Working Papers in Economics & Finance
2021-04

# Field Distance and Quality in Economists' Collaborations

Ali Sina Onder, University of Portsmouth
Sascha Schweitzer, University of Bayreuth
Hakan Yilmazkuday, Florida International University

Portsmouth Business School
https://www.port.ac.uk/about-us/structure-and-governance/organisational-structure/our-academic-structure/faculty-of-business-and-law/portsmouth-business-school

# Field Distance and Quality in Economists' Collaborations

Ali Sina Önder* Sascha Schweitzer† Hakan Yilmazkuday ‡

**Abstract**

We analyze economics PhDs' collaborations in peer-reviewed journals from 1990 to 2014 and investigate such collaborations' quality in relation to each co-author's research quality, field and specialization. We find that a greater overlap between co-authors' previous research fields is significantly related to a greater publication success of co-authors' joint work and this is robust to alternative specifications. Co-authors that engage in a distant collaboration are significantly more likely to have a large research overlap, but this significance is lost when co-authors' social networks are accounted for. High quality collaboration is more likely to emerge as a result of an interaction between specialists and generalists with overlapping fields of expertise. Regarding interactions across subfields of economics (interdisciplinarity), it is more likely conducted by co-authors who already have interdisciplinary portfolios, than by co-authors who are specialized or starred in different subfields.

**JEL Classification:** A11, A14, I23

**Key Words:** Collaboration; Distance; Team Formation; Research Productivity; Fields of Economics; Stratification; Specialization

*University of Portsmouth, Dept. of Economics and Finance, PO1 3DE, Portsmouth, UK; Tel:+44-2392-844277. *Email:*ali.onder@port.ac.uk

†University of Bayreuth, Dept. of Management, RW1, Universitätsstrasse 30, 95447 Bayreuth, Germany; Tel:+49-921-556063. *Email:*sascha.schweitzer@uni-bayreuth.de

‡Florida International University, Dept. of Economics, 11200 SW 8th Street, 33199 Miami FL, USA; Tel: +1-305-348-2316; *Email:*hyilmazk@fiu.edu

# 1 Introduction

Collaboration has become the dominant mode of research production in many disciplines in recent decades (Gazni et al., 2012; Wuchty et al., 2007). Collaboration may be motivated by career pressures to publish more and better (Ellison, 2002) as well as by the need to circumvent a gap of knowledge or expertise (McDowell and Melvin, 1983). Influence of research collaboration on citation impact is not uniform and varies largely across disciplines (Didegah and Thelwall, 2013; Haslam et al., 2008). However most disciplines, including economics, reveal a strong positive correlation between citation counts and the number of co-authors (Franceschet and Costantini, 2010). As far as economics research is concerned, co-authored papers not only have been dominating the publication scenary for several decades now (Laband and Tollison, 2000; Önder and Yilmazkuday, 2020) but also are more likely to get accepted for publication (Laband and Tollison, 2000) and receive more citations (Chung et al., 2009; Franceschet and Costantini, 2010) than sole author[1] papers.

In this paper, we focus on the outcome (in terms of the journal prestige and citation impact) of economists' collaborations and investigate how similarity and specialization of co-authors' research portfolios are related to the quality of collaboration. Focusing on economists provides a preferable environment for our analysis because research and collaboration in this field still takes place at a very personal level as opposed to laboratory driven research with large research teams as in many of the natural sciences. We use peer-reviewed economics journal articles between 1990 and 2014 of PhD graduates of US and Canadian economics departments whom we refer to as *North American PhDs* throughout this paper. This particular subset of economists can be controlled for educational background and potential social ties from the graduate school, because the American Economic Association provides full lists of all graduating North American PhDs sorted by their graduate department each year. We know that North American PhDs are a particularly influential group in academic publications: 20% of all EconLit papers, more than 50% of all papers in top general and top

[1] Hollis (2001) and Medoff (2003) suggest collaborations to have a negative overall effect on invididivual economists' research productivity, but Ductor (2015) shows that once the endogeneity of collaboration formation is accounted for, the effect of collaboration on invididual productivity becomes positive and significant.

field journals, and about 60% of all papers in the so-called top five have at least one North American PhD on board (Önder and Yilmazkuday, 2020).

Two important features in our study are co-authors' field distance and specialization levels. Co-authors with a very close field distance have publications in similar areas of economics, whereas co-authors with a large field distance have publications in different areas from one another.[2] Authors' specialization levels are calculated as the Herfindahl index of their research portfolios. Our analysis starts with a descriptive part that yields three stylized facts on co-authors' field distance and specialization: (i) Co-authors have become geographically more distant but much closer in terms of field distance over the last couple of decades; (ii) co-authors whose collaboration reveals better quality have a significantly smaller field distance; (iii) co-authors' specialization levels are little or not related to the overall quality of the collaboration. Assuming a two step process for collaborative research where co-authors search and match in the first stage and the quality of their collaboration is revealed in the second stage, we investigate the statistical significance of relations that are picked in these stylized facts. Our estimations reveal that the field distance between co-authors is negatively and significantly related to the quality of their collaborative output. This relation is robust to how quality is measured as well as whether it is the co-authors' first time collaboration or a subsequent collaboration.

Fafchamps et al. (2010) estimate the probability of potential collaborators to initiate and maintain collaboration using a logit framework. They show that the overlap of research areas between potential collaborators is a good indicator of collaboration, however, they do not look into the relation between co-authors' research overlap and the quality of their collaboration output. Our findings complement Fafchamps et al. (2010) such that we show, conditional on forming collaboration, a co-author pair's publications yield better quality when these authors have a close field distance to one another.

There is research documenting that distant collaborations are related to better quality research compared to same location collaborations (Adams et al., 2005; Franceschet and Costantini, 2010). Although the alleged importance of physical proximity between co-authors

[2] A concept similar to our field distance is being referred to as *cognitive distance* in informetrics literature (Nooteboom et al., 2007; Rafols et al., 2010).

is sensitive to the nature and technological context of the research in question (Frenken et al., 2010), and economists might still benefit from positive agglomeration effects that can be offered by large and prestigious departments with significant spillover for their colleagues in certain fields (Bosquet and Combes, 2017), distant collaborations have already become fairly common among economists as documented by Laband and Tollison (2000) and Hamermesh and Oster (2002) among others. Using publications in three general interest economics journals from 1970-1979 and 1992-1996, Hamermesh and Oster (2002) show that the sharp decrease in communication and travel costs over the last couple of decades coincides with an increase in long distance collaborations as these become more affordable. They find mixed results for the relation between collaborators' distance and publication quality. Sutter and Kocher (2004) investigate the effect of distance between collaborators in a gravity model and find no effect of collaborators' distance or other geographic variables on the quality of the collaboration output. Hoekman et al. (2010) as well as Freeman and Huang (2015) point to physical distance as an important ingredient of the quality of collaboration, yet they don't investigate further what is actually driving a distant collaboration. We find that distant and same location collaborations reveal significantly different field distances on average and since field distance is negatively related to research quality, same location collaborations are of less quality, on average. This finding complements the existing literature by providing a possible motive for engaging in distant collaboration, namely, co-authors that engage in a distant collaboration are significantly more likely to have a close field distance, and a close field distance is significantly related to having a high quality outcome for this collaboration.

Our research provides an important link between stratification of collaboration and sorting of researchers along research interests at the same time. This tendency has first been shown by Rosenblat and Mobius (2004) who estimate a significant increase in collaboration between researchers who are at distant locations from one another but work in similar fields. Heterogeneity in authors' types and the bias in their preferences to collaborate with their own type gives rise to authors' separation by type as opposed to location. The data sample of Rosenblat and Mobius (2004) contains co-authored publications only in top eight economics journals from 1980 to 1999, and the relation between the quality of collaboration output, collaborators' geographic distance, and exact field distance has not been investigated. Sorting of researchers along similar research interests is also captured indirectly by Fafchamps et al. (2010) and Ductor (2015) who show that researchers with large overlap of research interest are significantly more likely to engage in collaboration. We offer a channel as to why distant collaborations turn out better than same location collaborations by showing that field distance of co-authors plays a crucial role.

Our extensive data on co-authors' PhD background allows us to detect social networks above and beyond what can be captured in the usual co-author networks. As a result, we are able to show that the information content of geographical distance becomes very small and its relation to field distance turns out insignificant when co-authors' social distance is accounted for. This finding connects nicely to the innovation literature where geographical localization of knowledge spillovers is shown to be captured to a large extent by the social proximity of innovators (Jaffe et al., 1993; Singh, 2005). We show that this holds for economists' collaborations as well.

Specialization is an important aspect of researchers' portfolios and it can have long lasting effects on their careers (Corsi et al., 2019; Leahey, 2007). Specialization in a narrow research area has become an optimal response of academic researchers and industrial innovators to the increase in the amount of knowledge or expertise that is required to achieve a genuine innovation (Jones, 2010; Schweitzer and Brendel, 2020). This has been pointed out as a prompting factor underlying increased collaboration by Wuchty et al. (2007). Nevertheless, research on economists finds that specialization is not necessarily a good thing. Although Corsi et al. (2019) find that specialization (based on JEL codes) has no significant effect on getting promoted to associate professor in Italian economics departments, Ductor (2015) uses a more comprehensive dataset and shows a negative effect of specialization on economists' productivity. Bosquet and Combes (2017) find that an economist's diversity as opposed to specialization is significantly correlated with a better quality publication portfolio in any given field even after their department's overall specialization in that field has been accounted for. Using an article level analysis, Bramoulle and Ductor (2018) show that the specialization level of a co-author team has a positive effect on the journal quality but a negative effect on their citations.

Our contribution to this line of literature is to show how specialization works for and at the same time against the quality of collaboration. A high specialization level has an indirect positive effect on the quality of collaboration output because more specialized authors are more likely to team up with co-authors that have a very close field distance and such closeness is related to a high quality of collaboration output. However, once the indirect effect is accounted for, a high specialization level has a direct negative effect on the quality of collaboration output. The total effect of specialization is negative.

Besides deepening our understanding of economists' collaboration patterns, our analysis has the potential to help faculty hiring committees or research grant committees to make more educated choices and thus help to reach allocational efficiency. The remainder of this paper is structured as follows: We describe our data and main variables in Section 2, then we list and discuss three stylized facts in Section 3. In Section 4, we present our empirical findings. We discuss implications of our findings, put them in context and conclude in Section 5.

## 2 Data

We create the dataset for our analysis by merging complete lists of economics PhDs graduating from US and Canadian economics departments (*North American PhDs*) between 1970 and 2008 with records of peer-reviewed journal publications from 1990 to 2014.[3] We restrict our sample to those author pairs where each author has at least two publications prior to their collaboration. We identify 3,682 two author papers that embody a first time collaboration of two North American PhDs with one another and we analyze this subsample in the first part of the analysis. We analyze their subsequent life time collaborations in the second

[3] Appendix section A provides detailed information about data sources and how individuals in these data are identified and correctly merged.

part of our analysis. Variables[4] that we employ in our analysis are constructed as explained below.

**Quality of Collaboration Output** We measure the quality of a collaboration by the resulting publication's quality. Publication quality is captured by the quality of the journal where the paper is published or by the number of citations collected within five years after its publication. If a co-author pair publishes more than one article in a given year, we take the highest quality publication as the outcome of their collaboration in that year. We use index values of Combes and Linnemer (2010) and Kalaitzidakis et al. (2003) to capture journal quality, and we refer to these quality weights as CL-index and KMS-index, respectively, throughout this paper.[5] The CL-index takes values from zero to one, and it bundles journals into various groups by assigning the same quality weight to journals in the same group. For example, *American Economic Review*, *Econometrica*, *Quarterly Journal of Economics*, *Journal of Political Economy*, and *Review of Economic Studies* make up the highest ranked group, and their quality weight is one. The KMS-index is a continuous weighting scheme that assigns one to the *American Economic Review* and all other journals receive individual quality weights between zero and one. The KMS-index provides a more detailed ranking structure compared to the CL-index. However, it is possible that two papers have highly similar quality but cover completely different fields or use different methodology so that they may have very different suitability for a given journal, depending on that journal's field, focus, and style. The tiered structure of the CL-index captures exactly this and enables us to account for such differences.

The number of citations accumulated within five years after publication is another measure that we use for capturing an article's quality. Citation data was not provided by EconLit, and we obtained these from Aminer.org. We could trace citing articles of most publications

[4]Determinants of research quality is a significant topic on its own right and there is extensive research about it that covers a broad range of disciplines (Bonaccorsi and Secondi, 2017; Carayol and Matt, 2006; Didegah and Thelwall, 2013; Haslam et al., 2008; Vernon et al., 2018) as well as economics specifically (Bosquet and Combes, 2017; Bramoulle and Ductor, 2018; Card and DellaVigna, 2013; Conley et al., 2013; Corsi et al., 2019; Heckman and Moktan, 2018). Most variables that we use in our analysis are among fairly common controls in the above cited literature to capture authors' research and institutional background as well as authors' other major characteristics.

[5]Both rankings are fairly comprehensive in their coverage of existing economics journals. When a journal is not covered by a ranking, we assign it the lowest index value for that ranking.

in our database up to 2016. Nevertheless, publications with very incomplete data on citing articles had to be removed from our citation count analysis in order to avoid bias. We restrict our citation count analysis to articles published no later than 2011 so that the latest articles also have a five-year time window to accumulate citations.

**Field Distance** We calculate the field distance of two co-authors using their field profiles prior to their collaboration. Each author has a field profile consisting of JEL category codes[6] based on her prior publications. We use the classification of Card and DellaVigna (2013) where JEL codes are sorted into twelve major fields. Each author's field profile is a vector with twelve rows, each corresponding to a field. These twelve fields are microeconomics, macroeconomics, labor, econometrics, industrial organization, international, finance, public, health & urban, development, history, and experimental economics.[7] Suppose an author has two publications where, based on its JEL codes, one of these publications is in labor and international economics, and the other publication is in labor economics and economic history. Then rows corresponding to labor economics, international economics, and economic history in the fields vector of this author will have entries 1, 0.5, 0.5, respectively, and all other entries remain zero. We denote vectors representing field profiles of authors $a$ and $b$ by $\mathbf{A}$ and $\mathbf{B}$, respectively, and calculate the distance between field profiles of $a$ and $b$ as follows:

$$Field\ Distance(a,b) = 1 - \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} \in [0,1]$$

Rousseau et al. (2017) investigate various measures to capture the similarity between researchers' publication portfolios and their results are in favor of methods that make use of similarity-adapted publication vectors. Our field distance is a linear transformation of the

[6] JEL coding system relies on self-reporting of authors and editors. Although this may seem like a potential problem about the JEL coding system, Kosnik (2018) shows that JEL codes consistently represent papers that focus on topics one would expect to be assigned to these codes.

[7] There are many possible ways to map JEL codes to fields and Card and DellaVigna (2013) provide just one of them which is predated by the JEL-to-field mapping of Ellison (2002). Text search, machine learning, and topic modeling tools (LDA) are more flexible and alternative ways that allow endogenous formation of topics (Angrist et al., 2017; Fontana et al., 2019; Önder et al., 2020). Nevertheless, publication shares of major fields based on machine learning shown in Angrist et al. (2017) are fairly similar to those based on JEL codes shown in Önder and Yilmazkuday (2020). Furthermore, JEL codes have been employed in the analysis of Fafchamps et al. (2010), Ductor (2015), and Corsi et al. (2019) and we prefer to base authors' field activity in our analysis on JEL codes in order to connect (and for our results to be comparable) to these studies.

cosine of the angle between portfolios of two authors where each author's portfolio is defined as a twelve-dimensional vector. The term $\frac{\mathbf{A}\cdot\mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|}$ equals the cosine of the angle between vectors $\mathbf{A}$ and $\mathbf{B}$ and it is also a fairly standard measure of similarity in affiliation networks (Newman, 2010) and it has previously been used in the literature for calculating the degree of research overlap between co-authors by Fafchamps et al. (2010) and Ductor (2015). We calculate co-authors' field distance based on each author's research portfolio up to six years before the date of their collaboration. When investigating co-authors' subsequent collaborations, we ignore their previous mutual collaborations in the calculation of their field distance.

**Specialization** We calculate the Herfindahl index of each author's research portfolio for up to six years before their collaboration. Specialization may decrease over the life cycle of an author if they spend several years on one topic and then another several years on some other topic, as most authors probably do. As a result of this, older authors will look less specialized than younger authors simply by construction. In order to capture a more accurate state of authors' research specialization (as opposed to their long term switch of research interests) we restrict authors' research portfolios to include the last six years before the date of collaboration.[8]

Suppose author $a$ has a profile represented by a field vector $\mathbf{A} = (a_1, a_2, ..., a_{12})$. The Herfindahl index corresponding to this profile is $\sum_{i=1}^{12}(\frac{a_i}{\sum_{j=1}^{12} a_j})^2 \in [0,1]$. This is a fairly standard measure of specialization in trade and industrial organization literature (Rosenthal and Strange, 2003), also employed in the analysis of economists' publications by Corsi et al. (2019) and Bramoulle and Ductor (2018) among others. A larger value indicates that the author's publications are accumulated in the same field whereas a lower value indicates a rather equal spread of an author's research across different fields. We make use of two measures capturing specialization patterns of co-authors. Namely, we calculate the maximum degree of specialization and the difference between co-authors' degrees of specialization. Together, these two variables capture non-linearities in co-author pair's specialization.

[8]Although six years is obviously an arbitrary choice, it roughly corresponds to one generation of PhD cohorts, or the average time between finishing PhD and getting tenure. We obtain similar results when we use ten years instead.

**Geographical Distance and Location** We find co-authors' geographical distance using their affiliations. We pass the affiliation of each co-author provided by the EconLit to the application programming interfaces of one of the major online mapping services (Google Maps, Bing Maps, and OpenStreetMaps)[9]. Using the coordinates of co-authors, we calculate the great circle distance between them.

**Social Distance** We create several indicator variables based on co-author networks, authors' graduate background, and affiliations to capture the social distance between co-authors. *Common co-author* becomes one if two authors have a common co-author, and *six degrees* becomes one if they did not share a common co-author but have been at most six degrees apart from each other in the publication network of economists within the last six years before their collaboration. Proximity in co-author networks contains valuable information about co-authors' potential to start and sustain a collaboration as previously shown by Fafchamps et al. (2010).[10] We create additional social distance measures by exploiting data on current affiliation and graduate background of co-authors as follows: We control for PhDs' connections to their graduate department and introduce a dummy *collaboration with graduate department* that equals one if an author collaborates with someone who is a faculty member at their graduate department. We identify co-authors that work in the same institute (*same affiliation*), are graduates of the same institute within six years of one another (*same graduate department*). When investigating co-authors' subsequent collaborations, we ignore their previous collaborations in the calculation of their social distance.

**Authors' Research Quality** Each author's quality-weighted total number of publications is found by weighing each publication by the quality weight of the journal where it got published and then add them up. Each author's research quality at any given time is calculated by dividing her quality-weighted total number of publications by the raw number

[9]Google Maps tends to yield more accurate results when entering short affiliations without street address. Hence, Google Maps was used for a small number of the most frequent affiliations that cover about 80% of the dataset. Due to usage restrictions, we resorted to Bing and OpenStreetMaps for the geocoding of the remaining affiliations.

[10]Unlike Fafchamps et al. (2010), we don't use a continuous measure for two authors' geodesic distance. We are interested in co-author pairs that actually did collaborate and most co-author pairs have not been on the same component prior to their collaboration. As shown in Table 1, about 19% of co-authors were within six degrees of one another and this captures the most of the same-component pairs.

Table 1: Descriptive Statistics— First Time Collaboration in Two Author Papers

| | Count | Mean | St.Dev. | Min | Max |
|---|---|---|---|---|---|
| Collab.Quality (CL) | 3682 | 0.288 | 0.190 | .1133287 | .6931472 |
| Collab.Quality (KMS) | 3682 | 0.123 | 0.191 | .0119286 | .6931472 |
| Collab.Quality (Cites) | 1793 | 1.427 | 0.918 | 0 | 4.442 |
| Geo Distance | 3682 | 4.535 | 3.419 | 0 | 9.360466 |
| Field Distance | 3682 | 0.337 | 0.202 | 0 | .6931472 |
| Common Co-author | 3682 | 0.077 | 0.267 | 0 | 1 |
| Six Degrees | 3682 | 0.193 | 0.395 | 0 | 1 |
| Specialization Max | 3682 | 0.391 | 0.145 | .1527211 | .6931472 |
| Specialization Diff | 3682 | 0.162 | 0.146 | 0 | .6097655 |
| Quality Max | 3682 | 0.123 | 0.104 | .004324 | .6850646 |
| Quality Diff | 3682 | 0.071 | 0.076 | 0 | .616626 |
| Both Top 30 | 3682 | 0.549 | 0.498 | 0 | 1 |
| Both Nontop 30 | 3682 | 0.170 | 0.376 | 0 | 1 |
| Both Male | 3682 | 0.542 | 0.498 | 0 | 1 |
| Both Female | 3682 | 0.036 | 0.185 | 0 | 1 |
| Same Affiliation | 3682 | 0.292 | 0.455 | 0 | 1 |
| Same Graduate Dept | 3682 | 0.179 | 0.383 | 0 | 1 |
| Collab with GradDept | 3682 | 0.0003 | 0.016 | 0 | 1 |
| Same Age | 3682 | 0.555 | 0.497 | 0 | 1 |

Note: All variables except for dummies are in logarithms here, i.e. we report $ln(X + 1)$ for any variable $X$.

of her publications up to that time. We use two variables to capture how two co-authors' individual qualities reshape their collaborations: the maximum quality among co-authors and their quality difference.

**Other Characteristics** Dummy variable *both top30* (*both nontop30*) equals one if both co-authors are graduates of a top thirty (non-top thirty) institute —not necessarily the same institute, and zero otherwise. In addition, we introduce dummies for co-authors' gender (*both male* and *both female*)[11] and for graduating from PhD within six years of one another (*same age*). We control also for the year of publication and each author's year of graduation for their academic maturity.

Descriptive statistics of variables that we use in the first part of our analysis where we investigate first time collaborations of North American PhDs in two author papers are shown

[11] We run the gender assignment script from Conley et al. (2016) on authors' first and middle names.

in Table 1. All variables except for dummies are in logarithms, that is a variable X enters our analysis as $ln(X + 1)$. This is important to correct for long tails of productivity and geo distance variables so that they don't get to drive our results. An average paper resulting from two authors' first time collaboration yields a CL-index of 0.36 (this implies that it is published within the top 100 journals and below the level of top field journals) and it receives 5.7 citations within five years of publication. Almost 70% of first time collaborations are distant collaborations, that is, 30% of geographic distance entries are zeros. The average distance between first time collaborators in two author papers is 86 miles on average. The maximum distance in our sample is about 11,600 miles, which corresponds approximately to the direct flight distance between London (England) and Wellington (New Zealand).

54% of co-author pairs are all male whereas only less than 4% are all female. About 56% of co-author pairs are in the same cohort hence the same age. About 8% of co-author pairs have had a common co-author prior to initiating collaboration. About 19% of all co-authors that are collaborating for the first time with one another in a two author paper do not have a common co-author but they are within six degrees of separation from one another in the authors' publication network. About 55% of co-author pairs consist of two top thirty graduates whereas about 17% consist of two non-top thirty graduates.

# 3 Stylized Facts about Distance, Specialization, and Quality

In this section, we provide a descriptive discussion of first time collaborating North American PhDs' geographical and field distance, their specialization levels, and research quality. Our observations are grouped into three remarkable stylized facts.

**Stylized Fact 1** *Co-authors have become geographically more distant but much closer in terms of their research fields over the last couple of decades.*

The average geographical distance between first time collaborators has been increasing over the last decades. This finding is in line with the increasing trend of distant collaborations among economists from 1950s to 1990s shown by Laband and Tollison (2000). As shown in panel (a) of Figure 1, more than 35% of all first time collaborations in 1990s are initiated between authors in the same location, whereas this ratio drops below 30% after 2005. Distant collaborations constitute an increasing share of all first time collaborations. Decreasing costs in communication technologies may have a significant effect on these patterns as argued by Hamermesh and Oster (2002), Kim et al. (2009), and Rosenblat and Mobius (2004). Moreover, the average distance of distant collaborations grows. Panel (b) in Figure 1 reveals a noisy, nevertheless, significant upward trend in the average distance of first time collaborators conditional on collaborators being located at different locations.

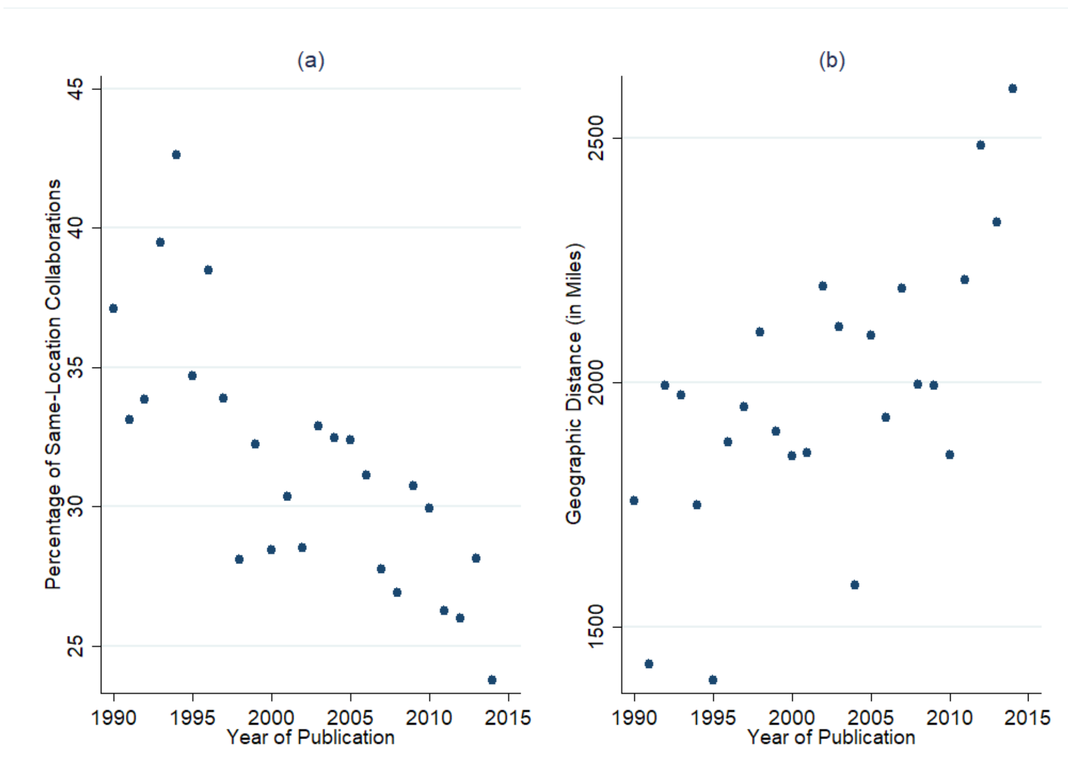


Figure 1: (a)Share of Same-Location Collaborations in all First Time Collaborations; (b)Average Distance of Distant Collaborations

The average field distance, on the other hand, has been diminishing over the same period. In Figure 2, we plot the annual average field distance against geographical distance. Average field distance and average geographical distance of first time collaborating PhDs in the early 1990s are located on the upper left corner of the diagram and they move southeast on the diagram over the years. Hence co-authors are getting apart in terms of geographical distance

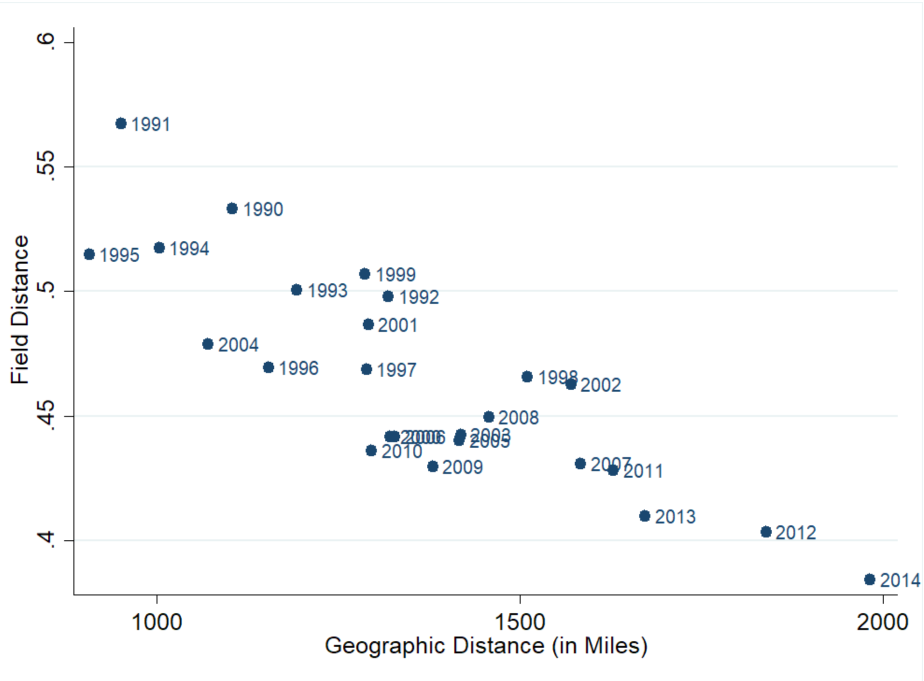


Figure 2: Geographic Distance vs. Field Distance at First Time Collaboration

while, at the same time, they are getting closer in terms of their specific research field over decades.

**Stylized Fact 2** *Co-authors whose collaboration reveals better quality have a significantly smaller field distance.*

Figure 3 shows the distribution of co-authors' field distance across journals of different quality. Panels (a) and (b) depict the distribution of field distance of first time collaborating PhDs in journals that have a CL-index greater than 0.5 and less than 0.5, respectively[12]. A CL-index larger than 0.5 corresponds to top twenty economics journals consisting of top general interest and top field journals. Comparing panels (a) and (b) of Figure 3, we find that a larger fraction of publications are accumulated at lower levels of field distance when highly ranked journals are considered. In panels (c) and (d) of Figure 3 we construct four groups of journals based on their CL-index. Panel (c) shows the mean and 95% confidence interval of field distances of co-authors in each group. Means of co-authors' field distance

[12]First time collaborators that have a field distance of one make up the largest fraction in panel (b) of Figure 3. This is mainly driven by collaborations of younger PhDs who do not have an overlap in terms of fields covered in their research.

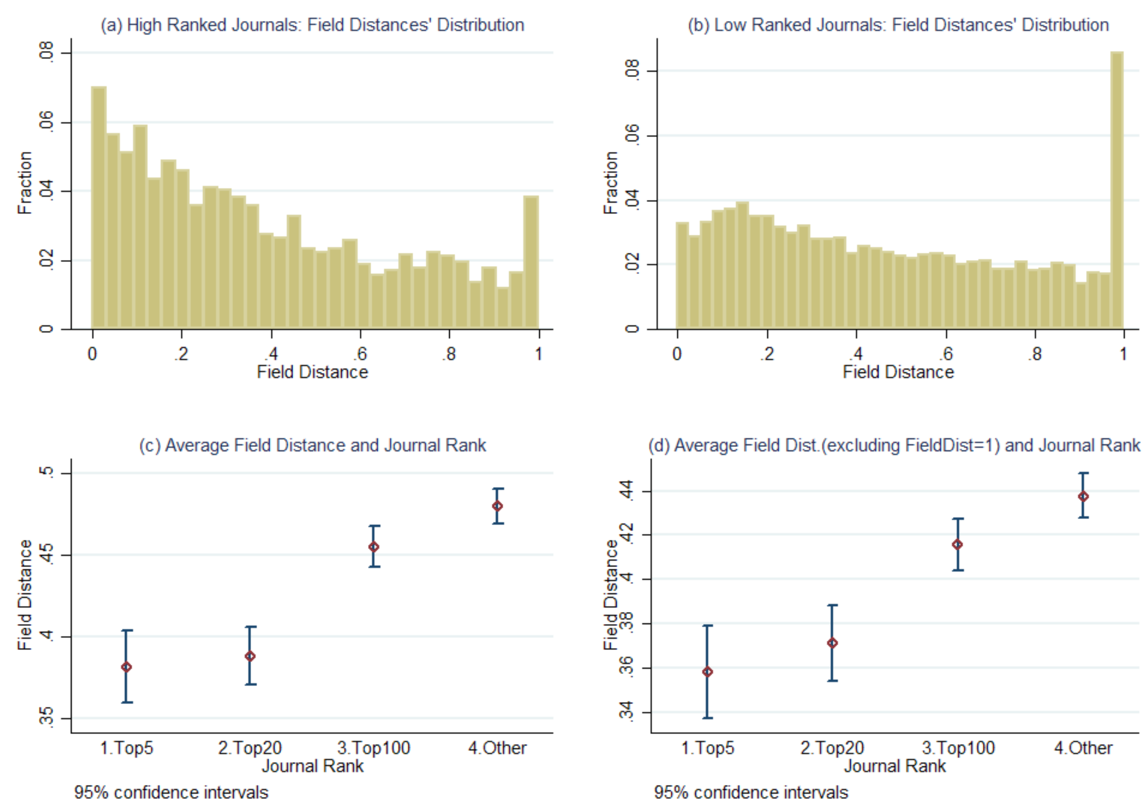


Figure 3: Field Distance and Journal Quality at First Time Collaboration

in top twenty journals are significantly lower than those in lower ranked journals. Panel (d) of Figure 3 is a reconstruction of panel (c) excluding author pairs that have maximum field distance. Results between panels (c) and (d) are fairly similar.

**Stylized Fact 3** *Co-authors' specialization levels are little or not related to the overall quality of the collaboration.*

In panel (a) of Figure 4, we sort journals into four categories based on the CL-index and we show for each category the mean and the 95% confidence interval of co-authors' specialization levels. We observe no statistically meaningful difference between specialization levels across journal categories. In panel (b) we drop collaborations of co-authors where either their field distance is maximum or one of the co-author's specialization level is maximum. In this case, specialization levels outside the top 100 journals are slightly lower on average but we do not observe significant differences in specialization levels within the top 100 journals.

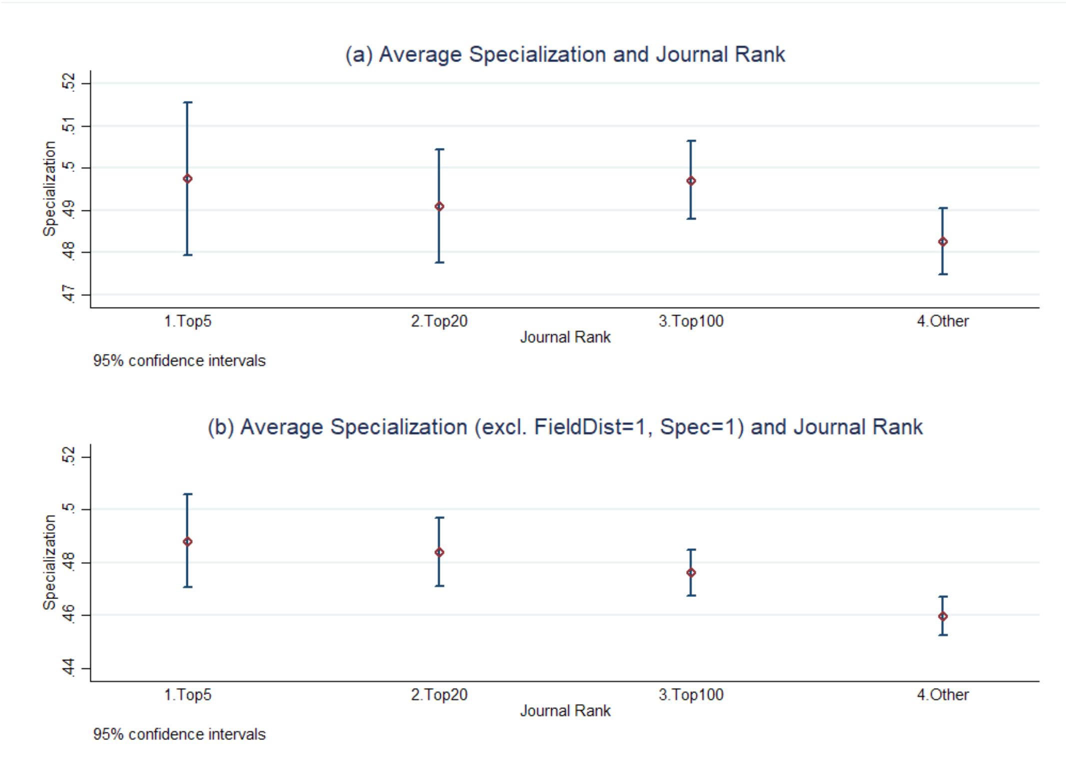


Figure 4: Average Degree of Specialization and Journal Quality

Summing up, we observe increasing geographical distance yet decreasing field distance between first time collaborating North American PhDs. A smaller field distance between co-authors is correlated with publication in a higher ranked journal. Yet, we find no clear evidence that specialization matters for publishing better.

# 4 Empirical Results

Stylized facts from Section 3 suggest that economists tend to collaborate more with those who have a similar research portfolio, hence a small field distance, to themselves. Moreover, co-authors with a small field distance are more likely to publish in better quality outlets compared to co-authors with larger field distance. Do these observations still hold when we control for various characteristics of collaboration? For this end, we investigate two author papers that embody a first time collaboration between these authors who may have collaborated with others before, and they may be at any point in their career. Two author papers make up more than 40% of all peer-reviewed journal publications of North American

PhDs in the past three decades (Önder and Yilmazkuday, 2020), so these collaborations make up a big share of published research. In addition, first time collaborations of co-author pairs offer a unique opportunity to assess the role of collaborators' field distance and specialization on the success of the collaboration. Any subsequent collaboration might be subject to different dynamics than those that led to the initiation of the collaboration as pointed out by Katz and Martin (1997), Heinze and Kuhlmann (2008), or Hollis (2001). While being aware of such differences between first time and subsequent collaborations, we analyze also subsequent collaborations of co-authors in subsection 4.2.

## 4.1 First Time Collaboration in Two Author Papers

It is plausible to assume that most authors aim to publish as good as they can and co-authorships are formed to support this aim. We denote the quality of the outcome of collaboration between co-authors $i$ and $j$ by $Quality.of.Collab_{ij}$ and we aim to estimate the following:

$$\begin{aligned} Quality.of.Collab_{ij} = \beta_0 + \beta_1(Field.Distance)_{ij} + \beta_2(Special.Max)_{ij} \\ + \beta_3(Special.Diff)_{ij} \\ + \beta_4(Quality.Max)_{ij} + \beta_5(Quality.Diff)_{ij} + \boldsymbol{\beta}\mathbf{X}_{ij} + \boldsymbol{\delta}\mathbf{FE} + \epsilon_{ij} \end{aligned} \quad (1)$$

where $\mathbf{X}_{ij}$ is a vector that captures pairwise characteristics of the co-author pair $ij$, $\mathbf{FE}$ denote fixed effects for publication year and each author's graduation year, and $\epsilon_{ij}$ is the error term.

Potential new collaborators can be met in the same department, in seminars, conferences, during academic visits, or simply by emailing directly to initiate contact. Social distance may play an important role in this process, for instance Fafchamps et al. (2010) show that having a common co-author increases the probability of collaboration between two authors who did not collaborate with one another before by 27%. In their investigation of interactions between inventors, Agrawal et al. (2008) show that social and geographical proximity are substitutes in their influence of knowledge diffusion between inventors and marginal benefit of geographical proximity is greater between inventors that are not socially close. Using economics publications, Fontana et al. (2019) show that knowledge spillovers between academic economists are geographically located. Hence, in addition to social distance, we suspect that geographical distance may also be an important factor that affects how economists interact whether the interaction is in terms of research discussion, dissemination of findings or co-authorship. As a result, geographical as well as social distance must be accounted for in an empirical model of co-authorship.

How do socially as well as geographically distant co-authors get in touch with one another? Campos et al. (2018) show that conference attendance has a causal effect on the creation of collaborations, especially when two potential collaborators are specialized in similar research fields. Fafchamps et al. (2010) and Ductor (2015) also show research similarity makes collaboration more likely. Similarity of co-authors' research fields (field distance in our paper's jargon) is a crucial aspect of co-authorship and it is connected to each of the co-authors' individual research portfolios. Let's assume that there are only two fields within economics, namely micro and macro. There are economists who work in only one field (they are specialists) and some economists do research in both fields (they are generalists). Specialists will be going to seminars and conferences in their own field, reading papers in their own field and thus becoming aware of potential co-authors in their own field. It is not that they actively avoid meeting authors from the other field, they simply don't get to meet them. Generalists, however, are more likely to meet potential co-authors in either of the two fields. In the real world, specialists in any given field are more likely to meet other specialists in the same field due to conferences and seminars they choose to attend or papers they choose to read. As a result, we would expect higher meeting probabilities[13] between specialists so that they get matched more to other specialists in their field rather than generalists or even less so specialists in other fields. Bottomline is that whom you are meeting as a potential new co-author is governed by your existing research interests, expertise, and either social or geographical (or both) proximity.

[13] Formal models of search and match in networks derive equilibrium meeting probabilities as the Nash equilibrium arising from each type's strategy to participate and stay in the matching process. The link formation model of Currarini et al. (2009) analyzing homophily in networks is an outstanding example of such a model. What we informally describe here captures the main intuition of their model's equilibrium in a nutshell.

Crafting an academic paper can be considered as a two stage production process: In the first stage, authors get matched based on their individual characteristics so that their field distance is the result of this match, as explained above. In the second stage, the quality of their collaboration is revealed and this quality is related to their above mentioned characteristics as well as their field distance. Co-authors may or may not care for their field distance per se, but in any way, field distance is a usual suspect for being related to the quality of collaboration[14] and according to what we describe above, it is not exogenous. Hence co-authors' individual research characteristics will be related to the quality of their collaboration via two channels: They may have a direct effect on the quality, and at the same time, they may have an indirect effect by influencing co-authors' field distance, which in return, may be related to the final quality. This, however, poses a simultaneity problem. In order to solve this problem, we employ a two stage least squares (TSLS) to estimate the equation 1 where we instrument co-authors' field distance by co-authors' geographical and social distance. We estimate the following equation in the first stage:

$$\begin{aligned} Field.Distance_{ij} = \alpha_0 + \alpha_1(Geo.Distance)_{ij} + \alpha_2(Social.Distance)_{ij} \\ + \alpha_3(Special.Max)_{ij} + \alpha_4(Special.Diff)_{ij} \\ + \alpha_5(Quality.Max)_{ij} + \alpha_6(Quality.Diff)_{ij} + \boldsymbol{\alpha}\mathbf{X}_{ij} + \boldsymbol{\gamma}\mathbf{FE} + \epsilon_{ij} \end{aligned} \quad (2)$$

Social distance between first time collaborators is captured by indicator variables *common co-author*, *six degrees*, *collaboration with graduate department*, *same affiliation*, and *same graduate department*. These variables are explained in section 2 in detail. Co-authors' characteristics such as graduate institute rankings and gender are captured in $\mathbf{X}$, and fixed effects are used for the year of publication and graduation years of collaborators.

### 4.1.1 Geographical, Social, and Field Distance between Co-authors

Columns (1) to (3) in Table 2 show estimation results of equation 2. Geographical distance between two co-authors collaborating for the first time is negatively and significantly related

[14] Nooteboom et al. (2007) show that field distance (or *cognitive distance* as they call it) is a significant factor in quality of an innovation.

Table 2: Field Distance at First Time Collaboration in Two Author Papers

| | Dependent Variable: Field Distance | | | | | | |
|---|---|---|---|---|---|---|---|
| | All Collaborations | | | Distant Collabs | | Same Location | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Geo Distance | -0.00531*** | -0.00434*** | -0.00115 | -0.00103 | -0.00173 | | |
| | (0.000980) | (0.000951) | (0.00168) | (0.00187) | (0.00185) | | |
| Specialization Max | | -0.211*** | -0.237*** | -0.288*** | -0.305*** | 0.0253 | -0.000189 |
| | | (0.0541) | (0.0538) | (0.0623) | (0.0626) | (0.0987) | (0.103) |
| Specialization Diff | | 0.364*** | 0.373*** | 0.410*** | 0.407*** | 0.239** | 0.218* |
| | | (0.0505) | (0.0503) | (0.0596) | (0.0596) | (0.0911) | (0.0936) |
| Quality Max | | -0.588*** | -0.515*** | -0.591*** | -0.518*** | -0.578*** | -0.523*** |
| | | (0.0560) | (0.0564) | (0.0667) | (0.0669) | (0.104) | (0.114) |
| Quality Diff | | 0.599*** | 0.507*** | 0.624*** | 0.524*** | 0.609*** | 0.505*** |
| | | (0.0724) | (0.0734) | (0.0852) | (0.0864) | (0.138) | (0.151) |
| Common Co-author | | | -0.0773*** | | -0.0732*** | | -0.0753*** |
| | | | (0.0110) | | (0.0131) | | (0.0213) |
| Six Degrees | | | -0.0626*** | | -0.0693*** | | -0.0402* |
| | | | (0.00821) | | (0.00958) | | (0.0167) |
| Same Affiliation | | | 0.0312* | | | | |
| | | | (0.0129) | | | | |
| Same Grad | | | 0.00795 | | 0.0140 | | -0.0271 |
| | | | (0.00879) | | (0.0101) | | (0.0197) |
| Collab w/GradDept | | | -0.133** | | -0.152** | | |
| | | | (0.0408) | | (0.0501) | | |
| Same Age | | | 0.0148* | | 0.0122 | | 0.0199 |
| | | | (0.00705) | | (0.00865) | | (0.0130) |
| Individual/Pair Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Pub.Year FE | Yes | Yes | Yes | Yes | Yes | No | Yes |
| Grad.Year FE | Yes | Yes | Yes | Yes | Yes | No | Yes |
| $N$ | 3682 | 3682 | 3682 | 2569 | 2569 | 1113 | 1113 |
| Adjusted $R^2$ | 0.033 | 0.105 | 0.126 | 0.107 | 0.129 | 0.079 | 0.091 |

Standard errors in parentheses

$^+$ $p < 0.10$, $^*$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

to their field distance in specifications (1) and (2). However, when social distance controls are added, geographical distance turns insignificant in column (3), probably because socially close authors are likely to get in touch despite of long geographical distance. Hamermesh and Oster (2002) explain that one reason why we observe large distances between co-authors is that they may actually know each other either from graduate school or from where they worked previously. Co-authors with six or less degrees of separation (based on the co-authorship network) are expected to have significantly less field distance; co-authors that are located in the same department have significantly larger field distance; authors collaborating with a faculty member from their graduate institute have significantly less field distance. The dummy variable for being graduates of the same department within five years of one another (*Same Grad*) turns out statistically insignificant. It is possible that co-authors' social distance already captures traits from sharing this common environment during very early stages of the academic career.

We restrict our subsample to distant collaborations in columns (4) and (5) in Table 2 and find no statistically significant relation between geographical distance and field distance, even when social distance controls are not included as in column (4). Distant collaborations account for about 70% of all collaborations and they have a significantly smaller field distance than those of same location collaborations. However, there is not much of a difference between field distance of a geographically very close (yet not exactly the same location) and a far far away collaboration.

The specialization levels of co-authors included in the specifications in columns (2) to (5) are related to their field distance in two ways: First, the higher is the degree of specialization of co-authors, the lower is their field distance. This indicates that a highly specialized author is more likely to collaborate with another highly specialized author if their fields are very close and two highly specialized authors in different fields are not likely to collaborate at all. Second, the difference in specialization levels of co-authors is positively related to their field distance. This is also the case when the subsample of distant collaborations are considered. Highly specialized authors get matched to authors with a small field distance, and collaboration of two specialists involves a very small field distance whereas generalists have a larger field distance on average. We do not find such patterns when the subsample is restricted to intra-departmental collaborations as shown in columns (6) and (7) hence the crucial difference between distant and same location (intra-departmental) collaborations.

In all specifications in Table 2 we find negative and significant correlation between co-authors' field distance and the maximum quality attained by either of the two co-authors whereas the quality difference between co-authors is positively and significantly related to their field distance. Authors with a high quality research track tend to collaborate with those who have a small field distance.

#### 4.1.2 Field Distance and Quality

Results for the second stage estimations are shown in Table 3. The quality of collaboration is measured in three different ways: In the first four columns of Table 3, we measure it by the quality of the journal where it got published. Columns (1) to (2) are based on the CL-index where journals are grouped and each journal within the same group gets the same quality weight. Columns (3) and (4) are based on the KMS-index where each journal gets an individual quality weight and these weights are subject to a severe discount as one moves down the ranking[15]. In the last two columns of Table 3 we use the number of citations[16] received by an article within five years after its publication.[17]

For the TSLS, we provide two diagnostic tests to verify the validity of using fitted values of field distance in the second stage. We report p-values associated with under-identification tests and we report Cragg-Donald Wald F statistic to diagnose weak identification. Test statistics are rejected at very low p-values, and comparing Cragg-Donald's F statistics to critical values provided by Stock and Yogo (2005), we observe that F statistics are larger than

[15]The *American Economic Review (AER)* has the highest quality weight in the KMS index. To give an idea how severe the discounting is, the *Review of Economics Studies* has about one third of the *AER*'s weight and most top field journals (such as the *Journal of Public Economics* or the *Journal of Labor Economics*) have about $15 - 25\%$ of the *AER*'s weight.

[16]When the quality of collaboration is measured by the number of citations, we include an additional control for the journal quality, because more prestigious journals provide greater visibility and pave the road to a larger citation count. Nevertheless, attributes that shoot a paper into a prestigious journal are the same attributes that lead to a large amount of citations as well, hence journal quality cannot be controlled for at its face value. We solve this problem by using residual journal quality which consists of the variation in journal quality that is not explained by controls we use in column (1) of Table 3. Residual journal quality is positively and significantly correlated with the amount of citations, meaning that a paper's citations may get boosted just for being published in a good ranked journal, and not due to its inherent qualities.

[17]When we restrict our whole analysis to those publications for which we have citation information, we obtain very similar results to those shown in columns (1) to (4) in Table 3.

critical values. Hence our instruments are highly relevant. While our analysis is obviously not based on some natural experiment, we make use of an extensive information set on authors' social connections that go beyond the conventional co-author networks that have been used in previous studies such as Fafchamps et al. (2010) since we link PhDs to publications. There is no obvious reason why social distance should directly affect co-author pairs' research quality. Our analysis has an indirect channel so that social distance may affect quality indirectly, namely via field distance. However, we cannot guarantee that there is absolutely no other indirect channel that links social distance and collaboration quality or that social distance does not correlate with an unobservable variable such as author's talent due to assortative matching in social networks.[18] We are well aware of these aspects and we refrain from reading too much into the causality of our findings.

Field distance is negatively and significantly related to journal quality as well as to number of citations in every specification in Table 3. Thus field distance and the quality of collaborative research are significantly related, whether we take the journal where it lands or the number of citations it receives as the revelation of its quality. Co-authors that have a smaller field distance are more likely to publish in higher ranked journals when journal quality is based on the CL-index. When journal quality is measured by the KMS-index, we obtain no statistical significance for fitted values of collaborators' field distance. This difference suggests that the variation of field distance between individual journals is not large enough whereas there is significant variation in the field distance when journals are bundled to form quality equivalence classes as the CL-index does. Point estimates of the instrumented field distance are larger than their OLS counterparts. This can be due to an omitted variable, which we suspect less, or rather due to large local average treatment effects that are captured by 2SLS. The first stage estimation is based on the assumption that authors search for co-authors via their networks or they contact potential co-authors based on the literature they read or conferences they attend. Although these may capture most plausible and (more importantly) measurable ways to get in touch with new co-authors, one cannot claim that

[18]More able authors collaborate with similarly able authors so that their co-authors of co-authors are also more able and so on. This means that social distance variables may capture the degree of assortative matching in authors' social networks. Similarly, one might argue that classmates from the same department may have similar talent etc..

Table 3: Publication Quality at First Time Collaboration in Two Author Papers

| | Dep.Var: CL Index | | KMS Index | | Citations | |
|---|---|---|---|---|---|---|
| | (1) | (2)IV | (3) | (4)IV | (5) | (6)IV |
| Field Distance | -0.0364** | -0.281*** | -0.0239+ | -0.0999 | -0.513*** | -1.858** |
| | (0.0138) | (0.0805) | (0.0139) | (0.0800) | (0.105) | (0.592) |
| Specialization Max | -0.105** | -0.157*** | -0.0579+ | -0.0739* | -0.636* | -0.936** |
| | (0.0341) | (0.0386) | (0.0348) | (0.0365) | (0.273) | (0.327) |
| Specialization Diff | 0.0753* | 0.165*** | 0.0581+ | 0.0858* | 0.620* | 1.024** |
| | (0.0327) | (0.0432) | (0.0333) | (0.0413) | (0.263) | (0.337) |
| Quality Max | 1.275*** | 1.130*** | 1.381*** | 1.335*** | 2.192*** | 1.365** |
| | (0.0562) | (0.0747) | (0.0633) | (0.0791) | (0.373) | (0.526) |
| Quality Diff | -0.771*** | -0.625*** | -0.913*** | -0.868*** | -0.932+ | 0.0722 |
| | (0.0755) | (0.0910) | (0.0824) | (0.0951) | (0.476) | (0.655) |
| Journal Quality[a] | | | | | 1.098*** | 1.096*** |
| | | | | | (0.127) | (0.129) |
| Individual/Pair Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Publication Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Graduation Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| $N$ | 3682 | 3682 | 3682 | 3682 | 1793 | 1793 |
| $Underidentification$ | | 4.02e-23 | | 4.02e-23 | | 9.68e-11 |
| $Cragg - Donald\ F$ | | 22.40 | | 22.40 | | 10.71 |

Standard errors in parentheses

$^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

[a] Journal Quality is captured by fitted residuals from (1)

these scenarios capture all possible ways to get in touch with new co-authors. It is very well possible that the field distance between co-authors who search for and match with new co-author in the way that is captured in our treatment make a greater difference for this subset of authors.

Specialization levels are negatively and significantly related to the quality of collaboration output. Together with results obtained in Table 2 we arrive at the following curious finding: Although high specialization levels among co-authors are significantly related to a small field distance between them, and small field distance is significantly related to high quality output, we find specialization levels to be negatively related to the quality of collaboration when they enter the second stage as a control. Hence co-authors' specialization levels work through two channels: First, a high specialization level has an indirect positive effect on the quality of collaboration output. The indirect channel works through co-authors' matching such that highly specialized authors team up with those that are close in field, and such closeness is related to a high quality of collaboration output. Second, high specialization level has a direct negative effect on the quality of collaboration output. This is certainly not driven by age differences within co-author pairs where more experienced authors publish high with their students. When specialization is calculated for the whole life cycle, older authors appear significantly less specialized than younger authors, but we restrict our specialization measure to account for the past six years only, not the whole career up to that point.

The total effect of specialization is the sum of the direct and indirect effects. Using coefficients from both stages, one can decompose the relation between specialization and quality. A one standard deviation increase in specialization is indirectly associated with a 0.6 unit increase in citations due to the relation between specialization and field distance in the first stage. The direct effect of such an increase in specialization in the second stage is, however, 1.2 units decrease in citations. Hence the total relation between specialization and quality is negative as also suggested by point estimates in not-instrumented specifications. More specialized authors find co-authors with a very small field distance, and this is associated with a better quality of their collaborative output, but the direct effect of specialization on quality is so large that the total relation between specialization and quality turns out to be negative. The difference in collaborators' specialization levels is positively and significantly related to the quality of the collaborative output, suggesting that high quality research is more likely to emerge as a result of an interaction between specialists and generalists.

Individual research qualities of co-authors are positively and highly significantly correlated with the quality of co-authors' joint research. Individual qualities work also through direct and indirect channels similar to specialization, but unlike specialization, individual quality works in the same direction in both stages. Authors with high quality research tend to team up with co-authors with a small field distance, which in turn is associated with a higher quality of their collaborative research. Moreover, individual research qualities are significantly and positively related to the quality of the collaboration even after we control for their indirect effect via the field distance and other pair characteristics (such both authors being top department graduates).

We provide a robustness check using an alternative measure for field distance that is based on JEL codes directly (at one-letter and one-digit level) instead of grouping them into fields in Table B.1 in the Appendix B. Direct and indirect channels for specialization and individual quality work in a very similar manner to those obtained in Table 3 when TSLS estimates are considered.

## 4.2 Subsequent Collaborations

In Table 4 we investigate subsequent collaborations of co-authors with one another that started off with a two author paper (hence these are the co-author pairs from the previous subsection) and collaborated at least once more after their first collaboration. In panel A of Table 4, we track each co-author pair from the year after their first collaboration with one another until 2014.[19] Subsequent collaborations are not restricted to two author papers. We record the quality of outcome for a year when a co-author pair did collaborate in that year; we record zero for years where they did not but could have done so. Field distance for subsequent collaborations is calculated by removing co-authors' joint publications from their research portfolios, otherwise co-authors' field distance will diminish upon collaboration by construction.

[19] Active years of an author are years from the first to last publication of this author in our data. These publications can be single authored or co-authored. Most authors don't publish every single year. If an author's last publication is within five years of 2014 we assume they could have been active in 2014 as well.

Table 4: Publication Quality at Subsequent Collaborations

A. Including years of actual and potential collaborations

| | Using Pairwise Fixed Effects | | | Using Graduate and Social Controls | | |
|---|---|---|---|---|---|---|
| | (1)CL | (2)KMS | (3)Cites | (4)CL | (5)KMS | (6)Cites |
| Field Distance | -0.0185** | -0.0139** | -0.0519** | -0.0210*** | -0.0131*** | -0.147*** |
| | (0.00605) | (0.00424) | (0.0158) | (0.00385) | (0.00261) | (0.0102) |
| Specialization Max | 0.151*** | 0.0619*** | 0.554*** | 0.0437*** | 0.0138* | 0.157*** |
| | (0.0145) | (0.0102) | (0.0386) | (0.00945) | (0.00685) | (0.0261) |
| Specialization Diff | -0.0189 | -0.00592 | 0.0162 | 0.00515 | 0.00547 | 0.0582* |
| | (0.0122) | (0.00858) | (0.0324) | (0.00951) | (0.00688) | (0.0258) |
| Quality Max | -1.087*** | -1.131*** | -3.305*** | 0.150*** | 0.196*** | 0.500*** |
| | (0.0474) | (0.0332) | (0.130) | (0.0128) | (0.0105) | (0.0282) |
| Quality Diff | 1.181*** | 0.968*** | 4.291*** | -0.00339 | -0.0818*** | 0.0727+ |
| | (0.0393) | (0.0276) | (0.106) | (0.0187) | (0.0148) | (0.0425) |
| Pairwise FE | Yes | Yes | Yes | No | No | No |
| Individual/Pair Controls | No | No | No | Yes | Yes | Yes |
| Publication Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Graduation Year FE | No | No | No | Yes | Yes | Yes |
| *N* | 29632 | 29632 | 26996 | 30260 | 30260 | 27401 |

Standard errors in parentheses

$^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

B. Including years of actual collaborations only

| | Using Pairwise Fixed Effects | | | Using Graduate and Social Controls | | |
|---|---|---|---|---|---|---|
| | (1)CL | (2)KMS | (3)Cites | (4)CL | (5)KMS | (6)Cites |
| Field Distance | -0.0524* | -0.0494* | 0.00719 | -0.0285* | -0.0213+ | -0.578*** |
| | (0.0217) | (0.0221) | (0.177) | (0.0126) | (0.0129) | (0.0915) |
| Specialization Max | 0.00245 | -0.0111 | -0.549 | -0.0619* | -0.0450 | -0.731** |
| | (0.0517) | (0.0527) | (0.433) | (0.0270) | (0.0288) | (0.225) |
| Specialization Diff | -0.00368 | 0.000944 | 0.682+ | 0.0165 | 0.0156 | 0.479* |
| | (0.0443) | (0.0451) | (0.380) | (0.0267) | (0.0287) | (0.227) |
| Quality Max | -1.161*** | -1.598*** | -0.556 | 0.913*** | 1.040*** | 1.868*** |
| | (0.124) | (0.126) | (0.908) | (0.0371) | (0.0415) | (0.222) |
| Quality Diff | 0.799*** | 1.028*** | 1.307 | -0.409*** | -0.596*** | -1.033** |
| | (0.106) | (0.108) | (0.801) | (0.0548) | (0.0619) | (0.318) |
| Pairwise FE | Yes | Yes | Yes | No | No | No |
| Individual/Pair Controls | No | No | No | Yes | Yes | Yes |
| Publication Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Graduation Year FE | No | No | No | Yes | Yes | Yes |
| *N* | 3777 | 3777 | 1692 | 3777 | 3777 | 2133 |

Standard errors in parentheses

$^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

We use pairwise fixed effects to account for time-invariant characteristics of co-author pairs in specifications from (1) to (3) in both panels of Table 4. As discussed in Fafchamps et al. (2010) in great detail, pairwise fixed effects capture time-invariant individual and pair-specific factors that may lead to forming and sustaining a collaboration such as having gone to the same graduate institute or having similar abilities, or having compatible views about how a research team should operate. Introduction of pairwise fixed effects enables a more robust analysis of time-variant characteristics of a co-author pair that may be linked to the quality of their collaboration after removing effects of any other characteristics that are time-invariant and specific to that particular pair.

In the previous subsection 4.1, we use information about authors' PhD background as instruments in the two stage analysis. Since these instruments are time-invariant, we can not use them in combination with pairwise fixed effects. Geographical distance and affiliation of co-authors are time-variant, of course, but they don't yield sufficient variation when pairwise fixed effects are included so we are left with co-author network distance variables as valid instruments. The idea behind the two stage estimation in subsection 4.1 is that the field distance is not exogenous to other qualities of the paper because these qualities may affect co-authors' matching in the first place. When investigating subsequent collaborations, however, there is no search-and-match argument to justify instrumenting of field distance. As a result, we do not employ IV in this subsection.

In panel A of Table 4, we include no-collaboration outcomes, i.e. *zeros*. The first three columns are with pairwise fixed effects and the last three columns are without pairwise fixed effects but include individual and pairwise controls for authors' PhD and social background. Whether we take the journal quality or the count of citations as the measure of the quality of subsequent collaboration, this significantly and negatively correlates with co-authors' field distance. Although this implies that co-author pairs with a close field distance are more likely to re-collaborate over their life cycle, it is not clear how strongly field distance correlates with subsequent collaborations' quality. In panel B, we drop years of non-collaboration. Since we are using pairwise fixed effects and there are no *zeros*, we restrict the subsample to co-author pairs who have collaborated at least twice after their first collaboration. Field distance turns out statistically significant and negative also in this setting (except for citations using pairwise fixed effects). Hence one cannot argue that field distance's negative coefficient in panel A is solely driven by co-authors with a close field distance being more likely to re-collaborate. We already know from Fafchamps et al. (2010) that co-authors with a small field distance are more likely to engage in subsequent collaboration. Our results, however, show that they are not only more likely to do so but their output is more likely to be of better quality on average compared to co-authors with a larger field distance who also choose to re-collaborate. Whether a close field distance has a causal impact on the quality of subsequent collaborations, is yet to be discovered, but strong and consistent correlation is already there.

The maximum specialization level among co-authors obtains a positive and significant coefficient for subsequent collaborations with *zeros* in every specification panel A of Table 4. However, we do not obtain statistical significance for specialization when *zeros* are dropped and pairwise fixed effects are used as in columns (1) to (3) in panel B. This means that more specialized authors are more likely to engage in subsequent collaboration with their existing co-authors, but we do not find any statistically significant relation between specialization and the quality of subsequent collaborations. Similar to our results for co-authors' first time collaboration in subsection 4.1 above, we obtain negative correlation between specialization and the quality of collaboration when we use individual social and education background controls instead of pairwise fixed effects.

The maximum individual quality level of co-authors has a significant and negative coefficient through most specifications with pairwise fixed effects in both panels of Table 4. Fafchamps et al. (2010) do not control for specialization but they do for individual quality and also obtain a negative coefficient for it when they include pairwise fixed effects. Based on panel A, one may claim that co-author pairs are more likely to re-collaborate when their publications from other collaborations or their sole author publications do not turn up as successful as they used to. The opportunity cost of re-collaboration is engagement in a new and possibly more promising collaboration or writing a sole author paper. Those authors who lack such fruitful opportunities or ideas go back to their old co-authors so that co-author pairs that re-collaborate are those who experience a downturn in their publication success. Moreover, their subsequent collaborations do not get published as high, and this drop is larger for co-authors with larger individual qualities. Using publication records of Nobel laureates, Chan et al. (2016) show that co-authors exhaust their most interesting and strong ideas in the beginning of their collaboration and subsequent collaborations don't yield as much success. A similar process might be taking place here as well.

In the Appendix B we present two further sets of results. Table B.2 repeats the analysis carried out in Table 4 using all subsequent collaborations of all co-authors whether their first-time collaboration has been in a two or more author paper. We obtain qualitatively very similar results to those shown in Table 4. Table B.3 uses the whole sample of collaborations (including first and subsequent) and presents results where not only the field distance but also its square enters the analysis in order to account for a possible non-linearity of the relation between field distance and quality.[20] Coefficients shown in Table B.3 are qualitatively similar to those shown in panel A of Table 4.

# 5 Discussion and Conclusion

We analyze North American economics PhDs' collaborations in peer-reviewed economics journals from 1990 to 2014 and observe three stylized facts: (i) Co-authors have become geographically more distant but much closer in terms of their research fields over the last couple of decades; (ii) co-authors whose collaboration reveals better quality have a significantly smaller field distance; (iii) co-authors' specialization levels are little or not related to the overall quality of the collaboration.

We find that the field distance between co-authors is negatively and significantly related to the quality of their collaborative output. This relation is robust whether we take the journal where the co-authored paper lands or the number of citations it receives, also whether we focus on co-authors' first time collaboration and their subsequent collaborations, whether we

[20] Rafols et al. (2010) show that successful collaborations are more likely to occur in a middle range of cognitive distance, probably because this is where co-authors can most successfully share their capabilities and expertise while still being able to understand one another. Nooteboom et al. (2007) analyze the idea of cognitive distance in innovation context and find that there is an inverted U-shaped effect of cognitive distance on innovation performance. Fafchamps et al. (2010) and Ductor (2015) also use a quadratic term to capture research similarity between co-authors. We chose not to do so in our main analysis because field distance is instrumented in subsection 4.1 and using polynomials of an instrumented variable would not only largely dilute channels for indirect effects but also greatly obscure the precision of direct effects.

instrument the field distance in first time collaboration by social distance, or we use pairwise fixed effects in subsequent collaborations to account for time-invariant characteristics.

Investigating how characteristics of co-authors' research portfolios are related to their field distance, we find that highly specialized authors get matched to authors with a small field distance, and collaboration of two specialists involves a very small field distance whereas generalists have a larger field distance on average. When we focus on intra-departmental collaborations, however, we do not find such patterns. This may hint that intra-departmental collaborations may be driven by different mechanisms, possibly due to the lack of the bias that exists in case of the search for a distant co-author created by underlying meeting probabilities. Our analysis starts from 1990, and distant collaboration has already become fairly common among economists at this point in time as shown by Laband and Tollison (2000) and Hamermesh and Oster (2002) among others. Kim et al. (2009) study the research quality of economics and finance faculty from 1970s until 2002 and show that the share of co-authored papers increases where co-authors are located at different universities and at least one co-author is at a top university. They further show that economists in non-elite universities are collaborating increasingly more in recent decades with economists in elite universities to create high quality publications. This suggests that positive spillovers of having colleagues with high quality research portfolio has moved beyond the physical limits[21] of a university. They explain this by advancements of internet and communication technologies as these make collaboration at a distance easier and disproportionately favor non-elite universities. Comparing outcomes of distant and same location collaborations, Franceschet and Costantini (2010) show that distant collaborations tend to receive more citations ceteris paribus.

We contribute to the literature on quality differences of same location and distant collaborations in two ways. First, co-authors that engage in a distant collaboration are significantly more likely to have a close field distance, and a close field distance is significantly related to having a high quality outcome for this collaboration. This provides an important link between the above mentioned literature on the geographical distance of collaborations and

[21] This does not necessarily mean that physical limits are completely irrelevant, of course. Characteristics of hiring departments are shown to have significant explanatory power on the quantity as well as quality of French economists' research (Bosquet and Combes, 2017). For an extensive survey of the literature on the effects of agglomeration on innovation in general, see Carlino and Kerr (2015).

the literature documenting the tendecy for research homophily[22] of co-authors, as shown by Rosenblat and Mobius (2004) and also indirectly by Fafchamps et al. (2010) and Ductor (2015). We show that the channel as to why distant collaborations turn out better than same location collaborations is the field distance.

Second, the significant correlation between geographical and field distance is lost when social controls are introduced. That is, once co-authors' social networks (not just co-author networks but graduate school and affiliation networks) are accounted for, physical distance becomes irrelevant. An important finding in the innovation literature is that knowledge spillovers are geographically highly concentrated (Jaffe et al., 1993), nevertheless, the geographical component is shown to be substituted by inventors' social proximity (Agrawal et al., 2008; Singh, 2005). What we are measuring in this paper is not explicitly knowledge spillovers, but collaboration can be considered as a specific form of knowledge diffusion between co-authors. While we find geographical distance to be significantly related to co-authors' field distance (and thus indirectly to the quality of their collaboration), this significance is lost once we account for co-authors' social ties. Hence our findings are in line with what has already been shown in the literature using patents and natural science publications, and we extent these findings to cover academic economists.

Specialization is certainly an important aspect of authors' portfolios. Previous research shows a negative effect of specialization on economists' productivity and impact (Bosquet and Combes, 2017; Bramoulle and Ductor, 2018; Ductor, 2015). We show that co-authors' specialization levels work through two channels: First, a high specialization level has an indirect positive effect on the quality of collaboration output which works through co-authors' matching such that highly specialized authors team up with those that are close in field, and such closeness is related to a high quality of collaboration output. Second, a high specialization level has a direct negative effect on the quality of collaboration output. The total effect of specialization is the sum of the direct and indirect effects and this turns out negative in most specifications. High quality research is more likely to emerge as a result of an interaction between specialists and generalists, yet, they should preferably have a close

[22] In this context this means preference to team up with co-authors who have similar research interests and agenda.

field distance. Although high specialization provides deep understanding of and strength in a topic, it does not seem to be sufficient for publishing in top journals or guarantee a high number of citations. As pointed out by Melero and Palomeras (2015), highly specialized research teams may lack the overall intuition, which is provided by a generalist rather than a specialist. The significantly positive coefficient of the difference in co-authors' specialization supports this narrative. Analyzing subsequent collaborations, we find that more specialized authors are more likely to engage in subsequent collaboration with their existing co-authors, but their specialization levels have no explanatory power for the quality of their subsequent collaborative work.

We find that authors with a high quality research track tend to collaborate with those who have a small field distance. Moreover, if two such established authors collaborate then it is very likely that they have a very small field distance. Co-authors' individual research qualities and the quality of their collaborative output are positively and highly significantly related via indirect (via their field distance) as well as direct channels. There may be an underlying mechanism that affects meeting probabilities of established authors, similar to that of highly specialized authors, so that they meet each other more frequently than they would meet less established authors. One reason for this could be that they attend rather exclusive conferences and seminars. Another reason may be that established authors prefer to collaborate with other established authors, because they worry about the quality of their work. Even so, they prefer to collaborate with established authors in their own field and not in distant fields.

An interesting interpretation of our results concerning the relation of authors' individual quality and the quality of collaboration is that subsequent collaborations do not get published as high, and this drop is larger for co-authors with larger individual qualities. This may have to do with the opportunity cost of re-collaboration compared to a new collaboration or other modes of research so that re-collaborating co-author pairs are those who lack more fruitful outside options. Using publication records of Nobel laureates, Chan et al. (2016) show that co-authors exhaust their most interesting and strong ideas in the beginning of a collaboration so that their subsequent collaborations don't achieve as much success. Similary, Bramoulle and Ductor (2018) find that authors with a diverse set of co-authors publish on average in better journals than authors who always publish with the same limited number of co-authors. A similar process might be taking place here as well.

Finally, our findings allow us to extrapolate some (however rough) conclusions about interdisciplinarity within economics research[23]. Collaborations between highly specialized authors or between well established authors are more likely if these have a very small field distance. Hence we do not observe interdisciplinary work in the form of collaboration between two specialists in distinct areas of economics and we do not find very often that two very well established economists in distinct fields join forces. Interdisciplinary research is created by a team of authors in which each author already has an interdisciplinary portfolio[24]. A possible explanation could be that interdisciplinary researchers are less siloed and thus able to speak to each other.

Interdisciplinarity is not an inherently good or bad attribute of collaboration. Using data from the UK's Engineering and Physical Sciences Research Council, Banal-Estañol et al. (2019) show that research teams with greater diversity in their research portfolio are less likely to receive funding, although they find no significant effect on research quality. Considering interdisciplinarity within economics, Corsi et al. (2019) do not find an adverse effect of interdisciplinarity on a researcher's chances of promotion. Our findings suggest that cross-field collaboration among economists takes place and is successful if each author's research expertise is already interdisciplinary. High quality interdisciplinary research that is created by the collaboration of authors in separate fields with completely separate expertise only exists in dreams of grant committees and in the strong imagination of faculty administrators, but not in the real world, at least not in our data.

[23] When we say interdisciplinarity within economics, we refer to collaborations between e.g. labor economists and economic historians or monetary and health economists, not collaborations between economists and mathematicians or political scientists.

[24] This is very much in line with Haeussler and Sauermann (2020) who use articles from Plos One (top quartile journal for interdisciplinary sciences) to show that a greater division of labor among co-authors in an interdisciplinary research can be achieved only if individual team members can draw on multiple disciplines as opposed to when they are specialized in different disciplines.

# References


Adams, J. D., G. C. Black, J. R. Clemmons, and P. E. Stephan (2005): "Scientific teams and institutional collaborations: Evidence from US universities 1981-1999," *Research Policy*, 34, 259–285.

Agrawal, A., D. Kapur, and J. McHale (2008): "How do spatial and social proximity influence knowledge flows? Evidence from patent data," *Journal of Urban Economics*, 64, 258–269.

Angrist, J., P. Azoulay, G. Ellison, R. Hill, and S. F. Lu (2017): "Economic research evolves: Fields and styles," *American Economic Review: Papers & Proceedings*, 107, 293–297.

Banal-Estañol, A., I. Macho-Stadler, and D. Pérez-Castrillo (2019): "Evaluation in research funding agencies: Are structurally diverse teams biased against?" *Research Policy*, 48, 1823–1840.

Bonaccorsi, A. and L. Secondi (2017): "The determinants of research performance in Europea universities: A large scale multilevel analysis," *Scientometrics*, 112, 1147–1178.

Bosquet, C. and P.-P. Combes (2017): "Sorting and agglomeration economies in French economics departments," *Journal of Urban Economics*, 101, 27–44.

Bramoulle, Y. and L. Ductor (2018): "Title length," *Journal of Economic Behavior and Organization*, 150, 311–324.

Campos, R., F. Leon, and B. McQuillin (2018): "Lost in the storm: the academic collaborations that went missing in hurricane Isaac," *Economic Journal*, 128, 995–1018.

Carayol, N. and M. Matt (2006): "Individual and collective determinants of academic scientists' productivity," *Information Economics and Policy*, 18, 55–72.

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

Carlino, G. and W. R. Kerr (2015): *Handbook of Regional and Urban Economics*, North-Holland, vol. 5A, chap. Agglomeration and Innovation, 349–404.

Chan, H. F., A. S. Önder, and B. Torgler (2016): "The first cut is the deepest: repeated interactions of coauthorship and academic productivity in Nobel laureate teams," *Scientometrics*, 106, 509–524.

Chung, K. H., R. A. Cox, and K. A. Kim (2009): "On the relation between intellectual collaboration and intellectual output: Evidence from the finance academe," *Quarterly Review of Economics and Finance*, 49, 893–916.

Combes, P.-P. and L. Linnemer (2010): "Inferring missing citations: A quantitative multi-criteria ranking of all journals in economics," *mimeo*.

Conley, J. P., M. J. Crucini, R. A. Driskill, and A. S. Önder (2013): "The effects of publication lags on life-cycle research productivity in economics," *Economic Inquiry*, 51, 1251–1276.

Conley, J. P., A. S. Önder, and B. Torgler (2016): "Are all economics graduate cohorts created equal? Gender, job openings, and research productivity," *Scientometrics*, 1–22.

Corsi, M., C. D'Ippoliti, and G. Zacchia (2019): "Diversity and background of ideas: The case of research evaluation in economics," *Research Policy*, 48, 103820.

Currarini, S., M. O. Jackson, and P. Pin (2009): "An economic model of friendship: Homophily, minorities, and segregation," *Econometrica*, 77, 1003–1045.

Didegah, F. and M. Thelwall (2013): "Which factors help authors produce the highest impact research? Collaboration, journal and document properties," *Journal of Informetrics*, 7, 861–873.

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

Ellison, G. (2002): "The slowdown of the economics publishing process," *Journal of Political Economy*, 110, 947–993.

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

Fontana, M., F. Montobbio, and P. Racca (2019): "Topics and gegeographic diffusion of knowledge in top economics journals," *Economic Inquiry*, 57, 1771–1797.

Franceschet, M. and A. Costantini (2010): "The effect of scholar collaboration on impact and quality of academic papers," *Journal of Informetrics*, 4, 540–553.

Freeman, R. B. and W. Huang (2015): "Collaborating with people like me: Ethnic coauthorship within the United States," *Journal of Labor Economics*, 33, S289–S318.

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

Gazni, A., C. R. Sugimoto, and F. Didegah (2012): "Mapping world scientific collaboration: Authors, institutions, and countries," *Journal of the American Society for Information Science and Technology*, 63, 323–335.

Haeussler, C. and H. Sauermann (2020): "Division of labor in collaborative knowledge production: The role of team size and interdisciplinarity," *Research Policy*, 49, 103987.

Hamermesh, D. S. and S. M. Oster (2002): "Tools or toys? The impact of high technology on scholarly productivity," *Economic Inquiry*, 40, 539–555.

Haslam, N., L. Ban, L. Kaufmann, S. Loughnan, K. Peters, J. Whelan, and S. Wilson (2008): "What makes an article influential? Predicting impact in social and personality psychology," *Scientometrics*, 76, 169–185.

Heckman, J. J. and S. Moktan (2018): "Publishing and promotion in economics: The tyranny of the top five," *National Bureau of Economic Research WP*, 25093.

Heinze, T. and S. Kuhlmann (2008): "Across institutional boundaries? Research collaboration in German public sector nanoscience," *Research Policy*, 37, 888–899.

Hoekman, J., K. Frenken, and R. J. W. Tijssen (2010): "Research collaboration at a distance: changing spatial patterns of scientific collaboration wiwith Europe," *Research Policy*, 39, 662–673.

Hollis, A. (2001): "Co-authorship and the output of academic economists," *Labour Economics*, 8, 503–530.

Jaffe, A. B., M. Trajtenberg, and R. Henderson (1993): "Geographic localization of knowledge spillovers as evidenced by patent citations," *Quarterly Journal of Economics*, 108, 577–598.

Jones, B. F. (2010): "Age and great invention," *Review of Economics and Statistics*, 92, 1–14.

Kalaitzidakis, P., T. P. Mamuneas, and T. Stengos (2003): "Rankings of academic journals and institutions in economics," *Journal of the European Economic Association*, 1, 1346–1366.

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

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

Kosnik, L.-R. (2018): "A survey of JEL codes: What do they mean and are they used consistently?" *Journal of Economic Surveys*, 32, 249–272.

Laband, D. N. and R. D. Tollison (2000): "Intellectual collaboration," *Journal of Political Economy*, 108, 632–662.

Leahey, E. (2007): "Not by productivity alone: How visibility and specialization contribute to academic earnings," *American Sociological Review*, 72, 533–561.

McDowell, J. M. and M. Melvin (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, 597–608.

Melero, E. and N. Palomeras (2015): "The Renaissance Man is not dead! The role of generalists in teams of inventors," *Research Policy*, 44, 154–167.

Newman, M. E. J. (2010): *Networks: An Introduction*, Oxford University Press.

Nooteboom, B., W. van Haverbeke, G. Duysters, V. Gilsing, and A. van den Oord (2007): "Optimal cognitive distance and absorptive capacity," *Research Policy*, 36, 1016–1034.

Önder, A. S., S. Popov, and S. Schweitzer (2020): "Leadership in scholarship: Editors' influence on the profession's narrative," *SSRN Working Paper*, 3105539.

Önder, A. S. and H. Yilmazkuday (2020): "Thirty-five years of peer-reviewed publishing by North American economics PhDs: Quantity, quality, and beyond," *Open Economics*, 3, 70–85.

Rafols, I., A. L. Porter, and L. Leydesdorff (2010): "Science overlay maps: a new tool for research policy and library management," *Journal of the American Society for Information Science and Technology*, 61, 1871–1887.

Rosenblat, T. S. and M. M. Mobius (2004): "Getting closer or drifting apart?" *Quarterly Journal of Economics*, 119, 971–1009.

Rosenthal, S. S. and W. C. Strange (2003): "Geography, industrial organization, and agglomeration," *Review of Economics and Statistics*, 85, 377–393.

Rousseau, R., R. Guns, A. J. Rahman, and T. C. Engels (2017): "Measuring cognitive distance between publication portfolios," *Journal of Informetrics*, 11, 583–594.

Schweitzer, S. and J. Brendel (2020): "A burden of knowledge creation in academic research: evidence from publication data," *Industry & Innovation*, DOI:10.1080/13662716.2020.1716693.

Singh, J. (2005): "Collaborative networks as determinants of knowledge diffusion patterns," *Management Science*, 51, 756–770.

Stock, J. H. and M. Yogo (2005): *Identification and inference for econometric model: essays in honor of Thomas Rothenberg*, Cambridge University Press, chap. Testing for weak instruments in linear IV regression, 80–108.

Sutter, M. and M. Kocher (2004): "Patterns of co-authorship among economics departments in the USA," *Applied Economics*, 36, 327–333.

Vernon, M. M., E. A. Balas, and S. Momani (2018): "Are university rankings useful to improve research? A systematic review," *Plos One*, e0193762.

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

# Appendices

## A Merging Datasets of PhDs and Publications

The American Economic Association keeps record of doctoral dissertations submitted in economics programs of the US and Canadian universities. Each year's graduates are listed in that year's December issue of the Journal of Economic Literature (JEL) since 1987, whereas the December issue of the American Economic Review (AER) used to be the designated outlet for these lists before 1987. We collect data on North American PhDs' names, graduate institutes and graduation years from the *Doctoral Dissertations in Economics* sections in December issues of the AER from 1970 to 1986 and the JEL from 1987 to 2008. Records of peer-reviewed journal publications between 1990 and 2014 are obtained from the EconLit database. Since the detection of authors' location is an important ingredient in our analysis and records on authors' affiliations (hence their locations) are incomplete in most publications before 1990, we start our analysis from 1990. All journals that are contained in the EconLit database between 1990 and 2014 are also contained in our analysis, hence we do not restrict our analysis to top journals only or to any arbitrarily determined set of journals. We cover publications up to 2014 to guarantee that the youngest cohort in our PhD dataset (graduates of 2008) have had six years after graduation to build up their publication record. Finally, we consider those first time collaborations between North American PhDs where all involved parties have had at least two publications (single authored or co-authored) prior to that collaboration.

A crucial step in merging the list of PhDs with the publication database is to create a correct mapping of names in the PhD list onto author names in the publication database. There are two major sources of caveats. First, multiple authors can have the same name. Second, a single author may use several different names in their publications. This problem occurs especially when an authors' publications are recorded with different variations of their middle name in the EconLit database. Author disambiguation algorithms typically deal with systematic recognizing and mapping of author names in publications. Our disambiguation procedure, which can be accessed online at *https://github.com/SaschaSchweitzer/persons* employs a graph theoretic approach and follows a hierarchical process.

In the first step, we identify sets of author names with identical last names. Within the set of a given last name, we construct a graph of the relationships of the corresponding first names to each other. We categorize first names as either identical, different, subsets of each other or partially compatible. In our terminology, *John A.* is a subset of *John*. This is, because *John A.* provides more specific information than *John*, making it incompatible with another subset of names that *John* would still be compatible with. If none of those three categories apply to an entry, we define it as a partial match. For example, we categorize *J. Adam* to be a partial match with *John A.*. After determining all binary relationships between the names given, we model the sets of first names as nodes and their relationships as edges in a graph. Finally, we eliminate shortcuts between nodes to determine the minimum graph and traversed the non-forking paths of subset relationships from the graph's leafs upwards. That is, we match two entries with identical last names and the first names *John Adam* and *John*, respectively, to the same person if these two are the only two forenames. We would not match them, however, if there is a *John Alex* with the same last name in our data.

# B Additional Tables

Table B.1: Publication Quality at First Time Collaboration in Two Author Papers (Field Distance based on JEL)

| | CL Index | | KMS Index | | Citations | |
|---|---|---|---|---|---|---|
| | (1)OLS | (2)IV | (3)OLS | (4)IV | (5)OLS | (6)IV |
| Field Distance | -0.0200 | -0.403** | -0.00767 | -0.140 | -0.478** | -1.890* |
| | (0.0186) | (0.123) | (0.0187) | (0.121) | (0.147) | (0.784) |
| Specialization Max | -0.103** | -0.219*** | -0.0552 | -0.0953* | -0.668* | -1.098** |
| | (0.0343) | (0.0495) | (0.0350) | (0.0471) | (0.271) | (0.362) |
| Specialization Diff | 0.0655* | 0.133*** | 0.0507 | 0.0739* | 0.537* | 0.745** |
| | (0.0323) | (0.0383) | (0.0329) | (0.0368) | (0.256) | (0.283) |
| Quality Max | 1.294*** | 1.232*** | 1.394*** | 1.372*** | 2.407*** | 2.111*** |
| | (0.0556) | (0.0600) | (0.0628) | (0.0652) | (0.368) | (0.396) |
| Quality Diff | -0.788*** | -0.704*** | -0.926*** | -0.897*** | -1.163* | -0.711 |
| | (0.0752) | (0.0815) | (0.0822) | (0.0857) | (0.473) | (0.523) |
| Journal Quality[a] | | | | | 1.112*** | 1.107*** |
| | | | | | (0.128) | (0.127) |
| Individual/Pair Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Publication Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Graduation Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| $N$ | 3682 | 3682 | 3682 | 3682 | 1793 | 1793 |
| $Underidentification$ | | 2.28e-18 | | 2.28e-18 | | 1.69e-12 |
| $Cragg - Donald\ F$ | | 19.09 | | 19.09 | | 12.35 |

Standard errors in parentheses

$^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

[a]Journal Quality is captured by fitted residuals from (3)

Table B.2: Publication Quality at Subsequent Collaborations –including those initiated in three or more author papers

A. Including years of actual and potential collaborations

| | Using Pairwise Fixed Effects | | | Using Graduate and Social Controls | | |
|---|---|---|---|---|---|---|
| | (1)CL | (2)KMS | (3)Cites | (4)CL | (5)KMS | (6)Cites |
| Field Distance | -0.00509$^{+}$ | -0.00416$^{*}$ | -0.0260$^{***}$ | -0.00855$^{***}$ | -0.00465$^{***}$ | -0.0643$^{***}$ |
| | (0.00308) | (0.00210) | (0.00748) | (0.00186) | (0.00120) | (0.00440) |
| Specialization Max | 0.0208$^{**}$ | 0.00453 | 0.0636$^{***}$ | -0.00114 | -0.000606 | -0.0490$^{***}$ |
| | (0.00787) | (0.00536) | (0.0191) | (0.00451) | (0.00312) | (0.0102) |
| Specialization Diff | 0.00619 | 0.00547 | 0.0613$^{***}$ | -0.00605 | -0.00366 | -0.00535 |
| | (0.00653) | (0.00444) | (0.0158) | (0.00432) | (0.00287) | (0.00968) |
| Quality Max | -0.328$^{***}$ | -0.409$^{***}$ | -1.035$^{***}$ | 0.118$^{***}$ | 0.120$^{***}$ | 0.506$^{***}$ |
| | (0.0299) | (0.0203) | (0.0738) | (0.00678) | (0.00538) | (0.0139) |
| Quality Diff | 0.329$^{***}$ | 0.337$^{***}$ | 1.175$^{***}$ | -0.107$^{***}$ | -0.0905$^{***}$ | -0.524$^{***}$ |
| | (0.0246) | (0.0167) | (0.0602) | (0.00952) | (0.00731) | (0.0193) |
| Pairwise FE | Yes | Yes | Yes | No | No | No |
| Individual/Pair Controls | No | No | No | Yes | Yes | Yes |
| Publication Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Graduation Year FE | No | No | No | Yes | Yes | Yes |
| $N$ | 67615 | 67615 | 64242 | 68523 | 68523 | 65209 |

B. Including years of actual collaborations only

| | Using Pairwise Fixed Effects | | | Using Graduate and Social Controls | | |
|---|---|---|---|---|---|---|
| | (1)CL | (2)KMS | (3)Cites | (4)CL | (5)KMS | (6)Cites |
| Field Distance | -0.0343$^{+}$ | -0.0420$^{*}$ | -0.153 | -0.0331$^{**}$ | -0.0254$^{*}$ | -0.372$^{***}$ |
| | (0.0188) | (0.0189) | (0.146) | (0.0107) | (0.0104) | (0.0776) |
| Specialization Max | 0.0291 | -0.0228 | -0.535 | -0.0120 | -0.0301 | -0.777$^{***}$ |
| | (0.0506) | (0.0508) | (0.406) | (0.0258) | (0.0269) | (0.184) |
| Specialization Diff | -0.0322 | 0.0175 | 0.367 | 0.00679 | 0.0205 | 0.484$^{*}$ |
| | (0.0447) | (0.0448) | (0.370) | (0.0275) | (0.0278) | (0.204) |
| Quality Max | -1.687$^{***}$ | -2.330$^{***}$ | -1.432 | 0.993$^{***}$ | 1.079$^{***}$ | 1.936$^{***}$ |
| | (0.142) | (0.143) | (1.111) | (0.0344) | (0.0384) | (0.201) |
| Quality Diff | 1.302$^{***}$ | 1.667$^{***}$ | 0.284 | -0.530$^{***}$ | -0.631$^{***}$ | -1.105$^{**}$ |
| | (0.129) | (0.130) | (0.974) | (0.0571) | (0.0629) | (0.353) |
| Pairwise FE | Yes | Yes | Yes | No | No | No |
| Individual/Pair Controls | No | No | No | Yes | Yes | Yes |
| Publication Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Graduation Year FE | No | No | No | Yes | Yes | Yes |
| $N$ | 4924 | 4924 | 2182 | 4924 | 4924 | 2708 |

Standard errors in parentheses

$^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

Table B.3: Publication Quality at First Time and Subsequent Collaborations

| | Using Pairwise Fixed Effects | | | Using Graduate and Social Controls | | |
|---|---|---|---|---|---|---|
| | (1)CL | (2)KMS | (3)Cites | (4)CL | (5)KMS | (6)Cites |
| Field Distance | 0.0924*** | 0.0126 | 0.480*** | 0.0403*** | 0.0131+ | 0.00420 |
| | (0.0143) | (0.00986) | (0.0391) | (0.0109) | (0.00754) | (0.0370) |
| Field Dist Square | -0.146*** | -0.0337** | -0.719*** | -0.0789*** | -0.0302** | -0.194*** |
| | (0.0180) | (0.0125) | (0.0495) | (0.0139) | (0.00956) | (0.0471) |
| Specialization Max | 0.179*** | 0.0570*** | 0.782*** | 0.0516*** | 0.0200*** | 0.0264 |
| | (0.00995) | (0.00688) | (0.0279) | (0.00627) | (0.00433) | (0.0215) |
| Specialization Diff | -0.0352*** | -0.00417 | -0.102*** | 0.00358 | 0.0000702 | 0.0989*** |
| | (0.00815) | (0.00563) | (0.0227) | (0.00594) | (0.00410) | (0.0201) |
| Quality Max | -1.221*** | -1.295*** | -3.660*** | 0.148*** | 0.202*** | 0.433*** |
| | (0.0302) | (0.0209) | (0.0871) | (0.00836) | (0.00683) | (0.0231) |
| Quality Diff | 1.276*** | 1.109*** | 4.568*** | 0.00835 | -0.0711*** | 0.0401 |
| | (0.0254) | (0.0176) | (0.0725) | (0.0123) | (0.00975) | (0.0350) |
| Journal Quality[a] | | | 4.069*** | | | 3.630*** |
| | | | (0.0128) | | | (0.0336) |
| Pairwise FE | Yes | Yes | Yes | No | No | No |
| Individual/Pair Controls | No | No | No | Yes | Yes | Yes |
| Publication Year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Graduation Year FE | No | No | No | Yes | Yes | Yes |
| *N* | 77676 | 77676 | 69831 | 80102 | 80102 | 70421 |

Standard errors in parentheses

$^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

[a]Journal Quality is captured by fitted residuals from (1) and (4), respectively