Innovation on Wings: Nonstop Flights and Firm Innovation in the Global Context

Dany Bahar, Prithwiraj Choudhury, Do Yoon Kim, and Wesley W. Koo [1]

Abstract

We study whether, when, and how better connectivity through nonstop flights leads to positive innovation outcomes for firms in the global context. Using unique data of all flights emanating from 5,015 airports around the globe from 2005 to 2015 and exploiting a regression discontinuity framework, we report that a 10% increase in nonstop flights between two locations leads to a 3.4% increase in citations and a 1.4% increase in the production of collaborative patents between those locations. This effect is driven primarily by firms, as opposed to by academic institutions. We further study the characteristics of firms and firm locations that are salient to the relation between nonstop flights and innovation outcomes across countries. Using a gravity model, we posit and find that the positive effect of nonstop flights on innovation is stronger for firms and subsidiaries with greater innovation mass (e.g., stocks of inventors and R&D spending), for firms and subsidiaries located in innovation hubs or in countries that are deemed technology leaders, and for firm and subsidiaries that are separated by large cultural or temporal distance.

[1] Bahar: Brown University and Harvard Growth Lab; Choudhury: Harvard Business School; Kim: Boston College; and Koo: INSEAD. Authors listed alphabetically and contributed equally.

1. Introduction

The strategy literature has long posited that firms might overcome the constraints of geographically localized search for knowledge and collaborators through employee geographic mobility (e.g., Rosenkopf and Almeida 2003). A more recent literature documents how better connectivity through roads and nonstop flights facilitates geographic mobility, enabling innovation outcomes in the United States (Agrawal et al. 2017, Catalini et al. 2020). However, for firms and inventors located across different countries, the relation between connectivity and innovation outcomes—such as cross-border knowledge spillovers (e.g., Singh 2005) and the production of global collaborative patents (GCPs) (e.g., Kerr and Kerr 2018)—might be more nuanced. This relates to two core insights in prior literature: (1) the persistence of cultural, temporal, and other dimensions of distance that affect firm innovation in the global context; and (2) the "spikiness" in spatial distribution of innovation around the world.

First, a long-standing literature, notably Ghemawat (2001) and Berry et al. (2010), documents the persistence of geographic, cultural, economic, and other dimensions of distance in the global context. While better connectivity through nonstop flights shrinks geographic distance between firm locations, in the global context, cultural, temporal, and other dimensions of distance between firm locations may influence the relation between nonstop flights and knowledge spillovers and collaborations. In particular, prior literature (e.g., Chua et al. 2015) has highlighted cultural distance between firms and inventors as a key friction for innovation in a global context. Researchers (e.g., Bahar 2020, Chauvin et al. 2020) have also documented how temporal distance between firms and subsidiaries constrains synchronous communication, a key facilitator for collaborative innovation.

Second, prior literature (e.g., Florida 2005, Furman and Hayes 2004, Kim and Aguilera 2015) documents that the distribution of innovation across countries is characterized by "spikes" (i.e., disproportionate spatial concentration of innovation in a select few locations around the world) and by differences between countries that are technology "leaders" versus those that are "followers." This raises the question of whether the relation between connectivity and innovation across countries depends on the characteristics of firms and firm locations.

While a broader literature in economics studies how nonstop flights affect global economic outcomes, such as firm productivity and regional economic activity (Campante and Yanagizawa-Drott 2018, Giroud 2013), to the best of our knowledge, we lack evidence on whether, when, and how nonstop flights affect innovation outcomes for firms in the global context. This leads to our research questions: *Do nonstop flights affect firm innovation outcomes across countries? If so, how, and what characteristics of firms and firm locations connected by nonstop flights are salient to the relationship between nonstop flights and innovation outcomes across countries?*

To motivate our empirical analysis, we borrow insights from the gravity model, which has been used extensively to predict trade, FDI, and migration flows (Anderson 1979, Frankel and Rose 2002, Vanderkamp 1977). We make two sets of predictions regarding how characteristics of firms and characteristics of firm locations affect the relation between flight connectivity and innovation outcomes. First, we predict that the relation between nonstop flights and innovation outcomes across countries is more salient for firms and subsidiaries with greater innovation mass (measured by R&D spending and number of inventors at firms/subsidiaries) and for firms and subsidiaries located within innovation hubs or within countries that can be deemed technology leaders. Second, we predict that the relation between nonstop flights and innovation outcomes across countries is more salient for firms and firm locations separated by temporal and cultural distance. We measure temporal distance by computing time zone differences between firm locations, and we measure cultural distance using the ethnic composition of inventors at firms/subsidiaries and using metrics from the World Values Survey of 2020 (WVS), measured at firm locations.

For our empirical analysis, we exploit a proprietary dataset comprising the universe of all active air routes globally from 2005 to 2015, sourced from the Official Aviation Guide (OAG). This unique dataset contains information on all flights, including nonstop flights, in each year emanating from 5,015 airports around the globe. We then geo-match these airports to the addresses of inventors based on the universe of patents filed with the United States Patent and Trademark Office (USPTO) for the aforementioned years. In doing so, we obtain two measures of innovative activities—global citations and global collaborations—for all patenting entities (firms and academic institutions) and inventors near each of these airports.

With these data, we arrive at our results through several steps. First, we apply a regression discontinuity design (RDD) to study the causal effect of nonstop flights on innovation outcomes of interest (i.e., citations and collaborations across locations). Here, we build on the work by Campante and Yanagizawa-Drott (2018), which leverages a discontinuity. They find that airport pairs less than 6,000 miles apart are more likely to be serviced by nonstop flights than pairs that are more than 6,000 miles apart, due to aviation regulations significantly increasing costs of operating nonstop flights for destinations that are more than 6,000 miles apart. Thus, in the context of innovation outcomes, this exercise enables the comparison of patent citations and collaborations between inventors in location pairs that are just below the 6,000 miles threshold (such as Shanghai and Milan, which are 5,650 miles apart) to those in location pairs that are just above the 6,000 miles threshold (such as Shanghai and Madrid, which are 6,350 miles apart). Insofar as airport designers and planners cannot precisely manipulate their distance from other airports (e.g., position themselves within 6,000 miles of other airports), the regression discontinuity implies that variation in between-airport distance near the 6,000-mile threshold is as good as random (Lee and Lemieux 2010). This feature allows us to interpret our results as causal. Our main estimation using this methodology finds that for airport pairs around the 6,000-mile threshold, a 10% increase in the number of yearly nonstop flights between two locations increases citations by 3.5% and collaborations by 1.5%.

We further show that the effects of nonstop flights on innovation are driven primarily by firms, not by academic institutions. For firms, a 10% increase in flights increases citations by 3.38% and collaborations by 1.48%; for academic institutions, the same increase in flights is 0.99% and 0.40% for citations and collaborations, respectively. Thus, for the rest of the paper, we focus squarely on firms (and firm subsidiaries) and examine when nonstop flights matter for firm innovation in the global context.[2]

Using the gravity model as a theoretical anchor, we argue that the innovation mass of firms or subsidiaries (measured as the number of inventors and R&D spending), the innovation mass at the

[2] We study citations between firms and collaborations between subsidiaries within a firm. This is because in the patenting dataset, a citation is usually between two different assignees (e.g., firms), and a global collaborative patent (collaboration) is between inventors within a single assignee. In other words, we assume that for a global collaborative patent assigned to a firm, collaborating inventors are located at different subsidiaries within the firm.

firm/subsidiary location (i.e., whether the location is a hub or whether the location is in a country deemed a technology leader), and non-geographic distances (i.e., temporal and cultural distances) between firms/subsidiaries are key for their joint innovation outcomes and, consequently, how nonstop flights can affect innovation outcomes. In particular, by re-estimating our RDD using different samples and subsamples based on characteristics of firms/subsidiaries and characteristics of firm/subsidiary locations, our empirical results show that the relation between nonstop flights and innovation outcomes is more salient for firms and subsidiaries: (1) with greater innovation *mass*; (2) located in innovation *hubs* or within countries that have been deemed *technology leaders*; and (3) separated by cultural and temporal *distance*.

We also explore the two mechanisms outlined in the airline connectivity literature to understand how nonstop flights affect innovation outcomes for firm innovation across countries: reduction in pecuniary travel costs (e.g., Catalini et al. 2020) and reduction in travel time (e.g., Bernstein et al. 2016). Using data from Google Flights, we find that travel time reduction, rather than reduction in pecuniary costs, is the most prominent factor explaining the relation between nonstop flights and global citations and collaborations at firms. In additional analyses, we show that our local average treatment effects at the airport pair level also extend to the airport level, suggesting that the documented effects represent aggregate increases in citations and collaborations.

Our findings contribute to several literatures. First, we contribute to the literatures in strategy and economics on knowledge spillovers and collaborative patents for firms in the global context (Bahar et al. 2020, Branstetter et al. 2014, Kerr and Kerr 2018, Miguelez and Fink 2013, Singh 2005). We find that knowledge frictions due to national borders, as discussed in Singh and Marx (2013),[3] are attenuated by the availability of international nonstop flights. This finding highlights how airline connectivity mitigates the frictions exerted by political borders and cross-country distance on international collaborations (Alcácer et al. 2017, Berry et al. 2010, Ghemawat 2007). Additionally, our results show that nonstop flights especially facilitate innovation outcomes for firms with high levels of innovation mass, for firms located near hubs or in

[3] Singh and Marx (2013) find that, even after accounting for geographic proximity between patent inventors, country and state borders still constrain knowledge diffusion in the form of citations.

countries deemed technology leaders, and for firms that are temporally and culturally distant. In light of Rosenkopf and Almeida (2003), this suggests that while nonstop flights help firms build bridges to temporally or culturally distant places, they also disproportionately help firms with more innovation mass or firms located within regions with greater innovative activities overcome the constraints of geographically localized search for knowledge and collaborators.

Second, by reporting a causal relation between easing human mobility through nonstop flights and the production and diffusion of knowledge at firms across country borders, we contribute to an evolving literature on connectivity/geographic mobility and economic outcomes (Agrawal et al. 2016, Bernstein et al. 2016, Campante and Yanagizawa-Drott 2018, Catalini et al. 2020, Choudhury 2017, Giroud 2013). While prior research documents the economic consequences of flight connectivity for domestic innovation (primarily in the U.S.),[4] we provide a parsimonious model and document causal evidence related to nonstop flights and firm innovation in a global context. Importantly, our paper also documents the conditions under which flight connectivity is likely versus unlikely to drive innovation at firms globally. By showing that flight connectivity is less effective for facilitating knowledge spillovers and collaborations at firms with smaller innovation masses, at firms located outside regions with significant innovative activity, and at firms and firm locations that are temporally or culturally proximate, this paper provides an important account of the *limitations* of transportation connectivity in driving innovation. Moreover, the insight on how the effects are stronger for firms vis-à-vis academic institutions is a novel insight for this literature. In summary, by focusing on firms in the global context, by considering how characteristics of firms and firm locations drive our results, and by documenting travel time savings as the underlying mechanism, our study departs from prior research (e.g., Catalini et al. 2020) that studies how flight connectivity and travel cost savings affect innovation in a domestic and academic context.

---

[4] For example, Catalini et al. (2020) find that after Southwest Airlines introduced new routes in the U.S., collaborations between academic scientists increased from 0.3 to 1.1 times. The authors highlight travel costs as an important hurdle to innovation collaborations. Agrawal et al. (2017) study roads and innovation. They find that a 10% increase in the availability of highways in a region is associated with a 1.3% increase in citation-weighted patents. Their study points to a mechanism through which roads drive innovation (within-region knowledge flows), as roads make it easier for innovators located in the same region to interact with one another.

Finally, this study also provides important and timely policy implications for firms and managers. As companies and knowledge workers debate whether to resume international travel after the pandemic (Weed 2021), our results shed light on the importance of business travel and nonstop flights for knowledge spillovers and collaborations, especially in a global context.

The paper is organized as follows: In Section 2, we introduce the empirical setting, data, and variables. In Section 3, we describe the regression discontinuity approach and present the causal estimates. In Section 4, we explore how characteristics of firms/subsidiaries and firm/subsidiary locations affect our results. Section 5 presents additional analyses, in which we estimate the effect of flight connectivity on innovation for a single location (as opposed to a pair of locations) to study whether nonstop global flights result in an aggregate increase in innovation outcomes. In Section 5, we also report results related to underlying mechanisms. Finally, in Section 6, we conclude with a set of research and managerial policy implications. Our paper is accompanied by an online Appendix.

2. Data and Setting

2.1 Dataset and Dependent Variables

Our data on commercial flights comes from OAG, a private company specializing in aviation analytics. This data contains information on 94,221 routes between 5,068 airports around the globe. We exclude 53 airports that serve only cargo flights or have no flights for the entire period from 2005 to 2015. This yields 5,015 airports. For each route, we calculate the geodesic distance between the origin and destination (using the *geodist* command in Stata).[5]

Next, we collect patent data and map each patent to nearby airports. We use the universe of patents filed in the U.S. Patent and Trademark Office (USPTO) from 2005 to 2015 from the Thompson Reuters patent dataset, as well as all of their forward citations. To assign patents to airports, we use inventor addresses contained in the patents. First, we map each inventor to *all* airports within a 50-mile radius of the inventor's

[5] As Campante and Yanagizawa-Drott (2018) point out, the *geodist* command in Stata computes the geodesic distance: the length of the shortest curve between two points along the surface of (a mathematical model of) the Earth, not the actual flight distance. As such, this proxy is exogenous to the geopolitical factors that determine actual flight distance.

location, within his/her own country.[6] This gives us a patent-inventor-airport-level dataset, with an inventor potentially assigned to multiple airports. Each patent is then assigned a location based on its inventors. If a patent has only one inventor, we code the patent's location as that inventor's location. If a patent has multiple inventors, we use the first inventor's location. Figure 1 shows the locations of all the inventors in our dataset. As expected, given that innovation activities occur predominantly in developed countries, most locations are in the United States, Europe, and East Asia. We categorized assignees into firm or academic institutions based on their names: We checked whether the assignee name contains words and phrases such as "university," "college," "institute of technology," or "school." Since many assignees are foreign, we translated each of those terms into all available languages on Google Translate.[7] Our sample contains 11,756 unique academic assignees and 198,327 unique firm assignees. Where applicable, we use the Duke DISCERN database (Arora et al. 2021) and the Compustat database to obtain the balance sheet information for the firm assignees.

[Insert Figure 1 here]

For the first dependent variable, citations, we measure activity at the airport pair level. We count a citation between two airports if there exists a patent citation from a patent mapped to one airport to a patent mapped to another airport. Our dataset is not directional—that is, the airport pair CDG-ORD (Paris Charles de Gaulle-Chicago O'Hare) in 2005 appears only once in the dataset, but there is no observation for the opposite direction, ORD-CDG, in 2005. This is because citations and collaborations may be driven by either flight direction (CDG-ORD or ORD-CDG). Therefore, we take the sum of all flights and the sum of citations/collaborations that occur between inventors located near two airports, regardless of flight direction, to measure the overall level of citation activity between the two locations. In cases where a patent inventor is assigned to more than one airport, we inversely weight the patent by the number of airports to avoid double counting citations. For example, if a citation involves one inventor assigned to ORD (Chicago O'Hare) and

[6] Specifically, we map inventors to all airports within a 50-mile radius with territorial contiguity and within their own country, with the exception of the Schengen Area, for which we relax the "within same country" rule as long as there is territorial contiguity between the location of the inventor and the nearby airport.
[7] The final list of keywords is available upon request.

another inventor assigned to both CDG and ORY (Paris Charles de Gaulle and Paris Orly), our approach assigns 0.5 citations for each of the ORD-CDG and ORD-ORY airport pairs.

For the second dependent variable, collaborations, we similarly measure those activities at the airport pair level. A collaboration between two airports corresponds to a collaborative patent with inventors from locations nearby those two airports. Specifically, for each year and each airport pair, we count the number of collaborative patents by inventors from both airports. In cases where an inventor is assigned to multiple airports, similar to how we count citations, we inversely weight each collaboration by the number of airports in order to avoid double counting.

Our final dataset, thus, consists of a yearly panel of citations, collaborations, and the number of flights at the airport pair level. Given that all collaborations pertain to a single assignee (i.e., multiple subsidiaries of the same assignee) and most citations are between assignees, we are able to conduct separate analyses based on the characteristics of firms/subsidiaries and of firm/subsidiary locations. An important caveat here is that we proxy firm/subsidiary location using the airport location—that is, airports proximate to the inventors working at the firm/subsidiaries. Any airport pair will appear for all 11 years of the sample (2005 to 2015), even if there were no flights or no patent information reported in some of those years (which are assumed to be zero), making it a balanced panel. A preliminary analysis using a counterfactual patent-matching method shows that more nonstop flights between inventors are associated with more citations and collaborations.[8]

2.2 Characteristics of Firms and Firm Locations

To study heterogenous effects, we collect information about innovation mass of firms/subsidiaries and distances between firms/subsidiaries. For innovation mass, we utilize three measures. First, we match firms to their balance sheets using Duke DISCERN data and Compustat data, and we obtain R&D spending at the firm level. Second, for each firm, we count the number of unique inventors who filed patents with the firm. A

[8] We conducted a correlational exercise using the counterfactual patent-matching methodology. This procedure helps map out the broad relation between flight connectivity and innovation outcomes across countries. In Appendix Sections A4 and A5, we show that nonstop flights provide an additional 2.6 percentage point increase in citations and a 2.9 percentage point increase in collaborations for international airport pairs.

third measure at the firm location level captures airports' proximity to hubs of academic science: for each airport (i.e., firm/subsidiary location), we check whether it is within a 50-mile radius of a scientific hub as defined in Bikard and Marx (2020). We also differentiate between firms/subsidiaries located in innovation-leader countries and those in innovation-follower countries (Furman and Hayes 2004).

We also collect data on distances between firm/subsidiaries and between their locations. First, we collect data on temporal distance, measured by the extent of working hour overlap between two firm/subsidiary locations (proxied by the airport location). For each airport, we obtained the time zone using its latitude and longitude coordinates and the *timezonefinder* package in Python.[9] Using the time zones, we calculate the time difference between the origin airport and the destination airport, then calculate the working hour overlap to be from 0 to 8 hours.[10] Next, we measure the cultural distance between two firm/subsidiaries in two ways. First, we obtain the ethnicities of the inventors in our sample through machine learning algorithms for name-matching, and we measure whether a group of citing or collaborating inventors are multiethnic or co-ethnic. The assumption is that a co-ethnic group of inventors who cite or collaborate with one another have shorter cultural distance (are more culturally similar) than a multiethnic group of citing/collaborating inventors. Second, we use data from a specific question in the World Value Survey—a question that measures how much each country's citizens believe immigrants play important roles in their society. This measure is particularly relevant to the context of global knowledge spillovers and collaborations because it gauges a specific cultural element—people's tendency to appreciate the work of foreigners.

3. Do Nonstop Flights Drive Firm Innovation in the Global Context? A Regression Discontinuity Approach

To causally estimate the impact of nonstop flights on innovation across countries, we utilize a unique feature of the airline industry: airport pairs that are less than 6,000 miles apart are more likely to be serviced by nonstop flights than are pairs that are more than 6,000 miles apart. This pattern exists because of higher

[9] More information can be accessed at https://pypi.org/project/timezonefinder/.

[10] Working hour overlap was defined as 8 hours if the time difference between the origin and destination airports is 0, 24, or -24 hours; 7 hours if the time difference is -1, 1, or 23 hours, and so forth.

operating costs to service routes that are more than 6,000 miles long; this expense is due to a combination of administrative and legal rules as well as technological factors. This paves the way for a regression discontinuity design. While Campante and Yanagizawa-Drott (2018) pioneered this approach, to our knowledge, we are the first to apply this method to explore innovation outcomes.

The 6,000-mile threshold corresponds to 12 hours of flight time given customary flight speeds (Campante and Yanagizawa-Drott 2018); flights above the 12-hour and 6,000-mile thresholds are known as ultra-long-haul (UHL) flights (McKenney et al. 2000). These UHL flights must meet special personnel availability requirements. For instance, the Federal Aviation Administration (FAA) requires flights that are more than 12 hours long to have an additional flight crew member as well as adequate sleeping quarters on the plane. Such requirements lead to greater operational costs for these flights, as the crew corresponds to about 36% of nonfuel costs (Federal Aviation Administration 2016). Technological advances in the 1980s and 1990s made this discontinuity more pronounced, as long-range airplane models introduced during this period made long-haul flights more fuel efficient, which accentuated the importance of minimizing nonfuel costs (e.g., crew costs).[11]

Using this feature of the data, we implement a fuzzy regression discontinuity analysis. The fuzzy design responds to the fact that while there are still nonstop flights above the 6,000-mile threshold, there are many fewer of these than there are flights below the 6,000-mile threshold. The unit of analysis, similar to the one adopted in Campante and Yanagizawa-Drott (2018), is at the airport pair level. The "treatment" in this setting corresponds to an airport pair being slightly below the 6,000-mile threshold, allowing for a higher likelihood of having nonstop flights between the two airports.[12] This is the key assumption: There is no reason to believe that innovation activities occurring between locations that are slightly more than 6,000 miles apart should be significantly different from those occurring between locations that are slightly less than 6,000 miles apart. In other words, arguably, whether the distance between any airport pair lies just above or just

[11] The Boeing 747-400 commenced commercial operations in 1989, followed by the Airbus A330 and A340 models as well as the Boeing 777 series. The 747-400 family was about 20% more fuel efficient than the previous best-selling planes, and the 777 pushed this gain to about 30% (Kharina and Rutherford 2015).

[12] For instance, 6,000 miles corresponds roughly to the distance between Los Angeles and Munich (slightly less than 6,000 miles) or from Cologne to Sao Paulo (slightly more than 6,000 miles).

below the 6,000-mile threshold is as good as randomly assigned. We provide summary statistics for our RDD dataset in Table 1.[13]

[Insert Table 1 here]

From Table 1, we see that the average airport pair has 2.14 citations and 2.00 collaborations in a given year, but the distribution is skewed to the left. Average firm citations between airport pairs is 1.96, and firm collaborations are 1.90. About half the airport pairs (in a given year) have nonstop flights, with the average number of nonstop flights at 611.95. The average distance between two airports in our dataset is around 1,111 miles. Hub-to-hub flights (flights between two airports that are both innovation hubs) are about 26% of the routes. The average location pair has a working hour overlap of 7.04 hours (1.83 hours for location pairs in the regression discontinuity sample). The average difference in immigrant friendliness scores between locations (explained in detail in Table 1's footnote) is 0.19. Finally, the average price for any ticket in our sample is $946.48, and the average travel duration is 13.72 hours.

A benefit of the regression discontinuity approach is that it is possible to visualize the effect of the discontinuity on innovation outcomes. Figure 2 presents a visual summary of our reduced-form results.[14] We see from both panels that airport pairs that are slightly less than the 6,000-mile threshold have more citations and collaborations than airport pairs that are slightly more than the 6,000-mile threshold.

[Insert Figure 2 here]

Our main analysis for the regression discontinuity quantifies the graphical relationship. Since this is a fuzzy regression discontinuity, the estimation involves two stages, where the first stage estimates the discontinuity in the number of flights around the 6,000-mile threshold and the second stage estimates the impact of the discontinuity on the outcomes of interest (innovative activities):

[13] We present summary statistics for the RDD sample in Appendix Table B1.

[14] In Appendix Section B7, we show our graphical results are robust to using higher-order polynomials to fit the data points to either side of the discontinuity.

$$
\begin{aligned}
&\operatorname{asinh}\left(Nonstop\ Flights_{a_o,a_d,t}\right) \\
&\qquad = \gamma_1 1\{Dist_{a_o,a_d} < 6{,}000\} + \gamma_2\left(Dist_{a_o,a_d} - 6{,}000\right) \\
&\qquad + \gamma_3 1\{Dist_{a_o,a_d} < 6{,}000\} \times \left(Dist_{a_o,a_d} - 6{,}000\right) + \phi_{c_o,c_d,t} + \epsilon_{a_o,a_d,t}
\end{aligned} \tag{1}
$$

$$
\begin{aligned}
\operatorname{asinh}\left(Y_{a_o,a_d,t}\right) &= \beta_1\, asinh\left(\widehat{TotalFlights}_{a_o,a_d,t}\right) + \beta_2\left(Dist_{a_o,a_d} - 6{,}000\right) \\
&+ \beta_3 1\{Dist_{a_o,a_d} < 6{,}000\} \times D\left(Dist_{a_o,a_d} - 6{,}000\right) + \phi_{c_o,c_d,t} + \epsilon_{a_o,a_d,t}
\end{aligned} \tag{2}
$$

Here, a_o and a_d refer to the origin and destination airports for a given route, and t refers to a calendar year. Our main variable of interest, $TotalFlights_{a_o,a_d,t}$, measures the number of nonstop flights between a_o and a_d in year t, which is estimated through the first stage in Equation (1).[15] Similarly, $Dist_{a_o,a_d}$ measures the distance in miles between the airports, a_o and a_d, and our running variable (the variable that determines which observations are "treated" with additional nonstop flights) is denoted as $(Dist_{a_o,a_d} - 6{,}000)$. The coefficient of interest is β_1, which measures the discontinuity at the 6,000-mile threshold. Intuitively, β_1 measures the jump in $Y_{a_o,a_d,t}$ at the 6,000-mile threshold from fitting separate regression slopes on either side of the discontinuity. To absorb the effects due to differences between two countries (e.g., language and time zone differences) on citations or collaborations, we include country-country-year fixed effects, marked as $\phi_{c_o,c_d,t}$. We utilize the inverse hyperbolic sine transformation (asinh) since it allows us to preserve observations with zeroes (MacKinnon and Magee 1990).[16]

First-stage regressions confirm the existence of discontinuity around the 6,000-mile threshold, where airport pairs just below the threshold have on average 260 to 550 more nonstop flights per year than airport pairs just above the threshold (see more details of the first stage in Appendix Section B2; Appendix Figure B1

[15] In this analysis, we use the total number of nonstop flights, instead of the binary variable for the existence of nonstop flights. Using a binary indicator as the instrument, we also conclude that the existence of a nonstop flight increases citations and collaborations, but the effect size is greater: The existence of a nonstop flight increases citations by 90.89% ($\beta_1 = 2.3964, p < 0.01$) and collaborations by 65.25% ($\beta_1 = 1.0571, p < 0.01$). Our preferred specification, to be conservative, is using the continuous number of flights variable, given that the very small changes in the binary variable exploited using the fuzzy regression discontinuity design might overestimate the Wald estimator.

[16] Appendix Section B12 shows that our results are robust to using raw counts, log+1 transformations, and Poisson quasi-maximum likelihood estimators.

provides visual evidence for the existence of the discontinuity). In Table 2, we present results of the second-stage specification.[17]

[Insert Table 2 here]

Columns 1 and 2 show that a 10% increase in the number of nonstop flights between two locations leads to an increase in patent citations of 3.4% and a 1.4% increase in the number of collaborations.[18]

3.1. Firms versus Academic Institutions

An important question is which entities—firms or academic institutions—benefit more from the presence of nonstop flights. An answer to this question will also shed light on which individuals (i.e., employees at firms versus academics) are likely taking the nonstop flights and contributing to knowledge spillovers. Firms, in particular, increasingly rely on global collaborations and learning in their innovative activities, as evidenced by the rising number of GCPs and the higher quality of GCPs relative to that of same-country patents for firms (Kerr and Kerr 2018).

We determine whether a patent's assignee is a firm or an academic institution by searching based on a set of keywords (e.g., "school," "university," "institute")[19] in its name. Then, we count citations and collaborations between locations for firms and academic institutions separately, and we test whether the marginal effect of nonstop flights is greater for firms or academic institutions using seemingly unrelated estimations (Mize et al. 2019) as well as bootstrapping. Table 2, Columns 3-6 present the results. We find that a 10% increase in the number of nonstop flights between two locations leads to an increase in citations of

[17] Generally, RD results are sensitive to which observations near the threshold are included. We provide two thresholds: a 500-mile bandwidth and an "optimal" bandwidth. The optimal bandwidth is calculated following the methodology described in Calonico et al. (2014), which builds on prior work on optimal bandwidth choice in RD by Imbens and Kalyanaraman (2012). In Appendix Sections B3-B5, we show our RD results are robust to varying the number of bins, the bin selection method, and kernel choice, as well as different levels of fixed effects and clustering.

[18] In Appendix Sections B8 and B9, we conduct permutation tests on the 6,000-mile threshold to check whether we see similar discontinuities at thresholds other than our 6,000-mile mark. We show that the RD coefficients are insignificant when using random thresholds far from the 6,000-mile mark, confirming the validity of our 6,000-mile threshold. We also conduct permutation tests on the running variable (e.g., the distances to 6,000-mile variable) to test whether airports strategically locate themselves closer to other airports. We find no discontinuities in our running variable and, thus, no precise manipulation of airport locations. Appendix Section B11 further shows the effects are indeed driven by nonstop flights, not one-stop flights.

[19] In addition to these keywords, we use Google Translate to translate the keywords across all available languages, and we include those keywords in our categorization as well.

3.38% between two firms in those two locations and 0.99% between two academic institutions. A comparison of the coefficients shows that the two coefficients are significantly different (Diff = 0.2398, s.e. = 0.0929). Similarly, a 10% increase in nonstop flights between two locations leads to an increase in collaborations of 1.48% between two subsidiaries within a firm and 0.40% between two entities within an academic institution. The point estimates are also significantly different (Diff = 0.1072, s.e. = 0.0527). These results show that nonstop flights mainly serve to facilitate citations between firms and collaborations between subsidiaries within a firm (and much less so between academic institutions or between their branches). In light of the relative importance of firms in driving innovation across countries through nonstop flights, the next section focuses squarely on further exploiting variation in firms/subsidiaries and between firm/subsidiary locations.

4. Firm Heterogeneity: A Gravity Model

What characteristics of the firms/subsidiaries and firm locations being connected by nonstop flights and of firm/subsidiary locations being connected by nonstop flights affect the relationship between nonstop flights and innovation across countries? To answer this question, we build on the gravity model, which scholars have used to explain migration, bilateral trade flows, and FDI between countries (Anderson 1979, Frankel and Rose 2002, Vanderkamp 1977). When applied to trade, the model states that bilateral trade volume is proportional to the product of the countries' masses (measured in countries' GDP) and inversely proportional to the distance between the countries. When applied to innovation, the gravity model states that knowledge flows and collaborations between firms are proportional to the product of the innovation masses in those firms (commonly measured as patenting activities or number of inventors in nearby locations) and inversely proportional to the distance between those firms (Montobbio and Sterzi 2013, Picci 2010). The following parsimonious equation illustrates the gravity model's key assumptions:

$$Y_{ij} \sim M_i \cdot M_j / D_{ij} \qquad (3)$$

which translates to the following equation after taking the logarithms of both sides and adding the temporal dimension:

$$log(Y_{ijt}) = \beta_0 + \beta_1\, log(M_{it}) + \beta_2\, log(M_{jt}) + \beta_3\, log(D_{ij}) + \epsilon_{ijt} \tag{4}$$

In these equations, *Y* is the outcome of interest (knowledge flows and collaborations between two firms), *M* represents the innovation masses in firm *i* and firm *j*, and *D* is the distance between firms *i* and *j*.[20] Whereas distance is usually the geographic distance, it can also stand for other types of distance (e.g., cultural, economic, and language distances). By the gravity model, we expect the coefficients on the mass terms (β_1, β_2) to be positive and the coefficient on the distance term (β_3) to be negative.

In this study, we modify the gravity model in Equation (4) to predict the conditions under which nonstop flights facilitate (do not facilitate) knowledge flows between firms and collaborations between subsidiaries within a firm. We present two arguments—one regarding innovation masses of a pair of firms and one regarding the distances between those firms/subsidiaries and their locations. Rather than focusing on mass and distance as the main effects, we discuss their roles as moderators that amplify or suppress the effect of nonstop flights on innovation outcomes. The following equation illustrates the modified gravity model:[21]

$$\begin{aligned} log(Y_{ijt}) = \beta_0 &+ \beta_1\, log(M_{it}) \\ &+ \beta_2\, log(M_{jt}) \\ &+ \beta_3\, log(D_{ij}) + \beta_4\, log(M_{it}) \cdot Flights_{ijt} \\ &+ \beta_5\, log(M_{jt}) \cdot Flights_{ijt} \\ &+ \beta_6\, log(D_{ij}) \cdot Flights_{ijt} + \beta_7 Flights_{ijt} + \epsilon_{ijt} \end{aligned} \tag{5}$$

First, regarding innovation mass, prior literature has shown that firm innovation in a global context is "spiky"—that is, the spatial distribution of innovative activities is highly concentrated in a few locations (Bresnahan et al. 2001, Florida 2005, Kerr and Robert-Nicoud 2020).[22] We extend this logic to firms. A lack

[20] To develop conceptual arguments, when we mention "firms," we imply both firms and their subsidiaries.

[21] We do not estimate the point estimates outlined in equation 5, but rather use the modified gravity model to motivate how innovation mass of firms and temporal/cultural distance between firms affect the relation between direct flights and innovation outcomes. As we will explain later, we employ subsample analyses and regressions for heterogeneous firms.

[22] Richard Florida's 2005 article documents a "spiky" map of innovation where the global patenting peaks are Tokyo, Seoul, New York, and San Francisco. Innovation activities are more concentrated in a few global locations than is economic activity or population. Bresnahan et al. (2001) present case studies that illustrate the necessary preconditions

of ex ante innovation mass at firms being connected (e.g., number of nearby inventors, level of R&D) acts as an innovation bottleneck that *cannot be* directly alleviated by nonstop flights. Just as connecting two countries with low GDPs and low populations is unlikely to create a significantly higher trade flow (because those two countries have few things to trade in the first place), connecting two firms with nonstop flights is unlikely to lead to increased citations and collaborations if those firms are characterized ex ante by low innovation masses. Conversely, nonstop flights connecting firms with large innovation masses are particularly likely to increase citations and collaborations between those firms because nonstop flights build a bridge between two firms with ex ante high innovation stocks, where inventors would otherwise have found it more difficult to meet each other face-to-face. Therefore, we posit that the positive effect of nonstop flights on innovation is stronger for firms and subsidiaries with greater innovation mass and for firms and subsidiaries located in innovation hubs or in countries that are deemed technology leaders.

In contrast to innovation mass, the lack of which is a bottleneck that cannot be alleviated by nonstop flights, the distance between two firms or subsidiaries is a bottleneck that might be alleviated by nonstop flights. International business scholars have developed various types of distance measures to study firms' internationalization decisions and other firm-level outcomes (Berry et al. 2010, Ghemawat 2007). Their key insight is that different types of distance across countries (e.g., geographic distance, cultural distance, temporal distance) create communication barriers that deter mobility, exchange, and collaboration. In our study, we focus on cultural and temporal distance, given the importance assigned to these two dimensions of distance in prior literature.

If two inventors are culturally distant, they are likely to carry different understandings of power relations and individualism (Hofstede 1980). Cultural distance prevents information flow, increases uncertainty in a relationship, and increases the cost of communication and collaboration (Berry et al. 2010, Lazear 1999). Two recent studies in the literature on innovation in a global context (Chua, Roth, and Lemoine

---

for the formation of new innovation hubs (concentration of firm-building capabilities and managerial skills, supply of skilled labor, and connections to markets). Recently, Kerr and Robert-Nicoud (2020) document the uneven distribution of innovation globally. The top 10 global innovation clusters in terms of patent count include large cities in Asia, the United States, and France. The first-place cluster (Tokyo-Yokohama) holds twice the patent count of the second-place cluster (Shenzhen-Hong Kong).

2015, Kerr and Kerr 2018), highlight the importance of cultural distance for our research question. While Kerr and Kerr (2018) posit that cultural sensitivity promotes global collaboration between innovators, Chua et al. (2015) provide a theoretical reasoning for why cultural distance matters for innovation in a global context. In particular, Chua et al. (2015) theorize that for global collaborative innovation to be successful, there has to be cultural alignment around the proposed solution between individuals in the two countries. They also theorize that cultural distance between the two countries, as well as the degree of "cultural tightness" in each country—that is, the degree of acceptance of "deviant" views of foreigners—determine whether or not there is cultural alignment.

Prior research also suggests the importance of studying temporal distance for our research question. Greater working hour overlap (short temporal distance) is associated with greater levels of collaboration and knowledge sharing because working hour overlap can reduce frictions in synchronous communication (Bahar 2020, Bircan et al. 2021, Chauvin et al. 2020, Espinosa et al. 2015, Mell et al. 2021). A long-standing literature has argued how being face-to-face facilitates the sharing of tacit knowledge and collaboration (e.g., Gaspar and Glaeser 1998, Nardi and Whittaker 2002). In the age of proliferation of synchronous communication technologies such as Zoom and Skype, knowledge can also be shared virtually, especially if individuals are in the same time zone and have common working hours. It is possible that direct flights help temporally distant innovators more than temporally proximate innovators, given that temporally proximate innovators have greater working hour overlap and opportunities to communicate synchronously using technologies, even without direct flights. Thus, we posit that the effect of nonstop flights on innovation is more positive for firms and subsidiaries separated by large cultural or temporal distance.

To test these predictions, we carry out analyses using different measures of mass and cultural/temporal distance in the context of the regression discontinuity framework.

4.1 Importance of Innovation Mass of Firms and Firm Locations

For innovation mass, we consider two firm-level variables: inventor mass and R&D spending. Inventor mass refers to the total number of inventors who have filed patents with a firm. R&D spending is also a proxy for the level of innovation productivity at a firm. To obtain the R&D spending data, we build on the Duke

DISCERN dataset (Arora et al. 2021) to match assignees to Compustat. For all matched assignees, we obtain the average of their R&D spending for 2005 to 2015.

To estimate the differential effect of flights for different levels of inventor mass and R&D spending, we split patent assignees into groups of high and low mass. A firm has high inventor mass if it has an above median number of inventors, and a firm has low mass otherwise. Similarly, a firm has high R&D spending if its average R&D spending from 2005 to 2015 is above the median in our sample. Once firms are categorized as high or low mass, we count the number of citations for high/low mass firms separately to the airport-pair-year level. Thus, we create two sets of variables: the number of citations by high mass firms and the number of citations by low mass firms. Finally, we twice run the same regression discontinuity specification as Equation (2), once using the number of citations at high mass firms and another time using the number of low mass firm citations. Similarly for collaborations, using the same high and low mass split, we aggregate the number of collaborations to high/low mass firm patents to the airport-pair-year level. We present the results below.[23]

[Insert Table 3 here]

We find support for predictions from the modified gravity model: The effect of nonstop flights on innovation across countries is greater for firms with high innovation mass. For firms with high inventor mass, a 10% increase in nonstop flights increases citations by 3.29%; the change is 1.85% for low inventor mass firms (Columns 1-2). Block bootstrapping tests reject the null of no difference in point estimates between the regressions for high and low mass firms (Diff = -0.143, p = 0.032).[24] Similarly for collaborations, a 10% increase in nonstop flights increases collaborations across subsidiaries by 1.41% for high mass firms, but 0.2% for low mass firms (Diff = 0.121, p = 0.027). Columns 5-8 show a similar picture: firms with high R&D spending benefit more from nonstop flights, but the difference is not significant for citations. For

[23] In Appendix Section C6, we present an additional set of mass results based on firm-level variables including revenue and employee count, in addition to R&D spending.

[24] Appendix Section E provides a detailed overview comparing RD coefficients across models. We use a seemingly unrelated estimates approach to compare effect sizes (Mize et al. 2019), and we also provide block bootstrapping results.

collaborations, firms with higher R&D spending benefit more, with a 10% increase in flights increasing collaborations by 0.93% for high R&D spending firms, but 0.31% for others (Diff = 0.062, p = 0.212).

Next, we study the importance of firms/subsidiaries being located in scientific hubs. Innovation hubs are locations (usually cities) where patenting activities in a technical field are geographically concentrated (Bikard and Marx 2020).[25] We mark an airport as belonging to an innovation hub if it is within a 50-mile radius of any hub.[26] We split the data into two subsamples: (1) routes that connect two firm/subsidiary locations (proxied by airport locations) that are both located near innovation hubs; and (2) routes in which at least one airport is not located near an innovation hub.[27] The two subsamples contain similar numbers of observations. For both subsamples, we repeat the regression discontinuity analysis to test whether the effect of nonstop flights on innovation depends on innovative activity levels. As with the RDD analysis, we conduct these subsample analyses at the airport pair-year level.

[Insert Table 4 here]

Table 4 shows the results for our subsample analysis. Columns 1 and 2 show that nonstop flights increase both citations and collaborations, respectively, for hubs. A 10% increase in nonstop flights increases citations by 3.2% and increases collaborations by around 2.5%. However, when either airport is not near an innovation hub, collaborations and citations are not impacted. Both coefficients in Columns 3 and 4 suggest that the effect sizes are close to zero and not statistically significant. This result indicates that nonstop flights enhance production of GCPs and knowledge flows mainly through connecting inventors located at various innovation hubs. This result should be interpreted with caution, as hubs and non-hubs differ inherently in terms of the likelihood of having nonstop flights and the ability to produce innovative ideas.

---

[25] The authors define innovation hubs as cities with significant patenting for a given subclass, for all subclasses. Specifically, they code a hub as being within a 50-mile radius of a city with (1) more than 5% of patents in a given subclass and (2) at least five patents within that subclass. Bikard and Marx (2020) provide additional details (including the location data for hubs).

[26] Appendix Section C1 contains examples of airports near hubs and those not near hubs. Section C2 tests the relationship between flight distance and the likelihood of a flight connecting two innovation hubs.

[27] We consider hub-to-hub connections versus routes with at least one non-hub airport in the main draft. Appendix C4 shows non-hub to non-hub routes do not benefit from nonstop flights.

Finally, we study knowledge diffusion between firms and subsidiaries located in countries at the technological frontier and those in follower countries. We borrow from Furman and Hayes (2004)'s categorization of leader versus follower countries in terms of their historical innovative productivity. According to their categorization, the leading innovating countries include Germany, Japan, Sweden, Switzerland, and the United States. These countries are categorized as "leaders," and the other countries (countries labelled as "middle tier," "third tier," and "emerging innovators" in Furman and Hayes (2004)'s terminology) are categorized as "followers." Then, we restrict the sample to citations and collaborations by firms located in these leader and follower countries, and we gauge the flights' effects on citations and collaborations that occur (1) between leaders, (2) between followers, and (3) between leaders and followers.[28] Table 5 shows that a 10% increase in nonstop flights between firms in two "leader" countries leads to a 17.95% increase in citations and a 4.96% increase in collaborations. We also find statistically nonsignificant effects for firms located in leader-follower and follower-follower country pairs.

[Insert Table 5 here]

4.2 Importance of Temporal and Cultural Distances

In this section, we explore how cultural and temporal distances between firms/subsidiaries and their locations (proxied by airport locations) affect the relation between nonstop flights and innovation outcomes. The analysis is done at the airport pair-year level. We consider the effects of two types of distance—temporal distance and cultural distance. For each type of distance, we compare the effect of nonstop flights on firms' innovation outcomes at airport pairs that are distant or close. Since our RDD setup includes only location pairs that are similar in terms of geographic distance, we will focus on non-geographic measures of distance. Whereas the location pairs in the RDD setup are of similar geographic distance (around 6,000 miles), there is significant variation in temporal and cultural distances.

First, we test whether the positive effect of nonstop flights on innovation is stronger for firms and subsidiaries separated by large temporal distance. To test this, we divide the sample into airport pairs with above median working hour overlap (greater than 1.5 hours) and those with below median working hour

[28] Appendix Section C5 contains the original list of leader countries.

overlap (1.5 hours or less). In Appendix Section D1, we also conduct robustness checks showing that a similar pattern emerges when using other cutoffs for high versus low temporal distances. In Appendix Section D2, we also show that the effect of nonstop flights on innovation outcomes is stronger for routes with shorter north-south distances (routes that cross over less longitudinal distance).

[Insert Table 6 here]

Table 6 (Columns 1-4) reports the results. We see that nonstop flights help overcome temporal distance: Nonstop flights increase citations and collaborations for firms in location pairs with below median working hour overlap (long temporal distance), but do not for above median pairs (short temporal distance). A 10% increase in flights for airport pairs with less than 1.5 hours of working hour overlap leads to a 5.25% increase in citations between firms and a 1.87% increase in collaborations between subsidiaries within a firm. For airport pairs with above median working hour overlap, we see the coefficient sizes are smaller and are statistically insignificant at conventional thresholds. Chow test results show that for citations, temporally distant location pairs benefit more than temporally proximate ones (Diff = 0.665, p = 0.003). For collaborations, we cannot reject the null of no difference (Diff = 0.088, p = 0.602) due to the large standard errors for our estimates for low-temporal distance pairs. Overall, nonstop flights do not seem to enable knowledge spillovers and collaborations between inventors that are temporally close to one another.

Second, we test whether the positive effect of nonstop flights on innovation is stronger for firms and subsidiaries separated by large cultural distance. We use two measures of cultural distance. First, we adopt a direct measure of cultural distance at the firm level, by using the ethnic composition of the inventors who cite one another at two geographically distant firms and that of the inventors who collaborate with one another at two geographically distant subsidiaries within a firm. We determine inventors' ethnicities using NamePrism, a tool based on machine learning algorithms that accurately predict a person's ethnicity based on the full name. Inventors in a citing or collaborating relationship are deemed co-ethnic if they share an ethnicity and multiethnic if there is more than one ethnicity in that relationship. Then, we compare the effect of nonstop flights on co-ethnic innovation and that on multiethnic innovation. Table 6 (Columns 5-8) shows that for inventor pairs that share the same ethnicity (low ethnic distance), co-ethnic collaborations increase by 0.87%, while co-ethnic citations increase by 1.43%. However, for inventor pairs with different ethnicities (high ethnic distance), multiethnic collaborations increase by 1.30% and citations by 3.54%. A Chow test for difference of coefficients shows that co-ethnic citations benefit more than multiethnic citations (Diff = 0.154, p = 0.001). Results for collaborations lack statistical significance (Diff = 0.043, p = 0.124). In summary, we find some evidence for the relationship that nonstop flights tend to facilitate innovation across countries for firms that are culturally distant.[29]

Next, we construct another measure for cultural distance. We utilize the data from the 2020 WVS to gauge individuals' attitudes toward foreign workers. Berry et al. (2010) and others have adopted the WVS data to create indices of cultural distance between countries. Instead of using an index, the interpretation of which is difficult to decipher, we utilize one specific question on the WVS that asks respondents to "evaluate the impact of immigrants on the development of your country," for which the answer choices range from "very bad" to "very good." We then aggregate the answers at the country level and obtain a value for each country, which we then match to the data on flights and patenting. The assumption is that immigrant friendliness is a good proxy for individuals' ability to appreciate the work of foreigners—a cultural dimension that is highly relevant to our context given that it relates to cultural tightness (Chua et al. 2015). We divide our sample into location pairs that are (1) both immigrant friendly, (2) both immigrant unfriendly, and (3) one immigrant friendly and one immigrant unfriendly. We find that nonstop flights increase citations and collaborations among firms in locations that are marked by a high degree of cultural distance: where one location is immigrant friendly and another location is immigrant *unfriendly*. Interestingly, where both locations are unfriendly to immigrants, nonstop flights do not facilitate innovation. Table 7 presents results of the regression discontinuity subsample analysis. For firms in location pairs with high cultural distance (in terms of friendliness toward immigrants), a 10% increase in nonstop flights leads to a 7.2% increase in citations. Finally, Appendix Section D4 presents an alternative immigrant analysis using firms' labor condition

[29] In addition to cultural distance, firm boundaries may serve to amplify institutional distance. Thus, nonstop flights may have different effects, depending on whether those citations are cross-assignee or within-assignee. However, as Appendix Section B10 shows, most citations (more than 95%) are cross-assignee, limiting our ability to test for differences.

applications (LCAs) as a proxy for firms' employment of immigrants, showing that nonstop flights drive knowledge diffusion among firms with high levels of immigrant labor.

[Insert Table 7 here]

5. Additional Analyses

5.1. Airport-level Instrument

Our regression discontinuity results show that at the airport pair level, pairs slightly below 6,000 miles apart have increased knowledge flows. However, an open question is whether this result extends more globally above and beyond the 6,000-mile threshold. One specific concern is that measurements at the airport pair level may be confounded by *redirection* of knowledge flows from other airport pairs to the focal pair. We analyze how direct flights affect citations and collaborations at a single airport. This approach calculates the net effect and mitigates concerns of compositional changes in innovation outcomes.

We implemented an instrumental variable-based identification strategy proposed by Campante and Yanagizawa-Drott (2018) to extend the results from the airport *pair* level to the airport level. At the airport level, we use exogenous variation in that airport's connectedness to measure its impact on the number of publications and citations. Variation in an airport's connectedness stems from the cost of operating 6,000+ mile flights: Airports with many other "potential" airports slightly less than 6,000 miles apart will be more "connected" in terms of number of flights. We use an instrumental variable approach where the share of airports slightly below 6,000 miles shifts the total number of realized connected airports, thus impacting innovation. Our first- and second-stage regressions are as follows.

$$ConnectedAirports_i^t = \alpha_0 + \alpha_1 ShareBelow6K_i + X_i + \varepsilon_i \tag{6}$$

$$Y_i^t = \beta_0 + \beta_1 \, \widehat{ConnectedAirports}_i^t + X_i + \varepsilon_i \tag{7}$$

$ConnectedAirports_i^t$ measures the number of airports with which airport i has a nonstop flight in year t. $ShareBelow6K_i$ counts the total number of airports (connected or unconnected) slightly below 6,000 miles and divides this by the total number of airports (again, connected or unconnected) around 6,000 miles. Y_i^t is our dependent variable, either the number of firms or academic publications near airport i in year t or the number of citations to those patents.[30] X_i includes control variables, including the total number of airports near 6,000 miles for airport i, its distance from the equator, and the time zone difference from GMT, as well as region fixed effects.[31]

[Insert Table 8 here]

Generally, we find that more connected airports lead to more citations and publications. Table 8 shows that an additional connected airport in 2015 increases the total number of citations by about 12.7% and the total number of publications by 10.6% for firms.[32] Given that the median number of connected airports in our sample is 4 (mean = 15.87, s.d. = 42.92), the economic magnitude of connectivity seems to be quite significant.[33] In Appendix Section F, we break down the effects across different years and find similar results. Appendix Section F also tests how connectivity affects the number of collaborators and duration and finds that both increase, suggesting the existence of intensive and extensive margins.

5.2. Mechanisms: Ticket Prices and Flight Duration

We next turn to the mechanisms by which nonstop flights affect collaborations and citations in our context. The two mechanisms of interest relate to travel cost reductions (e.g., Catalini et al. 2020) and travel time reductions (e.g., Bernstein et al. 2016). Advances in technology and increased competition have significantly impacted both the monetary cost as well as the time it takes to transport people. For instance, a typical trip from Los Angeles to Boston cost about $4,500 in 1941 (2015 dollars) and took more than 15 hours across 12

[30] We use the number of connected airports in 2015 and sum the number of publications and citations across all years in our patent sample (2005-2015). Appendix Section F1 shows our findings are robust to using alternate years.

[31] Region data from World Bank Development Indicators which categorize countries into seven regions: East Asia and Pacific, Europe and Central Asia, Latin America and the Caribbean, Middle East and North Africa, North America, South Asia, and Sub-Saharan Africa. Note that we are unable to add country-country fixed effects in this setting because the nature of the sample has no bilateral dimension (each observation is an airport).

[32] Since our dependent variables are inverse hyperbolic sine transformed, we can interpret them as log-transformed variables. Therefore, we use exp{coefficient}-1 to calculate the effect sizes.

[33] In general equilibrium, when looking at highly aggregated data, we cannot rule out the possibility of some firms self-selecting to be located near the most connected airports. However, our empirical design exploits exogenous variation to estimate the connectedness of airports, and therefore, given our assumptions, our estimates using the regression discontinuity framework are causal. Moreover, a longstanding literature in strategy outlines how firm location choice is driven by considerations such as access to resources (Alcacer and Chung 2007), location of competitors (Ghemawat and Thomas 2008), among others; there is less evidence on firms actively using closeness to connected airports as a key determinant of their location decisions.

stops. In contrast, in 2015, the same route could be travelled nonstop for just $480 (just 11% of the 1941 cost), and it would take only six hours (Garcia 2017).

Determining whether nonstop flights impact innovation across countries through travel duration or ticket price has important implications for theory and for informing policy. Prior work on collaboration over distance has focused primarily on travel costs as a key barrier to collaboration. For instance, pecuniary travel costs and the importance of face-to-face interactions in facilitating collaborative outcomes can explain why distance still matters for collaboration (Catalini et al. 2020). However, for many innovators, "time famine"—that is, a shortage of time—is a more salient constraint (Perlow 1999), especially for firm-employed inventors who, relative to academic inventors, might care more about time spent on international journeys than about monetary costs of travel. Delving into the components of travel costs is crucial to understanding the existence of barriers to collaboration other than, for instance, geopolitical borders (Singh and Marx 2013).

For each airport pair, we obtain the average ticket price and duration of travel sourced from Google Flights. Specifically, from January 5 to February 6, 2020, we queried Google Flights for flights leaving Thursday, June 18, 2020, and returning Sunday, June 21, 2020. The mid-June dates were chosen to represent a typical conference weekend.[34] Our query window for flight price and duration information (i.e., from January 5 to February 6, 2020) was before most COVID-induced travel bans took place (e.g., travel bans to and from China took place mostly in late January and early February 2020), which suggests that our flight data are representative of data that would be obtained during normal, non-pandemic times. We obtained ticket information for airport pairs that are more than 3,000 miles apart and that have more than 1,000 flights between them from 2005 to 2015 (i.e., 100 flights per year, or about one weekly round-trip flight), for a total of 3,708 routes. With this information, we calculate the average ticket price and the average time duration of all routes between an airport pair. By constructing the dataset as described above, we assume that flight durations and prices in 2020, adjusted for inflation, are similar to what they would have been in our sample period, 2005 to 2015.

[34] For instance, in 2019, the Association of Clinical Research Professionals Annual Meeting was held April 12-15 (Friday to Monday), and USPTO's Inventors Conference was held September 13-14 (Friday to Saturday).

A major appeal of nonstop flights is their ability to decrease flight time significantly, and customers frequently pay extra to take nonstop flights. Figure 3 plots how prices and travel duration vary with number of stops using the data we gathered from Google Flights. Each point on the graph represents the average difference in price and travel time between a nonstop flight and a flight with stops. The left panel shows that a one-stop flight is, on average, 5.1% cheaper than a nonstop flight. This constitutes an average price difference of about $40. However, the average one-stop flight is about 53% longer in terms of travel time than a nonstop flight, a time difference of about 5.8 hours. This trade-off between more expensive tickets and shorter flight durations is stronger for long-distance flights (Appendix Section G shows that the magnitude of the coefficient on duration increases for airport pairs more than 6,000 miles apart).

[Insert Figure 3 here]

To shed light on the relationship between firms' innovation activities and ticket prices/travel duration, we estimate the number of collaborations and citations using the following specification:

$$\begin{aligned} \operatorname{asinh}(Y_{a_o,a_d}) = \beta_0 + \beta_1 \operatorname{asinh}\left(Price_{a_o,a_d}\right) + \beta_2 \operatorname{asinh}\left(Duration_{a_o,a_d}\right) \\ + \delta \operatorname{asinh}\left(X_{a_o,a_d}\right) + \eta_{c_{a_o},c_{a_d}} + \epsilon_{a_o,a_d} \end{aligned} \quad (8)$$

where Y_{a_o,a_d} measures the total number of collaborations or citations between a_o and a_d. $Price_{a_o,a_d}$ is the average ticket price for flights between a_o and a_d, and $Duration_{a_o,a_d}$ measures the average number of hours for those flights.[35] In all specifications, we control for the distance between airport pairs $\left(X_{a_o,a_d}\right)$, as well as country-country fixed effects $\left(\eta_{c_{a_o},c_{a_d}}\right)$, which control for between-country differences such as language and time zone differences. We cluster our standard errors at the country-country level.

We present estimates of the specification in Table 9. Across all specifications, we see that flight duration has a negative and significant partial correlation with collaborations and citations. In our preferred specification for collaborations (Column 3), a 10% increase in duration (1.3 hours) is associated with an 8.9% decrease in collaborations (1.8 fewer collaborations per route, across the entire sample period). Similarly, for

[35] Our results are robust whether we average for direct and one-stop flights or whether we consider averages across all flights (2+ stops).

citations (Column 4), a 10% increase in duration is associated with an 8.7% decrease in citations (1.9 fewer citations per route). However, while prices are negatively correlated with citations and collaborations, the coefficients are not statistically distinguishable from zero for collaborations. This result hints that it is a cost—not price—reduction in terms of the time duration of nonstop flights that would facilitate the collaborative production of innovation and the diffusion of knowledge across firms globally.

[Insert Table 9 here]

6. Discussion and Conclusion

Firms continue to benefit from global knowledge diffusion and the production of global collaborative patents, but national borders remain relevant as a source of friction (Singh and Marx 2013). Alcacer et al. (2017) detailed how "figurative distances" stemming from political borders were creating frictions that impeded knowledge collaboration and spillovers, and Aguilera et al. (2019) bemoaned the de-globalization trend. In the context of this prior literature, this paper shows how in the global context, geographic mobility of individuals through nonstop flights boosts the diffusion of knowledge through patent citations and collaboration of inventors, especially at firms. To provide causal evidence, we use an RDD framework and find that a 10% increase in the number of nonstop flights between two locations increases citations by 3.4% and collaborations by 1.4%. This positive effect is driven primarily by firms, as opposed to academic institutions. We find the effects to be more salient at firms/subsidiaries: (1) with more inventors and R&D spending; (2) located in hubs or countries deemed as technology leaders; and (3) that are located in culturally distant or temporally distant places.

Our study has several limitations. Similar to Catalini et al. (2020), Bernstein et al. (2016), and most prior studies on airline connectivity and innovation/economic outcomes, we do not observe individuals traveling between locations and instead impute travel patterns by aggregating citations and collaborations to the level of airport pairs. One consequence of this is that we are unable to disaggregate who is travelling and for what reason.[36] A recent McKinsey report (Curley et al. 2020) documents that in 2018, international airline business travel spending exceeded $1.4 trillion; this travel encompassed transient travel and travel for meetings, incentives, conferences, and events (MICE), from large group offsite gatherings to industry-wide exhibitions. Future research should attempt to disaggregate the effects of airline travel for company meetings versus airline travel for attending conferences and contribute to the literature on temporary colocation and innovation outcomes (Boudreau et al. 2017, Chai and Freeman 2019). Additionally, there is an increasing adoption of alternative work arrangements and communication technologies such as Zoom (which might come to characterize the post-COVID world (Marr 2021)); future research may study whether the effects of transportation connectivity on innovation across countries are weakened or strengthened by the use of new communication technologies. Finally, due to data limitations, we examine only USPTO patents, while not including patents issued by other patent offices such as the European Patent Office or the Japan Patent Office. Even though our results hold while limiting the sample to airport pairs with at least one U.S. city, the analysis would be more complete if it included patents from the rest of the world.

Despite these limitations, our findings contribute to several streams of the strategy and innovation literature—notably the literature on connectivity/geographic mobility and economic/innovation outcomes. We contribute to this literature by showing whether, how, and when mitigating travel constraints can foster greater knowledge diffusion in a global setting (Agrawal et al. 2017, Baum-Snow 2007, Duranton and Turner 2012, Ghani et al. 2016). Our findings are related to Agrawal et al. (2017), who exploit historical data on planned highways, railroads, and exploration routes as sources of exogenous variation in order to estimate the effect of interstate highways on regional innovation.[37] Notably, we highlight the scope conditions of greater connectivity fostering innovation outcomes and document that for firms with relatively lower innovation

[36] To the best of our knowledge, the only prior study that uses actual international flight travel data for individuals is Choudhury (2017). Business travelers may include both managers who do not participate in patenting and firm-employed inventors who participate in patenting. We are unable to ascertain which types of travelers contributed more to outcomes of interest, and policy implications may differ depending on the traveler type.

[37] More broadly, our findings are also relevant to the literature on international labor mobility and knowledge diffusion (Agrawal et al. 2006, 2011, Almeida and Kogut 1999, Bahar et al. 2020, Choudhury 2016, Choudhury and Kim 2019, Foley and Kerr 2013, Ghani et al. 2014, Hovhannisyan and Keller 2015, Kapur 2001, Kapur and McHale 2005a, 2005b, Kerr 2008, MacGarvie 2006, Nanda and Khanna 2010, Obukhova 2009, Oettl and Agrawal 2008, Papageorgiou and Spilimbergo 2009, Rosenkopf and Almeida 2003, and Singh 2005).

mass, firms/subsidiaries that are located outside hubs/countries that represent the technological frontier, and for firms and firm locations that are culturally or temporally proximate, adding nonstop flights is less likely to enhance innovation outcomes. This result sheds light on why it may be difficult for firms and inventors in some "follower" countries to get to the technological frontier.

A relevant paper in this literature is Catalini et al. (2020), which uses a difference-in-differences empirical strategy combined with a series of robustness and falsification tests to document that the availability of cheaper options for airline travel has a causal effect on the probability, intensity, and direction of collaborations among academic scientists. However, while Catalini et al. (2020)'s study focuses on academic scientists within the United States (for whom temporal and cultural distance with collaborators or being located in a country that is a technological follower might be less salient) and focuses on savings in travel costs as the underlying mechanism, our study focuses on firms, how characteristics of firms/subsidiaries and their locations matter for the relation between nonstop flights and innovation outcomes across countries, and documents savings in travel time as the underlying mechanism.

Our findings also contribute to the literature on knowledge spillovers and collaborative patents for firms in the global context. Branstetter et al. (2014) document that multinational corporations (MNCs) from advanced economies are largely responsible for the "exponential" growth in U.S. patents filed from China and India. Kerr and Kerr (2018) cite analysis from the Bureau of Economic Analysis to state that the share of R&D for U.S. MNCs conducted by foreign subsidiaries rose from 6% in 1982 to 14% in 2004. Our findings contribute to this literature by outlining an important mechanism—that is, international travel and flight connectivity—that facilitates knowledge flows and GCP production across countries. To quote Kerr and Kerr (2018 p.F268), "[the] use of cross-border teams is a very attractive technique for multinationals conducting innovation abroad and careful thought by nations about short-term travel policies…may have a big impact as multinationals weigh their options." Our findings speak directly to this assertion and provide empirical evidence for whether, how, and when nonstop flights facilitate GCP production.

Finally, we contribute to the international business literature on distance. That research shows that interfirm alliances and employee geographic mobility create "bridges to distant contexts" that mitigate the constraints of geographically localized search for knowledge and collaborators (Rosenkopf and Almeida 2003). Our study suggests that flight connectivity is an important facilitator for firms to build bridges to distant contexts, but the effectiveness of the bridges depends on the characteristics of the firms and of the contexts being connected. Similarly, scholars have shown that temporal distance and a lack of working hour overlap impede knowledge-intensive communication in firms (Bahar 2020). Our study suggests that nonstop flights may feasibly overcome the temporal barrier and facilitate the spread of knowledge across temporally distant firms and subsidiaries within a firm by bringing individuals face-to-face. Another possibility is that nonstop flights may have facilitated knowledge diffusion across global firms by allowing inventors to work from similar time zones within business hours, while not necessarily meeting face-to-face. Appendix Section D3 offers support for the mechanism of similar-time zone work by showing that nonstop flights enable knowledge flows for firms that are highly temporally distant (8+ hours difference). This result is not conclusive, and it warrants more research to further unpack the difference between the mechanism of face-to-face meetings versus that of similar-time zone work. Additionally, our study relates to the interplay between geographic distance and non-geographic distances. Whereas geographic distance physically limits knowledge flows, non-geographic factors such as cultural frictions also constrain the point of contact and hamper interactions (Shenkar 2001, Shenkar et al. 2008). Our study suggests conditions under which geographic distance may not be a friction—that is, if firms and firm locations are culturally similar or temporally proximate, firms may not need a physical bridge (e.g., through nonstop flights) to exchange knowledge or collaborate.

Our study suggests several additional directions for future research. First, future research should explore whether the introduction of synchronous and asynchronous communication technology substitutes for or complements airline travel. Second, future research should explore the importance of global immigration policies as they relate to airline travel and how that affects the utility of choosing a nonstop flight versus a flight with more stops. In other words, while a nonstop flight avoids the need for securing a "transit visa," such visas might be salient for global travel that involves stopovers (O'Keefe 1993). Finally, future research should study the importance of global airline travel in an era of increasing distributed work and "work from anywhere" (WFA) (Choudhury et al. 2021). It would be interesting to study whether the importance of airline connectivity, travel, and temporary colocation increases when more firms adopt WFA and when workers become more globally distributed.

Our study also has several managerial and policy implications—notably that business travel to culturally and temporally distant places might be beneficial for innovation outcomes at firms with large innovation masses, especially when the travel connects two hubs. For decades, airports and policy makers have offered incentives to airlines to start nonstop flights.[38] Our study provides useful evidence for when policy makers should design incentives to attract airlines to start nonstop flights. Our study also points to the importance of business travel for fostering innovation and suggests conditions for when such travel might be more effective.[39] For example, direct flights may disproportionately benefit firms with greater innovation mass, compared to universities and smaller firms. As the McKinsey report published by Curley et al. (2020) documents, while business travel spending exceeded $1.4 trillion in 2018, "historically, business travel has been more volatile and slower to recover than leisure travel after economic downturns and other disruptions to travel patterns." Our study indicates that if indeed international flights exhibit a slow recovery in the aftermath of pandemics and economic downturns, cross-border knowledge spillovers and collaborations at some firms could be adversely affected.

In conclusion, this paper presents, to the best of our knowledge, the first set of causal evidence and boundary conditions for whether, when, and how nonstop flights positively affect firm innovation in a global context. Using unique data and a two-pronged empirical approach (including a cutting-edge RDD and tests of firm and firm location heterogeneity using a modified gravity model), we shed light on whether, when, and

[38] From 2012 to 2014, regional airports in the U.S. spent in excess of $171 million in incentives to attract new routes (CAPA 2018). Whether or not to operate nonstop flights between airports is a topic of active managerial discussion (Routes 2019). Many airports offered incentives to attract new international flights: Hartsfield-Jackson Atlanta International Airport offered to waive landing fees (Williams 2014); Tampa International Airport offered cash and airport fee waivers to attract Edelweiss (Thalji 2013); Indianapolis International Airport offered Delta $5.5 million in conditional incentives (Lange and Cook 2017); Pittsburgh was able to attract an international flight to London by offering British Airways $3 million in funding over two years (Belko 2019); New Orleans waived landing fees for British Airways (Buchanan 2016); and so forth. Cities in Ohio were unable to attract international airlines (Glaser 2018), even trying to kickstart local airlines that would serve international locations (Teasley 2018). Globally, Greece launched a fee waiver program to attract international routes during the winter season (GTP Editing Team 2018).

[39]

how nonstop flights affect knowledge spillovers (citations) and collaborations (GCP production) for firms in a global setting. Our study contributes to the literatures on connectivity/geographic mobility and innovation outcomes, knowledge spillovers and collaborative patents for firms in the global context, and how cultural and temporal distances affect innovation across countries. Finally, it provides policy and managerial implications on the value of business travel.

References

Agrawal A, Cockburn I, McHale J (2006) Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships. *Journal of Economic Geography* 6(5):571–591.

Agrawal A, Galasso A, Oettl A (2016) Roads and Innovation. *The Review of Economics and Statistics* 99(3):417–434.

Agrawal A, Galasso A, Oettl A (2017) Roads and innovation. *Review of Economics and Statistics* 99(3):417–434.

Agrawal AK, Catalini C, Goldfarb A (2011) *The geography of crowdfunding* (National bureau of economic research).

Aguilera R, Henisz W, Oxley JE, Shaver JM (2019) Special issue introduction: International strategy in an era of global flux. *Strategy Science* 4(2):61–69.

Alcacer J, Chung W (2007) Location strategies and knowledge spillovers. *Management science* 53(5):760–776.

Alcácer J, Kogut B, Thomas C, Yeung B (2017) Geography, Location, and Strategy. Retrieved (August 18, 2019), https://www.emerald.com/insight/content/doi/10.1108/S0742-332220170000036002/full/html.

Almeida P, Kogut B (1999) Localization of knowledge and the mobility of engineers in regional networks. *Management science* 45(7):905–917.

Anderson JE (1979) A theoretical foundation for the gravity equation. *The American economic review* 69(1):106–116.

Anon Routes Europe 2019 | Routesonline. *Routes Online*. Retrieved (July 20, 2020), https://www.routesonline.com/events/203/routes-europe-2019/.

Arora A, Belenzon S, Sheer L (2021) Knowledge spillovers and corporate investment in scientific research. *American Economic Review* 111(3):871–98.

Bahar D (2020) The hardships of long distance relationships: time zone proximity and the location of MNC's knowledge-intensive activities. *Journal of International Economics* 125:103311.

Bahar D, Choudhury P, Rapoport H (2020) Migrant inventors and the technological advantage of nations. *Research Policy* 49(9):103947.

Baum-Snow N (2007) Did highways cause suburbanization? *The quarterly journal of economics* 122(2):775–805.

Belko M (2019) Too much to ignore — British Airways takes a chance on a resurgent Pittsburgh. *Pittsburgh Post-Gazette*. Retrieved (July 20, 2020), https://www.post-gazette.com/business/development/2019/03/27/British-Airways-Pittsburgh-International-Airport-Christina-Cassotis/stories/201903270082.

Bernstein S, Giroud X, Townsend RR (2016) The Impact of Venture Capital Monitoring. *The Journal of Finance* 71(4):1591–1622.

Berry H, Guillén MF, Zhou N (2010) An institutional approach to cross-national distance. *J Int Bus Stud* 41(9):1460–1480.

Bikard M, Marx M (2020) Bridging academia and industry: How geographic hubs connect university science and corporate technology. *Management Science* 66(8):3425–3443.

Bircan C, Javorcik B, EBRD O, Pauly CS (2021) Creation and Diffusion of Knowledge in the Multinational Firm.

Boudreau KJ, Brady T, Ganguli I, Gaule P, Guinan E, Hollenberg A, Lakhani KR (2017) A field experiment on search costs and the formation of scientific collaborations. *Review of Economics and Statistics* 99(4):565–576.

Branstetter L, Li G, Veloso F (2014) The rise of international coinvention. *The changing frontier: Rethinking science and innovation policy*. (University of Chicago Press), 135–168.

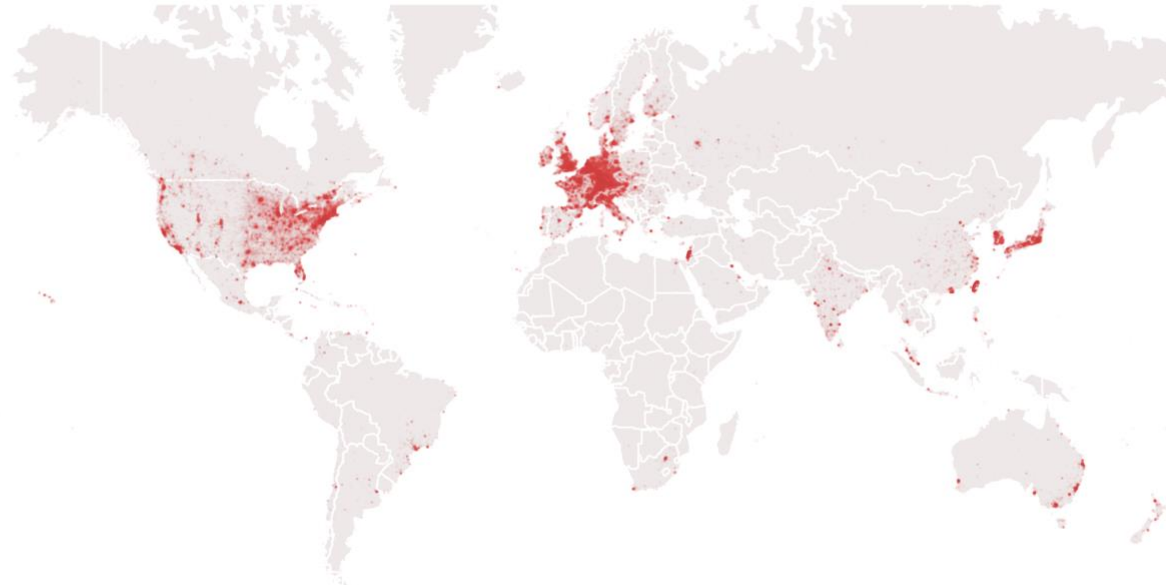
Bresnahan T, Gambardella A, Saxenian A (2001) ‘Old economy’inputs for ‘new economy’outcomes: Cluster formation in the new Silicon Valleys. *Industrial and corporate change* 10(4):835–860.
Buchanan S (2016) New Orleans attracts first, non-stop European flights in decades | New Orleans’ Multicultural News Source | The Louisiana Weekly. *The Louisiana Weekly*.
Calonico S, Cattaneo MD, Farrell MH (2020) Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal* 23(2):192–210.
Calonico S, Cattaneo MD, Titiunik R (2014) Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica* 82(6):2295–2326.
Campante F, Yanagizawa-Drott D (2018) Long-Range Growth: Economic Development in the Global Network of Air Links. *Q J Econ* 133(3):1395–1458.
CAPA (2018) US smaller airport using incentives to attract new LCC services | CAPA. Retrieved (July 20, 2020), https://centreforaviation.com/analysis/reports/us-smaller-airport-using-incentives-to-attract-new-lcc-services-409846.
Catalini C, Fons-Rosen C, Gaulé P (2020) How Do Travel Costs Shape Collaboration? *Management Science*.
Chai S, Freeman RB (2019) Temporary colocation and collaborative discovery: Who confers at conferences. *Strategic Management Journal* 40(13):2138–2164.
Chauvin J, Choudhury P, Pang TF (2020) *The Effects of Temporal Distance on Intra-Firm Communication: Evidence from Daylight Savings Time* (Harvard Business School).
Choudhury P (2016) Return migration and geography of innovation in MNEs: a natural experiment of knowledge production by local workers reporting to return migrants. *Journal of Economic Geography* 16(3):585–610.
Choudhury P (2017) Innovation outcomes in a distributed organization: Intrafirm mobility and access to resources. *Organization Science* 28(2):339–354.
Choudhury P, Foroughi C, Larson B (2021) Work-from-anywhere: The productivity effects of geographic flexibility. *Strategic Management Journal* 42(4):655–683.
Choudhury P, Kim DY (2019) The ethnic migrant inventor effect: Codification and recombination of knowledge across borders. *Strategic Management Journal* 40(2):203–229.
Chua RY, Roth Y, Lemoine JF (2015) The impact of culture on creativity: How cultural tightness and cultural distance affect global innovation crowdsourcing work. *Administrative Science Quarterly* 60(2):189–227.
Curley A, Garber R, Krishnam V, Tellez J (2020) For corporate travel, a long recovery ahead. *McKinsey & Company*. Retrieved (August 28, 2020), https://view.ceros.com/mckinsey/coronavirus-promo-video-desktop.
Duranton G, Turner MA (2012) Urban growth and transportation. *Review of Economic Studies* 79(4):1407–1440.
Espinosa JA, Nan N, Carmel E (2015) Temporal distance, communication patterns, and task performance in teams. *Journal of Management Information Systems* 32(1):151–191.
Federal Aviation Administration (2016) Economic Values for FAA Investment and Regulatory Decisions, a Guide--Final Report. Retrieved (March 4, 2021), https://www.faa.gov/regulations_policies/policy_guidance/benefit_cost/.
Florida R (2005) *Cities and the creative class* (Routledge).
Foley CF, Kerr WR (2013) Ethnic Innovation and U.S. Multinational Firm Activity. *Management Science* 59(7):1529–1544.
Frankel J, Rose A (2002) An estimate of the effect of common currencies on trade and income. *The quarterly journal of economics* 117(2):437–466.
Furman JL, Hayes R (2004) Catching up or standing still?: National innovative productivity among ‘follower’countries, 1978–1999. *Research policy* 33(9):1329–1354.

Garcia M (2017) What Flights Used to Cost in the "Golden Age" of Air Travel. *Travel + Leisure*. Retrieved (July 12, 2020), https://www.travelandleisure.com/airlines-airports/history-of-flight-costs.
Gaspar J, Glaeser EL (1998) Information technology and the future of cities. *Journal of urban economics* 43(1):136–156.
Ghani E, Goswami AG, Kerr WR (2016) Highway to success: The impact of the Golden Quadrilateral project for the location and performance of Indian manufacturing. *The Economic Journal* 126(591):317–357.
Ghani E, Kerr WR, Stanton C (2014) Diasporas and Outsourcing: Evidence from oDesk and India. *Management Science* 60(7):1677–1697.
Ghemawat P (2001) Distance still matters. *Harvard business review* 79(8):137–147.
Ghemawat P (2007) *Redefining global strategy: Crossing borders in a world where differences still matter* (Harvard Business Press).
Ghemawat P, Thomas C (2008) Strategic interaction across countries and multinational agglomeration: An application to the cement industry. *Management Science* 54(12):1980–1996.
Giroud X (2013) Proximity and Investment: Evidence from Plant-Level Data. *The Quarterly Journal of Economics* 861:915.
Glaser S (2018) Should Ohio pay to attract new international flights? Pittsburgh did. *cleveland.com*. Retrieved (July 20, 2020), https://www.cleveland.com/travel/2018/08/should_ohio_pay_to_attract_new.html.
GTP Editing Team (2018) Fraport: Incentive Program to Attract New Int'l Routes to Greece During Winter. *GTP Headlines*. Retrieved (July 20, 2020), https://news.gtp.gr/2018/06/14/fraport-greece-incentive-program-attract-new-intl-routes-greece-winter/.
Hofstede G (1980) *Culture's consequences: International differences in work-related values* (Sage Publications, Inc).
Hovhannisyan N, Keller W (2015) International business travel: an engine of innovation? *J Econ Growth* 20(1):75–104.
Imbens G, Kalyanaraman K (2012) Optimal Bandwidth Choice for the Regression Discontinuity Estimator. *Rev Econ Stud* 79(3):933–959.
Kapur D (2001) Diasporas and technology transfer. *Journal of Human development* 2(2):265–286.
Kapur D, McHale J (2005a) *Give us your best and brightest: The global hunt for talent and its impact on the developing world* (Center for Global Development Washington, DC).
Kapur D, McHale J (2005b) Mobile human capital and high-tech industry development in India, Ireland, and Israel. *From underdogs to tigers: The rise and growth of the software industry in Brazil, China, India, Ireland, and Israel* 236.
Kerr SP, Kerr WR (2018) Global Collaborative Patents. *Econ J* 128(612):F235–F272.
Kerr WR (2008) Ethnic scientific communities and international technology diffusion. *The Review of Economics and Statistics* 90(3):518–537.
Kerr WR, Robert-Nicoud F (2020) Tech clusters. *Journal of Economic Perspectives* 34(3):50–76.
Kharina A, Rutherford D (2015) Fuel efficiency trends for new commercial jet aircraft: 1960 to 2014.
Kim JU, Aguilera RV (2015) The world is spiky: An internationalization framework for a semi-globalized world. *Global Strategy Journal* 5(2):113–132.
Lange K, Cook T (2017) Indianapolis International Airport to offer nonstop flight to Paris. *The Indianapolis Star*. Retrieved (July 20, 2020), https://www.indystar.com/story/news/politics/2017/09/06/indianapolis-international-airport-offer-first-transatlantic-flight-paris/635535001/.

Lazear EP (1999) Globalisation and the market for team-mates. *The Economic Journal* 109(454):15–40.
Lee DS, Lemieux T (2010) Regression Discontinuity Designs in Economics. *Journal of Economic Literature* 48(2):281–355.
MacGarvie M (2006) Do firms learn from international trade? *Review of Economics and Statistics* 88(1):46–60.
MacKinnon JG, Magee L (1990) Transforming the dependent variable in regression models. *International Economic Review*:315–339.
Marr B (2021) How Technology Will Unlock New Ways Of Working In A Post-Covid World. *Forbes*. Retrieved (March 16, 2021), https://www.forbes.com/sites/bernardmarr/2021/03/01/how-technology-will-unlock-new-ways-of-working-in-a-post-covid-world/.
McKenney D, Prusak K, Robichaux L, Tweeden W (2000) United Airlines Crew Agumentation Study.
Mell JN, Jang S, Chai S (2021) Bridging temporal divides: Temporal brokerage in global teams and its impact on individual performance. *Organization Science* 32(3):731–751.
Miguelez E, Fink C (2013) Measuring the international mobility of inventors: A new database. *World Intellectual Property Organization-Economics and Statistics Division*.
Mize TD, Doan L, Long JS (2019) A General Framework for Comparing Predictions and Marginal Effects across Models. *Sociological Methodology* 49(1):152–189.
Montobbio F, Sterzi V (2013) The globalization of technology in emerging markets: a gravity model on the determinants of international patent collaborations. *World development* 44:281–299.
Nanda R, Khanna T (2010) Diasporas and domestic entrepreneurs: Evidence from the Indian software industry. *Journal of Economics & Management Strategy* 19(4):991–1012.
Nardi BA, Whittaker S (2002) The place of face-to-face communication in distributed work. *Distributed work* 83(112):10–7551.
Obukhova E (2009) *Does brain circulation promote international development? High-skilled migration and organizational performance* (Citeseer).
Oettl A, Agrawal A (2008) International labor mobility and knowledge flow externalities. *Journal of international business studies* 39(8):1242–1260.
O'Keefe C (1993) Immigration Issues and Airlines. *J. Air L. & Com.* 59:357.
Papageorgiou C, Spilimbergo A (2009) *Learning abroad and technology adoption* (Mimeo., IFM, Washington DC).
Perlow LA (1999) The time famine: Toward a sociology of work time. *Administrative science quarterly* 44(1):57–81.
Picci L (2010) The internationalization of inventive activity: A gravity model using patent data. *Research Policy* 39(8):1070–1081.
Rosenkopf L, Almeida P (2003) Overcoming local search through alliances and mobility. *Management science* 49(6):751–766.
Shenkar O (2001) Cultural distance revisited: Towards a more rigorous conceptualization and measurement of cultural differences. *Journal of international business studies* 32(3):519–535.
Shenkar O, Luo Y, Yeheskel O (2008) From "distance" to "friction": Substituting metaphors and redirecting intercultural research. *Academy of management review* 33(4):905–923.
Singh J (2005) Collaborative networks as determinants of knowledge diffusion patterns. *Management science* 51(5):756–770.
Singh J, Marx M (2013) Geographic Constraints on Knowledge Spillovers: Political Borders vs. Spatial Proximity. *Management Science* 59(9):2056–2078.
Teasley C (2018) Transportation: Columbus' Never-Ending Quest for Direct Flights. *Columbus CEO*. Retrieved (July 20, 2020),

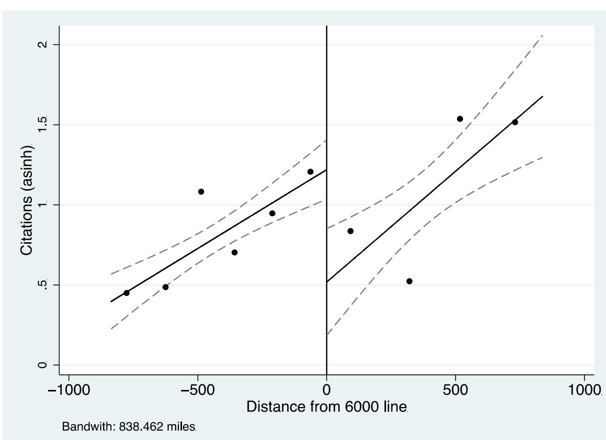
https://www.columbusceo.com/business/20181029/transportation-columbus-never-ending-quest-for-direct-flights.
Thalji J (2013) Edelweiss Air route a success story for Tampa International Airport. *Tampa Bay Times*. Retrieved (July 20, 2020), https://www.tampabay.com/news/business/airlines/edelweiss-air-route-a-success-story-for-tampa-international-airport/1268599/.
Vanderkamp J (1977) The gravity model and migration behaviour: An economic interpretation. *Journal of Economic Studies*.
Weed J (2021) Gauging the Prospects for International Travel. *The New York Times* (April 16) https://www.nytimes.com/2021/04/16/travel/international-summer-travel-coronavirus.html.
Williams T (2013) Airport Incentive Program Targets New Routes. *Global Atlanta*. Retrieved (July 20, 2020), https://www.globalatlanta.com/airport-incentive-program-targets-new-routes/.

Figure 1. Inventor Locations

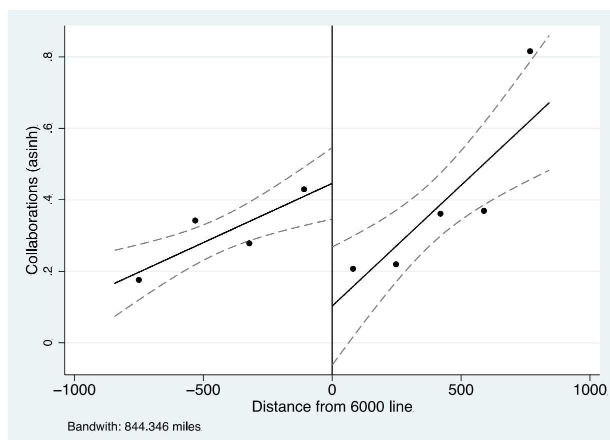


Note: Each inventor location (prior to mapping to the closest airport) is plotted as a red dot. Inventors with the same latitude and longitude are jittered so they are not plotted directly on top of each other.

Figure 2. Impact of Nonstop Flights on Citations and Collaborations



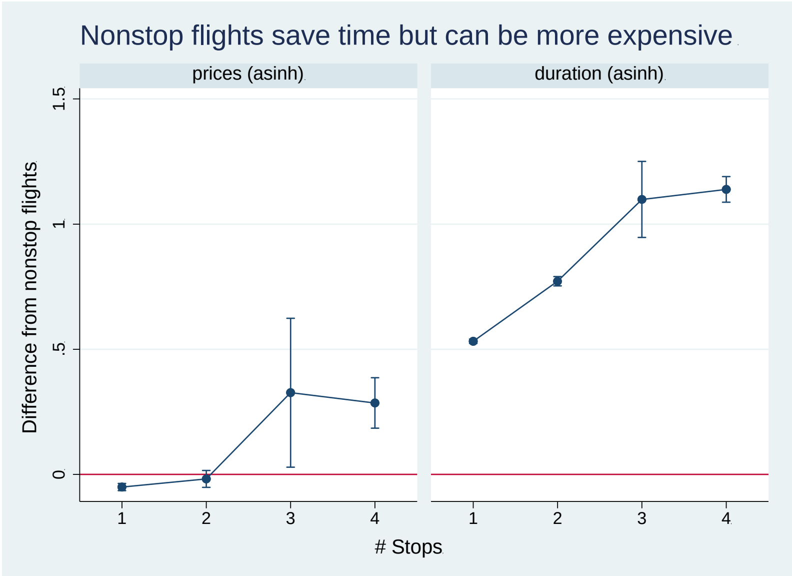
(a) Impact on citations



(b) Impact on collaborations

Note: These graphs use observations based on the optimal bandwidth, which corresponds to 844.35 miles for collaborations and 838.46 miles for citations, at either side of the 6,000-mile threshold. We use a triangular kernel to estimate the optimal bandwidth. It also uses a linear estimator as well as the mimicking variance evenly spaced method to define the number of bins, which results in relatively small bin sizes, reducing the possibility that few outliers on either side drive the discontinuity. Varying the number of bins does not alter the result. Appendix Sections B3-B5 show our graphical results are robust to changing the number of bins, as well as to using quantile-spaced binning methods, kernel choice, and different levels of fixed effects and clustering.

Figure 3. Nonstop Flights are More Expensive than One-stop Flights



Note: Coefficient plot of the relationship between number of stops and flight price and flight duration, obtained from Google Flights in January 2020. Each point measures the percentage difference between ticket price/duration for nonstop flights (omitted category) and flights with a given number of stops. Outcome variables are inverse hyperbolic sine transformed and are interpreted similarly to a log-transformation. Thus, the y-axis measures log-differences in price and duration against nonstop flights.

Table 1. Summary Statistics of the Full Sample

	count	mean	s.d.	min	max
# Citations	538,054	2.14	36.67	0.00	5,702.40
# Collaborations	538,054	2.00	56.47	0.00	7,228.03
# Citations (firms)	538,054	1.96	35.20	0.00	5,498.18
# Collaborations (firms)	538,054	1.90	55.07	0.00	7,048.09
# Citations (academic)	538,054	0.04	0.49	0.00	67.16
# Collaborations (academic)	538,054	0.07	1.41	0.00	239.69
Has nonstop	538,054	0.49	0.50	0.00	1.00
Nonstop flights (count)	538,054	611.95	1,799.85	0.00	74,002.00
Distance (miles)	521,477	1,110.86	1,246.99	0.00	11,873.40
Hub-to-hub flight	537,878	0.26	0.44	0.00	1.00
Working hour overlap	537,878	7.04	1.58	0.00	8.00
Immigrant friendliness distance	402,523	0.19	0.28	0.00	1.87
Average price	20,350	946.48	528.11	214.30	4,538.75

Note: Observations are at the route-year level, excluding average price and average duration, which are at the route-ticket level. Citations and collaborations are inversely weighted based on the number of airports within 50 miles (to avoid double counting). Cross-citations measures the number of citations across different assignees. Distance is in miles. Hub-to-hub measures (whether the origin and destination airports are within 50 miles of an innovation hub) are as defined in Bikard and Marx (2020). Inventor Mass and Publication Mass count the number of inventors/publications in either airport of an airport pair. Immigrant friendliness distance measures the difference between different countries' attitudes toward immigrants, which is a question (Q121) in Wave 7 of the World Value Survey (source: https://www.worldvaluessurvey.org/WVSDocumentationWV7.jsp).

Table 2. Regression Discontinuity: Effect of Nonstop Flights on Innovation in a Global Context is Stronger for Firms than for Academic Institutions

	Overall		Academic Institutions		Firms	
	(1) Citations (asinh)	(2) Collaborations (asinh)	(3) Citations (asinh)	(4) Collaborations (asinh)	(5) Citations (asinh)	(6) Collaborations (asinh)
Nonstop flights (asinh)	0.346***	0.152***	0.099***	0.040**	0.338***	0.148***
	(0.101)	(0.053)	(0.030)	(0.016)	(0.099)	(0.052)
(6,000-Distance)	-0.001**	-0.000	-0.000**	-0.000*	-0.001**	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
(6,000-Distance) x Under6,000	0.001*	0.000	0.001**	0.000**	0.001*	0.000
	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Observations	3,795	3,795	3,795	3,795	3,795	3,795

Note: Observations at the airport pair-year level, excluding singletons. Standard errors in parentheses, clustered at the country pair-year level. Variables inverse hyperbolic sine transformed are denoted by asinh. Bandwidth set at 550 miles. Observations weighted using a triangular kernel. All specifications include country pair-year fixed effects. Distance denotes the geodesic distance (in miles) between airport pairs. Under6,000 is an indicator variable equal to 1 if the distance is less than 6,000.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3. Regression Discontinuity: Effect of Nonstop Flights on Firm Innovation is Stronger for Firms with Greater Innovation "Mass"

	Inventor Mass				R&D Spending			
	Citations (asinh)		Collaborations (asinh)		Citations (asinh)		Collaborations (asinh)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Firm Mass:	High	Low	High	Low	High	Low	High	Low
Nonstop flights (asinh)	0.329***	0.185***	0.141***	0.020	0.180***	0.187***	0.093***	0.031
	(0.099)	(0.051)	(0.051)	(0.013)	(0.052)	(0.054)	(0.030)	(0.028)
(6,000-Distance)	-0.001**	-0.001**	-0.000	-0.000	-0.001***	-0.001***	-0.000*	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
(6,000-Distance) x Under6,000	0.001*	0.001**	0.000	0.000	0.001***	0.001***	0.000	0.000*
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	3,795	3,795	3,795	3,795	3,795	3,795	3,795	3,795

Note: Dependent variables are counts of citations/collaborations for firms of different types, all aggregated to the airport pair-year level. We divide firms into High/Low Inventor Mass firms (above/below median count of inventors, Columns 1-4), and firms with High/Low R&D Spending (above/below median R&D spending, Columns 5-8). Thus, Column (1) uses the number of citations to patents by firms with high inventor mass, while Column (2) uses the number of citations to patents by firms with low inventor mass, as dependent variables. Inverse hyperbolic sine transformed variables are denoted by asinh. RD Bandwidth set at 550 miles. All specifications include country pair-year fixed effects. R&D Spending for a firm is obtained through the DISCERN dataset (Arora et al. 2021). Distance denotes the geodesic distance (in miles) between airport pairs. Under6,000 is an indicator variable equal to 1 if the distance is less than 6,000. Standard errors in parentheses, clustered at the country pair-year level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4. Regression Discontinuity: Effect of Nonstop Flights on Firm Innovation is Stronger for Firms that are Both Located Near Innovation Hubs

	Both Airports are Hubs		One or More Non-hub Airports	
	(1)	(2)	(3)	(4)
	Citations (asinh)	Collaborations (asinh)	Citations (asinh)	Collaborations (asinh)
Bandwidth:	Optimal	Optimal	Optimal	Optimal
Nonstop flights (asinh)	0.321***	0.246**	0.014	0.057
	(0.092)	(0.079)	(0.032)	(0.041)
(6,000 – Distance)	-0.001	-0.001	-0.000	-0.000
	(0.001)	(0.001)	(0.000)	(0.000)
(6,000 – Distance) x Under6,000	0.002	0.002	0.000	0.000
	(0.003)	(0.001)	(0.000)	(0.000)
Country pair-year FE	Y	Y	Y	Y
Observations	1,034	1,485	1,199	1,507

Note: This table presents the results from the regression discontinuity design, divided into subsamples that consist of: (1) hub-hub location pairs; and (2) location pairs with at least one non-hub location. Standard errors are in parentheses, clustered at the country-country level. Dependent variables are inverse hyperbolic sine transformed. Optimal bandwidth calculation follows the methodology described in Calonico et al. (2020). Airports are located near an innovation hub if they are within a 50-mile radius of innovation hubs, as defined in Bikard and Marx (2020). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5. Regression Discontinuity: Effect of Nonstop Flights on Firm Innovation is Stronger for Firms in Innovation-Leading Countries

	Citations (asinh)			Collaborations (asinh)		
	(1) Leader-Leader	(2) Follower-Follower	(3) Leader-Follower	(4) Leader-Leader	(5) Follower-Follower	(6) Leader-Follower
Nonstop flights (asinh)	1.795***	2.225	0.118	0.496***	0.397	0.095
	(0.423)	(5.553)	(0.121)	(0.154)	(1.146)	(0.079)
(6,000-Distance)	-0.006*	-0.019	-0.001***	0.000	-0.004	-0.001***
	(0.003)	(0.052)	(0.000)	(0.001)	(0.011)	(0.000)
(6,000-Distance) x Under6,000	0.014**	0.029	0.002***	0.001	0.006	0.001***
	(0.005)	(0.077)	(0.001)	(0.002)	(0.016)	(0.000)
Observations	583	748	1,562	583	748	1,562

Note: Observations at the airport pair-year level. Standard errors in parentheses, clustered at the country pair-year level. Variables inverse hyperbolic sine transformed are denoted by asinh. Bandwidth set at 550 miles. All specifications include country pair-year fixed effects. "Leader" and "Follower" denote firms in countries that are defined in Table 6 in Furman and Hayes (2004). Leader countries are historically high in innovation productivity and include Germany, Japan, Sweden, Switzerland, and the United States.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6. Regression Discontinuity: Effect of Nonstop Flights on Firm Innovation is Stronger for Firm Locations Separated by Temporal and Ethnic Distance

	Temporal Distance				Ethnic Distance			
	Citations (asinh)		Collaborations (asinh)		Citations (asinh)		Collaborations (asinh)	
Distance:	(1) High	(2) Low	(3) High	(4) Low	(5) High	(6) Low	(7) High	(8) Low
Nonstop flights (asinh)	0.420***	0.328*	0.215***	0.190*	0.354***	0.143***	0.130***	0.087***
	(0.127)	(0.187)	(0.064)	(0.107)	(0.101)	(0.050)	(0.046)	(0.033)
(6,000-Distance)	0.000	-0.002	-0.000	-0.001	-0.001**	-0.001*	-0.000	-0.000*
	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
(6,000-Distance) x Under6,000	-0.001	0.004**	-0.000	0.002**	0.001*	0.001**	0.000	0.001*
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Observations	1,398	2,365	1,398	2,365	3,795	3,795	3,795	3,795

Note: Dependent variables are counts of citations/collaborations for firms of different types, all aggregated to the airport pair-year level. We divide firms into High/Low Temporal Distance firms (above/below median temporal distance, Columns 1-4), and firms/subsidiaries at airport pairs with High/Low Ethnic Distance (above/below median ethnic distance, Columns 5-8). Thus, Column (1) uses the number of citations to patents by firms/subsidiaries at airport pairs with high temporal distance, while Column (2) uses the number of citations to patents by firms with low temporal distance, as dependent variables. Standard errors in parentheses, clustered at the country pair-year level. Variables inverse hyperbolic sine transformed are denoted by asinh. Bandwidth set at 550 miles. All specifications include country-pair-year fixed effects. For temporal distance, “High” indicates two locations that are greater than one hour apart in time zone difference; “Low” indicates one hour or less in time zone difference. For ethnic distance, “High” indicates multiethnic composition of inventors across two locations; “Low” indicates co-ethnic composition of inventors.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7. Effect of Nonstop Flights on Firm Innovation is Stronger for Firms in Countries that are Culturally Distant in Terms of People's Attitudes Toward Immigrants (Immigrant Friendliness)

	Citations (asinh)			Collaborations (asinh)		
	(1) Friendly-Friendly	(2) Unfriendly-Unfriendly	(3) Friendly-Unfriendly	(4) Friendly-Friendly	(5) Unfriendly-Unfriendly	(6) Friendly-Unfriendly
Nonstop flights (asinh)	-2.698	-0.265***	0.789***	-0.897	-0.022	0.473**
	(3.064)	(0.081)	(0.292)	(1.136)	(0.033)	(0.194)
(6,000-Distance)	0.028	-0.004**	-0.000	0.009	-0.002*	-0.000
	(0.033)	(0.001)	(0.001)	(0.012)	(0.001)	(0.001)
(6,000-Distance) x Under6,000	-0.050	0.003*	-0.000	-0.017	0.002*	-0.000
	(0.060)	(0.002)	(0.001)	(0.022)	(0.001)	(0.001)
Observations	616	748	1,760	616	748	1,760

Note: Observations at the airport pair-year level, excluding airport pairs with missing immigrant friendliness measures (25.33% of the sample). Standard errors in parentheses, clustered at the country pair-year level. Variables inverse hyperbolic sine transformed are denoted by asinh. Bandwidth set at 550 miles. All specifications include country pair-year fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8. Firms Near More Connected Airports Have More Citations and Scientific Publications

	Firms		Academic Institutions	
	(1) Citations (asinh)	(2) Publications (asinh)	(3) Citations (asinh)	(4) Publications (asinh)
# Connected Airports (2015)	0.124***	0.105***	0.085***	0.061***
	(0.046)	(0.036)	(0.027)	(0.019)
Airports Near 6k	0.163	0.115	0.047	0.043
	(0.109)	(0.084)	(0.065)	(0.048)
Distance to Equator	-0.000	-0.000	-0.000	-0.000
	(0.001)	(0.000)	(0.000)	(0.000)
Time Zone Difference from GMT	-0.287	-0.197	-0.121	-0.091
	(0.362)	(0.281)	(0.217)	(0.153)
Region FE	Y	Y	Y	Y
Observations	4,956	4,956	4,956	4,956

Note: Coefficient estimates from 2 stage least squares. Observations at the airport level. Standard errors in parentheses, clustered at the country level. All specifications include region fixed effects. All dependent variables are asinh transformed. # Connected Airports counts the number of connected airports between 5,500-6,500 miles. Airports Near 6k counts the number of airports (connected or not) in the same bandwidth. We instrument for the # Connected Airports using ShareBelow6k, which divides the number of airports (connected or not) 5,500-6,000 miles away by the number of airports (connected or not) 5,500-6,5000 miles away.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 9. Shorter Flight Duration is Associated with More Citations and Collaborations

	(1) Citations (asinh)	(2) Collaborations (asinh)	(3) Citations (asinh)	(4) Collaborations (asinh)
Duration (asinh)	-0.628*	-0.659***	-0.808***	-0.764***
	(0.331)	(0.248)	(0.246)	(0.230)
Price (asinh)	-0.337	-0.240	-0.381**	-0.266
	(0.216)	(0.208)	(0.176)	(0.185)
Distance (asinh)			0.886	0.518
			(1.213)	(0.745)
Constant	6.826***	5.382***	-0.110	1.328
	(2.461)	(2.026)	(11.733)	(7.537)
Observations	1,247	1,247	1,247	1,247
R^2	0.742	0.590	0.744	0.591

Note: This table tests the validity of two mechanisms that potentially drive the connectivity-innovation relationship: flight duration and flight price. Standard errors are in parentheses, clustered at the country-country level. Dependent variables are inverse hyperbolic sine transformed. Both dependent variables are asinh transformed. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Online Appendix
for
Innovation on Wings: Nonstop Flights and Global Innovation

Table of Contents

Appendix A. Counterfactual Patent Matching

A1. Research questions and identification challenges

To test whether distance affects knowledge flows and whether knowledge is geographically localized, the typical way would be to carry out the following specification:

$$Citations_{ij} = \beta_0 + \beta_1 Distance_{ij} + \epsilon_{ij},$$

where i, j refer to two locations, $Citations_{ij}$ counts the number of citations between locations *i* and *j*, and $Distance_{ij}$ is the distance between those locations. If we see $\beta_1 < 0$, this would be consistent with localized knowledge spillovers.

A key omitted variable is technological similarity. For example, a relevant question to ask is are there more citations between inventors located in the Bay Area because of their geographic proximity to one another? Or is it because the Bay Area has a specific set of industries, and citations are more likely to occur within industries? Simply controlling for industry classification and comparing citation patterns within industries is also inadequate because technological similarity causes not only knowledge spillovers but agglomeration: inventors move to locations where they will be most productive. We cannot distinguish between the effects of technological similarity and proximity.

JTH 1993 suggest comparing patent-citation pairs that actually happen versus potential patent-citation pairs where a citation does not happen. Given a "focal" patent and a real "citing" patent that cites the focal patent, a "counterfactual" patent can be chosen such that it is similar to the real citing patent but does not actually cite the focal patent. There exists debate regarding how to accurately choose potential patent-citation pairs, but generally researchers control for counterfactual patents that have similar 1) application year and 2) patent classification. Along these lines, we aim to control for the underlying technological similarity by looking at how similar the keywords in the patent titles are. In section C.2, we describe the methodology for measuring technological similarity using text analysis.

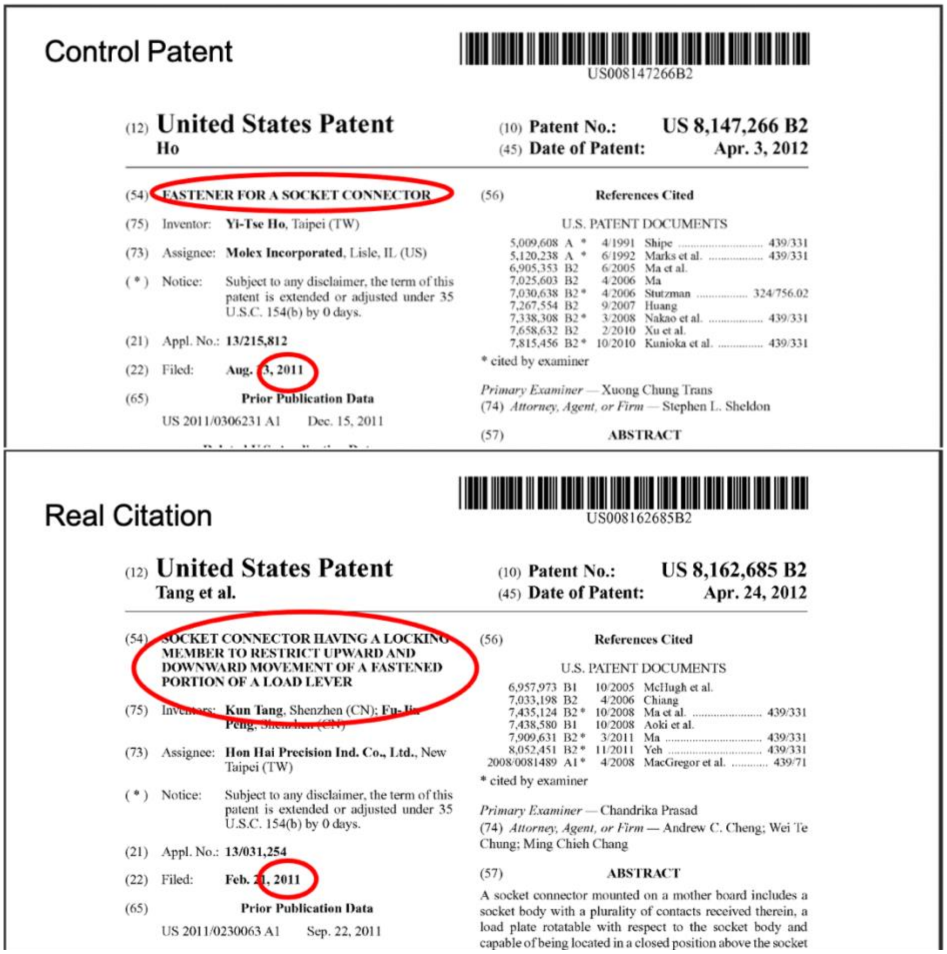
A2. Text similarity matching method

For each real citing patent, we collect a counterfactual patent that is similar to the real citing patent but does NOT cite the focal patent. We measure similarity based on the patents' titles. Intuitively, if patents contain similar words, they will have similar underlying technologies. Furthermore, if patents contain similar words that do not appear elsewhere (e.g., "socket connector"), they should be counted as more similar than two patents that share words that frequently occur elsewhere. Thus, we use weighted Jaccard similarity measure[1] with TF-IDF weights.

Because the text matching process is computationally intensive, we narrow down our sample of all patents between 2005 and 2015 to those that were published by the top 50 assignees with the most patents and their citing patents. For each citing patent in this subsample, we take three steps: 1) collect all patents that have the same application year but do not cite our focal patent (these are potential candidates for the counterfactual patent), 2) calculate similarity scores between the real citing patent and all the candidates, and 3) choose the patent that has the highest similarity score to be the counterfactual patent.

For example, Figure A1 shows a pair of patents, one of which actually cites a patent in our dataset (the focal patent), and another that is similar to that citing patent but does not cite any patents in our dataset. The two patent titles have similar keywords, such as "socket connector" or "fasten" which gave the pair high similarity scores. Also, the two patents have the same application year of 2011. However, the two patents have different locations: While the real citing patent's location is Shenzhen, mainland China, the counterfactual patent is in Taipei, Taiwan. The fact that one patent cites the focal patent but the other does not may be driven by the availability of nonstop flights. Specifically, the citation pattern may be driven by the possibility that there are nonstop flights between the focal patent's inventor location and the citing patent's inventor location, but not between the focal patent's inventor location and the counterfactual patent's inventor location. This is the relationship we aim to statistically test.

[1] We calculate $s(\boldsymbol{X},\boldsymbol{Y}) = \frac{\sum_{i=1}^{M}\min(X_i,Y_i)}{\sum_{i=1}^{M}\max(X_i,Y_i)}$ for vectors $\boldsymbol{X},\boldsymbol{Y}$. The vectors represent the TF-IDF scores of the words contained in the documents. TF-IDF refers to Term Frequency-Inverse Document Frequency and is calculated by counting the number of times a word occurs in a document (term frequency), and dividing that number by the log of the fraction of documents that contain the word (inverse document frequency).


Control Patent

US008147266B2

(12) United States Patent
Ho

(10) Patent No.: US 8,147,266 B2
(45) Date of Patent: Apr. 3, 2012

(54) FASTENER FOR A SOCKET CONNECTOR

(75) Inventor: Yi-Tse Ho, Taipei (TW)

(73) Assignee: Molex Incorporated, Lisle, IL (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: 13/215,812

(22) Filed: Aug. 23, 2011

(65) Prior Publication Data

US 2011/0306231 A1 Dec. 15, 2011

(56) References Cited

U.S. PATENT DOCUMENTS

5,009,608	A *	4/1991	Shipe	439/331
5,120,238	A *	6/1992	Marks et al.	439/331
6,905,353	B2	6/2005	Ma et al.	
7,025,603	B2	4/2006	Ma	
7,030,638	B2 *	4/2006	Stutzman	324/756.02
7,267,554	B2	9/2007	Huang	
7,338,308	B2 *	3/2008	Nakao et al.	439/331
7,658,632	B2	2/2010	Xu et al.	
7,815,456	B2 *	10/2010	Kunioka et al.	439/331

* cited by examiner

Primary Examiner — Xuong Chung Trans
(74) *Attorney, Agent, or Firm* — Stephen L. Sheldon

(57) ABSTRACT

Real Citation

US008162685B2

(12) United States Patent
Tang et al.

(10) Patent No.: US 8,162,685 B2
(45) Date of Patent: Apr. 24, 2012

(54) SOCKET CONNECTOR HAVING A LOCKING MEMBER TO RESTRICT UPWARD AND DOWNWARD MOVEMENT OF A FASTENED PORTION OF A LOAD LEVER

(75) Inventors: Kun Tang, Shenzhen (CN); Fu-Jin Peng, Shenzhen (CN)

(73) Assignee: Hon Hai Precision Ind. Co., Ltd., New Taipei (TW)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: 13/031,254

(22) Filed: Feb. 21, 2011

(65) Prior Publication Data

US 2011/0230063 A1 Sep. 22, 2011

(56) References Cited

U.S. PATENT DOCUMENTS

6,957,973	B1	10/2005	McHugh et al.	
7,033,198	B2	4/2006	Chiang	
7,435,124	B2 *	10/2008	Ma et al.	439/331
7,438,580	B1	10/2008	Aoki et al.	
7,909,631	B2 *	3/2011	Ma	439/331
8,052,451	B2 *	11/2011	Yeh	439/331
2008/0081489	A1 *	4/2008	MacGregor et al.	439/71

* cited by examiner

Primary Examiner — Chandrika Prasad
(74) *Attorney, Agent, or Firm* — Andrew C. Cheng; Wei Te Chung; Ming Chieh Chang

(57) ABSTRACT

A socket connector mounted on a mother board includes a socket body with a plurality of contacts received therein, a load plate rotatable with respect to the socket body and capable of being located in a closed position above the socket

Figure A1. Illustration of the matching method.
This figure juxtaposes two patents: one real patent that cites the focal patent (not shown here) and a control/counterfactual patent that does not cite the focal patent. This pair of patents help illustrate how text similarity is calculated in our matching method.

When testing for inventors and collaborations between inventors, we utilize a similar approach. The assumption is that all inventors on the counterfactual patent and on the real citing patent are working on similar technologies and are thus likely to collaborate with one another (relative to inventors who work on dissimilar technologies). However, collaborations occur between certain inventors but not between others. We estimate the extent to which this pattern is driven by the availability of nonstop flights between these inventors.

A3. Summary Statistics

In Table A1, we present summary statistics for the counterfactual patent matching data. Exactly half of the citations are real citations by construction. On average, 44.5% of patent-citation pairs are connected by a nonstop flight. Patent-citation pairs are separated by 3,635 miles, and on average have 2,453 of flights between them. Around 60% of the patent-citation pairs are located across country borders.

Similarly, for the counterfactual collaboration dataset, around 34.8% of inventor pairs are real collaborations. 53.2% of the inventors are connected by a nonstop flight, and are separated by 2,978 miles. On average, there are 3,550 flights between any two inventors, and 42.8% are located across country borders (i.e., are (potential) international collaborations). A more detailed breakdown is presented in Table A2.

Table A1. Summary statistics for the counterfactual patent matching data.

Part A. Counterfactual Citations					
	count	mean	s.d.	min	max
Real Citations	554884	0.50	0.50	0	1
Nonstop	554884	0.45	0.50	0	1
Distance (asinh)	554884	8.48	1.08	5.30	10.10
Nonstop Flights (asinh)	554884	3.99	4.61	0	12.60
Intl	554884	0.59	0.49	0	1
Total Flights	554884	2452.99	6099.44	0	74128
Application Year	554884	2011.75	2.27	2005	2016

Part B. Counterfactual Collaborations					
	count	mean	s.d.	min	max
Real Collaborations	3350653	0.35	0.48	0	1
Nonstop	3350653	0.53	0.50	0	1
Distance (asinh)	3350653	8.17	1.17	5.30	10.10
Nonstop Flights (asinh)	3350653	4.88	4.77	0	12.60
Intl	3350653	0.43	0.49	0	1
Total Flights	3350653	3550.44	7748.24	0	74355
Application Year	3350653	2011.82	2.29	2005	2016

Note: Observations are at the patent-citation pair level for Part A and at the inventor pair level for Part B. Exactly half of the patent-citation pairs are real citations in Part A. Nonstop is an indicator for whether there exists a nonstop flight between a patent's location and the citation's location (Part A) or between the inventors' locations (Part B). Intl is an indicator for whether the patent-citation pairs are located in different countries (Part A) or whether the inventors are located in different countries (Part B). Distance is in miles. Application year is the application year for the citation.

Table A2. Summary statistics for the counterfactual patent matching data, by whether a nonstop flight exists between two locations.

	count	mean	s.d.	min	max
Part A-1: Counterfactual Citations, RealCitation=1					
Has Nonstop	277442	0.47	0.50	0.00	1.00
Distance (asinh)	277442	8.42	1.09	5.30	10.09
Nonstop Flights (asinh)	277442	3.95	4.35	0.00	11.91
Intl	277442	0.55	0.50	0.00	1.00
Nonstop Flights	277442	5540.27	12916.75	0.00	148483.00
Year	277442	2011.75	2.27	2005.00	2016.00
Part A-2: Counterfactual Citations, RealCitation =0					
Has Nonstop	277442	0.42	0.49	0.00	1.00
Distance (asinh)	277442	8.53	1.07	5.30	10.10
Nonstop Flights (asinh)	277442	3.41	4.19	0.00	11.87
Intl	277442	0.64	0.48	0.00	1.00
Nonstop Flights	277442	4270.96	11400.96	0.00	142194.00
Year	277442	2011.75	2.27	2005.00	2016.00
Part B-1: Counterfactual Collaborations, RealCollaboration=1					
Has Nonstop	1166671	0.61	0.49	0.00	1.00
Distance (asinh)	1166671	7.79	1.23	5.30	10.07
Nonstop Flights (asinh)	1166671	5.30	4.44	0.00	11.91
Intl	1166671	0.29	0.45	0.00	1.00
Nonstop Flights	1166671	9128.62	17298.94	0.00	148483.00
Year	1166671	2011.76	2.33	2005.00	2016.00
Part B-2: Counterfactual Collaborations, RealCollaboration=0					
Has Nonstop	2183982	0.49	0.50	0.00	1.00
Distance (asinh)	2183982	8.38	1.08	5.30	10.10
Nonstop Flights (asinh)	2183982	4.09	4.37	0.00	11.91
Intl	2183982	0.50	0.50	0.00	1.00
Nonstop Flights	2183982	6016.53	14320.50	0.00	148483.00
Year	2183982	2011.86	2.27	2005.00	2016.00

A4. Counterfactual Citations

To construct the citation dataset, for each focal patent, we identify a citing patent that cites the focal patent and a counterfactual non-citing patent. That is, for each citing patent, we have identified a counterfactual patent from the same application year that does not cite the focal patent but has a title with high textual similarity to the title of the citing patent. We determine the degree of textual similarity using an algorithm based on the Jaccard index of similarity, which counts the number of common words ("tokens") in the two titles as a share of the total number of words (Niwattanakul et al. 2013).[2] More specialized words are weighted heavier than those that are common (for example, two patent titles that both contain the word "microprocessor" would be considered more similar than two titles that both contain the word "new"). We then study the effect of connectedness on patent citations using this set of matched counterfactual patents by creating a dataset in which each observation is a pair of patents. A patent pair contains either a focal patent and a real citing patent or a focal patent and a counterfactual patent matched to the citing patent, marked accordingly.

In the table below, we present summary statistics for the counterfactual patent-matching data. Exactly half of the patent-citation pairs are real citations by construction, with the other half consisting of counterfactual, non-citing patents. The dataset maps each patent pair (i.e., both patent-real citation pairs and patent-counterfactual citation pairs) to airport pairs to obtain information about the existence and the number of nonstop flights. On average, 45% of patent-citation pairs are connected by a nonstop flight (47% for real patent-citation pairs, 42% for counterfactual patent-citation pairs). The average patent-citation pair is separated by 3,635 miles in distance and has 2,453 flights per year (6.7 flights or around 3 round trips per day) between them.[3] Around 60% of the patent-citation pairs are located across national borders.

[2] Specifically, we calculate $s(\boldsymbol{X},\boldsymbol{Y}) = \frac{\sum_{i=1}^{M}\min(X_i,Y_i)}{\sum_{i=1}^{M}\max(X_i,Y_i)}$ for vectors $\boldsymbol{X},\boldsymbol{Y}$. The vectors represent the TF-IDF scores of the words contained in the titles. TF-IDF refers to Term Frequency-Inverse Document Frequency and is calculated by counting the number of times a word occurs in a title (term frequency) and dividing that number by the log of the fraction of titles that contain the word (inverse document frequency).

[3] The average number of nonstop flights between patent-citation pairs is larger than the average number of flights between airport pairs (which is 631.44 nonstop flights per year) because patent-citation pairs are likely to be located at busier airport pairs that have many flights.

Table A3. Summary statistics for the control patent matching data.

Part A. Counterfactual Citations

	count	mean	sd	min	max
realCitation	554884	.5	.5000005	0	1
Nonstop	554884	.4452696	.496996	0	1
Distance (asinh)	554884	8.475401	1.081952	5.298575	10.10236
Total Flights (asinh)	554884	3.99273	4.609691	0	12.60137
Intl	554884	.5942503	.491037	0	1
Total Flights	554884	2452.989	6099.436	0	74128
Application Year	554884	2011.751	2.270869	2005	2016

Part B. Counterfactual Collaborations

	count	mean	sd	min	max
realCollaboration	3350653	.3481921	.4763974	0	1
Nonstop	3350653	.5315071	.4990064	0	1
Distance (asinh)	3350653	8.17056	1.168847	5.298498	10.10236
Total Flights (asinh)	3350653	4.882422	4.766773	0	12.60137
Intl	3350653	.427522	.4947191	0	1
Total Flights	3350653	3550.44	7748.241	0	74355
Application Year	3350653	2011.821	2.294037	2005	2016

Note: Observations are at the patent-citation pair level for Part A, and at the inventor-pair level for Part B. Exactly half of patent-citation pairs are real citations in Part A. Nonstop is an indicator for whether there exists a nonstop flight between a patent's location and the citation's location (part A), or between the inventors' locations (part B). Intl is an indicator for whether the patent-citation are located in different countries (part A) or whether inventors are located in different countries (part B). Distance is in miles. Application Year is the application year for the citation.

With this dataset, we explore the question of whether flight connectivity between two locations (as measured by the availability of nonstop flights) can explain patent citations between those locations. We estimate the following specification using ordinary least squares:[4]

$$Real\ Citation_{p,c,t} = \delta_1 Has\ Nonstop_{p,c,t} + \delta_2 Distance_{p,c} + \omega_p + \gamma_c + \eta_t + \varepsilon_{p,c,t} \quad (1)$$

for focal patent p, citation c, and application year t, where c can be a real citation or counterfactual citation. $Real\ Citation_{p,c,t} = 1$ if c actually cites patent p and = 0 otherwise. $Has\ Nonstop_{p,c,t}$ is a binary variable for whether there exists a nonstop flight between the airports nearest to the locations of p, c in year t. Besides the binary variable for the existence of nonstop flights between the two locations underlying a patent-citation pair, we also use the count of nonstop flights between p and c as an alternative independent variable. We also control for $Distance_{p,c}$, which is the geographic distance between the primary contributors (inventors whose names are listed first) who filed the citing patent p and the counterfactual, non-citing patent.[5] We include focal patent fixed

[4] We are estimating a linear probability model, which approximates the marginal effects without assuming an arbitrary nonlinear relationship (Angrist and Pischke 2008). Furthermore, OLS allows us to calculate the correct cluster robust standard errors.

[5] Distance, similarly to other variables with long-tailed distributions, is rescaled using the inverse hyperbolic sine (asinh). The inverse hyperbolic sine approximates the natural logarithm while retaining zero-valued observations (MacKinnon and Magee 1990).

effects (ω_p) to control for unobserved characteristics associated with the focal patent. To control for geopolitical unobservables, we include country fixed effects for the countries in which the citing and counterfactual patents are located (γ_c), omitting focal patent country fixed effects since we already include focal patent fixed effects. Finally, we control for any year-specific shocks that may affect the relationship between nonstop flight existence and citations by including year fixed effects (η_t). Table A4 presents the estimates for this specification.

Table A4. Effect of the existence of a nonstop flight on the likelihood of patent citation.

Dep. Var.: Real citation (binary)				
	(1)	(2)	(3)	(4)
Has nonstop	0.0564		0.0395	
	(0.004) ***		(0.006) ***	
Nonstop flights (asinh)		0.0080		0.0073
		(0.000) ***		(0.001) ***
Has nonstop × Intl			0.0258	
			(0.008) ***	
Nonstop flights (asinh) × Intl				0.0010
				(0.001)
Distance (asinh)	-0.0251	-0.0196	-0.0139	-0.0091
	(0.002) ***	(0.002) ***	(0.003) ***	(0.003) ***
Intl			-0.0636	-0.0521
			(0.010) ***	(0.010) ***
N	554,864	554,864	554,864	554,864
R^2	0.064	0.065	0.065	0.065

Note: This table estimates the change in the likelihood of a patent citation given the existence of a nonstop flight between airports near the inventors. It also estimates the effect of an alternative independent variable (number of nonstop flights) on citations. All specifications include a cited patent fixed effect, a citing/counterfactual patent's country fixed effect, and a year fixed effect. Standard errors are clustered at the cited patent level.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

The results in Column 1 suggest that the existence of nonstop flights between two inventors who have patented similar technologies is associated with a 5.6 percentage point increase in the likelihood of a real citation, which represents an 11% increase in the likelihood of a real citation compared to the baseline citation likelihood of 50%. This estimate is obtained after controlling for the distance between the inventors and the characteristics of both countries of the inventors (in case they are different), as well as year fixed effects. To delve beyond the existence of nonstop flights between inventors, Column 2 of the same table uses an alternative independent variable—the number of flights between inventors. The result suggests that a 10% increase in the number of flights between two locations is associated with an increase of 0.08 percentage points in the likelihood of a patent citation. In terms of economic significance, given it is a continuous variable, we look in terms of standard deviations. In this case, a 1 standard deviation increase in flights corresponds to a 2 percentage point increase in the likelihood of a patent citation (or a 4% increase compared to the baseline).

Columns 3 and 4 allow for differential effects on whether a citation occurs across national borders by interacting the binary nonstop flight existence variable with a binary variable that equals one if p and c belong to inventors in different countries and equals zero otherwise. Our results suggest that the existence of nonstop flights is associated with an increase in the likelihood of citation, and this relation is, in part, driven when patents that are in different countries. When a nonstop flight exists across national borders relative to when there are no nonstop flights, the interaction term in Column 3 shows the likelihood of a citation increases by an additional 2.6 percentage points (a 5.2% increase based on the baseline citation likelihood). However, Column 4 shows that a 10% increase in the number of flights across international borders increases the likelihood of a citation by 0.01 percentage points, and the effect is not statistically significant. This result suggests that the existence of nonstop flights between two international locations is important for cross-country patent citations, but we are unable to find a distinguishable result for this continuous variable. In other words, the result suggests that the extensive margin is perhaps more important than the intensive margin in terms of flight connectivity. Thus far, our results suggest that connectedness between locations explains citations between patents, and this relationship holds when crossing national borders.

A5. Counterfactual Collaborations

We also employed a textual similarity-based counterfactual patent method to study how nonstop flights affect patent collaborations. To measure inventor collaborations, we replicate the idea above and create a dataset in which each observation is a pair of inventors. Given a focal patent, we identify all inventors who worked on both a real citing patent and its matched counterfactual, non-citing patent. For this exercise, we use two sets of patents: all citing patents used for analysis in Section 3.1 and all counterfactual citations used for analysis in Section 3.1. Since these two sets of matched patents are highly similar (as measured by textual similarity of patent abstracts) in content, the inventors on these patents are arguably potential collaborators whose work centers around similar technologies. We then create all pairwise combinations of inventors from those two sets of patents. We mark a pair of inventors as a real collaboration if the two inventors actually worked together and otherwise if it is a counterfactual pair. For both pairs of inventors (i.e., real collaborations and counterfactual collaborations), we impute the relevant airport pairs. Then, we compare the existence of nonstop flights between real collaborators against that between potential collaborators who were not collaborating. For this counterfactual collaboration dataset, 34% of inventor pairs are real collaborations.[6] Fifty-three percent of the inventor pairs are connected by a nonstop flight (61% for real inventor pairs, 49% for counterfactual inventor pairs) and are separated by 2,978 miles on average. Also, on average, there are 3,550 flights between an inventor pair, and 43% of the inventor pairs are located across national borders (i.e., are potential global collaborations).

Specifically, to study global collaborations, we estimate a slightly different specification from the citations regression:

$$\begin{aligned} &Real\ Collaboration_{p,i,j,t} \\ &\qquad = \beta_1 Has\ Nonstop_{i,j,t} + \beta_2 Distance_{i,j} + \omega_p + \eta_t + \varepsilon_{p,c,t} \end{aligned} \tag{2}$$

Again, p stands for the focal patent, i, j stand for inventors, and t stands for the year of collaboration. The value of $Real\ Collaboration_{p,i,j,t} = 1$ if inventors i, j are actual collaborators

[6] In contrast to citations, the collaborations dataset is not split evenly because real and counterfactual collaborations do not contain even numbers of inventors.

on a citing or counterfactual citing patent of focal patent p issued in year t and = 0 otherwise. $Has\ Nonstop_{i,j,t}$ is an indicator for whether a nonstop flight connects the two inventors in year t, and $Distance_{i,j}$ is the inverse hyperbolic sine of the distance between the inventors. We include focal patent fixed effects ω_p to control for the underlying technology and year fixed effects η_t to capture any time-specific effects.

Table A5. Effect of the existence of a nonstop flight on the likelihood of patent collaboration.

Dep. Var.: Real collaboration (binary)				
	(1)	(2)	(3)	(4)
Has nonstop	0.0546		0.0187	
	(0.002) ***		(0.003) ***	
Nonstop flights (asinh)		0.0069		0.0033
		(0.000) ***		(0.000) ***
Has nonstop × Intl			0.0290	
			(0.004) ***	
Nonstop flights (asinh)× Intl				0.0033
				(0.000) ***
Distance (asinh)	-0.1580	-0.1549	-0.0761	-0.0740
	(0.001) ***	(0.001) ***	(0.002) ***	(0.002) ***
Intl			-0.3988	-0.3960
			(0.005) ***	(0.005) ***
N	3,350,645	3,350,645	3,350,645	3,350,645
R^2	0.255	0.255	0.273	0.273

Note: This table estimates the change in the likelihood of a collaboration given the existence of a nonstop flight between airports near the inventors. It also estimates the effect of an alternative independent variable (number of nonstop flights) on citations. All specifications include a cited patent fixed effect, a citing/counterfactual patent's country fixed effect, and a year fixed effect. Standard errors clustered at the cited patent level.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A5 shows the results of estimating Appendix Equation (2). Columns 1 and 2 show that inventors connected by at least one nonstop flight between their locations are 5.5 percentage points more likely to be collaborators, which is a 16% increase based on the baseline collaboration likelihood of 34%, as compared to inventors without nonstop connections between their locations. Column 2 of the same table again uses the number of flights between inventors. The result suggests that a 10% increase in the number of flights between two locations is associated with an increase of 0.07 percentage points in the likelihood of a collaboration. In terms of economic significance, we find that a 1 standard deviation increase in flights corresponds to a 1.7 percentage point increase in the likelihood of a collaboration (or a 5.1% increase compared to the baseline collaboration likelihood of 34%).

Columns 3 and 4 present the results from interacting the flight variables in Equation (2) with a binary variable for whether the inventors are located in different countries. The positive and significant coefficient shows that for global inventor teams, there is an additional 2.9 percentage point increase (an additional 8.3% increase from the baseline collaboration probability of 34%) in the positive effect of the existence of nonstop flights on collaborations. In contrast to citations, the effect for international collaborations is significant for both independent variables—that is, whether or not two locations are connected by a nonstop flight and number of nonstop flights between the two locations. This result suggests that both the existence and the quantity/frequency of nonstop flights are important for driving collaborations between two locations.

A6. Alternate Empirical Specifications

Using the counterfactual patent information collected above, we test how nonstop flights correlate with the probability of a patent pair having a nonstop flight. If nonstop flights increase the likelihood of citations, we should see a positive relationship between the given patent pair is a real citing patent pair, and whether the two locations have a nonstop flight.

Our regression specification is as follows:

$$\begin{aligned} Real\ Citation_{p,c,t} \\ &= \delta_1 Nonstop\ Flights_{p,c,t} + \delta_2 Distance_{p,c} + \delta_3 Intl_{p,c} \\ &+ \delta_4 Intl_{p,c} \times Nonstop\ Flights_{p,c,t} + \omega_p + \gamma_c + \eta_t + \varepsilon_{p,c,t} \end{aligned} \quad (3)$$

for focal patent p and citation c, and application year t where c can correspond to a real citation or counterfactual citation. $Real\ Citation_{p,c,t} = 1$ if c actually cites patent p, and 0 otherwise. $Nonstop\ Flights_{p,c,t}$ counts the number of nonstop flights between the airports nearest to the locations of p, c in year t. We include focal patent fixed effects (ω_p) so that δ_1 measures the relationship between nonstop flights and real citations, keeping all the focal patent characteristics fixed. The variation in the number of nonstop flights comes from comparing the real citations and counterfactual citations. In some specifications, we include dummy variables for whether the flight is an international flight $Intl_{p,c}$, and an interaction term between international flights and the number of nonstop flights. To the extent that nonstop flights affect international knowledge flows, we should see $\delta_4 \neq 0$. To control for geopolitical unobservables, we include country fixed effects for the counterfactual patent (γ_c). Finally, we control for any year specific shocks that may affect the relationship between nonstop flights and citations (η_t).

While in our preferred specification, we include distance between airports as a control, we address concerns of potential multicollinearity between the number of nonstop flights and distance. We first show that while there is a negative correlation between distance and the number of flights $(\rho = -0.4481)$, nonparametric kernel estimates in Figure A2 suggest the relationship is nonlinear. For very short distances between two locations, the likelihood of having a nonstop flight between those locations is around 0.6. The likelihood of having a nonstop flight increases to 0.8 as airports become more distant. However, for longer flights, the probability of nonstop flights decreases.

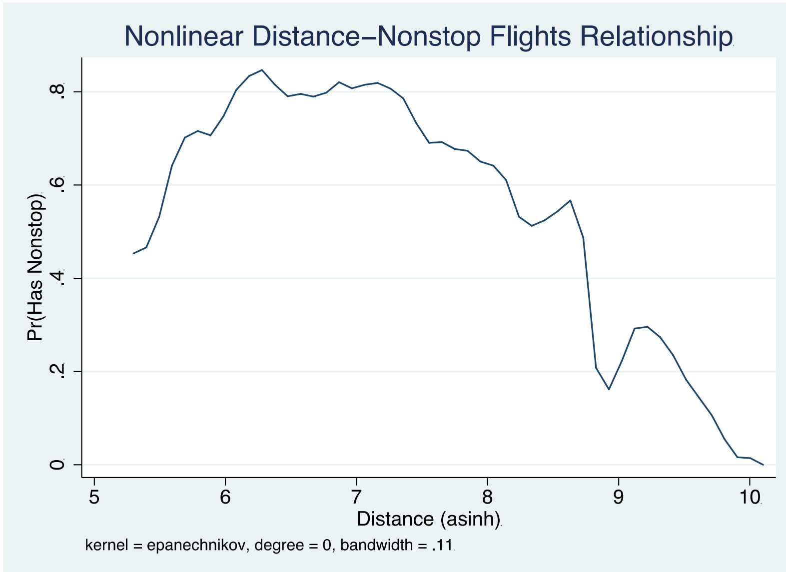


Figure A2. Local polynomial regression of nonstop flights on distance
Kernel-weighted local polynomial regression of nonstop flights on distance. Dataset uses all patent-citation pairs. Nonstop flights is a binary variable, denoting probability of any nonstop flights between a patent-citation pair. Distance between two airports is inverse hyperbolic sine transformed. Graph was created using the lpoly command in Stata.

We also estimate the specification above with and without distance controls, and compare the coefficients. Including controls for distance has little impact on our coefficient of interest δ_4 (the effect of nonstop flights on overcoming political borders) for citations, and a slight negative effect on collaborations. Overall, the results on reducing frictions from national borders hold even if we omit distance controls.

Table A6. Coefficients from the main specification results, with and without distance controls.

	Citations		Collaborations	
Coefficient	With Distance	Omit Distance	With Distance	Omit Distance
δ_1	0.0395*** (0.006)	0.0452*** (0.006)	0.0187*** (0.003)	0.0399*** (0.003)
δ_3	-0.0636*** (0.010)	-0.0926*** (0.007)	-0.3988*** (0.005)	-0.5692*** (0.004)
δ_4	0.0258*** (0.008)	0.0257*** (0.008)	0.0290*** (0.004)	0.0358*** (0.004)

Note: Standard errors in parentheses are clustered at the focal patent level. The coefficients in this table correspond to the same coefficients in Appendix Equation (3), which is the main specification in the paper.

Our standard errors are clustered at the focal patent level. Recent work shows that clustering is necessary when the sampling process and the assignment mechanism are clustered. In our context, this would correspond to how citations occur, and how counterfactual patents are assigned. Abadie et al. (2017)suggest researchers should cluster at the level of "treatment" assignment, hence our focal patent level clustered standard errors. If we believe instead that the assignment of counterfactual patents is completely random, we would not have a need for clustering; however, treatment is indeed not random, and assigned at the focal patent level. Alternatively, coarser levels of clustering (i.e., country-country level) would be appropriate if we have reason to believe non-random sampling of some countries. Again, however, our dataset comprises of all nonstop flights across the globe, diminishing concerns of such non-random sampling.

A7. Assumptions required for causal interpretation

Note that to interpret β as a causal effect, we would need to assume that $cov(Has\ Nonstop_{p,c,t}, \varepsilon_{p,c,t}) = 0$. This is a plausible assumption when thinking of the individual inventors: the existence of a nonstop flight between locations could be considered exogenous to patenting activities and citations. Of course, an important threat to our identification is whether the inventors relocate to places that have nonstop flights to locations with other inventors who they could collaborate with or cite. While this approach does not address some sources of endogeneity (e.g., we cannot distinguish between cases where the inventors involved in a particular citation or collaboration were in those locations before the existence of a flight), our regression discontinuity results are consistent with a causal interpretation.

A8. Inferring economic significance of the counterfactual results

The counterfactual analysis suggests adding 7,352 nonstop routes per year will increase citations by 6.26% and adding 6,670 routes will increase collaborations by 7.49%. Roughly, increasing the number of nonstop routes by 30.59% will increase citations by 0.24 per patent – also adding a nonstop route leads to 2.36 more citations between those airports. Similarly, increasing the number of nonstop routes by 43.14% will add 0.29 more inventors per patent.

From Tables A4 & A5, we see that the existence of nonstop flights increases the likelihood of a citation increases by 5.64 percentage points, and collaborations by 5.46 percentage points. The baseline likelihood of a citation is 50%, and thus a 5.64 percentage point increase would be a 11.2% increase. Similarly, the baseline likelihood of a collaboration is 34.82%, and thus a 5.46 percentage point increase would be a 15.68% increase in the likelihood of a collaboration. Taken together with the interaction terms for international pairs, this would be 8.22 percentage points for citations and 8.36 percentage points for collaborations. Thus, in percentage increases, this would be a 16.44% increase in citations, and 24.01% increase in collaborations.

To calculate standard deviation changes, a one standard deviation increase in nonstop flights is 6099.44, and compared to the mean 2452.99, this is a 248.6532% increase. Thus, that would be a 1.98923 percentage point increase in citations,

Interpreting the economic magnitude of these changes requires us to make assumptions about how many nonstop flights are added, which we detail below.

Table A4 shows that when estimating $Real\ Citation_{ij} = \beta_1 Has\ Nonstop_{ij}$ the coefficient is 0.0564. In conditional expectation form, we see $E[Real\ Citation_{ij} | Has\ Nonstop_{ij} = 1] - E[Real\ Citation_{ij} | Has\ Nonstop_{ij} = 0] = \mathbf{0.0564}$. This suggests the following interpretation: patent pairs ij with a nonstop flight are 5.64 percentage points more likely to be a real citation than patent pairs $i'j'$ without nonstop flights. Assuming this is causal, the implication is that for any patent pair ij that does *not* have a nonstop flight, adding a nonstop flight will make it 5.64 percentage points more likely to be a real citation. We will maintain this causal interpretation and calculate the increase in patenting when "adding" a nonstop flight between patent pairs.

Here, we calculate the magnitude of a 5.64 percentage point increase in patent citations. There are 554,884 patent pairs in our data, half of which are real. Alternatively, 247,073 patent pairs have nonstop flights, while 307,811 do not (See Table A3 below for a cross tabulation). The baseline probability of a real citation, given no nonstop flight, is 146,122 / 307,811 = 0.48. If nonstop flights were added to all patent-citation pairs, this baseline probability would increase by 5.64 percentage points to 0.53. Then, instead of 146,122 real citations, we would see 163,482.54 citations, an increase of 17,360.54 patents.[7] There are 72,509 unique publications in our counterfactual patent dataset, which is 3.83 citations per patent.[8] An increase of 17,360.54 patents is an increase of 0.24 citations per patent, which is a 6.26% increase in citations per patent.

Table A7. Cross tabulation of citations for patent pairs with and without nonstop flights.

	Counterfactual Citations	Real Citations	Total	Fraction Real Citations	Real citations if all routes have nonstop
No Nonstop	161,689	146,122	307,811	0.4747	163,482
Has Nonstop	115,753	131,320	247,073	0.5315	131,320
Total	277,442	277,442	554,884	0.5	294,802

We next calculate the magnitude of this increase in light of the number of nonstop flights. Adding a nonstop flight for all patent-citation pairs without nonstop flights (i.e., for 307,811 patent-citation pairs) is not equivalent to adding 307,811 routes because multiple patent-citation pairs exist for each route. A rough estimate would be to go to the route-level citations data and add nonstop flights to all routes with any number of citations between them.[9] There are 48,898 routes across 11 years for a total of 537,878 route-year observations. Of these, 176,704 route-years have citations, and of those, 80,871 route years have citations but no nonstop flights. To ensure all routes with citations have nonstop flights, we would need to add nonstop flights to 80,871 route-years, for about 7,351.90 routes per year. On average, 24,035.45 routes per year have nonstop flights, and thus 7,351.90 is a 30.59% increase in routes with nonstop flights.

Table A8. Tabulation of the number of route-years with and without citations and collaborations, divided by whether a route-year has nonstop flights.

	No citations	Has citations	No collaborations	Yes collaborations
No Nonstop	192,617	80,871	200,114	73,374
Has Nonstop	168,557	95,833	167,682	96,708
	361,174	176,704	376,796	170,082

The two tables above show that a 30.59% increase in nonstop flight routes will lead to a 6.26 percent increase in citations, or 0.24 citations per patent, per year. Alternatively, since adding 7,352 nonstop routes leads to an increase of 17,360 patents, we expect each new nonstop route to add 2.36 citations between those locations in a given year.

Finally, we repeat the exercise for collaborations. Again, Table A5 shows collaborations are 5.46 percentage points more likely if there exist nonstop flights. There are 3,350,653 potential inventor-pairs in our dataset, 1,166,671 of which are realized collaborations. If we add nonstop flights for all inventor pairs who don't have collaborations, we would see real collaborations increase

7 ((146,122/307,811)+0.0564)*307,811 = 163,482

8 277,447/72,509 = 3.8263

9 Since the results from the counterfactual analysis are derived from routes with patent-citation pairs, we take a conservative estimate and assume the impact may not carry over to routes without any patent-citation pairs.

from 454,515 to 540,223.73,[10] or by 85,708.73. This is 85,708.73 more collaborations, and an increase from 3.92 to 4.30 inventors on a patent, or about 0.29 inventors per patent.

Table A9. Cross tabulation of counterfactual versus real collaborations.

	Counterfactual collaboration	Real Collaboration	Total	Fraction Real	# Real if nonstop
No Nonstop	1,115,242	454,515	1,569,757	0.2895	540,224
Has Nonstop	1,068,740	712,156	1,780,896	0.3999	712,156
Total	2,183,982	1,166,671	3,350,653	0.3482	1,252,380

Calculating the number of flights we would need to add, we see that there are 73,374 route-years (6,670.36 routes per year) with no nonstop flights but some level of collaborations. Thus, adding 6,670.36 routes per year, which is a 43.14% increase in the number of routes will increase the average number of collaborators per patent by 0.29, or about 7.49%.

[10] The calculation is as follows: ((454,515/1,569,757)+0.0546)*1,569,757 = 540,223.732.

Appendix B. Regression Discontinuity

B1. Regression Discontinuity Sample Summary

Table B1. Summary statistics for the regression discontinuity sample.

	Count	Mean	Sd	Min	Max
# Citations	538,054	2.14	36.67	0.00	5,702.40
# Collaborations	538,054	2.00	56.47	0.00	7,228.03
# Citations (firms)	538,054	1.96	35.20	0.00	5,498.18
# Collaborations (firms)	538,054	1.90	55.07	0.00	7,048.09
# Citations (academic)	538,054	0.04	0.49	0.00	67.16
# Collaborations (academic)	538,054	0.07	1.41	0.00	239.69
# Citations (high inventor mass)	538,054	1.76	33.21	0.00	5,159.64
# Collaborations (high inventor mass)	538,054	1.87	54.54	0.00	6,945.35
# Citations (low inventor mass)	538,054	0.23	2.68	0.00	394.66
# Collaborations (low inventor mass)	538,054	0.10	1.73	0.00	200.72
# Citations (high R&D)	538,054	0.28	7.04	0.00	1,168.81
# Collaborations (high R&D)	538,054	0.37	12.79	0.00	2,028.88
# Citations (low R&D)	538,054	0.50	13.38	0.00	2,236.98
# Collaborations (low R&D)	538,054	0.38	15.61	0.00	2,389.79
# Citations (multi-ethnic)	538,054	1.53	24.00	0.00	3,733.47
# Collaborations (multi-ethnic)	538,054	1.47	41.30	0.00	5,526.10
# Citations (co-ethnic)	538,054	0.61	13.08	0.00	1,968.93
# Collaborations (co-ethnic)	538,054	0.53	15.86	0.00	2,085.88
Has nonstop	538,054	0.49	0.50	0.00	1.00
Total # of nonstop flights	538,054	611.95	1,799.85	0.00	74,002.00
Distance (miles)	521,477	1,110.86	1,246.99	0.00	11,873.40
Hub-to-Hub flight	537,878	0.26	0.44	0.00	1.00
Working hour overlap	537,878	7.04	1.58	0.00	8.00
Immigrant friendliness distance	402,523	0.19	0.28	0.00	1.87
Average price	20,350	946.48	528.11	214.30	4,538.75
Average duration (hours)	20,350	13.72	4.65	4.48	41.50

Note: This table provides the summary statistics for the regression discontinuity sample, which includes location pairs that are just above and below the 6000-mile threshold in terms of flight distance.

Table B2. Subsample correlations

	N	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) Hub-to-Hub	3,795	1.000							
(2) Leader-Leader	3,795	0.305	1.000						
(3) Follower-Follwer	3,795	-0.010	-0.211	1.000					
(4) Leader-Follower	3,795	0.077	-0.356	-0.207	1.000				
(5) Friendly-Friendly	3,795	0.119	-0.179	-0.171	0.051	1.000			
(6) Friendly-Unfriendly	3,795	0.012	0.225	0.005	-0.115	-0.635	1.000		
(7) Unfriendly-Unfriendly	3,795	-0.141	-0.090	0.176	0.089	-0.274	-0.568	1.000	
(8) High Business Hour Overlap	3,795	-0.437	-0.313	0.045	-0.089	0.012	-0.094	0.104	1.000

Note: Observations at the airport pair year level, excluding singletons.

B2. First Stage Regressions

In this section, we show evidence that the 6000-mile threshold is associated with a meaningful discontinuity in the number of available nonstop flights. This relationship is the "first stage" regression in the fuzzy regression discontinuity. Regressions in the table below show that airport pairs just below the 6000-mile threshold (to the left of the dotted line) have significantly more nonstop flights than airport pairs just above the threshold (to the right of the dotted line). Figure B1 plots the number of nonstop flights between routes, using a linear fit. It shows that the number of nonstop flights above the 6000-mile threshold exhibits a downward trend. The discontinuity uses bins computed through the IMSE-optimal evenly-spaced method using spacings estimators. Optimal bandwidth computation follows the methodology described in Calonico et al. (2020), who build on the work by Imbens and Kalyanaraman (2012).

Table B3. First stage of regression discontinuity.

	(1)	(2)	(3)	(4)
Dep. Var.: Nonstop Flights (asinh)				
Bandwidth:	250	500	750	Optimal
under6000	2.286***	1.313***	1.340***	1.418***
	(0.309)	(0.253)	(0.219)	(0.236)
dist6000	-0.007**	0.000	0.000	-0.001
	(0.003)	(0.001)	(0.000)	(0.001)
dist6000 x under6000	-0.003	-0.003***	-0.002***	-0.001
	(0.004)	(0.001)	(0.001)	(0.001)
Constant	2.198***	2.824***	2.654***	2.530***
	(0.276)	(0.208)	(0.169)	(0.189)
Country-pair-year FE	Y	Y	Y	Y
Observations	1375	3300	5368	4323
Adjusted R^2	0.439	0.403	0.379	0.392

Note: This table estimates the first stage, using several bandwidths for the estimation in terms of pair of airports at either side of the 6000 miles threshold: 250, 500 and 750 miles, as well as the optimal bandwidth. The Optimal bandwidth computation follows the methodology described in Calonico et al. (2020), who build on the work by Imbens and Kalyanaraman (2012). The estimation uses a triangular weight scheme, giving higher weight to observations closer to the threshold. All specifications include country-country-year fixed effects. Standard errors clustered at the country-country-year level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

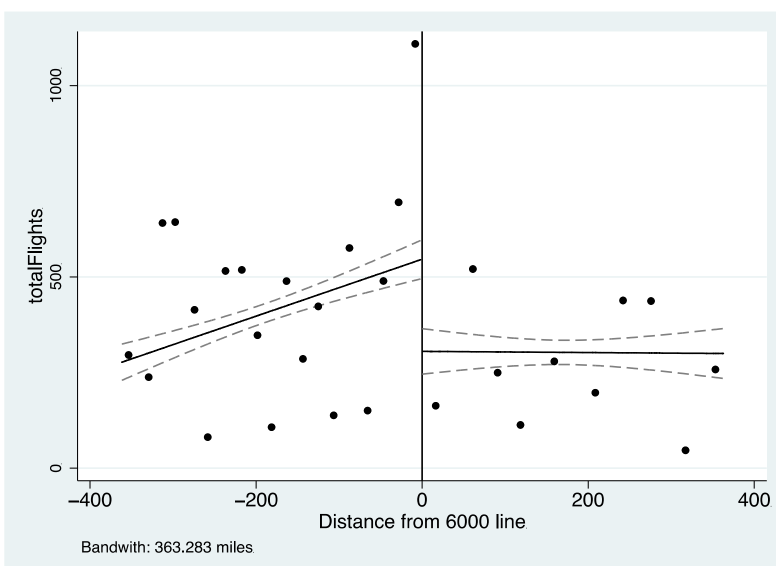


Figure B1. Discontinuity in number of flights with linear fit.

B3. Varying the number of bins

One concern regarding the regression discontinuity estimation is that the graphical results may depend on the number of bins used on either side of the threshold. The "optimal" number of bins attempt to minimize the integrated mean-squared error by balancing the trade-off between squared-bias and variance of the local sample means (Cattaneo et al. 2018). In the paper, we utilize the evenly-spaced, or ES method of choosing the number of bins. The ES method yields 15 bins to the left, and 9 bins to the right of the threshold for the RD on citations, and 5 bins to the left, and 15 bins to the right of the threshold for collaborations. Below, we show that the graphical results are robust to varying the number of bins from 25, 50, 100, and 150.

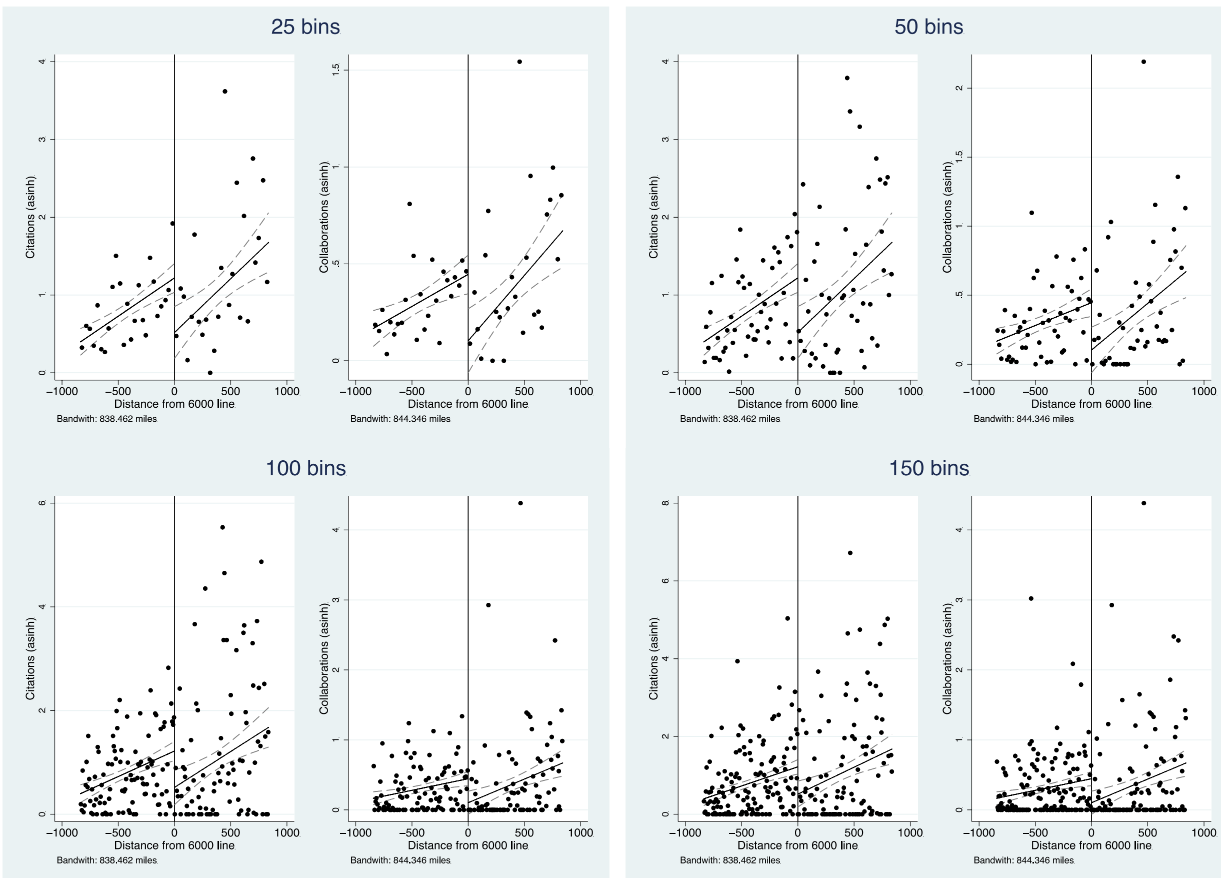


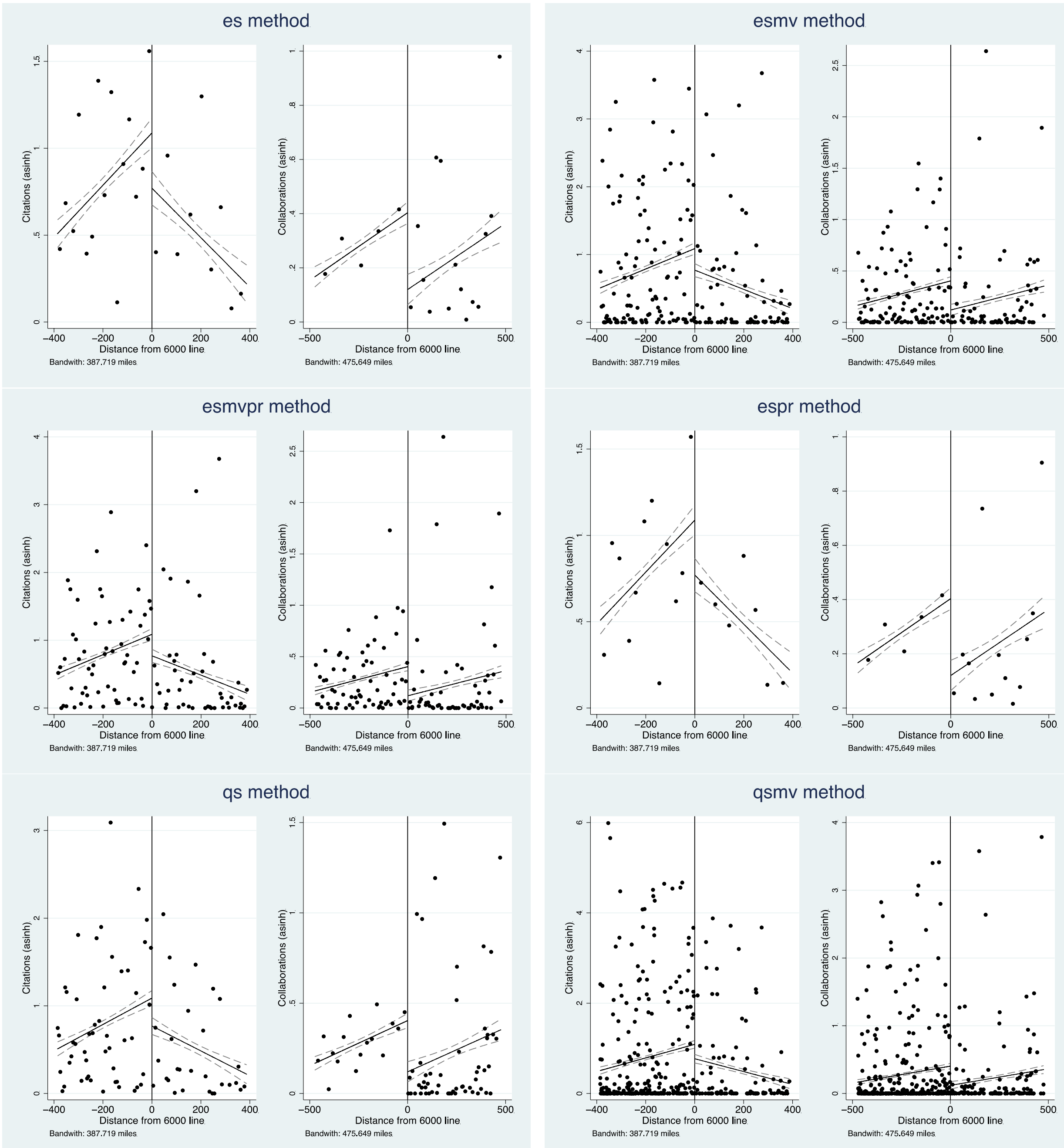
Figure B2. Collaborations and citations below and above the 6000-mile threshold, with varying numbers of bins.

B4. Varying the bin selection method

The results are also robust to changes in the bin selection method. The quantile-spaced (QS) method is a popular alternative to ES methods. In the QS method, bins are selected so that each bin contains the same number of observations. We present eight different types of bin selection methods:

- es: IMSE-optimal evenly-spaced method using spacings estimators.
- espr IMSE-optimal evenly-spaced method using polynomial regression.
- esmv mimicking variance evenly-spaced method using spacings estimators.
- esmvpr mimicking variance evenly-spaced method using polynomial regression.
- qs IMSE-optimal quantile-spaced method using spacings estimators.
- qspr IMSE-optimal quantile-spaced method using polynomial regression.
- qsmv mimicking variance quantile-spaced method using spacings estimators.
- qsmvpr mimicking variance quantile-spaced method using polynomial regression.

Since we are restricting the support to the optimal bandwidth, we use only a first-order polynomial to fit the regressions. This is to reduce concerns of overfitting and minimize boundary effects (Cattaneo et al 2019). Figure B3 presents the results with various bin selection methods. We see from the figures that the binning method does not visually change our results. Generally, the ESMV and QSMV are good choices to best depict the overall RD design. Since each point is based on quantiles, the QS method provides an accurate representation of the concentration of observations along our score (airport distance), while the ES method provides similar information, without the direct link to the concentration along the score support.


es method
esmv method
esmvpr method
espr method
qs method
qsmv method
Citations (asinh)
Collaborations (asinh)
Distance from 6000 line
Bandwith: 387.719 miles
Bandwith: 475.649 miles

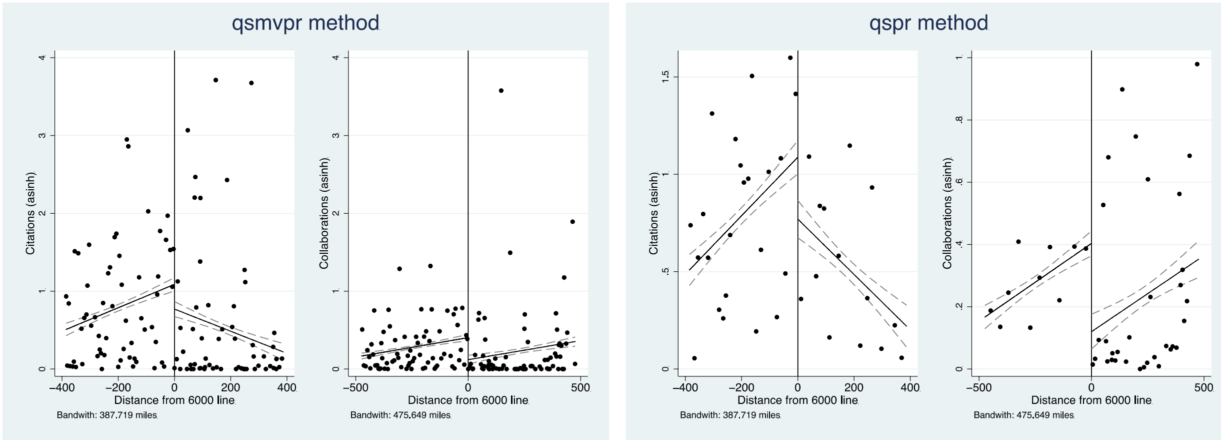


Figure B3. Collaborations and citations below and above the 6000-mile threshold, with different bin selection methods. Bin selections are based on combinations of evenly spaced or quantile based, mimicking variance, and whether polynomial regressions are used.

B5. Varying kernel choice

Next, we test whether the choice of kernel affects the graphical results. The choice of kernel affects how points near the threshold are weighted. Our preferred specification uses a triangular kernel, but graphical results are robust to using either the Epanechnikov kernel or a uniform kernel. For each of the plots below, we see that at the 6000-mile threshold, there exists a visual discontinuity in the number of citations as well as collaborations.

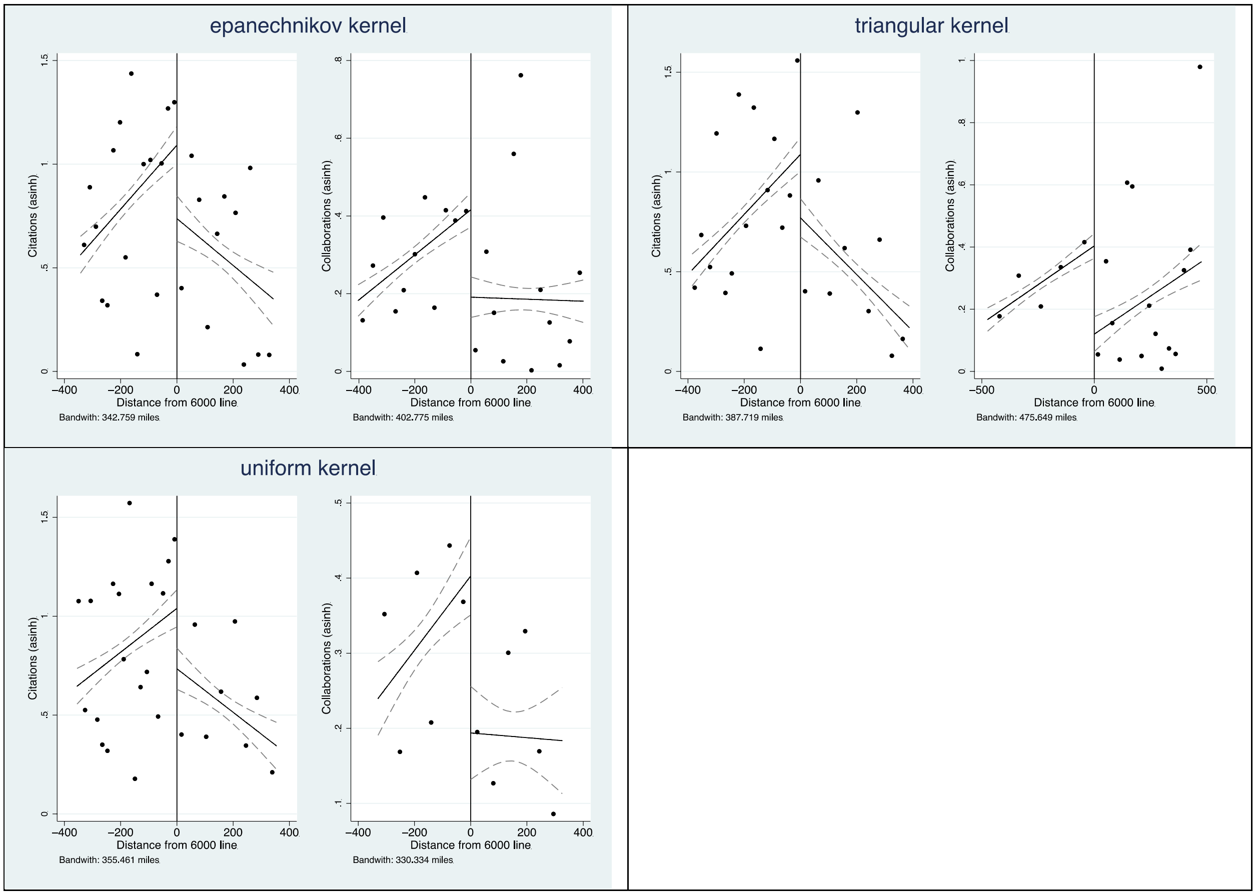


Figure B4. Collaborations and citations below and above the 6000-mile threshold, with different kernel choices. Kernel choices determine how much more weight to give to data points near the cutoff.

B6. Fixed Effects and Clustering

This section, we relax our use of fixed effects at the country-pair year level. Our preferred specification includes country-country-year fixed effects, absorbing the average innovation flows between two countries in a given year. Thus, for each route, we are comparing the effect of an additional flight above and beyond what can be explained by the country pair in a given year. Alternatively, we can include separate fixed effects for the origin country and destination country.

Table B4. Regression discontinuity with different fixed effects.

	Citations				Collaborations			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OC+DC+Y	OCY+DCY	OA+DA	OAY+DAY	OC+DC+Y	OCY+DCY	OA+DA	OAY+DAY
asinh(Flights)	0.605***	0.605***	0.451	0.451	0.454***	0.454***	0.755	0.755
	(0.212)	(0.221)	(0.376)	(0.304)	(0.151)	(0.171)	(0.966)	(0.933)
dist6000	0.001*	0.001*	-0.001	-0.001	0.000	0.000	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
dist6000 # under6000	-0.001	-0.001	0.001	0.001	0.000	0.000	-0.002	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.002)	(0.002)
Observations	3960	3498	3960	2156	4983	4521	4983	3069

Note: Standard errors in parentheses. Fixed effects denoted as following: OC=Origin Country, DC=Destination Country, OA=Origin Airport, DA=Destination Airport, OCY=Origin Country-Year, DCY=Destination Country-Year, OAY=Origin Airport-Year, DAY=Destination Airport-Year. Standard error clustered at the level of fixed effects except for Columns 3 and 7 which use robust standard errors. Both dependent variables are asinh-transformed. In this table, we have included fixed effects for origin country, destination country, and year (Columns 1 and 5), origin country-year, destination country-year (Columns 2 and 6), origin airport, destination airport (Columns 3 and 7), and finally origin airport-year, destination airport-year (Columns 4 and 8).

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B5. Regression discontinuity with different levels of clustering.

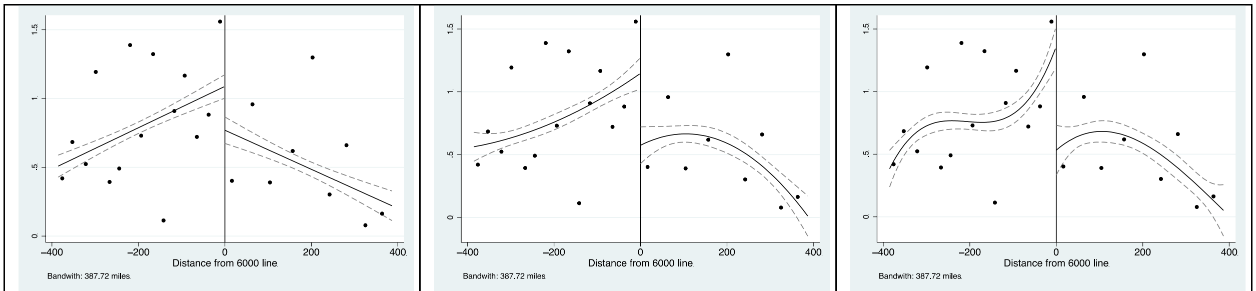
	Citations					
	(1)	(2)	(3)	(4)	(5)	(6)
	OC+DC	OCY+DCY	CCY	OAY+DAY	OA+DA	Robust
asinh(Flights)	0.285	0.285***	0.285***	0.285***	0.285	0.285***
	(0.227)	(0.081)	(0.076)	(0.090)	(0.255)	(0.082)
dist6000	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
dist6000 # under6000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.003)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)
Observations	2332	2332	2332	2332	2332	2332
R^2	0.612	0.612	0.612	0.612	0.612	0.612

Collaborations					
(7)	(8)	(9)	(10)	(11)	(12)
OC+DC	OCY+DCY	CCY	OAY+DAY	OA+DA	Robust
0.169	0.169***	0.169***	0.169**	0.169	0.169**
(0.145)	(0.056)	(0.059)	(0.070)	(0.192)	(0.067)
-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
0.000	0.000	0.000	0.000	0.000	0.000
(0.001)	(0.000)	(0.001)	(0.001)	(0.002)	(0.000)
3146	3146	3146	3146	3146	3146
0.442	0.442	0.442	0.442	0.442	0.442

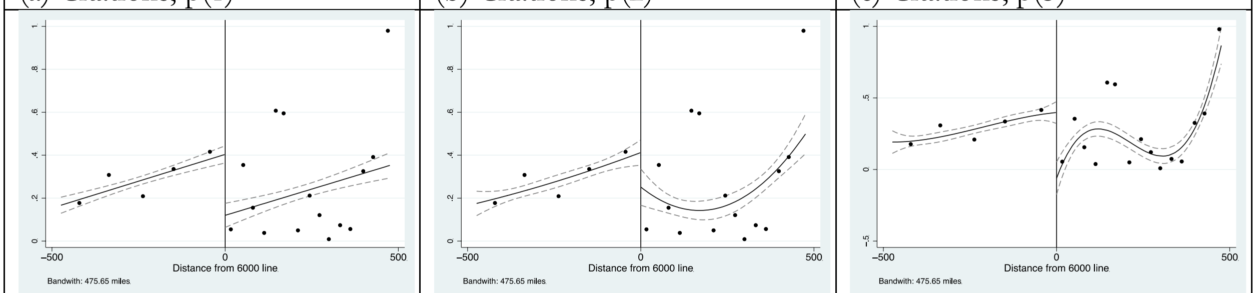
Note: Standard errors in parentheses, clustered at the country-pair-year level. All specifications include country-pair-year fixed effects. Both dependent variables are asinh-transformed. This table includes two-way clustering at the origin country plus year clustering (OCY) and at the destination country plus year clustering (DCY), as well as clustering at the origin airport plus year clustering (OAY) and at the destination airport with year clustering (DAY). After restricting the sample to around the 6000-mile threshold, we have 3,795 observations across 1,116 clusters.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B7. Higher order polynomials for the reduced form

We next check whether our RD results are sensitive to the polynomial order that we use. Lee and Lemieux (2010) suggest that plotting higher order polynomials may enhance the visual impact of the graph, and also suggest that it is essential to check that the RD results are robust to the inclusion of higher order polynomial terms. Thus, in this section, we check whether graphically, our results are robust to modeling higher order polynomials.



(a) Citations, p(1) | (b) Citations, p(2) | (c) Citations, p(3)



(d) Collaborations, p(1) | (e) Collaborations, p(2) | (f) Collaborations, p(3)

Figure B5. Higher order polynomial estimates of the regression discontinuity.
Each figure contains plots using the rdplot command in Stata, using polynomials of differing orders.

We see that for both citations and collaborations, we see a discontinuous jump at the 6000-mile mark. In all panels, we observe the 95% confidence intervals do not overlap, and the results are therefore statistically significant. The magnitude of the effects seem to change slightly: the impact of nonstop flights on citations increases when using higher order polynomials, but the effect on collaborations decreases when using a second order polynomial.

B8. Placebo thresholds

Next, we test whether our treatment of increased operating costs at the 6000-mile threshold is indeed meaningful. Towards this, we check whether we observe (or do not observe) similar discontinuities at placebo thresholds {Citation}. Since the 6000-mile threshold is an approximate cutoff for a 12-hour flight, it is a fuzzy discontinuity. Thus, far away from the 6000-mile threshold, we should see the treatment effect is zero. To test this, we estimate the regression discontinuity at various placebo thresholds. Specifically, for $c^* \in [4000, 9000]$ in 25 mile intervals, we estimate a 2SLS model with the first stage equation given by:

$$
\begin{aligned}
&Nonstop\ Flights_{a_o,a_d,t} \\
&\qquad = \gamma_1 1\{Dist_{a_o,a_d} < c^*\} + \gamma_2 (Dist_{a_o,a_d} - c^*) \\
&\qquad + \ \gamma_3 1\{Dist_{a_o,a_d} < c^*\} \times (Dist_{a_o,a_d} - c^*) + \phi_{c_o,c_d,t} + \epsilon_{a_o,a_d,t}
\end{aligned}
$$

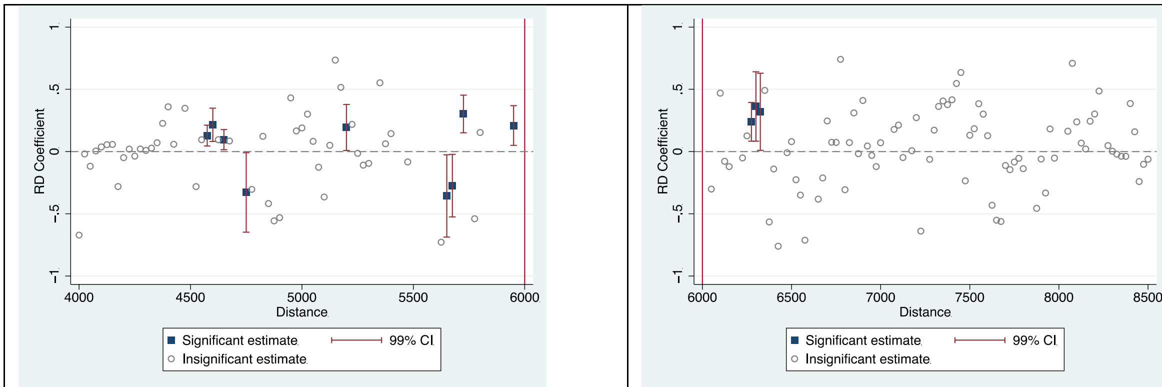
And the second stage is given by:

$$
\begin{aligned}
\log(Y_{a_o,a_d,t}) &= \beta_1 \ \widehat{Nonstop\ Flights}_{a_o,a_d,t} \quad + \beta_2 (Dist_{a_o,a_d} - c^*) \\
&+ \beta_3 1\{Dist_{a_o,a_d} < c^*\} \times D(Dist_{a_o,a_d} - c^*) + \phi_{c_o,c_d,t} + \epsilon_{a_o,a_d,t}
\end{aligned}
$$

The idea is that if there is a sharp discontinuity at the 6000-mile mark, for all $c^* \neq 6{,}000$, we should see that $\widehat{\beta_1} = 0$. Since our discontinuity is fuzzy, we should see instead that for cutoffs far away from the 6000-mile mark, we don't observe significant coefficients. As we will discuss, since citation and collaboration outcomes are driven by airport pairs with high innovation mass, we may observe some cutoffs that show significant coefficient estimates (e.g., Pacific flights connecting the U.S. with Asia). To avoid misspecification, we split the sample into [4000, 5950] and [6050, 9000] and estimate the discontinuity for those two samples. We first present results for citations and collaborations.

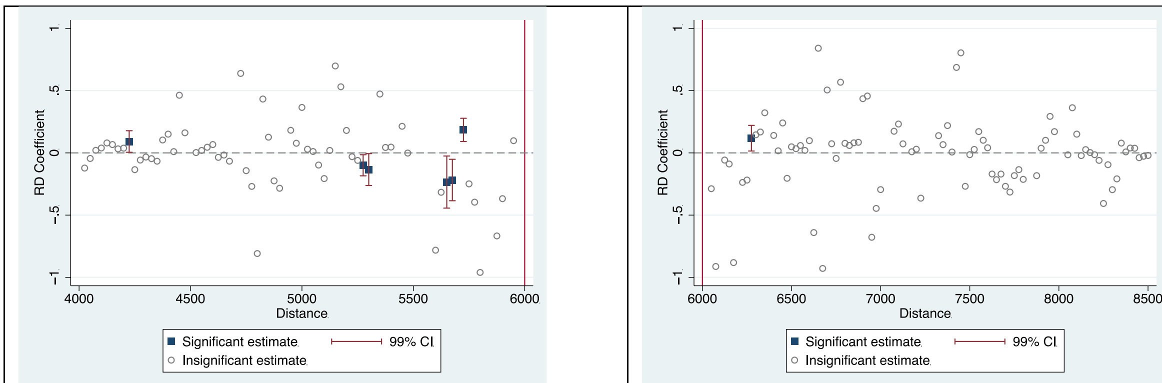
For citations (Figure B5), we see that most of the coefficients are statistically insignificant from zero. Since this is a fuzzy regression discontinuity, we see some significant placebo thresholds near the 6000-mile line. There appear to be a cluster of significant coefficients near the 4500 mile line, or a 9 hour flight. This effect is driven by the presence of airport pairs with high innovation mass, particularly between the US West coast and Asian countries, and is to be expected. Specifically, for citations, the placebo thresholds of 4575, 4600, 4650, 4750 miles are significant. Note that coefficients below the 4700-mile mark are positive, but become negative as the placebo threshold shifts right. As the placebo threshold moves, high-innovation mass routes between Asia and the US (e.g., SEA-HND, PDX-NRT) switch from control to treatment, and decreases the effect of additional flights. Thus, we are confident such changes are unrelated to our 6000-mile instrument.

Similarly, for collaborations (Figure B6), we see that most coefficients are statistically insignificant from zero, and the majority of significant coefficients can be found between 5500-6500. We also conduct the placebo test over the entire support of the dataset (Figure B7). Again, for collaborations, there is a small cluster of significant coefficients near the 4500-mile mark, but there is stronger evidence of a cutoff near the 6000-mile threshold.



(1) Airports closer than 6000 miles	(2) Airports farther than 6000 miles

Figure B5. Effect of placebo thresholds on citations
Each dot represents an estimate of β_1, the effect of a 1% increase in nonstop flights between two airports, assuming there exists a discontinuity at the given distance. Red lines denote 99% confidence intervals. Gray dots denote coefficients that are insignificant. Coefficients with absolute values greater than 1 are omitted for visualization purposes (all insignificant).



(1) Airports closer than 6000 miles	(2) Airports farther than 6000 miles

Figure B6. Effect of placebo thresholds on collaborations
Each dot represents an estimate of β_1, the effect of a 1% increase in nonstop flights between two airports, assuming there exists a discontinuity at the given distance. Red lines denote 99% confidence intervals. Gray dots denote coefficients that are insignificant. Coefficients with absolute values greater than 1 are omitted for visualization purposes (all insignificant).

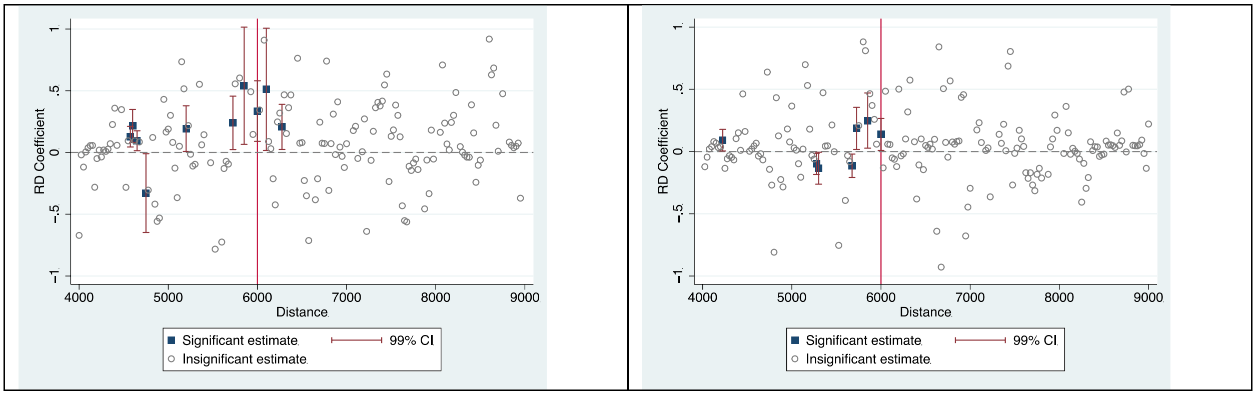


Figure B7. Effect of placebo thresholds across the entire support
Each dot represents an estimate of β_1, the effect of a 1% increase in nonstop flights between two airports, assuming there exists a discontinuity at the given distance. Red lines denote 99% confidence intervals. Gray dots denote coefficients that are insignificant. Coefficients with absolute values greater than 1 are omitted for visualization purposes (all insignificant).

B9. Density of running variable

An important assumption underlying the RDD is that the subjects cannot precisely manipulate their scores (Imbens and Lemieux 2008). This may happen, for instance, for birthdays and school year cutoffs. In the case of airport pairs, such concerns may arise if airport authorities decide locations so that they maximize the number of airports within 6000 miles. While unlikely, if it were the case, we would see a bump in an airport's "potential connections" just below the 6000-mile threshold. We do a nonparametric test using airport locations and all possible airport pair permutations, and plotting whether there are more potential airport pairs just below the 6000-mile line. From the figure below, we see that there is no evidence of bunching near the 6000-mile threshold.

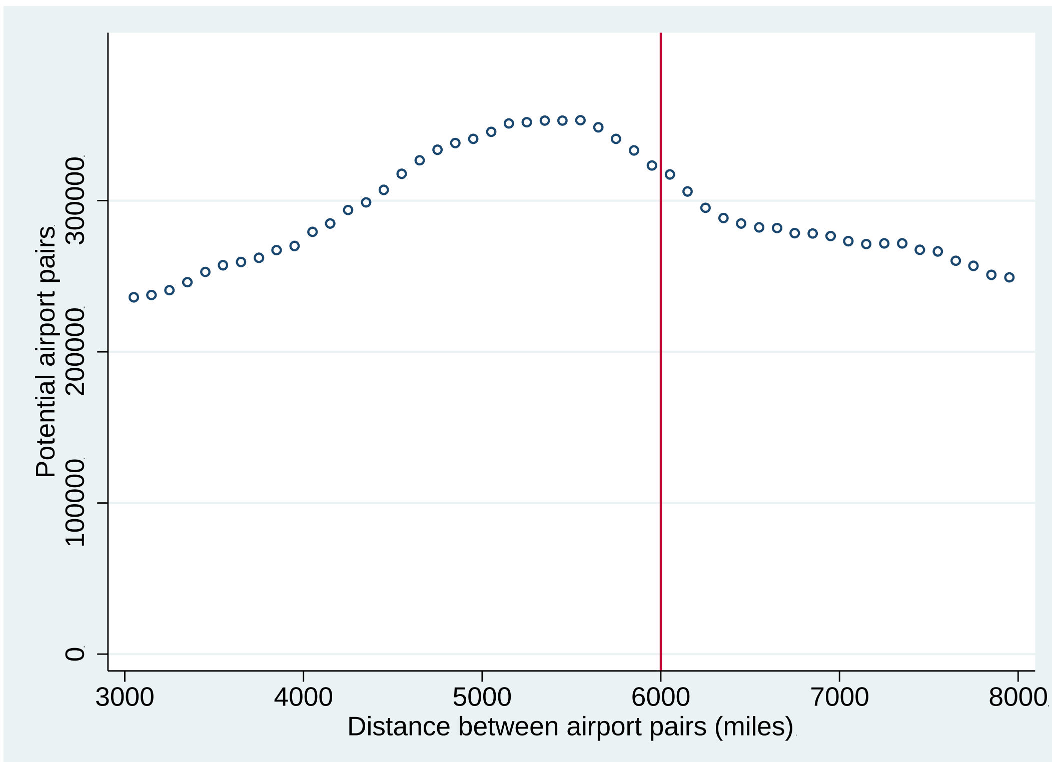


Figure B8. Nonparametric test of bunching near the 6000-mile threshold
Each point on this plot represents the "potential" number of airport pairs (y-axis) within a specified distance (x-axis). The "potential" number of airport pairs is taken from counting all pairwise combinations of all airport pairs in our dataset.

B10. Cross-assignee results

To better understand our results, we also study whether the effect of nonstop flights is greater within or across assignees. An assignee can be either an inventor or, in most cases, a firm to which a patent's ownership is assigned. One would expect that if indeed flight connectivity causes more knowledge flows, it should in fact disproportionately facilitate cross-assignee knowledge flows. This is because cross-assignee knowledge flows, which usually involve the crossing of geographic and/or organizational boundaries, are typically costlier than knowledge flows within firms.

To explore this possibility, we count the number of cross-assignee and within-assignee citations that happen between airports. To count the number of cross-assignee citations, we exclude patent-citation pairs that share the same assignee, then aggregate the number of citations to the airport-pair level. Similarly, we count only those patent-citation pairs that share the same assignee for within-assignee citations. For any airport pair, the sum of the cross-assignee and within-assignee citations is equal to the citations number defined above.

Table below shows the results from estimating the RD specification (equation *(1)* in the main paper) using cross-assignee and within-assignee citations as dependent variables. The dependent variable for Columns 1 and 2 is the number of citations across different assignees for a given airport pair, while the dependent variable for Columns 3 and 4 is the number of citations within the same assignee. The estimates for cross-assignee knowledge spillovers (Columns 1 and 2) are similar in magnitude to the estimates of the full count shown in the main results. This is because the 95% of the citations in our data are cross-assignee citations. With this caveat, nonstop flights have little effect on within-assignee knowledge flows as shown in Columns 3 and 4, where we cannot distinguish any statistical significance in the estimates. But this could be because in Columns 3 and 4 has very little variation to exploit.

Table B6. Regression discontinuity analysis of within-assignee and cross-assignee citations.

	Cross Assignee		Within Assignee	
	(1) Citations	(2) Citations	(3) Citations	(4) Citations
Bandwidth	500	Optimal	500	Optimal
Nonstop Flights (asinh)	0.3328***	0.2670***	0.0749**	0.0312
	(0.0937)	(0.0689)	(0.0351)	(0.0218)
(6000 – Distance)	0.0010**	0.0003	0.0005**	-0.0001
	(0.0004)	(0.0002)	(0.0002)	(0.0001)
(6000 – Distance) x Under6000	-0.0011	-0.0002	-0.0009**	-0.0001
	(0.0008)	(0.0003)	(0.0004)	(0.0001)
Observations	3300	4323	3300	4323
R^2	0.509	0.625	0.243	0.423

Note: Standard errors in parentheses, clustered at the country-country level. Both dependent variables are asinh-transformed. Optimal bandwidth calculation follows the methodology described in Cattaneo et al. (2018). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B11. Effects not driven by one-stop flights

This section further discusses how the existence of nonstop flights (and not other types of multi-stop flights) are driving the results. For the subsample of routes for which flight duration and prices are available, we test whether the 6000-mile threshold drives flights with layovers, and how innovation outcomes correlate with routes with nonstop/one-stop/multiple stops. Again, the subsample of routes includes those flights that are longer than 3000-miles apart, and have more than 1000 flights total in our 2005-2015 time period.

Figure B9 shows the results from a kernel-weighted local polynomial regression of the probability of a route having one or two stop flights, and the length of the route. We see that around the 6000-mile mark, there is no discernable difference in the likelihood of having one-stop flights, and two-stop flights seem to increase for longer flights.

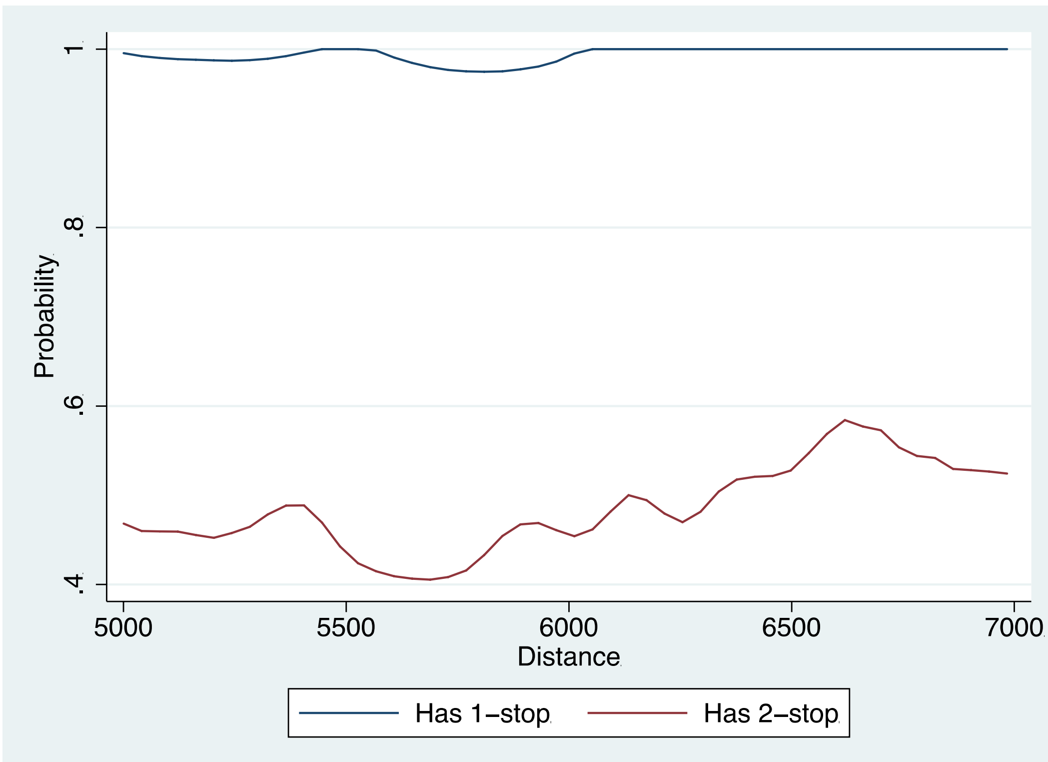


Figure B9. Probability of having 1-stop and 2-stop flights across distance

We confirm the visual results in regression form in the table below. We see that for all bandwidths, the 6000-mile discontinuity does not impact the probability of having a one-stop flight. Thus, routes just beneath and just above the 6000-mile threshold we use in the regression discontinuity differ only in the existence of nonstop flights, not one-stop flights. While not shown here, two-stop flights slightly increase after 6000 miles, but this would go against our findings.

Table B7. First stage using one-stop flights only.

	(1)	(2)	(3)	(4)	(5)
Dep. Var.: Has One-stop Flight					
Bandwidth (miles)	500	750	1000	1250	Optimal
Under6000	0.001	-0.027	-0.011	-0.007	-0.012
	(0.001)	(0.027)	(0.011)	(0.008)	(0.012)
6000 – Distance	-0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
(6000 – Distance) x under6000	0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.989^{***}	0.987^{***}	0.991^{***}	0.992^{***}	0.991^{***}
	(0.001)	(0.008)	(0.003)	(0.002)	(0.003)
Observations	78	122	197	269	184
R^2	0.494	0.340	0.332	0.331	0.332

Note: This table estimates the first stage, using several bandwidths for the estimation in terms of pair of airports at either side of the 6000 miles threshold: 500, 750, 1000, 1250 miles, as well as the optimal bandwidth. The optimal bandwidth is computed using the methodology described in Cattaneo et al. (2018) who build on the work by Imbens and Kalyanaraman (2012). In addition, the estimation uses a triangular weight scheme, giving higher weight to observations closer to the cutoff point. All specifications include country-country fixed effects. Standard errors clustered at the country-country level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B12. Poisson regressions, Log+1 transformations and raw counts

Our use of the asinh-transformed variables is motivated by Burbidge, Magee, and Robb (1988), MacKinnon and Magee (1990) who show the inverse hyperbolic sine transformation allows researchers to preserve observations with zero flights.

We used the Stata package PPMLHDFE to check whether our results hold in a sharp regression discontinuity setting. That is, we estimate the following specification

$$Y_{a_o,a_d,t} = f\big(\gamma_0 + \gamma_1 Under6000_{a_o,a_d} + \gamma_2 Dist6000_{a_o,a_d} + \gamma_3 Under6000_{a_o,a_d} \times Dist6000_{a_o,a_d} + X_{a_o,a_d}\xi\big)$$

Our coefficient of interest is γ_1, the magnitude of the discontinuity at the 6000-mile mark. The estimates are reported below.

Table B8. Regression discontinuity using PQML.

	Overall		Academic		Firms	
	(1)	(2)	(3)	(4)	(5)	(6)
	Citations	Collaborations	Citations	Collaborations	Citations	Collaborations
under6000	1.555***	0.662***	1.870***	1.599***	1.556***	0.634***
	(0.143)	(0.238)	(0.243)	(0.482)	(0.146)	(0.243)
dist6000	-0.004***	-0.002**	-0.003***	-0.002**	-0.004***	-0.002**
	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)
dist6000 # under6000	0.005***	0.001	0.004***	0.001	0.005***	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	1.921***	1.160***	-1.951***	-2.141***	1.856***	1.144***
	(0.154)	(0.247)	(0.225)	(0.387)	(0.158)	(0.250)
Observations	2472	1901	1074	731	2418	1870

Note: Standard errors in parentheses. Since these are Poisson models, the dependent variables are raw counts of citations and collaborations at the airport pair year level. Note that the number of observations for Poisson regressions is smaller than for linear regressions because of the separation problem (Correia, Guimaraes, and Zylkin, 2021; Santos Silva and Tenreyro, 2006). In short, maximum likelihood solutions of Poisson models may not have a solution when regressors are perfectly collinear over the subsample where the dependent variable is nonzero. The solution implemented by PPMLHDFE is to drop observations that are separated.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Instead of scaling by the number of flights in the first stage, we also scale by the existence of nonstop flights. Specifically, our first and second stages are

$$HasNonstop_{a_o,a_d,t} = \alpha_0 1(Dist6000 > 0) + \alpha_1 Dist6000 + \alpha_2 1(Dist6000 > 0) \times Dist6000 + \epsilon$$

$$Y_{a_o,a_d,t} = \beta_0\ Has\widehat{Nonstop}_{a_o,a_d,t} + \beta_1 Dist6000 + \beta_2 1(Dist6000 > 0) \times Dist6000 + \epsilon$$

The coefficient of interest is β_0, the predicted probability of having a nonstop flight. β_0 thus measures how how having any nonstop flight affects citations and collaborations. This is in contrast with our main specification, which measures the impact of increased flights. The table below shows the results.

Table B9. Discrete effect of having nonstop flights on knowledge diffusion.

	Citations		Collaborations	
	(1)	(2)	(3)	(4)
Bandwidth:	550	Optimal	550	Optimal
HasNonstop	3.140***	2.604***	1.376**	1.556**
	(1.167)	(0.878)	(0.584)	(0.682)
dist6000	-0.001**	-0.001	-0.000	-0.000
	(0.001)	(0.001)	(0.000)	(0.000)
c.dist6000#c.under6000	0.002	0.001	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
N	3795	2332	3795	3146

Note: Standard errors in parentheses, clustered at the country pair-year level. HasNonstop is equal to 1 if the number of total flights at the airport pair-year is greater than zero.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B10. Regression discontinuity using raw counts of citations and collaborations as dependent variables.

	Overall		Academic		Firms	
	(1)	(2)	(3)	(4)	(5)	(6)
	Citations	Collaborations	Citations	Collaborations	Citations	Collaborations
Nonstop Flights (asinh)	13.164***	0.722**	0.218***	0.058**	12.355***	0.660**
	(3.737)	(0.293)	(0.069)	(0.025)	(3.495)	(0.281)
dist6000	-0.083***	-0.004**	-0.001***	-0.000*	-0.078***	-0.004**
	(0.023)	(0.002)	(0.000)	(0.000)	(0.022)	(0.002)
dist6000 # under6000	0.128***	0.005*	0.002***	0.000*	0.120***	0.005*
	(0.036)	(0.003)	(0.001)	(0.000)	(0.033)	(0.003)
Observations	3795	3795	3795	3795	3795	3795

Note: Standard errors in parentheses. The dependent variables use raw counts of citations and collaborations.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B11. Regression discontinuity using log+1 transformation.

	Overall		Academic		Firms	
	(1)	(2)	(3)	(4)	(5)	(6)
	Citations	Collaborations	Citations	Collaborations	Citations	Collaborations
Nonstop Flights (asinh)	0.304***	0.124***	0.077***	0.032**	0.296***	0.121***
	(0.088)	(0.043)	(0.023)	(0.012)	(0.086)	(0.042)
dist6000	-0.001**	-0.000	-0.000**	-0.000*	-0.001**	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
dist6000 # under6000	0.001*	0.000	0.000**	0.000*	0.001**	0.000
	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Observations	3795	3795	3795	3795	3795	3795

Note: Standard errors in parentheses. Both dependent variables are transformed using "log + 1".
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B13. Pooling airports at cities

Table B12. Pooled airport analysis.

	Overall		Academic		Firms	
	(1) Citations	(2) Collaborations	(3) Citations	(4) Collaborations	(5) Citations	(6) Collaborations
Nonstop Flights (asinh)	0.303***	0.133***	0.089***	0.036***	0.295***	0.129***
	(0.082)	(0.045)	(0.025)	(0.013)	(0.081)	(0.043)
dist6000	-0.001*	-0.000	-0.000*	-0.000*	-0.001*	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
dist6000 # under6000	0.001	0.000	0.001**	0.000*	0.001	0.000
	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Observations	3322	3322	3322	3322	3322	3322

Note: Standard errors in parentheses, clustered at the country-country-year level. All specifications include country-country-year fixed effects. Observations are at the city pair-year level. Both dependent variables are asinh-transformed. In this table, we pool all airports in a given city. Specifically, we collected a list of cities served by multiple airports[11] and used it to aggregate from the airport pair to the city pair level. For cities served by multiple airports, we took the sum of all citations, collaborations, and flights to create those measures at the city-pair level. For distances between cities, we take the average distance between all airports between the cities. We keep all other airport pairs.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B13. Flight effects on knowledge diffusion are greater for flights with at least one U.S. city.

	Entire Sample		At Least One U.S. City	
	(1) Citations	(2) Collaborations	(3) Citations	(4) Collaborations
Nonstop Flights (asinh)	0.335***	0.137***	0.959***	0.445***
	(0.095)	(0.050)	(0.276)	(0.138)
dist6000	-0.001**	-0.000	-0.002**	-0.001
	(0.000)	(0.000)	(0.001)	(0.001)
dist6000 # under6000	0.001	0.000	0.007***	0.003**
	(0.001)	(0.001)	(0.002)	(0.001)
Observations	3300	3300	858	858

Note: Bandwidth at 500 miles. Standard errors in parentheses, clustered at the country-pair-year level. All specifications include country-pair-year fixed effects. Both dependent variables are asinh-transformed.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[11] https://en.wikipedia.org/wiki/List_of_cities_with_more_than_one_commercial_airport

Appendix C. Mass Variables

Our paper adopts several measures for innovation "mass": innovation hubs, leaders versus followers, and firm-level data such as R&D spending. This section delves into the mass variables in more detail.

C1. List of hubs

Of the 5,015 airports in our dataset, 965 are near innovation hubs. Again, we categorize an airport as being near innovation hubs if the airport is within a 50-mile radius of the innovation hubs listed in Bikard and Marx (2020). Below, we present the 20 largest hubs and 20 largest non-hubs in our dataset, based on the total number of outbound flights from those airports in 2005-2015.

Table C1. List of hubs.

a) Hubs

Airport Code	Country (ISO Code)	Airport Name
ATL	US	Hartsfield Jackson Atlanta International Airport
LHR	GB	London Heathrow Airport
PEK	CN	Beijing Capital International Airport
HND	JP	Tokyo Haneda International Airport
ORD	US	Chicago O'Hare International Airport
LAX	US	Los Angeles International Airport
CDG	FR	Charles de Gaulle International Airport
FRA	DE	Frankfurt am Main Airport
DFW	US	Dallas Fort Worth International Airport
HKG	HK	Chek Lap Kok International Airport
DEN	US	Denver International Airport
MAD	ES	Madrid-Barajas Adolfo Suárez Airport
JFK	US	John F Kennedy International Airport
SIN	SG	Singapore Changi Airport
AMS	NL	Amsterdam Airport Schiphol
PVG	CN	Shanghai Pudong International Airport
PHX	US	Phoenix Sky Harbor International Airport
CAN	CN	Guangzhou Baiyun International Airport
LAS	US	McCarran International Airport

b) Non-hubs

Airport Code	Country (ISO Code)	Airport Name

DXB	AE	Dubai International Airport
BKK	TH	Suvarnabhumi Airport
CGK	ID	Soekarno-Hatta International Airport
IST	TR	Istanbul Atatürk International Airport
MNL	PH	Ninoy Aquino International Airport
GRU	BR	São Paolo/Guarulhos - Governador André Franco Montoro International Airport
KMG	CN	Kunming Changshui International Airport
BNE	AU	Brisbane International Airport
JED	SA	King Abdulaziz International Airport
DOH	QA	Hamad International Airport
XIY	CN	Xi'an Xianyang International Airport
CGH	BR	Congonhas Airport
BOG	CO	El Dorado International Airport
CTS	JP	New Chitose Airport
PMI	ES	Palma De Mallorca Airport
SGN	VN	Tan Son Nhat International Airport
BSB	BR	Brasília Presidente Juscelino Kubistschek International Airport
OKA	JP	Naha Airport
LIS	PT	Lisbon Portela Airport
GIG	BR	Rio de Janeiro/Galeão, Antonio Carlos Jobim International Airport

C2. Probability of hubs across distances

In this section, we explore the likelihood of two locations connected by nonstop flights both being innovation hubs. First, note that the frequency of hub-to-hub connections changes with the distance between the two airports. The figure below shows a local polynomial regression of the probability of a hub-to-hub connection on distance with 95% confidence intervals. Thus, each point on the solid line denotes the fraction of flights that connect two airports in innovation hubs, for a given distance. We see that longer flights tend to be between two airports both located in innovation hubs. We detail the implications below.

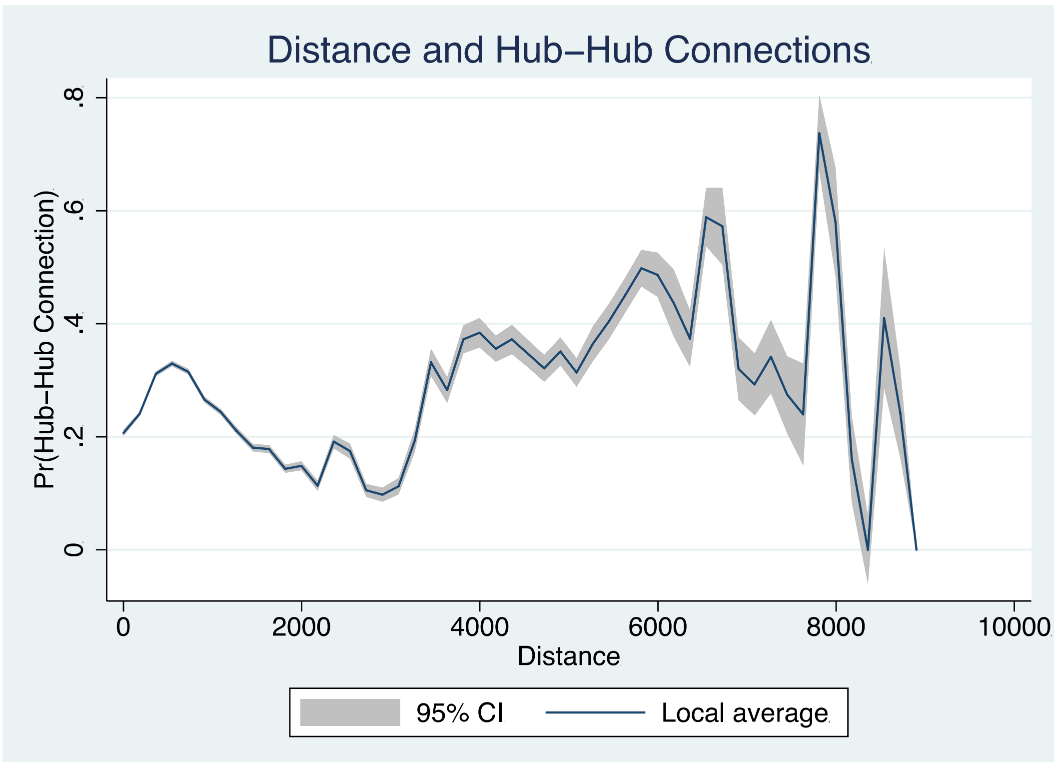


Figure C1. Probability of hub-to-hub connections
Local polynomial regression of the probability of a hub-to-hub connection against distance between airports.

Since hub-to-hub connections are more likely for airport pairs that are farther apart, we cannot distinguish between two mechanisms for the increase in innovation through nonstop flights. It is possible, for instance, that there are not enough ideas in non-hub locations, and thus is less knowledge spillovers. In contrast, it is also possible that nonstop flights tend to occur when both locations are innovation hubs, and we are capturing this effect indirectly.

C3. Innovation hubs and distance

In the main text, we see distance and innovation are positively correlated for airport routes that are more than 6000 miles apart, which is unexpected. However, a manual inspection of the airport pairs that have high innovation and are far apart shows these are likely driven by city pairs like Singapore-Newark, Singapore-Los Angeles, New York-Bangkok, Sydney-Dallas, Atlanta-Bombay, and so forth. These outliers in part drive the positive relationship between distance and innovations in our sample.[12] This relationship, however, is not present for shorter flights.

This section tests whether dropping routes with high levels of measured collaborations and citations impacts our results on travel duration and innovation. We repeat our analysis after dropping the top 10% of routes with most collaborations/citations. We see that increased travel duration is negatively correlated with innovation outcomes, especially for long-distance flights. Price, on the other hand, is mostly insignificant.

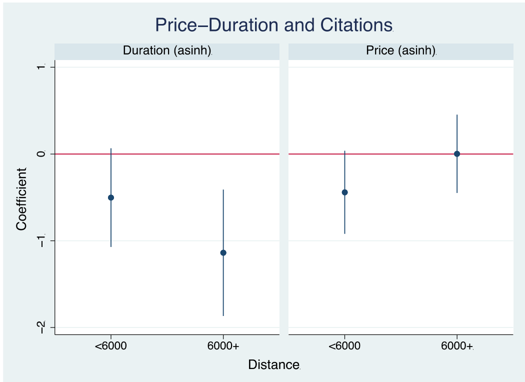
Table C2. Effects of flight duration and flight price on knowledge diffusion (without outliers).

	(1)	(2)	(3)	(4)
	Below 6000		Above 6000	
	Citations (asinh)	Collaborations (asinh)	Citations (asinh)	Collaborations (asinh)
Duration (asinh)	-0.298	-0.444**	-0.937**	-0.684**
	(0.193)	(0.188)	(0.408)	(0.302)
Price (asinh)	-0.210**	-0.044	-0.188	0.211
	(0.089)	(0.062)	(0.197)	(0.181)
Distance (asinh)	0.104	-0.201	2.368***	1.952***
	(0.599)	(0.611)	(0.866)	(0.659)
Constant	2.167	3.886	-15.670**	-16.183**
	(5.233)	(5.328)	(7.563)	(6.744)
Observations	477	512	317	417
R^2	0.737	0.610	0.807	0.581

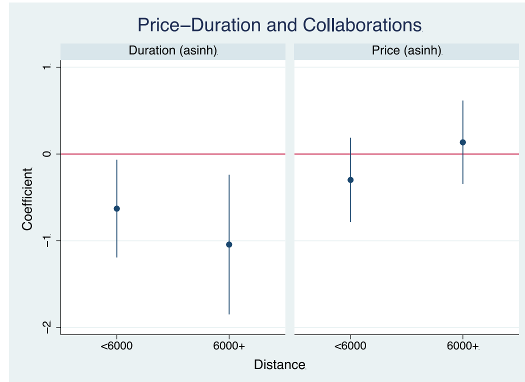
Note: Standard errors in parentheses, clustered at the country pair level. Outliers are defined as airport pairs with citations and collaborations above the 90% percentile. Inverse hyperbolic sine transformed variables are denoted with "asinh".
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We also plot this relationship using coefficient plots in the figure below.

[12] In section C3, we drop outlier routes (top 10%) that have very high levels of innovation. We see that the negative relationship between duration and collaboration/citations still hold, while the size of the coefficient on distance decreases. Price coefficients are mostly insignificant, but cheaper flights may increase citations in shorter flights.



(a) Coefficient plot of Duration and Price on Citations



(b) Coefficient plot of Duration and Price on Collaborations

Figure C2. Coefficients plots of flight duration and flight price on citations and collaborations

Association between price and duration on citations/collaborations changes with the distance. Each point on this graph is the coefficient for either Duration or Price from estimating equation (5) above for two subsamples. Subsamples are chosen based on how far apart airports are (i.e., less than or greater than 6000 miles apart). Vertical lines indicate 95% confidence intervals. All regressions include country-country fixed effects. Standard errors clustered at the country-country level.

C4. Hub to hub vs non-hub to hub

Table C3. Hub analysis further broken down into hub-hub, non-hub-non-hub, and non-hub-hub.

	Citations			Collaborations		
	(1) Hub-Hub	(2) Non-hub-Hub | Non-hub-Non-hub	(3) Non-hub-Hub	(4) Hub-Hub	(5) Non-hub-Hub | Non-hub-Non-hub	(6) Non-hub-Hub
Nonstop Flights (asinh)	0.532***	-0.045	-0.062	0.260***	0.051	0.076
	(0.180)	(0.032)	(0.047)	(0.092)	(0.036)	(0.057)
dist6000	-0.003**	-0.000	-0.000	-0.001	-0.000	-0.000
	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
dist6000 # under6000	0.006**	0.000**	0.000**	0.003*	0.000	0.000
	(0.003)	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)
Observations	1870	1760	1606	1870	1760	1606

Note: Standard errors in parentheses, clustered at the country-pair-year level. All specifications include country-pair-year fixed effects. Bandwidth at 550 miles. Hub denotes whether an airport is within a 50-mile radius of innovation hubs as defined in Bikard and Marx (2020). Both dependent variables are asinh-transformed.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C5. Leaders and followers

To differentiate between firms in innovation-leading countries ("leaders") and those in innovation-following countries ("followers"), we borrow from Furman and Hayes's 2004 Research Policy paper, which contains a list of countries that are categorized as leaders and followers based on their historical innovative productivity and advancement. Table 6 from Furman and Hayes (2004) contains the list of countries. This categorization sets precedence for measuring leaders versus followers and lends confidence to our analysis.

Leading innovating countries include the following: Germany, Japan, Sweden, Switzerland, and the United States. Firms and inventors in these countries are categorized as "leaders" in our analysis. Further, we categorize firms and inventors in the other countries in Furman & Hayes (2004) (*middle tier, third tier, and emerging innovators*) are categorized as "followers". Then, we restrict the sample to citations and collaborations by firms located in these leader and follower countries and conduct an analysis to gauge flights effects on citations and collaborations that occur 1) between leaders and 2) between a leader and a follower.

C6. Firm-level variables

In addition to R&D spending, which is used to measure firm-level innovation mass in the paper, we also have two additional firm-level measures for innovation mass: firm revenue and number of employees. The table below shows the split-sample analysis using these three variables.

Table C4. Firms with greater innovation mass benefit more from nonstop flights. (Collaborations)

Dep. Var.: Collaborations (asinh)	Revenue		Employees		R&D Spending	
	(1) High	(2) Low	(3) High	(4) Low	(5) High	(6) Low
Nonstop Flights (asinh)	0.086***	0.041	0.068***	0.053*	0.093***	0.031
	(0.028)	(0.027)	(0.025)	(0.028)	(0.030)	(0.028)
dist6000	-0.000	-0.000**	-0.000	-0.000**	-0.000*	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
dist6000 # under6000	0.000	0.000*	-0.000	0.001**	0.000	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	3795	3795	3795	3795	3795	3795

Note: Standard errors in parentheses, clustered at the country-pair-year level. All specifications include country-pair-year fixed effects. Bandwidth is 550 miles. Revenue, Employees, and R&D spending obtained from the Duke DISCERN database. "High" refers to above median firms in each mass category, while "Low" refers to below median. To generate this table, we use the Duke DISCERN dataset to match the assignees in our sample to Compustat firms. We use this Compustat matched data to categorize firms into large or small based on their revenue, R&D expenditure, and employee counts. Finally, we categorize assignees into large or small based on the number of publications and inventors. We present these results later in this response letter.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C5. Firms with more innovation mass benefit more from nonstop flights. (Citations)

Dep. Var.: Citations (asinh)	Revenue		Employees		R&D Spending	
	(1) High	(2) Low	(3) High	(4) Low	(5) High	(6) Low
Nonstop Flights (asinh)	0.197***	0.186***	0.197***	0.174***	0.180***	0.187***
	(0.053)	(0.056)	(0.053)	(0.054)	(0.052)	(0.054)
dist6000	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
dist6000 # under6000	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)
Observations	3795	3795	3795	3795	3795	3795

Note: Standard errors in parentheses, clustered at the country-pair-year level. All specifications include country-pair-year fixed effects. Bandwidth is 550 miles. Revenue, Employees, and R&D spending obtained from the Duke DISCERN database. "High" refers to above median firms in each mass category, while "Low" refers to below median. To generate this table, we use the Duke DISCERN dataset to match the assignees in our sample to Compustat firms. We use this Compustat matched data to categorize firms into large or small based on their revenue, R&D expenditure, and employee counts. Finally, we categorize assignees into large or small based on the number of publications and inventors.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix D. Distances

D1. Temporal distance

We test whether our results on temporal distance are robust to varying the threshold for temporal distance. Our main measure of temporal distance was business hour overlap, obtained using the *timezonefinder* package in Python. We first obtained each airport's time zones, then for each pair of airports we calculate the time zone difference (in hours) between the two. The average time zone difference across all routes is -0.011 hours (st. dev. 2.25), with AKL-LAX and HNL-MEL having time zones farthest apart. These airport pairs, however large their time zone difference is, will still have some working hour overlap, (3 hours, since 9AM in AKL is 2PM in LAX). Thus, we convert the time zone difference to "business hour overlap" (0-8 hours), as well as "time difference" (0-12 hours). For "time difference," we take the absolute value of time zone difference, and subtract it from 24 if time zone difference is 12+ hours. For "business hour overlap," we subtract "time difference" from 8, but set all negative values to 0. Some routes with large "time difference" are between LA and Moscow, which have zero business hour overlap. Routes such as Munich and Singapore have 1 business hour overlap, but are 7 hours apart. In the paper, we use the median business hour overlap to categorize airport pairs into high or low temporal distance routes.

In this section, we relax this assumption and use alternate cutoffs to bin routes into high or low temporal routes. Below, in Tables D1 through D5, we present regression discontinuity results from our two subsamples, low temporal distance and high temporal distance, based on the number of hours of business overlap. For each alternative threshold, we see that the effects are driven by airport pairs with high temporal distance (low business hour overlap).

Table D1. High temporal distance corresponds to >0 hours in business hour overlap.

	Citations		Collaborations	
	(1)	(2)	(3)	(4)
Business Hour Overlap:	Low	High	Low	High
asinh(Flights)	0.277***	0.828	0.242***	0.279
	(0.087)	(0.980)	(0.082)	(0.211)
dist6000	0.003***	-0.009	-0.000	-0.002
	(0.001)	(0.010)	(0.000)	(0.001)
dist6000 # under6000	-0.007***	0.010	-0.001	0.002
	(0.002)	(0.010)	(0.001)	(0.001)
Observations	828	1450	1101	2002

Note: Standard errors in parentheses, clustered at the Country pair-year level. All specifications include country-pair-year fixed effects. Dependent variables are inverse hyperbolic sine transformed. Optimal bandwidth calculation follows the methodology described in Cattaneo et al. (2018). Working hour overlap calculated from airport time zone data. Low working hour overlap indicates 0 working hour overlap while high working hour overlap indicates any business hour overlap. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D2. High temporal distance corresponds to >1 hours in business hour overlap.

	Citations		Collaborations	
	(1)	(2)	(3)	(4)
Business Hour Overlap:	Low	High	Low	High
asinh(Flights)	0.368***	-0.127	0.201***	0.133

	(0.077)	(0.126)	(0.057)	(0.214)
dist6000	0.001	0.003*	0.000	-0.001
	(0.001)	(0.002)	(0.000)	(0.002)
dist6000 # under6000	-0.001	-0.003*	-0.000	0.001
	(0.002)	(0.002)	(0.001)	(0.002)
Observations	1342	990	1859	1287

Note: Standard errors in parentheses, clustered at the Country pair-year level. All specifications include country-pair-year fixed effects. Dependent variables are inverse hyperbolic sine transformed. Optimal bandwidth calculation follows the methodology described in Cattaneo et al. (2018). Working hour overlap calculated from airport time zone data. Low working hour overlap indicates 0-1 hours of overlap while High working hour overlap indicates greater than 1 hour of business hour overlap. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table D3. High temporal distance corresponds to >2 hours in business hour overlap.

	Citations		Collaborations	
	(1)	(2)	(3)	(4)
Business Hour Overlap:	Low	High	Low	High
asinh(Flights)	0.392***	-0.096	0.220***	0.080
	(0.083)	(0.075)	(0.066)	(0.096)
dist6000	0.001	0.003**	0.000	-0.001
	(0.001)	(0.001)	(0.000)	(0.001)
dist6000 # under6000	-0.001	-0.003**	-0.000	0.001
	(0.002)	(0.001)	(0.001)	(0.001)
Observations	1595	737	2222	923

Note: Standard errors in parentheses, clustered at the Country pair-year level. All specifications include country-pair-year fixed effects. Dependent variables are inverse hyperbolic sine transformed. Optimal bandwidth calculation follows the methodology described in Cattaneo et al. (2018). Working hour overlap calculated from airport time zone data. Low working hour overlap indicates 0-2 hours of overlap while High working hour overlap indicates greater than 2 hours of business hour overlap. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table D4. High temporal distance corresponds to >3 hours in business hour overlap.

	Citations		Collaborations	
	(1)	(2)	(3)	(4)
Business Hour Overlap:	Low	High	Low	High
asinh(Flights)	0.299***	0.481	0.159***	-0.111
	(0.060)	(1.185)	(0.045)	(0.121)
dist6000	0.001	-0.005	0.000	0.002
	(0.001)	(0.017)	(0.000)	(0.002)
dist6000 # under6000	-0.001	0.006	-0.000	-0.002
	(0.001)	(0.020)	(0.001)	(0.002)
Observations	1712	620	2367	768

Note: Standard errors in parentheses, clustered at the Country pair-year level. All specifications include country-pair-year fixed effects. Dependent variables are inverse hyperbolic sine transformed. Optimal bandwidth calculation follows the methodology described in Cattaneo et al. (2018). Working hour overlap calculated from airport time zone data. Low working hour overlap indicates 0-3 hours of overlap while High working hour overlap indicates greater than 3 hours of business hour overlap. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table D5. High temporal distance corresponds to >4 hours in business hour overlap.

	Citations		Collaborations	
	(1)	(2)	(3)	(4)
Business Hour Overlap:	Low	High	Low	High
asinh(Flights)	0.292***	0.276	0.150***	-0.130
	(0.060)	(0.457)	(0.044)	(0.138)
dist6000	0.001	-0.004	-0.000	0.002
	(0.001)	(0.006)	(0.000)	(0.002)
dist6000 # under6000	-0.001	0.004	-0.000	-0.002
	(0.001)	(0.007)	(0.001)	(0.002)
Observations	1848	484	2576	561

Note: Standard errors in parentheses, clustered at the Country pair-year level. All specifications include country-pair-year fixed effects. Dependent variables are asinh-transformed. Optimal bandwidth calculation follows the methodology described in Cattaneo et al. (2018). Working hour overlap calculated from airport time zone data. Low working hour overlap indicates 0-4 hours of overlap while High working hour overlap indicates greater than 4 hours of business hour overlap. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D2. North-South analysis

Table D6. Flight effects on knowledge diffusion are greater for shorter North-South distances.

	Citations		Collaborations	
	(1)	(2)	(3)	(4)
	Above median North-South distance	Below median North-South distance	Above median North-South distance	Below median North-South distance
Nonstop Flights (asinh)	-0.078	0.599***	0.089	0.208***
	(0.079)	(0.138)	(0.094)	(0.063)
dist6000	0.001**	-0.001	-0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.000)
dist6000 # under6000	-0.001*	0.001	0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	1430	1815	1430	1815
R^2	0.339	0.028	0.099	0.413

Note: Standard errors in parentheses, clustered at the country-pair-year level. All specifications include country-pair-year fixed effects. Bandwidth at 500 miles. Both dependent variables are asinh-transformed. In this table, we study the differences between North-South routes (e.g., London - Johannesburg) and East-West routes (e.g., Los Angeles - Singapore) that cross many more time zones. To do this analysis, we calculate the difference in longitudes between airport pairs and run subsample analyses on above and below median longitudinal distance pairs. Some examples of routes that are slightly above the median North-South distance is London Gatwick-Hong Kong, or Nagoya-Charles DeGaulle. Routes that are slightly below the median are London Heathrow-Capetown, or Saigon-Frankfurt.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

D3. Nonlinear effects

Table D7. Nonlinear effects of temporal distance, by percentile.

	Citations				Collaborations			
	(1) Bottom 25%	(2) 25-50	(3) 50-75	(4) Top 25%	(5) Bottom 25%	(6) 25-50	(7) 50-75	(8) Top 25%
Nonstop Flights (asinh)	0.333	0.018	0.683***	0.420***	-0.176	-0.000	0.217**	0.215***
	(0.503)	(0.026)	(0.239)	(0.127)	(0.248)	(0.006)	(0.095)	(0.064)
dist6000	-0.003	-0.000	-0.004*	0.000	0.002	-0.000**	-0.001	-0.000
	(0.006)	(0.000)	(0.002)	(0.001)	(0.003)	(0.000)	(0.001)	(0.000)
dist6000 # under6000	0.004	0.001**	0.006**	-0.001	-0.003	0.000	0.003**	-0.000
	(0.008)	(0.000)	(0.003)	(0.001)	(0.004)	(0.000)	(0.001)	(0.001)
Observations	892	585	847	1398	892	585	847	1398

Note: Standard errors in parentheses, clustered at the country-pair-year level. All specifications include country-pair-year fixed effects. Bandwidth at 550 miles. Both dependent variables are asinh-transformed. Time zone distance measures the difference in time zones between airport pairs with a maximum of 12 hours and a minimum of 0 hours. In this table, we split airport pairs into quartiles based on the temporal distance. Bottom 25% corresponds to flights less than 6 hours, 25-50% lists flights greater than 6 hours but less than 7 hours long, 50-75% are flights less than 9 hours but greater than or equal to 7 hours, and the top 25% are flights 9 hours or longer.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D8. Nonlinear effects of temporal distance, by number of hours in time zone difference.

	Citations				Collaborations			
Time Zone Distance:	(1) <4 Hours	(2) 4-6 Hours	(3) 6-8 Hours	(4) 8+ Hours	(5) <4 Hours	(6) 4-6 Hours	(7) 6-8 Hours	(8) 8+ Hours
Nonstop flights (asinh)	0.280	0.051	1.196*	0.420***	-0.177	0.000	0.378	0.215***
	(0.375)	(0.032)	(0.644)	(0.127)	(0.217)	(0.005)	(0.230)	(0.064)
(6,000-Distance)	-0.003	0.000	-0.005	0.000	0.002	-0.000	-0.001	-0.000
	(0.005)	(0.000)	(0.003)	(0.001)	(0.003)	(0.000)	(0.001)	(0.000)
(6,000-Distance) x Under6,000	0.005	0.001	0.006*	-0.001	-0.003	-0.000	0.003*	-0.000
	(0.007)	(0.001)	(0.003)	(0.001)	(0.004)	(0.000)	(0.001)	(0.001)
Observations	637	398	1309	1398	637	398	1309	1398

Note: Standard errors in parentheses, clustered at the country-pair-year level. All specifications include country-pair-year fixed effects. Bandwidth at 550 miles. Both dependent variables are asinh-transformed. Time zone distance measures the difference in time zones between airport pairs with a maximum of 12 hours and a minimum of 0 hours. In this table, we split airport pairs roughly into quartiles based on the temporal distance. <4 Hours corresponds to the bottom 15% of the sample, 4-6 hours correspond to about 12% of the sample, 6-8 hours correspond to about 41% of the sample, and 8 hours or more corresponds to the top 31% of the sample.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D9. Nonlinear effects of cultural distance.

	Citations				Collaborations			
	(1) Bottom 25%	(2) 25-50	(3) 50-75	(4) Top 25%	(5) Bottom 25%	(6) 25-50	(7) 50-75	(8) Top 25%
Nonstop	0.246	-0.039	-0.273	1.169**	0.004	-0.126	-0.169	0.416*

Flights (asinh)								
	(0.292)	(0.157)	(0.211)	(0.553)	(0.039)	(0.130)	(0.111)	(0.211)
dist6000	0.000	0.001	0.002	-0.004*	-0.000**	0.001	0.001**	-0.000
	(0.001)	(0.001)	(0.001)	(0.002)	(0.000)	(0.001)	(0.001)	(0.001)
dist6000 # under6000	-0.001	-0.002	-0.002	0.011**	0.000	-0.001	-0.002**	0.002
	(0.002)	(0.001)	(0.002)	(0.005)	(0.000)	(0.001)	(0.001)	(0.002)
Observations	494	511	467	526	494	511	467	526

Note: Standard errors in parentheses, clustered at the country-pair-year level. All specifications include country-pair-year fixed effects. Bandwidth at 550 miles. Both dependent variables are asinh-transformed. Cultural distance measures are derived from Berry et. al., (2020).

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

D4. Immigrant Employees

Table D10. Flight effects on knowledge diffusion are greater for firms with more immigrant employees.

	Citations		Collaborations	
	(1)	(2)	(3)	(4)
	High	Low	High	Low
Nonstop Flights (asinh)	0.208***	0.073***	0.111***	-0.000
	(0.059)	(0.027)	(0.034)	(0.014)
dist6000	-0.001***	-0.001***	-0.000***	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
dist6000 # under6000	0.002***	0.001***	0.001**	0.000
	(0.001)	(0.000)	(0.000)	(0.000)
Observations	3795	3795	3795	3795

Note: Standard errors in parentheses, clustered at the country-pair-year level. All specifications include country-pair-year fixed effects. Both dependent variables are asinh-transformed. This table matches each firm to the number of labor condition applications (LCAs) they have submitted between the years 2008-2021. The number of LCAs approximates a firm's dependence on immigrants as a central part of their workforce. In the context of innovative firms, this likely means a higher number of immigrant inventors and business people. Depending on the number of LCAs, we categorize them into "High" LCA dependent firms and "Low" LCA dependent firms. Roughly one third of the sample are matched to the LCA dataset.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix E. Subgroup Analysis

In our broader research paper, there are two types of heterogeneity we try to uncover. The first is heterogeneity across different subsamples. For instance, testing whether the effect of nonstop flights is greater for airport pairs that are temporally distant. This would involve estimating regression discontinuities for two different subsamples: airport pairs with high temporal distance and another for low temporal distance and properly comparing the point estimates. A second type of heterogeneity we consider is heterogeneity across different outcome variables. For instance, to test whether the effect of flights is greater for firms, we estimate two regression discontinuities, both using the entire sample of airport pairs, but with different outcome variables: one outcome for firms and another for academics. We detail how to test for these two types of heterogeneity below.

Subgroup analysis of treatment effects in the RDD setting can be performed by modelling the effects of flights for each subsample. Specifically, we follow Wasserman (2021) and estimate the following regression specification using 2SLS:

$$\begin{aligned} Y_{a_o,a_d,t} = \alpha_0 &+ \alpha_1\, nonstop\widehat{Flights}_{a_o,a_d,t} + \alpha_2\left(D_{a_o,a_d} \times non\widehat{stop}Flights_{a_o,a_d,t}\right) + \alpha_3 D_{a_o,a_d} \\ &+ \alpha_4 dist6000_{a_o,a_d} + \alpha_5 Under6000_{a_o,a_d} \times dist6000_{a_o,a_d} \\ &+ \alpha_6 D_{a_o,a_d} \times dist6000_{a_o,a_d} + \alpha_7 D_{a_o,a_d} \times Under6000_{a_o,a_d} \times dist6000_{a_o,a_d} \\ &+ \varepsilon_{a_o,a_d,t} \end{aligned}$$

Here, $D_{a_o,a_d} = 1$ if the airport pair a_o, a_d are in the subsample of interest (e.g., airport pairs with above median temporal distance). All other variables are identical to our baseline specification. Our main variable of interest is then α_2, which denotes the differential effect of nonstop flights for group $D_{a_o,a_d} = 1$. We use 2SLS, using $Under6000$ to instrument for $nonstopFlights$, and $D \times Under6000$ to instrument for $D \times nonstopFlights$. As Wasserman (2021) and Hsu and Shen (2019) note, these estimates may lead to over-rejection. Thus, we use bootstrap resampling to estimate the variability of the estimates. We iteratively sample subsamples of our data with replacement, and run separate fuzzy RDD regressions for each subgroup to get instrumental variable estimates and standard errors of the group-specific effects.

The second test for heterogeneous effects involves comparing RD estimates for two different outcome variables. Here, we follow Mize et. al., (2019) in using Seemingly Unrelated Estimates (SUEST), as well as block bootstrapped estimates of the difference.

To test whether the effect of flights is different for academics vs firms, we follow Mize et. al., (2019) in using seemingly unrelated estimation (SUEST) to compare effect sizes. We compare the effects of the treatment on two separate outcomes by first, fitting a first-stage model to obtain "predicted" exposures for each unit. Then, we estimate separate second stage models for the two outcomes, while computing a cross-model covariance using the SUEST method. Testing cross-model difference can then be done using a Wald-like test or, alternatively, using a bootstrap method to simulate the distribution of the difference.

In practice, to test whether the effect of nonstop flights is different for firms and academics, we "stack" the data and fit the models simultaneously as in Mize et. al., (2019). Stacking allows us to estimate the covariance between the two estimates and adjust our test statistic. As the name stacking suggests, if the length of our original dataset is N, the stacked dataset is exactly 2N rows long. For columns, we create a stacked outcome variable, Y, of which the first N observations are the citation counts to academic patents, and the last N observations are the citation counts to firm patents. In addition to Y, we have a group of columns for observations regarding academic patents, and another group of columns for firm patents. For the academic columns, the first N rows replicate the original dataset variables asinh(Nonstop Flights), Dist6000, Under6000, and Dist6000*Under6000, while the last N rows are zeroes. For firm columns, the last N rows replicate the original dataset variables asinh(Nonstop Flights), Dist6000, Under6000, and Dist6000*Under6000, while the first N rows are zeroes. Thus, a roughly diagonal dataset is created as below:

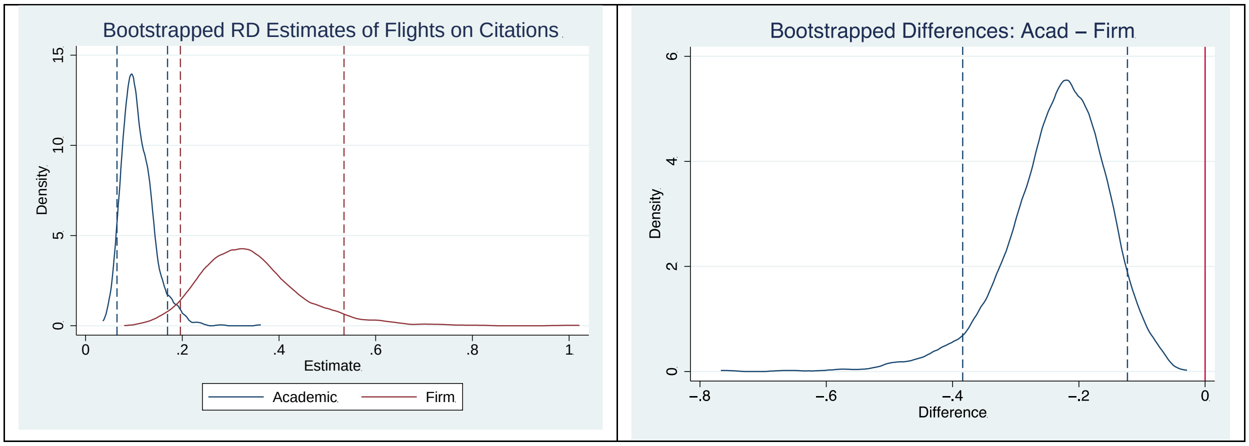
$$\begin{bmatrix} Y & Dist6000_a & Under6000_a & Dist6000_f & Under6000_f & academic \\ CiteAcad_1 & 948.3 & 1 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ CiteAcad_N & 10139.4 & 0 & 0 & 0 & 1 \\ CiteFirm_1 & 0 & 0 & 948.3 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ CiteFirm_N & 0 & 0 & 10139.4 & 0 & 0 \end{bmatrix}$$

Using the data structure as above, we estimate the following equation:

$$\begin{aligned} Y = \alpha_0\, \widehat{nonstopFlights}_a + \alpha_1 Dist6000_a + \alpha_2 Dist6000_a \times Under6000_a \\ + \beta_0 \widehat{nonstopFlights}_f + \beta_1 Dist6000_f + \beta_2 Dist6000_f \times Under6000_f \\ + \gamma Academic + X_a^{FE} + X_f^{FE} \end{aligned}$$

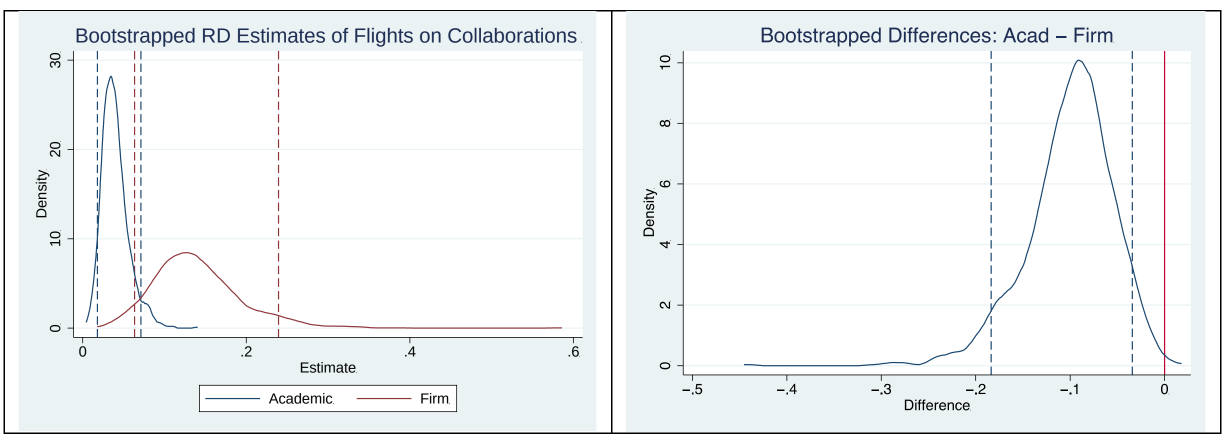
Here, the coefficients α_0 and β_0 will recover the estimates from separately running RDD for academic and firm citations. Note, we include country-pair year fixed effects for academic as well as firms. Finally, SUEST involves running a Wald test for $\alpha_0 = \beta_0$.We present the results below.

The SUEST results show that nonstop flights are more beneficial for firms than for academics: for academic assignees' patents, a 10% increase in nonstop flights leads to a 0.99% increase in citations, but for firm patents, citations increase by 3.38%. The difference of 2.39 percentage points has $z = 2.5813$ with a p-value of 0.010, thus we reject the null hypothesis that the effect of flights on academic and firm citations are equal. Below, we compare the SUEST results with our bootstrapping results.



Above, on the left-hand side, we show the distribution of 1,000 bootstrapped estimates of the two effects. The right-hand side shows the distribution of the estimates of the difference in effects. The mean of the difference is 0.235, with a 95% confidence interval of [-0.384, -0.123]. We see that the bootstrap differences are indeed similar to the SUEST results, but with tighter confidence intervals for the bootstrapped results.

We repeat the analysis above for collaborations and similarly find that the effect of nonstop flights is greater for firms than for academics. For academic assignees' patents, a 10% increase in nonstop flights leads to a 0.404% increase in collaborations, while for firms the effect is 1.48%. The mean of the difference is 0.107, with $z = 2.035$, and a p-value of 0.042. We thus reject the null hypothesis that the effect of flights on academic and firm collaborations are equal. Again, we compare our SUEST results with our bootstrap results.



The left-hand side panel shows the distribution of 1,000 bootstrapped estimates of the two effects, while the right hand side panel plots the distribution of their differences. The mean of the difference is 0.1003, with a 95% confidence interval of [-0.184, -0.034]. Again, we see that the mean difference is very similar, with narrower confidence intervals.

Appendix F. Airport level analysis and General Equilibrium concerns

This section provides additional details about estimating the impact of nonstop flights at the airport level, as well as ways to alleviate general equilibrium concerns.

F1. Instrumental variable approach

First, we implemented an instrumental variable-based identification strategy proposed by Campante and Yanagizawa-Drott (2015) to extend the results from the airport-*pair* level to the airport level. While the RD shows that at the airport pair level, pairs slightly below 6000 miles apart have increased knowledge flows, whether this carries over to the airport level is unclear. This effect may be a redirection of knowledge flows from other airport pairs to the focal airport pair. Analyzing how nonstop flights affect collaborations and citations at a single airport mitigates concerns of redirection as it would be the net effect.

Our approach is to use an instrument to create exogenous variation on the number of nonstop flights linked to an airport. A corollary to our identification strategy (ultra-long-haul flights are more expensive to operate) is that airports with many other "potential" airports slightly less than 6000 miles apart will be more "connected" in terms of number of flights. The identification assumption is that there is no reason for airports that happen to have relatively many airports sitting just under 6,000 miles away should be systematically different from airports that happen to have many just above that threshold. This statement is conditional on the total number of airports around 6,000 miles not explaining changes in innovation outcomes other than through the number of flights themselves. Thus, we weight the instrument with the information related to the potential of each connection; specifically, we proxy each airport's potential using its eigenvector centrality at the beginning of our sample (2005). We estimate the following equation for our first stage.

$$ConnectedAirports_{it} = \beta_0 + \beta_1 ShareBelow6K_i + X_i + \varepsilon_i$$

Here, $ConnectedAirports_i$ measures the number of airports with which airport i has a nonstop flight in year t. $ShareBelow6K_i$ counts the total number of airports (connected or unconnected) slightly below 6000 miles and divides this by the total number of airports (again, connected or unconnected) around 6000 miles. A positive β_1 is evidence that the share of airports slightly below 6000 miles can predict the number of airports to which nonstop flights exist. We present results of estimating this equation below.

Table F1. First stage results in the instrumental variable analysis to gauge the flight effects on knowledge diffusion at the airport level.

	(1) Connected airports, 2005	(2) Connected airports, 2010	(3) Connected airports, 2015	(4) Connected airports, 2015	(5) Network Centrality, 2005	(6) Network Centrality, 2010	(7) Network Centrality, 2015	(8) Network Centrality, 2015	(9) Total # of Connected Airports (2005-2015)
sharebelow6k (unweighted)	26.743**	29.610**	33.558**	37.026***					26.743**
	(12.680)	(14.525)	(16.393)	(11.528)					(12.680)
airportsnear6k (unweighted)	-0.007	-0.003	-0.001	0.003					-0.007
	(0.008)	(0.009)	(0.010)	(0.005)					(0.008)
dist_equator				-0.002				-0.000***	
				(0.002)				(0.000)	
gmt_timediff				-3.099***				-0.001***	
				(0.705)				(0.000)	
sharebelow6k					0.008***	0.008***	0.008***	0.011***	
					(0.002)	(0.002)	(0.002)	(0.003)	
airportsnear6k					0.000	0.000	0.000	0.000	
					(0.000)	(0.000)	(0.000)	(0.000)	
Constant	2.411	0.263	-1.489	14.700*	-0.000	-0.000	-0.000	0.008***	2.411
	(10.890)	(12.143)	(13.890)	(8.707)	(0.001)	(0.001)	(0.001)	(0.002)	(10.890)
Observations	4956	4956	4956	4956	4956	4956	4956	4956	4956

Note: Standard errors in parentheses, clustered at the country level. All specifications include region fixed effects. Sharebelow6k (unweighted) measures the fraction of airports within [5500,6000] miles over the number of airports within [5500,6500] miles. Sharebelow6k weights this measure by that airport's network centrality in 2005.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Our first stage results show that the unweighted instrument $ShareBelow6K$ is a good predictor of the number of connected airports in 2005, 2010, and 2015. We include geographic controls such as the distance to the equator and the time difference from GMT in Columns 4, 8 and 9. as well as the eigenvector centrality between 2005-2015. A one standard deviation increase in the unweighted share of airports below 6000 miles (0.084) increases the number of connected airports in 2015 by about 3.11. Similarly, a one standard deviation increase in the weighted share of airports below 6000 miles (0.144) increases network centrality by 0.002.

Next, we turn to estimating the impact of flights on publications and citations at the airport level. We estimate the following equation.

$$PatentPublications_i = \beta_0 + \beta_1 \ \widehat{ConnectedAirports_\iota} + X_i + \varepsilon_i$$

Where $\widehat{ConnectedAirports_\iota}$ is the predicted value of airports connected to i via nonstop flights from our first stage specification. The coefficient of interest β_1 thus measures the impact of an additional airport connection, or increased connectivity, on the number of publications at a given airport at a given year.

Table F2. Nonstop flights and connectivity increase the number of collaborations and citations at the airport level.

	Citations			Publications		
	(1) Total (2000-2015)	(2) 2004	(3) 2014	(4) Total (2000-2015)	(5) 2004	(6) 2014
Connected Airports, 2015	0.126***	0.099**	0.062***	0.105***	0.071**	0.057***
	(0.046)	(0.040)	(0.020)	(0.036)	(0.028)	(0.019)
Total airportsnear6k	0.160	0.113	0.060	0.114	0.064	0.057
	(0.109)	(0.093)	(0.046)	(0.085)	(0.063)	(0.042)
Distance from Equator	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Time zone difference from GMT	-0.282	-0.196	-0.061	-0.198	-0.107	-0.072
	(0.363)	(0.307)	(0.171)	(0.282)	(0.204)	(0.152)
Observations	4956	4956	4956	4956	4956	4956

Note: Standard errors in parentheses, clustered at the country level. All specifications include region fixed effects. Both dependent variables are asinh-transformed.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Column 1 shows that an additional connected airport in 2015 leads to a 12.6% increase in the total number of citations to patents near that airport. Columns 2 and 3 show that this effect is positive for 2004 and 2014, but decreasing, possibly because of insufficient time for citations to be realized. Similarly, Column 4 shows that an additional connected airport in 2015 increases the number of publications at that airport by about 10.5%, and that this effect is positive for 2004 and 2014.

Table F3. Instrumental variable analysis to gauge flight effects at the airport level, firms versus academic institutions.

	Firms		Academic Institutions	
	(1)	(2)	(3)	(4)
	Citations	Publications	Citations	Publications
conn2015	0.124***	0.105***	0.085***	0.061***
	(0.046)	(0.036)	(0.027)	(0.019)
airportsnear6k	0.163	0.115	0.047	0.043
	(0.109)	(0.084)	(0.065)	(0.048)
dist_equator	-0.000	-0.000	-0.000	-0.000
	(0.001)	(0.000)	(0.000)	(0.000)
gmt_timediff	-0.287	-0.197	-0.121	-0.091
	(0.362)	(0.281)	(0.217)	(0.153)
Observations	4956	4956	4956	4956

Note: Standard errors in parentheses, clustered at the country level. All specifications include region fixed effects. All dependent variables are asinh-transformed.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

An additional connected airport has positive effect on citations and publications. Furthermore, Columns 1-2 in the table above suggest that the effect is greater for patents by firms than for patents by academics. Note that we have instrumented for the number of connected airports in 2015 to more easily interpret the coefficient sizes. The same results hold when using eigenvector centrality instead of the number of connected airports.

Overall, results from section F1 suggest that flight connections between pairs are not changing the composition of citations / collaborations, but are increasing the pie overall.

F2. Forging new collaborations or strengthening existing collaborations?

Our main empirical strategy has no pre-post testing. Therefore, it is difficult to test whether there is a "change" in the composition of teams. Instead, we use an instrumental variable approach at the airport level to test whether additional flights affect the extensive margin (as measured by the number of collaborators) or the intensive margin (as measured by the duration of collaborations).

The main idea is that if flights have an effect on the extensive margin, we should see flights increase the *breadth* of collaborators, while the intensive margin would lead to increased *duration* of collaborations. First, we test whether more nonstop flights lead to more unique collaborators and/or longer collaborations between collaborators.

For each inventor in our sample, we find the unique number of collaborators. Then, for each inventor-collaborator pair, we find the collaboration duration (time difference in years between first and last collaboration). Then, for each airport in our sample, we take the average and maximum of 1) the number of collaborators, and 2) the average collaboration duration for each inventor. Thus, for each airport, we obtain 1) the average number of collaborators, 2) the average duration of collaborations, 3) the maximum number of collaborators, and 4) the maximum duration of collaborations across all inventors within a 50-mile radius of the airport.

We consider inverse hyperbolic sine transformed outcomes in the table below.

Table F4. Intensive versus extensive margins: mean and maximum number of collaborations and collaboration duration.

	(1) Mean # Collaborators	(2) Mean Collab. Duration	(3) Max # Collaborators	(4) Max Collab. Duration
conn2015	0.018^{***}	0.014^{**}	0.089^{***}	0.062^{***}
	(0.007)	(0.006)	(0.026)	(0.020)
airportsnear6k	0.032^{*}	0.013	0.104	0.064
	(0.018)	(0.016)	(0.069)	(0.050)
dist_equator	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
gmt_timediff	-0.053	-0.032	-0.135	-0.109
	(0.054)	(0.047)	(0.207)	(0.145)
Observations	4956	4956	4956	4956

Note: Standard errors in parentheses, clustered at the country level. All specifications include region fixed effects. Conn2015 measures the number of airports to which airport I is connected to in 2015. Airportsnear6k counts the total number of airports within 6000 miles of airport i. dist_equator measures the distance from airport I to the equator. Gmt_timediff measures the difference between airport I's time zone and GMT. Mean # collaborators measures the mean number of collaborators for all inventors within 50 miles of airport i. All dependent variables are asinh-transformed.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Columns 1 and 2 show the average number of collaborators and collaboration duration, while Columns 3 and 4 show the maximum number of collaborations and durations. We see that nonstop flights increase innovation in both the extensive and intensive margins, facilitating both meeting new collaborators as well as intensifying existing ones. Qualitatively, the coefficient magnitudes point to the extensive margin being larger, suggesting flights allow inventors to meet more new inventors, but the coefficients are not statistically significant from each other. Interestingly, the coefficient sizes are larger for Columns 3 and 4, suggesting that the effect of nonstop flights is greater for more productive inventors.

We also use the raw number of collaborations and duration without transforming in the table below.

Table F5. Intensive and extensive margins: mean and maximum number of collaborations / collaboration duration (raw value, not asinh-transformed).

	(1) Mean # Collaborators	(2) Mean Collab. Duration	(3) Max # Collaborators	(4) Max Collab. Duration
conn2015	0.027**	0.014	5.435**	0.487***
	(0.010)	(0.009)	(2.374)	(0.167)
airportsnear6k	0.041	0.011	11.444*	0.378
	(0.027)	(0.022)	(6.332)	(0.403)
dist_equator	-0.000	-0.000	-0.020	-0.001
	(0.000)	(0.000)	(0.028)	(0.002)
gmt_timediff	-0.072	-0.048	-2.481	-0.829
	(0.081)	(0.066)	(16.300)	(1.254)
Observations	4956	4956	4956	4956

Note: Standard errors in parentheses, clustered at the country level. All specification include region fixed effects. Conn2015 measures the number of airports to which airport I is connected to in 2015. Airportsnear6k counts the total number of airports within 6000 miles of airport i. dist_equator measures the distance from airport I to the equator. Gmt_timediff measures the difference between airport I's time zone and GMT. Mean # collaborators measures the mean number of collaborators for all inventors within 50 miles of airport i.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix G. Pecuniary Cost and Travel Time

Table G1. Duration matters more for airports further apart.

	Under 6000		Over 6000	
	(1)	(2)	(3)	(4)
	Collaborations	Citations	Collaborations	Citations
Duration (asinh)	-0.580**	-0.498*	-1.048**	-1.238***
	(0.278)	(0.291)	(0.404)	(0.390)
Price (asinh)	-0.291	-0.436*	0.142	0.044
	(0.245)	(0.242)	(0.240)	(0.244)
Distance (asinh)	-0.792	-1.835	3.128***	4.793***
	(1.046)	(1.715)	(0.742)	(0.819)
Constant	11.864	22.099	-24.567***	-37.367***
	(10.302)	(16.049)	(6.920)	(7.566)
Observations	651	651	572	572
R^2	0.592	0.700	0.637	0.831

Note: This table tests the two potential mechanisms (flight duration and flight price) for location pairs with longer versus shorter flight paths. Standard errors are in parentheses, clustered at the country-country level. Both dependent variables are asinh-transformed. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table G2. Duration matters more for international flights.

	Domestic Flight		International Flight	
	(1)	(2)	(3)	(4)
	Collaborations	Citations	Collaborations	Citations
Duration (asinh)	-0.052	0.067	-0.899***	-0.870***
	(0.385)	(0.188)	(0.249)	(0.244)
Price (asinh)	-0.704	-0.747	-0.159	-0.243**
	(1.261)	(1.003)	(0.116)	(0.113)
Distance (asinh)	-1.264	-2.646*	1.476***	2.743***
	(0.662)	(1.171)	(0.508)	(0.591)
Constant	18.272	31.045	-7.745	-17.737***
	(13.557)	(16.646)	(4.824)	(5.685)
Observations	219	219	1028	1028
R^2	0.283	0.424	0.684	0.854

Note: This table tests two potential mechanisms (flight duration and flight price) for location pairs with domestic versus international flights. Standard errors are in parentheses, clustered at the country pair level. All specifications include country pair fixed effects. Both dependent variables are asinh-transformed. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Appendix References

Abadie A, Athey S, Imbens GW, Wooldridge J (2017) *When should you adjust standard errors for clustering?* (National Bureau of Economic Research).

Angrist JD, Pischke JS (2008) *Mostly harmless econometrics: An empiricist's companion* (Princeton university press).

Calonico S, Cattaneo MD, Farrell MH (2020) Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal* 23(2):192–210.

Furman JL, Hayes R (2004) Catching up or standing still?: National innovative productivity among 'follower'countries, 1978–1999. *Research policy* 33(9):1329–1354.

Imbens G, Kalyanaraman K (2012) Optimal Bandwidth Choice for the Regression Discontinuity Estimator. *Rev Econ Stud* 79(3):933–959.

Imbens GW, Lemieux T (2008) Regression discontinuity designs: A guide to practice. *Journal of econometrics* 142(2):615–635.

MacKinnon JG, Magee L (1990) Transforming the dependent variable in regression models. *International Economic Review*:315–339.

Niwattanakul S, Singthongchai J, Naenudorn E, Wanapu S (2013) Using of Jaccard coefficient for keywords similarity. *Proceedings of the international multiconference of engineers and computer scientists*. 380–384.