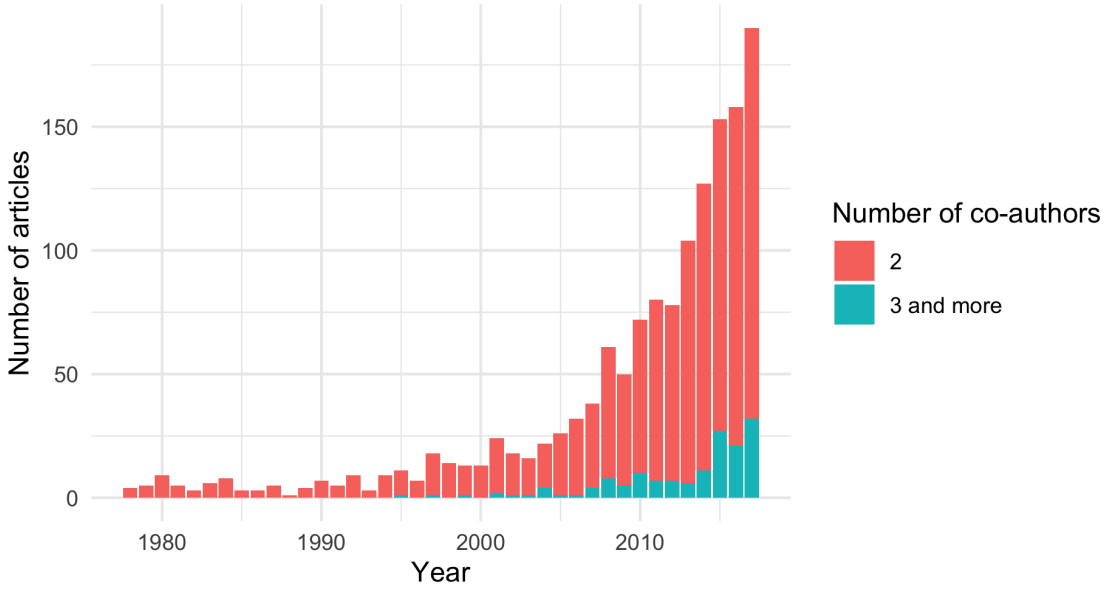


Table 1 presents the total national publication output and the share of international papers by a country for the observed period. The most active 40 countries are ranked in descending order by their total number and the share of international papers. Most publications are from English–speaking countries: the UK, the US, Australia, Canada, and New Zealand. Some European and Asian countries are also among the top publishing countries: China, the Netherlands, Germany, Norway, and Hong Kong.

The share of international publications in a country's output can be found in Table 2. The most internationally–oriented countries are Vietnam (more than 70% of international papers); Brazil, Romania, United Arab Emirates, Cyprus, China (more than 50%); South Korea, Switzerland, Czech Republic, Austria, Belgium, and Malaysia (more than 40%). There is at least one country in each region that accumulates the largest amount of international collaboration: Europe – the UK and the Netherlands; North America – the US and Canada; Asia – China and Hong Kong; Oceania – Australia and New Zealand; Africa – South Africa; South America – Brazil.

These two tables demonstrate that the larger a country's research effort, the smaller the proportion of its international co–authorship: scholars in smaller

Table 1: The number and the share of international papers in 2000–2017

N	Country	N of int. papers	Share (total)	N	Country	N of int. papers	Share (total)
1	the UK	391	0.31	21	Japan	30	0.02
2	the US	390	0.31	22	Singapore	29	0.02
3	Australia	276	0.22	23	Taiwan	28	0.02
4	Canada	142	0.11	24	Switzerland	26	0.02
5	China	113	0.09	25	France	25	0.02
6	the Netherlands	110	0.09	26	Ireland	23	0.02
7	Germany	90	0.07	27	UAE	21	0.02
8	New Zealand	78	0.06	28	Turkey	21	0.02
9	Norway	74	0.06	29	Brazil	19	0.02
10	Hong Kong	70	0.06	30	Austria	19	0.02
11	Sweden	60	0.05	31	Greece	16	0.01
12	Spain	50	0.04	32	Czech Republic	12	0.01
13	Portugal	49	0.04	33	Israel	12	0.01
14	Finland	49	0.04	34	Mexico	11	0.01
15	South Africa	48	0.04	35	Vietnam	10	0.01
16	Belgium	45	0.04	36	Romania	10	0.01
17	South Korea	39	0.03	37	Cyprus	10	0.01
18	Malaysia	37	0.03	38	Chile	10	0.01
19	Denmark	34	0.03	39	Russia	10	0.01
20	Italy	31	0.02	40	India	9	0.01

Only countries with 9 papers and more are shown in the table.

(resource–poor) countries are forced to look outside for co–authors, while researchers in the larger countries tend to collaborate more frequently with domestic partners.

The countries of Africa, Asia, Europe, North America, Oceania, and South America are brought together for further analysis. Figure 3 examines trends in international collaboration by a region of the world. If a country has a joint paper with a country that is not in the country's region, it counts as an internationally co–authored paper. Europe is the dominant region for collaboration, reflecting the total output of European countries (529 papers). The countries of North America (417) are second, followed by countries in Asia (361), Oceania (294), Africa (75), and South America (44). Figure 3 reveals the differences in growth by regions: collaboration with South America and Asia increased most rapidly, followed by Africa, Oceania, Europe, and North America.

Table 2: The share of international papers by country in 2000–2017

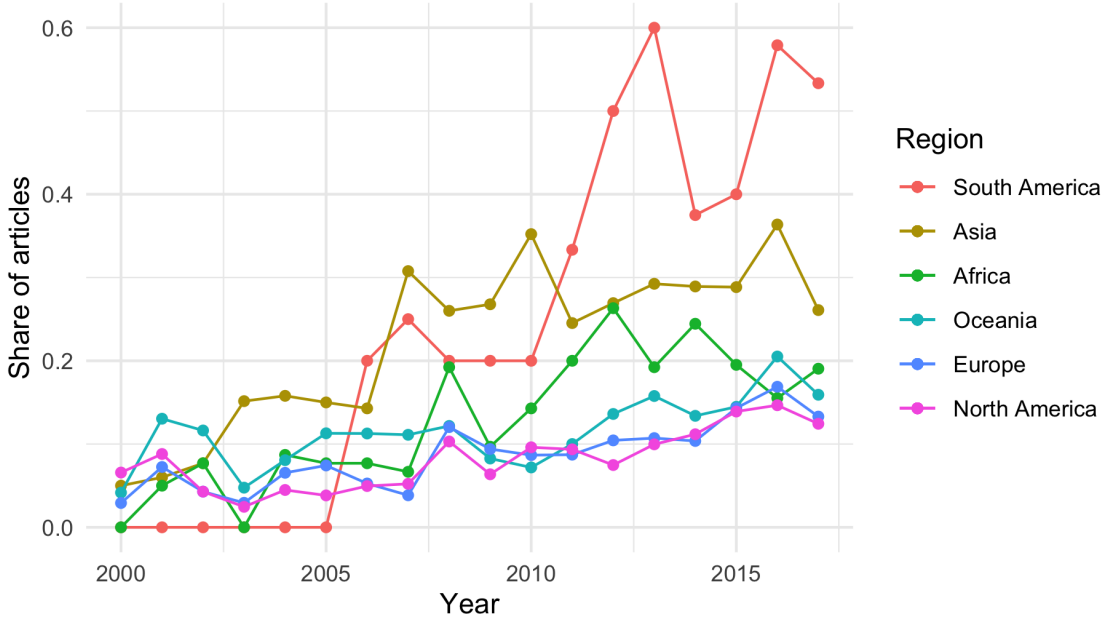
N	Country	Share	N	Country	Share
1	Vietnam	0.71	21	the Netherlands	0.33
2	Brazil	0.56	22	Russia	0.31
3	Romania	0.56	23	Turkey	0.30
4	UAE	0.54	24	Hong Kong	0.30
5	Cyprus	0.53	25	Japan	0.29
6	China	0.53	26	Portugal	0.28
7	South Korea	0.47	27	Canada	0.28
8	Switzerland	0.46	28	Taiwan	0.26
9	Czech Republic	0.44	29	Sweden	0.26
10	Austria	0.44	30	Italy	0.25
11	Belgium	0.41	31	Spain	0.24
12	Malaysia	0.40	32	India	0.24
13	Greece	0.39	33	New Zealand	0.23
14	Singapore	0.39	34	Finland	0.21
15	France	0.38	35	Ireland	0.21
16	Denmark	0.38	36	Australia	0.14
17	Germany	0.38	37	South Africa	0.14
18	Chile	0.37	38	the UK	0.14
19	Norway	0.36	39	Israel	0.12
20	Mexico	0.35	40	the US	0.09

Only countries with 9 papers and more are shown in the table.

European countries extended and collaborated more with researchers from non–European countries. The proportion of the collaboration within European countries is decreasing, while the proportion of the collaboration of European countries with co–authors from non–European countries is increasing. This illustrates that European countries have become more open towards non–European countries in international scientific collaboration.

Table 8 in the appendix shows the joint papers of the top 25 countries. There are no clear geographical preferences between countries: there are close geographical partners such as the pairs of Canada–US and Australia–New Zealand, and there are distant pairs of countries such as Australia–the UK, the UK–the US, and Australia–the US. These are the most productive countries, suggesting that the English language is the current lingua franca of science.

Figure 3: The share of international papers by region



4.2 Regression analysis

The summary statistics for the variables included in the model are presented in Table 3. The regression analysis includes international collaboration only between scholars from three top–publishing regions: Asia, Europe, North America. The average number of co–authored papers between countries is 1.46 per year. The average distance per paper is 6,472 km. About 22% of countries share the official or de facto official language/languages.

Table 3: The descriptive statistics for variables

Variables	N	Mean	St. Dev.	Min	Max
N of co–authored papers	1,157	1.46	1.30	1	21
N of authors in country i	1,157	30.92	50.79	1	233
N of authors in country j	1,157	99.08	111.56	1	336
Distance, km	1,157	6,472	5,143	60	19,576
Common language, 1 = Yes	1,157	0.22	0.41	0	1

The regression results for all countries are in Table 4. I start estimating the basic models without a dummy variable: Columns (1) and (2) – negative binomial,

columns (3) and (4) – zero–inflated count data. The coefficients for the number of authors in countries are 0.605 and 0.486 for the negative binomial model, and 0.491 and 0.434 for the zero–inflated count data model, statistically significant at the 0.01 level (see columns (1) and (3)). This shows that the higher the total number of scholars in a country, the higher the intensity of international collaboration.

Geographical proximity shows a negative statistically significant coefficient at the 0.01 level (see columns (1) and (3)). The coefficients are –0.262 and –0.208 for the negative binomial and zero–inflated count data models, respectively. This indicates a lower intensity of collaboration over longer distances.

Columns (2) and (4) include an additional regressor: common language. The coefficients are 0.325 and 0.364 and statistically significant at the 0.05 and 0.01 levels, respectively. This indicates that collaboration between countries that share a common language occur more often than collaboration between other countries in the world.

Table 4: The results for international collaboration between all countries

	N of co–authored papers (log)			
	negative binomial		zero–inflated count data	
	(1)	(2)	(3)	(4)
N of authors in country i (log)	0.605***	0.541***	0.491***	0.416***
	(0.050)	(0.057)	(0.051)	(0.055)
N of authors in country j (log)	0.486***	0.463***	0.434***	0.413***
	(0.046)	(0.047)	(0.042)	(0.040)
Distance, km (log)	−0.262***	−0.262***	−0.208***	−0.205***
	(0.053)	(0.053)	(0.050)	(0.042)
Common language, 1 = Yes		0.325**		0.364***
		(0.149)		(0.131)
Observations	1,157	1,157	1,157	1,157
Log Likelihood	−527.147	−524.780	−584.310	−580.905

Robust standard errors in parentheses.
*p<0.1; **p<0.05; ***p<0.01

Table 9 and Table 10 in the appendix show the regression results for all countries in two separated periods: 2000–2008 and 2009–2017. In line with previous results, I observe similar coefficients for the number of authors in countries. The coefficients for geographical distance and linguistic commonality show a slight decline over time,

suggesting that the intensity of international collaboration is less dependent on geography and language in the period 2009–2017 than in the period 2000–2008. All coefficients are statistically significant.

The regression results for European countries are presented in Table 5. I observe similar collaborative trends among European countries compared to the world. The total number of authors in a country is positive and statistically significant in both models (see columns (1) and (3)). The geographical patterns of collaboration for European countries are similar to the world average, suggesting that European collaboration do not occur more often over longer distances than the collaboration of other countries.

The coefficient for linguistic commonality is positive and statistically significant (see columns (2) and (4)). The coefficient for a common language is higher for European countries compared to the world which shows the greater importance of linguistic commonality for European scholars and their co–authors. European researchers collaborate significantly more often with co–authors who speak the same language.

Table 5: The results for international collaboration for European researchers

	N of co–authored papers (log)			
	negative binomial		zero–inflated count data	
	(1)	(2)	(3)	(4)
N of authors in country i (log)	0.636***	0.558***	0.645***	0.435***
	(0.069)	(0.075)	(0.066)	(0.151)
N of authors in country j (log)	0.432***	0.403***	0.245***	0.680***
	(0.059)	(0.060)	(0.068)	(0.117)
Distance, km (log)	−0.233***	−0.248***	−0.177***	−0.317***
	(0.069)	(0.069)	(0.066)	(0.109)
Common language, 1 = Yes		0.445**		1.031***
		(0.200)		(0.349)
Observations	776	776	776	776
Log Likelihood	−350.015	−347.626	−389.954	−178.125

Robust standard errors in parentheses.
*p<0.1; **p<0.05; ***p<0.01

The results for North American researchers are presented in Table 6. As in previous models, the coefficients for the total number of authors are positive and

statistically significant and the geographical distance is negative and statistically significant in both specifications (see columns (1) and (3)). However, the coefficient for a common language shows no statistically significant results for this region (see columns (2) and (4)). Even though English is the lingua franca in higher education research and English is the most widely spoken language in the region, this does not correlate with the intensity of international collaboration of North American scholars with other countries.

Table 6: The results for international collaboration for North American researchers

	N of co–authored papers (log)			
	negative binomial		zero–inflated count data	
	(1)	(2)	(3)	(4)
N of authors in country i (log)	0.559***	0.481***	0.514***	0.428**
	(0.072)	(0.095)	(0.120)	(0.194)
N of authors in country j (log)	0.390***	0.395***	0.629**	0.678**
	(0.114)	(0.117)	(0.304)	(0.276)
Distance, km (log)	−0.334***	−0.280**	−0.363***	−0.304*
	(0.106)	(0.113)	(0.137)	(0.169)
Common language, 1 = Yes		0.326		0.287
		(0.281)		(0.437)
Observations	350	350	350	350
Log Likelihood	−191.206	−190.540	−114.720	−114.795

Robust standard errors in parentheses.
*p<0.1; **p<0.05; ***p<0.01

The results for Asian scholars are presented in Table 7. The negative binomial and zero–inflated count models show similar results consistent with previous regions. The coefficients for the total number of authors are positive and statistically significant, those for distance are negative and statistically significant (see columns (1) and (3)). The higher coefficients for distance for Asian authors compared to European and North American ones suggest that geographical proximity is more important for this region. Asian scholars have significantly less collaboration with distant partners, suggesting that they prefer to choose co–authors from Asia. The coefficient for a common language shows no statistically significant results in both specifications (see columns (2) and (4)). Asian countries share common languages with few countries around the world, thus, it is not significant for the Asian region.

Table 7: The results for international collaboration for Asian researchers

	N of co–authored papers (log)			
	negative binomial		zero–inflated count data	
	(1)	(2)	(3)	(4)
N of authors in country i (log)	0.680***	0.670***	0.454***	0.442***
	(0.122)	(0.122)	(0.115)	(0.116)
N of authors in country j (log)	0.745***	0.760***	0.619***	0.627***
	(0.119)	(0.123)	(0.092)	(0.096)
Distance, km (log)	−0.570**	−0.540**	−0.457***	−0.427***
	(0.227)	(0.233)	(0.045)	(0.034)
Common language, 1 = Yes		0.312		0.274
		(0.351)		(0.321)
Observations	379	379	379	379
Log Likelihood	−150.860	−150.490	−166.353	−166.025

Robust standard errors in parentheses.
*p<0.1; **p<0.05; ***p<0.01

5 Concluding remarks

So far, several attempts have been made to reflect on the state of international collaboration in higher education research using different methods: a bibliometric approach with a content analysis (Kosmützky and Krücken, 2014) and a bibliometric approach with a social network analysis (Kuzhabekova et al., 2015). This paper enriches the current discussion by applying a spatial econometrics technique to contribute to the reflection of higher education researchers about the structure of their international collaboration (Yokoyama, 2016).

In contrast with previous papers, this paper focuses on the collaboration patterns in a multidisciplinary field (Rosenblat and Mobius, 2004; Leahey, 2016; Kliegl and Bates, 2011; Kuld and O'Hagan, 2018). The results show that the number of international publications has grown exponentially during the last two decades. This growth indicates a proliferation of international collaboration in higher education research in general. The findings are in line with the results by Kosmützky and Krücken (2014) and Kuzhabekova et al. (2015), who studied comparative and international higher education, respectively. However, the majority of the articles are still produced by co–authors from the same country, which is a well-known phenomenon

in social sciences (Waltman et al., 2011; Adams, 2013; Henriksen, 2016).

This paper confirms that the intensity of collaboration is negatively associated with geographical distance (Hoekman et al., 2010; Waltman et al., 2011). Notably, the geographical proximity differs to a great extent between various world regions: the dense clustering of European countries makes the intensity of international collaboration far less dependent on geography and more on language commonality. In contrast with previous research on international collaboration in Europe, this paper finds that despite the encouragement of multidisciplinary research by increasing funding opportunities, the substantial role of geography remains (Hoekman et al., 2010; Adams, 2013; Kwiek, 2020b). However, it should be noted that there is a slight decline in the association between the intensity of international collaboration and geographical distance between periods 2000–2008 and 2009–2017, which is in line with the findings by Wagner et al. (2017) and Kwiek (2020b).

Unsurprisingly, the main centres of producing higher education research papers are in major English–speaking countries, followed closely by Western Europe and China (Altbach, 2014; Marginson, 2020; Kwiek, 2020b). Linguistic proximity plays a significant role mostly for European researchers, leaving North American and Asian scholars less connected with international co–authors, which also align with previous evidence found by Hoekman et al. (2010), Adams (2013), and Kwiek (2017). These results confirm the findings by Kosmützky and Krücken (2014) that international collaboration in the higher education community include many places around the world, but still show a core-periphery picture.

The results of this paper show a higher rate of national rather than international collaboration in the North American higher education community, suggesting that they are more engaging in meso– and micro–level research, whereas European scholars emphasise analysis at the macro level and are more inclined to collaborate on the European level (Altbach, 2014; Wagner et al., 2015; Yokoyama, 2016). This study supports the findings by Kwiek (2017, 2020a) and Chen et al. (2019) that scholars in smaller countries are forced to look for outside authors as they benefit the most from it. This paper finds that European countries are the major international partners for non–European countries: the share of cooperation among European countries is decreasing while the share of the international collaboration of European countries with non–European countries is increasing (Wagner et al., 2015).

This paper analyses only two unique features of the field, e.g. geographical and

linguistic proximity, leaving other distinct features of the higher education studies out of the scope, thus, further research can focus on the following questions. How do international organisations such as the World Bank, OECD, and UNESCO influence the higher education community, for instance, research topics? How do the unique features of other social sciences influence the field of higher education, for example, quantitative or qualitative orientations of these fields? All these questions should be carefully analysed using the methods of causal inference design to make proper conclusions.

Ultimately, it is difficult to talk about coherence and solidarity in the higher education community. The scholars are still geographically and methodologically divided and form relatively isolated scientific communities. As a result, this paper offers several policy implications. First, in order to stimulate higher engagement of scholars to work in international partnerships, there should be more institutional support for early and junior higher education scholars from organisations, such as the Early Career Higher Education Researchers Network and the Consortium of Higher Education Researchers. Second, as higher education is a multidisciplinary field of research, one has to address methodological issues that researchers face while conducting such research projects, e.g. by facilitating methodological workshops, summer schools, and special conference sections that aim to promote more involved empirical methods such as casual inference, comparative analysis, etc.

References

Adams, J. (2012). The Rise of Research Networks. *Nature 490*, 335–336.

Adams, J. (2013). The Fourth Age of Research. *Nature 497*, 557–560.

Almeida, J., A. Pais, and S. J. Formosinho (2009). Science Indicators and Science Patterns in Europe. *Journal of Informetrics 3*(2), 134–142.

Altbach, P. G. (2014). The Emergence of a Field: Research and Training in Higher Education. *Studies in Higher Education 39*(8), 1306–1320.

Bartneck, C. and J. Hu (2010). The Fruits of Collaboration in a Multidisciplinary Field. *Scientometrics 85*(1), 41–52.

Chen, K., Y. Zhang, and X. Fu (2019). International Research Collaboration: An Emerging Domain of Innovation Studies? *Research Policy 48*(1), 149–168.

Correia, A., H. Paredes, and B. Fonseca (2018). Scientometric Analysis of Scientific Publications in CSCW. *Scientometrics 114*(1), 31–89.

Frenken, K., J. Hoekman, S. Kok, R. Ponds, F. van Oort, and J. van Vliet (2009). Death of Distance in Science? A Gravity Approach to Research Collaboration. In A. Pyka and A. Scharnhorst (Eds.), *Innovation Networks*, Chapter 3, pp. 43–57. Springer.

Gates, A. J., Q. Ke, O. Varol, and A.-L. Barabási (2019). Nature's Reach: Narrow Work Has Broad Impact. *Nature 575*(7781), 32–34.

Glänzel, W. (2001). National Characteristics in International Scientific Co-Authorship Relations. *Scientometrics 51*(1), 69–115.

Henriksen, D. (2016). The Rise in Co-Authorship in the Social Sciences (1980–2013). *Scientometrics 107*(2), 455–476.

Hoekman, J., K. Frenken, and R. J. W. Tijssen (2010). Research Collaboration at a Distance: Changing Spatial Patterns of Scientific Sollaboration within Europe. *Research Policy 39*(5), 662–673.

Hoekman, J., K. Frenken, and F. van Oort (2009). The Geography of Collaborative Knowledge Production in Europe. *The Annals of Regional Science 43*(3), 721–738.

Jung, J. and H. Horta (2013). Higher Education Research in Asia: A Publication and Co-Publication Analysis. *Higher Education Quarterly 67*(4), 398–419.

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

Kliegl, R. and D. Bates (2011). International Collaboration in Psychology is on the Rise. *Scientometrics 87*(1), 149–158.

Kosmützky, A. and G. Krücken (2014). Growth or Steady State? A Bibliometric Focus on International Comparative Higher Education Research. *Higher Education 67*(4), 457–472.

Kuld, L. and J. O'Hagan (2018). Rise of Multi-Authored Papers in Economics: Demise of the 'Lone Star' and Why? *Scientometrics 114*(3), 1207–1225.

Kuzhabekova, A., D. D. Hendel, and D. W. Chapman (2015). Mapping Global Research on International Higher Education. *Research in Higher Education 56*(8), 861–882.

Kwiek, M. (2017). Collaboration and International Research Orientation: Comparative Findings About European Academics. *Journal of Studies in International Education 22*(2), 1–25.

Kwiek, M. (2020a). Internationalists and Locals: International Research Collaboration in a Resource-Poor System. *Scientometrics 124*, 57–105.

Kwiek, M. (2020b). What Large-Scale Publication and Citation Data Tell Us about International Research Collaboration in Europe: Changing National Patterns in Global Contexts. *Studies in Higher Education*, 1–21.

Leahey, E. (2016). From Sole Investigator to Team Scientist: Trends in the Practice and Study of Research Collaboration. *Annual Review of Sociology 42*(1), 81–100.

Leydesdorff, L. and C. S. Wagner (2008). International Collaboration in Science and the Formation of a Core Group. *Journal of Informetrics 2*(4), 317–325.

Lovakov, A. and M. Yudkevich (2021). Multidisciplinary Roots of Higher Education Research: An Analysis of Citation Patterns in Journals in the Field.

Lovakov, A., M. Yudkevich, and S. Avdeev (2021). Is there a Consensus on Core Journals in the Field of Higher Education Research?

Marginson, S. (2020). The World Research System: Expansion, Diversification, Network and Hierarchy. In C. Claire, L. William, and S. Marginson (Eds.), *Changing Higher Education for a Changing World*, Chapter 3, pp. 25–51. Bloomsbury Academic.

Mongeon, P. (2015). The Journal Coverage of Web of Science and Scopus: A Comparative Analysis. *Scientometrics 106*, 213–228.

Newman, M. E. J. (2004). Coauthorship Networks and Patterns of Scientific Collaboration. *PNAS 101*(1), 5200–5205.

Rosenblat, T. and M. M. Mobius (2004). Getting Closer or Drifting Apart? *The Quarterly Journal of Economics 119*(3), 971–1009.

Tight, M. (2004). Research into Higher Education: An A-Theoretical Community of Practice? *Higher Education Research and Development 23*(4), 395–411.

Tinbergen, J. (1962). Shaping the World Economy. Suggestions for an International Economic Policy. *The 20th Century Fund: New York*.

Wagner, C. S. and L. Leydesdorff (2005). Network Structure, Self-Organization, and the Growth of International Collaboration in Science. *Research Policy 34*(10), 1608–1618.

Wagner, C. S., H. W. Park, and L. Leydesdorff (2015). The Continuing Growth of Global Cooperation Networks in Research: A Conundrum for National Governments. *PLOS ONE 10*(7), 1–15.

Wagner, C. S., T. A. Whetsell, and L. Leydesdorff (2017). Growth of International Collaboration in Science: Revisiting Six Specialties. *Scientometrics 110*(3), 1633–1652.

Waltman, L., R. J. W. Tijssen, and N. J. van Eck (2011). Globalisation of Science in Kilometres. *Journal of Informetrics 5*(4), 574–582.

Wang, L., B. Thijs, and W. Glänzel (2015). Characteristics of International Collaboration in Sport Sciences Publications and its Influence on Citation Impact. *Scientometrics 105*(2), 843–862.

Yokoyama, K. (2016). Reflections on the Field of Higher Education: Time, Space and Sub-Fields. *European Journal of Education 51*(4), 550–563.

6 Appendix

Table 8: The number of joint publications between top 25 countries in 2000–2017

	AU	BE	CA	CH	CN	DE	DK	ES	FI	FR	HK	IE	IT	JP	KR	MY	NL	NO	NZ	PT	SE	SG	UK	US	ZA
AU	0																								
BE	1	0																							
CA	16	1	0																						
CH	0	2	4	0																					
CN	11	3	4	0	0																				
DE	9	2	3	8	1	0																			
DK	2	0	1	0	2	6	0																		
ES	2	1	2	1	0	6	1	0																	
FI	3	5	3	0	7	5	2	3	0																
FR	0	2	2	1	0	1	0	3	0	0															
HK	12	1	5	0	18	1	0	0	0	0	0														
IE	4	0	0	2	0	1	0	1	0	1	0	0													
IT	2	2	0	3	1	1	1	3	1	3	0	0	0												
JP	3	0	3	0	1	1	0	1	0	0	2	0	0	0											
KR	1	0	0	0	2	0	0	0	0	0	2	0	0	0	0										
MY	12	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0									
NL	8	14	3	1	2	8	4	5	1	1	3	0	2	0	0	0	0								
NO	6	5	2	2	3	10	7	0	6	2	0	3	0	0	0	0	5	0							
NZ	33	0	3	0	3	0	1	0	3	0	4	1	0	2	1	3	3	0	0						
PT	6	1	1	0	4	2	1	3	3	0	3	3	2	2	1	0	2	2	1	0					
SE	6	0	2	1	3	1	5	3	9	1	1	0	0	2	0	0	4	12	0	2	0				
SG	6	0	1	0	3	0	0	0	0	1	5	0	0	2	0	1	4	0	1	0	0	0			
UK	96	10	31	6	29	14	6	15	6	8	10	9	5	6	3	7	24	15	19	9	16	1	0		
US	42	4	58	3	25	17	2	7	5	4	7	4	8	7	31	4	10	6	3	5	7	2	51	0	
ZA	8	0	3	0	0	1	1	1	0	0	0	1	0	0	2	0	3	3	3	2	5	0	17	5	0

AU – Australia; BE – Belgium; CA – Canada; CH – Switzerland; CN – China; DE – Germany; DK – Denmark; ES – Spain; FI – Finland; FR – France; HK – Hong Kong; IE – Ireland; IT – Italy; JP – Japan; KR – South Korea; MY – Malaysia; NL – the Netherlands; NO – Norway; NZ – New Zealand; PT – Portugal; SE – Sweden; SG – Singapore; UK – United Kingdom; US – United States; ZA – South Africa

Table 9: The results for all countries in 2000–2008 and 2009–2017, using the negative binomial model

	N of co–authored papers (log)			
	2000–2008		2009–2017	
	(1)	(2)	(3)	(4)
N of authors in country i (log)	0.641***	0.528***	0.632***	0.580***
	(0.141)	(0.172)	(0.055)	(0.065)
N of authors in country j (log)	0.635***	0.604***	0.514***	0.495***
	(0.169)	(0.171)	(0.050)	(0.051)
Distance, km (log)	−0.359***	−0.353***	−0.265***	−0.264***
	(0.135)	(0.134)	(0.058)	(0.058)
Common language, 1 = Yes		0.470***		0.242**
		(0.138)		(0.068)
Observations	252	252	905	905
Log Likelihood	−79.572	−79.001	−442.221	−441.200

Robust standard errors in parentheses.
*p<0.1; **p<0.05; ***p<0.01

Table 10: The results for all countries in 2000–2008 and 2009–2017, using the zero–inflated count data model

	N of co–authored papers (log)			
	2000–2008		2009–2017	
	(1)	(2)	(3)	(4)
N of authors in country i (log)	0.624***	0.495***	0.606***	0.541***
	(0.127)	(0.155)	(0.052)	(0.059)
N of authors in country j (log)	0.263***	0.314**	0.404***	0.361***
	(0.073)	(0.145)	(0.048)	(0.050)
Distance, km (log)	−0.299**	−0.302**	−0.208***	−0.220***
	(0.121)	(0.122)	(0.057)	(0.055)
Common language, 1 = Yes		0.411***		0.300*
		(0.105)		(0.158)
Observations	252	252	905	905
Log Likelihood	−89.711	−90.308	−490.300	−482.428

Robust standard errors in parentheses.
*p<0.1; **p<0.05; ***p<0.01