Essays in Applied Microeconomics

Research collaboration and dissemination in economics

Lukas Valentin Kuld

Supervised by

Professor John O'Hagan

A THESIS SUBMITTED TO THE UNIVERSITY OF DUBLIN, TRINITY COLLEGE

IN APPLICATION FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

2017

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

Chapter 2 and 3 are based on joint work. Details are given in the acknowledgements.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Lukas Kuld

Summary

This dissertation studies the production and dissemination of economic research in three essays. The chapters explore data on economic research articles and their authors to study, first, the effect of national borders on the international dissemination of economic research along cultural and technological links, second, the rise of collaborations between economists, and, third, the benefit of local research clusters.

Chapter 1 provides the introduction, along with a general background to this research, and then outlines the specific research questions that are explored in this dissertation.

Chapter 2 examines the existence of a border effect in the dissemination of knowledge in economics using a gravity model framework and novel data on domestic and international citation flows in economic research between 1970 and 2016. We link citation data to geographic and cultural distance measures, as well as novel indicators for virtual proximity and English proficiency. Our results show that (i) citation patterns follow the law of gravity; (ii) citations in economics exhibit a strong and significant home bias by an overall factor of 1.9 for all ten leading countries (a 90 percent higher propensity to cite domestic articles); (iii) bilaterally low levels of English proficiency are associated with a cost of up to 30 percent, while similarity in English proficiency is insignificant for the total sample; (iv) countries with closer early internet ties have significantly higher shares of bilateral citations of up to .25 percent for a 1 percent increase in internet hyperlinks. Over time, the estimated impact of home bias as well as geographic and cultural distance declines but remains significant.

Chapter 3 presents new evidence on several features of co-authorship in articles in economic journals. It builds on previous work by reviewing the key literature relating to the rise in co-authorship in economics. The empirical analysis draws on around 175,000 articles in the top 255 journals, over the period 1996 to 2014. The rises in quarto-plus and cross-country co-authored papers are striking, as are the differences in citations per article and citations per author. There is evidence of an alphabetical ordering of authors as the standard in co-authored papers in top journals with no downward trend evident over

time. A correlation between co-authorship and career stage is observed with young authors publishing significantly more solo-authored articles.

Chapter 4 studies benefits from local research linkages. Recent research observes a declining or no return from the quality of local colleagues on a researcher's productivity but positive spill-overs between co-authors. The findings of this chapter show benefits from research linkages between department colleagues below the level of co-authorship. Using data from the CVs of around 1,000 highly-cited economists, this chapter tests whether take-up of a research article by future research is increased if it links with the research of local colleagues. The estimates show that "high-profile" research that references articles by local colleagues receives significantly more citations than comparable work by the same authors. There is no dissemination benefit for less prominently published "routine-type" research or for colleagues without a thematic link.

Chapter 5 concludes the dissertation.

Acknowledgements

This research was funded by a Government of Ireland Postgraduate Scholarship from the Irish Research Council, and by teaching and research fellowships from the Department of Economics, Trinity College Dublin.

I am grateful to Professor John O'Hagan for the great support throughout the past four years. John was a great help in initiating research ideas and turning ideas into articles. Through untiring reading and editing of various drafts, organising collaborations and conferences, and pushing forward our joint article, this text profited greatly from John's supervision. He also added a wonderful personal site to my Trinity experience, including but not limited to very enjoyable invites to the Rubrics and Sandycove.

John made it possible together with the Department of Economics at Trinity College to collect detailed biographical data on around 1,5000 economists from 1700 until today. These data were collected by Kate Hayes, Greg Mangan, Andrew Morrow, Sarah Mortell, Conor Parle and Michael Stone during summer internships between 2014 and 2016. Parts of these data are used in Chapter 4.

I thank Professor Christiane Hellmanzik for the warm welcome in Dortmund and the chance for more collaboration.

I am thankful for the help and discussions from the members of the PhD working group, all three papers benefited from this exposure. I also want to thank Colette Ding for the great help in organising college life over the last four years.

Chapter 2 is based on joint work with Christiane Hellmanzik. Martin Schmitz, Danielle Kedan, Rogelio Mercado and Tara Bedi made helpful comments. Moreover, I thank Jan Lordick and Markus Friemann for their excellent research assistance.

Chapter 3 is based on a joint paper with John O'Hagan and also benefited from discussions with Tara Bedi and Rogelio Mercado. Two anonymous referees made helpful comments.

Chapter 4 benefited greatly from the comments by Tara Bedi, Agustin Benetrix, Christiane Hellmanzik, Martina Kirchberger, David Lagakos, Clemens Struck and Alan Walsh, and conference participants in Bilbao, Vienna, Valladolid, Dortmund, and Le Mans. The diligence and hard work of the summer interns named above made this paper possible.

Thank you Tara and Sonja for opening up your lives and enriching mine, showing me so much new and admirable. Thank you Rogelio, Michele, and Alan for great discussions and company, and painstaking proof reading. Thank you to all friends too far scattered around but nevertheless important, hopefully the next years allow more time together. And thank you Noemi, Judith, Mama, Papa, Levi and Juri, for always being there and supporting me.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Academic economic research has changed remarkably over the last 30 years. The English language journal article has become the dominant unit of economic research. Today, the global economic literature published in English includes prominently research from non-English language countries. In particular, the share of articles from continental European countries but also China has risen substantially. With the rise of the Internet and computers, the dissemination of economic knowledge and research knowledge has greatly simplified. In 1990, 70 percent of the articles that referenced American economic research were authored in the USA. Today, this share lies at 30 percent. Chapter 2 traces the international dissemination of economic research articles along cultural and technological ties.

With the change in the means of production, the organisation of economic research has changed, too. Where the majority of publications in the mid-90s was still single-authored, co-authorship has become the norm. This rise in formal collaboration is discussed in Chapter 3. Below the level of co-authorship, Chapter 4 studies potential benefits from local research linkages. The take-up of high-level research articles by future research is shown to be substantially higher if it references the research of a local colleague.

In Chapter 2, the dissemination of research articles is shown to follow a law of gravity. 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 home bias in quoting the works of other researchers. Even though any economic article is available worldwide, it is still more likely to be cited by a domestic researcher or a colleague from a country with strong cultural and linguistic ties. There are good reasons to assume frictions in the dissemination of economic research. Research interests might be local, the majority of collaboration networks stay within geographic borders, and schools

of economic thought vary from country to country. In this paper, we investigate the role of borders, as well as geographical, cultural, virtual and linguistic distances between countries for citation patterns in economics. We use domestic as well as bilateral information on citing articles between 1970 and 2016 to investigate what hinders a country's economic research output from travelling the globe.

While there is an expansive literature on co-authorship and citations in the sciences, little attention has been given to spatial biases in citations, other than noting differences in national performance (See Frenken et al. (2009) and Frenken and Hoekman (2014)). Our results show that national borders, as well as geographic, cultural, and linguistic distances impact negatively on the dissemination of research knowledge. However, we observe that the importance of distances and borders decreases over time and that countries with strong Internet links also cite each other's economic literature more often.

Chapter 3, first, summarises the key hypotheses on the rise of co-authorship brought forward in the economic literature of the last 30 years. Broadly, these hypotheses relate, first, to an increase in research specialisation and the resulting need to combine research knowledge; second, increased ease of co-authorship due to falling communication costs; and, third, changing incentives set by the research community to encourage co-authorship.

The rise of multi-authorship is then shown for a broad sample of 255 economics journals and, separately, for a more narrow sample of the twenty most cited economics journals. Against expectation, we show that the share of articles that list their authors alphabetically also increased slightly. Scientific fields with high average numbers of authors typically express the contribution of authors by name order, with first and last author contributing most.[1] An alphabetical listing of author names signals equal contribution of all authors. We take this as a sign that there is no trend in economic research towards more vertical research teams. Subsequently, form and take-up of articles are compared by number of authors. Co-authored articles receive substantially more citations, increasing for each additional co-author up to the highest category of four and more authors. In addition, co-authored papers are slightly longer and reference more articles. These differences are robust to controlling for journal and year means. Finally, we note a high share of single-authorship in the first career years. This is counter-intuitive if co-authorship is seen foremost as a response to the need to combine research knowledge.

While we observe these trends towards a more globally integrated research environment in Chapter 2 and 3, universities and national science federations design policies to create local clusters of economic research. For instance, to take a local example, the Department of Economics at Trinity College markets on its website the research of 12 out of 15 named staff

[1] See, for instance, Baerlocher et al. (2007) for contribution and order in medical research).

members along two strains, international macroeconomics (IM-TCD) and international development (TIME), both recent creations.[2]

The rise in international collaboration and the decreasing importance of borders and distances for the dissemination of economic research seen in Chapter 2 and 3 could suggest a decreasing importance of location in the production of economic research. Indeed, recent research has found the demise of benefits a researcher takes from their affiliation (Kim et al. (2009) and Bolli and Schläpfer (2015)).[3] If clusters do not improve the research performance of its members, the main remaining benefit would be in attracting new hires to an environment of like-minded researchers (Agrawal et al. (2014) and Waldinger (2016)). Though, there are negative implications by a strong specialisation, too. Few staff members are teaching in the area of their research, more diverse research could lead to cross fertilisation, and the wider interest in the department's research might be more volatile.

Chapter 4 addresses this question of local benefits by narrowing the focus of analysis. Instead of testing for department wide effects, the benefit of working on research questions that have links to the research of local colleagues is shown. The take-up by future research of an article that references the research of a department colleagues is substantially higher than the take-up of comparable work. However, this benefit is limited to 'difficult' work and no general dissemination benefit is observed for 'routine-type' work. The findings indicate that an increase in specialisation and decrease in collaboration within departments are further possible explanations for the elsewhere observed decline in local effects.

[2] See `https://www.tcd.ie/Economics/research/`, retrieved on September 25th, 2017.

[3] Focusing just on peer effects, recent published economic research suggest that these local peer effects never mattered, based on samples including physicists in Nazi Germany and mathematicians at the end of the Soviet Union (Waldinger (2012) and Borjas and Doran (2015)). Peer effects are, though, just one possible positive effect of being associated with a cluster or university, besides, for example, advantages in disseminating research, the sharing of infrastructure, and a social environment.

Chapter 2

No place like home: border effects in the dissemination of economic research articles

2.1 Introduction

Economic research articles do not travel easily. Where 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 home bias in quoting the works of other researchers. Even though any economic article is available worldwide, it is still more likely to be cited by a domestic researcher or a colleague from a country with strong cultural and linguistic ties. There are good reasons for such frictions in the dissemination of economic research; scientific interests might be local, the majority of collaboration networks stay within geographic borders, and schools of economic thought vary from country to country. In this paper, we investigate the role of borders, as well as geographical, cultural, virtual and linguistic distances between countries for citation patterns in economics. We use domestic as well as bilateral information on citing articles between 1970 and 2016 from Web of Science to investigate what hinders a country's economic research output from travelling the globe.

While there is an expansive literature on international collaboration networks and citations in the sciences, little attention has been given to spatial biases in citations, other than noting differences in national performance (see Frenken et al. (2009) and Frenken and

Hoekman (2014)). Pan et al. (2012) are the closest to our effort by linking citations to geographic distance in a simple gravity model. Nonetheless, the nature of citation data lends itself well to the application of a gravity model framework. This comes with the main advantage that we have a theoretical basis on which we can control for differences in quality and quantity of research output by employing two sets of fixed effects for citing and cited country. This double set of fixed effects accounts for omitted domestic influences such as the quality of articles. Using fixed effects and Poisson regressions to estimate the impact of distance has the added benefit of being consistent with estimating the effects of multilateral resistances (Fally (2015)).[1] Although this restricts our investigation to the inclusion of bilateral factors only, we will be able to draw conclusions on the existence of a home bias in citations as well as the role of common language and internet linkages. All of these factors are easily expressed as differences or similarities between and within countries.

Although one could argue that scientific research is not prone to any biases, it is reasonable to assume that citations in economics are subject to transactions costs, such as information costs, linguistic barriers or copyright restrictions, as well as cultural biases which increase with distance. In fact such costs are comparable to such findings as by McCallum (1995) and Anderson and van Wincoop (2003) who show that geographic as well as economic distance matter for international trade when controlling for other relevant factors. 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 comparable to patterns we observe in goods and services trade as well as international migration and capital flows (see for example, Felbermayr and Toubal (2010) or Kimura and Lee (2006), Francois and Hoekman (2010)). Closer to the consumption of scientific research, Blum and Goldfarb (2006) find that the gravity model also holds for taste-prone products such as music and games for a sample of American internet users.

In this paper we use a novel dataset on bilateral citation flows to identify and quantify any home bias in the take-up of economic research articles. Abstracting from restrictions such as language and culture differences, the dissemination process of a research article should be uniform world wide. Factoring out these differences allows us to estimate the home bias as the remaining preference for domestic research. In addition, we also aim to quantify the impact of geographic, cultural, and linguistic distances. Finally, we investigate whether the use of English as the lingua franca and the connectedness brought by internet links helps to bridge these distances. It is plausible to imagine these two factors as the driving force behind the integration of economic research globally.

[1] That is estimating the effect of relative resistances that separate two countries instead of direct bilateral distances.

Our results show that citation patterns can be described well by the gravity model framework with the number of citing articles decreasing by cultural, geographic, and linguistic distance. National borders play a significant role in the dissemination of economic articles. All studied countries show a significant home bias ranging from 1.9 to 7.2 if aggregated over the complete time period. Moreover, linguistic similarity bears a significant impact on citations, while similarity in English proficiency between citing and cited country has no significant impact overall. We observe, however, a significant cost between 15 and 30 percent for country pairs involving two countries with low English proficiency. Finally, countries which have closer internet ties also quote each other's work significantly more often. This effect is strongest for our earliest hyperlinks data wave of 1998. Over the observed period, the effects ranges from 0.04% to 0.25% for a 1% increase in internet hyperlinks between countries.

The data for this analysis are based on Web of Science's (WoS) indexation which encompasses all articles worldwide that cite any economics article from the ten leading countries in terms of eminent economists between 1970 and 2016. These ten production countries' publications account for a large share of the total output of economic research articles and citations received.[2] The resulting database encompasses 10 source and 124 citing countries for a total of 527,800 country and year pairs and 1240 observations when aggregating citation flows up to 2017.[3] Most importantly for our analysis we have information on both domestic as well as international citations which allows us to assess any prevalent home bias in citations.

We combine this novel dataset on citation in economics with a state of the art set of gravity model variables which we extend by two factors particularly relevant to economics research. First, next to the linguistic similarity indicator by Melitz and Toubal (2014), we employ a novel indicator for English similarity between countries in order to assess the relative proficiency in English between citing and cited country (if other than home) using the EF English Proficiency Index (`http://www.ef.edu/epi/`). In line with Melitz and Ottaviano (2008) and Melitz and Toubal (2014) who find that countries of the same linguistic roots tend to be closer trading partners, our hypothesis is that country pairs which share an affinity to English are also linked more closely in terms of their research and therefore citations.

Second, we capture the importance of the internet to overcome frictions between countries in terms of citations by the indicator employed by Hellmanzik and Schmitz (2015) capturing

[2]Following the indexation on Web of Science, these countries account for 59% of all economics articles. The share of citations is most likely substantially higher.

[3]However, to take into account that most observations (76 percent) are zero we aggregate the data by country pair (1240 observations) and by country pair and publication year (58,280 observations). This results in less than one percent of the aggregated country pairs having no citation flow.

bilateral hyperlinks on websites as an indicator of countries' international virtual linkages.[4] The idea is to reflect, for instance, how often British or French internet sites set links to websites from the United States (say the homepage of the New York Times) and vice versa. This measure is a proxy for informations costs and hurdles to accessing economics articles from remote countries. Although star authors and their articles have always had global appeal the internet decreased transactions costs in obtaining such articles considerably while also making these works available to an ever increasing audience. This is largely in line with the literature assessing the economic impact on economic activity, such as Freund and Weinhold (2002) who find that internet development abroad has resulted in increased exports of services to the United States, Choi (2010) who reports that a doubling of internet usage in a country leads to a 2% to 4% increase in services trade. Thus, our hypothesis is that bilateral hyperlinks have a significant, positive impact on citation flows in economics and decrease any home bias.

Concerning the analysis of the spatial dimension of citations, Matthiessen et al. (2002) identify the 40 strongest publishing regions in the world in terms of publication output from 1997 to 1999 and find that both citation as well as collaboration relations occur most frequently domestically and that citations are much less affected by distance than collaboration. In the same vein, Börner et al. (2006) assess the distance decay of the 500 most cited research institutions in the United States between 1982 and 2001 statistically. Their results suggest that there is a distance-decay in citation relations between research organisations, articles from nearby research organisations are more likely to be cited than articles from research organisations further away. In addition, Frenken et al. (2009) and Pan et al. (2012) focus on sciences more generally and using a simple model the latter find that citations decrease significantly with distance. Our paper expands on these analyses by using a state-of-the-art gravity model for novel data.

Although there is a considerable empirical literature on labour market aspects of Economics, not much of it takes an international angle and seeks to analyse the dissemination of Economics as measured by citations domestically as well as across the globe. While Frey and Eichenberger (1993) see a distinct division between US and European economists which they largely attribute to the different market conditions, Lazear (2000) highlights features of economics as a science which make it universally applicable and translatable across all specialities in what he dubs 'economic imperialism'. In terms of networks in economics, some work has been done on co-authorships and networks, such as Fourcade et al. (2015) who highlight the tight networks amongst US economists, while Goyal et al. (2006) find that social distance between co-authors has decreased between 1970 and 2000.

[4] Hellmanzik and Schmitz (2017) find that bilateral hyperlinks have a positive impact on investments of advanced economies as well as a positive impact on audiovisual services (Hellmanzik and Schmitz (2015)) using the same internet link indicator that we use in our paper.

Furthermore, Catalini et al. (2016) find a positive impact of cheaper flights on scientific collaborations.

In the remainder of the paper we proceed as follows. In Section 2.2 we introduce the various data sources and summary statistics, Section 2.3 presents the empirical gravity model, while the empirical results are shown in Section 2.4. Section 2.5 concludes.

2.2 Data

2.2.1 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 active in one of the ten leading countries in economics research between 1970 and 2016. These ten countries were chosen based on the highest number of eminent economists active and therefore the highest productivity rate in terms of research output in the respective country and they account for 59% of the international output of economic research articles indexed by WoS.[5]

Figure 2.1 illustrates the data collection and format. All observations are uniquely identified by the publication year and the articles' source country according to the main authors' affiliation. Based on that we count citing articles by country for each consecutive year after publication. In this example, three articles are indexed as being authored in the USA, in the field of economics (excluding business), and in 1993. Among the citing articles published in 1995, one article is indexed as being authored by a British author, two articles have an author in the US, and one article is indexed as Canadian.

We do not count total citations but citations to country-year aggregates. This might underestimate the relative citations, in particular to the United States. To understand why, imagine picking a random economics article written in 2017; 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 2017. By collecting year to year citations, we, therfore, have a measure of citation intensity by the number of cited years. All other countries are considerably smaller in

[5]For more details on sampling see Kuld (2017). Therefore, our sample covers many more citations received than articles authored. Notable missing countries who fell shy of the data inclusion threshold include China, Japan, and Spain.

terms of citation propensity and the issue is also less important in early years when the number of references given in an article was considerably lower.[6] However, we might still underestimate the variation if we are interested in citation intensity as opposed to citing articles and we further account for this issue by excluding the United States in alternative specifications.

The resulting database from this data collection effort encompasses 124 citing countries, excluding countries with less than 100 total citing articles, and 10 cited countries in order to focus our analysis on the core of economic research activity. The data comprises 527,800 country and year pairs in total and including both domestic and international citations which makes it possible to estimate a home bias in economics. However, as 76 percent of all citation pairs in our data are zero we conduct our main analysis using data aggregated by country pair (1240 observations) and data aggregated by country pair and publication year (58,280 observations). This has the advantage that less than one percent of the aggregated country pairs have no citation flow.

Citation data - stylised facts

Figure 2.2 depicts the evolution of economics research over our sample period. We find a strong increase in articles published since the 19070s and most notably the diagram reflects the dominance of US based economists in publications. Furthermore, Figure 2.3 depicts the number of citing articles between 1970 and 2016 for the five largest receiving countries. Figures on the left show the sum of articles citing publications from the respective source 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 source country.[7]

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.

[6] The average number of references given per paper almost doubled over the last twenty years from 25 to 40 (see Kuld and O'Hagan (2017)).

[7] For comparison, Gloetzl and Aigner (2017) 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).

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.

Figures 2.4 and 2.5 illustrate the evolution of citations over the course of the article's life-cycle with the lowest line in Figure 2.4a showing counts of articles that cite articles published between 1970 and 1974 in the year of publication, the following year, and so on. We see that for most years, the sum of citing articles is stable over a ten-year citation window after publication. Strikingly, articles published from 1995 onward are cited by more articles in each year after publication than in the preceding year. However, this increase in citations in absolute terms does not prevail when adjusting citing articles with a given time lag to the cited article by the yearly sum of all citing articles as shown in Figure 2.4b. In terms of the geographic spread of citations after publication, Figure 2.5 depicts the ratio of domestic to international citing articles over time; we see that domestic citations are much more important in the first years of publication while the article is expanding its academic reach over time with increasing global academic recognition. This is indicative of an initial home bias followed by slow international dissemination, however, the effect seems to gradually lose in importance for more recent publications.

Figure 2.6a further analyse the bi-directionality of citation flows between the top 10 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 source 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.[8] 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. Similarly, Israel exhibits a particularly strong citation link with the US. Table 2.1 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.

Together these stylised facts point toward an increase in the internationalisation of economic research.

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

2.2.2 Gravity model variables

We use the full set of gravity model variables in our analysis with two notable expansions, namely, a newly developed indicator for English proficiency in the general population as well as the virtual proximity indicator used by Hellmanzik and Schmitz (2015) and Hellmanzik and Schmitz (2017).

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 for each country take the linguistic proximity to the United Kingdom, setting all observations above 0.9 to 1 (the United States for instance). In a second step, we take the maximum of this value and the share of the population proficient in English (EF) given by the the 2016 round of the EF English Proficiency Index.[9] Based on this, we create *English Similarity* as follows:

$$English\ similarity_{i,j} = 1 - |EP_i - EP_j| \tag{2.1}$$

$$\text{with} \quad EP_k = \begin{cases} max\left(Language\ similarity_{k,l=UK}, EF_i\right) & \text{if k} \in \text{EF Index} \\ Language\ similarity_{k,l=UK} & \text{if k} \notin \text{EF Index} \end{cases}$$

In addition, we create a dummy variable which indicates that both countries have a relatively low English proficiency. We set this as an English proficiency lower than 40% in the citing country and lower than 60% in the cited country.[10]

We use use bilateral, inter-domain hyperlinks that internationally connect webpages in country A to webpages in country B as in Hellmanzik and Schmitz (2017) to capture information flows via the internet more generally. Their 'virtual proximity indicator' indicator is mainly based on hyperlink data provided by Chung (2011) 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.[11] Due to the bidirectional nature of the data, bilateral hyperlinks reflect the number of links from websites

[9] `http://www.ef.edu/epi/` lists the English proficiency of 72 countries.

[10] We have to set are relatively high cited country proficiency to include any country. With this threshold, we include Israel, France, and Italy. Israel's score is based on the proximity to the UK and, therefore, might have a possible downward bias when compared to the other top ten countries in economics research.

[11] To this end Chung (2011) uses Yahoo's search function and LexiURL Searcher, a social science web analysis tool developed by Thelwall (2009). At the time, Yahoo had indexed about 47 billion websites. For more detailed information on obtaining the measure of bilateral hyperlinks, please refer to Chung (2011).

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. Chung (2011)'s 2009 wave of data are more precise than most 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.[12] In addition, we obtained hyperlinks data for a smaller sample of countries in 1998 as reported by the OECD Communications Outlook 1999.

All geographic variables come from the CEPII datasets: *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 Melitz and Toubal (2014) which summarises evidence about linguistic influences including common official language and common native language and measures of linguistic proximity.

Gravity model variables - stylised facts

The variables employed in our gravity model framework and their summary statistics are reported in Table 2.2. It is worth noting that we have a relatively high English similarity between countries in our sample of 64% and the average geographical distance between cited and citing article is 5125 miles. In addition, Table 2.1 depicts the top 25 citation pairs 1970 to 2016, as well as distance, language, and hyperlinks in 1998, 2003, and 2009 the wave for which we have the fully '.com-cracked' information from Chung (2011). 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 2.1). Interestingly we see that the UK and the US are main drivers behind citations while most of the other top ten countries in terms of international interconnectedness are also the ones which are highly connected in terms of their economics research.

Table 2.3 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.[13] Not surprisingly, we find that English and overall language similarity are highly correlated, yet we believe that

[12]For the United States, usually the sum of the domains *.edu, .us, .mil and .gov* has been used Barnett et al. (2001) in the literature. In previous studies (e.g. Barnett and Sung (2005)), the *.com* domain had either been disregarded or completely attributed to the United States.

[13]Citing articles are shown in in the form log(citing articles + 1).

separating out the effect English has from the overall language effect makes the analysis richer in terms of understanding the major shifts in economics research. Reassuringly, neither variable is strongly correlated with aggregated citation counts. Besides the geographic distance measures, citing articles also show somewhat high correlations with early hyperlinks, while hyperlinks are autocorrelated.

2.3 Empirical strategy

Using our novel data on citations of the top research countries in economics between 1970 and 2016 and following the literature on bilateral trade flows (for example, Silva and Tenreyro (2006), Kimura and Lee (2006), or Hellmanzik and Schmitz (2015)), we estimate a gravity model for the aggregated domestic and bilateral citation flows in economics. We estimate the following quasi-Poisson regression with cluster robust standard errors:

$$\mathbb{E}\left(citing\ articles_{ij} | \alpha_i^I, \alpha_j^E, \mathbf{Z}_{ij}\right) = exp(\alpha_i^I + \mathbf{D}_{ij}\delta + \alpha_j^E) \tag{2.2}$$

We use the amount of citing articles by authors working in country i to articles by authors active in country j, $citingarticles_{ij}$ including domestic citations where citing and receiving country are identical as the dependent variable. Using citing articles as the dependent variable is analogous to the import of citations from articles published by authors who work in country j and reflects the extent to which economics knowledge is disseminated domestically as well as across the globe.

Next to bilateral resistance factors $\mathbf{D}_{ij}$, the estimations include citing (importer) (α_i^I) as well as cited (exporter) fixed effects (α_j^E) to control for any unobservable country-specific factors affecting citation flows in economics. By using Poisson regressions, we force the sums of expected citation flows to equal actual citation flows by including double-fixed effects for the relative nature of distances which is called multilateral resistances in the trade literature (Fally (2015)).[14]

In order to investigate the role of home bias, borders, as well as geographical, cultural, virtual and linguistic distances between countries, we include the following bilateral factors in our estimation:

[14]For specifications that use yearly data, we interact the year dummies with the cited year or the cited and citing year and cluster the standard errors accordingly. Details on this are given in the regression notes.

$$\begin{aligned}\mathbf{D}_{ij}\delta = \delta_1 home + \delta log(geographic\,distance) + \delta_3 time\,zone\,difference_{ij} + \delta_4 contiguity_{ij} \\ + \delta_5 colony_{ij} + \delta_6 common\,legal\,origin_{ij} + \delta_7 common\,religion_{ij} \\ + \delta_8 EU + \delta_9 language\,similarity_{ij} \\ + \delta_{10} English\,similarity_{ij} + \delta_{11} log(hyperlinks\,year)_{ij} \end{aligned} \tag{2.3}$$

As part of our gravity analysis we want to provide evidence on the importance of domestic citations relative to those from the rest of the world. This is captured by the indicator variable *home*. Conventional transportation costs and other biases associated inversely with distance are proxied by geographical distance. In addition to *geographical distance*, we include the full range of gravity model variables, namely *time zone difference*, an indicator for neighbouring countries *contiguous*, an indicator for former colonial ties *common colonial history* as well as indicators for *common religion* as well as *common legal origin* to capture any persisting historical, juridical or cultural links between countries which could bear an impact on citation flows. We also include an indicator variable for *EU* countries as these might per se have stronger ties. Moreover, we include the index for *language similarity* developed by Melitz and Ottaviano (2008) and Melitz and Toubal (2014) in our baseline estimations as typically done in the trade literature.

Two extensions to the gravity model seem particularly relevant in order to analyse 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. To that end, we use *English similarity* which reflects how well versed a country's citizens are in English typically and enters the estimation as the difference of that measure between countries. We hypothesis that *English similarity* has a positive impact on citations and decreases home bias as the standardisation to English as the language of investigation of economics research in most countries as well as in most if not all international research journals has accelerated the dissemination of Economics research eliminating previously existing linguistic costs.

In order to capture the impact of the internet on international citation flows we use a novel indicator for virtual proximity, namely the amount of hyperlinks set between countries. Thus, our second hypothesis is that virtual proximity 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.

In addition to the quasi-Poisson regression presented above, we estimate the above using negative binomial regressions and linear models using OLS with log transformed article

counts. Following Silva and Tenreyro (2006), the main arguments against using OLS in our empirical analysis is the strong assumption about the error needed for consistency. We also show estimates using a negative binomial distribution as the distribution is very skewed, in particular for extreme values in some country pairs that involve the United States.

Negative binomial regressions can be more efficient than Poisson regressions if the conditional variance is not proportional to the conditional mean. In its most common specification, the negative binomial model assumes a quadratic relation which tends to give relatively less weight to observations with a high expected citation count and we observe indeed much higher absolute deviations at the top using negative binomial regressions (See Figure 2.7). We do not use negative binomial regressions as the standard as we assume that similar to trade data, the data given are of higher quality for large countries (See Silva and Tenreyro (2006)). In addition, negative binomial regressions are rarely used for gravity model estimations and, therefore, not as comprehensively studied in this context as the standard estimation methods OLS and Poisson. We vary the sample selection by excluding the United States from our estimations as they likely drive most of our findings using Poisson and results could differ vastly for non-US countries.

Following Fally (2015), Poisson regressions are the only estimation method of the one presented that account consistently for multilateral resistances. Both, OLS and negative binomial regressions overestimate the actual sum of citing articles considerably. For instance, the total flows involving the United States are overestimated by 20 percent (OLS) and 31 percent (NegBin2) in the standard specification (Column (2) and (4) in Table 2.9). 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.

For our baseline estimations we aggregate our data over all years or to the year 2017 as our resistance variables Z_{ij} are time-invariant. However, as this procedure implicitly places more weight on recent years because of the strong growth in economic articles (as discussed in Section 2.2.1), we present alternative estimates exploiting the time variation in citation data.

2.4 Results

2.4.1 Gravity and home bias

Overall findings

Citations follow the law of gravity. Geographic, cultural, and linguistic distance are all estimated to impact negatively on the aggregation of articles that cite a countries research

output in the field of economics. Table 2.4 reports the results of our baseline gravity estimation including double fixed effects as well as clustered standard errors using quasi-Poisson regressions a laid out in the previous Section. Geographic distance is estimated to reduce citing articles by 0.19% 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 Silva and Tenreyro (2006)).

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 the full sample, the overall home bias is estimated between factor 1.59 and 2.96 ($e^{0.463}$ and $e^{1.086}$ respectively, implying an overcitation of domestic articles by 59 or 196%), meaning 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 found for citations is substantially lower than estimates in goods trade, which is estimated to be 3.74 by Chen (2004) and 4.22 by Head and Mayer (2002) for European samples but may be found to be as large as 30.88 by Balta and Delgado (2009).

Further to that, 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, a standard deviation increase in the language similarity index (Melitz and Toubal (2014)) increases the number of expected citing articles by 5% from the mean. If we exclude the USA, this value rises to 9 % but can be as big as 60 % for larger language differences. The estimates for the common religion index as well as dummies for a common colonial history and legal system are positive in all specifications. In the standard specifications (1) and (2), all three variables are estimated to have a substantial and significant impact on aggregated citation flows.

Finally, by considering a sample of almost 100% English language articles, we probably underestimate the home bias in non-English language countries. In addition, we might also underestimate the decrease in this home bias and the importance of language similarity as discussed in the next subsection.

Evolution over time and home bias cross-country

To explore the evolution of home bias over time we repeat our empirical analysis estimating separate regressions for each year thereby allowing covariates to vary for each year. Figure 2.8 presents our results.[15] Plot (a) shows that the home bias is indeed strongest in the early 80s and decreases markedly since the turn of the century. When excluding the US as

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

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 Nitsch (2000) who also provides evidence for a declining border effect in trade.

Plot (c) and (d) present the effect of distance over time. This effect appears to be constant since 2000 while decreasing before. Linguistic distance seems to loose in importance over time (Plots (e) and (f)). This possibly hints at an increase in global English proficiency levels and the rise of English as the language of economic research in non-English language countries.

Figure 2.9 displays the results of repeating the estimations for citations by economists in each cited sample country separately in order to shed more light on the geographical spread of a home bias in global citation patterns. Almost all countries' home bias declines over time with the exception of Israel. Aggregated over time, Israel also has the highest overall home bias at 7.2 ($e^{1.98}$ or 620%), while the USA have the lowest home bias aggregate at 1.9 (90%).[16] While the United States home bias was always considerably lower, most countries had very high values over 7, that is a 600 % over-citation of domestic articles before 2000. After 2010, the home bias factor falls to values around 2.5 in most countries. The UK shows the most striking decline to values around 1.2, surpassing the United States.

By considering only citations to ten countries, we might upward bias the home bias estimates for countries that deal extensively with third party countries. Judging by Figure 2.3d), this relates mostly to Australia. Conversely, we might downward bias Israel's home bias. Figure 2.3d) shows the share of citing articles from third party countries. If we assume that the outgoing citations flow reflects the incoming flow, we would underestimate relatively Israel's domestic traffic and overestimate Australia's propensity to refer to its own economic literature.

Home bias over the course of the article's life-cycle

Figure 2.10 shows how the overall home bias develops in the years after publication and its role in the dissemination time of economic research. 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 caused by low counts in the year of publication which are also mostly from recent years. 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.

[16] See Table 2.8 for home bias by country and different samples.

When looking at the results by country in Figure 2.11, we see that the pattern of a declining home bias over time is almost the same for all countries in our sample, albeit at different levels. The intensity with which domestic articles are favoured over those from abroad in the course of their publication history is equivalent in our estimates and the unadjusted ratio of domestic to foreign articles in Figure 2.5. The level differences between countries might be summarised as follows. We observe two low home bias countries, the UK and US, two high home bias countries, Israel and Belgium, and the other six countries similar at in between. However, these level differences are influenced by the sample selection. For instance, if we only use citations from the ten cited countries, the UK's home bias is estimated to be higher than Canada's. Table 2.8 shows the different home bias estimates by country.

2.4.2 The role of English and the internet

Typically, home bias in goods trade or international investment portfolios is 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.

English similarity

In order to shed light on the role of English proficiency we use a novel proxy for similarity in English levels between citing and cited country using the EF English Proficiency Index (`http://www.ef.edu/epi/`). In addition, we use a dummy for a bilaterally low level of English proficiency.[17] We find that the overall indicator for English similarity is only significant when we exclude English language countries from our estimation. However, the dummy for low English proficiency is associated with a decrease in bilateral citations by 15 to 30%.

[17] The construction of both variables is detailed in Section 2.2.2.

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 a barely significant factor in explaining citation patterns. Another possible explanation could be that we observe that countries with relative 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 overall more to the global economic research in English language journals, these level differences are factored out in our estimation.

In any event, our result for English similarity on citation patterns in economics is markedly different from the findings of Melitz and Toubal (2014) who find that linguistic similarity is conducive to trade more generally. This could be caused by our sample limitation to almost exclusively English language articles. Nevertheless, beyond language as a proxy for transactions costs, it is striking that English similarity does not have a significantly conducive impact on citation patterns.

Internet

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 Hellmanzik and Schmitz (2015) 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 significant, positive impact on citation flows in economics.

The results of this empirical exercise are presented in Table 2.6. We include each available wave of hyperlink information in the estimations separately. We find that the number of hyperlinks set between citing and cited country 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 .25% in 1998 for a one percentage point increase in bilateral hyperlinks, to .083% in 2009 while there is no significant impact of bilateral hyperlinks on citations in economics in 2003.

It seems sensible that the effect is largest in 1998 when the internet was still relatively new and therefore had its strongest innovative power and momentum to revolutionise how we

share information. As an alternative reading, the high early correlation of hyperlinks and citations might express the scientific character of the internet in its early phase such that hyperlinks in 1998 still represent largely scientific links between countries on the forefront of science.

This positive impact is largely in line with comparable literature in international economics, such as Choi (2010) who reports that a doubling of internet usage in a country leads to a 2% to 4% increase in services trade or Hellmanzik and Schmitz (2015) and Hellmanzik and Schmitz (2017) who find that bilateral hyperlinks have a positive impact on investments of advanced economies.

2.4.3 Robustness

Our main concern in the above estimations is the robustness across various estimation techniques in light of the many zeros in the data. We show various alternative estimations of our baseline estimation in Table 2.9. Columns (1) and (2) present the results of negative binomial regressions, Columns (3) and (4) repeat the same empirical exercise using an OLS techniques. In particular OLS is commonly used in gravity model papers and therefore provides insights into the stability of findings across estimation techniques. We conclude that the evidence presented above on home bias persists in all estimations. Distance and the significant proxies for cultural proximity - common colonial ties, common legal origin and most importantly language similarity - are stable across specifications. This is a strong indication that our finding of an existing home bias in economics research as well as the existence of cultural as well as linguistic barriers is robust.

Columns (5) and (6) present results on OLS estimates at different levels of aggregation. Disaggregating the data comes at the advantage that we have a much higher number of observations while the number of zeros in the dataset increases considerably. Nevertheless, adding the time dimension into the data and de facto conducting a panel rather than a cross-sectional analysis confirms the findings presented above.

Lastly, we run estimations for the internet specification using growth rather than levels as it seems likely that it is the dynamic in the rise of the internet rather than its level which is driving the main result. To that end we use levels and calculate the change of all time-varying variables from 1998 or 2003 to 2009 and re-run the regressions. We find that only the level of hyperlinks 1998 is significant but not the subsequent growth or subsequent levels. However, this findings supports our earlier result that the impact of the internet is particularly strong in its early days.

2.5 Conclusion

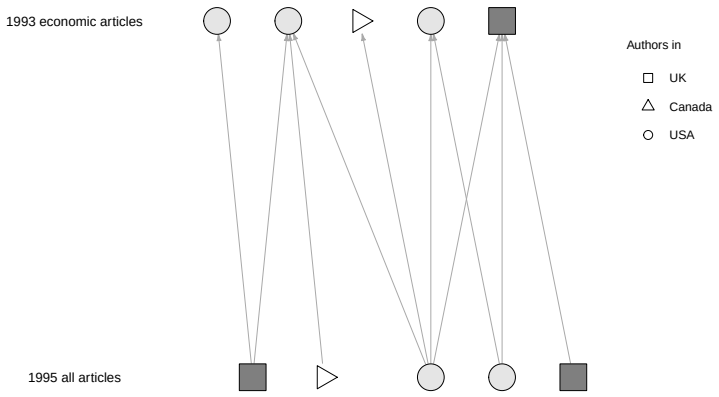
Using a gravity model framework this paper analyses international citation flows in Economic research between 1970 and 2016 seeking to shed light on the role of geographical, cultural, virtual and linguistic distances in the dissemination of knowledge in economics. Based on novel data, citation aggregates of country pairs including intra-national citations are linked to traditional measures for cultural proximity as well as new data on internet hyperlinks between countries as well as refined linguistic indicators.

Our results show that citation patterns follow the law of gravity with a factor that is smaller than commonly found in the literature on trade in goods and particularly services, nevertheless considerable if we take the the premise that research disseminates without borders or cultural preferences. Moreover, our paper provides evidence that citations in economics exhibit a strong and significant home bias in all ten leading countries amounting to a factor between 1.6 and 3 overall. This expresses an estimated propensity for domestic articles between 60 and 200 percent.

This ties in with our finding that while linguistic proximity more generally is significant for citations in economics, there is no significant English proficiency effect above a threshold of very low English proficiency levels. Thus, we find no direct evidence for the boosting effect of English for the exchange of economic research between countries. Lastly, our paper shows that countries with closer internet ties have significantly higher shares of bilateral citations ranging from .083 percent to a .25 percent for a 1 percent increase in internet hyperlinks. Though, our estimates for the impact of the distance measures are not changed substantially by the inclusion of English proficiency or internet links in our analysis. We find a strong home bias even for the native English speaking countries which dominate research in economics. Hence, we do not find comprehensive evidence that the home bias or geographic and cultural distance are resolved by the use of the internet and English as a universal scientific language.

Although our paper has provided evidence on the unequal dissemination of economic research between countries, we also observe the declining importance of such frictions over time. On a more general level and in spite of globalisation, one also has to acknowledge that the observed home bias does not necessarily reflect a shortcoming as many research questions indeed might be of a local nature and therefore of limited appeal to the rest of the world. It is left to further research to distinguish between an optimal and the observed home bias as it is common, for instance for the financial equity home bias (e.g. Park and Mercado (2014)).

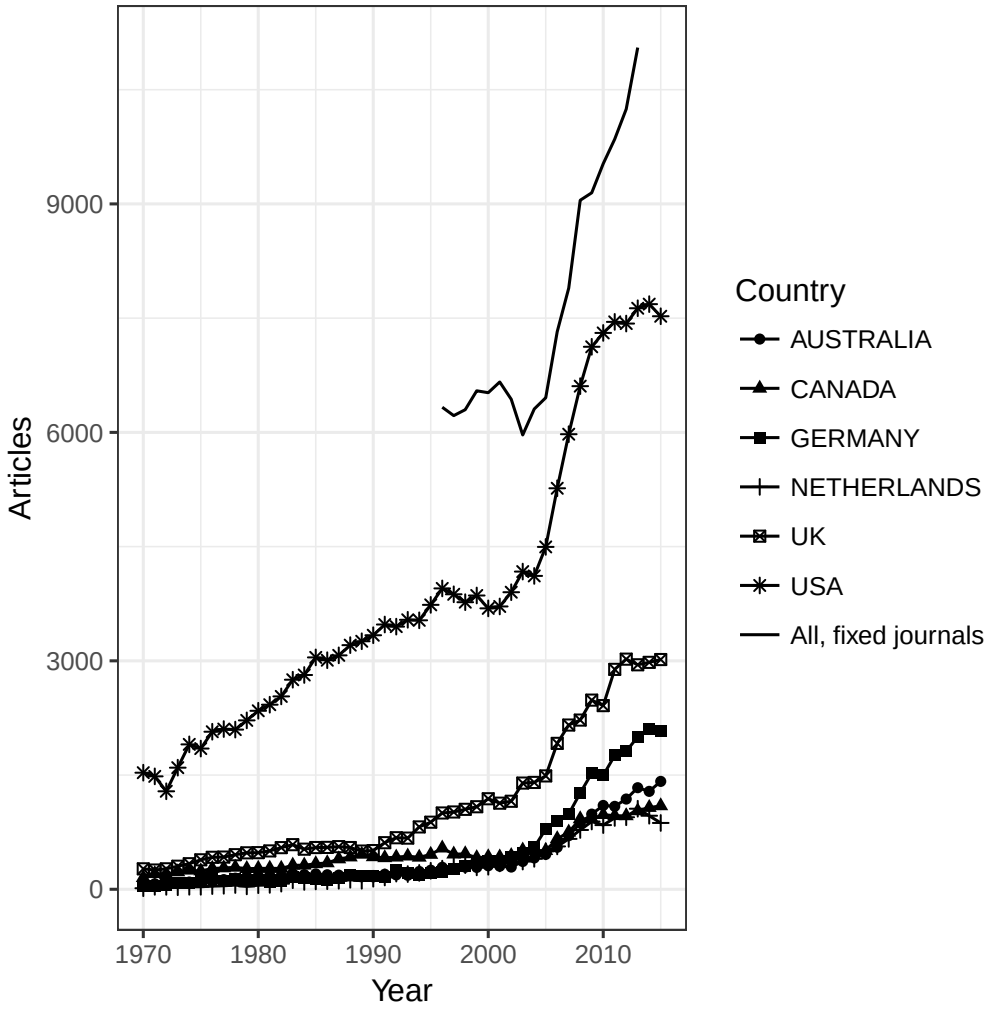
Figure 2.1: Citation data example



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

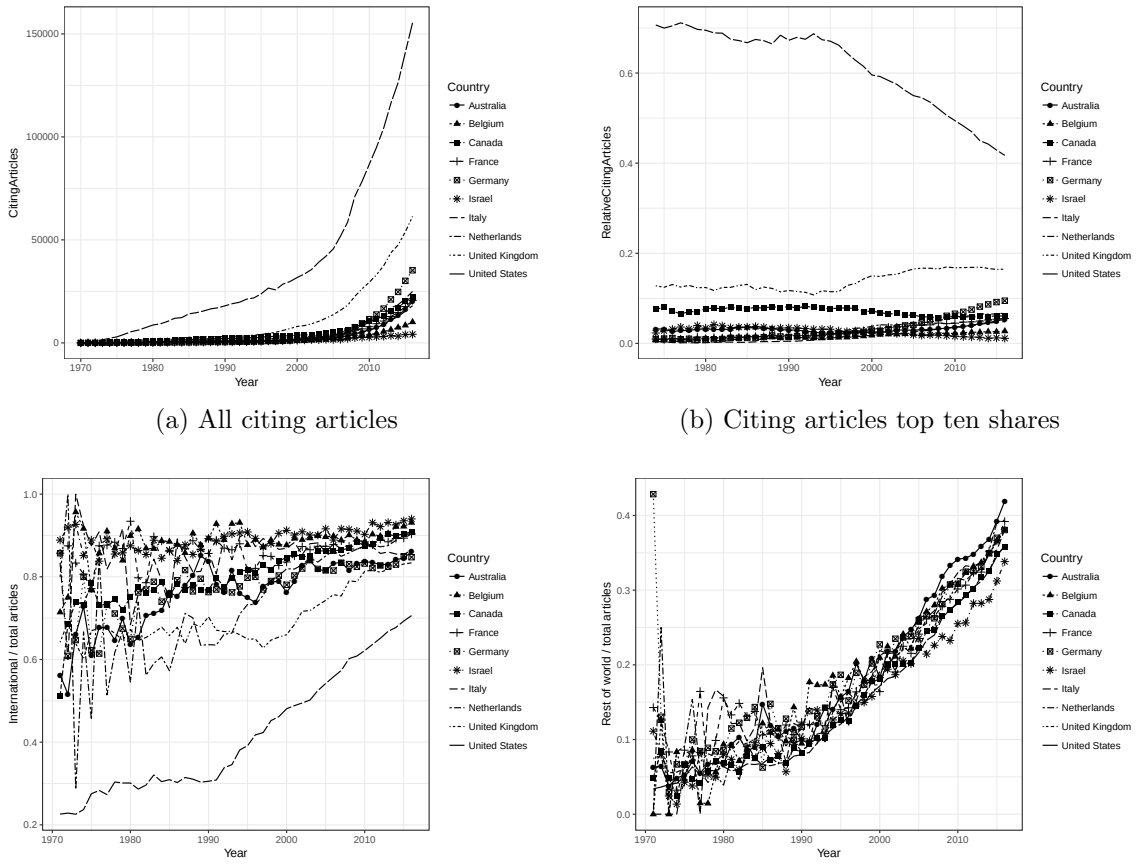
2.A Figures

Figure 2.2: 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 2.3: Citing totals for the top ten countries over time



(a) All citing articles

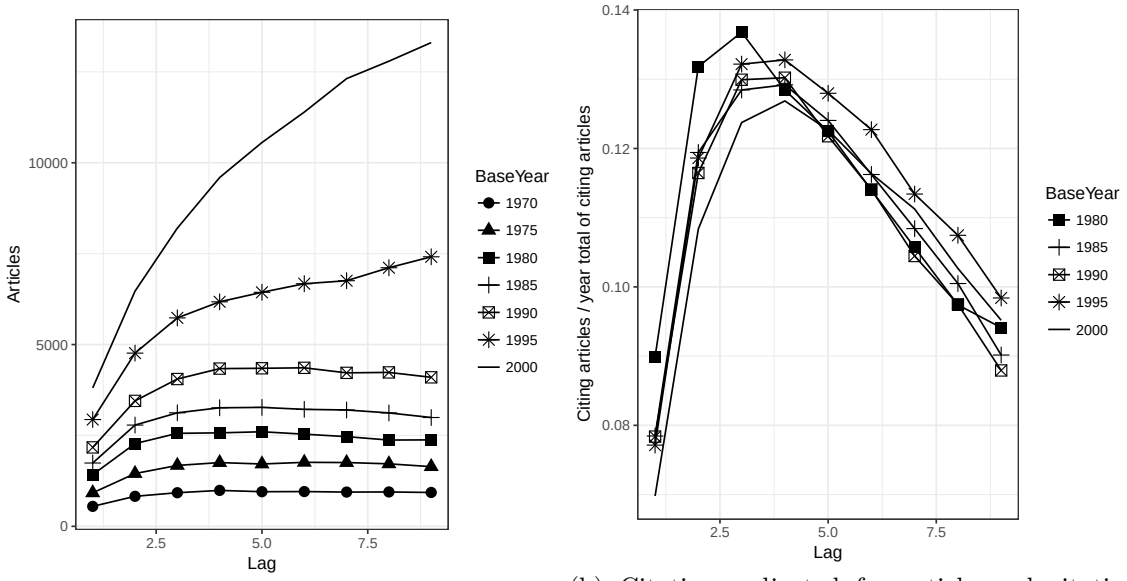
(b) Citing articles top ten shares

(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 2.4: Citation time windows - development of citations over time

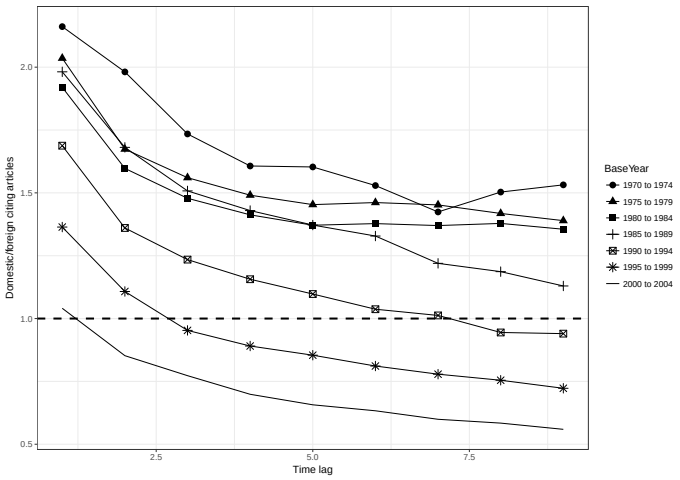


(a) Absolute number of citations

(b) Citations adjusted for article and citation growth

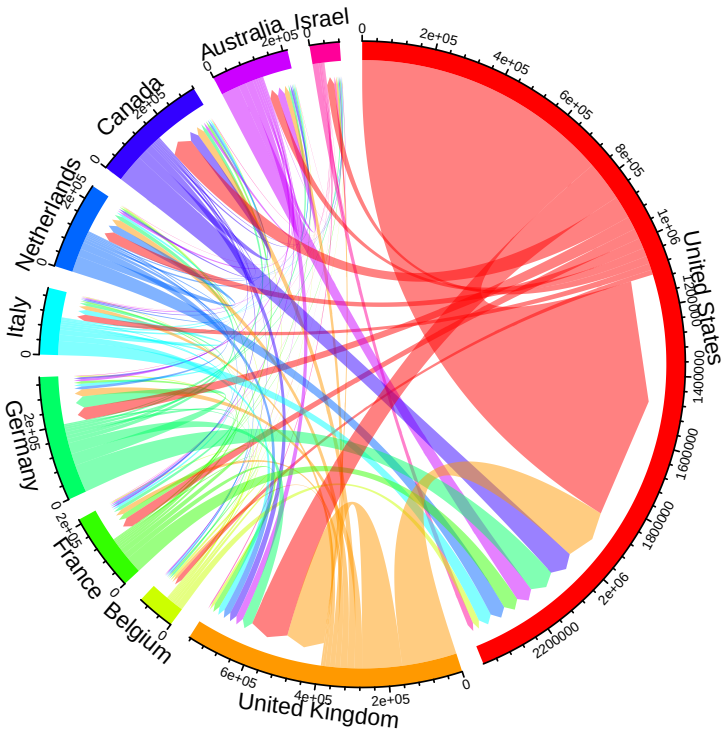
Source: Own calculations based on WoS data.

Figure 2.5: Domestic and Foreign citing articles over time

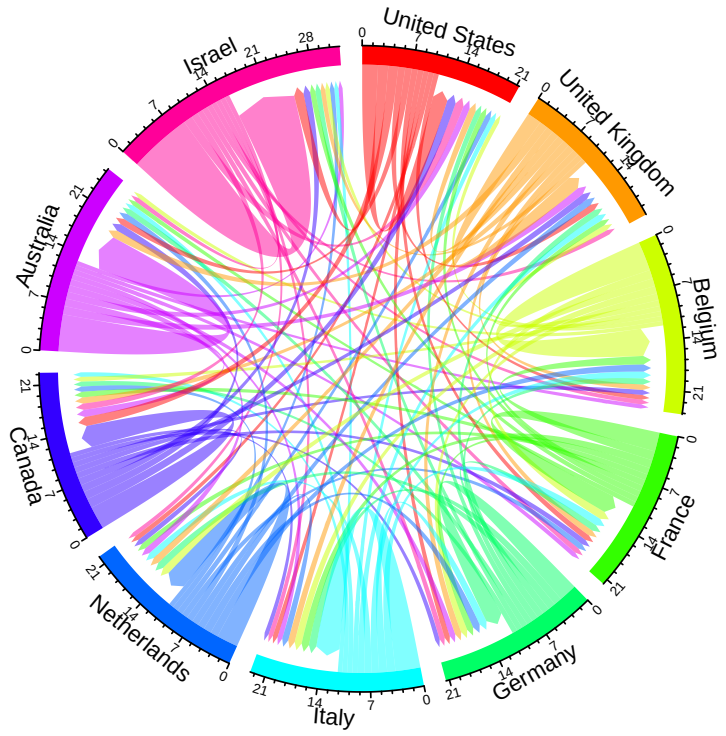


Notes: The ratio of domestic to foreign citing articles by time passed since publication.
Source: Own calculations based on Scopus data.

Figure 2.6: Citations between top ten countries



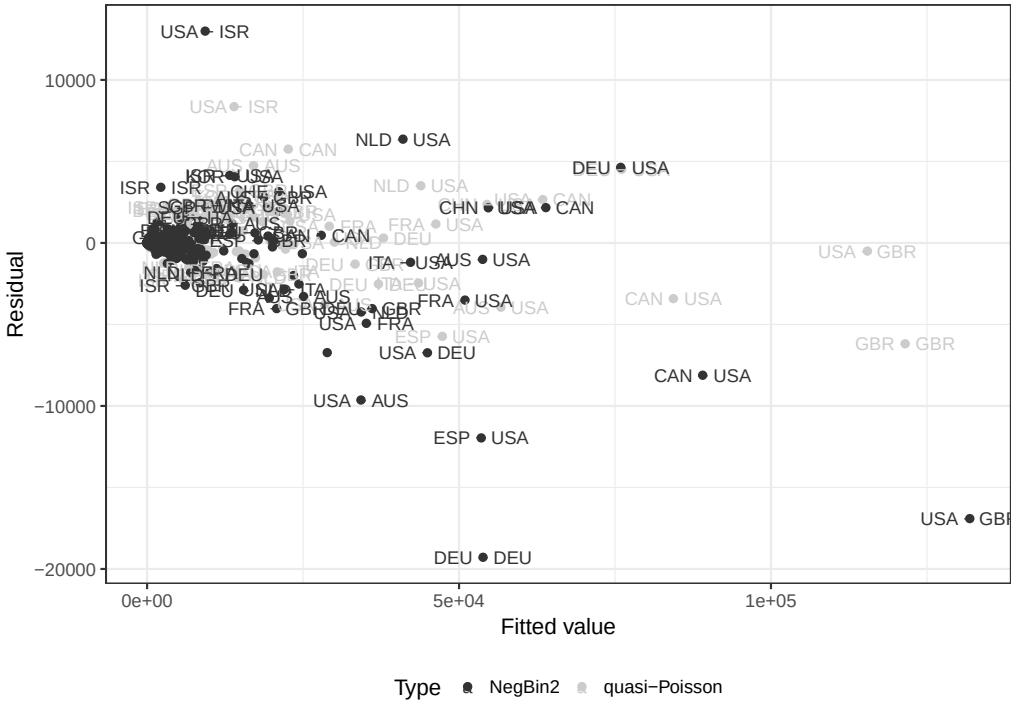
(a) Total citations



(b) Relative citations

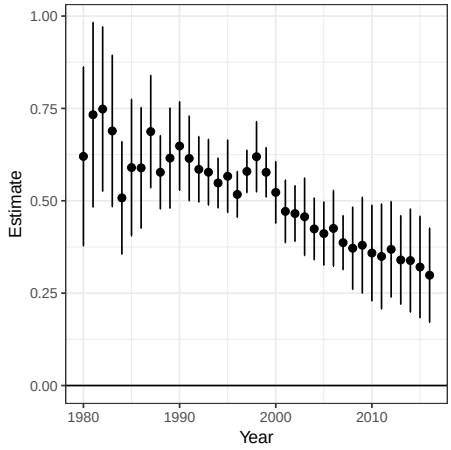
Notes: Citing articles between 1970 and 2016. *Source:* Own calculations based on WoS data

Figure 2.7: Residual plot for quasi-Poisson and negative binomial regression

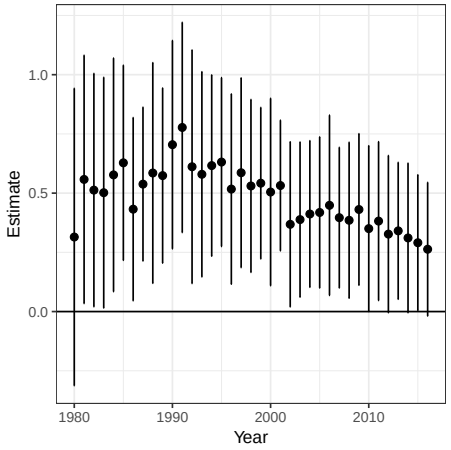


Note: Residuals and fitted values for the standard specification estimated using a quasi-Poisson regression as shown in Table 2.4, column (2) and a negative binomial regression as shown in Table 2.9, column (2). We omit fitted values over 150,000, that is both USA-USA and GBR-USA observations, for readability. The NegBin2 residuals are -645,675 (USA-USA) and -45,564 (GBR-GBR). The Poisson residuals are -11,182 (USA-USA) and -4,110 (GBR-GBR).

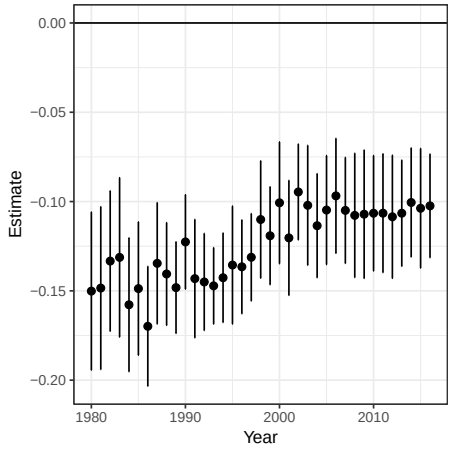
Figure 2.8: Distance and language estimates by citing year



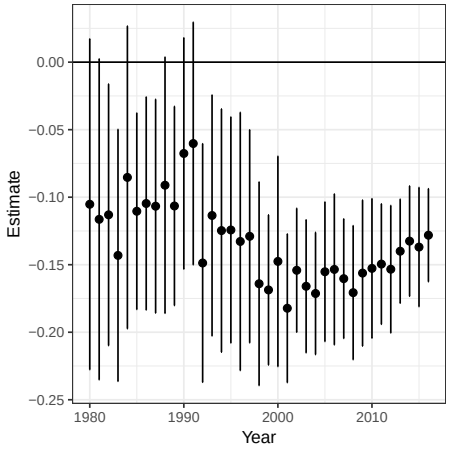
(a) Estimated home bias



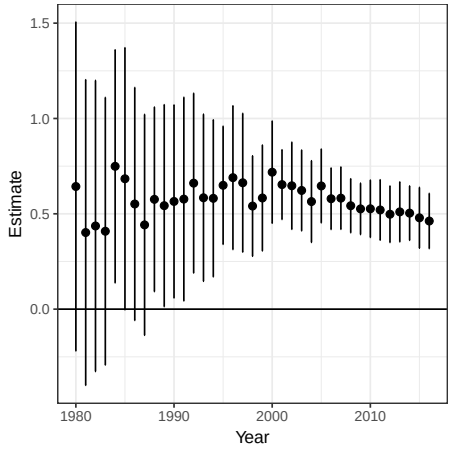
(b) Home bias similarity w/o USA



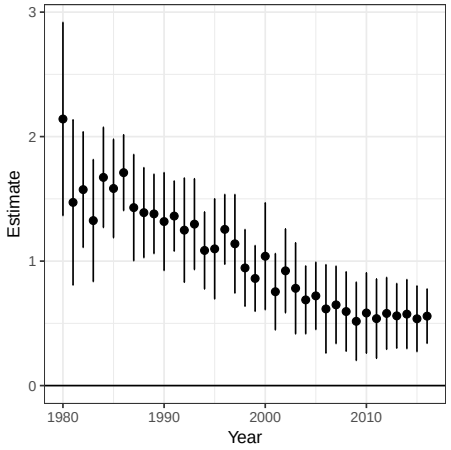
(c) Estimated effect of geographic distance



(d) Geographic distance w/o USA



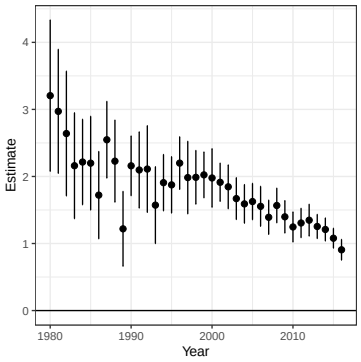
(e) Estimated effect of language similarity



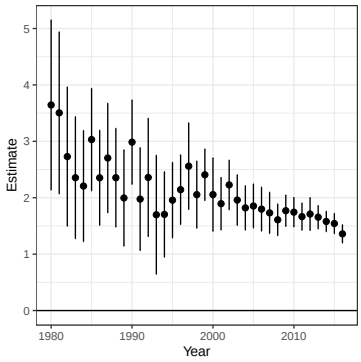
(f) Language similarity w/o USA

Notes: Estimated effects and 95 % percent confidence intervals from repeated quasi-Poisson regressions including the three shown independent variables and otherwise identical to the Regression shown in Table 2.4, Column (2). The regressions are repeated for each citing year from 1980 to 2016.

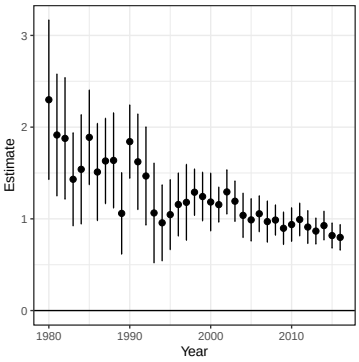
Figure 2.9: Home bias by citing year



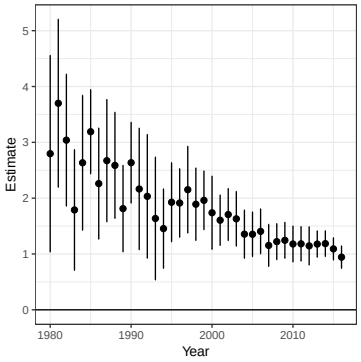
(a) Home bias Australia



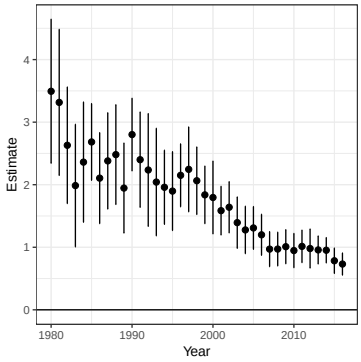
(b) Home bias Belgium



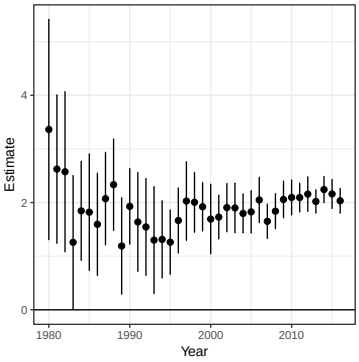
(c) Home bias Canada



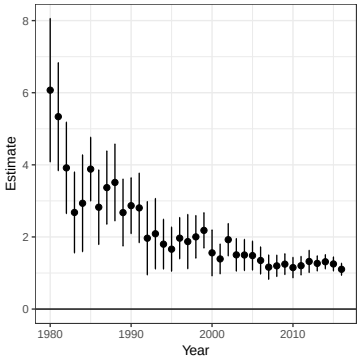
(d) Home bias France



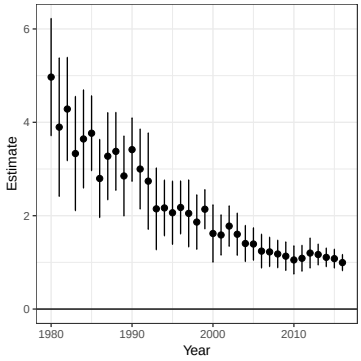
(e) Home bias Germany



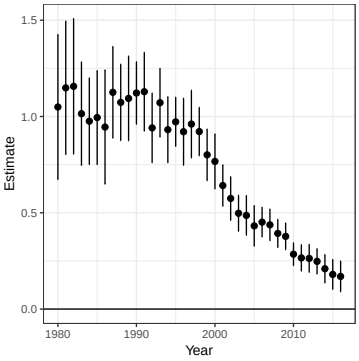
(f) Home bias Israel



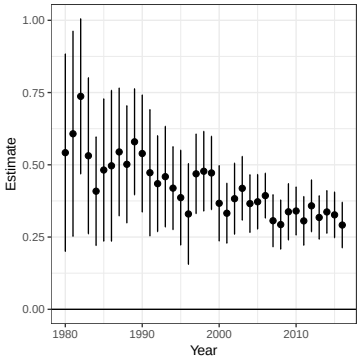
(g) Home bias Italy



(h) Home bias Netherlands



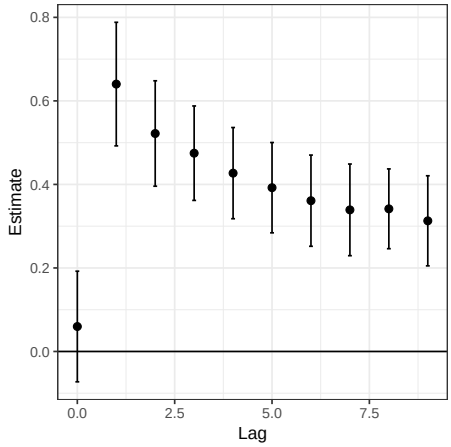
(i) Home bias UK



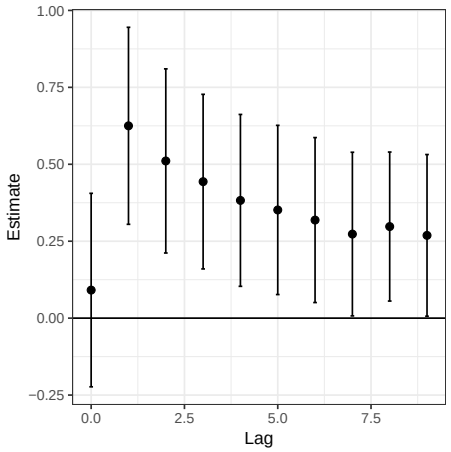
(j) Home bias United States

Notes: Estimated effects and 95 % percent confidence intervals from repeated quasi-Poisson regressions otherwise identical to the Regression shown in Table 2.4, Column (2). The regressions are repeated for each citing year from 1980 to 2016.

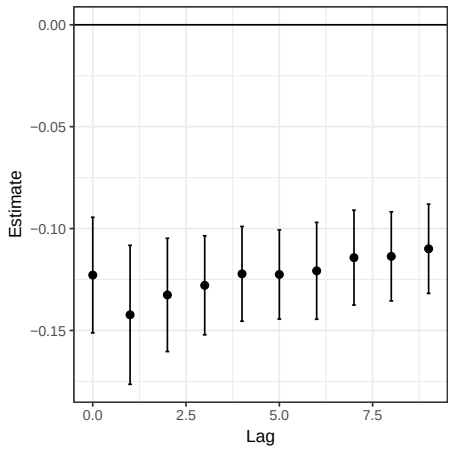
Figure 2.10: Distance and language over time between citing and cited article



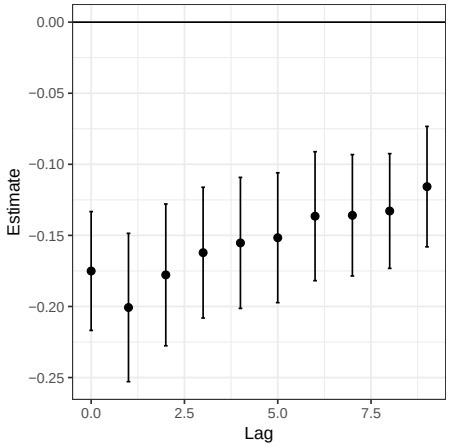
(a) Estimated home bias



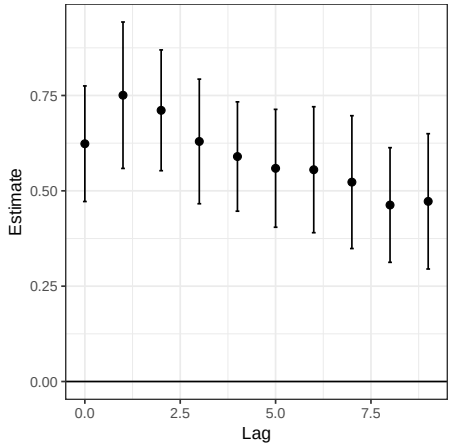
(b) Home bias similarity w/o USA



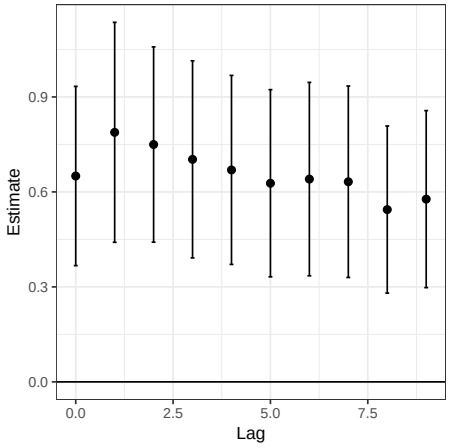
(c) Estimated effect of geographic distance



(d) Geographic distance w/o USA



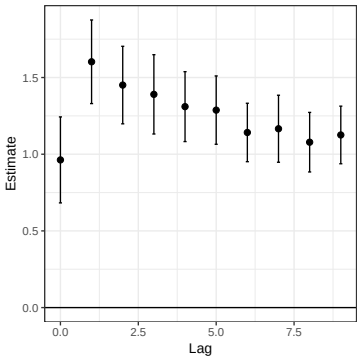
(e) Estimated effect of language similarity



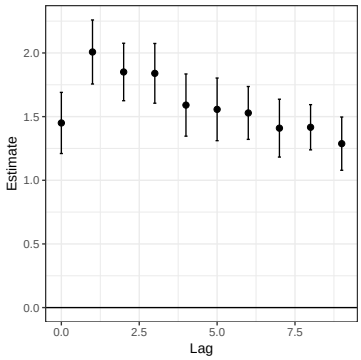
(f) Language similarity w/o USA

Notes: Estimated effects and 95 % percent confidence intervals from repeated quasi-Poisson regressions including the three shown independent variables and otherwise identical to the Regression shown in Table 2.4, Column (2). The regressions are repeated for per lag between citing and cited year ranging from 0 to 9 years.

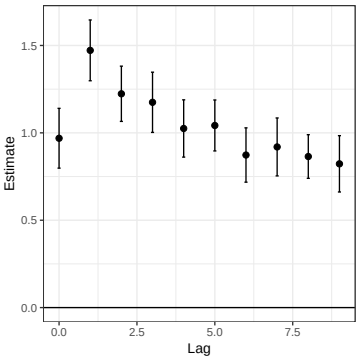
Figure 2.11: Home bias over time between citing and cited article



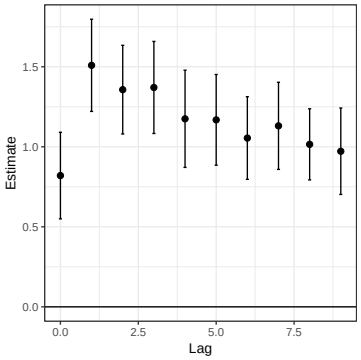
(a) Home bias Australia



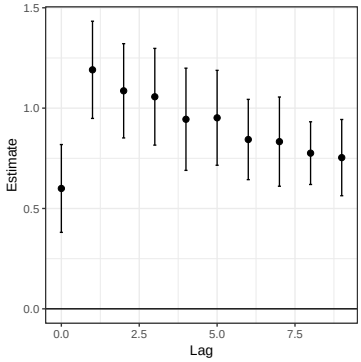
(b) Home bias Belgium



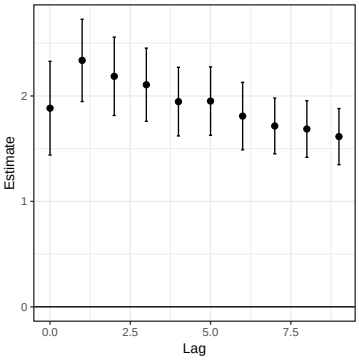
(c) Home bias Canada



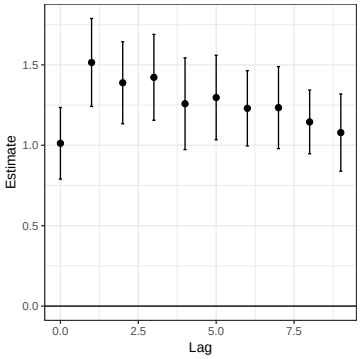
(d) Home bias France



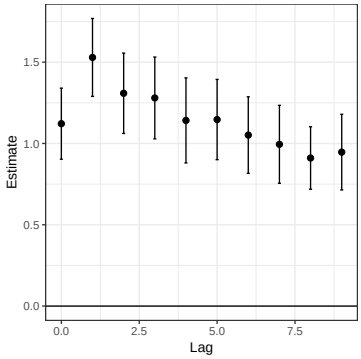
(e) Home bias Germany



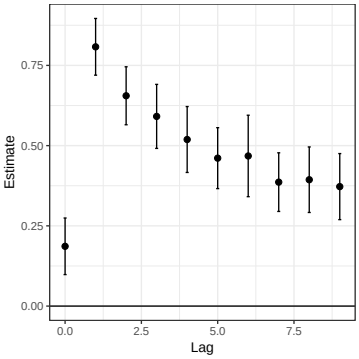
(f) Home bias Israel



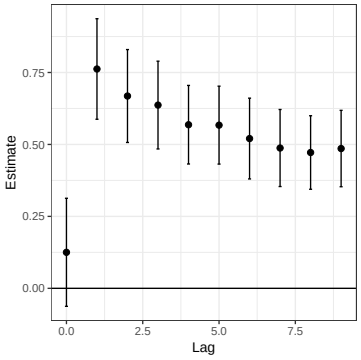
(g) Home bias Italy



(h) Home bias Netherlands



(i) Home bias UK



(j) Home bias United States

Notes: Estimated effects and 95 % percent confidence intervals from repeated quasi-Poisson regressions otherwise identical to the Regression shown in Table 2.4, Column (2). The regressions are repeated for per lag between citing and cited year ranging from 0 to 9 years.

2.B Tables

Table 2.1: Top 25 citation pairs

Citing country	Publication country	Citing articles	Distance	Language	Links98	Links03	Links09
United States	United States	746,197	1,161	1.00			59,269,708
United Kingdom	United States	158,753	5,570	0.79	149,667	1,931,700	31,314,578
United Kingdom	United Kingdom	115,325	186	1.00			5,770,814
United States	United Kingdom	114,959	5,570	0.79	212,106	24,936,200	48,878,700
Canada	United States	80,940	548	0.55	143,518	1,129,580	12,911,070
Germany	United States	80,555	6,035	0.34	140,936	1,049,700	20,119,072
United States	Canada	66,085	548	0.55	121,369	3,607,400	10,713,620
China	United States	56,843	10,994	0.06		195,143	34,859,492
Australia	United States	52,776	16,009	0.64	87,987	606,564	6,245,988
France	United States	47,448	5,838	0.19	33,027	442,321	20,979,438
Netherlands	United States	47,382	5,866	0.38	31,719	343,746	4,170,676
Spain	United States	41,599	5,770	0.25	19,238	476,771	14,314,497
Italy	United States	41,037	6,895	0.23	32,480	321,650	11,088,147
United States	Germany	38,198	6,035	0.34	82,739	20,946,400	40,771,812
Germany	Germany	34,565	225	1.00			3,744,129
Germany	United Kingdom	32,044	495	0.38	57,031	16,500,000	20,831,130
United States	France	30,211	5,838	0.19	33,349	4,701,510	18,235,724
United States	Netherlands	30,085	5,866	0.38	50,745	4,443,500	10,972,904
Canada	Canada	28,381	1,188	1.00			1,148,503
United States	Australia	24,645	16,009	0.64	85,841	6,291,230	13,225,842
Switzerland	United States	24,250	6,272	0.31	23,406	512,660	3,127,408
Japan	United States	24,226	10,856	0.07	123,278	471,260	34,111,636
Sweden	United States	23,000	6,323	0.41	41,679	357,827	4,287,421
United States	Israel	22,302	9,120	0.23		587,260	1,407,114
Netherlands	Netherlands	22,145	77	1.00			233,998

Table 2.2: Summary statistics for aggregated data

Statistic	N	Mean	St. Dev.	Min	Max
Home	1,240	0.008	0.089	0	1
LogDistance	1,150	8.516	0.968	4.013	9.866
Contiguity	1,150	0.025	0.157	0	1
TimeDifference	1,190	4.262	3.464	0.000	12.000
Colony	1,150	0.057	0.233	0	1
CommonLegal	1,190	0.364	0.481	0	1
CommonReligion	1,190	0.162	0.206	0.000	0.876
EU	1,240	0.126	0.332	0	1
LanguageSimilarity	1,150	0.179	0.169	0.000	1.000
EnglishSimilarity	1,180	0.643	0.268	0.000	1.000
LogHyperLinks98	252	8.238	1.493	4.431	12.265
LogHyperLinks03	460	12.100	2.040	6.182	17.032
LogHyperLinks09	760	11.988	2.561	4.290	17.898
Citations	1,240	2,681	22,949	0	746,197

Table 2.3: Correlation of distance measures demeaned

	Home	Dist	Cont	Time	Col	Law	Rel	EU	Lang	Engl	L98	L03	L09	Cit
Home		-0.36	-0.10	-0.16	-0.06	0.13	0.10	-0.14	0.52	0.18			-0.24	0.30
Dist	-0.36		-0.37	0.69	0.10	-0.12	-0.13	-0.44	-0.37	-0.12	-0.39	-0.33	-0.17	-0.36
Cont	-0.10	-0.37		-0.16	0.02	0.08	0.09	0.24	0.17	0.04	0.40	0.25	0.27	0.08
Time	-0.16	0.69	-0.16		0.03	-0.06	-0.02	-0.25	-0.18	-0.10	-0.42	-0.28	-0.14	-0.27
Col	-0.06	0.10	0.02	0.03		0.22	0.02	-0.12	0.16	0.11	0.13	0.07	0.10	0.11
Law	0.13	-0.12	0.08	-0.06	0.22		0.23	0.02	0.37	0.18	0.41	0.24	0.09	0.21
Rel	0.10	-0.13	0.09	-0.02	0.02	0.23		0.11	0.35	0.14	0.20	0.16	0.09	0.10
EU	-0.14	-0.44	0.24	-0.25	-0.12	0.02	0.11		-0.03	-0.08	0.26	0.27	0.30	0.12
Lang	0.52	-0.37	0.17	-0.18	0.16	0.37	0.35	-0.03		0.58	0.40	0.25	0.02	0.27
Engl	0.18	-0.12	0.04	-0.10	0.11	0.18	0.14	-0.08	0.58		0.26	0.06	0.01	0.06
L98		-0.39	0.40	-0.42	0.13	0.41	0.20	0.26	0.40	0.26		0.44	0.40	0.56
L03		-0.33	0.25	-0.28	0.07	0.24	0.16	0.27	0.25	0.06	0.44		0.83	0.29
L09	-0.24	-0.17	0.27	-0.14	0.10	0.09	0.09	0.30	0.02	0.01	0.40	0.83		0.04
Cit	0.30	-0.36	0.08	-0.27	0.11	0.21	0.10	0.12	0.27	0.06	0.56	0.29	0.04	

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 pairwise complete). Dist is LogDistance, Cont is Contiguity, Time is TimeDifference, Col is Colony, Rel is CommonReligion, Law is CommonLegal, Lang is LanguageSimilarity, Engl is EnglishSimilarity, L98 is LogHyperLinks98, L03 is LogHyperLinks03, L09 is LogHyperLinks09, and Cit is LogCitations. Citing artilces are aggregated counts from 1970 to 2016 as described in Section 2.2

Table 2.4: Common gravity distance measures and aggregated citation flows

	Dependent variable (link:log): Citing articles						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Home	0.755***	0.642***	1.086***	0.834***	0.577***	0.463***	0.606***
	(0.064)	(0.064)	(0.099)	(0.110)	(0.051)	(0.180)	(0.058)
LogDistance	-0.181***	-0.175***	-0.118**	-0.089**	-0.173***	-0.122***	-0.191***
	(0.032)	(0.032)	(0.046)	(0.039)	(0.035)	(0.045)	(0.034)
TimeDifference	0.023**	0.022**	0.003	-0.004	0.011	0.021	0.029***
	(0.011)	(0.011)	(0.014)	(0.012)	(0.009)	(0.014)	(0.010)
Contiguity	-0.039	-0.049	-0.026	-0.067	-0.065	0.080	-0.089*
	(0.042)	(0.045)	(0.040)	(0.041)	(0.045)	(0.064)	(0.046)
Colony	0.211***	0.168***	0.338***	0.249***	0.197***	0.089**	0.160***
	(0.049)	(0.048)	(0.053)	(0.051)	(0.051)	(0.039)	(0.048)
CommonLegal	0.135***	0.090**	0.031	0.007	0.039	0.110**	0.029
	(0.030)	(0.039)	(0.032)	(0.026)	(0.030)	(0.049)	(0.030)
CommonReligion	0.280***	0.194*	0.316***	0.178	0.277***	-0.080	0.278**
	(0.107)	(0.110)	(0.106)	(0.122)	(0.107)	(0.124)	(0.116)
EU	0.178***	0.186***	0.302***	0.314***	0.126*	0.118*	0.160***
	(0.065)	(0.064)	(0.086)	(0.081)	(0.070)	(0.065)	(0.061)
LanguageSimilarity		0.270*		0.472***	0.332**	0.352**	0.256*
		(0.154)		(0.147)	(0.133)	(0.169)	(0.133)
USA-ISR					0.596***		
					(0.190)		
N (df)	1140 (1009)	1140 (1008)	1017 (888)	1017 (887)	1140 (1007)	1140 (1008)	53,580 (47,790)
Countries	all	all	no USA	no USA	all	all	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 country level. All models include citing and cited country dummies. For regression model (5), the country dummies are interacted with the publication year (level of aggregation). The dependent variable is citation imports with one unit corresponding to the number of citing articles up to ten years after publication in any single country to articles written by authors in the ten leading research countries in a year between 1970 and 2016. 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. The regression models (1)-(5) use standard aggregated data between countries. For specification (6), all within distances are set to the minimum and similarities to the maximum. Specification (6) estimates, therefore, a lower bound on the home bias. Specification (7) uses data aggregated by country pair and cited year. (7) also includes cited and citing year fixed effects interacted with cited year. Standard errors are clustered by cited year and coting country in (7).

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

Table 2.5: English proficiency similarity

	Dependent variable (link:log): Citing articles					
	(1)	(2)	(3)	(4)	(5)	(6)
Home	0.708***	0.554***	0.716***	0.550***	0.572***	1.013***
	(0.070)	(0.058)	(0.073)	(0.058)	(0.056)	(0.150)
LogDistance	-0.183***	-0.164***	-0.178***	-0.164***	-0.148***	-0.101*
	(0.036)	(0.035)	(0.036)	(0.035)	(0.029)	(0.057)
TimeDifference	0.013	0.009	0.011	0.009	0.002	0.009
	(0.010)	(0.009)	(0.010)	(0.009)	(0.006)	(0.019)
Contiguity	-0.057	-0.058	-0.050	-0.060	-0.040	-0.026
	(0.043)	(0.049)	(0.042)	(0.049)	(0.041)	(0.059)
Colony	0.246***	0.185***	0.247***	0.184***	0.199***	0.240***
	(0.051)	(0.052)	(0.051)	(0.052)	(0.054)	(0.059)
CommonLegal	0.083***	0.044	0.096***	0.041	0.049*	0.034
	(0.029)	(0.031)	(0.026)	(0.031)	(0.028)	(0.033)
CommonReligion	0.367***	0.265**	0.376***	0.263**	0.335***	0.133
	(0.095)	(0.109)	(0.094)	(0.109)	(0.113)	(0.150)
USA-ISR	0.587***	0.571***	0.567***	0.575***		
	(0.189)	(0.183)	(0.198)	(0.183)		
EU	0.121*	0.124*	0.115	0.120*	0.093	0.313***
	(0.072)	(0.068)	(0.071)	(0.069)	(0.064)	(0.110)
EnglishSimilarity	0.072	-0.111		-0.098	-0.082	0.251**
	(0.081)	(0.102)		(0.103)	(0.092)	(0.109)
LanguageSimilarity		0.424**		0.422**	0.326**	0.237
		(0.169)		(0.167)	(0.142)	(0.175)
BadEnglish			-0.254***	-0.281***	-0.154**	-0.367***
			(0.057)	(0.068)	(0.061)	(0.061)
N (df)	1140 (1007)	1140 (1006)	1140 (1007)	1140 (1005)	1017 (885)	749 (625)
Countries	all	all	all	all	w/o Israel	w/o English language

Notes: This table reports estimated coefficients from quasi-Poisson with cluster robust standard errors in parentheses. Standard errors are clustered at the citing country level. All models include citing and cited country dummies. The dependent variable is citation imports with one unit corresponding to the number of citing articles in any single country to articles written by authors in the ten leading research countries in a year between 1970 and 2016. 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 2.6: Internet hyperlinks and citation flows

	Dependent variable (link:log): Citing articles					
	(1)	(2)	(3)	(4)	(5)	(6)
LogDistance	-0.055	-0.040	-0.009	-0.009	-0.026	-0.028
	(0.064)	(0.050)	(0.032)	(0.033)	(0.028)	(0.029)
TimeDifference	0.006	0.008	0.002	0.002	-0.011	-0.011
	(0.019)	(0.016)	(0.011)	(0.011)	(0.010)	(0.009)
Contiguity	0.144*	0.114	0.229***	0.229***	0.054	0.054
	(0.077)	(0.078)	(0.063)	(0.063)	(0.036)	(0.036)
Colony	0.204***	0.165**	0.121***	0.123***	0.140***	0.127***
	(0.070)	(0.067)	(0.046)	(0.046)	(0.040)	(0.046)
CommonLegal	0.040	-0.019	-0.038	-0.042	0.051	0.047
	(0.061)	(0.054)	(0.045)	(0.044)	(0.033)	(0.032)
CommonReligion	0.355*	0.320	-0.002	-0.004	0.215**	0.227**
	(0.212)	(0.224)	(0.167)	(0.168)	(0.092)	(0.092)
EU	0.332*	0.286*	0.349***	0.339***	0.270***	0.228***
	(0.194)	(0.166)	(0.105)	(0.112)	(0.079)	(0.077)
LanguageSimilarity	-0.186	-0.387	0.189	0.164	0.137	0.003
	(0.228)	(0.249)	(0.197)	(0.197)	(0.171)	(0.145)
EnglishSimilarity	0.079	0.174	-0.030	-0.023	-0.162	-0.103
	(0.215)	(0.210)	(0.153)	(0.150)	(0.115)	(0.105)
LogHyperLinks98		0.250***				
		(0.078)				
LogHyperLinks03				0.013		
				(0.022)		
LogHyperLinks09						0.083**
						(0.035)
N (df)	243 (198)	243 (197)	440 (377)	440 (376)	740 (647)	740 (646)
Citing year	1998	1998	2003	2003	2009	2009

Notes: This table reports estimated coefficients from quasi-Poisson with cluster robust standard errors in parentheses. Standard errors are clustered at the citing country level. All models include citing and cited country dummies. The dependent variable is citation imports with one unit corresponding to the number of citing articles in any single country to articles written by authors in the ten leading research countries in a year between 1970 and 2016. 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 2.7: Internet hyperlinks and citation flow growth

	Dependent variable (link:log): (Citing articles$_{t\geq 2009}$ - Citing articles$_{t\leq BaseYear}$)					
	(1)	(2)	(3)	(4)	(5)	(6)
LogDistance	-0.009	-0.005	-0.006	-0.010	0.001	0.002
	(0.018)	(0.017)	(0.019)	(0.017)	(0.015)	(0.014)
TimeDifference	-0.010**	-0.014**	-0.013**	-0.009**	-0.010**	-0.010**
	(0.005)	(0.006)	(0.006)	(0.004)	(0.004)	(0.005)
Contiguity	0.052**	0.050**	0.042**	0.023	0.011	0.008
	(0.025)	(0.021)	(0.021)	(0.025)	(0.024)	(0.023)
Colony	0.082***	0.081**	0.078**	0.077***	0.060**	0.060**
	(0.027)	(0.032)	(0.033)	(0.028)	(0.028)	(0.027)
CommonLegal	0.003	0.003	-0.006	-0.003	-0.020	-0.019
	(0.023)	(0.023)	(0.022)	(0.023)	(0.022)	(0.022)
CommonReligion	0.224***	0.249***	0.244***	0.216***	0.221***	0.228***
	(0.056)	(0.080)	(0.085)	(0.050)	(0.054)	(0.054)
EU	0.177***	0.088*	0.084*	0.146***	0.132**	0.133**
	(0.053)	(0.050)	(0.044)	(0.053)	(0.057)	(0.057)
LanguageSimilarity	0.018	0.008	-0.005	0.008	0.039	0.048
	(0.076)	(0.086)	(0.084)	(0.062)	(0.059)	(0.062)
LogCitations1998	0.214***	0.251***	0.238***			
	(0.027)	(0.048)	(0.050)			
LogHyperLinks98			0.055*			
			(0.029)			
LogHyperLinks9809			-0.011			
			(0.038)			
LogCitations2003				0.308***	0.382***	0.385***
				(0.032)	(0.033)	(0.033)
LogHyperLinks03						-0.005
						(0.010)
LogHyperLinks0309						-0.037
						(0.033)
N (df)	1130 (998)	225 (182)	225 (180)	1130 (998)	420 (359)	420 (357)
Base Year	1998	1998	1998	2003	2003	2003

Notes: This table reports estimated coefficients from quasi-Poisson with cluster robust standard errors in parentheses. Standard errors are clustered at the citing country level. All models include citing and cited country dummies. The dependent variable is citation imports with one unit corresponding to the number of citing articles in any single country to articles written by authors in the ten leading research countries in a year between 1970 and 2016. 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 2.8: Estimated home bias of top ten countries

Country	Home.Full	std.Error.Full	Home.Top.Ten	std.Error.Top.Ten	Home.Min.Dist	std.Error.Min.Dist
Australia	1.35	0.13	1.78	0.16	1.16	0.15
Belgium	1.71	0.12	2.11	0.23	1.69	0.12
Canada	1.11	0.09	1.25	0.14	0.92	0.09
France	1.29	0.15	1.81	0.28	1.20	0.15
Germany	1.02	0.12	1.52	0.20	0.91	0.14
Israel	1.98	0.13	2.05	0.26	1.89	0.13
Italy	1.36	0.14	1.78	0.28	1.29	0.15
Netherlands	1.23	0.12	1.71	0.21	1.17	0.13
United Kingdom	0.84	0.10	1.26	0.18	0.71	0.13
United States	0.65	0.11	0.96	0.16	0.46	0.11

Notes: Estimated home bias by country in regressions, similar to specification (2) in Table 2.4. Columns 4 and 5 show the estimated home bias and its standard error using only citations from the ten cited countries. For the estimates in the last two columns, the geographic distance within countries has been set to the overall minimum and religion similarity to the maximum, equivalent to column (6) in Table 2.4.

Table 2.9: Alternative estimations using OLS and negative binomial regressions

	Dependent variable (link:log): Citing articles					
	(1)	(2)	(3)	(4)	(5)	(6)
Home	1.056***	0.976***	1.022***	0.975***	1.241***	1.140***
	(0.176)	(0.172)	(0.181)	(0.182)	(0.191)	(0.244)
LogDistance	-0.108**	-0.105**	-0.114***	-0.111***	-0.008	0.020
	(0.044)	(0.043)	(0.043)	(0.043)	(0.032)	(0.031)
TimeDifference	-0.016*	-0.017**	-0.014*	-0.015**	-0.016*	-0.013*
	(0.008)	(0.008)	(0.007)	(0.007)	(0.008)	(0.007)
Contiguity	-0.047	-0.059	-0.011	-0.020	0.035	0.083
	(0.054)	(0.056)	(0.060)	(0.061)	(0.110)	(0.118)
Colony	0.205***	0.200***	0.224***	0.222***	0.133***	0.056
	(0.048)	(0.048)	(0.053)	(0.053)	(0.050)	(0.045)
CommonLegal	0.114***	0.111***	0.100***	0.098***	0.125***	0.072***
	(0.029)	(0.029)	(0.033)	(0.033)	(0.032)	(0.024)
CommonReligion	0.049	0.033	0.066	0.052	-0.061	-0.028
	(0.093)	(0.098)	(0.096)	(0.099)	(0.112)	(0.081)
EU	0.148**	0.136**	0.097	0.092	0.010	-0.007
	(0.065)	(0.065)	(0.070)	(0.070)	(0.062)	(0.046)
LanguageSimilarity	0.051	0.239	-0.025	0.115	0.046	-0.040
	(0.128)	(0.168)	(0.146)	(0.192)	(0.182)	(0.147)
EnglishSimilarity		-0.232*		-0.182	0.609***	0.674***
		(0.130)		(0.133)	(0.194)	(0.178)
N (df)	1140 (1008)	1140 (1007)	1140 (1008)	1140 (1007)	53,580 (47,789)	484,500 (432,215)
Projected-R^2			0.22	0.22	0.12	.13
Method	NegBin2	NegBin2	OLS	OLS	OLS	OLS
Aggregated by	citing and publication year	citing and publication year	citing and publication year	citing and publication year	citing year	no aggregation

Notes: This table reports estimated coefficients from negative binomial and OLS regressions with cluster robust standard errors in parentheses. Standard errors are clustered at the citing country level. All models include citing and cited country dummies. The dependent variable is citation imports with one unit corresponding to the number of citing articles in any single country to articles written by authors in the ten leading research countries in a year between 1970 and 2016. 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

Chapter 3

Rise of multi-authored papers in economics: demise of the 'Lone Star' and Why?

3.1 Introduction

Much has been written about the phenomenon of co-authorship in economics in the last thirty years, as prior to this sole-authored papers were the main output by far in economics journals.[1] More recently, Henriksen (2016) documented the rise in co-authorship in the social sciences in general over the last 30 years, using a very comprehensive data set. Rath and Wohlrabe (2016)), building on Nowell and Grijalva (2011), and also using very large data sets, established that the upward trend is visible across all sub-disciplines in economics and across different quality levels of journals.

The purpose of this paper is to develop on the work above in three ways. First, to review in some detail the economics literature relating to this phenomenon. Second, we outline new large and different data sets used in this paper to explore co-authorship in economics. We focus on two specific sets of economic research journals: a broad set including the top 255 and a top journal sample including the 20 best ranked journals for the period 1996 to 2014. Previous work on co-authorship in economics, bar Rath and Wohlrabe (2016) and Sommer and Wohlrabe (2017), is limited to small subsamples of economic research. While these two recent papers use RePEc as their source of data, this paper uses Scopus and additional sources, thereby providing different and supporting new evidence for their

[1] See for example McDowell and Michael (1983), Barnett et al. (1988), Hudson (1996), Wuchty et al. (2007), and Card and DellaVigna (2013).

work.[2] Third, we trace several aspects of co-authorship including overall shares, citations received and given, pages, career age, and the listing of authors for each author category, ranging from one to four-plus authors, separately. We provide these statistics by journal sample and over time. This is followed by a brief analysis of how the new data inform the debate in the literature on the causes for the rise in co-authorship.

The remainder of this paper is organised as follows. Section 3.2 reviews the key literature relating to co-authorship in economics. Section 3.3 provides information on the data used. The empirical findings are discussed in Section 3.4 and Section 3.5 concludes the paper.

3.2 Key literature

There is an extensive economics literature on the phenomenon of the rise in co-authorship in journal articles. The purpose of this section is to review this literature, and to demonstrate how this provided the context for the later data collection and empirical analysis.

The earliest substantive paper perhaps to look at the phenomenon of the rise in co-authorship in the economics literature was McDowell and Michael (1983), but using just ten journals in their sample. Barnett et al. (1988) widened the discussion considerably, but using an even narrower data set, namely the AER alone. Their starting point is what they term the 'division of labour' hypothesis, very similar to the specialisation focus of the earlier McDowell and Michael (1983) paper and the later paper of Jones (2009a), and put succinctly as follows.

> Individuals engaged in economic research have found it increasingly possible (and, indeed, necessary) to specialize in more narrowly defined areas within the profession. As such specialization has proceeded, it has become increasingly necessary to combine the skills of two or more scholars in the conduct of research projects. For example, one who is highly skilled in the testing of hypotheses may find it attractive to collaborate with one skilled in generating hypotheses. Both, in turn, may find it attractive to combine their efforts with one skilled in collecting and organizing the data required to implement empirical tests. Thus, as specialization proceeds, we should expect to observe, over time, an increase in the incidence of co-authorship. (Barnett et al. (1988), p. 539)

[2] The main reason for this is that the Scopus data allowed for exploration of other work (see Kuld (2017)) not possible using the RePEc data.

Neither of these papers though looked at the breakdown of co-authorship between different numbers of co-authors. Besides they used narrow data sets.

Another argument is the increasing emphasis on publication in refereed journals as a criterion for appointment and promotion. The days of books or chapters in books, or policy reports counting towards a person's research record have been it appears in decline since the 1970s and have been largely replaced by verifiable 'scientifically-ranked' journals and citation records. Barnett et al. (1988) argue that this allows less time to assist colleagues, the 'reward' of an acknowledgement or 'thank you' being replaced with the offer of co-authorship to elicit such assistance. This is their opportunity cost of time hypothesis.[3]

Another hypothesis relates to 'risk-aversion', which says it is better to spread your risks by submitting say four quarto-authored papers than one solo-authored paper. If the emphasis on journal article publication has increased, such considerations would have assumed greater importance over time. Barnett et al. (1988) argue that the variance and hence randomness of the process for assessing articles submitted has increased and hence so has the incentive to diversify through co-authorship. The key argument for this assertion is the huge increase in the number of journals and hence the difficulty of finding suitable editors and referees.[4]

Sauer (1988) tested the hypothesis of a higher return on co-authored papers but found that an individual's return from a co-authored paper with n authors is approximately 1/n times that of a single-authored paper.[5]

Medoff (2003) examines the widely held belief that researchers who collaborate produce higher quality research than those who are sole-authors. Like for the other articles he used a small number of journals, eight in this case. The empirical results he argues show that, controlling for article length, journal and author quality, and subject area, collaboration does not result in significantly higher quality research (as measured by the number of citations an article receives) in economics. The key question though is it citations per

[3]It is not clear though that if books and reports no longer count that the total demands on research time, and hence the opportunity cost of time, should have risen. Books and reports in many cases would have taken up a huge amount of research time, time now 'free' for journal article research and assistance to colleagues.

[4]However, while it is true that there has been a huge increase in the number of journals and articles submitted, there has also been a large increase in the number of economists upon which to draw on for editorial and refereeing purposes (see Osterloh and Frey (2014)), implying no increase in work-load per referee and editor.

[5]One wonders would that be the case today. From anecdotal evidence it appears that the return on a co-authored paper today might be considerably more than 1/n times that of a solo-authored paper, which in itself would be a very strong argument for the rise of co-authorship. It is likely that the more cross-country the co-authorship the more likely is the chance that there is no discounting of multi-authored papers (see later).

article or citations per author which matters. And if there are more citations for co-authored articles, which there are, does the increase in citations compensate for the fact that it took three to four authors, rather than one, to effect this increase?

Rosenblat and Mobius (2004) argue that advances in communication and transportation technologies have the potential to bring people closer together and create a "global village" and hence more collaborative work. They develop a model which they test by looking at the evolution of academic co-authoring between 1969 and 1999. Several new technologies decreased the cost of communication substantially starting around 1980. First fax technology became ubiquitous in the 1980s: second, emailing and file transfer through FTP was common by the beginning of the 1990s; third and perhaps most importantly, the rise of the Internet in the 1990s made it dramatically easier to publish and search for working papers.[6] Moreover, deregulation of the US airline and telephone industries in the 1980s drastically decreased the cost of traveling and making long distance telephone calls. Their data set contains 8,838 authors of whom 6,201 authors published at least one co-authored papers. It is possible though that this simply altered the nature of the co-authorship rather than the quantities of articles co-authored. A wider network from which to choose should change the pattern of co-authorship but it does not imply without further argument that the incidence of co-authorship increases.[7]

Jones (2009a) took up the division of labour argument also, but applied in this case to scientists and engineers. His starting point is that while physical stocks can be transferred easily, as property rights, from one agent to another, human capital, by contrast, is not transferred easily. The vessel of human capital-the individual-is born with little knowledge and absorbs information at a limited rate, so that training occupies a significant portion of the life cycle. Moreover, if innovation increases the stock of knowledge, then the educational burden on successive cohorts of innovators may increase. Innovators might confront this difficulty through two basic channels. First, they may choose to learn more. Second, they might compensate by choosing narrower expertise. Choosing to learn more will leave less

[6] Agrawal and Goldfarb (2008) examined the effect of a decrease in collaboration costs resulting from the adoption of Bitnet (an early version of the Internet) on university research collaboration in engineering, their interest being the broader question of how changes in collaboration costs may affect the structure of knowledge production. They examined 270 universities that published in seven top electrical engineering journals from 1981 to 1991 and found that a Bitnet connection did seem to facilitate a general increase in multi-institutional collaboration (by 40 percent, on average). Catalini et al. (2016) built on the explanation that links the increase in co-authorship to the drastic reduction in communication costs brought by the internet: as coordination and communication costs go down, scientists are able to sustain collaboration over distance in a more efficient way. In this paper, they test a complementary hypothesis: that the increase in distant collaboration may also be the result of the dramatic reduction in air travel costs that took place within the United States over the last 30 years.

[7] The focus of Fafchamps et al. (2010) was linking the extent of co-authorship to networks. The stronger the networks the greater the degree of co-authorship. They also note though that networks maybe are not as important in determining co-authorship given the greatly increased access to the web. They also address a potential problem, namely the time between when collaboration commenced and when it is noted, namely in a publications.

time in the life cycle for innovation. Narrowing expertise, meanwhile, can reduce individual capabilities and force innovators to work in teams, namely be involved eventually in co-authored patents or in the case of economics co-authored journal articles. His empirical work looks at three issues resulting from what he calls the 'burden of knowledge and death of Renaissance man', namely team size, date of first innovation and specialisation.

Hamermesh (2013) in a broad overview mused about possible broad explanations for the rise of co-authorship. He also examines the issue of multiple authors, one of the first to do so in economics and argues that co-authorship can be more fun, but why should this have increased over time,[8] and he also refers to the increased opportunity cost of time in the 'rat race' to publish more and more journal articles. In this situation as noted by others already, the 'price' of getting feedback on your work might be the offer of co-authorship.

Ossenblok et al. (2014) analyse co-authorship patterns in the social sciences and humanities in for the period 2000 to 2010. The basis for the analysis is the Flemish Academic Bibliographic Database for the Social Sciences and Humanities (VABB-SHW). Standing out in this case study is the fact that the Flemish research-funding system actively encourages co-authorship through its use of whole counts, that is giving- each institution full credit for an article that bears its name and address. This is opposed to systems that use fractional counts, that is counting an article as a single unit and fractionalising the publication credit. They do not indicate though how this might have changed in Flanders or indeed anywhere else, but yet it could be a vital factor, not only in research funding but in the global ranking of universities and hence for hiring and promotion. If by adding another person benefits him/her and takes nothing away from you, it is clear that there will be a huge incentive to be involved with co-authored papers, the more authors the better.

Across the social sciences, Henriksen (2016) examines the rise in co-authorship over a 34-year period. The paper investigates the development in co-authorship in different research fields and discusses how the methodological differences in these research fields together with changes in academia affect the tendency to co-author articles. The study is based on bibliographic data on about 4.5 million peer-reviewed articles published in the period 1980-2013 and indexed in the 56 subject categories of the Web of Science's Social Science Citation Index. The results show a rise in the average number of authors, share of co-authored and international co-authored articles in the majority of the subject categories. However, the results also show that there are great disciplinary differences in the extent of the rises in co-authorship. The subject categories with a greater share of international co-authored articles have generally experienced an increase in co-authorship, but increasing

[8] Önder and Schweitzer (2016) also examine trends in co-authorship, highlighting the rise in papers with more than two authors, applied to PhD graduates from German-speaking countries.

international collaboration is not the only factor influencing the rise in co-authorship. Hence, the most substantial rises have occurred in subject categories, where the research often is based on the use of experiments, large data set, statistical methods and/or team production models. This then provides more descriptive evidence for the Barnett et al. (1988) and Jones (2009a) hypotheses.

Two recent papers provide an extensive data analysis for co-authorship in economics. In the first of these focusing on 28,000 articles by American economists between 1985 and 2004, Nowell and Grijalva (2011) study co-authorship by economic sub-field (JEL-Code) and journal category. The key findings are, first, that quantitative papers, which the authors assume to be technically more demanding, are more likely to be co-authored, a fact that the authors attribute to specialisation effects (see Barnett et al. (1988)). Second, higher ranked journals tend to have more co-authors. Third, co-authorship increases over time and differs between sub-fields with the extremes being economic history at the lower and financial and agricultural economics at the higher end.

Rath and Wohlrabe (2016) expand on this using an international and a much larger data set based on over 700,000 articles in 1,615 economics journals in RePEc. Their findings mainly confirm the results obtained by Nowell and Grijalva (2011) for this wider sample. In addition, the authors show, first, the overall shares of articles by number of authors over time. Second, the average number of authors in articles that include an author without a previous registered publication on RePEc is found to be similar to the overall number and follow the general trend to increased co-authorship.

Four main strands then emanate from this literature with regard to the causes of the rise in co-authorship in economics. First is the increasing specialisation and division of labour hypothesis addressed in several papers. Second is the role of decreased communication and travel costs associated with increased co-authorship, especially across countries. This implies that the ease of co-authorship has increased. Third, the incentives set by the research community have changed. On the one hand, this leads to an increased opportunity cost of time which in turn leads to the addition of more authors than was previously the case, either because it forced researchers to be more efficient through co-authorship or because it reduced the willingness for collaboration or help without full recognition as co-author. On the other hand, co-authorship would be a response to avoid the risk of no publication. This depends though on how multiple authorship is discounted by hiring and funding agencies.

3.3 Data sources

The research objective was to generate data, the analysis of which might inform the above debate, using a large number of economics journals and differentiated by quality as mea-

sured by impact factor. To identify such economic research publications, all journals ranked in Kalaitzidakis et al. (2011) are used. The over two hundred journals in this list were supplemented with a number of other economic journals, if highly ranked in Ideas RePEc (`ideas.repec.org`). From this process we ended up with 255 of the most influential journals in economics by citations received.

In total, 174,266 research articles were published in these journals between 1996 and 2014 and listed on Scopus (`scopus.com`). All of these are included in the data analysis. For each article, Scopus includes information on the authors, journal title, number of pages, year and the number of citations received.[9] These data then allowed us to examine co-authorship in its various dimensions, in particular co-authorship involving two, three and four plus individuals and the trends in each by different rankings of journals, citations per paper and per author, and average page length.

In addition, these data are used to identify the 1,000 most cited economists, a small subgroup of the total.[10] Using on-line CV data for each, a career profile was then constructed for all, but these data are used in this paper only to a very limited extent, namely to examine the different career-age profiles with regard to preferences/outcomes in relation to different types of authorship (solo, duo, treble or quarto-plus). To do this, we looked at the 133 top economists who completed their PhDs 1996 and 1999, inclusive, and then plotted the trend in their publications from year of graduation by type of solo and co-authorship.

To add data on international collaborations, a search on Web of Science (`webofknowledge.com`) for economics articles with author affiliations in the US and other specific countries was carried out. We use this additional source to study the rise of increased cross-country co-authorship, potentially due to technological change and cheaper travel, as it has been posited as seen earlier as a reason for the steep rise in co-authorship.

3.4 Empirical analysis

3.4.1 Trends in charts

Figure 3.1a provides the picture of the overall trends in co-authorship in economics. Ninety-five per cent confidence intervals are provided for each year, relating to yearly means. As recently as 1996 solo-authored papers accounted for fifty per cent of all articles published

[9] Other data bases include Google Scholar, RePEc and EconLit. Scopus (`scopus.com`) has full coverage of the selected journals from 1996.

[10] The primary purpose of our data-collection exercise was to obtain information on the career paths the most published economists in of the last twenty years (see Kuld (2017)).

in our sample, with this number dropping to just over twenty-five per cent in 2014. While duo-authored papers share of the total remained steady, the huge pickup was in trio and quarto-plus authored papers, particularly the latter. By 2014 quarto-plus authored papers accounted for around eight per cent of the total, and trio authored papers for around twenty-five per cent of the total, more than double that of less than fifteen years earlier. This is a remarkable turnaround in a very short period. The picture is replicated whether the data relate to all journals (top 255, Figure 3.1a) or the top 20 journals (Figure 3.1b), but different trends are evident. The rise of trio and quarto-authorship is particularly marked in the top 20 journals, with just over twenty per cent now solo authored, with the number of trio and duo authored papers exceeding the total of single authored papers by a wide margin.[11] If present trends continue the number of quarto-plus authored papers could soon exceed the number of single authored papers.[12]

Turning now to the trends in co-authorship by country, we look at the pattern examining co-authorship between US researchers and economists from other countries.[13] This focus on American collaborations is motivated by the ongoing dominance of US universities in journal article output in economics. Figure 3.2 plots the percentage of each category of cross-country co-authored papers as a percentage of the total paper output of the two relevant countries combined. As such, this is a relative measure and hence a better indicator of trends.

Since 1990 there has been a huge rise in co-authorship across countries, especially between the US and the UK. The rises for the other country combinations though are large, especially as they are expressed in percentage terms. Of particular note is the rise in co-authored papers with China: their share expressed as a percentage of the total number of articles in both countries has risen from close to zero just fifteen years ago to almost four per cent in 2014, These trends may give us an insight again into the possible reasons for the rise in co-authorship to be looked at later. It is noteworthy though that the highest shares apply to US-UK and US-Canada articles, reflecting perhaps strong cultural connections, not least language. Still, the increases in US-China, US-Germany and US-France co-authored papers has been marked, with the gap between them and US-UK and US-Canada set to close in years to come perhaps.

Turning now to citations per article by co-authorship type, Figure 3.3a shows citations per article in the top 255 journals, *relative* to the yearly mean. It shows that citations per article are consistently highest for quarto-authored and lowest for single-authored articles.

[11] Rath and Wohlrabe (2016) broke the 1,615 journals used into quartiles using RePEc impact factors and found that the rises in all quartiles were the same, with the average number of authors in the top three quartiles very similar. The top quartile though includes over 400 journals.

[12] It is reassuring that similar trends were found in Rath and Wohlrabe (2016), even though using very different data and categories of authors.

[13] It would be interesting to expand this work to other pairwise comparisons (see Henriksen (2016)).

The picture with regard to the top 20 journals is similar but with differences (Figure 3.3b). As can be seen there is huge variability in the citations for quarto-authored papers (mainly due to the very small number of articles in the earlier years). Again, in almost all years their level is higher than those for trio and duo-authored papers, but with very wide confidence intervals evident at times. For every year citations are lowest for single-authored papers.

However, when citations per author are considered a very different picture emerges (Figure 3.4). As can be seen citations per article per author are much higher for single-authored papers and this is an alternative, arguably better, indicator of the contribution of an individual to the field. This is true no matter which category of journal is used. However, citation counts are no absolute indications of influence or even quality which makes such arithmetic difficult. The interesting question is how funding agencies and hiring/promotional bodies view co-authored versus single-authored papers.

Are journal articles getting longer and is there much variation by degree of co-authorship? Figure 3.5a illustrates that for the top 255 journals there was a large rise in the number of pages per article up to the early 2000s but large declines following this, with rises again in recent years. The pattern is very similar across all author types, with on balance no increase in the number of pages over the whole period regardless of author type. It is also noteworthy that the number of pages differs by just one to three pages across author types, differences which might be considered very small. This implies that the number of pages per author is much higher for solo-authored paper. The picture with regard to the top 20 journals is different in some respects (Figure 3.5b). Overall there have been increases in the number of pages regardless of author type.

A related issue is the number of cited references, the trends in which are outlined in Figure 3.6. There has been a large increase in the number of references, in a very short period, particularly in the trio and quarto-plus authored papers. For example, the number of references in the quarto-plus authored papers was around 23 with the number rising to over 40 by 2014. Similar increases are evident for the other categories in Figure 3.6, with the average number of references for solo authored papers rising from 22 to 35 in the same period. The observed differences in article length and references given are in all categories much smaller than the differences in citations received.[14]

Another issue related to the later discussion is the alphabetical ordering of author names on the articles by author type. Figure 3.7 shows that there is a high proportion of articles

[14]See Card and DellaVigna (2013) and Card and DellaVigna (2014), for discussions of trends in article length, but not from a co-authorship perspective).

using alphabetical ordering of names on the papers, even if adjusted for random alphabetical ordering.[15] The figure is around 60 per cent for duo and trio authored papers, but only around 40 per cent for quarto-plus authored papers. However, and of importance to the later argument, these percentages slightly increased over the period examined. The alphabetical ordering of names is particularly high in the top 20 journals, with no significant differences evident by number of authors (apart from one downward spike). This implies that the contribution of each author is *signalled* to be approximately equal which increases the costs of token adding of names. In addition, this makes it impossible to directly identify roles within the author team, for instance the lead author listed first or the group supervisor listed last.

One final chart constructed to throw light on the phenomenon of the rise in co-authorship relates to the career profile of 133 top economists who were awarded their PhD between 1996 and 1999, the data for which were discussed earlier. It is not clear though how representative this sample might be, but the trends are nonetheless instructive.[16] Figure 3.8 plots the articles by number of authors for these top 133 economists in the years following their PhD graduation. For each year and team size, the number of publications is divided by the total number of authors published in top 20 journals in the same year and with the same number of co-authors.[17] As can be seen in Figure 3.8, the share of solo-authored papers by these economists is highest and rises in the first five career years and thereafter declines steadily for the following ten years. In later years the shares of the categories converge which implies a publication pattern similar to the group of all top 20 journal authors.

3.4.2 Regressions: testing key trends

In this subsection, quasi-Poisson and logit regressions are used to build on the observed trends noted above. This allows to formally test the significance of observed differences while partly controlling for unobserved trends in economic publications using journal fixed effects. We estimate and show in Table 3.1 the relation between co-authorship and key outcomes, namely citations received, references given and pages published.

[15] Ordering by contribution can lead to unintentional alphabetical listings. For instance, a random order is alphabetical in 50 percent of duo-authored articles. The share of non-alphabetical duo-authored papers is, therefore, taken out of half of the sample. Accordingly, trio-authored papers are related to five-sixths of the sample and quarto-authored papers to 23/24.

[16] Rath and Wohlrabe (2016) looked briefly at this issue, from a different perspective. They used the first journal article to examine whether or not the average number of papers by 'scientific rookies' is smaller than the overall average but find that this is not the case. The analysis in our paper differs by following authors over their career.

[17] The top 20 journals are picked as the reference group as said economists are highly cited. However, taking all journals does not alter the conclusion.

$$\mathbb{E}\left(Y_i|\#A_i, T_i, J_i\right) = g^{-1}(\gamma_t T_i + \gamma_j J_i + f(\#A_i, J_i)) \tag{3.1}$$

Where, Y_i is either citations received, references given, pages published, or alphabetical order of article i. $\#A_i$ is the number of authors, and T_i and J_i are a dummies for publication year and the article's journal. g is the used link function, that is the natural logarithm for the quasi-Poisson regressions (columns (1) to (3)) and the logit function for the alphabetical listing in specification (4). f is a linear function of the number of authors interacted with a dummy to denote top journals.[18] We estimate the effect of two or more, three or more, and four or more authors for top 20 and other journals. In this specification, we estimates, therefore, whether an additional author has a significant effect compared to articles with one author less (up to the forth co-author). We also estimate whether there is a significant difference between top journals and other journals for each author team size. For convenience, we also express all co-author categories as differences to single authorship in an alternative specification (Table 3.2). The regressions are identical otherwise. All regressions include dummies for year and journal and robust standard errors clustered at the journal level.

Overall, co-authored articles are consistently longer and reference more articles than single-authored papers but the differences are small in size. The substantial increase in pages occurs with the addition of the first co-author. The estimated effect of this co-author is a six percent increase in pages (Table 3.1, column (1)). The addition of further authors is associated with only a modest increase of around one percent, that is less than a half a page on average. The estimated effects on references given are similar but again smaller (columnn (2)). The estimated 62 percent increase in references overall from 1996 to 2014 is substantially larger than the differences between author categories.[19] In addition, the interaction of a linear time trend with the number of authors indicates a slight decrease in the effect of co-authorship on the number of pages and a slightly increased effect on references given (see Table 3.3).

Duo, treble and quarto authored papers are associated with 30 to 90 percent more citations received than single-authored papers (Table 3.1, column (3)). The effect of the first co-author is again larger than the effect of further co-authors. The estimated effect of co-authorship does not differ significantly between publications in top 20 journals and other publications.[20] Since we use dummies to account for differences between journal means, the estimates indicate a surprisingly strong variation by author number in the expected mean

[18] $f(\#A_i, J_i) = \beta_1 DuoPlus + \beta_2 DuoPlus + \beta_3 TreblePlus : Top20 + \beta_4 DuoPlus : Top20 + \beta_5 DuoPlus : Top20 + \beta_6 TreblePlus : Top20$

[19] The estimated percentage effect for $\gamma_{2017} = 0.485$ is calculated as $(e^{0.485} - 1) * 100 = 62.4$.

[20] Table 3.3 indicates that this correlation between co-authorship and citations received weakens over the years. However, this might be caused by the overall compression of citation counts.

within a journal. We further investigate this relation using four individual logit regressions to estimate the probability of an article to be sole, duo, treble, or quarto-plus authored (Table 3.4) based on the citations they received. To allow for heterogeneous effects over the distribution of citations received, we create six relative citation brackets, that is, for instance, articles that received fewer citations than the median article in the publication year but more than the article at the first quartile. The estimated parameters for the associated dummies are plotted in Figure 3.10. This shows that the positive association between co-authorship and citations received holds throughout the distribution of citations received and when controlling for differences in journal and year means as well as references given and pages published.

These findings confirm the positive relation between co-authorship and citations received which was observed by Sommer and Wohlrabe (2017) for the full sample of RePEc publications.[21] We add that this relation holds within journals, for each additional co-author up to four, and over the full distribution of citations received up to the top percentile, that is the 95 most cited articles per year.

In addition, we repeat the regressions shown in Table 3.4 using only publication year and journal dummies. Arising from the regressions, odds ratios are constructed for the rise in articles with different number of authors and plotted for the publication year in Figure 3.9. Figure 3.9 again confirms and markedly so the very large rise noted earlier in the share of quarto-plus authored. For instance, the odds that a paper has four or more authors is 4.4 times as high in 2014 than 1996. The share of trio-authored papers also increased considerably with the odds that a 2014 article is trio-authored being 2.6 times as high as that in 1996. The share of duo-authored papers increased less (10 to 20 percent over the years) and a marked decline is observed in the share of solo authored papers where the odds that a 2014 publication is single-authored are only half the odds of a 1996 article to being single-authored.

3.4.3 Reflection on explanations

The empirical evidence above is used in this section to reflect on the explanatory hypotheses outlined in Section 3.2. First, the necessity and ease with which to specialise within a research team are discussed. Second, the role of decreasing communication costs is linked

[21] Sommer and Wohlrabe (2017) and Moosa (2017) examine the relation of co-authorship and citations received. Moosa (2017) studies the 300 most cited articles on RePEc and find no effect of co-authorship on citations received. Sommer and Wohlrabe (2017) repeat this exercise by varying the sample of articles by citations received and conclude that a negative effect of co-authorship is only visible for highly cited articles (though not significant). Taking the full sample of RePEc papers leads to a positive relation between co-authorship and citations received.

to the increase in international collaborations. Third, the pressure to publish is discussed in the light of the surprisingly high incidence of solo-authorship among young researchers.

Specialisation in the context of co-authorship relates to the benefit of specialised roles within an author team. These should be higher the easier it is to divide research roles and the wider is the necessary research knowledge. For instance, Nowell and Grijalva (2011) observe more co-authorship in more quantitative sub-fields of economics. It may be easier to separate quantitative research - or research that contains a quantitative and a theoretical part - than a purely theoretical work. Or, the former may necessitate two or more specialists because a single researcher is less likely to possess all necessary knowledge as discussed earlier in relation to the work of Barnett et al. (1988), McDowell and Michael (1983) and Jones (2009a).

Our empirical analysis though gives no strong evidence for this hypothesis. Comparing published research, we observe no increased return of co-authorship over time if measured by citations received or pages per article (see Figures 3.3, 3.4 and 3.5 and Table 3.3). Only the increased share of co-authored publications itself might indicate a higher return to co-authorship. We also do not find a trend towards research teams in which team roles or contribution are signalled by the ordering of names. Figure 3.7 shows the high and unchanging incidence of the alphabetical listing of authors.

The Internet and cheap flights did lower the costs of communication and, subsequently, co-authorship between distant researchers. The evidence in relation to co-authorship across countries and in relation to the huge rise in the number of citations would tend to support the argument that technology and transport costs may have been key factors (see Figures 3.2 and 3.6). However, as stated above, while a wider network from which to choose should change the pattern of co-authorship, it does not necessarily increase the incidence of co-authorship. More importantly, technological progress simplifies the sharing of research with all, including spatially closer, colleagues. As noted above the ease of dividing research into separate tasks should increase the incidence of co-authorship.

If the pressure on economists for more articles has increased over time, researchers can respond by increased co-authorship. First, shared work should be less time consuming than solo-authoring if there are gains from the division of labour. Second, if co-authored papers are not discounted by the number of co-authors in the assessment, for instance during hire or promotion decisions, this would provide incentives to co-author (see the discussion of Ossenblok et al. (2014) above). In addition, co-authorship diversifies the risk of individual research projects failing. However, the high incidence of solo authored papers in the early career stage points towards the ambiguity of co-authorship in the hiring process (see Figure 3.8). Young economists might react to a perceived disadvantage of co-authorship. Sarsons (2017b) shows that in tenure decisions, women receive less credit for

co-authored work than men. It is likely that young authors are equally perceived as not fully contributing to co-authored work and, therefore, choose to solo-author.[22] Again, other explanations such as a smaller network of potential co-authors are possible though.[23]

3.5 Concluding comments

The dramatic rise in multi-authored papers in economics as outlined in Section is at the very minimum of interest to economists. Explaining these trends is a different matter. As always in economics, several different factors are at work simultaneously and despite claims to the contrary, holding for fixed effects and using other techniques, simply cannot overcome the reality that when variables are all moving together it is nigh impossible to separate the effects.

A related problem is that many of the key variables cannot be measured and hence have to be excluded from the formal regression analysis. This is particularly the case given the huge variety of factors posited for the trends in co-authorship in science over the last forty years.

We can show though the substantial trend to multi-authorship in economics, with top ranked journals leading the way but increasing throughout the economic literature. Conversely, we do not observe a trend towards vertically differentiated author teams that signal role or contribution by order of names. There is also no evidence for time trends in the observed differences in pages, references given, and citations received by number of authors. This observed unchanging but high correlation between the number of authors and citations received even within journals is difficult to explain. However, it is important to note the evidence for a large share of single-authored work among the most cited articles (see Sommer and Wohlrabe (2017) and Moosa (2017)) and by young top economists shown in this paper.

What is needed more perhaps is more evidence on hiring, promotional and funding decisions with regard to solo versus multi-authored papers. The patchy evidence would seem to suggest that there is limited discounting of a published article by number of co-authors except perhaps for young or female authors (see above and Sarsons (2017b)). If this has

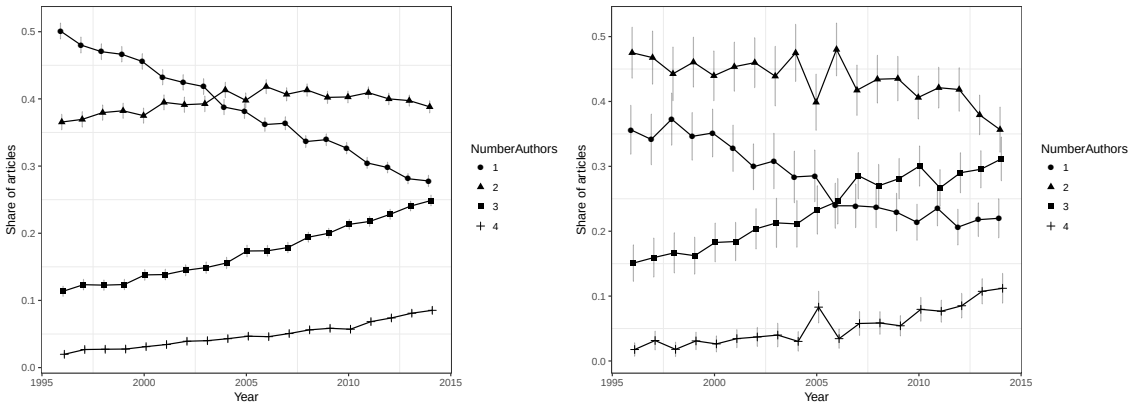
[22] In a follow-up version of this paper, Sarsons (2017a) finds no such discrimination in sociology, a discipline in which authors are listed according to contribution. Interestingly, sociology articles, though, have fewer authors on average (see Henriksen (2016)).

[23] The focus of Fafchamps et al. (2010) was linking the extent of co-authorship to networks. The stronger the networks the greater the degree of co-authorship. They also note though that networks maybe are not as important in determining co-authorship given the greatly increased access to the web. They also address a potential problem, namely the time between when collaboration commenced and when it is noted, namely in a publication.

further decreased over time then perhaps a key part of the explanation may be found here. A rising pressure to publish would suggest that the risk-sharing argument might have considerable validity, in particular if articles are not fully discounted for number of authors.[24] The increased ease and cheapness of electronic communication, plus greatly reduced cost of travel, that has opened up greater possibilities for collaboration. It is in relation to these factors that the most substantial evidence has been provided, but this may simply arise from the fact that these are easier to measure. As noted earlier, it is possible also to argue that while a wider network from which to choose should change the pattern of co-authorship it does not necessarily increase the incidence of co-authorship.

3.A Figures

Figure 3.1: Share of articles by number of authors

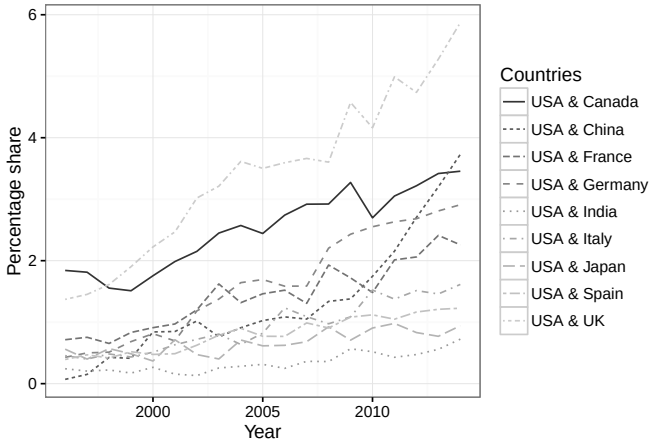


(a) Top 255 journals (b) Top 20 journals

Notes: Number of economic research articles published in a top 255 or top 20 journal as described in text, classed by number of authors and divided by the yearly total number of articles. 95 % confidence intervals as vertical lines.
Source: Own calculations based on Scopus data.

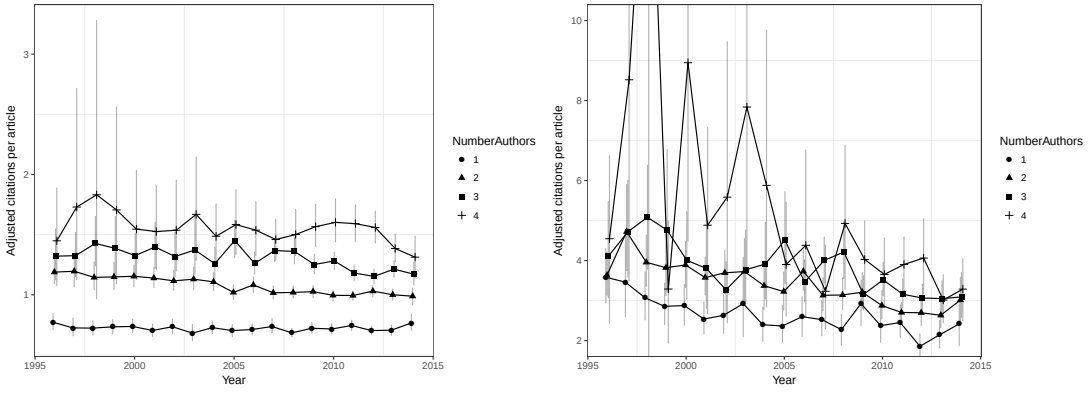
[24] See Osterloh and Frey (2014) for a general discussion on the use of citations and rankings in economics, in particular the randomness of some of the reviewing processes. Even if this always existed to a certain extent, the non-discounting of multi-authored papers would mean that the latter would be a very useful way of countering this randomness without any loss of individual/institutional ranking.

Figure 3.2: Co-authored papers across countries relative to combined output



Note: Number of economics articles with author affiliations in both respective countries divided by the yearly combined article output of these countries. Affiliation countries as indexed by Web of Science. *Source:* Own calculations based on Web of Science data.

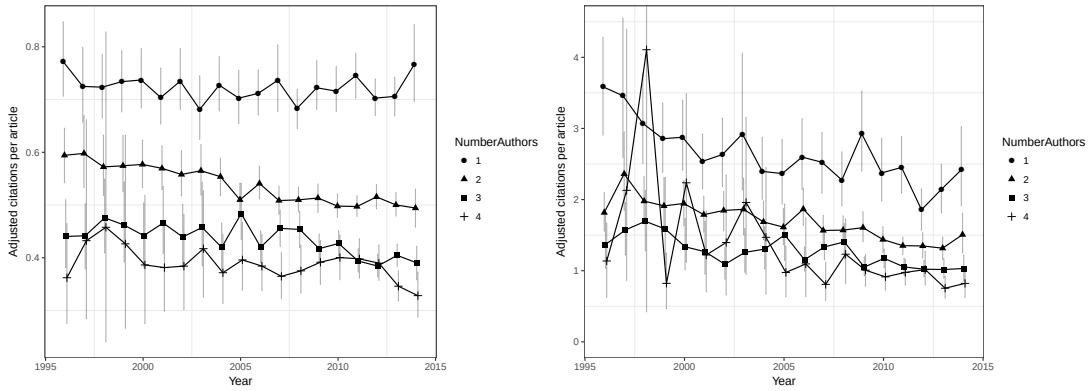
Figure 3.3: Citations received per article



(a) Top 255 journals (b) Top 20 journals

Notes: Means of citations to economic research articles published in a top 255 or top 20 journal as described in text, by number of authors and divided by the yearly mean of citations received per top 255 journal article. Bootstrapped 95 % confidence intervals as vertical lines. *Source:* Own calculations based on Scopus data.

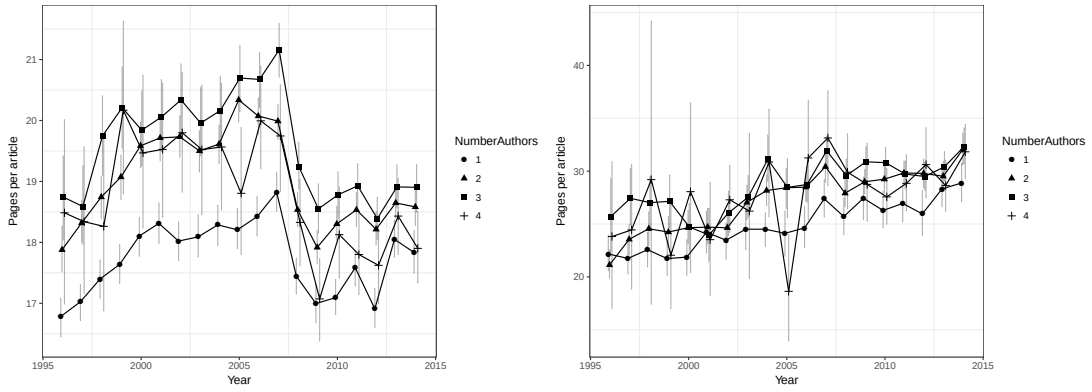
Figure 3.4: Citations received per author



(a) Top 255 journals (b) Top 20 journals

Notes: Means of citations to economic research articles published in a top 255 or in top 20 journal as described in text, by number of authors and divided by the number of authors and the yearly mean of citations received per top 255 journal article. Bootstrapped 95 % confidence intervals as vertical lines. *Source:* Own calculations based on Scopus data.

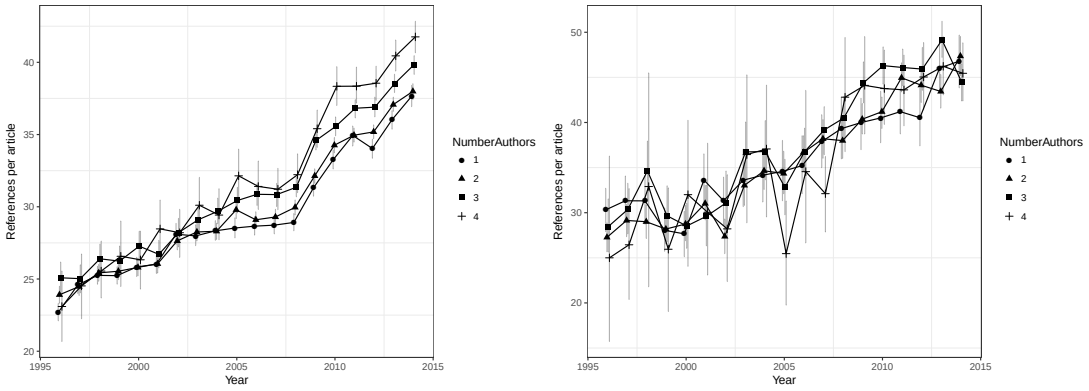
Figure 3.5: Number of pages



(a) Top 255 journals (b) Top 20 journals

Notes: Means of number of pages of economic research articles published in a top 255 or in a top 20 journal as described in text. Bootstrapped 95 % confidence intervals as vertical lines. *Source:* Own calculations based on Scopus data.

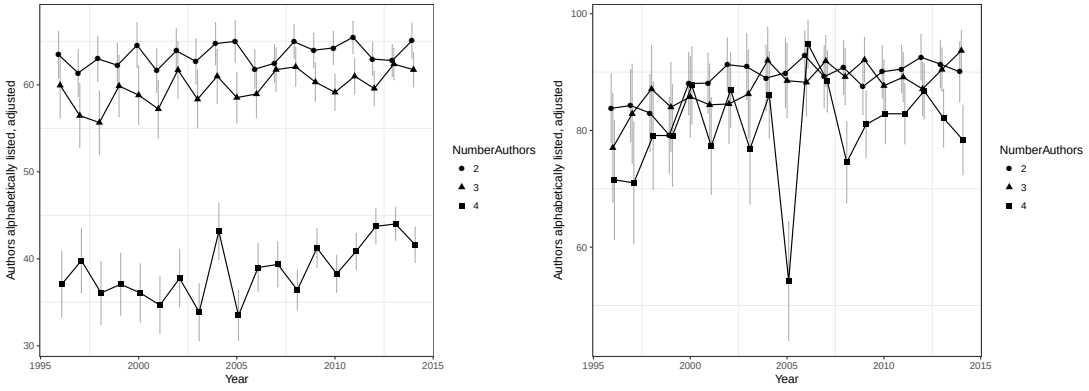
Figure 3.6: Number of references given



(a) Top 255 journals (b) Top 20 journals

Notes: Means of number of references in economic research articles published in a top 255 or in a top 20 journal as described in text. 95 % confidence intervals as vertical lines. *Source:* Own calculations based on Scopus data.

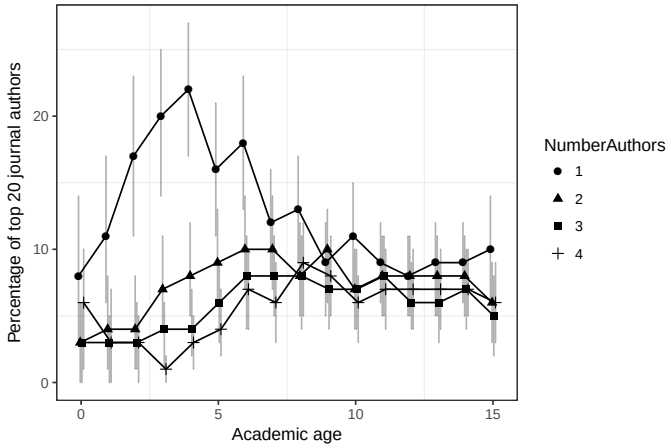
Figure 3.7: Alphabetical ordering of authors



(a) Top 255 journals (b) Top 20 journals

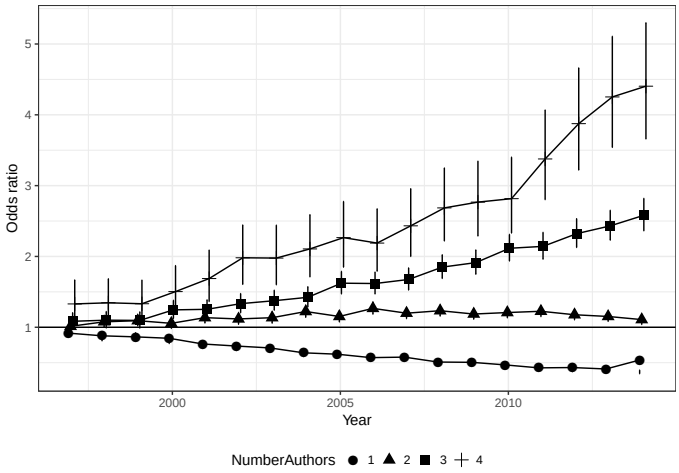
Notes: Percentages of alphabetically listed authors by number of authors and journal category. Percentage adjusted for different probabilities for random alphabetical ordering between different author group sizes. 95 % confidence intervals as vertical lines. *Source:* Own calculations based on online CV and Scopus data.

Figure 3.8: Relative shares by career age



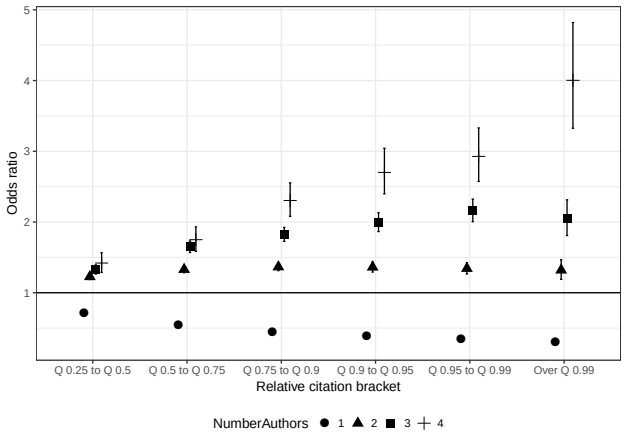
Note: Number of articles by year following award of PhD, as percentage of total authors of articles in top 20 journals. The number of total authors is the number of articles in top 20 journals multiplied by their umber of authors. Based on 3,874 articles by 136 highly cited economists who were awarded their first PhD between 1996 and 1999. *Source:* Own calculations based on the authors' on-line CVs and Scopus data.

Figure 3.9: Odds ratios for year of publication



Note: The graph shows estimated odds ratio for publication years estimated in logit reggressions as reported in Table 3.4. The x-axis lists the year of publication of an article. *Source:* Own calculations based on Scopus data.

Figure 3.10: Odds ratios for citations brackets



Note: The graph shows estimated odds ratio for dummies that refer to quantile brackets of citations received by an article relative to other articles with the same publication year. For instance, "Q.25-Q.5" indicates articles that received less citations than the median but more than the first quartile article in the publication year. More information is given in Table 3.4. *Source:* Own calculations based on Scopus data.

3.B Tables

Table 3.1: Number of pages, references given, citations received, and alphabetical listing

	Dependent variable (link: log (1-3) and logit (4)):			
	Pages	References given	Citations received	Alphabetical order
	(1)	(2)	(3)	(4)
DuoPlus	0.056***	0.035***	0.252***	
	(0.005)	(0.006)	(0.017)	
TreblePlus	0.013***	0.023***	0.133***	-0.877***
	(0.003)	(0.005)	(0.016)	(0.022)
QuartoPlus	0.013**	0.011	0.141***	-0.880***
	(0.007)	(0.007)	(0.031)	(0.050)
DuoPlus:Top20	0.041**	-0.017	0.017	
	(0.019)	(0.020)	(0.047)	
Top20:TreblePlus	0.033**	0.021*	-0.048	0.167**
	(0.016)	(0.012)	(0.031)	(0.081)
Top20:QuartoPlus	-0.027**	-0.052***	0.150	0.162
	(0.014)	(0.015)	(0.114)	(0.203)
1997	0.020*	0.048**	-0.019	-0.028
	(0.011)	(0.019)	(0.050)	(0.058)
1998	0.049***	0.075***	0.007	0.027
	(0.011)	(0.022)	(0.037)	(0.061)
1999	0.063***	0.082***	-0.016	0.051
	(0.011)	(0.024)	(0.044)	(0.063)
2000	0.079***	0.101***	-0.035	0.103
	(0.012)	(0.022)	(0.033)	(0.064)
2001	0.095***	0.110***	-0.058	-0.023
	(0.013)	(0.024)	(0.040)	(0.058)
2002	0.088***	0.171***	-0.124***	0.072
	(0.014)	(0.024)	(0.047)	(0.060)
2003	0.104***	0.200***	-0.174***	0.068
	(0.015)	(0.023)	(0.036)	(0.066)
2004	0.113***	0.200***	-0.294***	0.163***
	(0.016)	(0.023)	(0.043)	(0.062)
2005	0.119***	0.228***	-0.380***	0.093
	(0.015)	(0.023)	(0.047)	(0.061)

2006	0.116***	0.221***	-0.544***	0.086
	(0.017)	(0.025)	(0.040)	(0.065)
2007	0.126***	0.224***	-0.688***	0.133**
	(0.019)	(0.026)	(0.042)	(0.059)
2008	0.068***	0.252***	-0.822***	0.205***
	(0.026)	(0.024)	(0.043)	(0.056)
2009	0.019	0.319***	-1.056***	0.202***
	(0.030)	(0.024)	(0.047)	(0.059)
2010	0.035	0.371***	-1.363***	0.129**
	(0.030)	(0.023)	(0.047)	(0.064)
2011	0.055*	0.402***	-1.744***	0.213***
	(0.032)	(0.025)	(0.047)	(0.053)
2012	0.048	0.414***	-2.282***	0.136**
	(0.031)	(0.024)	(0.049)	(0.058)
2013	0.074**	0.451***	-3.111***	0.164**
	(0.031)	(0.025)	(0.058)	(0.065)
2014	0.078**	0.485***	-4.490***	0.198***
	(0.033)	(0.024)	(0.066)	(0.062)
N (df)	170470 (170192)	173449 (173171)	173449 (173171)	109541 (109265)

Notes: This table reports estimated coefficients from quasi-Poisson and logit regressions with cluster robust standard errors in parentheses. The estimations include journal dummies and standard errors are clustered at the journal level. Each observation is an economic research article. The variables relate to the number of references given to other articles the number of pages, a dummy for alphabetical listing of authors, as well as the publication year, and whether the publishing journal is among the twenty most cited journals (per article and relative to yearly citations).

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

Table 3.2: The estimates with single-authorship as reference

	Dependent variable (link: log (1-3) and logit (4)):			
	Pages	References given	Citations received	Alphabetical order
	(1)	(2)	(3)	(4)
TwoAuthors:OtherJournal	0.056***	0.035***	0.252***	
	(0.005)	(0.006)	(0.017)	
ThreeAuthors:OtherJournal	0.069***	0.058***	0.385***	-0.877***
	(0.006)	(0.009)	(0.023)	(0.022)
FourAuthors:OtherJournal	0.083***	0.069***	0.526***	-1.757***
	(0.009)	(0.011)	(0.037)	(0.061)
TwoAuthors:TopJournal	0.097***	0.018	0.269***	
	(0.018)	(0.019)	(0.043)	
ThreeAuthors:TopJournal	0.143***	0.062***	0.354***	-0.717***
	(0.029)	(0.023)	(0.046)	(0.074)
FourAuthors:TopJournal	0.129***	0.021	0.645***	-1.472***
	(0.031)	(0.022)	(0.092)	(0.185)
N (df)	170470 (170192)	173449 (173171)	173449 (173171)	109541 (109265)

Notes: This table reports estimated coefficients from quasi-Poisson and logit regressions with cluster robust standard errors in parentheses. The estimations include journal and year dummies and standard errors are clustered at the journal level. Each observation is an economic research article. The variables relate to the number of references given to other articles the number of pages, a dummy for alphabetical listing of authors, as well as the publication year, and whether the publishing journal is among the twenty most cited journals (per article and relative to yearly citations).

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

Table 3.3: Trends in pages, references, and citations received

	Dependent variable (link: log):			
	Pages	References given	Citations received	Citations adjusted
	(1)	(2)	(3)	(4)
Authors2	0.078***	0.020	0.346***	0.304***
	(0.011)	(0.015)	(0.024)	(0.030)
Authors2:Top20	0.006	-0.057*	-0.023	-0.002
	(0.031)	(0.034)	(0.084)	(0.083)
Authors2:Year	-0.002*	0.001	-0.013***	-0.005**
	(0.001)	(0.001)	(0.002)	(0.002)
Authors2:Top20:Year	0.002	0.003	0.007	0.001
	(0.003)	(0.002)	(0.008)	(0.007)
Authors3	0.116***	0.025	0.552***	0.450***
	(0.018)	(0.019)	(0.035)	(0.041)
Authors3:Top20	0.058	-0.033	-0.052	-0.040
	(0.051)	(0.035)	(0.093)	(0.099)
Authors3:Year	-0.004**	0.003**	-0.024***	-0.008***
	(0.002)	(0.001)	(0.003)	(0.003)
Authors3:Top20:Year	-0.002	0.003	0.007	0.004
	(0.004)	(0.003)	(0.009)	(0.009)
Authors4	0.174***	-0.006	0.759***	0.567***
	(0.028)	(0.026)	(0.066)	(0.071)
Authors4:Top20	-0.100	-0.122	0.297	0.333
	(0.068)	(0.092)	(0.205)	(0.211)
Authors4:Year	-0.007***	0.006***	-0.031***	-0.007
	(0.002)	(0.002)	(0.004)	(0.005)
Authors4:Top20:Year	0.007	0.005	-0.019	-0.021
	(0.005)	(0.006)	(0.016)	(0.015)
Year	-0.000	0.024***	-0.104***	0.010***
	(0.002)	(0.001)	(0.002)	(0.003)
Top20:Year	0.018**	0.000	-0.018**	-0.026**
	(0.008)	(0.005)	(0.009)	(0.010)
N (df)	170470 (170202)	173449 (173181)	173449 (173181)	173449 (173181)

Notes: This table reports estimated coefficients from quasi-Poisson regressions with cluster robust standard errors in parentheses. The estimations include journal dummies and standard errors are clustered at the journal level. Each observation is an economic research article. The variables relate to the number of references given to other articles, pages and citations received, as well as the publication year set to one in 1996, and whether the publishing journal is among the twenty most cited journals (per article and relative to yearly citations).

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

Table 3.4: Logit regressions for number of authors

	Dependent variable (link: logit):			
	1 Author	2 Authors	3 Authors	4+ Authors
	(1)	(2)	(3)	(4)
RelCit Q.25-Q.5	-0.271***	0.143***	0.144***	0.121***
	(0.019)	(0.018)	(0.028)	(0.045)
RelCit Q.5-Q.75	-0.531***	0.230***	0.353***	0.332***
	(0.022)	(0.020)	(0.026)	(0.044)
RelCit Q.75-Q.9	-0.710***	0.233***	0.450***	0.596***
	(0.028)	(0.023)	(0.029)	(0.049)
RelCit Q.9-Q.95	-0.754***	0.176***	0.477***	0.636***
	(0.037)	(0.028)	(0.031)	(0.058)
RelCit Q.95-Q.99	-0.918***	0.193***	0.572***	0.770***
	(0.048)	(0.031)	(0.038)	(0.078)
RelCit Over Q.99	-0.980***	0.218***	0.518***	1.014***
	(0.060)	(0.054)	(0.054)	(0.104)
NumberReferences	0.001*	-0.002***	0.001	-0.001
	(0.001)	(0.000)	(0.001)	(0.001)
NumberPages	-0.020***	0.010***	0.011***	0.013***
	(0.002)	(0.001)	(0.001)	(0.003)
N (df)	170470 (170190)	170470 (170190)	170470 (170190)	170470 (170190)

Notes: This table reports estimated coefficients from Logit regressions with cluster robust standard errors in parentheses. The estimations include journal and year dummies. Standard errors are clustered at the journal level. Each observation is an economic research article. RelCit variables refer to quantile brackets of citations received by an article relative to other articles with the same publication year. For instance, "Q.25-Q.5" indicates articles that received less citations than the median but more than the first quartile article in the publication year. The other variables relate to the number of references given to other articles and the number of pages, as well as the publication year.

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

Chapter 4

The university as a local idea space: benefits from research links

4.1 Introduction

Thematic research clusters are regularly designated at universities or set out in national research programs.[1] However, the merit of a local cluster is not obvious in fields without costly infrastructure. The recent economic literature observes no general localized productivity spill-overs in academic research such as benefits for a researcher's productivity from university quality or peer effects within departments. In turn, in the absence of local spill-overs, universities may be better off diversifying their research program to increase the variation in research and teaching.

The focus of this essay is, more narrowly, on the benefit of research linkages within a university department. Comparing the reception of a researcher's published articles, this paper tests whether articles are taken up more widely by future research than comparable work if the research of local colleagues is referenced. Hence, the conceptual definition of linked research is a research article that utilizes preceding research by a colleague or studies a related subject. The empirical analysis uses a sample of highly-cited economists. In turn, a leading researcher often constitutes a small cluster including, for instance, other less prominent researchers and graduate students. Any measured impact by a referenced top researcher might, therefore, be indirect. Thus, the empirical analysis focuses on the benefit added by a local cluster, that is thematically linked colleagues, in the production and dissemination of economic research.

[1] For instance, between 2006 and 2013, France restructured its public research and formed thematic and geographic clusters, among others the Paris and the Toulouse School of Economics.

The main findings in this essay indicate that the take up of a cutting-edge or high-profile (by the standard of the researcher) research article by future publications is significantly higher if the article references research by colleagues at the same university. This is in line with recent research that shows positive spill-overs between researchers who are both personally *and* thematically connected, in particular co-authors, on individual productivity.[2] The estimates in this paper complement these previous studies and show benefits from local research linkages below the level of co-authorship or PhD supervision. Similar to previous research, the paper finds no effect by not thematically linked local colleagues. Interestingly, the paper also finds no evidence for dissemination benefits for less prominently published "routine-type" research. Taken together, the results support a model of scientific production with positive spill-overs between thematically connected colleagues within a university. However, these benefits are limited to research that is by the standard of the researcher highly published. This empirical classification is motivated with a conceptual model in which an individual researcher produces both, "high-profile" or "cutting-edge" research which is understood to be more challenging and to require more novel research knowledge, and simpler "routine-type" research.

This paper expands on recent empirical work on location and peer effects in science. In the existing literature, the total annual article output of researchers is related to a change in the respective peer group or institution such as a change in department, co-author network of sub-field (see for example Agrawal et al. (2014), Borjas and Doran (2015), Azoulay et al. (2010), Waldinger (2010), Waldinger (2012) and Waldinger (2016))[3]. Most of these studies find no overall effect of researchers on the productivity of local colleagues or vice versa. With a focus on economics, two recent articles also observe no general localized effects within university departments in the last twenty years (Kim et al. (2009) and Bolli and Schläpfer (2015)).[4] This is contrasted by observed positive spill-overs between researchers

[2]This spill-overs occur, therefore, in the overlap of the 'idea space' and 'personal space' (Azoulay et al. (2010)).

[3]Agrawal et al. (2014) decompose the effects of a department hiring a star researcher in evolutionary biology into effects on overall research output by related and unrelated incumbents as well as new hires. While no positive effect is shown on incumbents overall, incumbents who work on similar research questions as the hired star increase their annual article output on average. The paper's main focus then is on the effect on new hires. Borjas and Doran (2015) study the potential negative effects resulting from the exodus of mathematicians after the end of the Soviet Union on the overall article output of collaborators left behind and previously geographically or thematically close researchers. Out of these groups, only former collaborators of highly productive emigrants appear to have been negatively affected. Azoulay et al. (2010) show the negative effects of the unexpected death of a scientific star on the productivity of collaborators in life science. Waldinger (2010) shows the importance of the quality of the supervisor on the lifelong productivity of PhD students in a sample of pre-World War II mathematicians. Waldinger (2012), in a later paper, finds no aggregated localized peer effects within German physics, chemistry or mathematics departments before the Second World War. Conversely, Waldinger (2016) finds negative short and long term impacts on the overall output of departments of the dismissed scientists.

[4]Kim et al. (2009) study the effect of being affiliated with a top American economics or finance department. In their empirical setting, regressions of the individual overall annual publication record on the characteristics of the researcher include university fixed effects. These affiliation fixed effects are positive for the 1970s but are insignificant in the 1990s. As a consequence, they conclude that localized peer effects

that are in a personal *and* in a thematic dimension closely connected, for instance co-authors or PhD students and their supervisors.

The empirical analysis is based on three key concepts. First, the paper compares research articles by an individual author rather than the annual productivity of researchers. This helps to address selection issues and differences in individual productivity. Second, a posited difference in the scope for quality improvement between "cutting-edge" (demanding) and "routine" (simpler) research is used to check for quality and reputation effects. Third, the estimates are compared to estimated effects of the researcher's network of former co-authors. This peer group is a close expression of past research and personal connectedness which allows to test for correlated effects, for instance reputation effects or strategically placed references. A rich and new biographical data set of around 1,000 highly-cited economists makes this empirical study possible.

The remainder of the paper is organized as follows. Section 2 describes the theoretical framework for local effects in knowledge production and dissemination. It sets a conceptual framework in which the impact of local colleagues can be empirically evaluated. Sections 3 and 4 present the data used and describe the empirical identification strategy. Section 5 presents empirical findings and Section 6 provides a discussion of the results and concludes.

4.2 Theoretical framework

The focus of this paper is on the benefit of research linkages within a university department. Comparing the reception of a researcher's published articles, this paper checks whether articles are taken up more widely by future research than comparable work if the research of local colleagues is linked. The empirical analysis is based on a sample of top economists. In turn, a leading researcher often constitutes a small cluster including, for instance, other less prominent researchers and graduate students. Any impact by the top researcher might, therefore, be indirect. In consequence, the empirical study focuses on whether linked specific local clusters add a benefit past the advantages offered by the host university in the production and dissemination of research.

Within a university department, a local research cluster can offer advantages to affiliated researchers. Following Duranton and Puga (2004), we can categorize advantages in the production of knowledge into benefits from sharing (for instance infrastructure), matching and learning. The sharing of costly infrastructure is less central in most fields of economic

disappeared with the decrease of communication costs. Similarly, Bolli and Schläpfer (2015) conclude that German economists' overall productivity did not profit on average from moves to new institutions between 2004 and 2008.

research. However, specialized research seminars and similar offer an advantage to thematically close researchers. The matching of skills is of rising importance upon the increases in research collaboration and specialization (see for instance Jones (2009b)). The largest potential benefit for research projects may arise though from personal connections of the authors to, and hence learning from, other researchers. In particular, other researchers who work on related topics might contribute through discussion and complementary knowledge even if not involved in the research project as authors. In the dissemination process, the cluster can act as a reputation signal for the research. Research that is produced in the vicinity of a leading scholar receives potentially more attention by researchers in the scholar's research field. In addition, the cluster can improve the dissemination through personal contacts, for instance in seminars.[5]

A model of cutting-edge and routine-type research

The model of scientific production tested in this paper assumes that a researcher engages in both of two types of research. First, "cutting-edge" research that leads to a high-profile publication and is challenging to the authors and, second, "routine-type" research that is less prominently published and does not require the same effort. The research process for either type starts with an idea. The researcher, then, decides whether to pursue the idea. Next, she decides on the level of effort for the project depending on the potential quality. Quality stands for the innovation step and relevance to the research area. Eventually, the effort and quality will correspond to the publication type: a high- or a low-profile journal.[6]

Peers can impact on this stylized research process. For example, common discussions can lead to more and better ideas and, thereby, increase the publication quantity. Focusing on an individual article, cutting-edge research that challenges the authors has a larger scope for quality improvements by local factors than routine-type research. First, the authors are more likely to seek feedback and over a longer time frame. Second, more parts of the research are based on new, potentially tacit knowledge and the authors are less likely to posses all necessary research knowledge beforehand. Third, more open questions in interpretation, theory or method give more scope for an impact by others, for instance through discussion or complementary knowledge. It is notable that the quality improvement itself can have a direct social component as, for example, the linking with concurrent and relevance to future research. On the other hand, routine-type research requires less time, effort, and innovation and has, therefore, less direct scope for quality improvement.

[5] Appendix 4.B provides a more in depth discussion on how local factors, in particular local colleagues, can impact on the influence exerted by a research article. The article's influence is set there as a function of quality, other dissemination factors and size of research area.

[6] The eventual quality and publication types are not to be seen *absolute* but *relative* to the author's standard.

Taken together, this motivates the assumption that, first, local research linkages increase the efficiency of the research process, and, second, a significant impact on quality is only expected for high-profile research (see also Bobtcheff et al. (2016)).[7] This assumption is posited for the empirical evaluation as follows:[8]

Assumption 1. *The quality of an article is set as a function of author ability and effort (A) and additional factors (X) as $h(A, X)$, with h strictly increasing in both up to a quality maximum $\widehat{Q}$:*

$$Q = \min\left(\widehat{Q}, h(A, X)\right).$$

In addition, for a routine-type article: $\forall x : h(A, x) \geq \widehat{Q}^{low}$ and, therefore, $Q = \widehat{Q}^{low}$ and for a cutting-edge type article: $\forall x : h(A, x) < \widehat{Q}^{high}$ and, therefore, $Q = h(A, x)$.

Assumption 1 states that the knowledge and effort of the authors are sufficient to produce the quality at the time of publication for a routine-type article. Conversely, the quality of a cutting-edge article benefits from additional input which the authors are also more likely to seek.

The take-up or influence of an article I is expected to depend positively on the quality of the article: $I'(Q) > 0$. Then, Assumption 1 leads to the empirical hypothesis that if a factor x helps the authors to improve the quality of an article behind their isolated efforts, then the semi-elasticity of the influence of the article at a given value of x is higher for a high-profile than a low-profile article.

Corollary 1. *Following from Assumption 1, if a variable x affects the measured influence (take-up) of a research article foremost through making the research process more efficient and enabling the authors to improve the article's quality behind their isolated efforts, then,the semi-elasticity of the influence I of an article at x is greater for a high-profile (cutting-edge) than a low-profile (routine) article, that is with quality $Q(x)$ and journal type J:*

$$\frac{\partial log(I(Q(x, j^{high}), j^{high}))}{\partial x} > \frac{\partial log(I(Q(x, j^{low}), j^{low}))}{\partial x} = 0$$

Conversely, personal connections may be correlated with the success of an article without contributing to the article's quality. A researcher's connectedness could, for instance,

[7] Bobtcheff et al. (2016) model a researcher's decision between early publication and quality maturation. Within their framework, a factor that makes the research process more efficient leads to more quality maturation and, eventually, to a higher quality of the published research.

[8] Appendix 4.C provides a weaker but less illustrative assumption for the quality of the article.

correlate with her or her department's (sub-field specific) reputation.[9] Then the opposite pattern is expected. Routine-type research in a low-profile journal should have a *relatively* larger reputation or visibility gain from the author or author's department. Cutting-edge research if published in a high-profile journal should benefit relatively less from the additional independent reputation gain. In the minimum, this reputation gain should not affect high-profile publications much more relatively than low-profile publications. This leads to the following stronger and the weaker versions of Assumption 2.[10]

Assumption 2. *The influence (take-up by future research) of an article is set as the function $I = I(R(Z), J)$, where Z is a variable that affects I foremost through the reputation R of the authors (or affiliation) and $J = \{j^{low}, j^{high}\}$ is the profile of the article's journal. Then, the following relations are posited.*

- *A high-profile article has more influence than a low-profile-article ceteris paribus:*

$$I(R(z, j^{low}), j^{low}) < I(R(z, j^{high}), j^{high})$$

- *A possible difference in the derivatives is dominated by this overall difference of a high- and low-profile article (stronger or weaker version).*

$$\frac{I(R(z, j^{high}), j^{high})}{I(R(z, j^{low}), j^{low})} > \text{ or } \gtrapprox \frac{\frac{\partial I(R(z, j^{high}), j^{high})}{\partial z}}{\frac{\partial I(R(z, j^{low}), j^{low})}{\partial z}}$$

Assumption 2 leads to the empirical hypothesis that if a factor z helps the general dissemination of an article without benefiting high-quality research more strongly, then the semi-elasticity of the influence of the article at a given value of z is lower for a high-profile than a low-profile article (or equal in the weaker version of Assumption 2).

Corollary 2. *Following from Assumption 2, if a variable z affects the measured influence of a research article foremost through correlation to the author's or department's reputation without a direct impact on quality, the semi-elasticity of the influence I of an article at z*

[9] Further examples of possible dissemination benefits include strategic citation behavior or a higher awareness of the research of personally connected researchers.

[10] The third, opposite case in which a reputation gain affects overwhelmingly high-profile publications is imaginable. In this case, we cannot distinguish between the two posited effects. To address this concern the estimates for local colleagues will be compared to analogous estimates using the network of former co-authors (See Section 4.4). Former co-authors serve here as an expression of former research and personal connectedness. Appendix 4.C also provides additional steps and an assumption for a functional form to illustrate the reputation effect of the article. Dietrich (2008) show the effect of visibility on citation counts using the ordering of arXiv lists of new astronomy articles.

is smaller for a high-profile (cutting-edge) than a low-profile (routine) article, that is with reputation $R(z)$ and journal type J:

$$\frac{\partial log(I(R(z, j^{high}), j^{high}))}{\partial z} < \text{ or } \lessapprox \frac{\partial log(I(R(z, j^{low}), j^{low}))}{\partial z}$$

Corollary 1 and 2 are central in testing the impact of local research links. The empirical analysis is centered on the distinction of high-profile and low-profile articles by an individual researcher at a given university. According to the two assumptions made, a positive impact of a local cluster on the quality of an article leads to a positive correlation between citations received and references to local colleagues in a high-profile article. However, no (strong) positive correlation is expected for low-profile publications.[11] Conversely, an impact by the cluster on the field-specific reputation would either lead to similar estimates for high- and low-profile publications or stronger correlation for low-profile publications.

4.3 Data

The empirical analysis is carried out using publications by the most cited academic economists between 1996 and 2014. The construction of this data set started with the economic journals listed by Kalaitzidakis et al. (2011). These over two hundred journals were supplemented with a number of other, highly ranked journals in Ideas RePEc, for instance the relatively recent AEA American Economic Journals, which brings the total number to 255 journals.

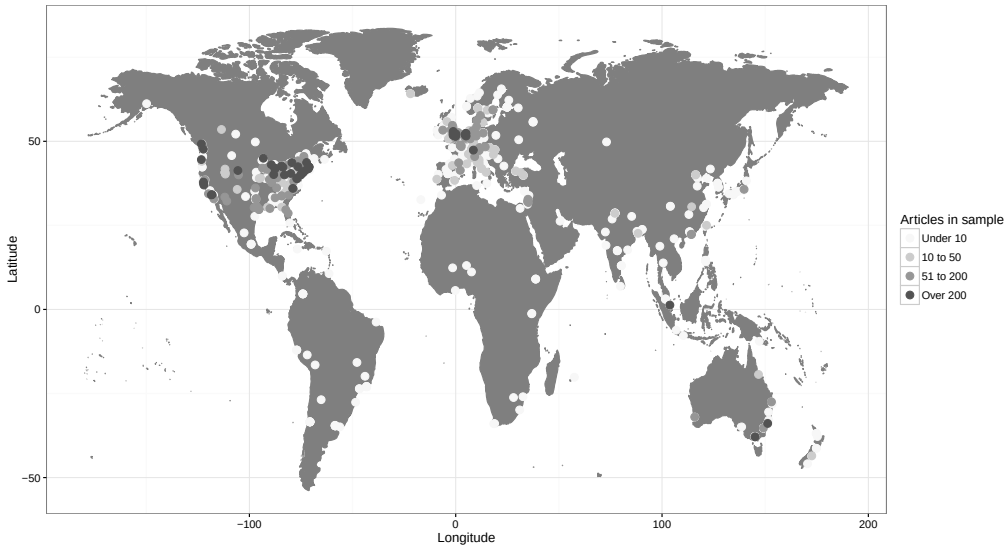
Next, the authors of these articles were ranked by the number of citations received as reported by Scopus.[12] From this exercise, a total of 967 economists were chosen based on work published in the period 1996 to 2014 and most highly cited in this period. All of these authors were ranked among the top 1,300 economists. The reasons that not all 1,300 top ranked economists are included is that CVs were not available or that their name details were not individual enough to be confidently attributed to a single economist.

As an illustration of the spatial distribution of the universities studied, the affiliations of the selected economists are shown in Figure 4.1. While many countries are observed, there is still a strong concentration on North-American universities. Within the sample, economists from Harvard and Berkeley account together for over ten per cent of the article output and the ten universities with the highest share of articles are in the USA. Education is even more concentrated; twenty-one per cent of the sample economists hold a PhD from Harvard or MIT and the top six universities account for forty-four per cent of the doctorates.

[11] The impact might be more on quantity than quality for this category.

[12] See `scopus.com`. Citations by published research articles in journals that are indexed on Scopus.

Figure 4.1: The affiliations of the authors in the sample.



For the economists selected, complete information on the research career from the undergraduate studies onward was compiled using on-line CVs and the encyclopedia 'Who's Who in Economics' (Blaug and Vane (2003)). This was complemented using name searches on Scopus to retrieve additional publication data. In most cases, the author's Scopus profile contains an extensive list of the author's available publications. The data sets used in the regressions below are restricted to publications between 1996 and 2014.[13] This yields 28,901 research articles that include at least one sample economist. The main regressions below use a more restricted data set of which all authors are in the sample. In addition, articles in the Journal of Economic Literature or articles that cite less than five or more than 80 other articles are excluded. This main data set, then, contains 7,291 research articles.

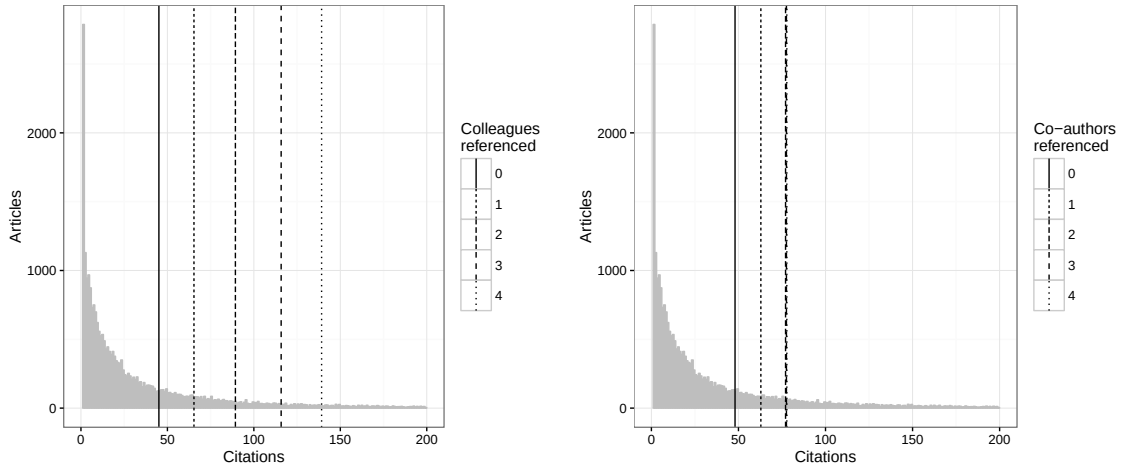
Before the following estimation, Figure 4.2 shows the citation averages of articles by the number of referenced eminent colleagues and former co-authors. This indicates in a preliminary, descriptive way a positive relation between the citations received by an article and the authors' personal connections to eminent researchers with related prior publications, something that is tested more formally below. The graph on the left in Figure 4.2 shows that on average every additional eminent colleague referenced is associated with an increase of slightly over 20 citations received per article with up to four colleagues referenced. In contrast, former co-authors are only associated with an increase in citations for the first two co-authors referenced (graph on the right in Figure 4.2). For both categories, higher counts of peers referenced than depicted are very uncommon, possibly, because self-references of the authors are excluded.

[13]Scopus provides full coverage of most journals from 1996 onward only.

Figure 4.2: Citations received per article.

(a) Citations by number of colleagues referenced.

(b) Citations by number of former co-authors referenced.



Notes: The background shows the citations received of 28,901 economic research articles between 1996 and 2014 of which at least one author are eminent economists as described in data section. The vertical lines show the means of citations received by number of eminent colleagues/co-authors referenced. The plots are truncated at 200 citations.

4.4 Empirical strategy

The take-up or influence of any article is measured in this paper by the number of citations.[14] The empirical identification of local effects is so based on the impact on the total sum of citations received per article (I). It is important to note that articles by the same authors are compared with a series of control variables. Conclusions based on citations would be problematic without the specific context.

Research links to local colleagues are estimated by classifying articles on whether they refer to articles by recent colleagues of the authors. Since the analyzed sample consists of the publications of around 1,000 top economists as described in Section 4.3, an impact of just one linked colleague is conceivable, in particular, if we see a top economist as a proxy for a small research cluster including further less eminent researchers and students. Possible spurious estimates due to correlated effects by proximity to eminent researchers are addressed, first, by focusing on linked research instead of university affiliation and using individual and university dummies, second, by comparing publication sub-groups with

[14] See, for instance, Tahamtan et al. (2016) for a literature review on factors that affect citation counts and Bornmann and Daniel (2008) or Osterloh and Frey (2014) for a discussion on impact and citation counts. Bornmann and Daniel (2008) give an overview of studies on the relation of citation counts and impact. While the citation behavior varies between researchers, the authors conclude that citation counts are generally a valid measure of impact.

different scopes for quality maturation and dissemination effects, and, third, by contrasting estimates using the network of former co-authors. More detail is given below.

The impact of colleagues referenced (N) in an article on the influence (I) of the article is estimated by:

$$\mathbb{E}\left(I|U,A,T,J,X,N\right) = \exp\left(U + A + T + J + X\gamma + N\beta\right) \tag{4.1}$$

Equation 4.1 shows the estimation of the mean of citations received, I, as an exponential function, including fixed effects for the universities U, authors A, time T and journal J, a vector of control variables X and indicator variables for a reference in the article to a colleague or former co-author, interacted with a dummy to indicate whether the article is published in a high- or low-profile journal, so:

$$\begin{aligned} N\beta =& (Colleague : HighProfile)\beta_1 + (Colleague : LowProfile)\beta_2 + \\ & (CoAuthor : HighProfile)\beta_3 + (CoAuthor : LowProfile)\beta_4 \end{aligned} \tag{4.2}$$

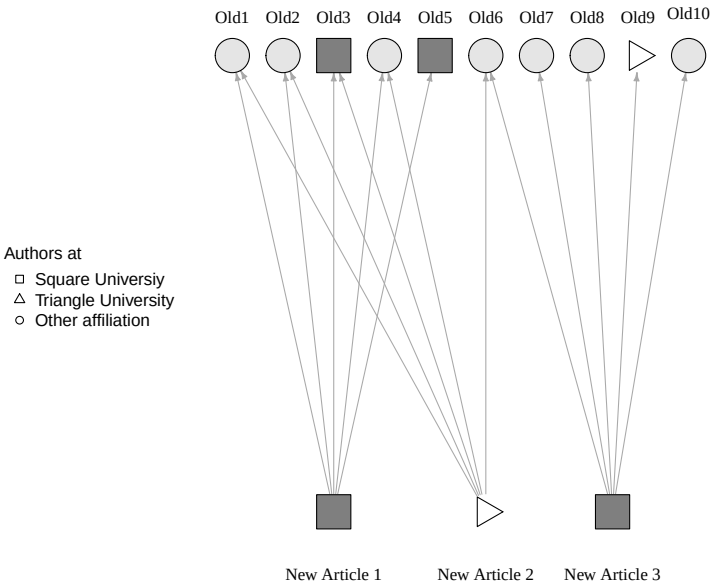
In the following, $HighProfile$ and $LowProfile$ denote whether the new article is published in a top 25 economics or finance journal.[15] The key variables of interest are the peer variables in N and these will be discussed first.

Peer variables

The peer variables in N, $CoAuthor$ and $Colleague$, denote articles that reference a recent co-author or colleague of the authors. Figure 4.3 sketches the applied measurement of personal and thematic links. For illustrative purposes, we assume that three new articles are written by authors at two different affiliations: Square University and Triangle University. Each article refers to five existing articles which defines their research area relatively. For example, New Article 1 and New Article 2 have a similar research focus based on their references while New Article 3 has a different topic. Of the three new articles, the only article for which the thematic and personal links overlap is Article 1, that is the article

[15] To rank the 255 economics journals, the set of all research articles between 1996 and 2014 as described in the data section is used. First, all citations are divided by the yearly median. Then, the mean of these adjusted citations per article is used to rank the journals. Finally, the first structural break in the mean of citations at 25 was used to classify the publications. In addition, the Journal of Economic Literature is excluded.

Figure 4.3: Thematic and personal relations in an article citation graph.



cites articles that are written by recent colleagues of the authors. In consequence, New Article 1 is the only article of the three that is classified as treated with the respect to the variable *Colleague*. The further empirical analysis is centered on the effect of this treatment.

Formally, S_i is constructed as the set of authors referenced in article i after excluding article references that are (co-)authored by an author of the article.[16] Next, this set of individuals is compared with the list of eminent economists in our sample[17] who have a known personal connection to an author around the time of production, that is are in the neighborhood N_i^x of the article. $N_i^{CoAuthor}$ and $N_i^{Colleague}$ are defined as the sets of economists in the sample who have shared an affiliation or co-authored an article one to five years prior to the publication with at least one of the authors of the article i. The time frame is chosen to reflect the likely production period of an article.

Subsequently, two variables are constructed that indicate the articles in which at least one colleague or former co-author from this time span is referenced: $Colleague = \mathbb{1}_A$, with $A = \left\{ i | N_i^{Colleague} \cap S_i \neq \emptyset \right\}$ and $CoAuthor = \mathbb{1}_B$, with $B = \left\{ i | N_i^{CoAuthor} \cap S_i \neq \emptyset \right\}$. Alternatively, the number of colleagues or co-authors is used instead of the respective

[16] This is varied in robustness check. The exclusion follows the argument that the co-authors should have complementary knowledge on the topic. Self-references are counted separately.

[17] The sample of 967 eminent economists as described in the data section.

dummy variable. These counts are truncated at three to account for the low number of higher values observed.

Control variables

The vector X contains a series of further characteristics of the article and its authors. First, the mean age of the authors and the squared mean age. This is included to account for career effects on productivity. Second, the number of affiliations, the number of authors and whether the authors are listed alphabetically is used to complement the individual and university dummies. A non-alphabetical ordering is unusual in economics[18] and can indicate a different background of the authors or authors added without a full contribution which would overestimate the number of authors. Third, the number of references and its squared value are included. More references can indicate a bigger project, more interest in the research area or increase the visibility of the publication independently.[19] The number of pages is not significant if the number of references is used and, subsequently, not used in the estimation.

Additional specifications introduce a control variable for the total number of eminent economists cited. Since the sample of economists was selected based on citations in economic journals, a high number of cited eminent economists indicates a research field that attracts a high interest by economic researchers. The interest in the topic could in turn cause a positive effect of peers referenced. However, as shown below, the main estimates for the influence of peers are not changed significantly after the introduction of this control variable. This variable is not included in the main regression models as it is difficult to rule out connections between citing researchers. Finally, the number of self-references is counted to indicate prior experience in the area: $NumSelfReferences_i = |M_i|$, where M_i is the set of references in i to articles by an author of i. Prior publications could indicate a higher visibility and linked prior experience should be helpful in the research process. On the other hand, self-references may be partially arbitrary or indicate follow-up work to main publications.

Identification

A causal interpretation of empirical effects of personal links to other researchers is not straightforward. The decision to link with the research of a colleague is not random. Most

[18] In the time span studied, 92 per cent of articles published in AER, Ecta, JPE, OJE and REStud with more than one author list the authors alphabetically.

[19] For example, on-line databases make it possible to search citing papers. Therefore, the more references an article lists the more such searches include the article.

importantly, more productive researchers are likely to be selected into universities with other productive researchers. Within a department, more productive researchers might also decide to work on common topics. This motivates the focus on the productivity of a fixed individual researcher at a fixed university. In this context, the main concern is that the researcher's forte corresponds to that of the department because of selection: The researcher selects or is selected into the department because of corresponding research interests.

To counter this concern, the author's network of former co-authors is used. This network relates closely to the researcher's past research record and personal links in a specific area.[20] The co-author network can, therefore, be used to control for the correlation between the author's department and their past research and connections. The co-author variable is also introduced as a control for other possible correlated effects in the local colleague variable. For example, both variables reflect the decision to reference a personally connected researcher. If the estimation process leads to mechanical correlations with other factors, then an equivalent estimation process of this peer group and local colleagues leads to similar estimates. Diverging estimates support that the estimates are not dominated by common correlated aspects of these peer groups or the estimation process.

Finally, the distinction made in the theoretical framework between a high-profile and a low-profile article is used to indicate possible general effects associated with the authors or universities that are expressed in the peer variables. For instance, the local colleague or co-author network may reflect the reputation of the authors which, in turn, increases the visibility of associated research. Relatively, this visibility gain would be more important for research that is published less prominently.

4.5 Empirical results

This section reports on the empirical estimation applied to the model discussed in Section 4.4, namely the estimates of the impact of local research linkage on the success of individual research articles by type of journal.[21]

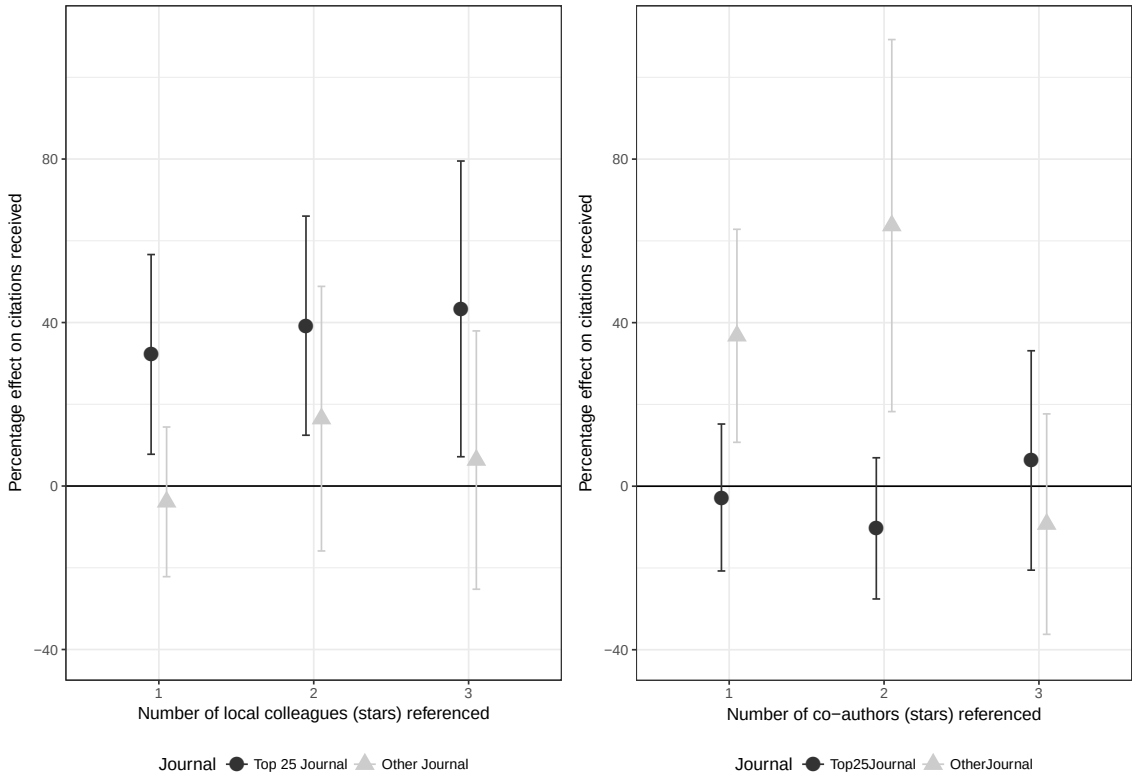
[20] While these former co-authors are closely linked to other work by the researcher, they are not authors of the new, analyzed article. In addition, since the co-author network is an expression of past research, the correlation with the success of an article does not necessarily indicate an influence of the co-authors.

[21] All main estimates are derived using quasi-Poisson regressions using the natural logarithm as the link function. There is no general consensus on the estimation of citation data models but none of the estimation models used (quasi-Poisson, negative Binomial, lognormal, normal with log transformed citation count) changes the key results qualitatively. Quasi-Poisson regressions account for the count nature of citations, the observed over-dispersion and is more robust than negative binomial regressions given the high number of dummies. The estimates are robust to different estimation models such as a using a log-transformed count or negative binomial models. The estimation is carried out using the glm function in R. The packages multiwayvcov (multi-way cluster-robust variance estimation as suggested by Cameron et al. (2011)) and lmtest (hypothesis testing) are used for standard error correction.

Figure 4.4: Estimated impact of local colleagues.

(a) Percentage effect of local colleagues referenced.

(b) Percentage effect of former co-authors referenced.



Notes: The plots show the estimated percentage effect of colleagues and co-authors referenced on citations received and the associated 95 % confidence intervals. The point estimates are transformed parameters of a quasi-Poisson regression. The standard errors are clustered at the journal-year level and transformed using the Delta method. The corresponding regression table can be found in the appendix (Table 4.3).

Figure 4.4 (a) shows that a high-profile article receives significantly more citations if it links with the research of local colleagues. The left plot depicts the estimated percentage change of citations received if the article references one, two or three and more local colleagues out of the sample of star economists.[22] The estimates show no significant correlation of local colleagues to the dissemination for low-profile publications. This concurs with the two empirical hypotheses 1 and 2 in the conceptual framework if local colleagues impact positively on the quality maturation of an article but are not associated with a general dissemination benefit.

These estimates are contrasted with analogous estimates for references to former co-authors in Figure 4.4 (b). This shows that the network of former co-authors is correlated with an increase of citations to low-profile articles but not to high-profile articles. More than two star co-authors referenced are very uncommon as self-citations are excluded which might explain the diverging estimates for the third category. Overall, these estimates show that factors that are common to both peer variables, such as the decision to reference a personally connected star researcher, cannot explain the observed effect of local colleagues. The estimates for local colleagues show no clear difference for the three categories. Seeing a star researcher as a small cluster instead of a single person might help explain this pattern. In consequence, the further analysis is carried out using dummies to indicate on or more linked colleague or co-author.

Table 4.1 repeats and varies the estimation shown in Figure 4.1 using dummies for articles that reference colleagues or co-authors. Similarly, the estimates show that a high-profile article receives significantly more citations if it links with the research of local colleagues. Column 1 indicates a 35 percent increase in citations received for a high-profile publication.[23] Conversely, the estimates indicate no significant influence of local colleagues on the take-up of low-profile publications. The opposite correlation pattern is observed for former co-authors: a high correlation with the take-up for low-profile articles and a low correlation with the take-up for high-profile articles. The estimates are, therefore consistent with the hypothesis that local colleagues impact on the quality of a high-profile publication and former co-authors are a reflection of the authors' field specific reputation.

So far, colleagues who are also former co-authors were included in both categories. Column (2) in Table 4.1 excludes colleagues that are also former co-authors from both groups and shows similar estimates as column (1). When we focus on this third group on its own (Table 4.6 in the appendix), it shows that their correlations are similar to non-local co-authors once

[22] The corresponding regression table can be found in the appendix (Table 4.3)

[23] The percentage effect for $\beta = 0.302$ is calculated as $(e^{0.302} - 1) * 100 = 35.26$, see (Wooldridge (2010), p. 726). All columns are variations of the baseline Equation 4.1 in Section 4.4 (Empirical Framework). All estimations also include yearly, individual, university, and top journal dummies, as well as a number of author and article specific control variables (reported in the respective tables in the appendix).

Table 4.1: Impact of local colleagues

	Dependent variable:				
	SumCitationsReceived *(link: log)*				
	(1)	(2)	(3)	(4)	(5)
Colleague:Top25Journal	0.30***	0.29***	0.19***	0.24***	0.27***
	(0.08)	(0.07)	(0.04)	(0.08)	(0.09)
Colleague:OtherJournal	-0.00	-0.01	0.01	-0.03	0.10
	(0.08)	(0.08)	(0.05)	(0.08)	(0.08)
FormerCoAuthor:Top25Journal	-0.02	0.12	0.04	-0.08	
	(0.10)	(0.17)	(0.06)	(0.10)	
FormerCoAuthor:OtherJournal	0.35***	0.26**	0.27***	0.33***	
	(0.08)	(0.13)	(0.05)	(0.08)	
N (df)	6958 (5730)	6958 (5746)	25149 (23863)	6958 (5729)	6958 (5732)
Pseudo-R^2	0.71	0.71	0.58	0.71	0.71

Notes: This table reports estimated coefficients from quasi-Poisson regressions with standard errors in parentheses. The semi-elasticity (percentage change in the dependent variable if a dummy is 1) is calculated as (exp(coefficient)-1)*100, for instance (exp(0.302)-1)*100=35.26. Standard errors are clustered at the journal-year level. Each regression also includes, yearly dummies, individual dummies, university dummies, a dummy for a top 25 journal publication, the number of authors, the number of universities, the mean and the squared mean of the authors' years since their PhD, the total number of references to other articles, the number of references to other articles by the authors, and a dummy for an alphabetical author name order. The baseline estimates in column (1) use publications of which all authors are in the sample (see Table 4.4). Column (2) differs from (1) by estimating colleague-co-authors separately (see Table 4.6). Column (3) differs from (1) by including publications that are co-authored outside the sample (see Table 4.9). Column (4) differs from (1) by including the number of eminent researchers referenced (see Table 4.13). Column (5) excludes the co-author variables (see Table 4.12).

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

individual dummies are introduced. Importantly, the introduction of individual dummies from column (1) to column (2) in Table 4.6 leads to a decrease in the correlation between co-author-colleagues and high-profile publications. The decrease due to individual dummies suggests that co-author-colleagues indicate very productive star researchers but that does not support a strong impact on an individual research article. On the other hand, local colleagues outside of the authors' co-author network are estimated to impact positively on high-profile research. This supports the hypothesis that co-location is important if the researchers are not already personally linked, for instance as co-authors.

Robustness checks and extensions

The first two columns in Table 4.1 show estimates that are based on publications of which *all* authors are in the sample of eminent economists. This raises questions on the representativeness for all publications of these authors and for publications of other less prominent authors. For example, the limitation to publications of which all authors are sample economists leads to an under-representation of multi-authored papers. To address representativeness within the sample of top economists, column (3) repeats the regressions with all publications of the sample economists. While this introduces new problems due to unknown authors, the estimates do not differ qualitatively from the other columns (see

also Table 4.9 in the appendix). This shows that the results are not limited to the more restrictive sample or caused by the over-representation of single-authored articles. Table 4.13 shows the estimates of the regressions presented in Table 4.4 while also controlling for the overall number of eminent economists cited. The estimates show that the estimated peer effects are not driven by the interest in the topic as evidenced by eminent colleagues working on it. However, the main specification does not include this control variable since a personal connection between the researchers cannot be ruled out. In addition, column (5) shows the estimated parameters for colleagues without including co-authors which indicates that the estimated effect does not come from the correlation between the two variables (see also Table 4.12).

Table 4.5 shows estimates for the number of not thematically connected recent star colleagues or co-authors. Again, the estimation includes individual and university dummies for the authors. In this setting, neither group is associated with an increase in citations received. In addition, the estimates for thematically connected colleagues and co-authors stay almost unchanged.

To indicate the effects on less eminent researchers, Table 4.10 (appendix) repeats the estimation of specification (1) in Table 4.1 leaving out publications by the most cited 100 (column 1) to 500 (column 5) economists. The literature stresses the importance of star researchers (see for instance Agrawal et al. (2014) and Oettl (2012)) on the productivity of peers, hierarchical effects on less productive researchers would, therefore, lead to higher estimates. Conversely, these slightly less cited researchers have on average less eminent sample economists in their departmental or co-author network which would lead to lower effects if the peer effects of stars are larger. Overall, the estimates in Table 4.10 do not show a clear trend, while the observed pattern persists that local colleagues strongly impact on high-profile publications.

High-profile articles are not more likely to refer to local colleagues if differences between individual researchers are taken into account (see Table 4.11). The estimates are, therefore, unlikely to be driven by strategic citation behavior.

The impact of colleagues is estimated to be stronger during the first five years following the award of a PhD. However, Table 4.7 also shows that the impact is persistent at later career stages. This is consistent with the hypothesis that young researchers profit most from the proximity to other researchers but that this proximity remains valuable at a lower level throughout the career.

Conclusion of empirical results

The estimation of the impact of local research linkage is based on references to local colleagues between top economists. Therefore, the estimates may relate to a local research cluster rather than the impact of an individual colleague.[24] These clusters or eminent colleagues are associated with a high relative increase in citations to high-profile (demanding) publications. Importantly, the number of eminent local colleagues is insignificant if their research is not linked. Therefore, the estimates do not support local peer effects without thematic research links. In addition, linked research does not necessarily study similar research questions or is classified as the same research area.

The empirical design implies further caveats in the interpretation of the results. First, the empirical analysis focuses on the general impact of research links to local colleagues. Endogenous peer effects and the impact of exogenous characteristics of colleagues or departments are not explicitly distinguished (see Manski (1993)). Therefore, the main focus is on assessing whether correlated effects such as selection effects are expressed in the estimates. Second, this paper focuses on individual articles instead of the overall article output of a researcher. The interpretation of the results is, therefore, limited on effects on the quality and dissemination of individual articles. Third, the estimates for the co-author network are used to support the causal interpretation of local colleagues. The estimates do not justify a causal interpretation between former co-authors and research quality.[25]

[24] A top researcher can be seen as a small cluster including less eminent researchers and graduate students. Any impact might, therefore, be indirect.

[25] Overall, the co-author network might impact more positively on individual productivity, in particular as direct co-authorship is not studied here (see for instance Azoulay et al. (2010), Borjas and Doran (2015), Ductor et al. (2014) and Ductor (2015)). In particular, learning through co-authorship may lead to an overall higher research performance which does not show in the comparison of individual research projects. In addition, the co-author network is arguably a stronger reflection of a researcher's past productivity than local colleagues, if the affiliation is controlled for. In turn, this lower dependence on past performance reaffirms any estimated positive impact of local colleagues outside the researcher's co-author network. The negligible impact of university dummies on the estimated effect of colleagues agrees with the posited low dependence. This causality in the estimated impact of local colleagues is further supported by the estimated pattern in which the effect is limited to high-profile research. First, as outlined in the conceptual framework (Section 4.2), a strong quality effect on low-profile research is unlikely given the sample of eminent researchers. Second, confounding factors that influence both publication types are ruled out. Third, the contrast to the co-author estimates rules out confounding factors that are similarly correlated with both groups. On the other hand, besides dissemination effects, additional arguments suggest a relatively higher importance of co-authors for lower-profile publications: First, in comparison to the individual average, co-authors are an expression of the authors' research record in different sub-fields and the indicated ability and reputation are, therefore, field-specific. As a consequence, co-authors have a higher expected impact on field journals which are typically lower ranked. Second, co-authors may be important for the generation of new ideas but that the quality of high-profile publications depends more on the efficiency of the maturation (Bobtcheff et al. (2016)) process which may be more strongly influenced by local colleagues. Alternatively, different researchers rely on different peer groups and publish in different journals. However, strong effects are estimated for colleagues throughout different subgroups of researchers in Table 4.10 which makes this explanation unlikely. Finally, reverse causality that colleagues are systematically cited in expected stronger publications can be excluded as a strong factor in the estimates. In Table 4.11, local colleagues are not cited significantly more often in top journal publications when individual dummies are used.

Fourth, the estimated impact of local colleagues is centered around articles that are already highly cited.[26] The type of article affected is likely correlated to the chosen sample of highly cited economists.

4.6 Conclusion

This paper argues that personal connections to thematically close researchers are crucial for the production and dissemination of research. To test this proposition, economic research articles are classified by reference to local colleagues. Furthermore, the publications are distinguished by the standing of the journal to separate a researcher's more and less high-profile research. Finally, the number of citations received is taken as an estimate for an article's influence or success.

In this framework, high-profile research is shown to receive a significantly higher number of citations due to linkage with research by local colleagues. These estimates control for differences between individual researchers and universities. Conversely, local colleagues are not associated with a dissemination benefit for low-profile research after controlling for the overall affiliation effect. The used sample of top economists is given the interpretation of a sample of research clusters including, for instance, less well-known researchers and graduate students.

Overall, this paper argues that local peer effects persist if researchers can link their research. This does not contradict the observed decline in localized peer effects but suggests that the decrease in communication costs may be just one of the causes for the observed decrease in department-wide effects in economic research, besides, for example, an increase in specialization (Jones (2009b)). The observed productivity effect of the interaction of local colleagues working within a common specific research area could be important in the organization of research entities. However, the emphasis is on benefits by research linkage and not by parallel research (the same research field). The results show, in particular, the benefit of thematic links to local colleagues outside the authors' co-author network. On the other hand, the absolute number of eminent colleagues is not associated with an increase in citations received without a thematic connection to their research.

4.A Tables

[26] For instance, more than 50 citations received on Scopus (Scopus typically reports fewer citations than for instance Google Scholar).

Table 4.2: Description of the regression variables

Name	Description	Range	Mean
Sum Received Citations	The sum of citations received up to January 2015.	⟦0,4200⟧	72.88
CoAuthor	Articles that refer to at least one researcher who was a co-author of any author one to five years prior to publication.	⟦0,1⟧	0.11
Colleague	Articles that refer to at least one researcher who was a colleague of any author one to five years prior to publication.	⟦0,1⟧	0.18
Top25Journal	A publication in one of the 25 most frequently cited economics journals (citations per article/divided by the yearly median).	⟦0,1⟧	0.26
Mean Academic Age	The mean time since completion of the PhD of the sample authors.	[0,60]	17.2
Number Uni	The number of universities of sample authors.	⟦0,4⟧	1.23
Num Authors	The total number of authors.	⟦1,5⟧	1.32
Num References	The number of articles referenced, truncated at 5 and 80.	⟦5,80⟧	30.8
Num Self References	The number of articles referenced that are written by an author of the article, truncated at 20.	⟦0,20⟧	3.8
Order Authors	Share of multi-authored articles that list authors alphabetically.	⟦0,1⟧	0.94
MaxAge5	None of the authors has completed their postgraduate studies more than five years prior to the publication of the article.	⟦0,1⟧	0.13

Notes: The Table shows a description, the range and the mean of all variables included in the regressions presented. The statistics refer to the main data set of 7,291 articles of which all authors are in the sample as described in the data section.

Table 4.3: Effect of local colleagues by number

	Dependent variable:			
	SumCitationsReceived *(link: log)*			
	Estimate	Std. Error	z value	Pr(>|z|)
1 Colleague:Top25Journal	0.28	0.09	2.96	0.00
2 Colleagues:Top25Journal	0.33	0.10	3.37	0.00
3 Colleagues:Top25Journal	0.36	0.13	2.80	0.01
1 Colleague:OtherJournal	-0.04	0.10	-0.41	0.68
2 Colleagues:OtherJournal	0.15	0.14	1.08	0.28
3 Colleagues:OtherJournal	0.06	0.15	0.41	0.68
1 CoAuthor:Top25Journal	-0.03	0.10	-0.27	0.79
2 CoAuthors:Top25Journal	-0.11	0.23	-0.48	0.63
3 CoAuthors:Top25Journal	0.06	0.19	0.32	0.75
1 CoAuthor:OtherJournal	0.31	0.09	3.50	0.00
2 CoAuthors:OtherJournal	0.49	0.18	2.67	0.01
3 CoAuthors:OtherJournal	-0.10	0.40	-0.24	0.81
Top25Journal	0.80	0.05	15.38	0.00
NumAuthors	0.18	0.40	0.44	0.66
NumberUnis	0.11	0.20	0.55	0.58
MeanAcademicAge	-0.02	0.01	-1.30	0.19
MeanAcademicAge^2	0.00	0.00	1.49	0.14
NumReferences	0.02	0.00	10.17	0.00
NumSelfReferences	-0.01	0.01	-0.75	0.45
OrderAuthors	0.38	0.09	4.12	0.00

Notes: This table reports estimated coefficients from quasi-Poisson regressions with standard errors in parentheses. Standard errors are clustered at the journal-year level (variables as described in Table A.1). Each observation is an article of which all authors are in the sample described of 967 eminent economists. The estimations include dummies for the year, for each author, and for universities with more than two articles. The estimates correspond to the shown percentage effects in Figure 4.4. The plotted percentage effect is calculated for any β as (exp(β)-1)*100.

Table 4.4: Impact of local colleagues

	Dependent variable:				
	SumReceivedCitations *(link: log)*				
	(1)	(2)	(3)	(4)	(5)
Colleague:Top25Journal	0.167***	0.308***	0.302***	0.313***	0.226**
	(0.062)	(0.079)	(0.080)	(0.119)	(0.089)
Colleague:OtherJournal	0.091	0.048	-0.001	0.016	-0.045
	(0.104)	(0.092)	(0.079)	(0.108)	(0.075)
CoAuthor:Top25Journal	0.026	0.043	-0.017	0.068	-0.070
	(0.081)	(0.097)	(0.099)	(0.163)	(0.108)
CoAuthor:OtherJournal	0.264***	0.356***	0.349***	0.297**	0.292***
	(0.102)	(0.095)	(0.081)	(0.136)	(0.084)
Top25Journal	0.778***	0.808***	0.798***	0.785***	
	(0.054)	(0.053)	(0.052)	(0.072)	
NumAuthors	0.558***	0.172	0.170		0.281
	(0.080)	(0.366)	(0.384)		(0.333)
NumberUnis	-0.203***	0.057	0.180	0.296	0.103
	(0.072)	(0.079)	(0.206)	(0.474)	(0.242)
MeanAcademicAge	-0.002	-0.015	-0.017	-0.363***	-0.011
	(0.007)	(0.012)	(0.012)	(0.137)	(0.013)
MeanAcademicAge^2	0.000	0.000	0.000	0.001**	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
NumReferences	0.015***	0.016***	0.017***	0.016***	0.016***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
NumSelfReferences	-0.001	-0.009	-0.005	-0.004	0.001
	(0.006)	(0.007)	(0.007)	(0.009)	(0.007)
OrderAuthors	0.306***	0.367***	0.349***	0.189	0.256***
	(0.054)	(0.068)	(0.070)	(0.181)	(0.063)
Year dummies	Yes	Yes	Yes	Yes	Yes
Individual dummies	No	Yes	Yes	No	Yes
Authorgroup dummies	No	No	No	Yes	No
University dummies†	No	No	Yes	Yes	Yes
Journal dummies	No	No	No	No	Yes
N (df)	6958 (6928)	6958 (5993)	6958 (5730)	6958 (4904)	6958 (4884)
Pseudo-R^2	0.45	0.68	0.71	0.78	0.79

Notes: This table reports estimated coefficients from quasi-Poisson regressions with standard errors in parentheses. Standard errors are clustered at the journal-year level except for model (4) which is clustered at the author group level (variables as described in Table 4.2). Each observation is an article of which all authors are in the sample described of 967 eminent economists. †: Universities with more than two observations.

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

Table 4.5: Thematically not connected colleagues

	Dependent variable:		
	SumReceivedCitations *(link: log)*		
	(1)	(2)	(3)
Colleague:Top25Journal	0.301***	0.317***	
	(0.081)	(0.083)	
Colleague:OtherJournal	0.001	0.010	
	(0.079)	(0.078)	
CoAuthor:Top25Journal	-0.027	-0.018	
	(0.099)	(0.106)	
CoAuthor:OtherJournal	0.354***	0.372***	
	(0.082)	(0.088)	
LogNumColleagues		-0.029	0.010
		(0.029)	(0.029)
LogNumCoAuthors		-0.027	0.013
		(0.060)	(0.055)
Top25Journal	0.794***	0.795***	0.812***
	(0.052)	(0.052)	(0.044)
NumAuthors	0.187	0.249	0.218
	(0.396)	(0.405)	(0.421)
NumberUnis	0.170	0.156	0.319
	(0.205)	(0.204)	(0.213)
MeanAcademicAge	-0.016	-0.018	-0.018
	(0.012)	(0.013)	(0.013)
MeanAcademicAge^2	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
NumReferences	0.017***	0.017***	0.018***
	(0.002)	(0.002)	(0.002)
NumSelfReferences	-0.006	-0.006	-0.006
	(0.007)	(0.007)	(0.007)
OrderAuthors	0.356***	0.370***	0.371***
	(0.094)	(0.094)	(0.098)
N (df)	6958 (5730)	6958 (5728)	6958 (5732)
Pseudo-R^2	0.71	0.71	0.71

Notes: This table reports estimated coefficients from quasi-Poisson regressions with standard errors in parentheses. The estimations include year, university and individual dummies (Universities with more than two observations). Standard errors are clustered at the journal-year level. Variables as described in Table 4.2. In addition, LogNumColleagues and LogNumCoAuthors are the logarithm of the number of colleagues or co-authors +1. Each observation is an article of which all authors are in the described sample of 967 eminent economists. Universities with more than two observations.

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

Table 4.6: Impact of local colleagues with a differentiation between co-authors and other local colleagues

	Dependent variable:				
	SumReceivedCitations *(link: log)*				
	(1)	(2)	(3)	(4)	(5)
Colleague:NoCoAuthor:Top25Journal	0.127*	0.307***	0.290***	0.324***	0.241***
	(0.067)	(0.076)	(0.074)	(0.125)	(0.079)
Colleague:NoCoAuthor:OtherJournal	0.028	0.041	-0.009	0.031	0.009
	(0.110)	(0.084)	(0.076)	(0.113)	(0.078)
CoAuthor:OtherUni:Top25Journal	0.065	0.179	0.120	0.225	-0.023
	(0.148)	(0.155)	(0.169)	(0.142)	(0.155)
CoAuthor:OtherUni:OtherJournal	0.148	0.253	0.256**	0.259	0.222*
	(0.118)	(0.175)	(0.126)	(0.178)	(0.118)
Colleague:CoAuthor:Top25Journal	0.223*	0.002	-0.038	0.052	-0.075
	(0.133)	(0.146)	(0.153)	(0.206)	(0.162)
Colleague:CoAuthor:OtherJournal	0.511***	0.366**	0.349**	0.350*	0.213*
	(0.160)	(0.186)	(0.154)	(0.181)	(0.127)
Top25Journal	0.777***	0.792***	0.785***	0.793***	
	(0.053)	(0.051)	(0.050)	(0.070)	
NumAuthors	0.583***	0.247	0.283		0.341
	(0.091)	(0.440)	(0.474)		(0.423)
NumberUnis	-0.203***	0.059	0.142	0.131	0.109
	(0.070)	(0.080)	(0.177)	(0.407)	(0.206)
MeanAcademicAge	-0.001	-0.014	-0.016	-0.311**	-0.008
	(0.007)	(0.012)	(0.012)	(0.150)	(0.012)
MeanAcademicAge^2	0.000	0.000	0.000	0.001**	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
NumReferences	0.016***	0.016***	0.016***	0.016***	0.016***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
NumSelfReferences	-0.001	-0.010	-0.007	-0.006	0.000
	(0.006)	(0.007)	(0.007)	(0.009)	(0.007)
OrderAuthors	0.160**	0.374***	0.356***	0.080	0.285***
	(0.079)	(0.088)	(0.095)	(0.193)	(0.087)
Year dummies	Yes	Yes	Yes	Yes	Yes
Individual dummies	No	Yes	Yes	No	Yes
Authorgroup dummies	No	No	No	Yes	No
University dummies†	No	No	Yes	Yes	Yes
Journal dummies	No	No	No	No	Yes
N (df)	6958 (6926)	6958 (5991)	6958 (5746)	6958 (4917)	6958 (4900)
Pseudo-R^2	0.45	0.68	0.71	0.78	0.79

Notes: This table reports estimated coefficients from quasi-Poisson regressions with standard errors in parentheses. Standard errors are clustered at the journal-year level except for model (4) which is clustered at the author group level (variables as described in Table 4.2). Each observation is an article of which all authors are in the sample described of 967 eminent economists. †: Universities with more than two observations.

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

Table 4.7: Impact of local colleagues by age and journal profile

	Dependent variable:				
	SumReceivedCitations *(link: log)*				
	(1)	(2)	(3)	(4)	(5)
Colleague:Top25Journal:MaxAge5	0.300**	0.394***	0.417***	0.561***	0.308**
	(0.123)	(0.126)	(0.133)	(0.200)	(0.146)
Colleague:Top25Journal:OtherAgeProfile	0.129*	0.283***	0.265***	0.232*	0.197**
	(0.069)	(0.085)	(0.084)	(0.132)	(0.091)
Colleague:OtherJournal:MaxAge5	-0.038	0.062	-0.006	0.122	-0.117
	(0.139)	(0.126)	(0.135)	(0.182)	(0.145)
Colleague:OtherJournal:OtherAgeProfile	0.115	0.040	-0.006	-0.015	-0.033
	(0.127)	(0.109)	(0.091)	(0.125)	(0.085)
CoAuthor:Top25Journal:MaxAge5	-0.125	-0.221	-0.189	0.087	-0.161
	(0.215)	(0.245)	(0.237)	(0.424)	(0.243)
CoAuthor:Top25Journal:OtherAgeProfile	0.056	0.077	0.008	0.088	-0.053
	(0.082)	(0.105)	(0.108)	(0.167)	(0.122)
CoAuthor:OtherJournal:MaxAge5	0.527**	0.519**	0.596**	0.518*	0.548**
	(0.219)	(0.209)	(0.238)	(0.305)	(0.260)
CoAuthor:OtherJournal:OtherAgeProfile	0.235**	0.342***	0.320***	0.265*	0.259***
	(0.112)	(0.103)	(0.087)	(0.144)	(0.089)
Top25Journal	0.774***	0.805***	0.794***	0.783***	
	(0.054)	(0.054)	(0.053)	(0.072)	
MaxAge5	0.078	0.049	0.038	0.012	0.037
	(0.086)	(0.087)	(0.087)	(0.142)	(0.092)
NumAuthors	0.569***	0.169	0.173		0.280
	(0.080)	(0.370)	(0.388)		(0.336)
NumberUnis	-0.205***	0.057	0.179	0.339	0.106
	(0.072)	(0.078)	(0.208)	(0.481)	(0.246)
MeanAcademicAge	0.005	-0.008	-0.009	-0.319**	-0.004
	(0.009)	(0.015)	(0.014)	(0.140)	(0.016)
MeanAcademicAge^2	0.000	0.000	0.000	0.001	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
NumReferences	0.016***	0.016***	0.017***	0.016***	0.016***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
NumSelfReferences	-0.001	-0.009	-0.005	-0.004	0.001
	(0.006)	(0.007)	(0.007)	(0.009)	(0.007)
OrderAuthors	0.310***	0.370***	0.352***	0.175	0.259***
	(0.054)	(0.068)	(0.070)	(0.181)	(0.063)
Year dummies	Yes	Yes	Yes	Yes	Yes
Individual dummies	No	Yes	Yes	No	Yes
Authorgroup dummies	No	No	No	Yes	No
University dummies†	No	No	Yes	Yes	Yes
Journal dummies	No	No	No	No	Yes
N (df)	6958 (6923)	6958 (5988)	6958 (5725)	6958 (4899)	6958 (4879)
Pseudo-R^2	0.45	0.68	0.71	0.79	0.79

Notes: This table reports estimated coefficients from quasi-Poisson regressions with standard errors in parentheses. Standard errors are clustered at the journal-year level except for model (4) which is clustered at the author group level (variables as described in Table 4.2). Each observation is an article of which all authors are in the sample described of 967 eminent economists. †: Universities with more than two observations.

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

Table 4.8: Overall impact of peers

	Dependent variable:				
	SumReceivedCitations *(link: log)*				
	(1)	(2)	(3)	(4)	(5)
Colleague	0.235***	0.196***	0.195***	0.219**	0.124*
	(0.055)	(0.067)	(0.064)	(0.090)	(0.073)
CoAuthor	0.103	0.122	0.075	0.059	0.065
	(0.065)	(0.082)	(0.079)	(0.126)	(0.076)
NumAuthors	0.679***	0.325	0.326		0.315
	(0.082)	(0.359)	(0.372)		(0.341)
NumberUnis	-0.196***	0.106	0.243	0.357	0.180
	(0.073)	(0.082)	(0.251)	(0.599)	(0.252)
MeanAcademicAge	0.000	-0.018	-0.019	-0.408***	-0.010
	(0.008)	(0.013)	(0.013)	(0.141)	(0.013)
MeanAcademicAge^2	0.000	0.000	0.000	0.001*	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
NumReferences	0.017***	0.018***	0.018***	0.017***	0.016***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
NumSelfReferences	-0.013**	-0.017**	-0.013*	-0.011	0.001
	(0.006)	(0.007)	(0.007)	(0.008)	(0.007)
OrderAuthors	0.351***	0.401***	0.383***	0.261	0.261***
	(0.057)	(0.073)	(0.072)	(0.173)	(0.063)
Year dummies	Yes	Yes	Yes	Yes	Yes
Individual dummies	No	Yes	Yes	No	Yes
Authorgroup dummies	No	No	No	Yes	No
University dummies†	No	No	Yes	Yes	Yes
Journal dummies	No	No	No	No	Yes
N (df)	6958 (6931)	6958 (5996)	6958 (5733)	6958 (4907)	6958 (4886)
Pseudo-R^2	0.39	0.63	0.67	0.75	0.79

Notes: This table reports estimated coefficients from quasi-Poisson regressions with standard errors in parentheses. Standard errors are clustered at the journal-year level execept for model (4) which is clustered at the authorgroup level. Variables as described in Table 4.2. Each observation is an article of which all authors are in the described sample of 967 eminent economists. †: Universities with more than two observations.

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

Table 4.9: Overall impact of peers in less restrictive sample

	Dependent variable:				
	SumReceivedCitations *(link: log)*				
	(1)	(2)	(3)	(4)	(5)
Colleague:Top25Journal	0.251***	0.181***	0.189***	0.234***	0.146***
	(0.048)	(0.044)	(0.044)	(0.066)	(0.043)
Colleague:OtherJournal	0.093	0.013	0.014	0.019	0.007
	(0.060)	(0.053)	(0.054)	(0.061)	(0.047)
CoAuthor:Top25Journal	0.033	0.055	0.037	0.052	0.009
	(0.067)	(0.057)	(0.055)	(0.074)	(0.054)
CoAuthor:OtherJournal	0.307***	0.283***	0.267***	0.258***	0.186***
	(0.059)	(0.053)	(0.049)	(0.060)	(0.044)
Top25Journal	0.784***	0.698***	0.690***	0.685***	
	(0.036)	(0.040)	(0.038)	(0.041)	
NumAuthors	0.079**	0.068*	0.071*		0.009
	(0.033)	(0.040)	(0.039)		(0.014)
NumberUnis	0.333***	-0.049	0.057	-0.162	0.073
	(0.038)	(0.062)	(0.132)	(0.277)	(0.121)
MeanAcademicAge	-0.006	-0.013*	-0.012	-0.245***	-0.007
	(0.005)	(0.008)	(0.008)	(0.084)	(0.007)
MeanAcademicAge^2	0.000*	0.000***	0.000**	0.000**	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
NumReferences	0.016***	0.015***	0.015***	0.015***	0.015***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
NumSelfReferences	0.007*	-0.001	-0.003	-0.002	0.005
	(0.004)	(0.004)	(0.004)	(0.005)	(0.003)
OrderAuthors	0.045*	0.031	0.020	0.005	0.002
	(0.027)	(0.029)	(0.029)	(0.040)	(0.022)
Year dummies	Yes	Yes	Yes	Yes	Yes
Individual dummies	No	Yes	Yes	No	Yes
Authorgroup dummies	No	No	No	Yes	No
University dummies†	No	No	Yes	Yes	Yes
Journal dummies	No	No	No	No	Yes
N (df)	25149 (25119)	25149 (24159)	25149 (23863)	25149 (22372)	25149 (22018)
Pseudo-R^2	0.42	0.57	0.58	0.65	0.69

Notes: This table reports estimated coefficients from quasi-Poisson regressions with standard errors in parentheses. Standard errors are clustered at the journal-year level except for model (4) which is clustered at the author group level (variables as described in Table 4.2). Each observation is an article of which all authors are in the sample described of 967 eminent economists. †: Universities with more than two observations.

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

Table 4.10: Impact of peers excluding the most cited economists

	Dependent variable:				
	SumReceivedCitations *(link: log)*				
Without the Top	100	200	300	400	500
Colleague:Top25Journal	0.458***	0.324***	0.249**	0.278**	0.457**
	(0.069)	(0.106)	(0.111)	(0.120)	(0.214)
Colleague:OtherJournal	0.048	-0.017	-0.095	-0.127	-0.018
	(0.098)	(0.114)	(0.136)	(0.165)	(0.277)
CoAuthor:Top25Journal	0.018	0.035	0.040	0.248**	-0.075
	(0.103)	(0.112)	(0.125)	(0.112)	(0.229)
CoAuthor:OtherJournal	0.174*	0.143	0.224*	0.228	0.258
	(0.090)	(0.125)	(0.124)	(0.160)	(0.314)
Top25Journal	0.759***	0.760***	0.739***	0.728***	0.746***
	(0.046)	(0.061)	(0.076)	(0.101)	(0.108)
NumAuthors	0.245	-0.242	-0.015	-0.071	0.324*
	(0.362)	(0.311)	(0.249)	(0.172)	(0.179)
NumberUnis	-0.713**	-0.324	-0.230	-0.113	-0.127
	(0.307)	(0.266)	(0.325)	(0.338)	(0.352)
MeanAcademicAge	-0.026*	-0.035**	-0.039**	-0.012	-0.045***
	(0.015)	(0.015)	(0.016)	(0.017)	(0.014)
MeanAcademicAge^2	0.000	0.000	0.000	0.000	0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
NumReferences	0.018***	0.018***	0.017***	0.016***	0.017***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
NumSelfReferences	0.014**	0.012*	0.008	0.012	0.019
	(0.006)	(0.007)	(0.007)	(0.009)	(0.013)
OrderAuthors	0.287***	0.332***	0.508**	0.618***	0.191
	(0.095)	(0.123)	(0.215)	(0.213)	(0.222)
Year dummies	Yes	Yes	Yes	Yes	Yes
Indiviudal dummies	Yes	Yes	Yes	Yes	Yes
University† dummies	Yes	Yes	Yes	Yes	Yes
N (df)	5428 (4342)	4353 (3396)	3424 (2589)	2612 (1908)	1982 (1396)
Pseudo-R^2	0.71	0.71	0.72	0.73	0.73

Notes: This table reports estimated coefficients from quasi-Poisson regressions with standard errors in parentheses. Standard errors are clustered at the journal-year level (variables as described in Table 4.2). Each observation is an article of which all authors are in the sample described of 967 eminent economists. †: Universities with more than two observations.

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

Table 4.11: Logistic regressions: top journal publications

	Dependent variable:		
	Top25Journal (link: logit)		
	(1)	(2)	(3)
CoAuthor	0.050	0.099	-0.002
	(0.090)	(0.102)	(0.142)
Colleague	0.577***	0.261***	0.082
	(0.079)	(0.089)	(0.136)
MeanAcademicAge	0.021*	0.011	0.008
	(0.011)	(0.014)	(0.030)
MeanAcademicAge^2	-0.001***	-0.001***	-0.001
	(0.000)	(0.000)	(0.001)
NumAuthors	0.556***	0.449***	0.190
	(0.128)	(0.150)	(0.570)
NumberUnis	-0.051	-0.465**	0.017
	(0.095)	(0.217)	(0.215)
NumReferences	0.017***	0.018***	0.023***
	(0.003)	(0.003)	(0.004)
NumSelfReferences	-0.079***	-0.071***	-0.063***
	(0.010)	(0.012)	(0.018)
OrderAuthors	0.605***	0.624***	0.736***
	(0.130)	(0.157)	(0.271)
Year dummies	Yes	Yes	Yes
Individual dummies	No	No	Yes
University dummies†	No	Yes	No
N (df)	6958 (6931)	6958 (6661)	6958 (5996)
Pseudo-R^2	0.10	0.21	0.39

Notes: This table reports estimated coefficients from a logistic regressions with clustered standard errors in parentheses. Variables as described in Table 4.2. Each observation is an article of which all authors are in the described sample of 967 eminent economists. †: Model(3) contains dummies only for universities with more than three observations.

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

Table 4.12: Impact of colleagues and co-authors seperately

	Dependent variable:			
	SumReceivedCitations *(link: log)*			
	(1)	(2)	(3)	(4)
Colleague:Top25Journal			0.173***	0.270***
			(0.066)	(0.089)
Colleague:OtherJournal			0.175*	0.104
			(0.098)	(0.079)
CoAuthor:Top25Journal	0.085	0.068		
	(0.086)	(0.101)		
CoAuthor:OtherJournal	0.315***	0.303***		
	(0.095)	(0.083)		
Top25Journal	0.814***	0.856***	0.766***	0.764***
	(0.052)	(0.048)	(0.052)	(0.050)
NumAuthors	0.615***	0.253	0.608***	0.200
	(0.097)	(0.429)	(0.099)	(0.396)
NumberUnis	-0.206***	0.268	-0.211***	0.238
	(0.072)	(0.211)	(0.073)	(0.214)
MeanAcademicAge	-0.004	-0.022*	-0.000	-0.013
	(0.007)	(0.012)	(0.007)	(0.012)
MeanAcademicAge^2	0.000	0.000*	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
NumReferences	0.016***	0.017***	0.016***	0.017***
	(0.002)	(0.002)	(0.002)	(0.002)
NumSelfReferences	-0.001	-0.006	0.000	-0.006
	(0.006)	(0.007)	(0.006)	(0.007)
OrderAuthors	0.142*	0.369***	0.144*	0.360***
	(0.084)	(0.094)	(0.083)	(0.096)
Year dummies	Yes	Yes	Yes	Yes
Individual dummies	No	Yes	No	Yes
University dummies†	No	Yes	No	Yes
N (df)	6958 (6930)	6958 (5732)	6958 (6930)	6958 (5732)
Pseudo-R^2	0.44	0.71	0.44	0.71

Notes: This table reports estimated coefficients from quasi-Poisson regressions with standard errors in parentheses. Standard errors are clustered at the journal-year level. The Variables are described in Table 4.2. Each observation is an article of which all authors are in the sample described of 967 eminent economists. †: Universities with more than two observations.

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

Table 4.13: Impact of peers by journal profile with overall cited eminent economists

	Dependent variable:				
	SumReceivedCitations *(link: log)*				
	(1)	(2)	(3)	(4)	(5)
Colleague:Top25Journal	0.130**	0.250***	0.239***	0.227*	0.156*
	(0.062)	(0.081)	(0.080)	(0.117)	(0.089)
Colleague:OtherJournal	0.054	0.011	-0.034	-0.031	-0.088
	(0.105)	(0.094)	(0.080)	(0.105)	(0.076)
CoAuthor:Top25Journal	-0.017	-0.017	-0.079	0.000	-0.123
	(0.081)	(0.099)	(0.098)	(0.168)	(0.107)
CoAuthor:OtherJournal	0.251**	0.332***	0.328***	0.275**	0.271***
	(0.103)	(0.096)	(0.081)	(0.130)	(0.083)
Top25Journal	0.751***	0.781***	0.772***	0.762***	
	(0.055)	(0.053)	(0.051)	(0.071)	
StarCited	0.018***	0.024***	0.027***	0.033***	0.030***
	(0.004)	(0.005)	(0.005)	(0.007)	(0.006)
NumAuthors	0.607***	0.246	0.225		0.334
	(0.097)	(0.351)	(0.367)		(0.313)
NumberUnis	-0.216***	0.045	0.135	0.205	0.050
	(0.074)	(0.083)	(0.204)	(0.492)	(0.238)
MeanAcademicAge	-0.001	-0.016	-0.018	-0.352***	-0.012
	(0.007)	(0.012)	(0.012)	(0.131)	(0.013)
MeanAcademicAge^2	0.000	0.000	0.000*	0.001**	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
NumReferences	0.013***	0.012***	0.012***	0.011***	0.011***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
NumSelfReferences	-0.001	-0.012*	-0.008	-0.005	-0.001
	(0.006)	(0.007)	(0.007)	(0.009)	(0.007)
OrderAuthors	0.121	0.380***	0.372***	0.025	0.294***
	(0.081)	(0.086)	(0.094)	(0.200)	(0.086)
N (df)	6958 (6927)	6958 (5992)	6958 (5729)	6958 (4903)	6958 (4883)
Pseudo-R^2	0.45	0.68	0.71	0.79	0.79
Year dummies	Yes	Yes	Yes	Yes	Yes
Individual dummies	No	Yes	Yes	No	Yes
Authorgroup dummies	No	No	No	Yes	No
University dummies†	No	No	Yes	Yes	Yes
Journal dummies	No	No	No	No	Yes
N (df)	6958 (6927)	6958 (5992)	6958 (5729)	6958 (4903)	6958 (4883)
Pseudo-R^2	0.45	0.68	0.71	0.79	0.79

Notes: This table reports estimated coefficients from quasi-Poisson regressions with standard errors in parentheses. Standard errors are clustered at the journal-year level except for model (4) which is clustered at the author group level (variables as described in Table 4.2). Each observation is an article of which all authors are in the sample described of 967 eminent economists. †: Universities with more than two observations.

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

4.B Discussion of article influence

To study the impact of research links within a department formally, the influence that a research paper exerts on future research is used as the key definition of success in this paper. It is assumed that a research article's influence depends on the quality of the research, the dissemination process and the current and future interest in the research topic. Quality of a research article is seen as its *potential* contribution to the specific area regardless of the overall interest in the topic or dissemination and visibility effects: the innovation step and how it links to past, concurrent and future research. Dissemination denotes other factors that increase the influence of an article, in particular the reputation of authors and journals in the research area. The reputation gives a quality signal and increases visibility.[27] In an efficiently working scientific field, quality and dissemination should be highly correlated but it is possible, for example, for a high-quality article to have little influence because of very low visibility.[28]

$$I_i = f(Q_i, Dis_i, Int_i) \tag{3}$$

Equation 3 then sets the influence of a research article i as a function of its quality, Q_i, the dissemination process, Dis_i, and the overall interest in the research area, Int_i. In turn, quality and dissemination are potentially helped by the co-location with thematically connected researchers. As it is posited here that local peers potentially can affect the quality of an article during its production, the quality of article i in Equation 4 is set as a function of the researchers in the neighborhood N_i of the article, for instance the departments of the authors. In addition, A_i denotes the authors and U_i are characteristics of the authors' institutes.

$$Q_i = h(U_i, A_i, g(A_i, N_i)) \tag{4}$$

The influence of peers depends potentially on the form of the relationship between the authors and the peers which is expressed by $g(A, N)$. For example, this paper focuses on research proximity and co-location and argues that peers are more helpful if their research is closely related and the researchers co-locate during the research process.

[27] See Dietrich (2008), Feenberg et al. (2015) and Judge et al. (2007) for the importance of visibility for the influence of scientific publications.

[28] A recent example for temporary obscurity of an important research result is the proof of the Gaussian correlation inequality by Thomas Royen (Royen (2014)) that was published in a low-impact journal by the then retired Royen.

Two main arguments suggest potential quality effects from close personal contact to thematically related researchers. First, most knowledge to produce research is tacit and not easily accessible but apprehended, that is transmitted and created, by personal contact (see Polanyi (1958)).[29] Second, new ideas and knowledge are at first circulated in personal connections or created within personal connections.[30] As a result, personal connections lead to a more efficient research process which is particularly important in a dynamic research field: In the trade-off between maturing a research project and ensuring priority by publication , the help and critique of colleagues lead to a more efficient maturation process and, therefore, a higher quality at the time of publication (see Bobtcheff et al. (2016)). In this process, peers can, for example, improve how the research process links with related research and make it, therefore, more relevant to a wider audience or they can provide feedback and directions for technical or conceptual questions.

The importance of tacit knowledge may change over the career of a researcher. Established researchers possess most of the relevant tacit knowledge within their discipline. However, this shared tacit knowledge facilitates, subsequently, the creation and discussion of new ideas. Therefore, interactions with other researchers are more fruitful if the research focus is similar. As a consequence, the importance of personal connections is expected to be at its highest at the very beginning of a career. Lower but stable effects are expected in later career stages.[31]

Besides quality effects, personal connections can impact also on the dissemination process more directly. First, local colleagues are potentially more aware of and influenced by ongoing research by the authors. Second, connected researchers can influence the wider dissemination process, for example, as journal editors. Furthermore, a high number of personally connected researchers working in a given sub-field may indicate the authors' or

[29] While all knowledge is somewhat tacit or rooted in tacit knowledge, tacit knowledge is defined here as knowledge that is not easily codified and made accessible to other scientists, for example, the appreciation of a mathematical theory by scientists before experimental observation and the theory's relation to the observation thereafter (Polanyi (1958), p.60).

[30] See for instance Borowiecki (2013), Hellmanzik (2010), and Mitchell (2016) for evidence for this concept in artistic production.

[31] As argued by Kim et al. (2009), cheap air travel and the Internet can change the setting of this knowledge transfers and creation as it allows for an easier connection with distant colleagues. These connections though have a higher cost and are less likely. In addition, lower communication costs, higher mobility of scientists and the increased specialization of research, can threaten departments as a common social and work space. As a result, the departments' role in enabling new research connections and mutual assistance between researchers is potentially weakened. Through the Internet and air travel, scientists can find a better matching of skills with researchers at distant universities and subsequently stay clear of department colleagues. This has potentially negative effects for less experienced researchers without a network and recalls the image of Schopenhauer's "republic of geniuses" (as opposed to the "republic of scholars") in which according to Nietzsche "One giant calls to the other across the bleak intervals of time and the conversation of the great minds goes on undisturbed by the mischievous, noisy dwarfs who creep among them."(Nietzsche (1872), p.269)). Analyzing research articles that are explicitly influenced by a colleague allows us to assess peer effects without being as strongly affected in the analysis by this development.

department's prominent standing in the sub-field. From the perspective of a potentially influenced researcher, this reputation is a quality signal which reduces the effort for quality screening. In addition, aligning with influential research, which is signaled by reputation, makes research more likely to be of interest to (or commensurable with) future research.

Therefore, the reputations of the authors A and of their affiliations U and peers, and the type of journal impact positively on the dissemination of a research article. In turn, the journal depends among others on the quality of the article. Taken together, dissemination can be set as $Dis = Dis(A, U, N, J(Q))$. As a consequence, peers in the neighborhood N can impact an article's influence through quality and dissemination which leads to the following overall relation between influence and peers in Equation 5, ceteris paribus.

$$I(N) = f\Big(Q\big(N\big), Dis\big(N, Q(N)\big), Int\Big) \tag{5}$$

For a given set of authors and universities, the take-up of a research article by future research can be improved by connected researchers within the department through two channels: the article's quality and its dissemination.

4.C Alternatives for the assumptions in the conceptual framework

4.C.1 Quality - Assumption 1 and Corollary 1

Assumption 3. *The influence of an article is set as the function $I = I(Q(x), J)$, where x is a variable that affects I foremost through enabling the authors to increase the quality Q of the article and $J \in \{highprofile, lowprofile\}$ is the reputation of the publishing journal (relative to the author's reputation). Then, the following relation is posited.*

- *In contrast to innovative cutting-edge research (high-profile), a routine-type (low-profile) article has no significant scope for quality improvement behind the authors' own effort and ability.*

$$\frac{\partial I(Q(x), j^{high})}{\partial x} > \frac{\partial I(Q(x), j^{low})}{\partial x} \approx 0$$

- *The ratio of this two derivatives is, therefore, close to zero and smaller than the ratio of the overall influence of the high- and low-profile article.*

$$\frac{\frac{\partial I(Q(x), j^{low})}{\partial x}}{\frac{\partial I(Q(x), j^{high})}{\partial x}} < \frac{I(Q(x), j^{low})}{I(Q(x), j^{high})}$$

4.C.2 Reputation - Assumption 2 and Corollary 2

This reminder of this subsection provides, first, additional steps from Assumption 2 to Corollary 2 and, second, an alternative assumption. It follows from Assumption 2 that:

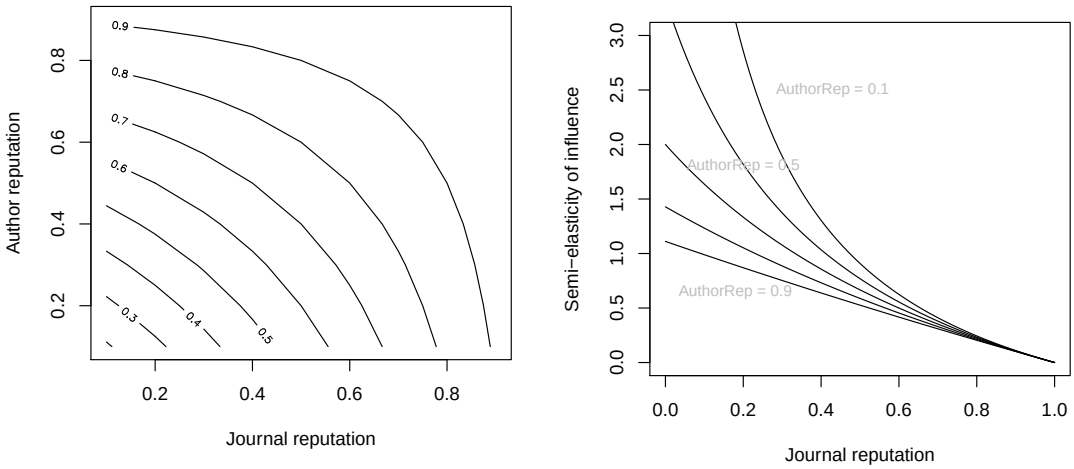
$$\begin{aligned} \frac{I_z(R(z), j^{high})}{I_z(R(z), j^{low})} &< \frac{I(R(z), j^{high})}{I(R(z), j^{low})} \\ \frac{I_z(R(z), j^{high})}{I(R(z), j^{high})} &< \frac{I_z(R(z), j^{low})}{I(R(z), j^{low})} \\ \frac{\partial log(I(R(z), j^{high}))}{\partial z} &< \frac{\partial log(I(R(z), j^{low}))}{\partial z} \end{aligned} \tag{6}$$

which corresponds to Corollary 2.

Figure 4.5 illustrates the relation between author and journal reputation as alternative to the stronger version of Assumption 2. Assume, article i is published in a research area with n future publications. A part $p, p \in [0, 1]$, of the n future authors reads article i and a further part $pQ, Q \in [0, 1]$, decides to relate it to their research, that is cite article i. Therefore, the article's influence is set as $I = pnQ$, with $p(AuthorRep, JournalRep)$ and $\frac{\partial^2 p(AuthorRep, JournalRep)}{\partial AuthorRep\, \partial JournalRep} < 0$. To illustrate this condition, Figure 4.5 shows the function $p(AuthorRep, JournalRep) = AuthorRep + JournalRep - AuthorRep * JournalRep$ and the consequent semi-elasticity of the influence I for a change in author reputation. In this example, the reputation effect of the author is close to zero for a very prestigious journal, for instance JournalRep=.9[32] while the visibility of publications in low-profile journals benefit strongly from an eminent author.

[32] That is 90 percent in the research area consider publications in this journal as relevant.

Figure 4.5: Example for the relation between author and journal reputation



(a) Probability that a researcher in the area considers the publication relevant based on journal and author reputation.

(b) Semi-elasticity of an article's influence I for a change in author reputation

Notes: Based on the influence function $I = pnQ = (AuthorRep + JournalRep - AuthorRep * JournalRep)nQ$

Chapter 5

Conclusion

The three essays in this thesis show that economic research has become more internationally integrated. The role of borders, and geographic and cultural distance in the dissemination of research knowledge has decreased and co-authorship has increased substantially within and across borders. However, research links between department colleagues help their research output which shows local effects.

Chapter 2 has provided evidence on the unequal dissemination of economic research between countries along cultural and technological links. However, the observed home bias does not necessarily imply a shortcoming, as many research questions indeed might be of a local nature and, therefore, of limited appeal to the rest of the world. It is left to further research to conceptualise an optimal or natural home bias for comparison as for instance for the financial equity home bias in Park and Mercado (2014).

In Chapter 3, several hypotheses from the economic literature for the rise of co-authorship are summarised. Other literatures emphasise different factors, for instance the higher epistemological authority of co-authored research (Beaver (2004)).[1] Though this advantage would have preceded the rise in co-authorship. If changing incentives bring economists to realise productivity gains, co-authorship might be seen as a sign for the professionalisation of economic research. We also have to be careful not to overstress the role of technological change for the rise of co-authorship in economics which began two or three decades before the internet and the common use of computers. Empirically, it is difficult to single out reasons for the broad development towards more co-authorship across all scientific fields.

In Chapter 4, the take-up of an article by future research is shown to be higher if linked with the research of local colleagues. An increase in specialisation and a decrease of collaboration

[1] "Philosophically (and sociologically), it seems that the chief epistemological advantage of collaboration is that it confers the benefits of intersubjective verifiability" (Beaver (2004)). This term means to establish truth through exchange between people .

within departments could, therefore, explain the elsewhere observed decline of local effects (Kim et al. (2009)). Conversely, this chapter points to benefits for a department's research output by increased local collaboration. However, the benefit arises from realised research links and not necessarily specialised clusters. Chapter 4 provides so an argument for universities and research institutes as a location of common research and scientific exchange between colleagues.

Bibliography

Agrawal, A. and Goldfarb, A. (2008). Restructuring research: Communication costs and the democratization of university innovation. *American Economic Review*, 98(4):1578–90.

Agrawal, A. K., McHale, J., and Oettl, A. (2014). Why stars matter. Working Paper 20012, National Bureau of Economic Research.

Anderson, J. E. and van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review*, 93(1):170–192.

Azoulay, P., Zivin, J. S. G., and Wang, J. (2010). Superstar extinction. *The Quarterly Journal of Economics*, 125(2):549–589.

Baerlocher, M. O., Newton, M., Gautam, T., Tomlinson, G., and Detsky, A. S. (2007). The meaning of author order in medical research. *Journal of Investigative Medicine*, 55(4):174–180.

Balta, N. and Delgado, J. (2009). Home bias and market integration in the eu. *CESifo Economic Studies*, 55(1):110–144.

Barnett, A. H., Ault, R. W., and Kaserman, D. L. (1988). The rising incidence of co-authorship in economics: further evidence. *Review of Economics and Statistics*, 70:539–543.

Barnett, G. A., Chon, B.-S., and Rosen, D. (2001). The structure of the internet flows in cyberspace. *NETCOM*, 15(2):61–80.

Barnett, G. A. and Sung, E. (2005). Culture and the structure of the international hyperlink network. *Journal of Computer-Mediated Communication*, 11(1):217–238.

Beaver, D. d. (2004). Does collaborative research have greater epistemic authority? *Scientometrics*, 60(3):399–408.

Blaug, M. and Vane, H. (2003). *Who's Who in Economics*. Edward Elgar Pub.

Blum, B. S. and Goldfarb, A. (2006). Does the internet defy the law of gravity? *Journal of International Economics*, 70(2):384–405.

Bobtcheff, C., Bolte, J., and Mariotti, T. (2016). Researcher's dilemma*. *The Review of Economic Studies*.

Bolli, T. and Schläpfer, J. (2015). Job mobility, peer effects, and research productivity in economics. *Scientometrics*, 104(3):629–650.

Borjas, G. J. and Doran, K. B. (2015). Which peers matter? the relative impacts of collaborators, colleagues, and competitors. *Review of Economics and Statistics*, 97(5):1104–1117.

Börner, K., Penumarthy, S., Meiss, M., and Ke, W. (2006). Mapping the diffusion of scholarly knowledge among major us research institutions. *Scientometrics*, 68(3):415–426.

Bornmann, L. and Daniel, H.-D. (2008). What do citation counts measure? a review of studies on citing behavior. *Journal of Documentation*, 64(1):45–80.

Borowiecki, K. J. (2013). Geographic clustering and productivity: An instrumental variable approach for classical composers. *Journal of Urban Economics*, 73(1):94 – 110.

Card, D. and DellaVigna, S. (2013). Nine facts about top journals in economics. *Journal of Economic Literature*, 51(1):144–161.

Card, D. and DellaVigna, S. (2014). Page limits on economics articles: Evidence from two journals. *Journal of Economic Perspectives*, 28(3):149–68.

Catalini, C., Fons-Rosen, C., and Gaule, P. (2016). Did cheaper flights change the direction of science? CEPR Discussion Paper 11252.

Chen, N. (2004). Intra-national versus international trade in the european union: why do national borders matter? *Journal of international Economics*, 63(1):93–118.

Choi, C. (2010). The effect of the Internet on service trade. *Economics Letters*, 109(2):102–104.

Chung, C. J. (2011). The geography of global internet hyperlink networks and cultural content analysis. *Dissertation, University at Buffalo*.

Dietrich, J. P. (2008). Disentangling visibility and self-promotion bias in the arxiv:astro-ph positional citation effect. *Publications of the Astronomical Society of the Pacific*, 120(869):801.

Ductor, L. (2015). Does co-authorship lead to higher academic productivity? *Oxford Bulletin of Economics and Statistics*, 77(3):385–407.

Ductor, L., Fafchamps, M., Goyal, S., and van der Leij, M. J. (2014). Social networks and research output. *The Review of Economics and Statistics*, 96(5):936–948.

Duranton, G. and Puga, D. (2004). Micro-foundations of urban agglomeration economies. In Henderson, J. V. and Thisse, J. F., editors, *Handbook of Regional and Urban Economics*, volume 4 of *Handbook of Regional and Urban Economics*, chapter 48, pages 2063–2117. Elsevier.

Fafchamps, M., van der Leij, M. J., and Goyal, S. (2010). Matching and network effects. *Journal of the European Economic Association*, 8(1):203–231.

Fally, T. (2015). Structural gravity and fixed effects. *Journal of International Economics*, 97(1):76–85.

Feenberg, D. R., Ganguli, I., Gaule, P., and Gruber, J. (2015). It's good to be first: Order bias in reading and citing NBER working papers. NBER Working Papers 21141, National Bureau of Economic Research, Inc.

Felbermayr, G. J. and Toubal, F. (2010). Cultural proximity and trade. *European Economic Review*, 54(2):279–293.

Fourcade, M., Ollion, E., and Algan, Y. (2015). The superiority of economists. *Journal of Economic Perspectives*, 29(1):89–114.

Francois, J. and Hoekman, B. (2010). Services trade and policy. *Journal of Economic Literature*, 48(3):642–92.

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

Frenken, K. and Hoekman, J. (2014). Spatial scientometrics and scholarly impact: A review of recent studies, tools, and methods. In *Measuring scholarly impact*, pages 127–146. Springer.

Freund, C. and Weinhold, D. (2002). The internet and international trade in services. *American Economic Review*, 92(2):236–240.

Frey, B. S. and Eichenberger, R. (1993). American and european economics and economists. *Journal of Economic Perspectives*, 7(4):185–193.

Gloetzl, F. and Aigner, E. (2017). Six dimensions of concentration in economics: Scientometric evidence from a large-scale data set. Ecological Economics Papers ieep15, Institute of Ecological Economics.

Goyal, S., van der Leij, M. J., and Moraga-Gonzalez, J. L. (2006). Economics: An emerging small world. *Journal of Political Economy*, 114(2):403–432.

Hamermesh, D. S. (2013). Six decades of top economics publishing: Who and how? *Journal of Economic Literature*, 51(1):162–72.

Head, K. and Mayer, T. (2002). *Illusory border effects: Distance mismeasurement inflates estimates of home bias in trade*, volume 1. CEPII Paris.

Hellmanzik, C. (2010). Location matters: Estimating cluster premiums for prominent modern artists. *European Economic Review*, 54(2):199–218.

Hellmanzik, C. and Schmitz, M. (2015). Virtual proximity and audiovisual services trade. *European Economic Review*, 77:82–101.

Hellmanzik, C. and Schmitz, M. (2017). Taking gravity online: the role of virtual proximity in international finance. *Journal of International Money and Finance*, 77:164–179.

Henriksen, D. (2016). The rise in co-authorship in the social sciences (1980–2013). *Scientometrics*, 107(2):455–476.

Hudson, J. (1996). Trends in multi-authored papers in economics. *Journal of Economic Perspectives*, 10(3):153–158.

Jones, B. F. (2009a). The burden of knowledge and the "death of the renaissance man": Is innovation getting harder? *The Review of Economic Studies*, 76(1):283.

Jones, B. F. (2009b). The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder? *Review of Economic Studies*, 76(1):283–317.

Judge, T. A., Cable, D. M., Colbert, A. E., and Rynes, S. L. (2007). What causes a management article to be cited: Article, author, or journal? *The Academy of Management Journal*, 50(3):491–506.

Kalaitzidakis, P., Mamuneas, T. P., and Stengos, T. (2011). An updated ranking of academic journals in economics. *Canadian Journal of Economics*, 44(4):1525–1538.

Kim, E. H., Morse, A., and Zingales, L. (2009). Are elite universities losing their competitive edge? *Journal of Financial Economics*, 93(3):353–381.

Kimura, F. and Lee, H.-H. (2006). The gravity equation in international trade in services. *Review of World Economics*, 142(1):92–121.

Kuld, L. (2017). The university as a local idea space: benefits from research links. Working paper, Trinity College Dublin.

Kuld, L. and O'Hagan, J. (2017). Rise of multi-authored papers in economics: demise of the 'lone star' and why? Working paper, Trinity College Dublin.

Lazear, E. P. (2000). Economic imperialism*. *The Quarterly Journal of Economics*, 115(1):99–146.

Manski, C. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, 60(3):531–542.

Matthiessen, C., Schwarz, A., and Find, S. (2002). The top-level global research system, 1997-99: Centres, networks and nodality. an analysis based on bibliometric indicators. *Urban Studies*, 39(5-6):903–927.

McCallum, J. (1995). National borders matter: Canada-u.s. regional trade patterns. *The American Economic Review*, 85(3):615–623.

McDowell, J. M. and Michael, M. (1983). The determinants of co-authorship: an analysis of the economics literature. *Review of Economics and Statistics*, 65:155–160.

Medoff, M. H. (2003). Collaboration and the quality of economics research. *Labour Economics*, 10(5):597–608.

Melitz, J. and Toubal, F. (2014). Native language, spoken language, translation and trade. *Journal of International Economics*, 93(2):351–363.

Melitz, M. and Ottaviano, G. (2008). Market size, trade, and productivity. *Review of Economic Studies*, 75:295–316.

Mitchell, S. (2016). London calling? Agglomeration Economies in Literature since 1700. Working paper, Trinity College Dublin.

Moosa, I. A. (2017). Citations, journal ranking and multiple authorships: evidence based on the top 300 papers in economics. *Applied Economics Letters*, 24(3):175–181.

Nietzsche, F. (1954, 1E 1872). *Werke in drei Bänden. Band 1.* Hanser.

Nitsch, V. (2000). National borders and international trade: evidence from the european union. *Canadian Journal of Economics/Revue canadienne d'économique*, 33(4):1091–1105.

Nowell, C. and Grijalva, T. (2011). Trends in co-authorship in economics since 1985. *Applied Economics*, 43(28):4369–4375.

Oettl, A. (2012). Reconceptualizing stars: Scientist helpfulness and peer performance. *Management Science*, 58(6):1122–1140.

Önder, A. S. and Schweitzer, S. (2016). Catching up or falling behind? promising changes and persistent patterns across cohorts of economics PhDs in German-speaking countries from 1991 to 2008. Working paper, University of Bayreuth.

Ossenblok, T. L., Verleysen, F. T., and Engels, T. C. (2014). Coauthorship of journal articles and book chapters in the social sciences and humanities (2000-2010). *Journal of the Association for Information Science and Technology*, 65(5):882–897.

Osterloh, M. and Frey, B. S. (2014). Academic rankings between the 'republic of science' and 'new public management'. In Lanteri, A. and Vromen, J., editors, *The Economics of Economists*, pages 77 – 103. CUP.

Pan, R. K., Kaski, K., and Fortunato, S. (2012). World citation and collaboration networks: uncovering the role of geography in science. *Scientific reports*, 2.

Park, C.-Y. and Mercado, R. V. (2014). Equity home bias, financial integration, and regulatory reforms: implications for emerging Asia. In *Global Shock, Risks, and Asian Financial Reform*, Chapters, chapter 9, pages 347–376. Edward Elgar Publishing.

Polanyi, M. (1958). *Personal Knowledge: Towards a Post-critical Philosophy*. The University of Chicago Press.

Rath, K. and Wohlrabe, K. (2016). Recent trends in co-authorship in economics: evidence from repec. *Applied Economics Letters*, 23(12):897–902.

Rosenblat, T. S. and Mobius, M. M. (2004). Getting closer or drifting apart?*. *The Quarterly Journal of Economics*, 119(3):971.

Royen, T. (2014). A simple proof of the Gaussian correlation conjecture extended to multivariate gamma distributions. *ArXiv e-prints*.

Sarsons, H. (2017a). Recognition for group work. *Working paper*.

Sarsons, H. (2017b). Recognition for group work: Gender differences in academia. *American Economic Review: Papers and Proceedings*.

Sauer, R. D. (1988). Estimates of the returns to quality and coauthorship in economic academia. *Journal of Political Economy*, 96(4):855–866.

Silva, J. M. C. S. and Tenreyro, S. (2006). The Log of Gravity. *The Review of Economics and Statistics*, 88(4):641–658.

Sommer, V. and Wohlrabe, K. (2017). Citations, journal ranking and multiple authorships reconsidered: evidence from almost one million articles. *Applied Economics Letters*, 24(11):809–814.

Tahamtan, I., Safipour Afshar, A., and Ahamdzadeh, K. (2016). Factors affecting number of citations: a comprehensive review of the literature. *Scientometrics*, 107(3):1195–1225.

Thelwall, M. (2009). Introduction to webometrics: Quantitative web research for the social sciences. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 1(1):1–116.

Waldinger, F. (2010). Quality matters: The expulsion of professors and the consequences for phd student outcomes in Nazi Germany. *Journal of Political Economy*, 118(4):787–831.

Waldinger, F. (2012). Peer effects in science: Evidence from the dismissal of scientists in Nazi Germany. *Review of Economic Studies*, 79(2):838–861.

Waldinger, F. (2016). Bombs, brains, and science: The role of human and physical capital for the creation of scientific knowledge. *Review of Economics and Statistics*, 98(5):811–831.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*, volume 1 of *MIT Press Books*. The MIT Press, 2 edition.

Wuchty, S., Jones, B. F., and Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316(5827):1036–1039.