No Place Like Home - Geography and Culture in the Dissemination of Economic Research Articles

Christiane Hellmanzik*and Lukas Kuld†

13 March 2020

Abstract

This paper examines the existence of distance and border effects in the dissemination of knowledge in economics research using a state-of-the-art gravity model for domestic and international citations between 1970 and 2016 for the top 20 source countries. We extend the model with two novel indicators, English proficiency and bilateral internet ties - two key forces in the dissemination of research and knowledge more generally. Our results show that (i) citations decrease with distance (ii) citations exhibit a significant home bias greater than 1.68, i.e. a more than 50 % higher propensity to cite domestic articles (iii) home bias as well as geographic and cultural distance measures remain significant and at persistent levels over time (iv) bilaterally high levels of English proficiency are insignificant for citations beyond the measure of general language similarity (v) countries with closer internet ties have higher shares of bilateral citations (vi) geographic proximity is insignificant for citations to econometric articles while cultural linkages are significant.

*Department of Economics, TU Dortmund, Vogelpothsweg 87, 44227 Dortmund, Germany. christiane.hellmanzik@tu-dortmund.de

†Department of Economics, TU Dortmund, Vogelpothsweg 87, 44227 Dortmund, Germany. lukas.kuld@tu-dortmund.de

The authors thank Martin Schmitz, Sara Mitchell, Rogelio Mercado and Tara Bedi and the audiences at the Irish Economics Associations annual conference and the annual congress of the European Economic Association for helpful comments. Moreover, the authors thank Jan Lordieck and Markus Friemann for their excellent research assistance.

This version is as accepted for publication in Empirical Economics. The final article can be found at https://doi.org/10.1007/s00181-020-01860-0.

JEL-codes: A14 • F16 • O34

Keywords: gravity, border effect, economics, citations, knowledge dissemination

1. Introduction

While Karl Kraus has likened the translation of a poem to it being skinned at the border, we show that economic research articles have a hard time travelling, too. Although the internet has made research papers readily available to peers around the world and English is widely accepted as the lingua franca of economics, we still find evidence of a substantial bias in quoting research articles in economics. That is, even a research article written in English is more likely to be cited domestically or by colleagues from a country with strong cultural, linguistic or virtual ties. While there are good reasons for such frictions in the dissemination of economic research, such as local topical interests and varying schools of economic thought between countries, it is still striking that in light of globalisation, digitisation and mostly barrier-free access to economic research there is still a significant home bias in citations.

In this paper, we use data on scientific citations to articles written in the 20 top economic research countries from 126 citing countries between 1970 and 2016. The cited articles alone account for 71% of economic research articles indexed by Web of Science. As we have information on both domestic as well as international citations we can assess any prevalent home bias in citations. Moreover, we link these citation data with various distance measures in a gravity model to trace the role of linguistic barriers as well as cultural biases which typically increase with distance.

Our gravity model encompasses the set of commonly used distance variables such as geographical distance, common border, legal system and religion as well as colonial history (CEPII dataset) as well as the language similarity index by **?**. In addition to these data, we include two novel extensions to the gravity model which seem particularly relevant in order to analyse the dissemination of economics research in the last 30 years. First, a novel measure for similarity in English proficiency and, second, a variable for internet linkages based on **?**. We investigate whether the use of English as the lingua franca and the connectedness brought by internet links helps to mitigate barriers to the dissemination of scientific knowledge. It is plausible to imagine these two factors as the driving force behind the integration of economic research globally. Factoring out these differences in a gravity model framework allows us to identify and quantify any home bias in the take-up of economic research articles as the remaining preference for domestic research.

Our contribution to the literature of knowledge diffusion is two-fold: first, we empirically assess the role of borders as well as cultural and linguistic links on scientific

citations using a state-of-the-art gravity model. While this model has been used to study patent citations (Peri, 2005 and Li, 2014), it has not been employed in the literature on scientific citations. The exception are, to a degree, **?** who show a correlation between scientific citations and geographic distance in a simpler setting. Second, we expand the existing literature on knowledge dissemination by introducing two novel factors: English similarity and internet linkages. That is, how similar countries are in their English proficiency as well as how connected they are on the web as measured by hyperlinks between countries.

Our results show that for our sample of citations to economics research articles, the number of citing articles decreases by cultural, geographic, and linguistic distance. That is, national borders play a significant role in the dissemination of economic articles and citations exhibit a strong, significant home bias by an overall factor of more than 1.68, i.e. more than a 50 % higher propensity to cite domestic articles. Moreover, while linguistic similarity bears a significant impact on citations, similarity in English proficiency between citing and cited country somewhat surprisingly has no significant impact overall when controlling for language similarity. Countries with closer internet ties, however, quote each other's work significantly more often, that is bilateral citations increase by up to .12% for a 1% increase in internet hyperlinks.

Finally, we repeat the analysis using a sub-sample of econometrics articles as a comparison group, as these tend to involve more formal mathematical writing which is a type of knowledge that is less language-sensitive. In our analysis we find no significant geographic distance effect. However, national borders and cultural distances still have a significant impact. This could be indicative of local scientific interests, traditions, and networks driving citation patterns across borders when language and topic are less of an obstacle to dissemination.

Although there is an expansive literature on international collaboration networks and citations in the sciences, little attention has been given to spatial biases in scientific citations, other than noting differences in national performance (see **?** and **?**). **?** statistically assess the 500 most cited research institutions in the United States between 1982 and 2001; articles from nearby research organisations are more likely to be cited than articles from research organisations further away. Neither of these papers explicitly assess the channels behind the geographical distance effect.

Slightly different, **?** find that citation, as well as collaboration, relations occur most frequently domestically. Yet, they also find that citations are less correlated with distance than collaborations for a sample of 40 regions between 1997-99. Our paper ex-

pands on these analyses by employing a rich gravity model to scientific citations. We also include two new and highly relevant factors which have shifted the dissemination of knowledge in academia considerably, namely English similarity - as a proxy for how well versed each country is in English therefore reducing transactions costs in terms of understanding the article and decoding the information - as well as virtual proximity - as a proxy for digital connectedness between country pairs.

We explain citations as the consumption and dissemination of scientific knowledge, based on costs and preferences that correlate with our cultural, geographical, and language distance measures. Thereby, we contribute to the understanding of drivers and inhibitors of international citation patterns beyond geographic distance in the field of economics. Unlike this strand of literature, our focus is not on the direct collaboration of scientists but on the diffusion of knowledge as evidenced by citations.

Conceptually, the measurement of knowledge dissemination by scientific citations is closely related to knowledge exchange in more commercially oriented research as captured by patent citations. Similar to our study, **?** uses a gravity model for European patent data and finds significant border and distance effects while common language and culture are conducive to patent collaborations. In the same vein, **?** finds that only 20% of average knowledge as measured by patent citations is learned outside the patent's region of origin with 9% being learned outside the country of origin.[1]

These findings are contrary to **?** who show that the impact of country borders and distance on patent citations decreased until mid-1990s followed by growth until 2009. Furthermore, **?** in a study of internationalisation of technology as measured by patent data find that countries which are closer to each other and share similar specialisations as well as language also tend to cooperate more in terms of doing collaborative research. This is supported by the fact that **?** finds that more science-heavy patents are less likely to be traded across borders. However, according to **?** patent citations might be a bad indicator for actual knowledge flows and rather reflect common techniques and sources. Thus, scientific citations might similarly reflect more factors than just knowledge flows even if they depend less on underlying technologies.

Altogether, despite better integration in economics (**?**) and increased co-authorship as suggested by **?** it is not clear ex-ante what the impact of various distances on citations - which are a looser form of scientific connection similar to patent citations - would be. This strand of literature indicates that economics research due to its internationality as

[1]More generally, **?** provides evidence for strong geographic concentration of patent citations which is further confirmed by **?** who links this concentration of knowledge back to the lack of mobility of inventors who are likely to remain in their co-inventor network.

well as generalisability is a good case for our analysis of knowledge dissemination based on citation counts.

Moreover, **?** and **?** show that tastes and preferences between countries correlate with bilateral trade. This motivates the inclusion of "softer" cultural proximity indicators next to geographic distances. Moreover, the relatively close ties in research between a handful of relatively rich and research-inclined countries in economics, with a noticeable US dominance, is very similar to what we observe in goods and services trade, as well as international migration and capital flows (see for example **?**, **?**).

In the remainder of the paper we proceed as follows. In Section 2 we describe the various data sources and provide summary statistics. In Section 3 we introduce a gravity model of citations and the corresponding empirical gravity model, while the empirical results are shown in Section 4. Section 5 concludes.

2. Citation and gravity model data

2.1. Citation data

The citation data for this analysis are based on Web of Science's (WoS) indexation encompassing all articles worldwide that cite an economics article authored by an economist affiliated with an institution in one of the twenty leading countries in economics research between 1970 and 2016.[2] Our sample covers 71% of economic research articles indexed by WoS. We chose the top twenty countries based on the number of articles indexed by WoS, not citations. However, the resulting sample is consistent with the one resulting from an exercise of locating the most cited economists (see **?**).

The resulting citation data connects four nodes. First, a citation links the country of the university of the citing article's author to the country of the university of the cited article.[3] Second, the citation links the publication year of the citing article to the publication year of the cited article. For instance, an American article from 1996 may cite a Spanish article from 1992. Figure A.1 illustrates the data collection and format. In this example, three articles are indexed as being authored in the USA, in the field of economics in 1993. These articles are cited by articles written in 1995 as follows: one citing article is written by a British author, another by a Canadian, and two more by

[2]The twenty leading countries are Australia, Belgium, Canada, China, Czech Republic, Denmark, France, Germany, India, Israel, Italy, Japan, Netherlands, Republic of Korea, Romania, Spain, Sweden, Switzerland, United Kingdom, and USA.

[3]Articles with authors in multiple countries are attributed to each country.

American researchers. This gives us gives us a count of two for the specific link USA 1995 to USA 1993.

The result of this data collection effort encompasses 126 citing countries and 20 cited countries.We, thereby focus our analysis on the citations to the main economic research countries.[4] The data comprises 1,122,000 country and year pairs in total and including both domestic and international citations which makes it possible to estimate a home bias in economics. 76 % of all citation pairs in our year-to-year data set are zero. However, since our main variables of interest are time invariant, we conduct our main analysis using data aggregated by country pair (2,380 observations). In this sample, we observe less than four % of the aggregated country pairs with no citation flow.

In addition, we collected a smaller comparison sample encompassing econometrics articles by the top ten leading countries only in order to assess whether more formal and mathematical economic articles travel the globe more easily. This sub-sample comprises 53 citing and 10 cited countries. Aggregating by country pair yields 908 observations in total.

A caveat of our data is that we do not count total citations but the number of articles that set a set of articles published in a specific country and year. It is important, however, that we observe not only the two countries but also the publication years of the citing article and of the cited article. This gives as a better measure for citation intensity than just country aggregates. Conceptually, this difference is vaguely similar to counting importing companies instead of total imports in a world with many small companies. In the example given in Figure A.1, columns 5 and 6 show possible differences between counting citing articles and citations. For instance, a British article might cite American articles from eight of the last ten years but Canadian articles from only two of the last ten years. Still, this idiosyncrasy of the data might lead us to underestimate relative citations, in particular those to the United States as the greatest producer of economics articles.[5] Moreover, the issue is less important in earlier years when the number of references given in an article was considerably lower.[6]

[4]We exclude citing countries with less than 100 total citing articles from our analysis.

[5]To understand why, imagine picking a random economics article written in 2016; it is reasonable to assume that the article cites at least one article from the United States. By just counting citing articles for the United States overall, we would end up simply counting all 2017 economics articles. However, a new article might not reference American articles from each of the last ten years and is very unlikely to reference 47 American articles covering each year from 1970 to 2016. Technically, we only observe an article's first citation to a given year-country which underestimates countries and years that have relatively high citation averages.

[6]The average number of references given per paper almost doubled over the last twenty years from 25 to 40 (see **?**).

2.2. Gravity model data

All geographic variables come from the CEPII data sets (2016): *distance* captures the bilateral geographical distance between two countries' capitals, while *time difference* refers to the time zone difference between two countries. We also use indicators for countries which share a common border (*contiguity*), for a shared colonial past (*colony*), a similar legal system (*common legal*), and an index for religious similarity (*common religion*). In our estimations we also include the aggregated index for *language similarity* constructed by **?** which summarizes evidence about linguistic influences including common official language and common native language and measures of linguistic proximity. Moreover, we include an indicator variable to reflect if the article from both citing as well as cited country is a member state of the EU.

In addition to these data, two extensions to the gravity model seem particularly relevant in order to analyze the dissemination of economics research in the last 30 years, namely the importance of English as lingua franca for research as well as the rise of the internet in the late 90s and its impact on publications.

2.2.1. English Similarity

In order to investigate the impact of English as the *lingua franca* in economics more specifically, we construct a novel index based on differences in English proficiency in the general population. To measure English proficiency (EP), we start from the overall language index and take the linguistic proximity to the United Kingdom for each country, setting all observations above 0.9 to 1 (the United States for instance). In a second step, we take the share of the population proficient in English (EF) given by the EF English Proficiency Index.[7] Based on this, we create *English Similarity* as follows:

$$English\ similarity_{i,j} = 1 - |EP_i - EP_j| \tag{1}$$

$$\text{with } EP_k = \begin{cases} 1 \text{ if k an English language country} \\ EF_k \qquad\qquad\qquad \text{if k } \in \text{ EF Index} \end{cases}$$

We hypothesize that *English similarity* has a positive impact on citations. For instance, Sweden and the Netherlands are historically open to publish research in English and have similarly high levels of English proficiency. Economists from these countries

[7] `http://www.ef.edu/epi/`

might be more likely to read the same journals and be able to communicate using English instead of their native languages. Conversely, a Spanish economist in the 1980s might have had less exposure to English language journals than their Northern European colleagues.

2.2.2. Internet linkages

In order to capture the impact of the internet on international citation flows we use an indicator for internet linkages, namely the amount of hyperlinks set between countries as employed in **?**. To this end, we use use bilateral, inter-domain hyperlinks that internationally connect webpages in country A to webpages in country B as in **?** to capture information flows via the internet more generally. Their 'virtual proximity indicator' indicator is mainly based on hyperlink data provided by **?** who covers the years 2003 and 2009 for up to 87 countries for which Chung found more than 9.3 billion hyperlinks included in 33.8 billion sites from 273 different top-level domains.[8]

Due to the bidirectional nature of the data, bilateral hyperlinks reflect the number of links from websites with domain .xx (i.e. from the country with domain .xx) to domain .yy (i.e. to the country with domain .yy) and vice versa. **?**'s 2009 wave of data are more precise and granular than most comparable data sources as Chung developed an attribution method which 'cracks', and thereby uniquely identifies, the host country of a *.com* domain for his sample of 87 countries in addition to encompassing country top-level domains (ccTLD), such as *.it* for Italy, thus providing a more accurate reflection of internet linkages than other data sources.[9]

Thus, our second hypothesis is that the amount of internet linkages positively impacts citations and decreases home bias considerably as transactions costs have become virtually zero in accessing the state of the art in economics research locally as well as around the globe.

2.3. Stylised facts on economics' citations

Figure A.2 depicts the evolution of economics research over our sample period. We find a strong increase in articles published since the 1970s and most notably the diagram

[8]To this end **?** uses Yahoo's search function and LexiURL Searcher, a social science web analysis tool developed by **?**. At the time, Yahoo had indexed about 47 billion websites. For more detailed information on obtaining the measure of bilateral hyperlinks, please refer to **?**.

[9]For the United States, usually the sum of the domains *.edu, .us, .mil and .gov* has been used **?** in the literature. In previous studies (e.g. **?**), the *.com* domain had either been disregarded or completely attributed to the United States.

reflects the dominance of US based economists in publications. Furthermore, Figure A.3 depicts the number of citing articles between 1970 and 2016 for the ten largest receiving countries. Figures on the left show the sum of articles citing publications from the respective origin country published in the same year and up to nine years earlier. The right column shows these counts divided by the yearly sum of the top ten countries. Not surprisingly, international citations reflect the dominance of the United States as a production country with about three times the number of citing articles of British articles, the next biggest origin country.[10]

While there is a constant upward trend in international citations for all top production countries there is a marked increase in citations in the late 1990s and early 2000s. This pattern persists when we exclude domestic citing articles and citing articles from any of the top production countries which is indicative of the fact that increased production of articles goes hand in hand with a similar trend in citations both domestically and from abroad. The latter could be explained by the fact that the internet facilitated international visibility as well as integration of the research world by allowing for better availability, accessibility and information exchange from the late 1990s onwards.

Figure A.4a analyse the bi-directionality of citation flows between the top 20 countries in a circle diagram depicting the entire observation period from 1970 to 2016 for domestic as well as citations from abroad. This diagram underscores the dominance of the US as the largest origin as well as recipient of citations in economics for each of the depicted partner countries. Next to that, it is interesting that there is a relatively high share of domestic citations as well as 'neighbourly citations' for demeaned citation counts as depicted in the second circle diagram.[11] This gives a first indication on the role of language and distance and cultural biases in citation patterns. For instance, France and the Netherlands are Belgium's preferred source of economic knowledge when accounting for differences in national production levels while Israel exhibits a particularly strong citation link with the US. Table B.1 confirms the most prevalent citation pairs and again the US's role as the strongest country in economics research becomes apparent as well as the strong tendency for domestic citations.

The variables employed in our gravity model framework and their summary statistics

[10] For comparison, **?** count articles in 441 economics journals from 1980 to 2014. In this sample, North American researchers alone authored half of the world's economics articles indexed by WoS between 1980 and 2014 which then received 75% of total citations. 98.4% of the economics articles from the top ten countries are written in English (WoS classification). The next two languages are French with 0.8% and German with 0.3% of the total indexed article output of these countries (WoS).

[11] To obtain positive counts, we show the exponential of log demeaned by citing and cited country (thereby using the geometric mean).

are reported in Table B.2. It is worth noting that we have a relatively high English similarity between non English-Language countries in our sample of 91% while the average geographical distance between cited and citing article is 5,125 miles. In addition, Table B.1 depicts the top 25 citation pairs 1970 to 2016, as well as distance, language, and hyperlinks in 2009 the wave for which we have the fully '.com-cracked' information from **?**. In 2009, the largest number of bilateral hyperlinks arose from webpages hosted and visited from within the US with about 59 million links, followed by links set from UK to US websites (Table B.1). Interestingly, we see that with the UK and the US dominating in both citations as well as hyperlinks most of the other top twenty countries which display a high degree of international interconnectedness are also the ones which are highly connected in terms of their economics research.

Table B.3 displays the correlation matrix of our variables of interest with the range of gravity variables used in this paper. All variables are demeaned by citing and cited country, and given in the form in which they enter the regression analysis. Not surprisingly, we find that English and overall language similarity are highly correlated, yet we believe that separating out English from the overall language effect makes the analysis richer in terms of understanding major shifts in economics research.

3. A gravity model of citations in economics: theoretical and empirical framework

In this section we describe the theoretical foundation and empirical specification of this paper. In order to assess border and distance effects for citations in economics, we translate a standard multi-country trade model into the realm of citations. That is, we define countries as research communities who exchange research articles which represent product varieties of highly differentiated goods similar to goods or services trade or indeed patents. A gravity model framework lends itself particularly well for an analysis of citations between and within countries as it offers a theory as well as an empirical solution to deal with unobserved (or hard to measure) quality of research while explicitly modelling preferences as well as transactions costs. Thus our research strategy is similar to Picci (2010) and Peri (2005) and others who apply the gravity model to patent citations.

3.1. The gravity model - theory

The key idea of our model is that any research community can choose to consume research from home and abroad in order to produce their novel contribution to the economics literature. During this process they incur a cost of understanding and incorporating the existing knowledge and this cost is related to various measures of distance in physical, linguistic as well as cultural space.

The research community in country i consumes research articles from research communities abroad as well as domestic present and past research. Using these articles to create new research knowledge, it derives a utility expressed in the following Equation 2:

$$U_i = \left(\sum_{j=1}^{N} \alpha_{ij}^{\frac{1-\sigma}{\sigma}} m_{ij}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \tag{2}$$

Here, m_{ij} is the number of articles cited that are written by researchers in country j. $\alpha_{ij} \geq 0$ is a weight for the preference of researchers in country i for research from country j.

$$C_i = \sum_{j=1}^{N} p_{ij} m_{ij}, \text{ with } p_{ij} = p_j t_{ij} \tag{3}$$

The research community in country i spends its time and attention budget C_i on reading and using research from all countries incurring both search and application costs. For researchers in any country i, research articles from country j have the same ex-'lab' price p_j. However, transactions costs ($t_{ij} \geq 1$) arise depending on the cultural, historical, or geographic distance between the scientists' countries. That implies that articles by scientists from a different research tradition in terms of methods or topics or a combination thereof might be more difficult to incorporate. For instance, a country's research on a certain topic might be more or less theory-driven and differ in favoured modelling choices from research in other countries. [12] Thus, we have a final price p_{ij} in country i for an article written in country j: $p_{ij} = p_j t_{ij}$.

Maximizing the research community's utility (Equation 2) subject to time and attention costs in Equation 3 yields country i's demand m_{ij} for economics articles written in

[12] In addition, incentives to cite articles within local networks create opportunity costs for citations to unconnected researchers.

country j, where P_i is the price index in country i.[13]

$$m_{ij} = \left(\frac{\alpha_{ij}}{P_i}\right)^{1-\sigma} p_{ij}^{-\sigma} C_i \tag{4}$$

For our empirical analysis we are interested in the role of the various dimensions of geographical and cultural proximity on bilateral citation patterns as specified in Equation 4 where it appears in two parameters: transactions costs t_{ij} and the preference weight α_{ij}. Cultural proximity, namely linguistic, religious and virtual ties, are negatively correlated with transactions costs as linguistic similarity, trust (Guiso et al, 2009) as well as general ease of networking, for example, result in lower information and search costs and therefore enhance bilateral citations. For α_{ij}, a stronger preference in country i (a higher α_{ij}) for economics articles from country j leads to higher citation levels (larger imports) from country j for $\sigma < 1$.

3.2. The gravity model - estimation

In our estimation, we capture the impact of cultural proximity on economic research affinity between and within countries. To this end, we use state-of-the-art proxies for cultural proximity, such as common spoken language and religion similar to, for example, **?**, **?**, or **?** using bilateral service or capital flows. In addition, we introduce two novel variables into the gravity model which seem particularly relevant for citations, namely bilateral internet links between countries (as also employed in **?**) and common English proficiency (see Section 2).

We link these cultural distance proxies to citation aggregates between countries to estimate an empirical gravity model for domestic and bilateral citation flows in economics.

$$\mathbb{E}\left(citing\ articles_{ij} | \mathrm{IM}_i, \mathrm{EX}_j, D_{ij}\right) = exp\left(\mathrm{IM}_i + \mathrm{EX}_j + \delta D_{ij}\right) \tag{5}$$

As dependent variable we use the count of citing articles by authors working in country i to articles by authors active in country j, $citing\ articles_{ij}$. This includes domestic citations where citing and receiving country are identical. Using a trade analogy, we observe the import of citations from articles published by authors who work in country j

[13] $P_i = \left(\sum_{j=1}^{N} (a_{ij} p_{ij})^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$

The bilateral resistance factors in D_{ij} provide and estimation for the ratio of costs t_{ij} and preferences shown in Equation 4. This includes a wide range of variables such as the indicator variable *home* to reflect whether the article cited is by a domestic colleague, physical distance ($distance_{ij}$), time zone difference ($time_{ij}$) and the existence of common borders ($contiguous_{ij}$) between citing and cited country.

Moreover, we use indicator variables for former colonial ties *common colonial history* as well as *common religion*, *common legal origin* and the index for *language similarity* developed by **?** in our baseline estimations as typically done in the trade literature to capture any persisting historical, juridical or cultural links between countries which could bear an impact on citation flows. Thus, we use the following set of bilateral resistance factors D_{ij} in our estimation:

$$\begin{aligned}\delta D_{ij} =& \delta_1 home + \delta_2 log(geographic\, distance) + \delta_3 time\ zone\ difference_{ij} \\ &+ \delta_4 contiguity_{ij} + \delta_5 colony_{ij} + \delta_6 common\ legal\ origin_{ij} \\ &+ \delta_7 common\ religion_{ij} + \delta_8 EU + \delta_9 language\ similarity_{ij} \\ &+ \delta_{10} English\ similarity_{ij} + \delta_{11} log(hyperlinks\ year)_{ij} \end{aligned} \tag{6}$$

Next to this, the estimations include citing (importer) (IM_i) as well as cited (exporter) fixed effects (EX_j) to control for any unobservable country-specific factors, such as GDP or size and budget of the research community as well as supply and quality of research output. All these affect citation flows in economics by affecting the domestic supply of economics research, however, they are very hard to observe. Thus, we control for them using double fixed effects in our specifications. Moreover, we use time fixed effects when using disaggregated data to control for any trends in citation networks.

3.3. The gravity model - estimation method and data aggregation

We estimate variations of our baseline specification given in Equation 5 using quasi-Poisson regressions. Doing so, we account for the resistances being multilateral, that is bilateral costs and preferences relative to the rest of the world, and their multiplicative impact (see Equation 4). By using Poisson regressions and double-fixed effects, we force the sums of expected citation flows to equal actual citation flows. In turn the estimates are consistent with a structural adjustment for multilateral relative distances (**?**). Furthermore, we use re-estimated, clustered standard errors (often called PPML,

see **?**) to correct for country heterogeneity.

For our baseline estimation we aggregate citations received in the year of publication to 2016, that is, we regard citations as cumulative over time and estimate a cross-section as the resistance variables summarised in D_{ij} are time-invariant. This strategy eliminates the risk of inflating the data in the absence of variation over time.[14] However, we also provide estimates for the key distance measures from repeated yearly regressions and disaggregated data.

When using aggregated, time-invariant data, we might obscure heterogeneous effects over time. Also, fixed effects for citing and cited countries cannot control for change in omitted factors over time. Therefore, we repeat the baseline estimation by decade to address these concerns. In addition we plot yearly estimates for the three main variables, home bias, geographic distance, and language similarity.

For comparison with other research and as robustness test, we estimate negative binomial regressions and linear models using OLS with transformed article counts (inverse hyperbolic sine or asinh)[15] and varying samples.[16]

4. Results

4.1. Gravity and home bias

4.1.1. Overall findings

Table B.4 reports the results of our baseline gravity specification varying sample and estimation method. These we analyse in detail to motivate our baseline estimation to which we add our novel variables for English proximity as well as international hyperlinks.

Column (1) shows the results of a quasi-Poisson estimation of the gravity model on aggregate citation flows in economics between 1970 and 2017 including fixed effects for citing and cited country, while Columns (2) and (3) present results for the same regression but using negative binomial and OLS estimations, respectively. Geographic,

[14] The results are unchanged for Poisson with clustered errors but other methods are less robust with respect to the aggregation level.

[15] $Asinh(x) = ln(x + \sqrt{x^2 + 1})$. For $x \geq 2, asinh(x) \approx ln(x) + ln(2)$, but $asinh(0) = 0$. Also suggested for citation counts by **?**, for instance.

[16] However, both, OLS and negative binomial regressions, have stronger identification assumptions and are not consistent with the theoretical multilateral resistance model. As an indication, they overestimate the actual sum of citing articles considerably. For instance, the total flows involving the United States are overestimated by 20 % (OLS) and 31 % (negative binomial) in the standard specification. While smaller countries are underestimated, the ratios of total estimated flows to actual flows are 1.15 (OLS) and 1.22 (NegBin2) as opposed to 1 using Poisson.

cultural, and linguistic distance are all consistently estimated to impact negatively on the aggregation of articles that cite a country's research output in the field of economics. Geographic distance is estimated to reduce citing articles 0.23% or less across models for each 1% increase in distance. This effect is not particularly large relative to what is typically found in the trade literature (where it usually is estimated to be closer to unit-elasticity, for example, **?**).

Moreover, economic research is cited significantly more often by articles from within the country than by articles that are authored behind a national border. When looking at all three specifications, the overall home bias is estimated in the range of 1.68 and 3.86 ($e^{0.52}$ and $e^{1.35}$ respectively, implying that domestic articles that domestic articles are cited up to three times as often as those from abroad even if accounting for the included distance measures. In terms of its magnitude, the home bias we find for scientific citations is substantially lower than estimates in goods trade, which are estimated to be 3.74 by **?**, 4.22 by **?** for European samples but may be found to be as large as 30.88 by **?**. Moreover, our finding on home bias is in line with **?** who finds that only 20% of average knowledge as measured by patent citations is learned outside the patent's region of origin with 9% being learned outside the country of origin.

As our results are consistent across specifications, Columns (1) to (3) provide insights into the stability of findings across estimation techniques. We conclude that the evidence presented above on home bias and distance persist in the aggregated data is not sensitive to the estimation methods. This is a strong indication that our finding of an existing home bias in economics research as well as the existence of cultural barriers is robust. This is confirmed in Column (8) where we repeat our baseline Quasi-Poisson estimation without the US, the dominating producer of scientific output.

Columns (4) to (7) present estimations to further investigate whether this baseline result is robust across specifications. Column (4) repeats the OLS analysis presented in Column (3) using disaggregated data and a model without country fixed effects. However, it includes time fixed effects and unilateral mass variables, that is the total number of citations, population as well as GDP of citing and cited country, respectively. The results are consistent with the previous models indicating that the effect of absolute distance estimated in this specification is similar to the relative distance estimated in the main specification in Column (1).

In Column (5) and (6) we present the results of a logit and a quasi-Poisson analysis of disaggregated scientific citations to assess the impact of the many zero observations in our sample. Thus, we effectively analyse the extensive and intensive margin of scientific

citations. That is, we first asses the probability of a positive flow in the disaggregated data and, in a second step estimate the intensity of citation flows without zeros. Although the results are slightly different in magnitude, the overall results are still in line with previous estimate. In particular, this shows that the step from zero to positive counts is not systematically different with respect to the considered covariates.

4.1.2. Evolution of home bias and geographic distance over time and over the article's life-cycle

To explore the evolution of home bias over time we repeat our empirical analysis estimating separate regressions for each decade thereby allowing covariates to vary over time. The results of this exercise are presented in Table B.5. Figure A.4 shows the yearly evolution for three key variables.[17] We find that indeed the home bias is strongest in the early 80s and decreases markedly since the turn of the century as depicted in Plot (a) of Figure A.4. However, this might be partly driven by more references per article as discussed in the data section.

When excluding the US as both citing and cited country from our analysis in panel (b), the results are comparable. This decline in home bias over time is in the same vein as the findings of **?** who also provides evidence for a declining border effect in trade and **?** who show that the impact of country borders and distance on patent citations wanes until the mid-1990s followed by a period of growth until 2009.

When focusing on the results for the evolution of the distance effect over time (Figure A.4, Plot (c) and (d)), we find that the distance effect appears to be slightly increasing in the 1980s and 90s while being constant after the millennium. Thus, we do not find evidence for 'the death of distance' for economic citations.

Figure A.5 shows how the overall home bias develops in the years after publication. First, we see that in the year of publication of a novel article it is cited almost equally domestically as well as abroad. This is probably a spurious observation caused by relatively low citation counts in the year of publication. In the first years following, however, there is a marked home bias in citation patterns which wanes slowly over time. This holds true for both the full sample and when excluding US authors. This is mirrored in the impact of geographic and linguistic distance over time which slowly loses in importance over the years following publication. When comparing to the literature on patent citations, our findings for the dissemination of economic knowledge across the globe are in line with the findings by **?**, for example, who shows that border and distance

[17] The estimation tables underlying the results for the figures presented are available upon request.

effects increase over time for same-age citations among patents.

4.2. The role of English and the internet for citation patterns and considering econometrics citations only

Typically, home bias in goods trade or international investment portfolios as well as trade in patents are attributed to transactions costs, in particular information asymmetries. Both these factors might also be relevant to economics research and its dissemination despite the fact that information in this context should be more broadly available and that there are more or less agreed upon quality standards in most economic journals. The many university, journal and individual researcher rankings can be understood as efforts for transparency in this context. Nevertheless, whatever their prohibitive effect may be, transactions costs have undergone a shift over the course of our sample period with the rise of the internet and therefore widely available access to economic publications, as well as with the convention of English as the universal scientific language. In the following we expand the above analysis by these two possible avenues to explain the above findings.

4.2.1. English similarity

The results of the impact of linguistic barriers in scientific ciations and therefore on the dissemination of knowledge are presented in B.6. When exploring the role of linguistic barriers we start with the inclusion of linguistic similarity as defined by **?** in order to understand the role of language for the dissemination of scientific citations more generally. We find that linguistic similarity is significant and positive indicating that countries which share the same or a language from the same linguistic family are also more likely to reference each other's work.

In the standard specification in Column (1), a standard deviation increase in the language similarity index (**?**) increases the number of expected citing articles by 13 % from the mean. [18]

We extend the set of gravity model variables by a novel indicator for similarity in English proficiency in Columns (4), (5) and (6) of Table B.4 as English could be one of the key barriers to international citations, in addition to the more general language similarity indicator. However, our estimations indicate that English similarity is not

[18] Additionally we ran a specification excluding the US (results not presented but available upon request). If we do so, this value rises to 16 % but can be as big as 90 % for larger language differences.

significant when including language similarity more generally. In line with the literature, language can be viewed as a proxy for both information asymmetries as well as cultural proximity more generally. In our case, it seems that the role of language in citation patterns for economists reflects cultural proximity which could explain why English similarity is insignificant in explaining citation patterns. Another possible explanation could be that we observe that countries with relatively low levels of English proficiency prefer research from native English speaking countries, in particular, the United States and the United Kingdom.

While there is good evidence that countries with high English proficiency contribute more to global economic research in English language journals, these level differences are factored out in our estimation. Our result for English similarity on citation patterns in economics are in line with the findings of **?** who show that linguistic similarity is conducive to trade more generally. Nevertheless, beyond language as a proxy for transactions costs, it is striking that English similarity does not have a significant impact on citation patterns since we only consider English language articles in our analysis.

4.2.2. Internet linkages

The second factor which had a major impact on scientific research and economics, more specifically, is the rise of the internet in the late 1990s which decreased transactions and information costs to virtually zero. In order to capture the importance of the internet to overcome frictions between countries in terms of citations we use an indicator employed by **?** capturing bilateral hyperlinks on websites as an indicator of countries' international virtual linkages. A caveat in the data is that we cannot assess the impact of virtual proximity on home bias as intra-national hyperlink data are not available. Our hypothesis is that bilateral hyperlinks have a positive impact on citation flows in economics.

The results of this empirical exercise are presented in Table B.7. We find that the number of hyperlinks set between citing and cited country in both 2003 (Column 2) and 2009 (Column 3) have a significant, positive impact on citations indicating that countries which are virtually more integrated also share closer research ties in economics. In terms of its magnitude the effect amounts to an increase in the amount of citations by .12% for a one percentage point increase in bilateral hyperlinks for 2009 while the effect is .04% for 2003 hyperlinks.

The positive impact of virtual integration on economics citations is largely in line with comparable literature in international economics, such as **?** who reports that a doubling

of internet usage in a country leads to a 2% to 4% increase in services trade or **?** and **?** who find that bilateral hyperlinks have a positive impact on portfolio investments and trade in audiovisual services. In terms of dissemination of knowledge our finding that digital ties decrease distance effects is important as it indicates that indeed digital access to scientific knowledge is conducive to its diffusion.[19]

In Columns (5) and (6), we run estimations for the growth of citing articles from 2003 to 2009 as this is the time window for which data for the bilateral hyperlinks indicator is available. Citation growth should be less prone to persistent trends which might be driving our results and as citations might be endogenous we re-estimate our analysis with citing articles after 2008 as the dependent variable while including citing articles before 2003 in logarithmic levels thereby effectively lagging the impact of hyperlinks - in both level and growth terms - on citation growth.

Column (6) estimates the same including the level of 2003 hyperlinks and the growth of bilateral hyperlinks between 2003 and 2009 as explanatory variables. The exercise presented in Column (5) shows that the growth in citing articles largely follows the same pattern as the level of citing articles before 2003; column (6) shows that the 2003 level of hyperlinks is significant while the coefficient for the growth rate indicates that internet hyperlinks have developed differently from economics citations in this time period.

These estimations are indicative of the importance of early internet links for fostering research ties between countries. Maybe it suggests that the internet's role in international visibility and networking in economics research is particularly strong in its early days.

4.2.3. Citation patterns for econometrics articles

As an extension and control, we repeat the estimation presented above with citations to econometrics articles only.[20] The underlying idea is that geographically focused research interests such as the impact of Brexit or East Asian trade are less important in econometric research. In addition, neither language nor cultural preferences should be as important for the more technical and method-focused discipline of econometrics as they perhaps are for economics in general. Although the sample is much smaller we can shed further light on the role language and cultural preferences play when contrasting

[19] The lower effect estimated for geographic distance reflects the change in the sample size and year. Table B.6, for instance, shows that the omission of the home variable alone does not affect the same distance estimate significantly in the baseline sample.

[20] These articles are classified by WoS as economics as well as mathematics or statistics and probability. The data are collected for articles written in the top ten leading countries between 2004 and 2008.

results based on econometrics articles only to a comparable sub-sample of economics articles.

Table B.8 presents the results of our empirical exercise. Columns (1) and (2) present estimates for citations to the then leading econometric research countries. The other four columns exclude Russia which is a leader in econometrics but not in the top 20 of economics for comparability. Interestingly, we find that for econometrics articles the coefficient on geographic distance is not significant in any of the estimations while we find comparable coefficients for cultural distances and language as well as national borders for international citation patterns for econometrics articles. Indeed, the coefficients on *common religion* as well as *language similarity* are larger for econometrics articles than for economics. At the same time, the indicator for *bilateral hyperlinks* in 2009 is insignificant, however, this is also the case for the sub-sample of economics articles. All in all, these findings suggest that there are local research preferences and interests which prevent research traveling the globe even in areas where language and culture are less relevant due to the mathematical nature of the articles.

5. Conclusion

In a novel theoretical application, this paper analyses international citation flows in economics research between 1970 and 2016 using a gravity model framework. This sheds light on the role of geographical, cultural, virtual and linguistic distances in the dissemination of knowledge in economics. To this end, we use data on citation aggregates between and within countries which we link to measures for geographic and cultural proximity, as well as novel data on internet hyperlinks between countries and refined linguistic indicators.

Our results show that citations to economic articles decrease with the geographic distance with a factor that is smaller than commonly found in the literature on trade in goods and particularly services, albeit considerable if we take the premise that research disseminates without borders or cultural preferences. In light of this, a key finding of our paper is that researchers have a more than 50% higher propensity to cite domestic articles when controlling for other factors suggests that citations in economics exhibit a strong and significant home bias in all twenty leading countries. As such, scientific citations are somewhat comparable to patent citations as discussed in **?**.

This ties in with our finding that, contrary to linguistic proximity more generally, there is no significant English proficiency effect. Thus, we find no direct evidence for

a boosting effect of similarly good or bad levels of English for the affinity in economic research. In addition, our paper shows that countries with closer internet ties have significantly higher shares of bilateral citations. Lastly, a sub-sample of econometric articles shows similar citation patterns except for the pure geographic distance. This suggests that the attention given to other researchers is affected by cultural and other distances even for a literature in which the barriers to the dissemination of research might be smaller due to its technical nature.

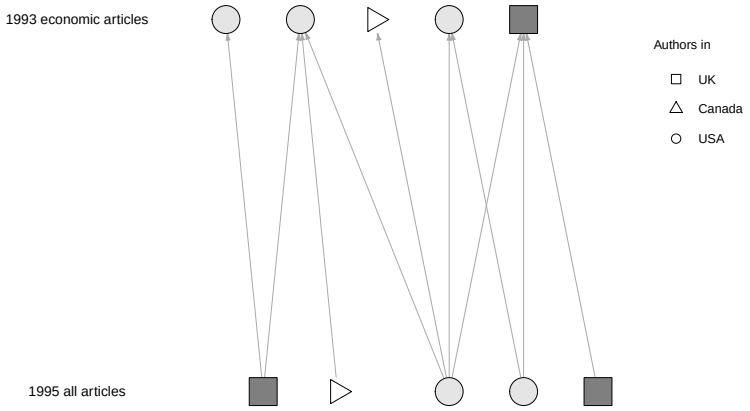
Thus, we do not find comprehensive evidence that the home bias or geographic and cultural distances have lost their importance by the use of the internet and English as a universal scientific language. In particular, we observe the persistent influence of language and geography on citation patterns up to today.

6. Compliance with Ethical Standards

This article does not contain any studies with human participants or animals performed by any of the authors. The authors declare that they have no conflict of interest.

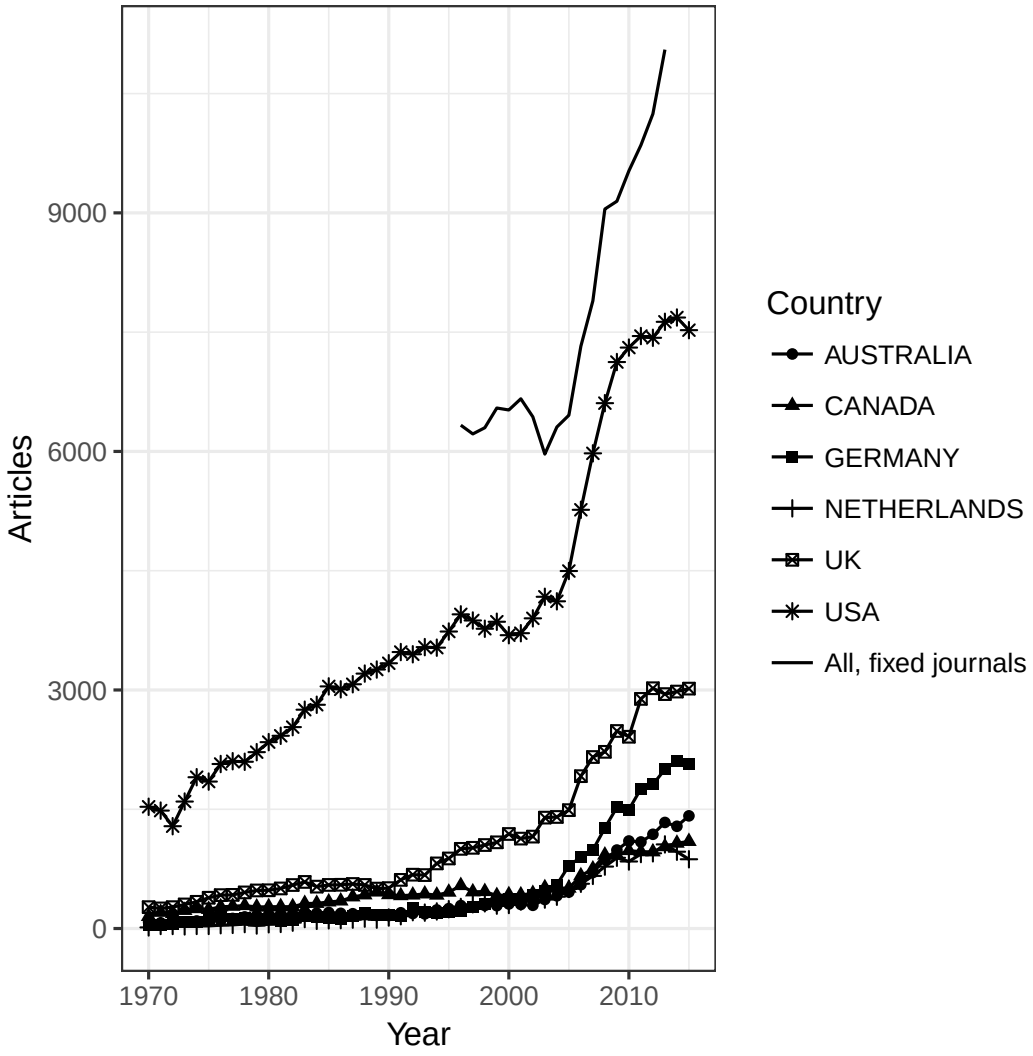
A. Figures

Figure A.1: Citation data example



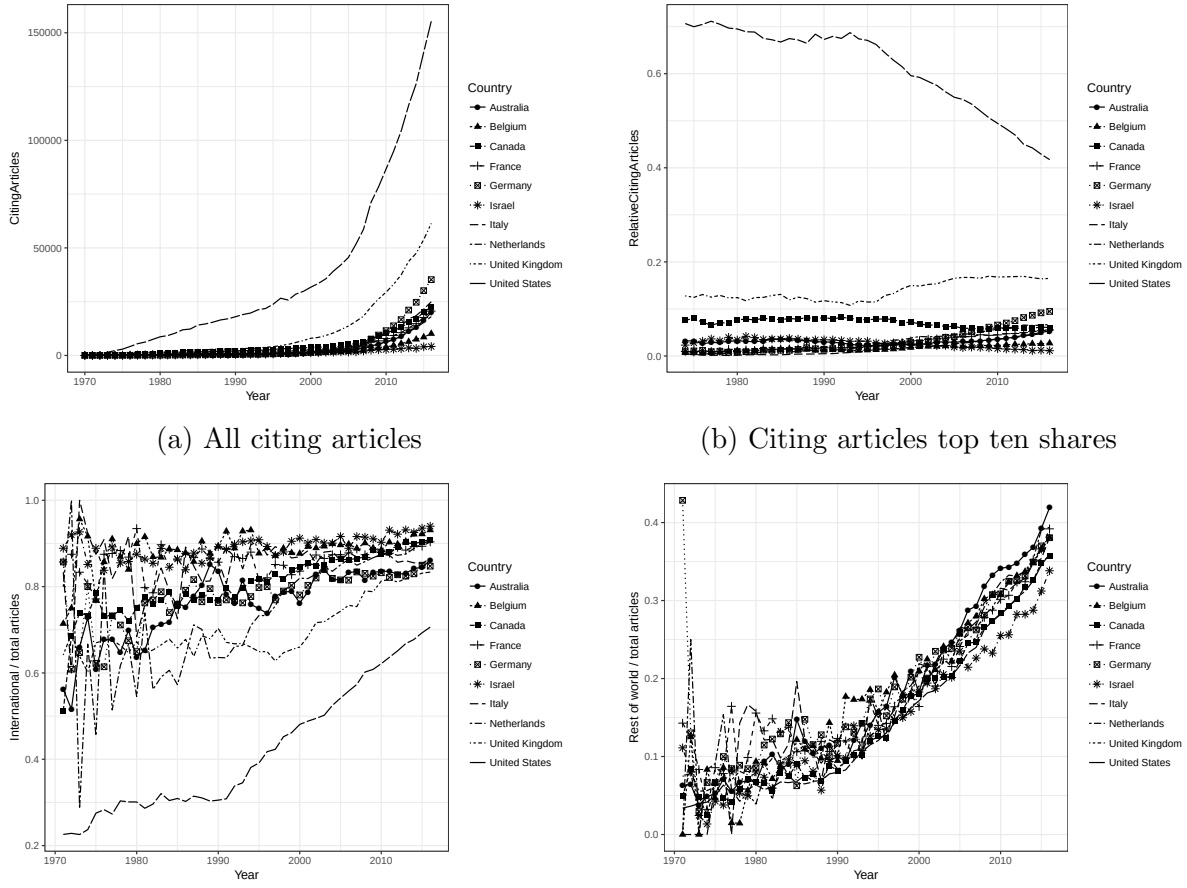
Publication Country	Citing Country	Publication Year	Citing Year	Citing Articles	Citations
UK	UK	1993	1995	1	1
UK	USA	1993	1995	2	2
UK	Canada	1993	1995	0	0
USA	UK	1993	1995	1	2
USA	USA	1993	1995	2	3
USA	Canada	1993	1995	1	1
Canada	UK	1993	1995	0	0
Canada	USA	1993	1995	1	1
Canada	Canada	1993	1995	0	0

Figure A.2: Article output for selected countries



Note: Countries as indexed by WoS and between 1970 and 2016. "All" is the sum articles in top 255 journals that are indexed by Scopus in every year between 1996 and 2014. *Source:* WoS and Scopus.

Figure A.3: Citing totals for the top ten countries over time



(a) All citing articles

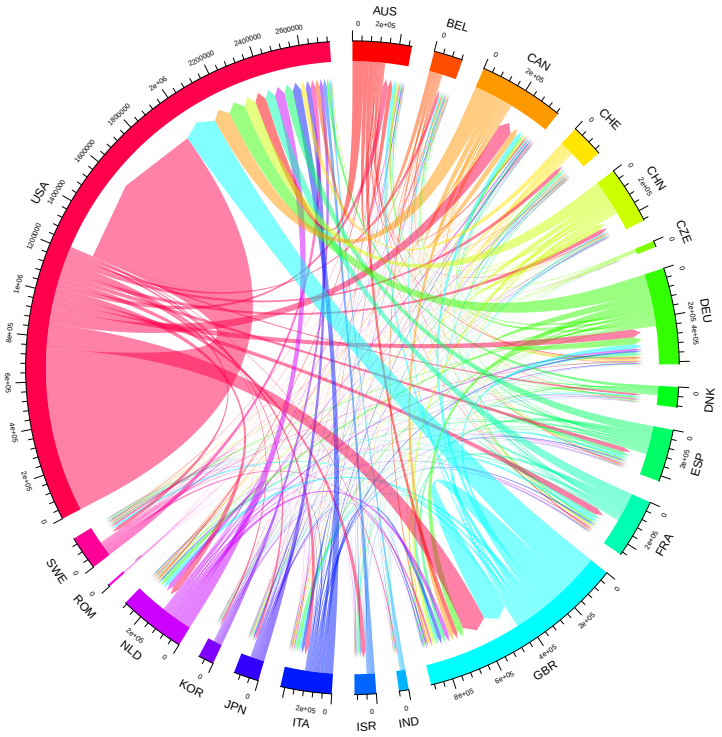
(b) Citing articles top ten shares

(c) International citing articles to total

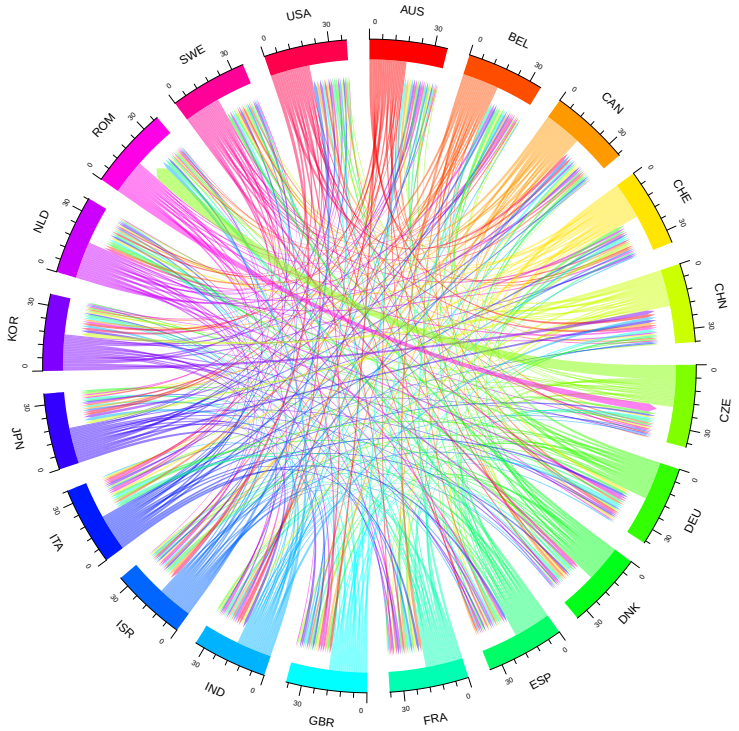
(d) Citing articles from rest of world to total

Notes: Clockwise, from top left: Citing articles; citing articles relative to the yearly sum of citing articles to top ten countries; citing articles from other countries relative to total of country; and citing articles from outside the top ten countries relative to total of country.

Figure A.3: Citations between top twenty countries



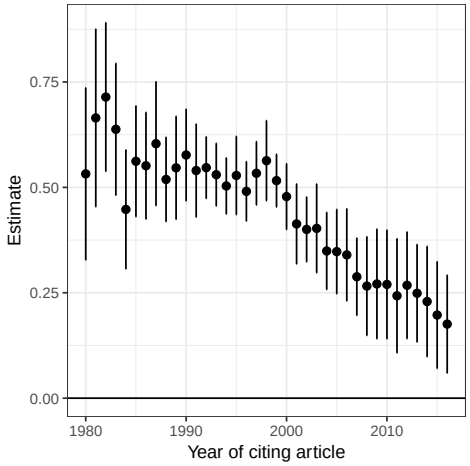
(a) Total citations 1970-2016



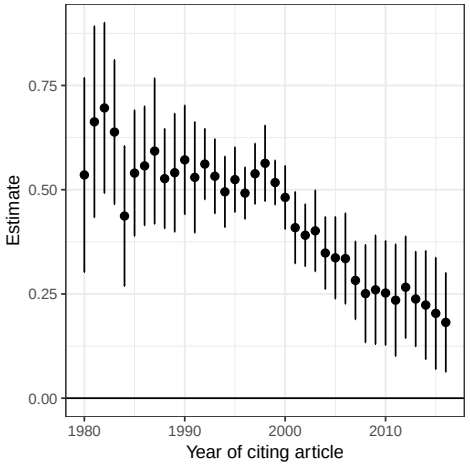
(b) Relative citations without self-citations

Notes: Citing articles between 1970 and 2016. The lower plot shows the exponential of the demeaned (citing and cited country) natural logarithm of citing articles. *Source:* Own calculations based on WoS data

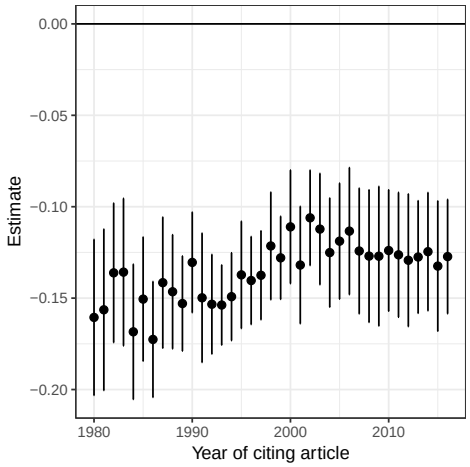
Figure A.4: Distance and language estimates by citing year



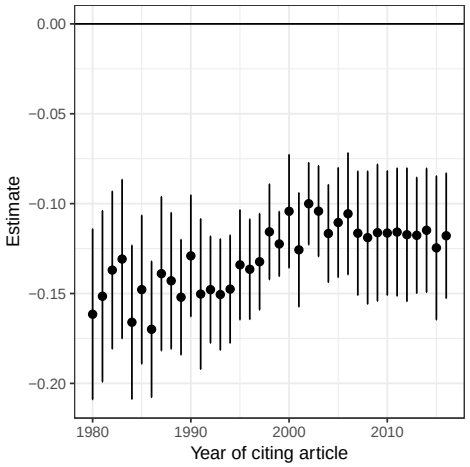
(a) Home bias



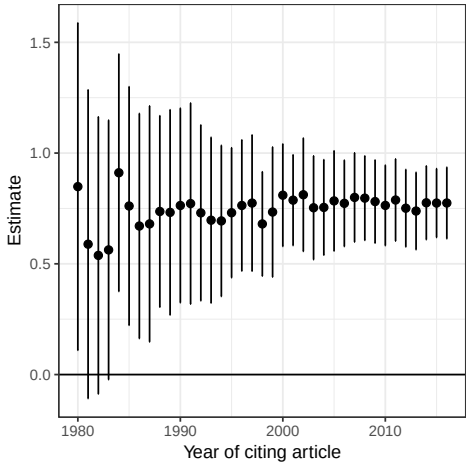
(b) Home bias without USA



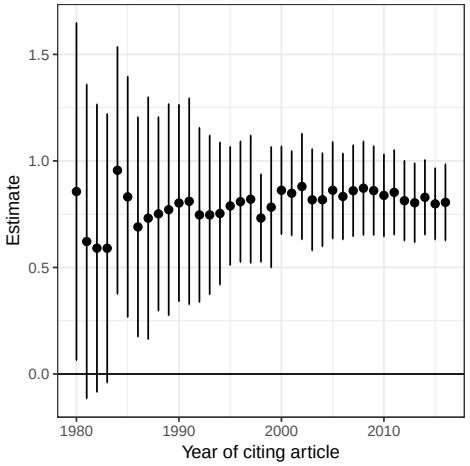
(c) Geographic distance



(d) Geographic distance without USA



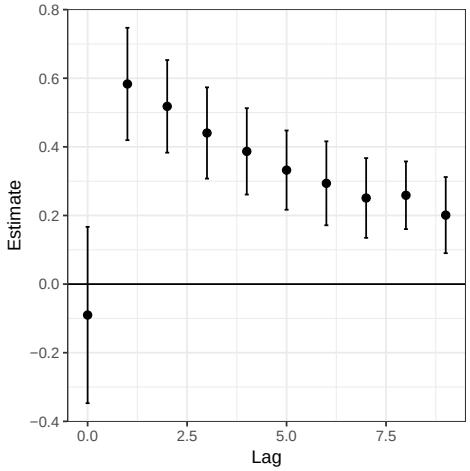
(e) Language similarity



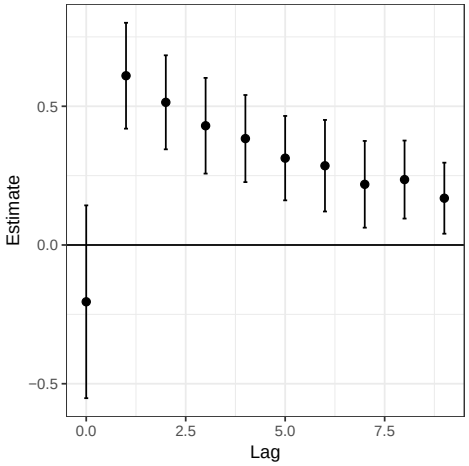
(f) Language similarity without USA

Notes: Estimated effects and 95% confidence intervals from repeated quasi-Poisson regressions including the three shown explanatory variables and otherwise identical to the Regression shown in Table B.4, Column (2). The regressions are repeated for each citing year from 1980 to 2016.

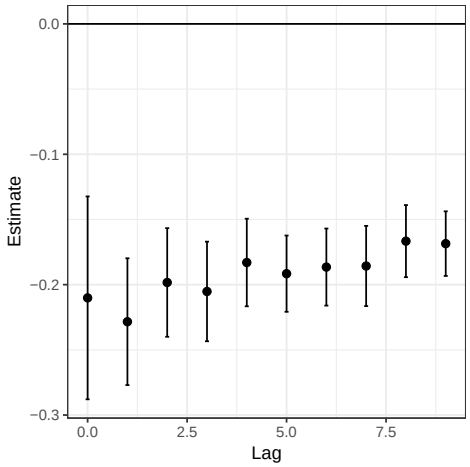
Figure A.5: Distance and language over time between citing and cited article



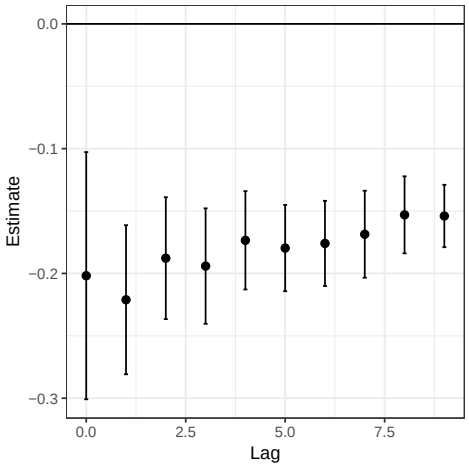
(a) Home bias



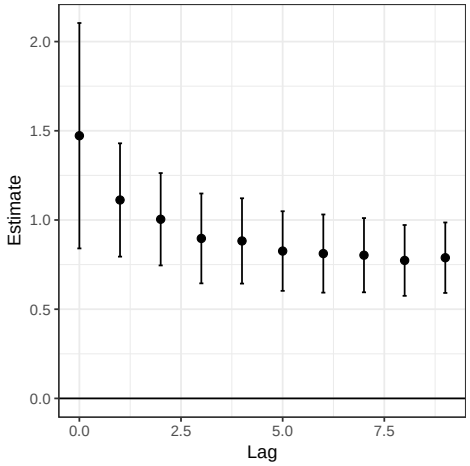
(b) Home bias without USA



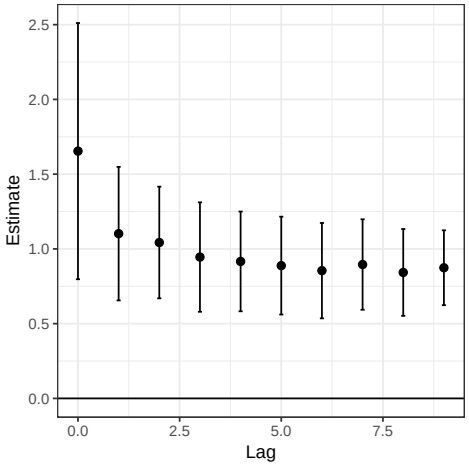
(c) Geographic distance



(d) Geographic distance without USA



(e) Language similarity



(f) Language similarity without USA

Notes: Estimated effects and 95% confidence intervals from repeated quasi-Poisson regressions including the three shown explanatory variables and otherwise identical to the Regression shown in Table B.4, Column (2). The regressions are repeated for per lag between citing and cited year ranging from 0 to 9 years.

B. Tables

Table B.1: Top 25 citation pairs

Citing country	Publication country	Citing articles	Distance	Language	Links09
United States	United States	746,197	1,161	1.00	
United Kingdom	United States	158,753	5,570	0.79	31,314,578
United Kingdom	United Kingdom	115,325	186	1.00	
United States	United Kingdom	114,959	5,570	0.79	48,878,700
Canada	United States	80,940	548	0.55	12,911,070
Germany	United States	80,555	6,035	0.34	20,119,072
United States	Canada	66,085	548	0.55	10,713,620
China	United States	56,843	10,994	0.06	34,859,492
Australia	United States	52,776	16,009	0.64	6,245,988
France	United States	47,448	5,838	0.19	20,979,438
Netherlands	United States	47,382	5,866	0.38	4,170,676
Spain	United States	41,599	5,770	0.25	14,314,497
Italy	United States	41,037	6,895	0.23	11,088,147
United States	Germany	38,198	6,035	0.34	40,771,812
Germany	Germany	34,565	225	1.00	
Germany	United Kingdom	32,044	495	0.38	20,831,130
United States	France	30,211	5,838	0.19	18,235,724
United States	Netherlands	30,085	5,866	0.38	10,972,904
Canada	Canada	28,381	1,188	1.00	
United States	Australia	24,645	16,009	0.64	13,225,842
Switzerland	United States	24,250	6,272	0.31	3,127,408
Japan	United States	24,226	10,856	0.07	34,111,636
Sweden	United States	23,000	6,323	0.41	4,287,421
United States	Israel	22,302	9,120	0.23	1,407,114
Netherlands	Netherlands	22,145	77	1.00	

Table B.2: Summary statistics for aggregated data

Statistic	N	Mean	St. Dev.	Min	Max
Home	2,640	0.008	0.087	0	1
LogDistance	2,400	8.511	0.934	4.013	9.885
Contiguity	2,400	0.025	0.155	0	1
TimeDifference	2,500	4.216	3.369	0.000	12.000
Colony	2,400	0.038	0.192	0	1
CommonLegal	2,500	0.292	0.455	0	1
CommonReligion	2,500	0.125	0.196	0.000	0.943
EU	2,640	0.108	0.311	0	1
LanguageSimilarity	2,400	0.148	0.168	0.000	1.000
AverageEnglishSimilarity	1,109	0.914	0.059	0.701	0.995
LogHyperLinks09	1,580	11.620	2.612	4.290	17.898
Citations	2,640	1,474.167	15,815.570	0	746,197

Table B.3: Correlation of distance measures demeaned

	Home	Dist	Cont	Time	Col	Rel	Law	Lang	Engl	L09
Home		-0.39	-0.05	-0.16	-0.02	0.12	0.16	0.54		
Dist	-0.39		-0.36	0.73	0.01	-0.24	-0.14	-0.44	-0.29	-0.20
Cont	-0.05	-0.36		-0.19	0.04	0.15	0.07	0.18	0.08	0.26
Time	-0.16	0.73	-0.19		-0.01	-0.17	-0.06	-0.23	-0.12	-0.16
Col	-0.02	0.01	0.04	-0.01		0.14	0.24	0.23	0.08	0.18
Rel	0.12	-0.24	0.15	-0.17	0.14		0.18	0.37	0.24	0.18
Law	0.16	-0.14	0.07	-0.06	0.24	0.18		0.34	0.07	0.07
Lang	0.54	-0.44	0.18	-0.23	0.23	0.37	0.34		0.29	0.15
Engl		-0.29	0.08	-0.12	0.08	0.24	0.07	0.29		0.18
L09		-0.20	0.26	-0.16	0.18	0.18	0.07	0.15	0.18	

Notes: The table shows Pearson correlations using pairwise complete observations. The variables are demeaned by citing and cited country using all for the variable available observations (which is more than complete variable pairs). Dist is LogDistance, Cont is Contiguity, Time is TimeDifference, Col is Colony, Rel is CommonReligion, Law is CommonLegal, Lang is LanguageSimilarity, Engl is EnglishSimilarity, and L09 is LogHyperLinks09.

Table B.4: Baseline distance measures and aggregated citation flows

	Dependent variable: Citing articles						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Home	0.52**	1.35***	1.05***	1.54***	1.83***	0.56***	0.83***
	(0.18)	(0.27)	(0.21)	(0.09)	(0.27)	(0.15)	(0.12)
LogDistance	−0.23***	−0.22**	−0.23**	−0.10*	−0.17	−0.18***	−0.19***
	(0.03)	(0.08)	(0.09)	(0.04)	(0.08)	(0.03)	(0.03)
TimeDifference	0.03*	−0.01	−0.01	0.02*	−0.03	0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Contiguity	−0.08***	0.24	0.15*	0.23*	0.11	−0.04	−0.03
	(0.02)	(0.15)	(0.06)	(0.12)	(0.07)	(0.02)	
Colony	0.26***	0.23***	0.22***	0.28**	0.44***	0.26***	0.41***
	(0.07)	(0.05)	(0.04)	(0.09)	(0.07)	(0.05)	(0.06)
CommonLegal	0.14**	0.12***	0.12***	0.02	0.21**	0.14**	0.03
	(0.05)	(0.03)	(0.03)	(0.04)	(0.07)	(0.05)	(0.03)
CommonReligion	0.32	0.19	0.22**	0.06	0.19	0.28	0.27
	(0.23)	(0.11)	(0.08)	(0.12)	(0.12)	(0.20)	(0.14)
EU	0.16**	0.17*	0.10	−0.01	0.05	0.11	0.22*
	(0.06)	(0.08)	(0.12)	(0.11)	(0.10)	(0.06)	(0.09)
asinh(CitedCountryAgg)				0.08**			
				(0.03)			
asinh(CitingCountryAgg)				0.18***			
				(0.03)			
log(pop_cited)				−0.01			
				(0.03)			
log(pop_citing)				−0.02			
				(0.02)			
log(gdp_cited)				0.09			
				(0.06)			
log(gdp_citing)				0.07***			
				(0.02)			
Estimation	Q-P	NB	OLS	OLS	Logit	Q-P	Q-P
Sample	agg.	agg.	agg.	full	full	> 0	no US
N	2381	2381	2381	872266	1011925	184988	2243
CitedCountry	20	20	20		20	20	19
CitingCountry	120	120	120		120	120	119
CitedYear FE				47	47	47	
CitingYear FE				47	47	47	
R^2 (full model)			0.96	0.50			
R^2 (proj model)			0.20	0.43			

Notes: This table reports estimated coefficients and cluster robust standard errors in parentheses (H3). Standard errors are clustered at the citing and at the cited country level. The dependent variable is citing articles. In models 1,2,6, and 7, we use the natural logarithm as link function and in model 5 we use a logit link for $P(y > 0)$. Models 3 and 4 are linear but the dependent variable is transformed using the inverse hyperbolic sine (asinh). $Asinh(x) = ln(x+\sqrt{x^2+1})$. For $x \geq 2, asinh(x) \approx ln(x)+ln(2)$, but $asinh(0) = 0$. The explanatory variables are distance between capitals (in natural log form), geographic contiguity between the countries, time zone difference, similarity of spoken languages and English proficiency, common membership of the European Union, and common legal origins. Specifications 1,2,3, and 7 estimate citations flows between 1970 and 2016 aggregated by country pair. Specifications 4 and 5 use all available year and country pairs without aggregating. Specification 6 uses citations larger than 0 only.

*p<0.1; **p<0.05; ***p<0.01

Table B.5: Evolution of distance over time

	Dependent variable (link:log): Citing articles			
	(1)	(2)	(3)	(4)
Home	1.78***	1.15***	0.58**	0.33
	(0.34)	(0.18)	(0.18)	(0.20)
LogDistance	−0.32***	−0.32***	−0.27***	−0.24***
	(0.05)	(0.04)	(0.03)	(0.03)
TimeDifference	0.09***	0.07***	0.04***	0.03**
	(0.01)	(0.01)	(0.01)	(0.01)
Contiguity	0.07	−0.08	−0.13**	−0.13***
	(0.14)	(0.10)	(0.05)	(0.02)
Colony	0.11	0.14	0.26***	0.29***
	(0.10)	(0.08)	(0.07)	(0.08)
CommonLegal	0.41***	0.26***	0.13***	0.06
	(0.05)	(0.04)	(0.03)	(0.04)
CommonReligion	−0.78*	−0.21	0.26	0.41
	(0.37)	(0.19)	(0.22)	(0.25)
EU	0.00	0.06	0.14*	0.13*
	(0.09)	(0.08)	(0.05)	(0.06)
CitingYears	1979-1988	1989-1998	1999-2008	2009-2016
N	1673	2261	2381	2361
CitedCountry FE	19	20	20	20
CitingCountry FE	89	114	120	119

Notes: This table reports estimated coefficients from quasi-Poisson with cluster robust standard errors in parentheses (H3). Standard errors are clustered at the citing and at the cited country level. All models include citing and cited country dummies. The dependent variable is citing articles from different decades. The time frames are indicated above. The explanatory variables are distance between capitals (in natural log form), geographic contiguity between the countries, time zone difference, similarity of spoken languages and English proficiency, common membership of the European Union, and common legal origins.

*p<0.1; **p<0.05; ***p<0.01

Table B.6: Language and English proficiency

	Dependent variable (link:log): Citing articles					
	(1)	(2)	(3)	(4)	(5)	(6)
Home		0.755***	0.462***	1.162***	0.777***	0.761***
		(0.069)	(0.060)	(0.109)	(0.101)	(0.094)
LogDistance	-0.248***	-0.225***	-0.204***	-0.107***	-0.082**	-0.205***
	(0.030)	(0.024)	(0.024)	(0.036)	(0.033)	(0.025)
TimeDifference	0.013	0.024**	0.024**	-0.049***	-0.040***	0.014*
	(0.008)	(0.010)	(0.010)	(0.012)	(0.012)	(0.008))
Contiguity	-0.288***	-0.070**	-0.093***	0.116***	0.049	-0.075**
	(0.047)	(0.029)	(0.032)	(0.041)	(0.033)	(0.031))
Colony	0.031	0.263***	0.143***	-0.093	-0.064	0.301***
	(0.045)	(0.058)	(0.045)	(0.102)	(0.071)	(0.063)
CommonLegal	-0.009	0.113***	0.008	0.058**	0.041***	0.072**
	(0.030)	(0.033)	(0.030)	(0.026)	(0.015)	(0.031)
CommonReligion	0.203***	0.512***	0.280***	0.233***	0.120	0.637***
	(0.069)	(0.074)	(0.073)	(0.086)	(0.098)	(0.091)
EU	0.051	0.169***	0.187***	0.151*	0.162**	0.138*
	(0.050)	(0.057)	(0.051)	(0.085)	(0.064)	(0.073)
LanguageSimilarity	1.004***		0.665***		0.652***	
	(0.108)		(0.122)		(0.165)	
EnglishSimilarity				0.707**	0.239	-0.017
				(0.335)	(0.383)	(0.058)
N	2380	2380	2380	1049	1049	1385
CitedCountry	20	20	20	16	16	20
CitingCountry	120	120	120	72	72	80
Countries	all	all	all	no Engl	no Engl	all

Notes: This table reports estimated coefficients from quasi-Poisson with cluster robust standard errors in parentheses (H3). Standard errors are clustered at the citing and at the cited country level. All models include citing and cited country dummies. The dependent variable is citing articles. The explanatory variables are distance between capitals (in natural log form), geographic contiguity between the countries, time zone difference, similarity of spoken languages and English proficiency, common membership of the European Union, and common legal origins. Columns (5) and (6) do not include English language countries.

*p<0.1; **p<0.05; ***p<0.01

Table B.7: Internet Hyperlinks

	Dependent variable (link:log): Citing articles					
	(1)	(2)	(3)	(4)	(5)	(6)
LogDistance	−0.04	−0.03	−0.09**	−0.08*	−0.04**	−0.03*
	(0.02)	(0.02)	(0.03)	(0.03)	(0.01)	(0.01)
TimeDifference	−0.00	−0.00	−0.01	−0.01	−0.01**	−0.01*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Contiguity	0.16***	0.16**	−0.01	−0.03	0.00	−0.00
	(0.05)	(0.05)	(0.05)	(0.05)	(0.02)	(0.02)
Colony	0.17***	0.17***	0.17***	0.13**	0.09***	0.08***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.02)	(0.03)
CommonLegal	−0.02	−0.04	0.04*	0.02	0.04*	0.02
	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
CommonReligion	0.26	0.24	0.39***	0.34***	0.29***	0.26***
	(0.14)	(0.13)	(0.11)	(0.10)	(0.08)	(0.07)
EU	0.34***	0.32***	0.19***	0.15**	0.11**	0.10**
	(0.09)	(0.09)	(0.05)	(0.05)	(0.04)	(0.03)
LogHyperLinks03		0.04*				0.05*
		(0.02)				(0.02)
LogHyperLinks09				0.12***		
				(0.03)		
LogCitations03					0.29***	0.28***
					(0.03)	(0.03)
LogHyperLinks09m03						−0.04
						(0.05)
Citing articles in	2003	2003	2009	2009	after 2008	after 2008
N	861	861	1521	1521	821	821
CitedCountry FE	20	20	20	20	20	20
CitingCountry FE	44	44	77	77	42	42

Notes: This table reports estimated coefficients from quasi-Poisson with cluster robust standard errors in parentheses (H3). Standard errors are clustered at the citing and at the cited country level. All models include citing and cited country dummies. The dependent variable is citing articles. The explanatory variables are distance between capitals (in natural log form), geographic contiguity between the countries, time zone difference, similarity of spoken languages and English proficiency, common membership of the European Union, and common legal origins. In addition, these estimates use counts of internet hyperlinks in two years: 2003 and 2009. The sample of citing articles is restricted to the respective year of hyperlink counts. The last two models include the count of citing articles in 2003 to explain the citations after 2008. In addition, model 6 includes hyperlinks in 2003 and the growth of hyperlinks in the form $log(Links_{2009} - Links_{2003})$.

*p<0.1; **p<0.05; ***p<0.01

Table B.8: Econometric citations

	Dependent variable (link:log): Citing articles					
	Econometrics				Economics	
	(1)	(2)	(3)	(4)	(5)	(6)
Home	0.364***	0.270**	0.370***	0.282**	0.368***	0.353***
	(0.090)	(0.114)	(0.090)	(0.115)	(0.078)	(0.086)
LogDistance	-0.044	-0.043	-0.051	-0.050	-0.149***	-0.147***
	(0.032)	(0.031)	(0.032)	(0.032)	(0.028)	(0.029)
TimeDifference	0.002	-0.002	0.004	0.000	0.014	0.013
	(0.014)	(0.015)	(0.014)	(0.015)	(0.009)	(0.009)
Contiguity	-0.024	-0.035	-0.032	-0.043	-0.099**	-0.098**
	(0.075)	(0.078)	(0.078)	(0.081)	(0.043)	(0.044)
Colony	0.078	0.106*	0.083	0.109*	0.126***	0.129***
	(0.050)	(0.056)	(0.052)	(0.058)	(0.043)	(0.041)
CommonLegal	-0.104***	-0.101***	-0.102***	-0.099***	-0.059	-0.063*
	(0.032)	(0.030)	(0.031)	(0.030)	(0.037)	(0.036)
CommonReligion	0.514***	0.574***	0.499***	0.557***	0.421***	0.429***
	(0.165)	(0.167)	(0.169)	(0.172)	(0.084)	(0.090)
EU	0.228**	0.284**	0.223**	0.276**	0.174**	0.182**
	(0.103)	(0.118)	(0.102)	(0.120)	(0.079)	(0.076)
LanguageSimilarity	0.684***	0.723***	0.671***	0.705***	0.584***	0.599***
	(0.140)	(0.137)	(0.148)	(0.148)	(0.134)	(0.123)
LogHyperLinks		-0.116		-0.109		-0.013
		(0.071)		(0.073)		(0.050)
N (df)	510 (441)	470 (404)	459 (391)	423 (358)	459 (391)	423 (358)
CitedCountry FE	10	10	9	9	9	9
CitingCountry FE	51	47	51	47	51	47

Notes: This table reports estimated coefficients from quasi-Poisson with cluster robust standard errors in parentheses (H3). Standard errors are clustered at the citing country level. All models include citing and cited country dummies. The dependent variable is citing articles. These are aggregates of citing articles up to ten years after publication in any single country to articles written by authors in the ten leading econometric research countries in a year between 2004 and 2008. The explanatory variables are distance between capitals (in natural log form), geographic contiguity between the countries, time zone difference, similarity of spoken languages and English proficiency, common membership of the European Union, and common legal origins. Columns (1) and (2) present estimates for citations to the then leading econometric research countries. The other four columns exclude Russia which is in the econometrics sample but not in the main economics sample.

*p<0.1; **p<0.05; ***p<0.01