NBER WORKING PAPER SERIES

HOW DO TRAVEL COSTS SHAPE COLLABORATION?

Christian Catalini
Christian Fons-Rosen
Patrick Gaulé

Working Paper 24780
http://www.nber.org/papers/w24780

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2018

We appreciate helpful comments from Alberto Galasso, Ina Ganguli, Avi Goldfarb, Xavier Giroud, Jeff Furman, Stepan Jurajda, Mara Lederman, Nikolas Mittag, Alex Oettl, Erin Scott, Paula Stephan, Scott Stern, Jane Wu, and seminar participants at Charles University, the EPFL, Harvard Business School, Imperial College London, Stockholm School of Economics, the Academy of Management, Universidad Carlos III, and National University of Singapore for helpful discussions and advice. All errors are our own. This research contributes to the agenda Strategie AV21. Fons-Rosen acknowledges financial support by the Spanish Ministry of Economy and Competitiveness (ECO2014-55555-P). Christian Catalini acknowledges the support of the Junior Faculty Research Assistance Program at the MIT Sloan School of Management. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2018 by Christian Catalini, Christian Fons-Rosen, and Patrick Gaulé. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

How Do Travel Costs Shape Collaboration?
Christian Catalini, Christian Fons-Rosen, and Patrick Gaulé
NBER Working Paper No. 24780
June 2018
JEL No. L93,O18,O3,O31,O33,R4

ABSTRACT

We develop a simple theoretical framework for thinking about how geographic frictions, and in particular travel costs, shape scientists' collaboration decisions and the types of projects that are developed locally versus over distance. We then take advantage of a quasi-experiment - the introduction of new routes by a low-cost airline - to test the predictions of the theory. Results show that travel costs constitute an important friction to collaboration: after a low-cost airline enters, the number of collaborations increases by 50%, a result that is robust to multiple falsification tests and causal in nature. The reduction in geographic frictions is particularly beneficial for high quality scientists that are otherwise embedded in worse local environments. Consistent with the theory, lower travel costs also endogenously change the types of projects scientists engage in at different levels of distance. After the shock, we observe an increase in higher quality and novel projects, as well as projects that take advantage of complementary knowledge and skills between sub-fields, and that rely on specialized equipment. We test the generalizability of our findings from chemistry to a broader dataset of scientific publications, and to a different field where specialized equipment is less likely to be relevant, mathematics. Last, we discuss implications for the formation of collaborative R&D teams over distance.

Christian Catalini
MIT Sloan School of Management
100 Main Street, E62-480
Cambridge, MA 02142
and NBER
catalini@mit.edu

Christian Fons-Rosen
Universitat Pompeu Fabra and CEPR
Barcelona GSE
Carrer Ramon Trias Fargas, 25-27
08005 Barcelona
SPAIN
christian.fons-rosen@upf.edu

Patrick Gaulé
University of Bath
patrickgaule@gmail.com

1 Introduction

The drastic reduction in communication costs brought by the diffusion of the internet initially led to claims about a future in which technology could overcome geographic frictions and facilitate the rapid exchange of ideas, goods and services independent of distance (Cairncross, 1997; Friedman, 2005). Empirically, this "death of distance" hypothesis has found limited support, as most evidence points to agglomeration mattering more, not less than before across a variety of settings (Leamer & Levinsohn, 1995; Blum & Goldfarb, 2006; Forman et al., 2005; Agrawal et al., 2015). Instead of substituting for co-location, digital interactions often complement it (Agrawal & Goldfarb, 2008),[1] resulting in non-obvious changes in how teams and organizations structure collaborations and develop new ideas when communication costs are low, but teamwork and R&D require specialized expertise and resources that are geographically dispersed (Adams et al. 2005; Jones, Wuchty & Uzzi 2008; Wuchty, Jones & Uzzi 2007)[2].

Moreover, not all types of interactions have benefited in the same way from improvements in communication technology. Co-location plays a disproportionate role in the serendipitous discovery of new collaborators and ideas (Catalini, 2017), and in the absence of offline opportunities for interaction, search frictions can prevent individuals from finding ideal collaborators even within the boundaries of the same institution (Boudreau et al., 2017). Similarly, exchanges that require the transfer of complex information and tacit knowledge (Polanyi, 1958; Von Hippel, 1994) still heavily rely on face-to-face interactions (Rosenthal & Strange, 2001; Gaspar & Glaeser, 1998; Storper & Venables 2004). As a result, firms, communities of experts and teams invest substantial amounts of time, effort and resources to ensure that the right individuals can be co-located – even if only temporarily – to discuss ideas, make

[1] Agrawal & Goldfarb's (2008) study of Bitnet, an internet predecessor, finds that as more academic institutions joined the network, collaboration among affected scientists increased. Interestingly, their results hint at the technology being a complement to offline interactions, as co-authorship increases disproportionately among university pairs that are co-located. Other studies have found an effect of bitnet on collaborations in the academic life sciences (Ding et al. 2010), and of the internet on cooperative R&D between firms (Forman & Zeebroeck 2012).

[2] By 2000, less than 20% of papers in science and engineering were single authored. Similar patterns, and in particular the rise of coauthorship and distant coauthorship, have been documented in economics. See Gaspar & Glaeser (1998), Hamermesh & Oster (2004), Rosenblatt & Mobius (2004)

progress on projects, and develop the relationships that can later support more effective interactions over distance. Such temporary forms of co-location have been shown to foster both idea diffusion and the formation of new collaborations (Chai & Freeman, 2018).[3]

If face-to-face interactions are instrumental in finding and evaluating new collaborators, establishing trust, and advancing joint work, then as communication costs drop, this complement to remote interactions becomes not only more valuable, but possibly the key friction in the formation and operation of geographically distributed teams. Ironically, by making online communication extremely efficient, the internet may have enhanced the role that travel technology plays in the economy.

The objective of this paper is to develop and test a simple theoretical framework for thinking about how geographic frictions, and in particular travel costs, shape collaboration decisions and the types of projects that are developed locally versus over distance. The model highlights a key trade-off individuals face when deciding if they should work with a local versus a distant collaborator: whereas the global pool of potential collaborators is often deeper and may therefore offer an ideal match, collaboration over distance incurs additional communication and travel costs. We build on this basic tension in a context where individuals endogenously allocate effort to projects based on their potential, and where a project's riskiness or the need for complementary expertise, equipment or resources can influence with whom a project is pursued. While simple, we believe the framework captures an increasingly relevant challenge: to be able to solve problems of increasing complexity, teams of specialized experts have to be put together (Jones, 2009), but this often involves collaboration over distance.

We take advantage of a quasi-experiment – the introduction of new routes by a major low-cost airline – to test the predictions of the theory within the context of collaborations between scientific labs. The setting allows us to observe the full set of scientists at risk of

[3] Chai & Freeman (2018) compare collaboration patterns among attendees of the Gordon conference before and after the event in a difference-in-differences framework using a carefully constructed control group of qualitatively similar non-participants. They find that attendees are more likely to be cited by and collaborate with other participants, especially if they were new to this community of experts. In a related paper, Campos, de Leon & Mcquilin (2018) document that a conference cancellation led to a decrease in individuals likelihood of co-authoring together.

collaboration in any given year as well as important characteristics about them such as their age, career stage, past productivity, area of specialization, and departmental funding.

The cheaper fares brought by the expansion of the low-cost airline (Southwest Airlines)[4] are part of a broader, 50% reduction in the cost of air travel that took place in the United States over the last 30 years (Perry 2014).[5] Furthermore, they provide a source of plausibly exogenous variation in the cost of conducting research between scientists at the affected airports.

Using a difference-in-differences empirical strategy we are able to recover a causal estimate of the effect of a reduction in travel costs not only on the rate of collaboration, but more importantly on the type of projects scientists pursue. Results show that travel costs are an important friction to collaboration: after Southwest entry, the number of collaborations increases by 50%, a result that is robust to multiple falsification tests and causal in nature. The reduction in geographic frictions is particularly beneficial for high quality scientists that are otherwise embedded in worse local environments. Consistent with the theory, lower travel costs also endogenously change the types of projects scientists engage in locally versus over distance. After the shock, we observe an increase in higher quality and more novel projects, as well as projects that take advantage of complementary knowledge and skills between subfields, or that rely on specialized equipment. We test the generalizability of our findings within chemistry to a broader dataset of scientific publications, and to mathematics, a field where specialized equipment is less likely to be relevant. Last, we discuss implications for the formation of collaborative R&D teams in the presence of geographic frictions.

The rest of the paper is as follows: in Section 2, we introduce our theoretical framework. Section 3 provides additional institutional details about collaboration in chemistry and describes the data. We present the main results in Section 4, and a series of extensions targeted at testing the generalizability of our findings in Section 5. Section 6 concludes.

[4] Southwest has been described as the most significant development in the market structure of the U.S. airline industry by the Transportation Research Board (1999) and by industrial economists (Morrison 2001, Borenstein & Rose 2007, Goolsbee & Syverson 2008).

[5] Kim, Morse & Zingales (2009) and Freeman, Ganguli & Murciano-Goroff (2014) note that secular declines in both communication costs and air travel costs may have facilitated long distance collaborations.

2 Theoretical Framework

The objective of this section is to develop a simple theoretical framework to highlight key trade-offs scientists face when deciding if they should collaborate with a local or a distant co-author, and how much effort they should dedicate to a collaboration based on its intrinsic potential. The model generates novel predictions about how travel costs shape collaboration decisions, which we then test using our data.

We start by assuming that because the global pool of potential co-authors offers more variety than the local one, it is on average possible to find better matches when team formation is not constrained by geographic distance. The quality of a match may depend on complementary ideas, knowledge, skills, equipment, and resources that a co-author brings to a project. Of course, because of agglomeration forces, as the size, specialization and quality of a region's local pool increases, scientists will rely less on distant co-authors. To account for this, in an extension of the baseline model we allow for the share of 'first best' co-authors available locally to vary.[6]

Our setup is straightforward: ideas are born with intrinsic quality q, but require effort e to be developed and achieve their full potential v. Since scientists observe a noisy signal of q before starting a project, they will allocate more effort, time and resources to projects that have higher potential (i.e. in our model, effort is endogenous to potential). At the same time, since research constitutes an uncertain endeavour, even when scientists apply effort projects are only successful with probability p, which depends on the quality of the co-author match. Thus, the realized value of a project can be expressed as $v = p_i q e$, where p_i (with $i = G, B$) is higher when a good match between co-authors is achieved (p_G), relative to a bad match (p_B).

Whereas the global pool may offer a better match between co-authors (i.e. p_G) and

[6] The fraction of first best co-authors in the global pool is assumed to be z. Since the global pool can be seen as an average over all possible local pools, the fraction of first best co-authors in a given local pool w can be either higher, lower, or equal to z. If $w > z$, then scientists will never collaborate over distance, as they would incur additional costs but would not be more likely to find an ideal co-author over distance. Therefore, the range of values of w that provides a meaningful trade-off is $0 \leq w < z$. To simplify the exposition, in the paper we will assume $w = 0$. More general cases are discussed in the Appendix.

increase the chances of realizing a project's full potential v, collaborating over distance introduces additional costs, as scientists have to travel for face-to-face interactions, and may be less effective at communicating complex information remotely. As a result, scientists face a trade-off between less choice locally, and increased communication and travel costs over distance. In the next sections, we perform comparative statics and explore this tension in more detail.

2.1 Local versus Distant Collaborations

The scientist's payoff from developing an idea with a local co-author for a given level of effort e is:

$$\pi_L(e) = p_B q e - c(e) \tag{1}$$

where $c(e)$ is the cost of effort which we assume for tractability to have the following convex function: $c(e) = \frac{\alpha}{2}e^2$. Thus, equation (1) can be re-written as:

$$\pi_L(e) = p_B q e - \frac{\alpha}{2}e^2 \tag{2}$$

The first order condition yields an optimal effort level of $e_L^* = \frac{p_B q}{\alpha}$, which is increasing both in project quality q, and in the quality of the co-author match p_i. Intuitively, scientists are more willing to apply effort to projects with higher potential, and to projects they are working on with better matched co-authors. Inserting e_L^* back into (2), we obtain a scientist's payoff for a local collaboration given the optimal effort level as:

$$\pi_L^* = \frac{(p_B q)^2}{2\alpha}. \tag{3}$$

How does this compare to a distant collaboration? In our setup, over distance scientists have a higher chance of securing the ideal co-author because the global pool offers more variety. At the same time, this does not happen all the time, and scientists have to incur additional communication and travel costs t_i to develop a project over distance. We assume that with probability z scientists find a first best co-author and secure p_G, and with probability $(1 - z)$ they land a co-author of the exact same level they would have found in the local pool p_B. Thus, the payoff for a distant collaboration can be written as:

$$\pi_D(e,t) = (1-z)[p_B q e_B - \alpha\frac{e_B^2}{1+t_B} - \beta t_B^2] + z[p_G q e_G - \alpha\frac{e_G^2}{1+t_G} - \beta t_G^2], \tag{4}$$

where e_i and t_i with $i = G, B$ are the optimally chosen levels of effort and travel for perfectly matched co-authors (p_G) versus imperfectly matched ones (p_B).

Traveling enters as a convex cost ($t_i = [0,1]$, scaled by a parameter β)[7], but also increases the chances of success because it improves the ability to communicate complex information, coordinate work and make progress on a project through face-to-face interactions. This trade-off allows for interesting cases to emerge where temporary co-location between distant co-authors is expensive but also helpful, and can therefore lead to both higher and lower payoffs relative to a collaboration on the same project with a local co-author. For simplicity, we assume that once a local versus distant co-author has been chosen for a project, it is too costly to switch type without starting a completely new project.[8] We also assume that before a substantial amount of effort and travel is dedicated to a project, the quality of a co-author match has been revealed. The first order conditions with respect to effort and traveling are, respectively:

$$e_D(t_D) = \frac{p_i q(1+t_D)}{2\alpha}$$

$$\alpha\frac{e_D^2}{(1+t_D)^2} - 2\beta t_D = 0,$$

where $i = B, G$ depends on whether the distant co-author has led to a first best or a second best match. Combining both first order conditions, one can show that the optimal levels of travel and effort for a distant collaboration are:

$$t_i^* = \frac{(p_i q)^2}{8\alpha\beta} \tag{5}$$

$$e_i^* = \frac{p_i q}{2\alpha}(1+t_i^*) = \frac{p_i q}{2\alpha}(1+\frac{(p_i q)^2}{8\alpha\beta}) \tag{6}$$

[7] If $t_i = 1$, face-to-face communication is always available (as with a local collaborator), and the cost of effort would be the same under both scenarios. Advancements in communication technology and virtual reality can be therefore thought of as changes in t_i.

[8] One intuitive way to think about changing a co-author within our simple framework is to imagine the original project failing, and a new one being launched with a different team.

If we plug these back into the payoff function we obtain:

$$\pi_i^* = \frac{(p_i q)^2}{4\alpha}(1 + \frac{1}{2}t_i^*) = \frac{(p_i q)^2}{4\alpha}(1 + \frac{(p_i q)^2}{16\alpha\beta}) \tag{7}$$

Thus, the overall payoff over distance is:

$$\pi_D^* = (1 - z)\pi_B^* + z\pi_G^* \tag{8}$$

Comparing payoff equation (3) for local collaborations with equation (8) for distant ones is informative independent of travel costs (which we discuss in detail in the next section). For example, it allows us to explore how the relative appeal of a local versus distant collaborations changes as the comparative advantage of the global pool (z) over the local one varies:

$$\frac{\partial[\pi_D^* - \pi_L^*]}{\partial z} = \pi_G^* - \pi_B^* > 0 \tag{9}$$

Intuitively, an increase in the likelihood of finding a first best co-author in the global pool will lead to a relative increase in the payoff for distant over local collaborations. Similarly, if scientists enjoy a high quality local environment with good matches (e.g. they are in an agglomerated research cluster), they will find limited benefits from collaborating over distance.

Until now, we have assumed that all scientific projects have the same risk of failure. At the same time, more novel and exploratory projects – such as those that recombine knowledge across disciplines – typically entail substantially more risk than incremental ones. To account for this, we introduce risk as γ, and link it to the overall probability of success through $p_G = (1+\gamma)p_B$. For a given p_B, a low γ means that the quality of the match between co-authors will have a minor influence on the chances of realizing a project's full potential. One can think of low γ projects as relatively more straightforward ones where most of the techniques and ideas are established (or everyone has access to similar infrastructure to work on them), and the gap between working with the best possible co-author versus anyone else is small. As a result, when γ is low, the relative appeal of the global talent pool is more limited. When a project is instead extremely risky, scientists will be more willing to travel to work with the ideal co-author and increase their chances of success:

$$\frac{\partial[\pi_D^* - \pi_L^*]}{\partial\gamma} = z\frac{\partial[\pi_G^*]}{\partial\gamma} > 0 \tag{10}$$

An extreme example of this is a project for which there are only a few leading experts or key labs with the right equipment (e.g. CERN, LIGO etc.), and the difference between working with them relative to working with a local alternative is large.

Last, when comparing local versus distant collaborations, it is useful to point out that increases in the underlying, intrinsic project quality (q) have an ambiguous effect on the choice of co-author type. As shown in the Appendix, which type of collaborations prevail still depends on the basic trade-off between the quality of the match between scientists and travel costs (since a distant collaboration can still leave a scientist with a match of similar quality to the local alternative).

2.2 Reductions in Travel Costs

How does a reduction in travel costs affect the types of collaborations scientists engage in? In this section, we perform comparative statics to see how cheaper fares like the ones brought by a low-cost airline change the relative attractiveness of local versus distant collaborations, and how this effect varies for projects of different type (more versus less risky, and higher versus lower potential). To simplify the notation and exposition, we define $\theta = \frac{1}{\beta}$ (which is the inverse of travel costs) as the "ease of travel". One can think of an improvement in θ as better infrastructure that allows scientists to meet with their distant co-authors at a lower cost and with lower frictions. The derivative of relative profits with respect to θ is:

$$\frac{\partial[\pi_D^* - \pi_L^*]}{\partial\theta} = (\frac{q^2}{8\alpha})^2[(1-z)p_B^4 + zp_G^4] > 0 \tag{11}$$

θ does not matter for the returns to local collaborations (π_L^*) as no travel is required, but makes face-to-face interactions with distant co-authors less expensive. Therefore, it is intuitive that with better travel technology the relative attractiveness of the global talent pool increases,[9] as accessing it is now more cost effective.

[9] Notice that this holds for the general case of $0 \leq w < z$, and is not limited to cases where $w = 0$.

But how does this effect vary with the ex-ante relative competitiveness of the local pool? I.e., how does this vary for regions that offered better versus worse alternatives to begin with? Remember that this is captured in our framework by the share of first best co-authors that are in the global pool z. Taking the first order condition with respect to θ and z we obtain:

$$\frac{\partial[\pi_D^* - \pi_L^*]}{\partial\theta\partial z} = (\frac{q^2}{8\alpha})^2(p_G^4 - p_B^4) > 0 \tag{12}$$

Which leads to the following prediction:

Prediction 1: *A reduction in travel costs will be especially beneficial for researchers that have access, ex-ante, to a relatively worse pool of local co-authors.*

If we instead take the derivative with respect to quality:

$$\frac{\partial[\pi_D^* - \pi_L^*]}{\partial\theta\partial q} = q(\frac{q^2}{4\alpha})^2[(1-z)p_B^4 + zp_G^4] > 0 \tag{13}$$

we see that after an improvement in ease of travel, higher quality projects are more likely to be undertaken with better matched co-authors. Since these are on average more abundant within the global pool, it follows that:

Prediction 2: *A reduction in travel costs will be especially beneficial for distant collaborations on higher quality projects.*

Last, if we do not assume that all projects have the same risk of failure and explore how the effect changes with project riskiness, we obtain:

$$\frac{\partial[\pi_D^* - \pi_L^*]}{\partial\theta\partial\gamma} = z(\frac{q^2}{4\alpha})^2(1+\gamma)^3 p_B^4) > 0 \tag{14}$$

which shows that a reduction in travel costs makes the global pool disproportionately more appealing as the riskiness of projects increases,[10] or restated:

Prediction 3: *A reduction in travel costs will be especially beneficial for distant collaboration on higher risk projects.*

Empirically, to proxy for riskiness we will rely on how novel the keywords used by the authors on a focal paper are, as well as explore results for collaborations that span different sub-fields of chemistry versus not. We now turn to our data to test these predictions.

[10]Intuitively, in our model this is a result of the complementarity between risk and the quality of a co-author match.

3 Data

3.1 Collaboration in Chemistry

Scientific research in chemistry, similar to other fields of science, is an increasingly collaborative endeavour, resulting in a higher number of co-authors per paper over time. However, it largely remains a lab-based science, and has not embraced the large scale, big science projects observed, for example, in physics. In our sample, the median number of co-authors per paper is 4, and many of the authors are graduate students, post-docs or technicians. These perform most of the experiments and day-to-day work on a project. They are employed by the lab of a faculty member (principal investigator) who obtains funding for the lab, directs research projects, appears as a co-author on all publications, oversees resource allocation and effectively decides whether to collaborate or not with other labs. While many research projects involve a single principal investigator, collaborations between labs and principal investigators are common as well. Consistent with the findings from large scale surveys of scientists (Freeman, Ganguli, and Raviv Murciano-Goroff 2014), in our conversations with U.S. principal investigators in chemistry, complementary expertise, skills, materials or new types of experiments are all mentioned as reasons for collaboration between labs. As in other fields of science, collaborations are sourced through the principal investigators and junior members' professional networks, serendipitous interactions at conferences, email, etc. In the paper, we focus on collaborations between principal investigators, which are essentially collaborations between different labs.

3.2 Data Sources

To examine the effect of changes in travel costs on scientific collaboration, we combine data on scientists with publication records and air transportation information. Within the chemistry sample, biographical information on scientists enables us to effectively disambiguate publication data, while also allowing us to separate faculty members from other types of authors.

Air Transportation Data - To recover information on when Southwest operated flights between different routes, as well as information on prices, passengers and miles flown, we use data from the Airline Origin and Destination Survey (DB1B) of the U.S. Bureau of Transportation Statistics. The DB1B is a 10% random sample of airline tickets from reporting carries in each quarter. For each itinerary, the DB1B records all connecting airports (including origin and destination), the itinerary fare, and other information. This data is available only from 1993, hence we will focus on Southwest entry decisions that occur after 1993.

Match Between Airports and Universities. We compute distances between airports and universities using Google Maps. The matching between universities and airports is complicated by the fact that the same metropolitan area could be served by multiple airports (e.g. O'Hare and Midway in Chicago), or that a college town could be half-way between two airports. We chose to match universities to all airports within a 50 miles radius. We code the year of Southwest entry for a pair of universities as the first year in which Southwest operates a flight on any route whose endpoints (airports) are within 50 miles of the respective universities. Results are robust to narrowing this definition further (e.g. 25 miles, 10 miles), see Table A-9.

Data on Scientists. Our focus is on collaborations between faculty members (and therefore effectively across labs) in the discipline of chemistry[11], in part due to data availability, and in part because of the short publications cycles in this discipline. For biographical information on scientists, our data source is the directory of graduate research published by the American Chemical Society. Intended as a source of information for prospective graduate students, this directory provides comprehensive listings of faculty affiliated with U.S. departments granting PhDs in chemistry, chemical engineering and biochemistry. Besides faculty names and departmental affiliations, the directory provides information on year of birth, gender and education. The directory is published biannually in print and since 1999

[11]Chemistry, which focuses on the composition, structure, transformations and properties of matter, is a large discipline, with chemistry PhD graduates accounting for 30% of U.S. PhD science graduates.

on the web.[12] We combine the directories from 1991 to 2013 to build a longitudinal panel of over 20,000 scientists. We complement this information with department-level R&D expenditures from The National Science Foundation (NSF) Survey of Research and Development Expenditures at Universities and Colleges.

Publication Data. We match faculty names to publication data from Scopus covering more than 200 chemistry journals (including all journals from the American Chemical Society), multidisciplinary journals and major journals in neighboring disciplines.[13] Within chemistry, the match between publications and scientists is facilitated by the fact that we know institutional affiliations from the American Chemical Society faculty data. We match publications to faculty based on last name, first and (if non-missing) middle initials, department and university affiliation. From publication data, we construct for each scientist time-varying measures of past productivity (with a moving average over the last three years of publication counts weighted by journal impact factor). We also infer our main outcome, copublications, from bibliometric data combined with the faculty data.

A key strength of our data is that we know when individuals enter and exit the profession and therefore are at risk of collaborating with others. If we were inferring copublications from publication data only, we could hardly distinguish between active scholars and individuals that have retired or are not doing research in the field. Papers are counted as a copublication between all pairs of faculty members involved.[14]

Additional Key Outcomes. For part of the analysis, we will weight copublications by the citations they have received as a proxy for their impact and quality. Citation counts

[12] The American Chemical Society also produced a CD-ROM for the years 1991-1993.

[13] Scopus is one of the two major bibliometric databases (along with ISI Web of Science). Our set of chemistry journals includes all journals from the American Chemical Society, as well as any chemistry journal with impact factor above 2. Our set of multidisciplinary journals includes *Nature*, *Science*, *Cell* and the *Proceedings of the National Academy of Sciences*. Our set of major journals in neighboring disciplines includes all journals with impact factor above 6 in physics, biology, material science and nanotechnology.

[14] The majority (75%) of papers matched to a faculty member have exactly one faculty author, 21% percent have two, and less than 4% have more than two authors. Both papers with one faculty author and papers with multiple faculty authors typically have several non-faculty authors. We focus on faculty authors because they are the ones usually making the decision to collaborate. Papers in chemistry journals that are not matched to any of our U.S. faculty authors are likely to be from foreign scientists, scientists working in corporate environments and federal labs.

originate from Scopus, are at the article level and are counted from the year of publication until 2013. We also construct measures of novelty based on author keywords. These are based on the entire corpus of articles within chemistry journals and related fields. For each keyword, we calculate the share of papers in a given year that contains the keyword – a proxy for how popular it is at any point in time. We then calculate the first and second derivative of this measure relative to the previous year. If both the first and second derivative are positive, then the keyword is classified as novel since its use is quickly accelerating. Additionally, if the first derivative is zero, and the second is positive, then we are at a local minimum right before a keyword takes off, which we also consider as novel use. Aggregating up at the paper level, a publication is considered novel if it has an above the median number of novel keywords. Similarly, we constructed proxies for the equipment-intensiveness of a focal publication by first collecting a large-scale list of keywords associated with chemistry equipment,[15] and then checking this list against the keywords used in each paper. Papers with an above the median number of equipment-related keywords are classified as equipment-intensive.

3.3 Descriptive Statistics

Our dataset covers over 20,000 scientists and their collaborations. However, we focus on a specific subset of pairs of scientists who experience Southwest entry and for whom we have variation in collaboration over time. Since all regressions include scientist-pair fixed effects, pairs that never collaborate drop out of the sample. In the Appendix, we show that our main result is robust to replacing scientist-pair fixed effects with city-pair fixed effects and including a random sample of non-collaborating pairs.

We have 15,244 pairs of scientists who collaborate at least once.[16] Excluding co-authors that are in the same department, we have 8,311 pairs of scientists in our sample. Only a minority (1,158) of these pairs experience Southwest entry during our analysis period of 1993-2012, either because for the other 7,153 pairs Southwest is already operating a flight,

[15] This was built by scraping and compiling an inventory of equipment for sale in online catalogues and stores targeted at a wide range of chemistry labs.

[16] Our dyadic data is not directed, and thus is symmetric: the pair between i and j is the mirror image of the pair between j and i. The 15,244 figure is after dropping an equal number of symmetric observations.

or because Southwest never flies between the relevant endpoints. We drop pairs in locations where Southwest enters but then leaves within two years, as well as pairs where Southwest entry coincides with the move of a scientist.[17] Finally, we also exclude pairs that are within less than 200 miles of each other as air travel is unlikely to be their main travel option.[18] Our final analysis covers 758 pairs of scientists corresponding to 845 individuals.

[Insert Table 1 about here]

Table 1 displays descriptive statistics for our chemistry sample at different levels of analysis: individual, individual-pair and individual-pair-year. Most individuals in the sample are male (90%) with an average age at the time of Southwest entry of 49. We do not observe individual research budgets but as a proxy we use departmental R&D expenses divided by the number of faculty members in the department. The average in our sample is $279,000 at the time of Southwest entry. According to the NSF Survey, R&D expenses include compensation for R&D personnel, equipment and indirect costs. In terms of specialization[19], the largest area is physical chemistry (32%), followed by biochemistry (22%), inorganic chemistry (13%), organic chemistry (14%) and material science (11%).

We observe the 758 pairs for 17 years on average,[20] corresponding to 13,147 observations at the individual-pair-year level. Southwest entry events map to 413 distinct new routes. The median pair experiences Southwest entry in 1999, but we observe Southwest entry from 1994 to 2011. The mean number of copublications over the whole period is 1.9, but the majority of pairs copublishes once. Only 9% of pairs collaborates both before and after Southwest entry.

[17] Scientists in our sample may move from one department to another, in some cases leading to a change in whether they are connected by Southwest or not. We want changes in Southwest status to be driven by Southwest entry decisions rather than by location decisions, and thus exclude pairs who happen to move in the same year as Southwest enters, the year before or the year after.

[18] Results are robust to decreasing this threshold to 100 or 50 miles.

[19] Specialization is inferred based upon the journals in which a scientist publishes. For instance, a faculty member who often publishes in the Journal of Biological Chemistry is assumed to be specialized in biochemistry.

[20] A pair is in our sample for a maximum of 22 years (from 1991 to 2011). We observe some pairs for less than 22 years due to pair members starting their first faculty appointment after 1991, retiring before 2011, or otherwise no longer being listed in the ACS faculty directory (e.g. moving to industry or to a foreign country).

[Insert Table 2 about here]

It is useful to compare our analysis sample to other distant pairs that do not experience Southwest entry. We have approximately 6,000 such pairs. These include pairs where Southwest is already present in the relevant market prior to 1993 when our sample starts, or has not entered by 2012 when it ends. They also include cases where one of the pair members is a new faculty hired after Southwest has already entered. The comparison is shown in Table 2. The pairs that experience Southwest are not statistically different from the others in terms of publications, but are slightly older (51 versus 49 years) and are observed on average for a slightly longer period of time (17 versus 14 years).[21] Importantly, there is no significant difference in terms of R&D budgets or propensity to be in different subfields of chemistry.

4 Empirical Strategy and Main Results

Our empirical specification is a straightforward difference-in-differences framework at the scientist-pair level where we exploit variation in Southwest entry across different airport pairs over time. It includes scientist-pair fixed effects and is estimated using a Poisson model:

$$Y_{ijt} = \beta AfterSW_{ijt} + \mu_t + \gamma_{ij} + \epsilon_{ijt}$$

where Y_{ijt} is the number of copublications between scientist i and scientist j in year t, $AfterSW_{ijt}$ is an indicator variable that takes value 1 after Southwest entry, μ_t is a year fixed effect, γ_{ij} is a pair fixed effect to control for unobservable, time-invariant differences between pairs of scientists, and ϵ_{ijt} is an idiosyncratic error term.

Our analysis examines the change in the rate of collaboration and in the types of papers that emerge over time for pairs that co-author at least once. The pair fixed effects completely capture pairs of scientists for which we never see activity, and thus we remove these from the analysis without empirical consequences. Robust standard errors are clustered at the pair level.

[21] This makes sense since a longer observation period mechanically increases the chances of experiencing Southwest entry.

4.1 Southwest Entry and Changes in Passengers, Prices, Miles and Transfers

Before our main analysis, we check how the arrival of Southwest affects some of the key passenger and fare metrics of interest in the air travel industry. In this exercise, we run regressions at the airport-pair level, and compare a number of outcomes before and after Southwest entry. Regressions include airport-pair fixed effects and year fixed effects. The coefficients in Table 3 reflect the types of changes one would expect to take place after the arrival of a low-cost competitor: the number of passengers increases by approximately 44%, and prices drop by around 20%. We do not find any effect on the average miles flown[22] or on direct flights, and the reduction in the number of transfers is extremely small. Overall, results are consistent with Southwest lowering the cost of air travel without drastically changing the types of routes available or the number of miles passengers have to fly to connect between two endpoints.

[Insert Table 3 about here]

4.2 Changes in Collaboration and Evidence for a Causal Interpretation

As discussed in the theoretical framework, after a reduction in travel costs the relative attractiveness of the global talent pool increases, since accessing it becomes more cost effective. This should lead to an increase in collaboration between the affected locations. As can be seen in Column 1 of Table 4 (which uses the main econometric specification we described at the beginning of Section 4), after Southwest entry we observe a large and significant increase in collaboration of approximately 50% between scientists at the connected end points.[23] While the magnitude is large, it is off a small base (the mean of the dependent variable

[22] The data from the Bureau of Public Transportation includes the number of miles flown for each itinerary. Differences in miles flown arise from the number of connections an itinerary involves. We compute average miles flown as the average across all passengers travelling between two airports in a given year.

[23] Collaboration between scientists is increasing over time. In our regressions, this trend is captured by the inclusion of year fixed effects. Therefore, one can interpret our estimates as the relative percentage increase in collaboration due to Southwest entry once the underlying increasing trend in collaboration has been accounted for.

is approximately 0.1), and comparable with previous studies on the impact of communications, search costs and co-location on scientific collaboration: Agrawal & Goldfarb (2008) find that Bitnet increased the likelihood of collaboration between pairs of universities by 40%; Boudreau et al. (2017) find that a 90-minute structured information sharing session led to a 75% higher probability of co-applying for a grant; Catalini (2017) estimates that exogenous co-location increased the chance of a collaboration between labs on the Jussieu campus of Paris by 3.5 times.

[Insert Table 4 about here]

One may worry that Southwest entry is systematically correlated with time-varying factors such as growth of the universities (or the regional economies) at both ends of the routes, and therefore that collaboration would have increased even in the absence of a reduction in travel costs. While our main specification already controls for aggregate time trends through year fixed effects, the validity of our results could be threatened by systematic, time-varying factors that affect the target locations around the time of Southwest entry. In Column 2, we mitigate these concerns by controlling for two possible time-varying confounders: the age of the scientist pair, and the (log of) departmental R&D budget per faculty member. The first one accounts for changes in the incentives to collaborate as scientists progress in their careers, the second for changes in the local economies. Whereas the coefficients for the controls are positive and significant, our main result is unaffected. In Column 3, we additionally control for the number of years that have passed since both scientists obtained their PhD, a proxy for their ability to both decide who they want to collaborate with. This estimated coefficient is negative and significant but again does not affect the estimate of *Southwest entry*. In Column 4 we study the dynamic effects of the reduction in travel costs by replacing the treatment indicator for *Southwest entry* from Column 1 with a set of four dummy variables capturing the years around the treatment. For example, the indicator *Southwest entry (-1)* is equal to one if the focal scientist-pair observation is recorded one year prior to the treatment. The other indicator variables are defined analogously with respect to the year of treatment

(0), the first year after treatment (1), and two or more years after treatment (2+).[24] The coefficient for *Southwest entry (-1)*, which would capture any 'effect' of the new airline routes before their introduction, is insignificant, suggesting that there is no collaboration pre-trend in the data, i.e. it is only once travel costs are reduced that the coefficients turn positive and statistically significant.

[Insert Figure 1 about here]

A graphical version of a similar exercise with a full set of coefficient estimates for the 5 years before and 5 years after Southwest entry is displayed in Figure 1. There is again no collaboration pre-trend before Southwest launches a route, and it is only after the new route is available that the estimated coefficients are positive and steadily increasing in magnitude.[25] It is useful to highlight that publication lags in chemistry are substantially shorter than in the social sciences: when studying the 10 major analytical chemistry journals (1985-1999), Diospatonyi et al. (2001) find median lags between submission and publication of 3 to 10 months, with some journals publishing papers within 2 months of first submission.

In Column 5 of Table 4, we conduct a placebo test where we randomly allocate Southwest entry events to scientist pairs. The coefficient for *'Fake Southwest entry'* is not significant and close to zero, suggesting that it is not the structure of the panel or changes in the data over time that are driving the result. In Column 6 of Table 4, we conduct one more falsification test by looking at entry events (not included in the other regressions) where Southwest withdraws from the market within two years. For these cases, the point estimate of Southwest entry is close to zero and insignificant.

Overall, we believe results in Table 4 and Figure 1 provide robust support for a causal interpretation of our main effect, and reassure us that we are not simply measuring some underlying, unobservable process that takes place with each entry event[26] and drives both Southwest decisions and the increase in scientific collaboration.

[24] We adopt this particular specification because it is the same used by Bernstein, Giroud, and Townsend (2015) in their study of venture capital monitoring and air travel costs, but show robustness to specifications with additional years in Figure 1.

[25] We repeat the same graph within the large sample at the CBSA-pair level in Figure A-1.

[26] Since we observe Southwest arrival across multiple locations and years.

While Southwest is the largest U.S. low-cost carrier in terms of number of passengers transported, there are other low costs airlines operating within the same market. In Appendix Table A-1, we explore how our results vary depending on whether a low-cost airline is already operating on a route, as well as whether they differ when other airlines (low-cost or not) start operating a flight in the same year as Southwest. Consistent with the impact of Southwest on travel costs being largest when no low-cost alternatives existed on the same route, estimates are approximately 12% larger when Southwest is the first low-cost to enter (Table A-1, Column 2),[27] and are positive, but non-significant when another low-cost was already operating between the same airports (Column 3). Results are instead essentially unchanged if we exclude cases where other low-cost airlines enter at the same time (Column 4), other major airlines[28] enter at the same time (Column 5), or any other airline enters at the same time (Column 6). We conclude that our results are robust to considering concurrent entry by other airlines.[29]

Results are also not driven by the fact that our sample includes only pairs that ever collaborate: when we include a random sample of non-collaborating pairs and replace individual-pair fixed effects with university-pair fixed effects[30], we find comparable effects of Southwest entry (see Table A-6). In Appendix Table A-8 we decompose the main effect by pairs of scientists who collaborate both before and after Southwest entry (intensive margin pairs) versus pairs of scientists who collaborate either before or after entry, but not both (extensive margin pairs). We find a stronger effect for intensive margin pairs (Column 3), although the cheaper fares also seem to enable experimentation in the form of new collaborations over distance (Column 2).

[27] Our list of low-cost airlines includes AirTran Airways Corporation, JetBlue, Frontier Airlines, Spirit Air Lines, ATA Airlines, Allegiant Air, Virgin America, Sun Country Airlines, ValuJet Airlines, Vanguard Airlines

[28] We classify as major airlines: Delta, American Airlines, United Airlines, US Airways, Northwest Airlines, Continental, America West Airlines, Alaska Airlines, Trans World Airlines and Envoy Air. These correspond to the 10 companies with the largest numbers of passengers carried between 1993 and 2012.

[29] One might also wonder about additional modes of transportation. As shown in Appendix Table A-2, we find no effect of Southwest entry in the Northeast corridor, where train travel has been a consistent alternative to flying.

[30] If we were to run this regression with individual-pair fixed effects, the non-collaborating pairs would be dropped from the estimation.

In the Appendix, we perform additional robustness to different econometric approaches, functional forms, clustering of standard errors, treatment of outliers and inclusion in the sample of non-collaborating pairs. Briefly, we obtain qualitatively and quantitatively similar results using ordinary least squares instead of Poisson (see Table A-3, Column 2). We also obtain a positive and significant coefficient for Southwest entry (though of a somewhat smaller magnitude) when we run a linear probability model with an indicator variable for any copublication in the focal year as the dependent variable (see Table A-3, Column 3). Clustering at the city-pair level, rather than at the individual-pair level, hardly impacts the standard errors (see Table A-4). The coefficient on Southwest entry remains significant when we exclude pairs that have more than two copublications over the entire observation period, or winsorize observations with more than two copublications (see Table A-5).

4.3 Changes in the Types of Research Projects

Having shown evidence that links Southwest entry to a plausibly causal increase in collaboration between the affected scientists, we now take advantage of this source of exogenous variation to test the theoretical framework.

The first prediction of the model focuses on the impact travel costs have on scientists embedded within better versus worse local research environments. Intuitively, more agglomerated regions with a greater number of potential collaborators offer on average better local matches to begin with, which makes the global scientist pool relatively less appealing. Since it is difficult to build accurate proxies for the number of ideal co-authors a specific scientist may have access to without traveling, we rely on past productivity to assess if a scientist from a given department is more versus less likely to find a good match locally.

As can be seen in Column 1 of Table 5, the increase in collaboration we observe after the arrival of Southwest is driven by scientist pairs where at least one member is more productive than her local peers, and is even more pronounced when both scientists are more productive than their colleagues. Thus, the cheaper fares seem to be particularly helpful for individuals that are talented, but potentially do not have access to co-authors of comparable quality within their local environment. They might be in peripheral institutions because of imperfections in the labor market, or simply because of their geographic preferences. With lower travel costs, these individuals are able to find and sustain better matches over distance.

The next set of predictions of the model link the reduction in geographic frictions to an increase in the amount of time and effort allocated to higher quality (Prediction 2) and riskier projects with distant co-authors (Prediction 3). As discussed in Section 3.2, we proxy for the quality of projects using citations, and for riskiness by looking both at projects that span different sub-fields, as well as research that uses novel keywords. If riskier projects are more likely to fail, then our estimate will likely underestimate the full impact of a reduction in travel costs on this set of more novel and experimental projects, as many will be abandoned and never turn into a publication to begin with.

In terms of riskiness, in Column 2 of Table 5 we see that after Southwest enters, collaborations between different sub-fields increase disproportionately relative to other types of collaborations.[31] These across sub-field projects may benefit more from face-to-face interactions because of a greater need to exchange complex information which may be new to at least one of the participants, and because these pairs cannot rely on a shared, discipline-specific vocabulary to streamline communications over distance.

[Insert Table 5 about here]

In terms of project quality, in Column 1 of Table 6 we condition on collaboration and weight the dependent variable, copublications, by citations received (a proxy for scientific impact and quality). Consistent with Prediction 2 and with the idea that lower travel costs induce scientists to allocate disproportionately more effort to distant collaborations as quality increases, we observe a larger effect of Southwest entry on right tail projects.

In Column 2 of Table 6, we instead test the hypothesis that better matches over distance could be driven not only by complementarities in ideas and skills (as we have seen

[31] The result from the first two columns of Table 5 is unchanged when we control for both interactions in Column 3. In Appendix Table A-7 we also present a larger set of interactions beyond those predicted by the theory. We use university-pair fixed effects instead of individual fixed effects so that we can display and interpret the main effect of the moderators. The estimates show that researchers in more distant disciplines or further away from each other are less likely to collaborate, but that Southwest entry helps compensate for this greater distance in knowledge space and geography. Other moderators such as age, relative productivity in the department, and department R&D budgets are non-significant.

in the across sub-field collaborations), but also by complementarities in equipment and infrastructure. The dependent variable in Column 2 is the number of equipment-intensive copublications. Although these publications are more rare, the effect of Southwest entry is large and significant, suggesting that at least within chemistry equipment may play a key role in how scientists select into distant collaborations.

Last, in Column 3 we directly look for novelty in the paper keywords as an additional proxy for project riskiness, finding further support for Prediction 3. In the next section we perform robustness on all of the model's predictions within a larger dataset of chemistry and chemistry-related journals, as well as in a different fields of science (mathematics).

[Insert Table 6 about here]

5 Extensions and External Validity in Different Samples

The analysis and results presented in the last section describe the effect of Southwest entry on the rate and type of collaborations between chemistry faculty members. While this approach has the advantage of leveraging rich individual-level data and offers a cleaner identification strategy, one may also be interested in studying the same effects within chemistry more generally, as well as to test their validity in other fields.

In this section, we present three additional empirical exercises, starting from an analysis of the effect of Southwest entry in different fields of science. The regressions are conducted at the region-pair level since we do not have individual level data for biology, physics and engineering. Results show that the effect we have identified within chemistry also applies to other fields.

Next, we expand our chemistry analysis to include all publications in chemistry and chemistry-related journals at the region-pair level. This allows us to test all the hypothesis of our model within a larger set of collaborations, which is particularly helpful for studying more rare types of projects (e.g. equipment-intensive, very high quality, novel). Consistent with the predictions of the model, we see a disproportionate increase in novel, cross-disciplinary and equipment-intensive copublications.

Last, we perform an additional, deep-dive within an additional domain for which we have collected individual-level data: mathematics. Although we find an effect of Southwest entry, the magnitude is lower, possibly because mathematics does not rely on specialized equipment in the same way as chemistry – and equipment-intensive collaborations were among the most affected in that field – or because collaborations over distance are less useful to begin with since mathematics relies on a more explicit, codified language. As in chemistry, effects are stronger for possibly riskier collaborations between pairs of scientists that are active in different sub-fields of mathematics.

5.1 Aggregate Changes in the Rate of Collaboration in Different Fields Across U.S. Regions

To test if the availability of cheaper flights had an effect on scientific collaboration across fields, we use a large-scale publication dataset covering close to a million papers matched to U.S. regions (defined in terms of CBSAs – Core-Based Statistical Areas).[32] Specifically, we explore how collaboration between any two CBSAs changed after Southwest starts operating a new route between them. The unit of analysis is the CBSA-pair-year (48,274 pairs), and we include CBSA-pair fixed effects and year fixed effects to respectively control for underlying differences across regions that are consistent over time, and overall time trend.[33] The regressions also include linear time trends for the origin and destination CBSA. For the estimation, we use a Poisson model with standard errors clustered at the CBSA-pair level.

[32] The starting point for the construction of this sample is the population of scientific articles published in the top 477 scientific journals in biology, chemistry, physics and engineering between 1991 and 2012. We have a total of 2,773,560 papers, of which 1,169,458 had at least one author with a U.S. address. Out of all papers with U.S. addresses, we were able to successfully map 994,672 (85%) to a U.S. CBSA using a combination of three different geocoding services (Google Maps API, Bing Maps API, and the Data Science Toolkit). This allows us to link the vast majority of U.S. papers to the geographic regions involved in their production.

[33] While this approach has the advantage of considering different fields of science, it also has important limitations. We can no longer include scientist-pair fixed effects and account for idiosyncratic, unobservable, and time invariant reasons that may drive collaboration between any two scientists. Core-based statistical areas (CBSAs) may also be too large as a unit of analysis for correctly measuring the effects of interest. Finally, our ability to test the full set of predictions of the model is limited.

[Insert Table 7 about here]

Results are displayed in Table 7. In Column 1 we find that Southwest entry is associated with an increase in collaborations of 65% in the full sample, with comparable effects across the different fields of science: 17% increase in chemistry (Column 2), 64% in biology (Column 3), 15% in physics (Column 4), and 27% in engineering (Column 5). Since chemistry, according to the estimates in Table 7, is not an outlier in either direction (if anything, the effects within the field are on the lower end of the spectrum), we believe the main results of the paper are generalizable to other fields.

5.2 Aggregate Changes in the Type of Chemistry Collaborations Across U.S. Regions

Next, we study how Southwest entry changes the rate and types of chemistry collaborations at the regional level. The analysis of collaborations between chemistry faculty members already suggested that lower travel costs are particularly beneficial for higher-quality, riskier, and equipment-intensive projects. However, the estimates are significant, but noisy when looking at rare outcomes such as novel copublications. We therefore replicate our analysis within a broader set of papers in chemistry and related fields. As in the previous subsection, the analysis is at the CBSA-pair-year level and includes CBSA-pair fixed effects and year fixed effects.

[Insert Table 8 about here]

The effects (see Table 8) are consistent with our previous findings, and highlight that lower travel cost have a disproportionate effect on the right tail of the quality distribution, and on more novel and cross-disciplinary ideas. The impact of these changes is large, with increases in aggregate output of 21% in terms of copublication counts, 24% for novel ideas, and 26% for projects that span different fields or that are equipment-intensive. These correspond to roughly 300 extra copublications per year.[34]

[34] The sample mean for copublications at the CBSA-pair-year level is around 2.1, leading to an increase of 0.42 per CBSA-pair-year. We have around 2,100 pairs per year, of which around one third are in treatment status. So a back of envelope estimate is 0.42*2100*0.33=294.

5.3 Changes Within Mathematics

Last, we explore the effect of Southwest entry within mathematicians. The dataset includes all faculty members in U.S. mathematics departments that have advised at least one PhD student.[35] We observe 431 pairs of individuals that experienced Southwest entry between 1993 and 2012 and have at least one copublication in that period. We adopt the same empirical strategy as in the chemistry sample and regress copublications on an indicator variable for Southwest entry, controlling for individual-pair fixed effects and year fixed effects.

[Insert Table 9 about here]

Results show that Southwest entry increases copublications by 25%. While the point estimate is significant, the magnitude is smaller than the one we find for collaborations between chemists. Consistent with the chemistry sample, effects are stronger for mathematicians who are more productive than their peers, and for pairs of scientists working in different sub-fields of mathematics.

6 Conclusions

The paper explores how geographic frictions, and in particular travel costs, shape the rate and direction of scientific research. Our theoretical framework builds on a tension between lower collaboration costs when co-located, and the availability of a broader set of potential collaborators over distance. We start from this basic trade-off and then explore some of the key choices scientists face when deciding if they should collaborate locally versus over distance, how much effort to allocate to projects of different potential, and how much risk to take. Whereas previous work has mostly focused on communication costs and their impact on the rate of collaboration, our paper emphasizes other effects distance-related frictions can have on innovative outcomes even before projects are started.

[35] This database is based on MathSciNet, an abstracting service run by the American Mathematical Project and the Mathematics Geneaology Project, which is targeted at tracking PhD theses in mathematics. We construct a sample of US-based mathematicians who advise at least one PhD student, and deduce their location from the institution their students graduate from.

We test the predictions from the model by taking advantage of a source of plausibly exogenous variation in travel costs: the differential timing of entry by a low-cost carrier across multiple U.S. airports. Our difference-in-differences empirical strategy, combined with a series of robustness and falsification tests, supports the idea that the availability of lower fares had a causal effect on the probability and intensity of collaboration between scientists.[36] The effect is particularly pronounced for scientists that are less likely to find co-authors of the same quality within their local environment, is present across multiple fields of science (chemistry, physics, biology, engineering, mathematics), and is robust to controlling for idiosyncratic scientist-pair characteristics, trends in collaboration over time, and department R&D budgets. Moreover, we do not observe a pre-trend in collaboration between scientist pairs that are going to experience lower air travel costs in the future.

Consistent with the theory, the reduction in geographic frictions also transforms the types of projects that emerge: our estimates suggest a sizable increase in higher quality papers, in projects that span different sub-disciplines, are more intensive in their use of specialized equipment, and are more risky and novel. Comparisons between our findings in chemistry and mathematics suggest that complementarities in specialized equipment – while important for collaboration decisions between distant labs – are not the only driver behind the observed increase in joint projects over distance. Scientists also launch more experimental projects and projects that seem to take advantage of the complementary skills, ideas and knowledge that a distant lab may contribute to a collaboration.

Beyond the lower fares introduced by the low-cost airline we study in the paper, the cost per mile in the United States has dropped by over 50% in the last 30 years (Perry,

[36]A back of the envelope calculation suggests that Southwest entry induced close to 400 copublications among chemistry faculty pairs. The sample mean in our sample is 0.1 copublications per year increasing by 50% to 0.15 copublications per year after Southwest entry. We have 750 pairs and 10 post-entry years on average, leading to a back of envelope estimate of 0.05 * 750 * 10=375 copublications. While this number is sizeable, it is small relative to the total number of copublications among chemistry faculty members in this period. However, Southwest entry corresponds to a 20% price reduction affecting only a fraction of faculty pairs (a large fraction of pairs are served by Southwest or other low-cost carriers before our observation period). Over the last 30 years, the cost per mile for air travel across all routes within the U.S. dropped by 50% (Perry 2014). This suggests that reductions in air transportation overall could have had a substantial aggregate effect on collaboration, above and beyond the particular source of variation in air travel cost we use in this paper.

2014),[37] and convenience and routes have greatly improved. Our results should be therefore interpreted within this broader context of improvements in our ability to travel and work with distant co-authors. Whereas we cleanly estimate the impact of only part of these changes, improvements in air travel are likely to affect a much larger population of individuals. This includes inventors and researchers working within firms that have multiple sites, which will face a similar trade-off between the possibility to form the ideal team when not constraining the search for members to one site, and the additional communication, coordination and travel costs this may add.[38] Further exploring the trade-offs geographic frictions introduce in these different contexts is a fruitful area for subsequent research.

[37] We would expect similar effects in Europe, where low-cost airlines had even more of an effect on market structure and competition, as well as on uniting different economies.

[38] There is an interesting parallel here with the literature on communications costs and collaboration: while Agrawal & Goldfarb (2008) focused on academic collaborations, Forman & Zeebroeck (2012) subsequently found that the internet fostered R&D collaborations within firms. In principle, one could make progress on this related question using patenting and co-invention data together with our empirical strategy.

References

Adams, J. D., Black, G. C., Clemmons, J. R., & Stephan, P. E. (2005). Scientific teams and institutional collaborations: Evidence from US universities, 1981-1999. Research policy, 34(3), 259-285.

Agrawal, A., & Goldfarb, A. (2008). Restructuring Research: Communication Costs and the Democratization of University Innovation. American Economic Review, 98(4), 1578-90.

Agrawal, A., Catalini, C., & Goldfarb, A. (2015). Crowdfunding: Geography, social networks, and the timing of investment decisions. Journal of Economics & Management Strategy, 24(2), 253-274.

Bernstein, S., Giroud, X., & Townsend, R. R. (2015). The Impact of Venture Capital Monitoring. The Journal of Finance.

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

Borenstein, S., & Rose, N. L. (2007). How airline markets work... or do they? Regulatory reform in the airline industry. No. w13452. National Bureau of Economic Research.

Boudreau, K. J., Brady, T., Ganguli, I., Gaule, P., Guinan, E., Hollenberg, T., & Lakhani, K. (2017). A Field Experiment on Search Costs and the Formation of Scientific Collaborations. Review of Economics and Statistics. 99(4):565-576

Cairncross, F. (1997). The death of distance: How the communications revolution will change our lives. Boston: Harvard Business School Press

Campos, de Leon & Mcquilin (2018). Lost in the Storm: The Academic Collaborations that Went Missing in Hurricane Issac. Economic Journal, 128(610):995-1018.

Catalini, C. (2017). Microgeography and the direction of inventive activity. Management Science, 10.1287/mnsc.2017.2798

Chai, S. & Freeman, R. (2018) Knowledge Spillover and Collaboration through Temporary Colocation. Mimeo, ESSEC Business School

Ding, W. W., Levin, S. G., Stephan, P. E., & Winkler, A. E. (2010). The impact of information technology on academic scientists' productivity and collaboration patterns.

Management Science, 56(9), 1439-1461.

Diospatonyi, I., Horvai, G., Braun, T. (2001). Publication Speed in Analytical Chemistry Journals. Journal of Chemical Information and Computer Sciences, 41, 1452-1456.

Forman, C., Goldfarb, A., & Greenstein, S. (2005). How did location affect adoption of the commercial Internet? Global village vs. urban leadership. Journal of urban Economics, 58(3), 389-420.

Forman, C., & Zeebroeck, N. V. (2012). From wires to partners: How the Internet has fostered R&D collaborations within firms. Management science, 58(8), 1549-1568.

Friedman, T. L. (2005). The world is flat: A brief history of the twenty-first century. New York, NY: Farrar, Straus and Giroux Publishers

Freeman, R. B., Ganguli, I., & Murciano-Goroff, R. (2014). Why and wherefore of increased scientific collaboration (No. w19819). National Bureau of Economic Research.

Gaspar, J., & Glaeser, E. L. (1998). Information technology and the future of cities. Journal of Urban Economics, 43(1), 136-156.

Glaeser, E. L., & Ponzetto, G. A. (2007). Did the death of distance hurt Detroit and help New York? (No. w13710). National Bureau of Economic Research.

Goolsbee, A., & Syverson, C. (2008). How Do Incumbents Respond to the Threat of Entry? Evidence from the Major Airlines. The Quarterly Journal of Economics, 123(4), 1611-1633.

Guimaraes, P., & Portugal, P. (2010). A simple feasible procedure to fit models with high-dimensional fixed effects. Stata Journal, 10(4), 628.

Hamermesh, D., & Oster, S. (2002) "Tools or Toys? The Impact of High Technology on Scholarly Productivity" Economic Inquiry 40(4):539-555

Jacobs, J. (1969). The Life of Cities. New York, NY: Random House.

Jones, B. F. (2009). The burden of knowledge and the death of the renaissance man: Is innovation getting harder? The Review of Economic Studies, 76(1), 283-317.

Jones, B.F., Wuchty, S., Uzzi, B. (2008). Multi-University Research Teams: Shifting Impact, Geography and Stratification in Science. Science 322(5905), 1259-1262

Katz, J. (1994). Geographical proximity and scientific collaboration. Scientometrics, 31(1), 31-43.

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

Leamer, E. E., & Levinsohn, J. (1995). International trade theory: the evidence. Handbook of International Economics, 3, 1339-1394.

Mairesse, J., & Turner, L. (2005). Measurement and explanation of the intensity of copublication in scientific research: An analysis at the laboratory level. NBER Working Paper No 11172. National Bureau of Economic Research.

Morrison, S. A. (2001). Actual, adjacent, and potential competition: Estimating the full effect of Southwest Airlines. Journal of Transport Economics and Policy, 239-256.

Perry, M. (2014) "The cost of air travel in the US has been remarkably stable for the last decade, and 17% cheaper than 20 years ago" Carpe Diem, Blog of American Entreprise Institute. https://www.aei.org/publication/the-cost-of-air-travel-in-the-us-has-been-remarkably-stable-for-the-last-decade-and-17-cheaper-than-20-years-ago/ accessed March 17, 2016

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

Rosenblatt T. & Mobius M. (2004) Getting closer or drifting apart? Quarterly Journal of Economics 119(3):971-1009

Rosenthal, S. S., & Strange, W. C. (2001). The determinants of agglomeration. Journal of Urban Economics, 50(2):191-229.

Stephan, P. E. (2012). How Economics Shapes Science. Cambridge, MA: Harvard University Press.

Storper, M., & Venables, A. J. (2004). Buzz: face-to-face contact and the urban economy. Journal of Economic Geography, 4(4), 351-370.

Transportation Research Board, National Research Council (1999): Entry and Competition in the US Airline Industry: Issues and Opportunities. Washington, DC: National

Academy Press

Von Hippel, E. (1994). "Sticky information and the locus of problem solving: implications for innovation. Management science, 40(4), 429-439.

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

Tables

Table 1: **Summary Statistics**

Variable	Obs	Mean	Std. Dev.
Individual Scientist Level			
Age	845	49.6	11.0
Female	845	.10	.30
Average R&D budget in dept. (1000s USD)	845	279.88	226.75
Speciality:			
Physical chemistry	845	.32	.47
Biochemistry	845	.22	.41
Inorganic chemistry	845	.14	.34
Organic chemistry	845	.13	.34
Material science	845	.11	.31
Other	845	.08	.27
Individual-Pair Level			
Year of SW entry	758	2001	4.5
Distance (in miles)	758	1232	808.6
Years in sample	758	17.3	4.6
Total copublications	758	1.9	3.4
Copub. both before and after	758	.09	.28
Copub. before SW entry	758	.49	.50
Copub. after SW entry	758	.60	.49
In different field of chemistry	758	.45	.50
Individual-Pair-Year :evel			
Copublications	13,147	.11	.41
Dummy for any copublication	13,147	.09	.29
Individual-pair-year level conditional on copublication			
Cites weighted copublications	1,177	44.9	71.8
Novel copublications	1,177	.09	.32
Equipment intensive copublications	1,177	.27	.53

Table 2: **Comparing Pairs in the Analysis Sample to Pairs Not Experiencing Southwest**

	Distant Pairs Not Experiencing SW	Analysis Sample	P-Value For Equality of Means
Total copublications	1.78	1.90	0.19
Number of years observed	13.85	17.51	<0.01
Age (average in pair)	49.16	51.23	<0.01
Different type of chemistry	0.46	0.45	0.79
Average R&D budget in dept.	288.9	278.1	0.17
Observations	5,954	758	

Table 3: **Effects of Southwest Entry on Price, Passengers and Routes**

	(1) Passengers (log)	(2) Mean Price (log)	(3) Average Miles Flown (log)	(4) Direct Flight	(5) Number of Transfers
Southwest Entry	0.4437***	-0.1910***	0.0007	0.0002	-0.0174***
	(0.0050)	(0.0024)	(0.0006)	(0.0004)	(0.0017)
Airport-Pair Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Mean of Dep. Variable	4.238	5.454	7.066	0.007	1.239
Number of Pairs	55750	55750	55739	55750	55750
Number of Observations	956029	956029	955983	956029	956029

Notes: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, Southwest entry is an indicator variable that takes value 1 if Southwest has started operating a flight between airports. All specifications include airport-pair fixed effects and year fixed effects. Estimation by ordinary least squares.

Table 4: **Effect of Southwest Entry on Copublications at the Individual-Pair level**

DV=Copublications	(1) Baseline	(2) Controls	(3) Controls	(4) Timing	(5) Placebo 1	(6) Placebo 2
Southwest Entry	0.505***	0.526***	0.526***			-0.029
	(0.121)	(0.121)	(0.121)			(0.216)
Mean Age		0.153***	0.268***			
		(0.008)	(0.015)			
Dept R&D Budget per Faculty (log)		0.364***	0.364***			
		(0.127)	(0.127)			
Years Since Both Have a PhD			-0.230***			
			(0.022)			
Southwest Entry (-1)				0.078		
				(0.152)		
Southwest Entry (0)				0.485***		
				(0.150)		
Southwest Entry (1)				0.518***		
				(0.166)		
Southwest Entry (2)				0.582***		
				(0.181)		
Fake Southwest Entry (Random Timing)					0.095	
					(0.121)	
Individual Pair FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Number of Pairs	758	758	758	758	758	171
Number of Obs.	13,147	13,147	13,147	13,147	13,147	2,945

Notes: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the number of copublications between pairs of scientists. Southwest entry is an indicator variable that takes value 1 if Southwest has started operating a flight from airports close to the respective scientists. All specifications include individual-pair fixed effects and year fixed effects. Column 1 is our baseline specification. Column 2 adds controls for the age of the pair members and departmental R&D budget per faculty (both variables are means across the two pairs members). Column 3 additionally controls for the numbers of years that have passed since both pairs members obtained their PhD. Column 4 replaces Southwest entry by a set of indicator variables corresponding to different times from or since entry: SW entry (-1) is an indicator variable if the observation is in the year preceding SW entry; SW entry (0) SW entry (1) SW entry (2+) are defined analogously for the year of SW entry, the year after SW entry, and two years or more after SW entry. Column 5 is a placebo where we pretend Southwest entry has occurred in a random year for each pair. Column 6 is a placebo where we look at the set of pairs (not included in the baseline specification) who experienced Southwest entry followed by a Southwest exit event shortly thereafter. Estimation by Poisson Quasi-Maximum Likelihood.

Table 5: **Effect of Southwest Entry on Copublications: Which Pairs Are Most Affected?**

DV=Copublications	(1)	(2)	(3)
Southwest Entry	0.119	0.272*	-0.075
	(0.166)	(0.143)	(0.197)
SW Entry X One More Productive than department average	0.346*		0.314*
	(0.187)		(0.180)
SW Entry X Both More Productive than department average	0.695***		0.662***
	(0.249)		(0.237)
Southwest X Different Type of Chemistry		0.543**	0.513**
		(0.233)	(0.216)
Pair Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
Number of Pairs	758	758	758
Number of Observations	13,147	13,147	13,147

Notes: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: **Effect of Southwest Entry on the Type of Collaborations**

	(1) Cites Received	(2) Equipment-Intensive Copublications	(3) Novel Copublications
Southwest Entry	0.420*	0.024***	1.175*
	(0.234)	(0.280)	(0.687)
Pair Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
Number of Pairs	189	80	37
Number of Observations	606	276	137

Notes: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. These regressions are conditional on the pair having collaborating in the focal year. The dependent variables are the number of cites received (column 1), the number of equipment-intensive collaborations (column 2), and the number of novel copublications (column 3). Pairs that never have a non-zero value of the dependent variable are dropped from the regressions, hence the lower number of observations in columns 2 and 3. All specifications include individual-pair fixed effects and year fixed effects. Estimation by Poisson Quasi-Maximum Likelihood.

Table 7: **Southwest Entry and Collaborations Between U.S. Cities (CBSAs)**

	(1) All copubs	(2) Chemistry	(3) Biology	(4) Physics	(5) Engineering
Southwest Entry	0.503***	0.159***	0.494***	0.141***	0.238***
	(0.020)	(0.033)	(0.032)	(0.031)	(0.055)
CBSA Pair Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
City Trends	Yes	Yes	Yes	Yes	Yes
Number of Pairs	48,274	15,303	22,079	15,872	7,635
Number of Observations	965,480	306,060	441,580	317,440	152,700

Notes: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ These regressions are run at the CBSA-pair level. The dependent variable is the number of copublications between pairs of CBSAs. Southwest entry is an indicator variable that takes value 1 if Southwest has started operating a flight from airports close to the respective cities. Column 1 is based on copublications in all journals in our sample. Columns 2, 3, 4, 5, are based on chemistry, biology, physics and engineering journals respectively. All specifications include CBSA-pair fixed effects, year fixed effects, an origin-CBSA time trend and a destination-CBSA time trend. Estimation by Poisson Quasi-Maximum Likelihood.

Table 8: **Southwest Entry and Types of Collaboration Across CBSAs**

	(1) Copublications	(2) Citation-Weighted	(3) Novel	(4) Across Fields	(5) Equipment
Southwest entry	0.209***	0.240**	0.235***	0.266**	0.264***
	(0.065)	(0.120)	(0.071)	(0.116)	(0.087)
Year FE	Yes	Yes	Yes	Yes	Yes
CBSA-Pair FE	Yes	Yes	Yes	Yes	Yes
Observations	40,227	39,517	34,847	27,379	29,767
CBSA Pairs	6,172	5,970	4,941	3,300	3,789

Notes: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. These regressions are run at the CBSA-pair level and are based on a large sample of publications in chemistry and chemistry-related fields. The dependent variables in column 1 is the number of copublications between pairs of CBSAs. Column 2 uses cites-weighted copublications as dependent variable. Column 3, 4 and 5 count the number of novel, across-fields and equipment intensive collaborations, as inferred from the keywords associated with the papers.

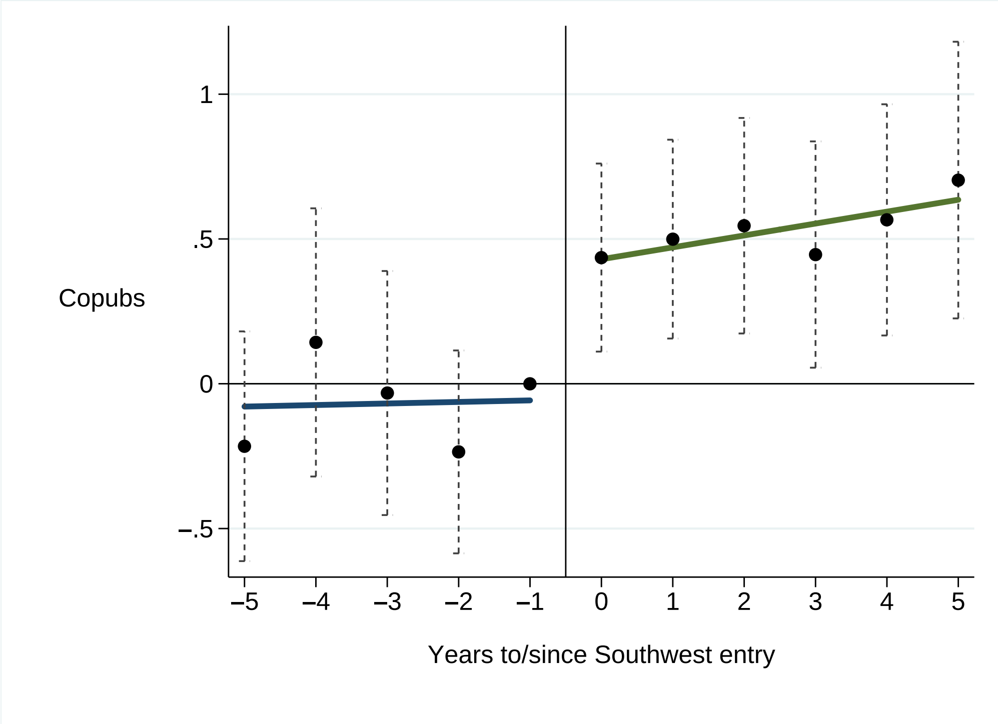
Table 9: **Effect of Southwest Entry on Collaboration Among Mathematicians**

DV=Copublications	(1)	(2)	(3)	(4)
Southwest Entry	0.247**	0.123	-0.370	-0.426
	(0.123)	(0.131)	(0.309)	(0.310)
SW Entry X Different Types of Mathematics		0.469***		0.471***
		(0.177)		(0.181)
SW Entry X One More Productive than Dept Average			0.555*	0.435
			(0.333)	(0.336)
SW Entry X Both More Productive than Dept Average			0.728**	0.673**
			(0.315)	(0.316)
Pair Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Number of Pairs	431	431	431	431
Number of Observations	5,514	5,514	5,514	5,514

Notes: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. These regressions are based on a dataset of U.S. mathematicians constructed using MathSciNet and the Mathematics Geneaology Project. The empirical specifications are analogous to the ones we use when studying collaborations among chemistry faculty members.

Figures

Figure 1: **Dynamics of the Effect of Southwest Entry: Individual-Pair Level**



Notes: To generate this graph, we regress individual copublications on year fixed effects, pair effects and a set of indicator variables corresponding to 5 years before SW entry, 4 years before SW entry, ..., 4 years after SW entry, 5 years after SW entry (1 year before SW entry is omitted). We then plot the coefficients associated with these indicator variables against time from/to Southwest entry, superimposing a linear fit line before entry and after entry. The vertical bars represent 95% confidence intervals. The coefficient for the year immediately before entry is set to zero and displayed without confidence interval since it our baseline year.

Online Appendix

Appendix A: Tables

Table A-1: **Presence and Concurrent Entry of Other Airlines**

DV=Copubs	(1) Baseline	(2) No LC before SW	(3) LC before SW	(4) excl. entry by other LC	(5) excl. entry by MC	(6) excl. entry by LC and MC
Southwest Entry	0.505***	0.625***	0.291	0.498***	0.477***	0.441***
	(0.121)	(0.157)	(0.199)	(0.131)	(0.145)	(0.157)
Pair Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of pairs	758	479	279	649	556	480
Number of obs.	13,147	8,349	4,798	11,261	9,711	8,367

Notes: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Column 1 is the baseline regression. Column 2 limits the sample to pairs in locations where no other low-cost airline was operating in the year before Southwest entry. Conversely, column 3 limits the sample to pairs in locations where another low-cost airline was operating in the year before Southwest entry. Column 4 exclude cases where another low cost company entered in the same year as Southwest, column 5 excludes where another major company entered in the same year as Southwest and column 6 excludes both of these groups.

Table A-2: **Northeast Corridor Falsification Test**

DV= Copublications	(1) All	(2) NE corridor only	(3) excluding NE corridor
Southwest Entry	0.505***	-0.745	0.559***
	(0.121)	(0.597)	(0.123)
Pair Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
Number of pairs	758	31	727
Number of observations	13,147	564	12,583

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A-3: **Alternative Functional Forms**

	(1) Poisson Copublications	(2) OLS Copublications	(3) OLS Any copublication
Southwest Entry	0.505***	0.052***	0.030***
	(0.121)	(0.015)	(0.010)
Pair Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
Number of pairs	758	758	758
Number of obs.	13,147	13,147	13,147

Notes: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Southwest entry is an indicator variable that takes value 1 if Southwest has started operating a flight from airports close to the respective scientists. Column 1 is the baseline regression at the individual pair level (estimated by Poisson Quasi-Maximum Likelihood.) Column 2 estimates the same specification with ordinary least squares. Column 3 is a linear probability model with an indicator variable for any copublication.

Table A-4: **Inference with City-Pair Clustering**

DV=Copublications	(1)	(2)	(3)
Southwest Entry	0.505***	0.505***	0.492***
	(0.121)	(0.100)	(0.134)
Pair Fixed Effects	Individual pair	Individual pair	City Pair
Year Fixed Effects	Yes	Yes	Yes
Clustering	Individual pair	City pair	City pair
Number of obs.	13,147	13,147	13,147

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the number of copublications between pairs of scientists. Southwest entry is an indicator variable that takes value 1 if Southwest has started operating a flight from airports close to the respective scientists. Column 1 is our baseline regressions with individual pair fixed effects and individual pair clustering. In column 2, we keep individual pair fixed effects but cluster at the city pair level, using the POI2HDFE Stata command from Paulo Guimaraes that implements the algorithm Guimaraes & Portugal (2010). In column 3, we replace individual pair fixed effects with city pair fixed effects and cluster by city pair fixed effects

Table A-5: **Sensitivity to Outliers**

DV=Copublications	(1)	(2)	(3)
Southwest Entry	0.505***	0.308***	0.424***
	(0.121)	(0.118)	(0.112)
Pair Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
Comment	Baseline	Excluding outlier pairs	Winsorizing outliers
Number of pairs	758	732	758
Number of obs.	13,147	12,701	13,147

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, The dependent variable is the number of copublications between pairs of scientists. Southwest entry is an indicator variable that takes value 1 if Southwest has started operating a flight from airports close to the respective scientists. Column 1 is our baseline regressions with individual pair fixed effects and individual pair clustering. In column 2, exclude pairs that have more than two copublications in any given years. In column 3, we winsorize observations that have more than two copublications to two.

Table A-6: **Robustness to Including Non-Collaborating Pairs**

DV=copublications	(1)	(2)	(3)
Southwest Entry	0.505***	0.500***	0.337***
	(0.121)	(0.129)	(0.124)
Pair Fixed Effects	Individual pair	University Pair	University Pair
Year Fixed Effects	Yes	Yes	Yes
Sample includes non-collaborating pairs	No	No	Yes
Number of obs.	13,147	13,147	1,425,523

Notes: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Southwest entry is an indicator variable that takes value 1 if Southwest has started operating a flight from airports close to the respective scientists. Estimation by Poisson Quasi-Maximum Likelihood. Column 1 is our baseline regression, which includes only pairs of scientists who collaborates at some point. Column 2 keeps the same sample but replaces individual pair fixed effect by university pair fixed effects. Column 3 adds a 10% random sample of non-collaboration pairs to the sample of collaborating pairs and is run with university pair fixed effects.

Table A-7: **Heterogeneous Effects of Southwest Entry**

DV=Copublications	(1)	(2)	(3)	(4)	(5)	(6)
Southwest Entry	0.024	-0.006	-0.016	0.174	0.555***	-0.186
	(0.088)	(0.144)	(0.093)	(0.113)	(0.147)	(0.203)
Both below 50	0.154					0.151
(at time of entry)	(0.117)					(0.119)
SW X both below 50	0.465***					0.528***
(at time of entry)	(0.149)					(0.152)
One more productive		-0.059				-0.052
than dept average		(0.131)				(0.130)
Both more productive		0.045				0.069
than dept average		(0.141)				(0.139)
SW X one more productive		0.240				0.266
than dept average		(0.167)				(0.166)
SW X both more productive		0.441**				0.490***
than dept average		(0.187)				(0.186)
Different type			-0.362***			-0.400***
of chemistry			(0.097)			(0.099)
Southwest X different			0.581***			0.588***
type of chemistry			(0.131)			(0.131)
Both in dept.				0.186		0.230**
with below median R&D budget				(0.115)		(0.115)
One in dept.				0.101		0.138
with below median R&D budget				(0.121)		(0.121)
SW X Both in dept.				0.115		0.067
with below median R&D budget				(0.157)		(0.162)
SW X One in dept.				-0.161		-0.278*
with below median R&D budget				(0.159)		(0.160)
Distance less than					0.155	0.140
1000 miles					(0.123)	(0.123)
SW X distance less than					-0.377**	-0.420**
1000 miles					(0.166)	(0.167)
Distance between 1000					0.091	0.066
and 2000 miles					(0.138)	(0.141)
SW X distance between 1000					-0.474**	-0.575***
and 2000 miles					(0.186)	(0.197)
University Pair Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of pairs	758	758	758	758	758	758
Number of observations	13,147	13,147	13,147	13,147	13,147	13,147

Notes: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Distance above 2000 miles is the omitted distance category.

Table A-8: **Intensive and Extensive Margin of Southwest Entry**

DV= Copublications	(1) Baseline	(2) Extensive margin	(3) Intensive margin
Southwest entry	0.505***	0.410***	0.806***
	(0.121)	(0.135)	(0.234)
Pair Fixed Effects	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes
Number of pairs	758	692	66
Number of observations	13,147	11,969	1,178

Notes: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the number of copublications between pairs of scientists. Southwest entry is an indicator variable that takes value 1 if Southwest has started operating a flight from airports close to the respective scientists. Column 1 is the baseline specification. Column 2 restricts the sample to pairs of scientists who collaborate either before or after entry, but not both; column 3 restrict the sample to pairs of scientists who collaborate both before and after entry. All specifications include individual-pair fixed effects and year fixed effects. Estimation by Poisson Quasi-Maximum Likelihood.

Table A-9: **Changing the Definition of Proximate Airport**

	(1) 10 miles	(2) 25 miles	(3) 50 miles	(4) 100 miles
Southwest entry	0.896***	0.583***	0.505***	0.323***
	(0.206)	(0.116)	(0.095)	(0.073)
Pair Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Number of pairs	150	433	758	1,127
Number of obs.	2,600	7,275	13,147	19,292

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

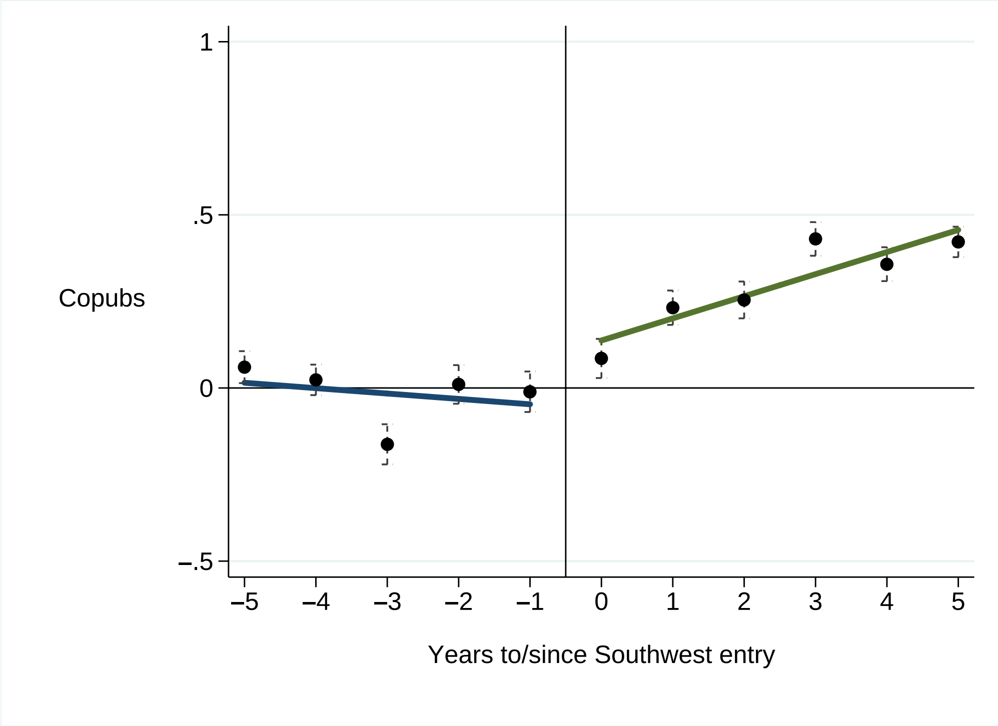
Table A-10: **Using Journals to Define Field of Specialization**

Field	Journal (examples)
Biochemistry	*Journal of Biological Chemistry, Biochemistry*
Inorganic Chemistry	*Inorganic Chemistry*
Material Science	*Macromolecules, Advanced Materials*
Physical Chemistry	*Journal of Physical Chemistry*
Organic Chemistry	*Journal of Organic Chemistry, Organic Letters*

The area of specialization for a given faculty member is inferred from the journals s/he publishes in. For instance, a faculty member who publishes often in the *Journal of Biological Chemistry* is assumed to be specialized in biochemistry.

Appendix A: Figures

Figure A-1: **Dynamics of The Effect of Southwest Entry: CBSA-Pair Level**



Notes: To generate this graph, we regress CBSA copublications on year fixed effects, pair effects, origin-CBSA and destination-CBSA time trends, and a set of indicator variables corresponding to 5 years before SW entry, 4 years before SW entry, ..., 4 years after SW entry, 5 years after SW entry (5 years before SW entry omitted). We then plot the coefficients associated with these indicator variables against time from/to Southwest entry, superimposing a linear fit line before entry and after entry.

Appendix B: Theoretical Model Under w=0

B1: $\pi^*_D > \pi^*_{G,L}$

Intuitively, if scientists go on the global market for co-authors and end up with someone of the same quality as a local co-author (the outside option), then profits will be lower because they additionally have to pay travel costs:

$$\frac{(p_L q)^2}{2\alpha} > \frac{(p_L q)^2}{4\alpha}(1 + \frac{1}{2}t^*_L) > 0$$

$$\iff \frac{1}{2} > \frac{1}{4} + \frac{1}{8}t^*_L > 0, \tag{15}$$

since the maximum value of t^*_L is 1. Q.E.D.

B2: $\pi^*_{G,H} > \pi^*_D$ requires an assumption on parameters

$$\pi^*_{G,H} - \pi^*_D > 0 \tag{16}$$

$$\iff \frac{(p_H q)^2}{4\alpha}(1 + \frac{1}{2}t^*_H) - \frac{(p_L q)^2}{2\alpha} > 0$$

$$\iff (p_H)^2(1 + \frac{1}{2}t^*_H) > 2(p_L)^2 \tag{17}$$

Intuitively, the benefit of a better co-author has to compensate for higher travel costs.

B3: $\pi_G^* > \pi_D^*$ requires an assumption on parameters

$$(1-z)\pi_{G,L}^* + z\pi_{G,H}^* - \pi_D^* > 0 \tag{18}$$

$$\iff (1-z)\frac{(p_L q)^2}{4\alpha}(1+\frac{1}{2}t_L^*) + z\frac{(p_H q)^2}{4\alpha}(1+\frac{1}{2}t_H^*) - \frac{(p_L q)^2}{2\alpha} > 0$$

$$\iff (1-z)(p_L)^2(1+\frac{1}{2}t_L^*) + z(p_H)^2(1+\frac{1}{2}t_H^*) - 2(p_L)^2 > 0$$

$$\iff z[(p_H)^2(1+\frac{1}{2}t_H^*) - (p_L)^2(1+\frac{1}{2}t_L^*)] - (p_L)^2(1-\frac{1}{2}t_L^*) > 0 \tag{19}$$

Intuitively, the benefit of a better co-author has to compensate for higher travel costs. Conditional on this being the case, the more likely this will happen (i.e. z high), the more attractive the choice of a distant co-author becomes.

B4: $\frac{\partial[\pi_G^* - \pi_D^*]}{\partial q} > 0$ requires an assumption on parameters

First, $\frac{\partial[\pi_{G,H}^*]}{\partial q}$:

$$\begin{aligned}\frac{\partial \pi_{G,H}^*}{\partial q} &= \frac{(p_H)^2 q}{2\alpha}(1+\frac{1}{2}t_H^*) + \frac{(p_H q)^2}{4\alpha}\frac{1}{2}\frac{(p_H)^2 q}{4\alpha\beta} \\ &= \frac{(p_H)^2 q}{2\alpha}(1+t_H^*)\end{aligned} \tag{20}$$

Second, $\frac{\partial[\pi_{G,L}^*]}{\partial q}$:

$$\frac{\partial \pi_{G,L}^*}{\partial q} = \frac{(p_L)^2 q}{2\alpha}(1+t_L^*) \tag{21}$$

Third, $\frac{\partial[\pi_D^*]}{\partial q}$:

$$\frac{\partial \pi_D^*}{\partial q} = \frac{(p_L)^2 q}{\alpha} \tag{22}$$

Finally, use the three previous results to show: $\frac{\partial[\pi_G^* - \pi_D^*]}{\partial q}$

$$(1-z)\frac{\partial \pi_{G,L}^*}{\partial q} + z\frac{\partial \pi_{G,H}^*}{\partial q} > \frac{\partial \pi_D^*}{\partial q} \tag{23}$$

$$\iff (1-z)\frac{(p_L)^2 q}{2\alpha}(1+t_L^*) + z\frac{(p_H)^2 q}{2\alpha}(1+t_H^*) > \frac{(p_L)^2 q}{\alpha}$$

$$\iff \frac{q}{2\alpha}[(1-z)(p_L)^2(t_L^* - 1) + z[(p_H)^2(1+t_H^*) - 2(p_L)^2]] > 0 \tag{24}$$

The **first term** is negative since t_L^* is between 0 and 1. This reflects the idea that if you use the global market you can end up with a co-author of the same quality as a domestic co-author, and then you also have to incur travel costs. The **second term** is positive and reflects the idea that if you use the global market you can end up with a co-author of higher quality with probability z. Ignoring the first term $\frac{q}{2\alpha}$, we simplify and re-write:

$$\iff (1-z)(p_L)^2(t_L^* - 1) + z[(p_H)^2(1+t_H^*) - 2(p_L)^2] > 0$$

$$\iff z[(p_H)^2(1+t_H^*) - (p_L)^2(1+t_L^*)] - (p_L)^2(1-t_L^*) > 0 \tag{25}$$

Note that the condition to satisfy $\frac{\partial[\pi_G^* - \pi_D^*]}{\partial q}$ is very similar to the condition to satisfy $\pi_G^* - \pi_D^* > 0$. The subsection below will show that if the level condition ($\pi_G^* - \pi_D^* > 0$) is satisfied, then the difference condition ($\frac{\partial[\pi_G^* - \pi_D^*]}{\partial q}$) is automatically satisfied. Therefore, no further assumptions are needed in addition of the level condition.

B5: Comparing the level condition to the difference condition

Level condition: $\pi_G^* - \pi_D^* > 0$ means:

$$z[(p_H)^2(1+\frac{1}{2}t_H^*) - (p_L)^2(1+\frac{1}{2}t_L^*)] - (p_L)^2(1-\frac{1}{2}t_L^*) > 0 \tag{26}$$

Difference condition: $\frac{\partial[\pi_G^* - \pi_D^*]}{\partial q} > 0$ means

$$z[(p_H)^2(1+t_H^*) - (p_L)^2(1+t_L^*)] - (p_L)^2(1-t_L^*) > 0 \tag{27}$$

Therefore, by combining both conditions one can show that:

$$
\begin{aligned}
&\frac{\partial[\pi_G^* - \pi_D^*]}{\partial q} > \pi_G^* - \pi_D^* \\
\iff\ & z[(p_H)^2(1+t_H^*) - (p_L)^2(1+t_L^*)] - (p_L)^2(1-t_L^*) > \\
& z[(p_H)^2(1+\frac{1}{2}t_H^*) - (p_L)^2(1+\frac{1}{2}t_L^*)] - (p_L)^2(1-\frac{1}{2}t_L^*) \\
\\
\iff\ & z[(p_H)^2(t_H^*) - (p_L)^2(t_L^*)] + (p_L)^2(t_L^*) > \\
& z[(p_H)^2(\frac{1}{2}t_H^*) - (p_L)^2(\frac{1}{2}t_L^*)] + (p_L)^2(\frac{1}{2}t_L^*) \\
\\
\iff\ & 2z[(p_H)^2(t_H^*) - (p_L)^2(t_L^*)] + 2(p_L)^2(t_L^*) > \\
& z[(p_H)^2(t_H^*) - (p_L)^2(t_L^*)] + (p_L)^2(t_L^*) \\
\\
\iff\ & z[(p_H)^2(t_H^*) - (p_L)^2(t_L^*)] + (p_L)^2(t_L^*) > 0
\end{aligned}
$$

Therefore, if a given project is preferred under a distant co-author, this effect is magnified when project quality rises.

Appendix C: Theoretical Model Under $z > w > 0$

Here we address the general case where $z > w > 0$ (in the main text we solved for the specific case where $w = 0$).

Profits under a *local co-author* are:

$$\pi_L^* = (1-w)\frac{(p_B q)^2}{2\alpha} + w\frac{(p_G q)^2}{2\alpha} = \frac{q^2}{2\alpha}(wp_G^2 + (1-w)p_B^2). \tag{28}$$

Profits under a *distant co-author* are:

$$\pi_D^* = \frac{q^2}{4\alpha}(zp_G^2(1+\frac{t_G^*}{2}) + (1-z)p_B^2(1+\frac{t_B^*}{2})), \tag{29}$$

where $t_i^* = \frac{(p_i q)^2}{8\alpha\beta}$.

Therefore, the difference between both profit levels can be shown to be:

$$\pi_D^* - \pi_L^* = \frac{q^2}{4\alpha}[p_G^2(z(1+\frac{t_G^*}{2}) - 2w) + p_B^2((1-z)(1+\frac{t_B^*}{2}) - 2(1-w))]. \tag{30}$$

By again using the definition of $\beta = \frac{1}{\theta}$, one can show that:

$$\frac{\partial[\pi_D^* - \pi_L^*]}{\partial\theta} = (\frac{q^2}{8\alpha})^2(zp_G^4 + (1-z)p_B^4) > 0, \tag{31}$$

and from here it is easy to see that the three predictions of the main text will be satisfied again.