LONG-RANGE GROWTH: ECONOMIC DEVELOPMENT IN THE GLOBAL NETWORK OF AIR LINKS*

FILIPE CAMPANTE AND DAVID YANAGIZAWA-DROTT

We study the impact of international long-distance flights on the global spatial allocation of economic activity. To identify causal effects, we exploit variation due to regulatory and technological constraints, which gives rise to a discontinuity in connectedness between cities at a distance of 6,000 miles. We show that improving an airport's position in the network of air links has a positive effect on local economic activity, as captured by satellite-measured night lights. We find that air links increase business links, showing that the movement of people fosters the movement of capital. In particular, this is driven mostly by capital flowing from high-income to middle-income (but not low-income) countries. Taken together, the results suggest that increasing interconnectedness induces links between businesses and generates economic activity at the local level but also gives rise to increased spatial inequality locally, and potentially globally. *JEL Codes*: F15, F21, F23, F63, O11, O18, O19, O47, R11, R12, R40.

I. INTRODUCTION

Our age of globalization is unique in that it is now far cheaper and faster than ever to transport people, which has made it possible to travel back and forth between distant places as never before. This is the direct consequence of the explosion in air travel. Of course, it was possible to travel long distances before air travel, but the cost was so high that few actually did, and those who did, for the most part, would not travel frequently. Now, for the first

time in human history, the whole world is effectively connected in a global network that enables a constant flow of people between countries and continents far apart.

This article studies the impact of direct long-distance air links, to present the first evidence of causal effects of that transformation on economic development at the local level. A long intellectual tradition has posited that proximity—in particular its most fundamental aspect, face-to-face contact—is a key driver of the transmission of knowledge and information (Storper and Venables 2004; Glaeser 2011), which in turn underpins the increases in productivity without which sustained economic growth is impossible. People have been able to move between Shanghai and London or New York for a long time, and goods have been moving for just as long, and now people can go back and forth between these places. This opens up new possibilities of exchange and interaction, with potentially transformative effects for development.

How important might these possibilities be? Consider Shanghai and Jakarta, cities that as of the early 1990s would seem at a similar level of economic development. In the two decades between 1990 and 2010, Jakarta added 13 new direct (at least weekly) long-distance flight connections, and Shanghai added 34. Over that period, Jakarta grew substantially, and Shanghai grew substantially more.

Of course this does not tell us whether or how much of Shanghai's extra growth was caused by those additional connections. There are myriad differences between those two cities and what happened to them over this period, which might entail differences in economic performance. It could just as well be performance driving connections: lots of people want to go to and interact with prosperous and/or fast-growing places. Perhaps "investment goes to cities that are attractive in their own right, rather than because they are easy to get to" (*The Economist* 2015).

We tackle this empirical challenge by establishing and exploiting a key feature of the network of air links: cities that are just under 6,000 miles apart are distinctly more likely to have direct air links, compared with cities slightly above that threshold. This is the result of an interaction between flight regulations and the evolution of airplane technology. Regulatory requirements on things like maximum flight time and crew accommodations increased costs substantially for flights of more than 12 hours, which corresponds to a distance of 6,000 miles—a little over what separates Milan from Shanghai, or Istanbul from Jakarta.

Although the regulations have been in place for decades, the introduction of landmark long-range airplane models (Boeing's 747-400 and 777, in 1989 and 1995, respectively, and Airbus's A330 and A340, in 1993–1994) made that discontinuity increasingly meaningful.

This discontinuity grants us a strategy to identify the causal effect of air links, using a sample of 819 cities with major airports. First, we compare pairs of cities that are just below 6,000 miles apart (such as Shanghai and Milan, 5,650 miles) to pairs that are just above (Shanghai and Madrid, 6,350), under the assumption that those pairs across the 6,000-mile threshold are not systematically different, using a standard regression discontinuity (RD) design.

Because we are also interested in outcomes at the local level, as opposed to the city-pair level, we show that the discontinuity translates into plausibly exogenous variation across different airports. Specifically, we can compare places near airports that happen to have a large share of potential destinations just below the threshold, such as Shanghai, to those near airports with relatively many just above it, such as Jakarta.[1] We show that this variation is not correlated with outcomes as of about 1989, when the discontinuity was weaker, underscoring that developments since then are what the empirical strategy is designed to capture.

We first show a set of basic, "first-stage" results: pairs of places like Shanghai and Milan (connected by a nonstop flight since 2003) are indeed more likely to be connected than pairs like Shanghai and Madrid (no nonstop flights before 2016).[2] Similarly, having a greater share of potential links just below the threshold—and particularly taking into account the quality of those links—indeed predicts a larger number and improved quality of connections. We show that this is because connections induce further connections:

1. Specifically, over three out of four major international airports in our sample which are between 5,500 and 6,500 miles from Shanghai happen to be below 6,000 miles. This places Shanghai in the top decile of the distribution. In contrast, for the same range relative to Jakarta, about two in three are above 6,000 miles, placing it in the bottom decile. Note that this controls for the total number of destinations around the threshold, which captures broader patterns of geographical location and isolation. We also incorporate information on the quality of each of those potential links, which slightly improves Jakarta's position but still keeps it in the bottom sixth of the distribution.

2. As we will discuss, there is reason to believe that the 6,000-mile discontinuity is in the process of disappearing, after regulatory changes implemented starting in 2014.

there is a spillover from the initial shock to shorter distances, as additional long-haul links increase a city's desirability for other connections, as well as an increase in the total flow of passengers.

We show that these connections matter for economic development. First, using granular data at the level of grid cells, we find that places close to airports with a larger share of potential quality-weighted connections just below the 6,000-mile threshold grew faster, as captured by satellite-measured night lights, between 1992 (when the data first become available) and 2010. This holds with different definitions of closeness, as well as excluding the specific location of the airports, and is not driven by specific countries. The effect is also economically significant: improving an airport's position in the network by one standard deviation, which would take the median airport in the sample about 300 spots up in the network centrality rankings, is associated with a one standard deviation increase in the growth in night lights over the period. Using Henderson, Storeygard, and Weil's (2012) estimate for the elasticity of GDP growth with respect to night lights growth, this boils down to about 0.8% in annual GDP growth.

We exploit the spatial richness of the data to show that the effect cannot be fully explained by spatial reallocation of economic activity from the hinterland to the airport's vicinity: while the positive effect dissipates with distance from the airport, as expected, and we cannot rule out the possibility of a negative net effect at longer distances, the magnitude of the latter would not be enough to quantitatively match the increased activity near the airport. This also means that connections induce spatial inequality, as the places that get connected grow faster.

We study how long-distance air links shape economic outcomes and development, focusing on the role of businesses. There is widespread circumstantial evidence that businesses care about ease of connection and the availability of flight links.[3] Yet some would argue that direct links do not matter so much for businesses, perhaps because "air travellers do not mind having to

3. For instance, the effort exerted by airports, airlines, and countries in getting direct flights, often justified as a way of attracting business investment; the fact that nonstop flights command higher prices, indicating that business passengers, the least price-conscious kind of traveler, do value them; the fact that businesses tend to locate disproportionately near airports (Bel and Fageda 2008; Kasarda and Lindsay 2011; Stilwell and Hansman 2013). Last but not least, there is growing empirical evidence of the business value of direct flight links (Giroud 2013; Bernstein, Giroud, and Townsend 2016).

connect flights in a foreign hub as much as they did in the past, because it is now easier to work on the go" (*The Economist* 2015).

To answer this question, we turn to data on business links. We start with firm-level information on foreign direct investment (FDI)—more specifically, on majority ownership of companies across different countries, where one would expect the possibility of face-to-face contact to be particularly important. Using the Orbis database, we geolocate over half a million foreign-owned companies all over the world, as well as their ultimate owners. For instance, the data allow us to find more than three times as many ownership links between Shanghai and Milan as between Shanghai and Madrid.

We show that this illustrates a general pattern, indicative of a causal impact of the availability of direct flights in facilitating the emergence of connections between firms in different locations. First, we again find a discontinuity right at the 6,000-mile threshold—pairs of cities just below 6,000 miles apart have substantially more business ownership links. From this we estimate that a given increase in connections generates about a similar proportional increase in ownership links. We find that this cannot be explained solely by the relocation of businesses from the areas surrounding competing major airports.

In addition, the evidence suggests that most of this increase constitutes capital flowing from relatively richer to relatively poorer countries: two thirds of the increase in business connections could be attributed to companies in high-income countries owning companies in middle-income ones, and one third in the opposite direction. This suggests that a lack of connections can be part of the explanation for the Lucas (1990) paradox of why capital does not flow from rich to poor countries.

In sum, the evidence shows that improving an airport's position in the global network of air links has a significant impact on economic activity at the local level. This seems to come along with an intensification of business links, consistent with the idea that the ability to interact in person is crucial for the establishment of those links. In other words, the movement of people fosters the movement of capital, even though there is no technological reason capital would need airplanes to move around.

This suggests that policy interventions that increase the number of connections could spur development at the local level. That said, our empirical strategy is not suited for an aggregate assessment, and we show that there are important caveats as to

whether such interventions would be desirable from that aggregate perspective: any policy maker whose concerns extend beyond the immediate vicinity of the airport would have to take into account that spatial inequality increases, and that we cannot rule out meaningful negative spillover effects at the local level and across major airports.

The results also highlight the potential for affecting inequality on a global scale. The first-stage relationship linking potential long-haul connections just below 6,000 miles and additional actual connections turns out not to hold for places too poor to begin with: Vientiane (Laos) gets a good draw in terms of potential, but this does not translate into actual connections when a place is too poor to be worth connecting to. As a result, low-income countries, unlike middle-income ones, get shut out of the increase in business links and capital flows. In sum, globalization, in its long-range air dimension, has seemingly helped the Shanghais of the world achieve convergence and increased the distance between them and the Vientianes. Whether the overall effect increases or decreases inequality depends on which of these two forces is seen as more important from that perspective.

This article relates to the broad empirical literature on the effects of globalization on economic outcomes (e.g., Frankel and Romer 1999; Dollar and Kraay 2004; Dreher 2006; Bacchetta and Jansen 2011; Hummels 2007; Ortega and Peri 2014). In particular, some of the work in that vein has looked at the effect of transportation technologies, such as steamships (Pascali 2017), railroads (Donaldson and Hornbeck 2016; Donaldson forthcoming), and airplanes (Feyrer 2009). This literature has focused largely on the effects of trade and openness, and as such it is mostly at the cross-country level or within one country. In contrast, we focus on a different aspect of globalization, namely, the movement of people through the network of air links, which also allows us to look at economic outcomes at a global yet granular level.[4] By doing so, we shed light on the substantial debate on globalization and inequality (e.g., Dollar 2005; Bourguignon 2015), which has focused on the contrast between inequality decreasing between countries

---

4. It is possible that increased connections have an impact on trade—both directly, as a substantial part of merchandise trade is transported by air, and indirectly, as the business links we detect could easily induce more trade as well. the data do not allow us to separately identify the impact of this trade channel, as we do not have city or city-pair information on trade.

while increasing within. We show that air links can contribute to that pattern while also helping explain why some places end up left behind (Collier 2007).

The idea that air links may have an impact on local development is quite natural, and it is unsurprising that a body of literature has looked into the connection (e.g., Brueckner 2003; Green 2007; Mukkala and Tervo 2013). However, the attempts to identify a causal impact have been limited, given the empirical challenges involved. An exception is Redding, Sturm, and Wolf (2011), who look at the impact of hub airports on the location of industries, using the natural experiment from the postwar division of Germany. Relatedly, others have looked at the impact of air travel and proximity on collaboration and productivity in various domains, such as business (Giroud 2013; Bernstein, Giroud, and Townsend 2016) or science (Catalini, Fons-Rosen, and Gaule 2016), using the introduction of air links in the United States as a source of variation, and also on trade (Cristea 2011; Poole 2013; Yilmazkuday and Yilmazkuday 2014). We differ in that this approach allows for causal identification at a global level, and for studying the impact on economic activity and the potential channels via business links.

The article is organized as follows. Section II provides background on the recent evolution of long-haul air travel to lay out the foundations of our identification strategy. Section III describes the data and develops a model showing how the regulatory and technological characteristics described in Section II translate into the specifications that implement that strategy. Section IV empirically establishes the presence and effect of our discontinuity on air links. Section V contains the results establishing the key effects on economic activity. Section VI focuses on the impact on business links. Section VII concludes.

II. Background on Long-Haul Air Travel

Ever since the advent of the so-called Jet Age, turbine-powered aircraft have made air travel increasingly common and far reaching (Proctor, Machat, and Kodera 2010). The technological evolution of commercial airplanes (Anderson 2002, chap.7) enabled greater and greater distances to be covered: from the Boeing 707, which started flying transatlantic routes in 1958, to the Boeing 747 (a.k.a. "Jumbo Jet"), which enabled, for instance, the route between San Francisco and Sydney which, at just

under 7,500 miles, in 1976 became the longest regularly scheduled nonstop flight in the world.

The introduction of the Boeing 747 in 1970 brought about the era of ultra long-haul (ULH) commercial aviation. There is no single definition of what constitutes ULH, but a common practical one singles out flights that take longer than 12 hours (McKenney et al. 2000). Given customary speeds, a 12-hour flight translates into about 6,000 miles, corresponding to the distance between London or Paris and Tokyo.[5] The distinction is apparent in the range of modern commercial aircraft by Airbus and Boeing: there is a set of aircraft models designed to fly up to 4,000 nautical miles (about 4,600 miles), and another designed to fly at least 6,000 miles.

The crucial import of the ULH distinction is not in the technical feasibility of flights by different kinds of aircraft—in fact, the shorter-haul planes cannot fly the 9–12-hour range anyway.[6] Instead, the 12-hour threshold is meaningful because of its impact on the cost of a given flight, as very long flights impose requirements on the availability of pilots and crew. For instance, the U.S. Federal Aviation Authority (FAA) had required since the 1950s that a two-pilot crew could fly at most 12 hours within a 24-hour period: flights above that limit require at least three pilots and an additional flight crew member (double augmentation), as well as "adequate sleeping quarters" on the plane (Code of Federal

5. Specifically, the air distance between Heathrow and Narita airports is 5,966 miles (as per www.airmilescalculator.com), and nonstop flight hours are estimated at 11:40 (as per a simple Google flight search). By contrast, from Chicago O'Hare to Narita takes 6,267 miles, and an estimated 13:15. More generally, although we do not have information on flight duration in the data, we can plot distance versus duration for the set of longest flights for each airline (as of 2016). As can be seen in the Online Appendix (Figure A1), the relationship is very tight, and there are essentially no flights above 6,000 miles under 12 hours. Note also that this benchmark can be applied into the past, as there has been very little evolution in speed over time–the 747-100B, from 1970, reached Mach 0.84 (644 mph) cruise speed, compared to the Mach 0.85 (652 mph) of the modern Boeing 787 from 2011.

6. More broadly, the range of a given plane is not exactly a clear-cut number: there is a trade-off between so-called payload (the sum of passenger and cargo weight) and fuel, and an airplane typically becomes uneconomical to fly way before it is technically infeasible to do so. The range numbers we use correspond to so-called MTOW (maximum takeoff weight) range at maximum payload (i.e., with the plane carrying as much passengers and cargo as possible). Planes can fly longer than that if they are willing to reduce payload to have more fuel. (For a discussion on this, see Clark 2002, chap.5).

Regulations, Section 121.485).[7] Similarly, European regulators adopted in 1991 a daily maximum of 13 hours for a flight crew member's flight duty period working in a basic (unaugmented) crew. Since the regulator also imposed that pre- and postflight duties included in that period could not be less than one hour, there would necessarily be additional crew in any flight of more than 12 hours (EU-OPS Subpart Q, Council Regulation (EEC) no. 3922/91).[8]

This type of regulation entails that ULH flights are discontinuously costly, and significantly so given that personnel constitutes a substantial share of the costs of a flight. In fact, the cost patterns documented by the U.S. FAA (FAA 2016, Table A-6) show that, for long-haul planes (wide-body, 300-plus seats) in passenger air carriers, crew corresponds to about 36% of nonfuel costs (11% of total costs). On top of that, additional crew and sleeping quarters imply less space and weight available for carrying payload, thus reducing revenue potential.

How important is this discontinuity in practice? Using the sample of cities with major international airports from the International Civil Aviation Organization, shown in the map in

7. Other restrictions apply once additional crew members (augmented crew) are required. For instance, pilots receive two local nights off on a layover after an augmented sector, and augmented pairings require an extra day off at home after the pairing (McKenney et al. 2000). Note also that in practice there are limitations even if flight time requires an augmented crew in only one direction of a pairing, as this is often required by the pilot's working agreement (contract) (McKenney et al. 2000, 54), although this contract varies by airline.

8. See https://www.law.cornell.edu/cfr/text/14/121.485 and http://www.vcockpit.de/fileadmin/dokumente/presse/2003/ECAFTLpositionFeb2002.pdf. The pattern holds beyond these examples. For one thing, U.S. regulations also have an impact elsewhere. For instance, in 1992 the Indian regulator introduced its own rules (AIC 28) essentially adopting U.S. standards, and Chinese regulations closely mimic these as well (CCAR Section 121.483). Other countries have very similar limits: daily flight duty time is capped at 14 hours in Canada, Australia limits two-pilot crews to 11 hours extendable to 12, and so on. (For a comparative account, see the Report of the Zaidi Committee, written for the Indian Ministry of Civil Aviation, available at http://dgca.nic.in/reports/Report_FDTL.pdf) Similarly, Singapore Airlines, which along with United Airlines were the top two airlines in terms of weekly flights above 12 hours, required double-augmented crews above 12.5 to 14 hours of flight duty time (depending on departure time) (McKenney et al. 2000, 48). Finally, other costs would also respond discontinuously for ULH flights: for instance, the United Kingdom's "air passenger duty," an excise duty levied on a per passenger basis, would become about 13% higher at the 6,000-mile threshold.

Figure I, we can calculate the distance between all city pairs and the number of nonstop flights between cities. (We describe the data in detail in Section III.)

The discontinuity manifests itself clearly in Figure II, Panel A, which depicts the number of city pairs connected to one another, defined as having at least weekly service between the cities, as of 2014. Each dot corresponds to the number of pairs within a 200-mile bin in terms of distance. We can see that although the number of flights unsurprisingly declines with distance, there is a sharp drop right at 6,000 miles: there are substantially more connected city pairs in which the two cities sit at a distance between 5,800 and 6,000 miles, say, than is the case for pairs situated between 6,000 and 6,200 miles apart. Interestingly, if we break down the connections by the type of aircraft, it is clear that the same models are flying below and above the threshold—consistent with the change in incentives introduced by the ULH range.[9]

This discontinuity, however, has not always been quite this pronounced. In fact, Figure II, Panel B displays the same information as in Panel A, except that it superimposes the data for 1989, and Panel C shows the change in connections from 1989 to 2014. It is apparent that the decline in the number of connected city pairs with respect to distance was a lot smoother back in 1989. From a purely descriptive perspective, this is due to the fact that, while the number of connections goes up between the two dates at just about any distance, the magnitude of the increase is noticeably larger between 4,600 and 6,000 miles. Panel D further clarifies that the discontinuity is not an artifact of potential connections due to geography: there is no sharp drop in total city pairs (using all possible permutations in our sample) around the same distance.

What explains this pattern? The late 1980s and 1990s witnessed important shifts in the long-haul civil aviation landscape, both in terms of technology and market structure, with the introduction of three very successful families of aircraft. The Boeing 747-400 started commercial operations in 1989; by 1990 there had already been over 100 units delivered, and the 747-400 went on to become the best selling subset in the 747 family, with more than 1,400 units delivered. A few years later, in 1993–1994, Airbus

---

9. For instance, as of 2014, in the range between 5,500 and 6,000 miles (resp. 6,000 to 6,500): 84 A330 (resp. 18), 89 A340 (24), 14 A380 (3), 91 Boeing 747 (19), 48 Boeing 767 (3), 127 Boeing 777 (76), and 27 Boeing 787 (16).
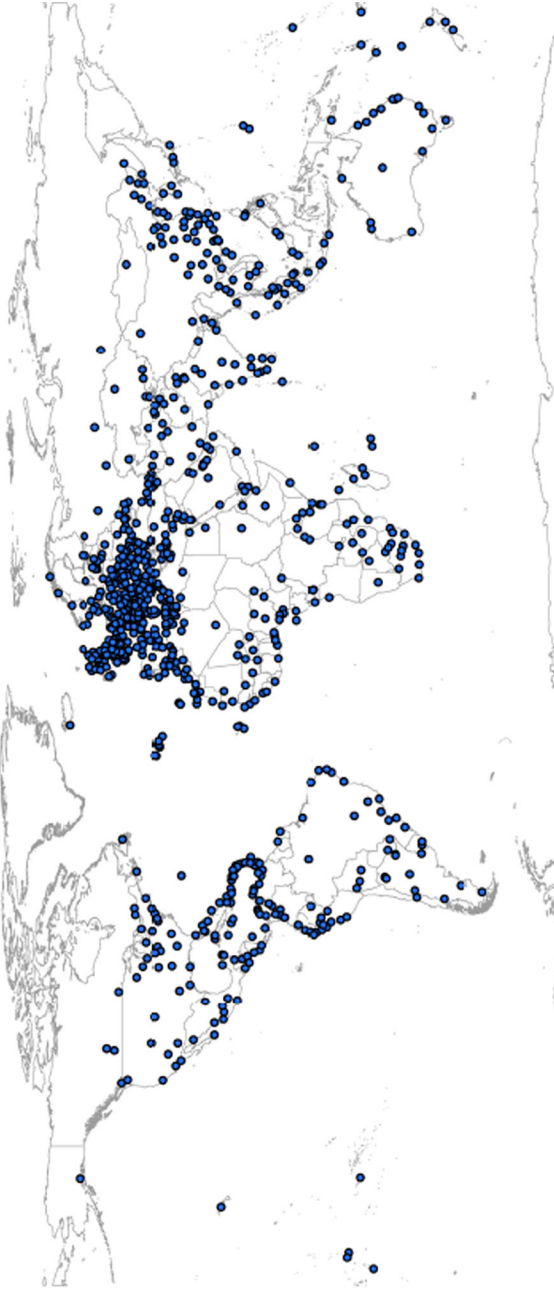
FIGURE I

Cities with Major International Airports

The map plots the locations of the 819 cities in the sample, all with major international airports as defined by the International Civil Aviation Organization (ICAO).
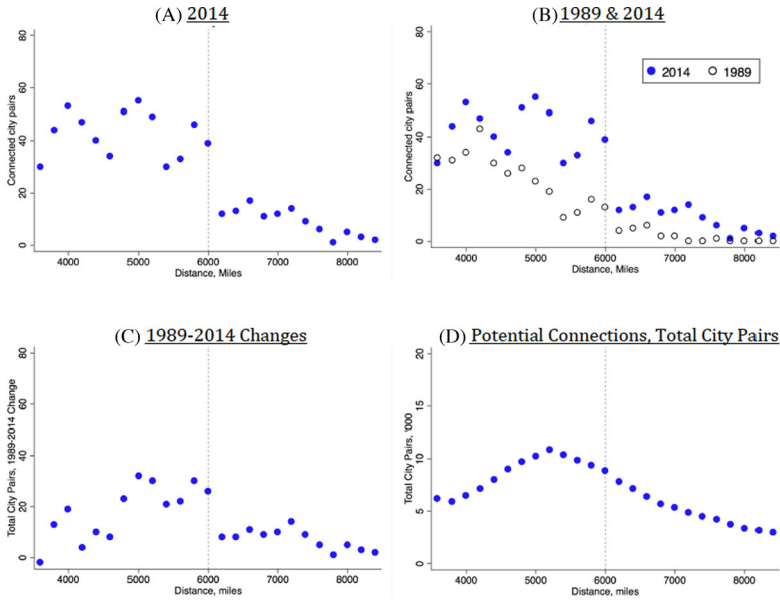
FIGURE II

Connections between City Pairs, by Distance

A connected city pair (airport pair) is defined as having at least weekly nonstop flights between the two cities. The data consists of the 819 cities in our baseline sample. Panel A displays the total number of connected city pairs in 2014 by distance. Panel B adds connected pairs in 1989. Panel C shows the change in connected pairs from 1989 to 2014. Panel D shows the total city pairs by distance, across all possible permutations of city pairs. The x-axis bin size is 200 miles. In each bin, the dot represents the number of city pairs in the preceding 200 miles. Together, the graphs show there is a clear discontinuity in connections around 6,000 miles in 2014, that this relationship is primarily driven by changes in connections after 1989, and that there is no sharp discontinuity for potential connections around 6,000 miles.

introduced its A330 and A340 models, which made the company into a serious competitor for Boeing; the A330/A340 have combined to sell more than 1,800 units. Finally, in 1995, the Boeing 777 family went into operation, eventually delivering nearly 1,500 planes.[10]

    These plane models made ULH flights substantially cheaper, because they combined long ranges with much improved fuel

---

    10. Full lists of plane deliveries by family of aircraft are available at https://www.planespotters.net/production-list/index. The numbers are as of the end of 2016.

efficiency. Kharina and Rutherford (2015, 14) show that the 747-400 family was about 20% more fuel-efficient than the preceding best-selling family of twin-aisle planes, from the early 1970s, and the 777 pushed that gain further to a total of about 30%.

This helped make long-haul flights commercially more viable. The rise of long-haul flights in turn made the discontinuity around the ULH threshold more meaningful over time, leading to the pattern that is apparent as of 2014. We will see in the next section, in the context of a simple conceptual framework, how the combination of a stable discontinuity in fixed costs and declining marginal costs of distance can generate exactly this kind of pattern over time.

This evolution took place very rapidly, as can be seen when we break down the number of routes by year and aircraft manufacturer, as we show in Online Appendix Figure A2. The number of long-haul flights (above 4,500 miles) goes up sharply right after 1989, and this is largely pushed by the range below 6,000 miles. This is in turn driven by Boeing aircraft, matching the introduction of the 747-400. Airbus then enters the long-haul market in 1993, exactly as the A330 and A340 come into the picture, and the increase in its presence is overwhelmingly in the below-6,000 mile range as well.[11]

As it happens, the period over which this discontinuity has existed provides us with a unique window to identify the causal impact of long-haul connections.[12] To the extent that cities that, as of the late 1980s, happened to have many airports lying just below 6,000 miles of distance do not differ systematically from cities that happened to have many airports just above that threshold, this distinction constitutes a source of exogenous variation in the number of places to which a city gets connected over the subsequent period.

11. The speed with which the increase takes place is not surprising: the 747-400 had been planned since 1984, and many orders were in place by 1985, leaving plenty of time for immediate adoption right on availability.

12. There is reason to believe that this discontinuity may be in the process of disappearing. Both U.S. and European regulators have adopted major revisions to flight time limit regulations in 2014 (known respectively as FAR 117 and new EU FTL), which in the European case required compliance by early 2016 at the latest. These impose stricter limits: FAR 117 in essence implies that two-pilot crews cannot fly for more than 9 hours, roughly speaking, whereas flight duty period in the new EU FTL is generally capped between 11 and 12 hours (depending on start times). See https://www.law.cornell.edu/cfr/text/14/part-117 and https://www.eurocockpit.be/sites/default/files/ftl_commission_regulation_83_2014.pdf.

III. Empirical Framework

We now describe how we implement this idea in practice. We begin with our key data sources and then motivate and discuss the empirical specifications we use.

III.A. Data

1. Air Links. Our key source of data comes from the International Civil Aviation Organization (ICAO). Specifically, we use the "Traffic by Flight Stage" (TFS) data set, which gives, for the period between 1989 and 2014, annual traffic on-board aircraft on individual flight stages of international scheduled services, and includes information on aircraft type used, number of flights operated, and traffic (passengers, freight, and mail) carried. This data set contains information on city and country names, which we use to merge it with information on the coordinates of major international airports, from the "Airport Traffic" data set.[13]

We are left with 819 cities with major international airports, from 200 different countries, which are shown in the map in Figure I. (Descriptive statistics can be found in Online Appendix Table A1.) We can see that the cities in our sample are spread all over the world, with some concentration in Europe due to a combination of its level of development and small country size. We use the definition of a "major" airport as given by the ICAO data, because we do not want to include small airports that would distort the picture we are trying to build: for instance, the key reference for a business located in Orange County, CA, is most likely the Los Angeles (LAX) airport, even though the local John Wayne Airport has a handful of international flights. Note that selection into the major airport category being correlated with the source of variation does not appear to be an issue: if that were the case and airports had been picked based on the number of potential connections around the 6,000-mile threshold, we would expect there to be a sharp discontinuity in the number of airport pairs at that point, which Figure II, Panel D shows not to be the case.

For each city pair in our sample, we can flag whether the pair are connected in any given year. The baseline analysis

---

13. For cities with more than one airport, we use the average coordinates of all airports in question. While keeping that in mind, for simplicity we use "airports" and "cities" interchangeably.

defines two cities as being connected if they have at least weekly flights between them.[14] We can either consider the snapshot of whether there is a connection in a specific year or aggregate the information over a multiyear period, which we do by adding the number of years within that period in which the two cities were connected to create a measure of connection-years. We study both measures, depending on whether the outcome of interest refers to a specific point in time, or to changes over a longer period. We also compute the shortest distance between the cities using their coordinates.[15]

We aggregate the information to the level of cities. For each of the cities in the sample, we calculate the number of other cities to which it is connected in a given year and the aggregate number of connection-years over a longer period. Similarly, we compute the total number of flights to and from the city, as well as the seats and passengers in them, and the number of countries to which the city is connected. We focus on flights of more than 2,000 miles to concentrate on the range over which airplanes are essentially the only relevant means of transporting people.[16]

We can summarize the quality of those links with a measure of network centrality. We focus on eigenvector centrality (Bonacich 1972), which relies on the idea that the prestige of a node in a network is related to the prestige of the nodes to which it is directly linked (Bloch, Jackson, and Tebaldi 2017). This seems important in this context, because it seems natural that a direct link to a well-connected airport is more economically meaningful than a link to a poorly connected one. To implement this idea, we describe the structure of the network of air links, across all distances, with the "adjacency matrix" A, in which each entry a_{ij} takes a value

14. We define a weekly connection as having at least 52 flights back and forth in a year. We show that the results are essentially identical, qualitatively speaking, using alternative definitions, such as twice-weekly connections (104 flights) and daily (365). There is clear bunching in the data around these values, making them natural definitions.

15. Specifically, using the *geodist* command in Stata, we compute the geodesic distance: the length of the shortest curve between two points along the surface of (a mathematical model of) the Earth. (This can be thought of as the "great-circle" distance, except that the latter term refers to a perfect sphere, which the Earth is not.) This is not the actual flight distance, in practice, but the latter is obviously endogenous to economic and geopolitical factors, so we choose to use the exogenous (and easily calculated) proxy.

16. Results are very similar if we use 3,000 miles as the threshold instead.

of 1 if cities i and j are connected, and 0 otherwise. If one assumes that the centrality of node i is proportional to the sum of centrality of node i's neighbors, then it follows that the nth component of the eigenvector associated with the greatest eigenvalue of A gives the measure of centrality of airport n, with the proportionality weight being the corresponding eigenvalue. Intuitively, this procedure thus assigns relative scores to all cities in the network while ensuring that connections to high-scoring cities contribute more to the score of a given city than equal connections to low-scoring cities.[17]

2. Economic Activity. To capture economic activity at the local level, on a global scale, we use the now standard information on light density measured by satellites at night, available from the National Centers for Environmental Information at the National Oceanic and Atmospheric Administration (NCEI-NOOA). This has become a widely used proxy for economic activity at the local level, as exemplified by a number of recent publications (Bleakley and Lin 2012; Henderson, Storeygard, and Weil 2012; Michalopoulos and Papaioannou 2013). We follow the data-cleaning procedure suggested by Lowe (2014), then aggregate the data into grid cells of size 0.25×0.25 degrees. We focus on growth over the two decades following the introduction of the Boeing 747-400 (and the start of the sample of air links), so we compute average nights lights in the cell for 1992 (first year available) and 2010.[18]

3. Business Links. To shed light on potential mechanisms behind the effect of air links on economic activity, we look at their impact on business links over long distances. For that, we make use of two data sets with spatial information and global coverage.

i. Firm ownership. We use the Orbis online data from Bureau van Dijk (BvD). Orbis is a database of firms that contains detailed financial, ownership, employment, location, and industry data on

---

17. As noted by Bloch, Jackson, and Tebaldi (2017, 6), this measure "is closely related to ways in which scientific journals are ranked based on citations, and also relates to influence in social learning." This underscores its appropriateness in our context, as we think of air links as facilitating face-to-face contact and the transmission of knowledge.

18. In the Online Appendix (Figures A3 and A4) we map the distribution of night lights for those years, around the world and in Asia (as an example for greater detail). One can clearly see substantial growth as well as changes in the geographical distribution of economic activity over the intervening period.

over 195 million firms in 229 countries. The sample consists of all of the one- and two-way business ownership links between cities located in different countries that are available in the online database.

To construct the network of foreign ownership links at the city pair level, we first consider the universe of firms that are owned by a foreign global ultimate owner (GUO). We define the GUO of a given firm as any company that owns a stake of 50% or more in the firm in question and is located in a country other than the one in which the firm is registered under by Orbis. The GUO is also an ultimate owner, which implies that it is not in turn owned by another company.[19]

We identify approximately 1.1 million firms that have a foreign GUO. For each firm we collect the following variables: name, BvD identification number, and spatial information (country, city) and the same information for the owner. Out of the initial set, we were able to obtain coordinates for 523,702 companies, in a total of 55,135 company cities in 181 countries, and 29,648 GUO cities in 183 countries.[20]

Since the Orbis online database is continuously updated, the data captures a cross-section of ownership as of the most recent update. The data was downloaded from April–June of 2016, and hence reflects a snapshot of ownership patterns as of that point in time.[21]

19. We also have information on whether the firm is owned by a foreign immediate shareholder (ISH). The ISH of a given firm is defined identically except that it may be owned by a GUO. For example a company in Sri Lanka may have an ISH in India, whose GUO is a holding company in the Netherlands. The Dutch company is therefore the GUO of both the Sri Lankan and the Indian companies. In 52% of the cases, the GUO and ISH are identical, and results are very similar using the ISH definition of ownership instead.

20. Specifically, we georeferenced the list of firms to provide latitude and longitude points, using an algorithm that searches inputted strings on Here Maps (https://maps.here.com/). By default, the search string describes a city, and the search yields the center point of the city in question. If information about the firm location beyond city was provided, the coordinates will identify specific districts, neighborhoods, or addresses within a city. In cases where an administrative unit larger than a city was provided in the data, the center point of the appropriate subnational unit is used.

21. Figures A5–A7 in the Online Appendix map the distribution of ownership links in space, with the caveat that Orbis coverage varies considerably from country to country. The first panel in each figure captures the total number of foreign-owned companies located in a given grid cell; the second panel, in contrast, displays the total number of companies located abroad that are owned by

ii. Major business events. We use the GDELT data set, which automatically codes online information on the occurrence and location of events mentioned in broadcast, print, and web news reports worldwide, in any of more than 100 languages. (For a more detailed description, see Leetaru and Schrodt 2013; Manacorda and Tesei 2016.) We use the "Historical Backfile" collection within GDELT, which consolidates information on events from January 1, 1979, through March 31, 2013. For each event the data report the exact day of occurrence and precise location (latitude and longitude of the centroid) at the level of city or landmark. We discard all events that cannot be located at that level of precision. Each event record codes two actors involved in the event, with latitude and longitude coordinates for each of them. We restrict attention to events where each actor is located in different places, using observations located within 100 miles of an airport in the ICAO data set. To capture our focus on business links, we look at observations coded as involving at least one "business" or "multinational corporation" actor, and as having to do with "material cooperation."[22]

We end up with a total of around 31,000 fully geocoded observations of major business events for the entire world. We aggregate the data into pre-1990 (1979–1989) and post-1990 (1990–2013) subperiods, and match the data to all the city pairs in the air links dataset. A graphical representation of the post-1990 events across the globe is provided in the Online Appendix (Figure A8), and from this it is apparent that these cooperation events are more evenly distributed than the ownership links. This

---

individuals or firms located in that grid cell. (We show pictures for the full set of foreign-owned firms, as well as for the subset of firms located at least 3,000 miles away from their owners, and again in the case of Asia for greater detail.) It is apparent, from the comparison between the two panels, that the latter is more geographically concentrated, indicating that owners are more unevenly distributed over space than the owned. This is in itself unsurprising, but the extent seems striking nonetheless.

22. Specifically, we rely on the Conflict and Mediation Event Observations (CAMEO) taxonomy used by GDELT. According to this taxonomy, the first step in coding each actor is to classify them as either "international" or "domestic." Out of the former, we look at events where at least one of the actors is described as "multinational corporations" (MNC); out of the latter, we consider those where at least one of the actors is classified as "business" (BUS), encompassing "businessmen, companies, and enterprises, not including MNCs." CAMEO also classifies events into 20 primary categories, and we restrict attention to those coded as "Engage in material cooperation."

is consistent with the fact that they represent weaker links between two businesses, and thus capture a different dimension of business interaction across distances.

4. Additional Variables. We use a number of variables as controls and/or for robustness checks. At the level of airports, we use the aforementioned ICAO sources to obtain airport characteristics as of 1989: numbers of daily, twice-weekly, and weekly flights; number of connected cities; number of connected countries (twice-weekly); total number of seats; total number of passengers; and total number of flights. We use distance to the equator and time zones, motivated by the evidence that time zone differences can affect the emergence of economic links across distances (Stein and Daude 2007).

At the level of grid cells, we use data on population from the Gridded Population of the World (GPW) version 4, which we obtain from the PRIO-GRID website, and we use the first available year (1990). We use geographic characteristics as controls (distance to equator, time zone, precipitation, temperature). Finally, we use real GDP per capita at the country level from the Penn World Tables 8.0, as well as World Bank country classifications into regions and income groups.

III.B. Identification Strategy and Specifications

To identify a causal effect of air links, we rely on the discontinuity in the likelihood of links as a function of the distance between two cities, at 6,000 miles. The article now discusses how we use it to implement the empirical strategy. For that, it is useful to lay out a simple conceptual framework describing how a discontinuity in costs in line with what was discussed in Section II can lead to a discontinuous drop in the likelihood that two airports are connected. We use this framework to motivate the specifications for two distinct levels of analysis, depending on the nature of the outcomes of interest: city pairs and grid cells.

1. A Simple Model of Connections. Consider the decision of whether to establish a direct connection between two cities, i and j, located at a distance d_{ij} from one another. We model the profit coming from this connection as:

$$(1) \qquad \pi_{ij} = m_{ij} - f - d_{ij} * c.$$

The revenue generated by the connection depends on the "economic potential" involved in linking the cities, denoted m_{ij}. This term could depend on, say, the product of (a function of) the GDPs of the two cities, or the countries they are in—as in the gravity equation describing trade between countries—as well as myriad other factors, running from market power to cultural proximity. Let $G(m; d)$ denote the distribution of m_{ij}, conditional on d_{ij}, for all the pairs of airports in our sample.

The cost component in equation (1), in turn, has two parts. First are costs—most importantly, fuel—that vary continuously with the distance between the two cities. For simplicity, we assume that they increase linearly, $d_{ij} * c$. Second are "fixed" costs that do not vary smoothly with distance: for instance, one cannot adjust the size of the flight crew or operational costs at the airport continuously with the flight distance. These are denoted by f.

In the context of this framework, we can model the two key takeaways from the discussion in Section II as follows. First, the regulations requiring additional crew for ULH flights entail that f is a discontinuous function of distance:

$$(2) \qquad f(d_{ij}) = \begin{cases} f_{high}, & \text{if } d_{ij} > 6,000 \\ f_{low}, & \text{if } d_{ij} \leqslant 6,000, \end{cases}$$

with $f_{high} > f_{low}$. Second, we can capture the technological transformation making long-distance flights cheaper, with the introduction of the Boeing 747-400, and later the Airbus A330/A340 and Boeing 777, as reducing the marginal cost of distance, c. In other words, the initial technological environment has c_{high} and the introduction of the new aircraft changes the parameter to c_{low}, with $c_{high} > c_{low}$.

A connection between i and j will be established whenever $\pi_{ij} \geqslant 0$. As a result, we can write the probability of such a connection, p_{ij}:

$$p_{ij} = Pr\left(\pi_{ij} \geqslant 0\right) = Pr\left(m_{ij} \geqslant f(d_{ij}) + d_{ij} * c\right)$$
$$(3) \qquad = 1 - G(f(d_{ij}) + d_{ij} * c; d_{ij}).$$

Although extremely simple, this model already conveys the impact the two key features can have over the pattern of connections. To illustrate that, we run a set of simulations using figures designed to calibrate actual cost patterns as described in Section II: a 30% reduction from c_{high} to c_{low}, and an increase in

f of 11% at the 6,000-mile discontinuity.[23] The increase in f is calibrated for a 33% increase in crew costs, given that crew corresponds to about one-third of nonfuel costs (FAA 2016, Table A-6). However, because fuel corresponds to more than two-thirds of costs for long-haul planes (wide-body, 300-plus seats) in passenger air carriers, this entails an increase of a mere 3.3% in total costs for a 6,000-mile flight.

Yet the simulation results, depicted in Online Appendix Figure A9, show that this modest increase has substantial effects on the presence of connections around the discontinuity.[24] Not only is there a sizable drop in the number of connections across the 6,000-mile threshold in the c_{low} distribution, but the discontinuity is much more pronounced in that case than in the c_{high} distribution, even though the regulation over the fixed costs is the same in the two cases. The intuition for that can be seen quite clearly from the limit case in which c grows without bound: G will tend to 1 for any distance—no flight is economically viable, and hence no discontinuity will emerge regardless of the presence of regulations over f.

2. City-Pair Analysis. The key identification assumption is that $G(m; d)$ is continuous in a neighborhood of the $d = 6,000$ threshold: there is nothing about the economic fundamentals of city pairs in the sample that changes discontinuously around 6,000 miles of distance. In other words, whether the bilateral distance between any two cities happens to be just above or just below 6,000 miles is as good as randomly assigned.

If that is the case, for any outcome of interest defined at the level of city pairs, Y_{ij}, we can perform a regression discontinuity (RD) analysis:

$$(4) \qquad Y_{ij} = \alpha + \beta * Below6K_{ij} + g(d_{ij}) * \gamma + \varepsilon_{ij},$$

where $Below6K_{ij}$ is a dummy equal to 1 if d_{ij} is less than 6,000 miles. It is well known that higher-order polynomials in $g()$ can result in approximation errors due to overfitting or biases at boundary points, so in the baseline specification we use a parsimonious

23. Details of the calibration can be found in the Online Appendix.
24. The figure assumes that m_{ij} is log-normally distributed, which is a good approximation for the actual distribution of the product of night lights (as of 1992) in the sample of airport pairs, as described in the Online Appendix. The results depict the average outcome of 100 simulation rounds.

specification allowing for different linear slopes above and below the 6,000 mile threshold. We provide robustness tests using a second-order polynomial, as well as estimates using various sample bandwidths, including the optimal bandwidth that minimizes the mean squared error of the point estimator, using the algorithm developed by Calonico, Cattaneo, and Titiunik (2014).[25] Following Imbens and Lemieux (2008), we adopt robust standard errors as our baseline specification. However, we also check robustness by showing standard errors clustered at the country-pair level, thereby allowing for correlation between city pairs located in the same country pair.

To test for a first-stage relationship, with respect to p_{ij}, we use an outcome variable indicating whether the city pair is connected or the number of connection-years between them, defined by at least weekly flights. If the 6,000-mile threshold is meaningful, we expect β to be positive. We estimate reduced form effects on other city-pair outcomes. Under the exclusion restriction that outcomes around the threshold are affected only through a change in the likelihood of getting connected, we present scaled instrumental variable estimates—the marginal effect of getting connected on outcomes—using a "fuzzy" RD approach.

3. Grid-Cell Analysis. To test whether connections affect economic activity, we must expand the analysis beyond city-pair outcomes and translate the empirical strategy to use data at the grid-cell level.

Let $\delta > 0$ denote a relatively small number (compared to $6,000$ miles); we refer to a "δ neighborhood" of 6,000 miles as the set of distances between $6,000 - \delta$ and $6,000 + \delta$ miles. For each city i, we can define the conditional distribution describing the economic potential of all the pairs involving i, which with a slight abuse of notation we call $G_i(m; d)$. This distribution contains information about the broader economic potential of the city: for instance, if we consider two cities, a and b, such that G_a dominates G_b in the first order stochastic sense, we can state that the economic potential of a is greater than that of b. In consonance with the basic identification assumption underlying the RD analysis, G_i is assumed to be continuous in a neighborhood of the $d = 6,000$ threshold.

---

25. This is implemented in Stata using the *rdrobust* routine.

Given this identification assumption, we can write for any city i, as shown in the Appendix, the number of connections to cities within a δ neighborhood of 6,000 miles away from i, $K_{i\delta}$, as:

$$(5) \qquad K_{i\delta} = N_{i\delta} * \left(1 - z^i_{high}\right) + \left(z^i_{high} - z^i_{low}\right) * N_{i\delta-} + \nu_i,$$

where $N_{i\delta}$ denotes the total number of cities within a δ neighborhood of 6,000 miles away from i; $N_{i\delta-}$ is the number, out of those cities, that happen to be closer than 6,000 miles; and ν_i is an error term. The parameters z are defined as $z^i_{low} \equiv G_i(f_{low} + 6,000 * c; 6,000)$ and $z^i_{high} \equiv G_i(f_{high} + 6,000 * c; 6,000)$; because $f_{high} > f_{low}$, and G_i is a probability distribution, it follows that $0 \leqslant z^i_{low} \leqslant z^i_{high} \leqslant 1$.

The first key takeaway from equation (5) is that, controlling for the total number of cities around 6,000 miles away from i, the number of i's air links around that range is an increasing function of the number of cities that happen to be just below that threshold. This is an intuitive implication of flights below 6,000 miles being discontinuously cheaper, which makes them discretely more likely to be profitable.

A second takeaway has to do with when we should expect there to be a link between actual connections and the number of potential connections just below the 6,000-mile threshold. First, it is clear from the definitions that z_{low} (and also, necessarily, z_{high}) will approach 1 as c grows larger. From equation (5) it follows that, when the cost of flying long distances is very high, we should expect the link to disappear. Similarly, for any given cost parameter values, we should expect the link to be absent in places with sufficiently low economic potential—namely, if G_i is such that $z_{low} = 1$.

When it comes to the potential impact of adding connections, it is natural to expect that it would depend on whom exactly one is getting connected to: linking up with London would presumably matter more than linking up with Stockholm, not least because from the former it is possible to get to many more places. One way to capture that in the context of this model is to weight each destination j by the economic potential available in linking city i to it, m_{ij}. In the Appendix, we show that this weighted measure of connections, $\hat{K}_{i\delta}$, can be written as:

$$(6) \qquad \hat{K}_{i\delta} = M_{i\delta} * \left(1 - z^i_{high}\right) + \left(z^i_{high} - z^i_{low}\right) * M_{i\delta-} + \mu_i,$$

where $M_{ik} \equiv \sum_{j \in \Omega_i^k} m_{ij}$ is the total economic potential associated with links to all the cities in Ω_i^k. This has an analogous

interpretation to that of equation (5), with each city being weighted by its associated economic potential.

Equations (5) and (6) lay the groundwork for the empirical specification, because they flag a source of variation for the number and quality of connections. The identification assumption is that there is no reason airports that happen to have relatively many major 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: $N_{i\delta}$ or $M_{i\delta}$ could be correlated with characteristics of city i that also affect its economic fundamentals—after all, they should contain information on where that city is located in the globe, and hence its degree of isolation and other geographical features. Our assumption is that conditional on $N_{i\delta}$ or $M_{i\delta}$, the share of those that happens to fall below 6,000 miles is as good as randomly assigned.

To implement this logic, we define the unweighted instrument, $ShareBelow6K_i^U$, as the number of airports 5,500 to 6,000 miles away from airport i, divided by the number of airports 5,500 to 6,500 miles away.[26] To build intuition for how this instrument is constructed, Figure III provides the graphical example of San Francisco (SFO).

Our baseline instrument, $ShareBelow6K_i$, is the weighted version that incorporates the information related to the potential of each connection, as per equation (6). Specifically, we proxy each airport's potential using its (eigenvector) centrality at the beginning of the sample (1989), and sum the centrality measures for all airports within the 5,500–6,000-mile range from airport i. We divide that sum by the sum over the entire 5,500–6,500-mile range, to define $ShareBelow6K_i$. (The list of top and bottom 50 cities, ranked by $ShareBelow6K^U$ and $ShareBelow6K$, can be found in the Online Appendix, Table A2.) In the spirit of making the best use of the available information, we focus on $ShareBelow6K$ as the preferred instrument.

In our specifications, we control for the total number and network centrality (as of 1989) of airports in the range of 5,500 to 6,500 miles. This accounts for factors related to the general isolation or broad location of the airport, so that we only make use

---

26. We provide robustness tests showing that using alternative windows does not qualitatively alter the main results.
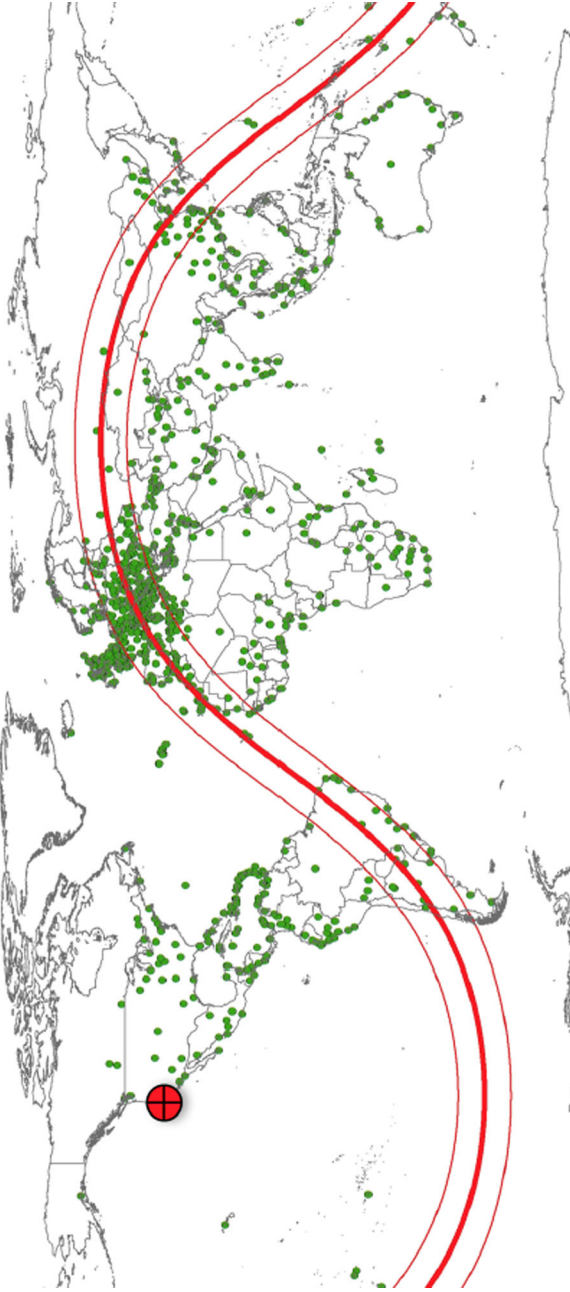
Figure III

Constructing the Airport-Level Instrument, SFO Example

The thick red line is drawn 6,000 miles from San Francisco International Airport (SFO) (color artwork is available at the online version of this article). The buffer around the thick line indicates which other airports (cities) are located within 5,500–6,500 miles from SFO. For each of the 819 observations, the airport-level instrument is the share of other cities within the buffer that are located below 6,000 miles, weighted by the centrality of the city as of 1989.

of the residual, arguably idiosyncratic variation.[27] The graphical representation in Figure IV displays this residual variation in *ShareBelow*6*K* after controlling for the total network centrality of airports in the 5,500–6,500-mile range and region fixed effects. We can see that the places with high and very low draws in the "lottery" of potential connections just below the threshold are spread all over the world and in a relatively random manner within regions. Also notably, we can see that there are places with very positive and very negative shocks located close to one another. This reassures us that the variation is essentially idiosyncratic, and not driven by specific parts of the world.

We start with the following reduced-form specification:

$$(7) \qquad Y_{ic} = \alpha^r + \beta^r * ShareBelow6K_{ic} + X_{ic}\gamma^r + \varepsilon_{ic}^r,$$

where c denotes a grid cell, i denotes the closest airport (within the same country) in the sample, Y_{ic} is an outcome of interest (e.g. night lights in the cell), $ShareBelow6K_{ic}$ is the value of the instrument at the closest airport, X_i is a vector of control variables. If connections foster economic growth in areas close to the airport, for instance, we expect β^r to be positive.

All regressions include in the vector X the total number of airports between 5,500 and 6,500 miles away, weighted by network centrality in the baseline specification, as discussed already, as well as the log distance in miles from the grid cell c centroid to the airport i, and region fixed effects to ensure that the results are not driven by variation across regions. We further control for grid-cell night lights as of 1992 (earliest data available) and population as of 1990, to reduce residual variation and increase the precision of our estimates given persistence in the data over time, and for robustness purposes.[28] In addition, we use various predetermined covariates to ensure that the results are robust.

---

27. As an example, Philadelphia and Boston will naturally have a similar number of airports located between 5,500 and 6,500 miles away (66 and 57, as it happens), because they are close to each other (about 280 miles). However, the share of those that happens to fall just below the 6,000-mile threshold is 64% larger for the former than for the latter, with Boston being in the bottom decile of that distribution and Philadelphia just below the median.

28. These variables may have been affected by our discontinuity themselves, given the timing of the regulations and the entry of the Boeing 747-400, as we have discussed. However, we show that our variation is essentially uncorrelated with them, consistent with the idea that important effects would have taken some years to be felt.
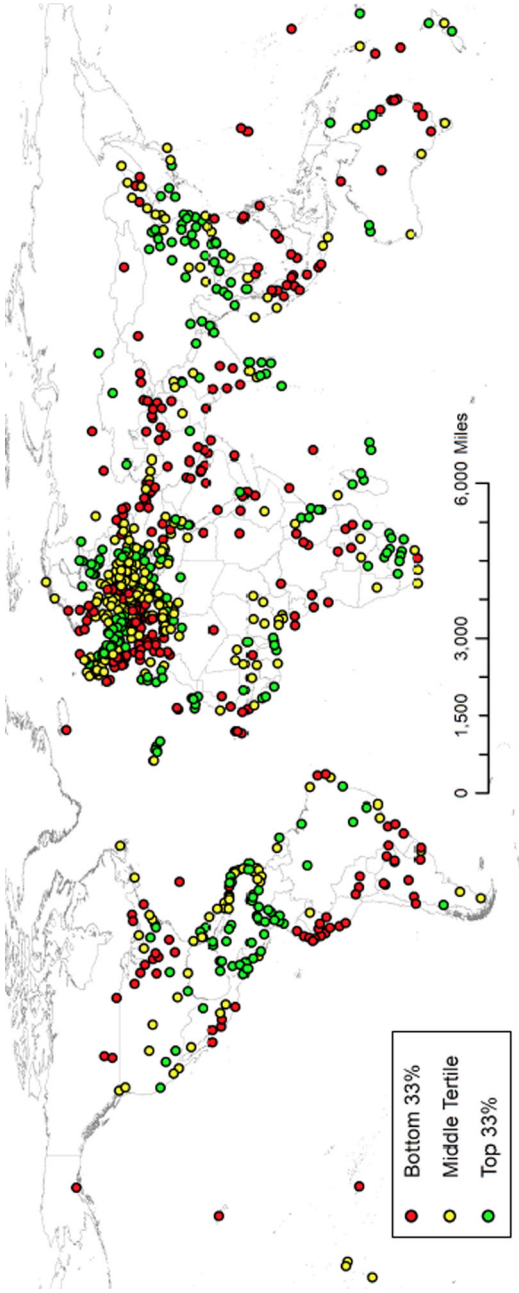
Figure IV

Identifying Variation, Airport-Level

The map depicts the identifying variation across the 819 airports. The identifying variation for each airport/city is the OLS residual of the instrument—the centrality-weighted share of airports below 6,000 miles in the 5,500 to 6,500 miles buffer—after controlling for region fixed effects and the centrality-weighted number of other airports (i.e., total network centrality) in the buffer. The map shows that there is meaningful variation within regions.

To estimate the magnitude of the effects in ways that are more easily interpretable, we scale the reduced-form estimates with a first-stage estimate using two-stage least squares (IV/2SLS). The endogenous variable may be simply the number of connections, as in equation (5), or account for their quality, as in equation (6). In either case, we obtain an estimate that captures the effect of improving an airport's position in the network of air links—induced by long-distance connections—on variables of interest at the local level. The corresponding first-stage specifications are:

$$(8) \quad Connections_{ic} = \alpha^{f0} + \beta^{f0} * ShareBelow6K_{ic} + X_{ic}\gamma^{f0} + \varepsilon_{ic}^{f0},$$

$$(9) \quad Centrality_{ic} = \alpha^{f1} + \beta^{f1} * ShareBelow6K_{ic} + X_{ic}\gamma^{f1} + \varepsilon_{ic}^{f1},$$

where $Connections_{ic}$ is the number of cities the airport i is connected to (at least weekly flights), $Centrality_{ic}$ is the network centrality of airport i and all other variables are defined as in equation (7).

We can also exploit the granularity of the data to uncover spatial patterns in our effects. Intuitively, we would expect the economic activity in cells around the airport to be affected, if at all, only if they are relatively close by. The estimations of equation (7) include grid-cells within 100 miles of the airport, as it seems plausible ex ante that such cells are potentially affected. (We will show robustness with respect to other thresholds.) However, because we would expect the effects on economic activity to depend on how close a cell is to the airport, we can estimate the reduced-form effects as a function of the spatial distance to the airport:

$$
\begin{aligned}
Y_{ic} = {} & \alpha + \beta_1 * ShareBelow6K_{ic} \\
(10) \quad & + \beta_2 * ShareBelow6K_{ic} * Distance_{ic} + X_{ic}\gamma + \varepsilon_{ic},
\end{aligned}
$$

where $Distance_{ic}$ is the log distance in miles from the grid cell c centroid to the airport i.

This specification allows us to test whether any positive effects dissipate with distance and at what rate. More precisely, β_1 captures the reduced form effect of connections for grid cells that are located in the immediate vicinity of the airport, since $Distance_{ic}$ takes values around 0 in those cases. By contrast, β_2 captures the marginal effect of distance to the airport on the

treatment effect. If connections result in positive effects that are maximized in areas in and around the city, and dissipate with distance, we expect β_1 to be positive and β_2 to be negative. The combination of the two estimates will, in turn, allow us to probe at which distance the effects are no longer positive. We provide results using more flexible estimations of these spatial relationships.

In all of the specifications, we cluster the standard errors at the country level to allow for the possibility of correlated shocks across cities in the same country. We also show robustness to other approaches to computing the standard errors—namely, clustering at the level of airports to deal explicitly with the fact that our key variation is at that level, and implementing the Conley (1999) correction for spatial correlation.

IV. Impact on Air Links

IV.A. *Establishing the Discontinuity*

The first step in the analysis is to show how the variation translates into more and better air links. While Figure II has provided graphical evidence for the existence and evolution of the discontinuity in the likelihood of connection between city pairs at a distance of 6,000 miles, we now turn to the task of establishing this more systematically.

The key evidence is in Table I, implementing versions of equation (4) with both robust and country-pair-clustered standard errors reported. We see a robust pattern where city pairs just over 6,000 miles apart are about 0.3–0.4 percentage point less likely to be connected by at least weekly flights, as of 2014, as compared with those separated by slightly less than 6,000 miles. Because the overall likelihood of a given pair in our sample being connected is around 1%, this entails a quantitatively substantial difference.

The result holds with a first-order polynomial for $f(Distance_{ij})$ (columns (1)–(4)), as well as with a second-order polynomial (columns (5) and (6)). It is not affected by different bandwidth choices either: we start off with a narrow window of 500 miles (column (1)), which we expand to 1,000 miles (column (2)), before presenting the optimal-bandwidth baseline (column (3)). It is also robust to controlling for time zone differences and for whether the pair was already connected in 1989, as well as a set of 1989 covariates measuring the extent of connections between the two countries in the pair (columns (4) and (6)).

TABLE I

REGRESSION DISCONTINUITY REGRESSIONS AROUND 6,000 MILES, CITY-PAIR LEVEL

	Dep. var:: Connected, dummy						1989	Dep. Var: Years Connected		
Bandwidth, Miles	2014 500 (1)	2014 1,000 (2)	2014 Optimal (3)	2014 Optimal (4)	2014 Optimal (5)	2014 Optimal (6)	Optimal (7)	1990–2000 Optimal (8)	2000–2010 Optimal (9)	1990–2010 Optimal (10)
Distance < 6,000 Miles	0.0031	0.0035	0.0038	0.0029	0.0040	0.0030	0.0010	0.0139	0.0187	0.0344
Robust std. err.	(0.0012)***	(0.0008)***	(0.0009)***	(0.0008)***	(0.0010)***	(0.0009)***	(0.0005)*	(0.0043)***	(0.0075)**	(0.0114)***
Cluster std. err.	(0.0012)***	(0.0009)***	(0.0010)***	(0.0008)***	(0.0011)***	(0.0009)***	(0.0006)*	(0.0045)***	(0.0077)**	(0.0117)***
Polynomial order	1st	1st	1st	1st	2nd	2nd	1st	1st	1st	1st
Baseline covariates	No	No	No	Yes	No	Yes	No	Yes	Yes	Yes
Observations, total	334,954	334,954	334,954	334,954	334,954	334,954	334,954	334,954	334,954	334,954
Observations, effective	41,477	81,842	65,323	72,243	113,565	123,522	60,595	103,716	71,120	73,846
Dep. var. mean	0.0103	0.0103	0.0103	0.0103	0.0103	0.0103	0.0057	0.0678	0.0938	0.1554

Notes. Local polynomial regression discontinuity estimates, using the rdrobust command in Stata (default options unless otherwise stated). Running variable is distance between airports, discontinuity at 6,000 miles. Optimal bandwidth selected using one common MSE-optimal bandwidth selector. "Polynomial order" refers to the order of the polynomial in distance. The dependent variable is a dummy indicating whether the city pair had (at least) weekly flights in 2014. Standard errors are in parentheses. "Cluster std. err." refers to cluster-robust nearest neighbor variance estimation at the country-pair level, using a minimum of three nearest neighbors. The point estimates and the number of effective observations refers to the number of observations within the bandwidth when using the robust standard errors. The covariates are all measured in 1989 and consist of: a dummy for having weekly flights, time zone difference in hours, log of total city pairs within the country pair, log total connections within the country pair, and log total passengers within the country pair; where the logged variables add one to deal with undefined log function at zero . ***$p < .01$, **$p < .05$, *$p < .1$.
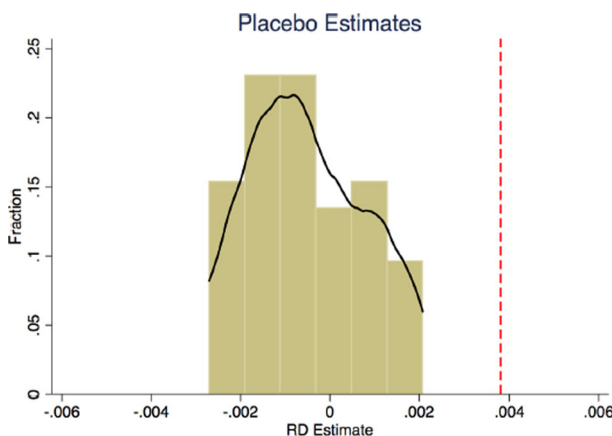
FIGURE V

Placebo RDD Estimates on City Pair Connections, 2014

Histogram (and kernel approximation) for regression discontinuity estimates computed using each 50-mile point between 4,500 and 5,750 miles, and between 6,250 and 7,500 miles as distance thresholds. Specifications use first-order polynomial and optimal bandwidth. Vertical dashed line depicts the estimate using the 6,000-mile threshold. The plot shows that the estimate at the 6,000-mile threshold is a clear outlier.

We probe this result by running specifications where we arbitrarily impose "placebo" discontinuity thresholds other than 6,000—specifically, every 50 miles between 4,500 and 7,500, leaving aside the range between 5,750 and 6,250 miles as the match between the regulations and the specific distance is an approximation.[29] Figure V shows that the estimate at the 6,000-mile threshold is much larger than the placebo alternatives, falling far to the right of the distribution computed for the latter. This reassures us that the effect we pick up is unlikely to be spurious. We can also ask how the discontinuity evolved over time. Column (7) in Table I implements the baseline specification, but with a dummy for the presence of a connection in 1989 as the dependent variable. The coefficient is relatively small (p-value $= .099$), indicating that in the year of the launch of the Boeing 747-400,

29. Notably, the absolute value of the coefficient is maximized precisely at 6,000 (0.0038), with the next-highest value at 0.0023 for the discontinuity set at 6,050. The specification here includes a first-order polynomial in distance, and optimal bandwidth, as well as standard errors clustered at the level of country pairs.

the likelihood of connection just below the threshold was perhaps higher but not strongly so—consistent with a scenario of relatively high costs of flying long distances, as laid out in the context of our model.

The remainder of the table aggregates the information for the subsequent two decades, using connection-years as the outcome of interest. We see that the effect is already strongly significant in the 1990s (column (8)), and gets stronger in the 2000s (column (9)). All in all, the presence of connections over the entire period is markedly higher just below the 6,000-mile threshold (column (10)). In short, the discontinuity seems to have magnified rapidly upon the technological developments of the late 1980s and 1990s, becoming further established over time.

Having established the presence of the discontinuity at the level of city pairs, we turn to how it translates into the level of airports, which we use for the grid-cell analysis. Put simply, does the share of potential connections just below the threshold predict the total number and quality of connections that are actually available in airport i?

Table II answers in the affirmative. We start off with the un-weighted instrument, $ShareBelow6K^U$, and whether it predicts the total number of (at least) weekly connections, initially with only the baseline controls for the total number of airports in the 5,500–6,500-mile range and region fixed effects. Column (1) shows that there was no significant correlation as of 1989, again consistent with the relatively high-cost scenario.[30] In contrast, we see in column (2) that as of 2014 there is a strong correlation.

The magnitude of the effect is largely unaffected, and precision is improved, when we control for airport characteristics as of 1989, including initial connections (column (3)). Connections in 1989 and 2014 are highly correlated, as would have been expected, and this helps account for the increased precision. The same is true when we control for distance to the equator and time zone (Column (4)), or population as of 1990 and night lights as of 1992 (Column (5)), indicating that even within regions the effect is not driven by location features or initial development.

---

30. In fact, in the Online Appendix (Table A3) we show that $ShareBelow6K$ is not significantly correlated with any of the 1989 airport characteristics, with quantitatively small standardized effects, again indicating that the effect of the discontinuity was weak at best at the time of the introduction of the Boeing 747-400.

TABLE II

EFFECT ON AIR LINKS, AIRPORT LEVEL

Dependent variable:	Connected cities, 1989	Connected cities, 2014	Connected cities, 2014	Connected cities, 2014	Connected cities, 2014	Network centrality, 1989	Network centrality, 2014	Network centrality, 1990–2000	Network centrality, 1990–2010	Network Centrality, 1990–2010	Network Centrality, 1990–2010
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Share of cities < 6,000 miles, unweighted	−0.46	4.88**	4.89***	5.05***	5.14***					0.27**	0.05
	(1.56)	(2.28)	(1.43)	(1.38)	(1.45)					(0.11)	(0.13)
Share of cities < 6,000 miles, centrality-weighted						−0.14	0.37***	0.24***	0.26***		0.24***
						(0.18)	(0.10)	(0.06)	(0.07)		(0.09)
Observations	819	819	819	819	777	777	777	777	777	777	777
R-squared	0.04	0.06	0.62	0.62	0.63	0.19	0.68	0.93	0.89	0.89	0.89
Controls											
Number of cities around 6,000 miles	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region fixed effects	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Airport controls, 1989	No	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Geographic controls	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional controls	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dep. var. mean	1.44	2.83	2.83	2.83	2.94	0.02	0.02	0.02	0.02	0.02	0.02

Notes. A unit of observation is an airport/city. The outcome in columns (1)–(4) is the number of cities connected to, where a connection is defined as long-distance weekly flights (at least) to and from the city. In columns (5)–(10), the outcome is the network centrality, which is the standardized eigenvector centrality of the city based on international connections to cities (i.e., at least weekly flights) at all distances. In columns (1)–(6) the outcome is measured in a given year, whereas in columns (7)–(10) it refers to the yearly mean over the specified time period. The unweighted share of cities < 6,000 miles is simply the number of cities 5,500–6,000 miles away, divided by the total number of cities 5,500–6,500 miles away, whereas the centrality-weighted version weights each city by their network centrality as of 1989. All regressions control for the total number of cities 5,500–6,500 miles away, unweighted in columns (1)–(4) and centrality-weighted in columns (5)–(10). The successive controls: region fixed effects are a set of region dummies; airport controls, 1989 refers to measures as of 1989 (Number of weekly, twice-weekly, and daily flights, respectively; number of connected cities; number of connected countries (twice weekly); log of total number of seats; log of total number of passengers; log of total number of flights; Network centrality); Geographic controls consist of log distance to the equator and time zone difference from the airport to GMT in hours; additional controls consist of average 1989 night lights and average 1990 population in the cells within 100 miles of the airport. Robust standard errors in parentheses, clustered at the country level. *** $p <.01$, ** $p <.05$, * $p <.1$

We turn to our preferred, weighted instrument, *ShareBelow*6K. Column (6) shows that the instrument is uncorrelated with the position of the airport in the network as of 1989. Again, a strong correlation is present in 2014, even including the full set of controls (column (7)). Columns (8) and (9) then consider the evolution of that pattern over time. We see that the correlation was already present when the dependent variable is the average network centrality over the 1990–2000 period, and if anything became stronger subsequently, as the coefficient increases when we consider the two-decade average over 1990–2010.[31]

It is interesting to note that the unweighted instrument already predicts the quality of the airport's position in the network, and not just the number of connections (column (10)). That said, when the two instruments are jointly included (column (11)), it is clear that the predictive power of the weighted instrument is far greater, unsurprisingly given that it incorporates more information.

The estimated magnitudes indicate a substantial effect. (The measures of centrality are standardized to facilitate the comparison of magnitudes.) A coefficient of 0.26 (column (9)) entails that going from the 25th percentile to the 75th percentile in *ShareBelow*6K (0.471 and 0.707, respectively) would translate into an improvement of 0.06 standard deviations in the average quality of the airport's position in the network. This corresponds to the median airport moving 69 spots up the centrality rankings.

---

31. Note that we avoid using country fixed effects primarily because it greatly reduces the variation in the instrument, leading to substantially larger standard errors and a statistically insignificant first-stage relationship. This is unsurprising given that fixed effects effectively drop the variation in the 99 countries that have only one airport–not to mention the fact that our estimates would then only use the rather selected subsample of countries with multiple major airports (which tend to be wealthier). The instrument is also naturally highly correlated across airports within countries, by construction, since it exploits the location of airports around 6,000 miles away. This further reduces the variation. It is worth noting that the point estimates remain similar, as we show in the Online Appendix (Figure A10 and Table A6), indicating that unobservable country-level characteristics are indeed unlikely to drive the relationship. Because there is no a priori reason to think that the variation is correlated with omitted variables that would distort the results, and since the stability of the coefficient suggests this does not happen to be the case, we prefer the specification with region fixed effects to avoid discarding relevant information.

Last, we also check that the results are robust to different ways of implementing our variation. In particular, we show in the Online Appendix (Table A5) that they still hold when considering the number of cities between 5,500 and 6,000 miles, instead of the share, as well as when we define connections based on the presence of twice-weekly or daily flights, or when we construct *ShareBelow6K* over different windows (5,700–6,300, 5,200–6,800).

IV.B. Network Spillovers

The key source of variation affects directly the availability of air links over a specific range, yet we found effects on the total number and quality of connections available at a given airport. We now ask how we can go from that specific shock to these broad effects.

We start by looking at how a favorable draw in terms of *ShareBelow6K* affects connections in the range around 6,000 miles. The result is in column (1) of Table III. We see that places with better potential connections just below the threshold indeed add more connections over the 5,500–6,500-mile range.

In column (2) we consider the 2SLS estimate of the impact of an additional connection over that range, and find evidence of important spillover effects: about five to six total long-haul connections overall. Column (3) shows that the result is unaltered when we add the full set of control variables from Table II. What's more, the contrast between Columns (4)–(6) and (7)–(9) shows that the spillovers are essentially coming from the shorter range between 2,000 and 5,500 miles. All in all, this is eminently consistent with the idea that having more direct flights increases the value of an airport for others to connect to: connections induce further connections.

Those better and more plentiful connections also increase the flow of people. Of course, it would be rather surprising if that were not the case, and the last three columns in Table III confirm that intuition. The 2SLS estimate indicates that increasing the airport's centrality in the network by one standard deviation— which, for the sake of comparison, would move the median airport about 300 positions up the rankings—increases the yearly number of passengers going through the airport by roughly 1.3 million, or five-sixths of a standard deviation.

Taken together, these results establish that the impact of a shock yielding more connections within a relatively narrow range

TABLE III
NETWORK SPILLOVERS, AIRPORT LEVEL

	Direct effect	Total effect		Spillover effects						Total passengers in 2014, millions		
	5,500–6,500 miles	All distances		2,000–5,500 miles			> 6,500 miles					
	FS	2SLS	2SLS	RF	2SLS	2SLS	RF	2SLS	2SLS	RF	2SLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Share of cities < 6,000 miles	0.51*** (0.18)			2.12*** (0.72)						0.456** (0.185)		
Connected cities in 2014, 5,500–6,500 miles		5.88*** (1.63)	6.43*** (2.08)		4.15*** (1.37)	4.76*** (1.77)	0.24 (0.17)	0.47 (0.28)	0.50 (0.37)			
Network centrality, 2014											1.318** (0.420)	1.274** (0.470)
Anderson-Rubin p-value	NA	0.002	0.004	NA	0.004	0.005	NA	0.168	0.241	NA	0.015	0.028
Sanderson-Windmeijer F-stat	NA	8.01	6.97	NA	8.01	6.97	NA	8.01	6.97	NA	11.12	11.88
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional controls	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Observations	819	819	777	819	819	777	819	819	777	819	819	777
R-squared	0.69	NA	NA	0.59	NA	NA	0.50	NA	NA	0.75	NA	NA

Note. A unit of observation is an airport/city. The outcomes are: in column (1), the number of connected cities 5,500–6,500 miles away; in columns (2) and (3), the number of connected cities at all distances above 2,000 miles away; in columns (3)–(6), at distances 2,000–5,000 miles away; in columns (7)–(9), at distances above 6,500 miles away; and in columns (10)–(12), the total number of passengers in 2014 in millions. Share of cities < 6,000 miles is the centrality-weighted share of cities below 6,000 miles, in the 5,500–6,500 miles range, and network centrality is the standardized eigenvector centrality of connections in 2014. All regressions include the controls: weighted number of cities around 6,000 miles, region fixed effects, time zone difference between airport and GMT, log distance to the equator, and airport controls 1989. Airport controls 1989 include: numbers of daily, twice-weekly, and weekly flights (total and 2,000–5,500 5,500–6,500 and above 6,500 ranges), number of connected countries (twice-weekly), log of total number of connected cities at any flight frequency, total number of seats, log of total number of passengers, log of total number of flights, and network centrality. Additional controls consist of average 1989 night lights and average 1990 population, in the cells within a 100-mile radius of the airport. FS = first stage, RF = reduced form, 2SLS = two-stage least squares. Additional test statistics provided by the Stata ivreg2 command: Anderson-Rubin p-value refers to the weak instrument robust inference on the endogenous regressor using the Anderson-Rubin Wald test (F-stat version) and the Sanderson-Windmeijer F-stat refers to the first-stage test for weak identification of excluded instruments. Robust standard errors in parentheses, clustered at the country level. *** p < .01, ** p < .05, * p < .1 (using weak instrument robust inference when applicable).

(A) <u>Air Links</u>
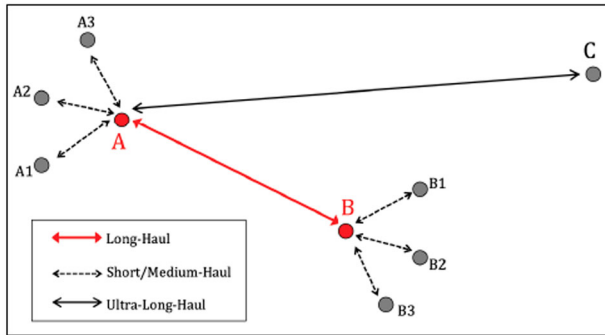
(B) <u>Causal Framework</u>
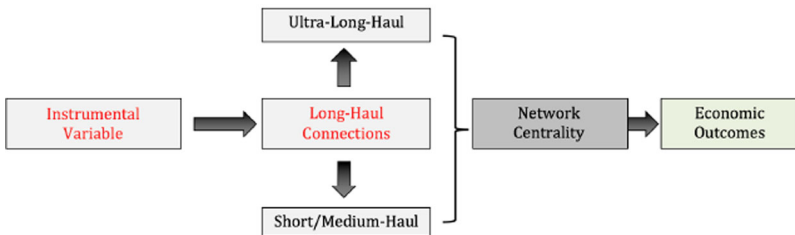
FIGURE VI

Conceptual Framework

The graph illustrates how our instrument may influence connections, directly and indirectly, thereby increasing the network centrality of a city, which in turn may affect economic development in and around the city.

of distance is magnified by the ripple effect that this has over shorter distances, with additional connections inducing yet more additional connections and an increased flow of people. This in turn translates into a substantial increase in the flow of people going through a city. This logic is summarized in Figure VI, Panel A: when airport A gets a long-haul connection to airport B, this offers not only the ability to go directly to B, but also the possibility of reaching other airports from there (B1, B2, B3). As a result, other airports located in a short- or medium-range distance (A1, A2, A3) and potentially farther afield (C) will have an incentive to connect to A. It follows that linking up with a better connected airport has a stronger impact on A's position.

The fundamental causal chain underlying our empirical strategy is thus conveyed in Figure VI, Panel B. Because city pairs that are just under 6,000 miles apart are indeed more likely to be connected than those just over the threshold, airports with a large share of better potential connections just below the threshold have more and better long-haul connections. This in turn makes them more attractive, yielding more and better connections over other ranges, in a process that yields a stronger overall position in the network. Thus, our empirical strategy—which ultimately instruments for network centrality—identifies how a greater position in the global network of air links affects economic activity in and around cities.

V. Air Links and Economic Development

V.A. Baseline Results

We turn to studying the impact of air links on economic activity. Table IV shows results using grid-cell level night lights within a 100-mile radius of any of the 777 airports we are able to match to the night lights data. We start with reduced-form results linking night lights to the weighted share of potential connections just below the 6,000-mile threshold, controlling for the total number and quality within 500 miles of that threshold. Column (1) shows the correlation with a parsimonious set of controls for night lights as measured in 1992. We see no significant correlation yet, suggesting that the increase in long-haul connections unleashed by the introduction of new planes was too recent for there to be a significant effect on economic activity.

Column (2) shows that by 2010, in contrast, a significant correlation had emerged: places close to lucky airports display greater levels of economic activity. Because there is substantial persistence in levels of local economic development, we look at the change in measured night lights between 1992 and 2010 (column (3)), which in essence tests whether the difference between the coefficients in columns (1) and (2) is statistically significant. We find that it clearly is, showing that those places saw larger increases in economic activity over those decades.

Column (4) adds controls for the 1992 level of night lights, population as of 1990, and baseline airport characteristics, in order to account for possible convergence effects and to increase precision. Column (5) further shows that the result is essentially unaltered

TABLE IV

EFFECT ON NIGHT LIGHTS (100-MILE RADIUS), GRID-CELL LEVEL

Dependent variable: Night lights	1992 RF (1)	2010 RF (2)	92–10 change RF (3)	92–10 change RF (4)	92–10 change RF (5)	92–10 change 2SLS (6)	92–10 change, std. 2SLS (7)	92–10 growth RF (8)	92–10 growth RF (9)	92–10 growth 2SLS (10)	92–10 growth, std. 2SLS (11)
Share of cities < 6,000 miles	1.08 (0.84)	2.79** (1.40)	1.71** (0.76)	1.71*** (0.65)	1.77** (0.70)			0.247*** (0.092)	0.227** (0.098)		
Network centrality, 1990–2010						4.10** (1.75)	0.81** (0.34)			0.525** (0.236)	1.01** (0.46)
Anderson-Rubin *p*-value						0.012	0.012			0.022	0.022
Sanderson-Windmeijer *F*-stat						12.98	12.98			12.98	12.98
Airports	777	777	777	777	734	734	734	777	734	734	734
Observations	39,487	39,487	39,487	39,487	37,766	37,766	37,766	39,487	37,766	37,766	37,766
R-squared	0.27	0.30	0.22	0.34	0.35	NA	NA	0.23	0.24	NA	NA
Controls											
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Night light in 1992	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Population in 1990	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Airport controls, 1989	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic and GDP controls	No	No	No	No	Yes	Yes	Yes	No	Yes	Yes	Yes

Note. A unit of observation is a grid cell, within 100 miles of the closest airport in the country. The outcomes are: in column (1), night light in 1992, in column (2), night light in 2010; in columns (3)–(7), the change in night light from 1992 to 2010 (standardized in column (7)); in columns (8)–(11), the growth in night lights from 1992 to 2010 (standardized in column (11)). The independent variables refer to the nearest airport: Share of cities < 6,000 miles is the centrality-weighted share of cities below 6,000 miles, in the 5,500–6,500 miles range; network centrality, 1990–2010 is the standardized average eigenvector centrality during 1990–2010. All regressions include the controls: weighted number of cities around 6,000 miles, region fixed effects, time zone difference between airport and GMT; log distance between the cell and the airport, and log distance to the equator. The successive controls are: night light in the cell in 1992; population in the cell in 1992; airport controls 1989 (numbers of daily, twice-weekly, and weekly flights, number of connected countries (twice-weekly), log total number of connected cities at any frequency, log of total number of seats, log of total number of passengers, log of total number of flights, eigenvalue centrality); geographic and GDP controls (average yearly precipitation 1980–2014, average yearly temperature, 1980–2014, and log real GDP per capita 1990). Additional test statistics provided by the Stata ivreg2 command: Anderson-Rubin *p*-value refers to the weak instrument robust inference on the endogenous regressor using the Anderson-Rubin Wald test (*F*-stat version) and the Sanderson-Windmeijer *F*-stat refers to the first-stage test for weak identification of excluded instruments. Robust standard errors in parentheses, clustered at the country level. ***p <.01, **p <.05, *p <.1 (using weak instrument robust inference when applicable).

if we also include geographical controls and initial GDP at the country level.[32] Note in particular that the coefficient changes very little, and if anything increases in magnitude as we add covariates.

We then turn to 2SLS specifications, scaling the reduced-form results so as to interpret their implications in terms of the impact of network centrality on local economic activity. Column (6) reproduces the full set of controls from Column (5), showing a positive and statistically significant effect of having a better position on the network, as measured by average network centrality over the period. Columns (8)–(10) show that the picture that emerges if we focus on growth rates instead is similar across the board.

To facilitate the interpretation of the quantitative implications of the estimates, columns 7 and 11 present standardized results. An increase of one standard deviation in the average network centrality of an airport would correspond to an increase of roughly 0.8 standard deviations in the change of night lights over the period, or one standard deviation in the growth rate. (Again, for the sake of comparison: that increase in centrality would improve the median airport's position by about 300 spots in the centrality rankings.) For a sense of what this might imply in terms of GDP growth, consider the elasticity around 0.3 between night lights and GDP growth, as estimated by Henderson, Storeygard, and Weil (2012). One standard deviation in the distribution of night lights growth corresponds to 52% over the entire period, or about 15.6% in GDP growth using the aforementioned elasticity. This boils down to about 0.8% in annual GDP growth rates, as this is the annual rate that compounds to 15.6% over the period.

The 2SLS estimates mask considerable heterogeneity in the extent to which the link between the quality of potential connections just below the threshold and the quality of actual connections materializes for different places. In column (6), the first-stage coefficient implies that the impact of a one standard deviation change in *ShareBelow*6*K* on network centrality is 0.432 ($p < .001$), for the overall sample. In contrast, if we focus on the bottom quartile of

---

32. Table A4 in the Online Appendix examines to what extent *ShareBelow*6*K* is correlated with various potential determinants of economic growth at baseline. In general the relationships are weak, with relatively small standardized coefficients, and statistically significant only in the case of population. The results are, again, essentially identical when population is included as a control, which is unsurprising given the weak relationship with our instrument.

airports according to how developed they were in 1992 (as measured by night lights over the 100-mile radius around the airport), we see a very different picture: the first-stage coefficient for that subsample is -0.031 ($p = .382$).[33]

This is consistent with the conceptual framework described in Section III.B, where we noted that a city can be so unattractive in terms of its economic potential that the link between connections and our instrument disappears. In other words, a place like Vientiane gets an excellent draw when it comes to the "lottery" around the 6,000-mile threshold, but that does not translate into more connections. Note that these results do not imply that connections do not have an effect in those less developed locations: the absence of a first-stage link prevents us from reaching any conclusion in that regard.

In sum, we see increased economic activity over the period of analysis in places that are closer to airports that get additional flights induced by the exogenous variation in potential long-haul connections. This indicates a causal impact of air links on economic activity at the local level, but one that is available only to places that were developed enough, to begin with, that they could indeed get connected.

We check the robustness of our main findings in a number of different ways. For brevity, all the results are shown in the Online Appendix. We first experiment with the different implementation of the key source of variation: the results remain very similar if we consider different windows around 6,000 miles, as well as the unweighted instrument, $ShareBelow6K_i^U$ (Online Appendix Table A7). We ask whether our results are reliant on specific places. We have shown that our variation, controlling for the number of cities around the threshold, does not seem particularly concentrated in a given region, and we control for region fixed effects throughout. Still, we go one step further and redo the estimation, dropping each region at a time, in Online Appendix Table A8. The 2SLS results are qualitatively robust, with relatively strong first-stage relationships in all cases, indicating that the results are not driven by any specific region, and thus not any specific country.

We further consider different subsamples in Online Appendix Table A9. We show that the results are unaltered if we leave out,

33. The threshold for the bottom quartile is Abidjan, and the subsample is disproportionately in Sub-Saharan Africa: the region has about 10% of airports in the sample, but 34% of those in the bottom quartile.

for each airport, the cell whose centroid is closest to the airport coordinates or cells within 10 or 20 miles from the airport. This indicates that the effect on night lights is not being driven by the airport itself, as opposed to a broader increase in economic activity. We experiment with alternative thresholds: the results still hold when we choose 150 or 200 or 300 miles instead. The point estimates tend to decrease in magnitude as we expand the area, which is unsurprising given that we expect any positive effects to diminish with distance to the airport. We limit the sample to airports that we verified to have been operational as of 1989, to allay possible concerns regarding the selection of airports, and the results remain (Online Appendix Table A10).[34]

V.B. Spatial Patterns

How far-reaching is the impact of connections on economic activity? We can study this question by exploiting the spatial richness and granularity in the available data. Specifically, we do not have to restrict attention to the immediate vicinity of the airports, but can examine how the impact of additional connections might change as we move away from them.

We start by considering a simple linear interaction specification for the reduced form, where we regress the change in night lights on *ShareBelow*6*K* and its interaction with grid-cell distance to the nearest airport in the country. Based on the previous results, we would expect a positive coefficient for the main effect of *Share-Below*6*K*, indicating the positive effect of potential connections on economic activity around the airport. The interaction, however, could well be negative, as the effect gets weaker with distance.

The results are in Table V and align with that intuition. Column (1) displays the increase in night lights between 1992 and 2010 as the dependent variable, controlling for baseline covariates and the initial levels of night lights and population. It shows a positive effect around the airport, which declines with distance. One concern is that the results might be unduly affected by very remote places, which tend to be located in only a few sparsely populated regions of territorially large countries—not many places will be, say, 1,000 miles from the closest major airport in the country. Column (2) shows that the result is essentially identical, and in fact somewhat stronger, when we restrict the analysis to

---

34. We also show that the first-stage relationship between city pairs, in the RD design, is robust to dropping these airports (Online Appendix Table A11).

TABLE V

EFFECT ON NIGHT LIGHTS, SPATIAL PATTERNS, GRID-CELL LEVEL

	92–10 Change All	92–10 Change <500	92–10 Change <500	92–10 Change <100	92–10 Change 100–200	92–10 Change 200–300	92–10 Change 300–500	92–10 Growth <100	92–10 Growth 100–200	92–10 Growth 200–300	92–10 Growth 300–500
Dependent variable: Night lights Sample (miles from closest airport)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Share of cities < 6,000 miles	5.16**	5.96**	5.65**	1.77**	0.29	−0.08	−0.27	0.23**	0.07	−0.00	−0.06
	(2.20)	(2.40)	(2.51)	(0.70)	(0.25)	(0.19)	(0.17)	(0.10)	(0.07)	(0.05)	(0.04)
Share of cities < 6,000 miles X log distance	−0.92**	−1.10**	−0.99**								
	(0.39)	(0.44)	(0.46)								
Airports	777	777	734	734	439	268	182	734	439	268	182
Observations	229,382	166,012	157,223	37,766	43,904	33,255	42,298	37,766	43,904	33,255	42,298
R-squared	0.34	0.35	0.38	0.35	0.35	0.23	0.14	0.24	0.20	0.10	0.10
Controls											
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Night light in 1992	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Population in 1990	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Airport controls, 1989	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic and GDP controls	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Network centrality, 1990–2010, 2SLS Estimate				4.10**		−0.20	−0.92	0.52**	0.16	−0.00	−0.22
				(1.75)		(0.52)	(0.67)	(0.24)	(0.15)	(0.14)	(0.16)
First stage F-stat				12.98	9.90	7.79	6.62	12.98	9.90	7.79	6.62

Note. A unit of observation is a grid cell. The samples vary depending on the distance from the centroid of the cell to the nearest airport, as denoted above. Log distance refers to the distance between the cell and the airport. All regressions include the main effect of log distance. All other variables and controls are defined the same as in Table IV. The bottom three rows report the corresponding instrumental variable estimation results. Robust standard errors in parentheses, clustered at the country level. ***p < .01, **p < .05, *p < .1.

grid cells at most 500 miles away from the closest airport. Using the same sample, column (3) adds airport and geographic controls, confirming the same message.

The point estimates from columns (2) and (3) imply that the effect turns negative at a distance of around 300 miles, which would seem to suggest that the potential for additional connections could hinder the economic prospects of sufficiently remote places. This simple linear specification, however, obviously imposes that this would be the case at some distance. To better assess the issue of spatial reallocation, we then estimate the effect of distance on the potential impact in a more flexible, semi-parametric way. Specifically, we run the baseline regressions, as in Table IV, repeatedly in a rolling window: first restricting the sample to 0–100 miles from the closest airport, then 5–105 miles, 10–110 miles and so forth up to 300–400 miles.

We plot the results in Figure VII. Panel A displays the first-stage coefficients from regressing the average network centrality on *ShareBelow*6*K*. We see a stable coefficient, but the associated *F*-statistic eventually declines with distance, as airports drop from our sample.[35] Panel B shows a positive reduced-form effect that gets weaker with distance, and becomes a precisely estimated 0 roughly by the 100–200-mile range. Panel C in the figure shows the 2SLS version of the same exercise, with an unsurprisingly similar message.

Although we cannot rule out that there are negative spillovers to more distant cells, and these could be potentially meaningful, the evidence stands against the hypothesis of pure spatial reallocation: the positive effect near the airport cannot be fully accounted for by relocation from the hinterland. This notwithstanding, it is still true that the effect is to exacerbate inequalities over space, as connected cities grow faster relative to their hinterlands.[36]

35. Recall that a given airport will drop out of the sample if there are no cells over a given range for which that airport happens to be the closest one in the country.

36. A full account of the spatial impact would take into account population movements, as the air links may induce some people to move closer to the city. When we look at grid-cell level population, we fail to find evidence of population increases (decreases) closer to (farther away from) the airport. However, we deem this analysis unreliable due to severe measurement error problems in the population data, as discussed in Henderson et al. (2018), precluding us from reaching firm conclusions on that front. (The results are available on request.)
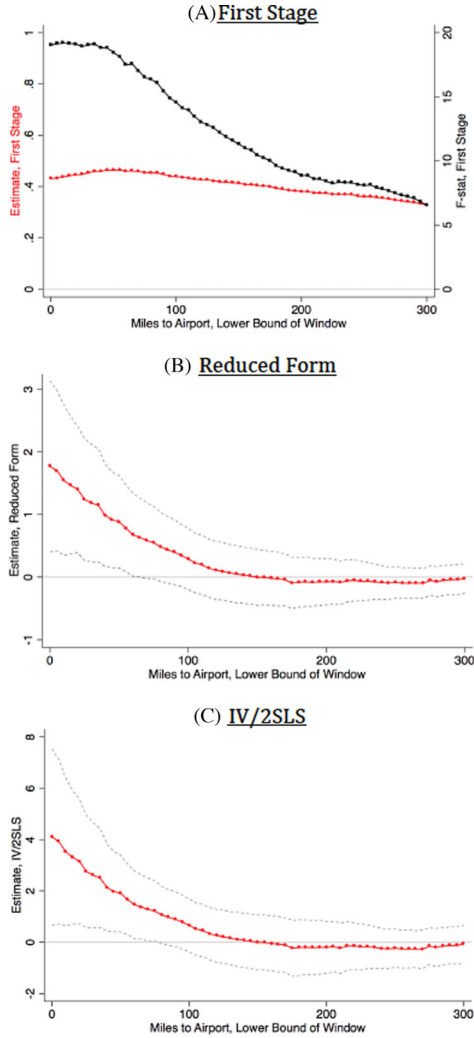
FIGURE VII

Effects on Night Lights by Distance to Airport

The graphs plot estimates from separate regressions using subsamples of cells at different distances from the airport. Each subsample consists of a 100-mile wide donut of cells. The first estimate starts with a lower bound of zero miles to the airport, that is, 0–100 miles, followed by 5–105 miles, 10–110 miles, 300–400 miles. Panel A shows that the first-stage point estimate is relatively stable, but with weaker precision for subsamples further away as the F-stat decreases (the sample size becomes smaller with distance). Panels B and C plot point estimates and 95% confidence intervals, and show that the effects on night lights are positive relatively close to the airport, decreasing in distance, and eventually reach 0.

One potential concern with these spatial patterns is whether spatial autocorrelation in the data leads to incorrect inference. To address this issue, we check robustness with respect to different ways of computing standard errors. First, the results remain qualitatively unaltered if we cluster standard errors at the level of airports. This deals specifically with the fact that the variation we use is at that level. Second, and more important, we use the Conley (1999) approach allowing for spatial correlation across different cells. Because the severity of the spatial correlation is unknown, we allow for relationships to exist at various distances, within 100 miles up to 1,000 miles. Across the board, all the reduced-form results are qualitatively similar and the main conclusions remain (Online Appendix Table A13).[37]

Last but not least, it is important to note that the comparison here is done with respect to distance to the closest major airport in the country. A different but related question pertains to what happens to other major airports in response to shocks hitting a given airport. We address this issue in the context of studying the impact of air links on business links, to which we now turn.

VI. AIR LINKS, BUSINESS LINKS, AND CAPITAL FLOWS

The key differential of air transportation relative to other alternatives, as we have argued, is its implications for the ease of transporting people over long distances—much more so than goods, as trade still flows mostly by sea. It stands to reason that an important part of the impact of air links would likely come from fostering connections between people.[38]

---

37. For completeness, in Online Appendix Table A14 we also show that the first-stage relationship at the airport level is robust to using standard errors adjusting for spatial correlation. Overall, results are very similar. The one exception is that there is some evidence of negative effects on night lights at very long distances from the airport, more than 300 miles away. However, the effective sample size in airports is small in this case (182) and the statistical significance fluctuates depending on the specification. (See Online Appendix Table A13, columns (7) and (11).)

38. It is nevertheless possible that part of the effect could work through a potential impact on trade–both directly, since merchandise is also transported by airplane, and indirectly, since establishing connections between people and businesses could itself impact trade. We do not examine the impact on trade because to the best of our knowledge, there are no data on trade at the level of cities or city pairs.

What are these people bringing—be they locals flying abroad and returning, or outsiders flying in—that could have the substantial effect on economic activity that we have found? We have argued that the ability to interact face-to-face is at the heart of what flight connections make possible.

It seems plausible that the ability to interact face-to-face could be particularly important for business relationships.[39] There are many pieces of circumstantial evidence suggesting that businesses care deeply about access to direct flights. First, there is the effort exerted by airports and policy makers in obtaining such connections, often justified as a way of attracting businesses.[40] Then there is simply revealed preference: nonstop flights typically command a substantial premium over the alternatives. Those flights save time on average, of course, but they also reduce risk: no chance of missed connections, one fewer aircraft to have technical issues, one fewer airport to have logistical issues and so on. In the same vein, businesses tend to locate disproportionately close to airports.[41] Last but not least, there is also growing empirical evidence of the business value of direct flight links (Giroud 2013, Bernstein, Giroud, and Townsend 2016).

As a result, increasing the number and quality of direct air links to a given city could spur the development of connections

39. Another possibility is that the impact we find is driven by leisure tourism flows, but this sector seems too small to justify the sizable impact. The World Travel and Tourism Council, an industry organ, claims that 9.8% of world GDP corresponds to "tourism and travel," but it stands to reason that business travel accounts for a large part of that. It would be interesting to investigate the impact of leisure tourism, but we do not have extensive data on that at the level of cities and city pairs.

40. The president of Alitalia captured the sentiment as he announced his airline's new Italy-China ventures: "China represents a fundamental market for our country, which must aim at growth of Chinese investments and tourism, [and] we believe that the Milan-Shanghai flight ... will establish a very important bridge." (*China Daily*, May 5, 2015)

41. Stilwell and Hansman (2013, 69) show that in the United States, the headquarters of over 50% of Fortune 500 companies, and about 37% of corporate headquarters more broadly, are located within 10 miles of an airport hub–numbers that go to 84% and 66% if we extend the radius to 20 miles. This compares to about 29% of all business establishments and 26% of population within the 10-mile range. As noted by *The Economist* (2005), "with so much emphasis on just-in-time manufacturing and some professionals needing to jump on planes almost daily, airports are becoming the centres of cities of their own," or "aerotropolises" (Kasarda and Lindsay 2011)—and this in spite of obvious drawbacks (noise, traffic, height restrictions on buildings). See also Bel and Fageda (2008).

linking businesses in that city to other businesses elsewhere, which would in turn foster economic activity at the local level, via increased productivity or access to capital.

VI.A. *Air Links and Business Links*

We start off by asking whether connecting two cities has an impact on the links between businesses located in these cities. One straightforward kind of business link relates to foreign direct investment (FDI).[42] It is natural to expect that proximity and face-to-face contact would matter most when FDI involves a majority stake, so we ask whether, given a pair of connected cities, one would see more companies located in one being owned by companies or individuals based on the other.

For that we turn to the Orbis data recording companies with a foreign-based majority owner. We compute, for each company in the data, the distance between the airport in the sample that is closest to its location and the one that is closest to the location of its owners. Figure VIII depicts the total number of firms measured against distance, in 200-mile bins—Panel A with all firms, and Panel B considering only those for which both company and owner are within 100 miles of one of the airports in the sample. We see a substantial drop in the number of ownership links around the 6,000-mile threshold: city pairs just below the threshold have, in total, about twice as many links as those just above it. This naturally suggests the possibility of a causal impact of the availability of direct air links on business connections between cities.

To assess more systematically whether that is the case, we match the Orbis data to all possible airport dyads in the data set. Specifically, focusing again on companies with both parties within 100 miles of one of the airports in the sample, we attribute each company to the corresponding airport dyad: a firm in Shanghai with a majority owner in Milan is attributed to the Shanghai-Milan pair. We add all the companies for each of the nearly 335,000 possible pairs.

---

42. There is evidence that proximity matters for venture capital investors (Bernstein, Giroud, and Townsend 2016) and that migration links matter for FDI across locations in the United States (Burchardi, Chaney, and Hassan 2016). Similarly, the trade literature has studied how distance can affect FDI decisions and offshoring through knowledge flows and direct communication (e.g., Antràs, Garicano, and Rossi-Hansberg 2006; Keller and Yeaple 2013), which would be very affected by the presence of direct flights.

(A) <u>All Firms</u>
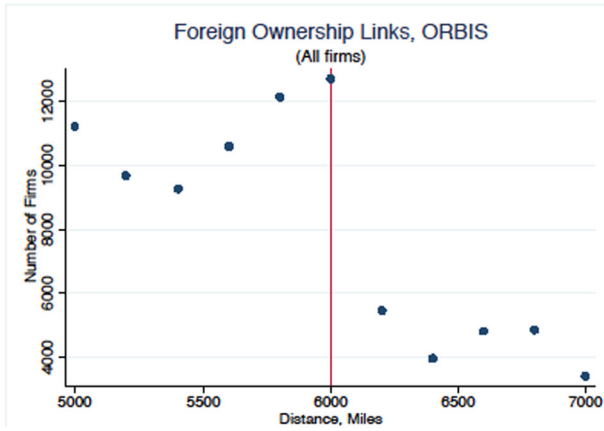
(B) <u>Firms within 100 miles of Airport</u>
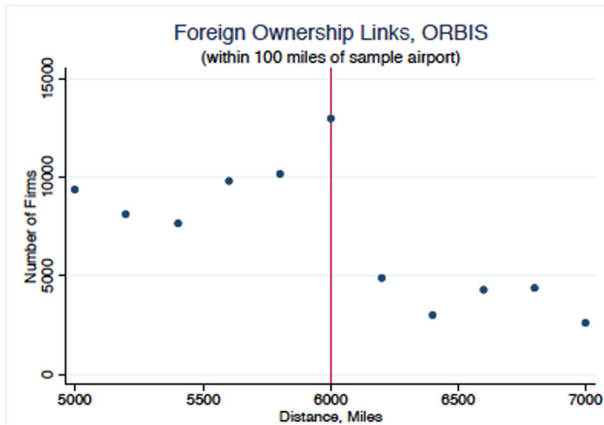
FIGURE VIII

Number of Firms with Cross-Ownership Links, by Distance between Closest Airports

This graph depicts the total number of firms with cross-ownership links as per the Orbis data, according to the distance between the airport in our sample that is closest to the location of the company and the airport in our sample that is closest to the location of the owner. Panel A includes all firms in the data set of georeferenced companies and owners. Panel B restricts the attention to companies that are within 100 miles of one of the 819 airports in the sample. The x-axis bin size is 200 miles. In each bin, the dot represents the number of city pairs in the preceding 200 miles. The graphs show there is a clear discontinuity in the number of cross-ownership links around 6,000 miles.

This allows us to resort again to RD methods to estimate the reduced-form impact of distance around the 6,000-mile threshold on the number of ownership links. The results are in Table VI, where we first consider a sharp RD design to study the "reduced-form" relationship between distance and the number of cross-owned firms in an airport pair. Columns (1)–(6) show a consistent message: there is a significant drop in the number of ownership links upon crossing the threshold, regardless of whether we use different bandwidths, including the optimal bandwidth, include a second-order polynomial, control for baseline covariates, or cluster standard errors by country pair. We also run a test with "placebo" discontinuity thresholds, similar to the one for the discontinuity in flight connections that we showed in Figure V.[43] As can be seen in the Online Appendix (Figure A11), it is once again the case that the estimate at the 6,000-mile threshold is an outlier in the distribution computed for the placebo estimates. Quantitatively, we estimate a drop of around 0.8–0.9 firms comparing the two sides of the discontinuity, which corresponds to about 65% of the average, or 0.04–0.05 standard deviations.

What is the magnitude of the impact we find, in terms of the effects of additional connections? A simple visual comparison gives us a useful benchmark: Figure II shows a drop in the number of connected city pairs, around the 6,000-mile threshold, by a factor of roughly one-third. Figure VIII, in turn, shows a drop in the number of ownership links by a factor of roughly one-third. This suggests that a given increase in connections generates about a similar proportional increase in ownership links.[44] In absolute numbers, this translates into roughly 250 companies for an additional connected pair. This is around the number of ownership links between London and Minneapolis in the data, and an increase of that magnitude is comparable to taking this number to the level of links between London and Malmo (Sweden). This result is confirmed by columns (7) and (8) in Table VI, which exploit a fuzzy RD design where the independent variable of interest is a

43. We also show that the results are robust to using the subsample dropping airports not operational as of 1989, in Online Appendix Table A12.

44. To use more precise numbers: there are 107 connected city pairs between 5,500 and 6,000 miles, and 34 between 6,000 and 6,500 (a factor of 0.32), against 27,964 and 10,229 ownership links (a factor of 0.36). Since $\frac{0.32}{0.36} \approx 0.9$, this means that increasing the number of connected pairs by 10% leads to an increase in ownership links by about 9%.

TABLE VI

EFFECT ON NUMBER OF CROSS-OWNED COMPANIES, AIRPORT-PAIR LEVEL

Dep. var.: Number of cross-owned companies

RD Estimate:	Distance < 6,000 miles						Connected, 2014	
	RF	RF	RF	RF	RF	RF	Fuzzy IV	Fuzzy IV
Bandwidth, miles:	500	1,000	Optimal	Optimal	Optimal	Optimal	Optimal	Optimal
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
RD estimate	0.890	0.922	0.926	0.766	0.987	0.851	248.1	260.4
Robust std. err.	(0.330)***	(0.197)***	(0.221)***	(0.173)***	(0.220)***	(0.230)***	(63.73)***	(80.63)***
Cluster std. err.	(0.432)**	(0.301)***	(0.326)***	(0.256)***	(0.339)***	(0.316)***	(78.17)***	(95.88)***
Polynomial order	1st	1st	1st	1st	2nd	2nd	1st	1st
Baseline covariates	No	No	No	Yes	No	Yes	No	Yes
Observations, total	334,954	334,954	334,954	334,954	334,954	334,954	334,954	334,954
Observations, effective	70,366	134,799	81,842	71,708	130,764	112,703	67,778	69,164
First stage							0.0037***	0.0030***
							(0.0009)	(0.0008)

Note. Local polynomial regression discontinuity estimates, using rdrobust command in Stata (default options unless otherwise stated). The running variable is distance between airports, discontinuity at 6,000 miles. The optimal bandwidth was selected using one common MSE-optimal bandwidth selector. "Polynomial order" refers to the order of the polynomial in distance. The dependent variable is the number of cross-owned companies, the total number of companies within a 100-mile radius of one of the airports in the pair that are owned by individuals/firms located within a 100-mile radius of the other airport in the pair. Connected, 2014 in columns (7) and (8) is a dummy equal to 1 if there were at least weekly flights between the two airports in 2014. The point estimates and the number of effective observations refers to the number of observation within the bandwidth when using the robust standard errors. The covariates are all measured in 1989 and consist of a dummy for having weekly flights, time zone differences, log of total city pairs within the country pair, log total connections within the country pair, and log total passengers within the country pair, where the logged variables add 1 to deal with the undefined log function at 0. Standard errors are in parentheses. "Cluster std. err." refers to cluster-robust nearest neighbor variance estimation at the country-pair level, using a minimum of three nearest neighbors. ***$p < .01$, **$p < .05$, *$p < .1$.

dummy that takes a value of 1 if the airport pair happened to be connected via (at least) weekly flights in 2014.[45]

An alternative way to get a handle on the magnitudes involved is to consider how these ownership links relate to passenger flows. The discontinuity gives rise to about a 0.03 standard deviation increase in passenger flows, which is matched by roughly a 0.04 standard deviation increase in ownership links (see Online Appendix Table A17). Scaled in this manner, a 1 standard deviation increase in passengers between city pairs is associated with slightly more than a 1 standard deviation increase in cross-ownership links.

This pattern extends to other kinds of business interactions beyond ownership, and for that we turn to the GDELT data on geolocated business collaboration events. This has the advantage of going back in time, which will let us exploit the timing of the increased importance of the 6,000-mile discontinuity in air links. We begin by constructing a plot analogous to Figure VIII: we count the number of events where each party is located within 100 miles of airports in our data set and plot the resulting totals against distance, in 200-mile bins. The results are in Figure IX and show a pattern that is very much consistent with the ownership data. The dark dots correspond to the sum of events recorded between 2000 and 2014, and once again they suggest a substantial discontinuity around the 6,000-mile threshold. Interestingly, the white dots depicting the pre-1990 events are very much in contrast with the subsequent period, displaying little sign of a discontinuity. Put simply, we have yet another independent source of data displaying a pattern in line with a causal effect of air links on business connections having evolved since around 1990.

We can then pursue a similar RD-based exercise using the GDELT data. (Full results are left to Online Appendix Table A15 in the interest of brevity.) Not surprisingly in light of Figure VIII, the results mirror the findings using the ownership data, indicating a causal impact of air links. In particular, pairs of cities just below 6,000 miles apart have more instances of business collaboration after 1990, and witnessed a larger increase relative to the pre-1990 period, compared to those pairs just above the threshold.

In sum, we find substantial evidence that establishing direct air links between two cities has a causal impact on the strength

45. Results are essentially identical if we consider weekly connections at some point between 2005 and 2014.
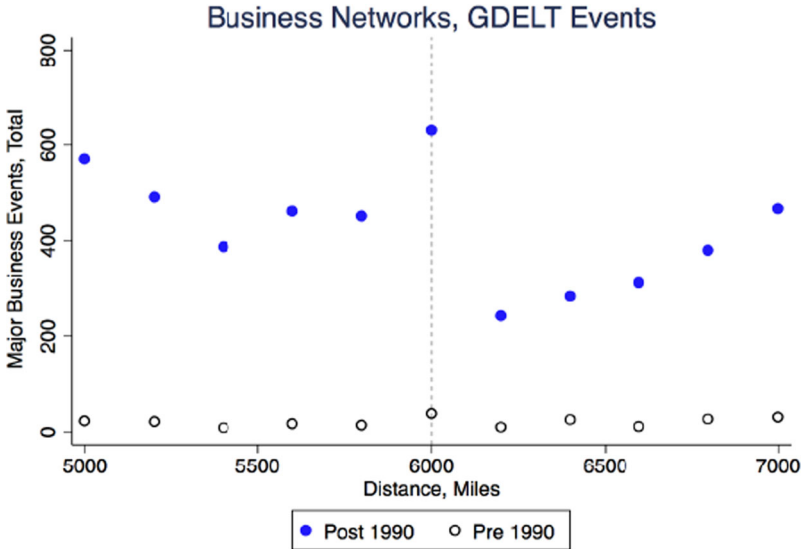
FIGURE IX

Major Business Events (GDELT) Pre- and Post-1990, by Distance between Closest Airports

This graph depicts the total number of major events in the GDELT data involving at least one party (source or target) classified as "business" or "multinational corporation," according to the distance between the airport in our sample that is closest to the source and the airport in our sample that is closest to the target of the event. "Post-1990" refers to events recorded after and including 1990, "Pre-1990" refers to events recorded before 1990. The *x*-axis bin size is 200 miles. In each bin, the dot represents the number of events in the preceding 200 miles. The graphs show there is a clear discontinuity in the number of events linking locations around 6,000 miles.

of the connections between businesses located in each of them, consistent with an enhanced ability to engage in face-to-face interactions fostering those connections.

VI.B. Spillovers across Airports

The evidence above begs the question of to what extent links arising in one city might be displacing those that had or would have been formed with other cities. For example, investors may decide to acquire firms in a certain region or country, but be relatively indifferent with respect to the specific city. In this case, connecting Milan and Shanghai leads to more business links between those two cities, but this might be coming at the expense of

Madrid or Rome.[46] This certainly matters from a policy perspective.

To shed light on this issue, we ask what happens to the business links around a given airport when the closest major airports receive a shock to their position in the network. Specifically, we consider whether two key airport-level variables predict the number of foreign-owned companies located within 100 miles of a given airport. The first is our baseline instrument, *ShareBelow*6*K*, which predicts improvements in the airport's own position in the network. To that we add a second instrument, *ShareBelow*6*K_Closest*10: the average *ShareBelow*6*K* for the ten airports in our sample that are closest to the airport in question, while at least 300 miles away.[47] (The latter constraint ensures that the instruments contain meaningful variation and are not too highly correlated with *ShareBelow*6*K* by construction, and that the airports' regions of influence do not overlap.) This will capture shocks that improve the position of those nearby competitors.

Table VII displays the results, starting off with the first-stage regressions. (All variables are standardized for ease of interpretation.) Column (1) confirms that the instrument *ShareBelow*6*K* is predictive of the airport's network centrality, and Column (2) shows that this remains true when *ShareBelow*6*K_Closest*10 is included. In contrast, the latter is not predictive of the airport's network centrality. The pattern is reversed in Column (3), where the dependent variable is the centrality of the ten closest airports: our second instrument does predict that centrality.[48]

The remainder of the table then puts the two instruments to use. Columns (4)–(6) investigate the effects on business links

46. More broadly, one could imagine that the discontinuity effect in Table VI is driven by investments moving from places just above 6,000 miles away to those just under 6,000 miles. While Figure VIII does not indicate that there is a particularly sharp drop just above 6,000 miles, compared to, say, several hundred miles above that threshold, ultimately one cannot rule out this possibility. We thank an anonymous referee for pointing this out.

47. Results are similar if we look at different numbers of airports, say twenty or twenty-five (available upon request).

48. It could well be the case that the centrality of a given airport would be affected by shocks to nearby competitors, without affecting the validity of the respective instruments. For instance, if an airport becomes more central in the network, airports that are connected to it become more central as well; similarly, in the opposite direction, there could be a business-stealing effect. Our result suggests that these forces are weak, and/or that they cancel each other out.

TABLE VII

SPILLOVER EFFECTS: FIRMS WITH FOREIGN OWNERS < 6,000 MILES AWAY, AIRPORT LEVEL

	First-stage regressions			Reduced form and IV/2SLS regressions								
				Number of firms owned by foreigners, just below 6,000 miles away			Number of firms owned by foreigners, 2,000–6,000 miles away			Night lights growth, 92–10		
Dependent variable, standardized:	Network centrality	Network centrality	Network centrality, 10 Nearest airports									
	FS (1)	FS (2)	FS (3)	RF (4)	RF (5)	2SLS (6)	RF (7)	RF (8)	2SLS (9)	RF (10)	RF (11)	2SLS (12)
Share of cities < 6,000 miles	0.361*** (0.099)	0.364*** (0.112)	−0.041 (0.241)	0.603*** (0.160)	0.706*** (0.170)		0.473*** (0.130)	0.502*** (0.128)		0.537*** (0.204)	0.572*** (0.197)	
Share of cities < 6,000 miles, 10 nearest airports		−0.014 (0.198)	1.632*** (0.355)		−0.481* (0.283)			−0.137 (0.191)			−0.161 (0.378)	
Network centrality						1.909*** (0.641)			1.372*** (0.481)			1.561** (0.696)
Network centrality, 10 nearest airports						−0.279 (0.274)			−0.072 (0.190)			−0.086 (0.305)
Observations	777	777	777	777	777	777	777	777	777	777	777	777
R-squared	0.685	0.685	0.331	0.500	0.502	NA	0.683	0.683	NA	0.345	0.345	NA
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First-stage S-W *F*: network centrality	NA	NA	NA	NA	NA	10.08	NA	NA	10.08	NA	NA	10.08
First-stage S-W *F*: network centrality, 10 nearest	NA	NA	NA	NA	NA	17.20	NA	NA	17.20	NA	NA	17.20

Note. A unit of observation is an airport/city. All regressions include a full set of controls: centrality-weighted number of cities around 6,000 miles, region fixed effects, airport controls 1989, average night lights in 1992 within 100 miles of the airport, average population in 1990 within 100 miles of the airport, time zone difference to GMT, and log distance to the equator. All outcomes are standardized: in columns (1) and (2) the dependent variable is network centrality in 2014; in column (3) the average network centrality of 10 nearest airports; in column (4)–(6) the number of firms within a 100 miles of the airport that are owned by an individual/firm located 5,000–6,000 miles away; in columns (7)–(9), those located 2,000–6,000 miles away; in columns (10)–(12) the total night light growth in cells within 100 miles of the airport. Share of cities < 6,000 miles is defined as before, whereas Share of cities < 6,000 miles, 10 nearest airports is the average value of the 10 nearest airports. To get meaningful variation, the 10 nearest airports are measured among those located at least 300 miles away, for all measures. The firm ownership data is measured in the Orbis database (2016) and the network centrality is defined in the most recent air links data set (2014). The bottom two rows report the Sanderson-Windmeijer *F*-stat from the first stage, for weak identification of excluded instruments. Robust standard errors in parentheses, clustered at the country level. ***$p < .01$, **$p < .05$, *$p < .1$.

with places located just under 6,000 miles away (5,000–6,000 miles). This is where one would expect negative spillover effects among marginal investors to be the strongest, given the RD results showing a sharp increase for pairs just below 6,000 miles apart. The evidence in columns (4)–(6) show that the airport's own network centrality has a positive causal impact on firm ownership. In contrast, the impact of shocks to the centrality of other airports nearby is negative, but smaller in magnitude and statistically insignificant in the 2SLS specification. Note in particular that $ShareBelow6K\_Closest10$ is the average network centrality of the set of ten nearby airports; this means that the impact of a one-standard-deviation increase in the network centrality of a single competing airport, keeping the other airports' constant, would be an order of magnitude smaller than the reported coefficient. Columns (7)–(9) show a similar pattern when we extend the window to consider businesses owned by foreigners located between 2,000 and 6,000 miles away. Columns (10)–(12) then establish the same result when the outcome is economic growth around the airport, as measured by night lights.

Of course, the estimates are not precise enough that we can rule out the presence of meaningful negative spillovers across airports. However, the impact of connections on outcomes at the local level does seem to go beyond mere relocation from nearby airports.[49]

VI.C. Air Links and Convergence: Where Does Capital Flow To?

We can dig deeper into the nature of the business links we study by turning back to the Orbis data and considering, within a given city pair, the direction of each ownership link. In particular, we are interested in whether the increase in cross-ownership is driven by a relatively richer party investing in the relatively poorer one, or vice versa. This seems particularly relevant if we want to understand whether air links foster convergence or divergence across different places.

To study that question, we classify parties in each pair of airports as "richer" or "poorer" according to the relative (PPP-adjusted) income per capita of the country they are in, as of 1990,

---

49. In the Online Appendix, Table A16, we show that all the first-stage and reduced-form results are similar when adjusting the standard errors for spatial correlation. There is again some suggestive evidence of negative spillovers, but one cannot consistently reject the null hypothesis of no effect.

measured by the Penn World Tables (version 8.0). Columns (1)–(4) in Table VIII show that the impact of connections seems to be larger for the number of companies owned by the richer country in the poorer country, rather than vice versa. In fact, if we compare the magnitudes of the two IV estimates, we can conclude that about $\frac{2}{3}$ of the effect on total cross-ownership links comes from capital flowing from rich to poor.

Columns (5) and (6) break it down further by focusing on pairs such that the countries in which they are located are distinctly in an asymmetric position, as measured by the World Bank country classification of income groups (as of 2016): "high income," "upper middle income," "lower middle income," "low income." (This avoids flagging a German firm opening a subsidiary in Norway as an example of capital flowing from the poor to the rich.) The results are very much the same, indicating that the flows are originating largely in that wealthiest tier, and the picture remains essentially unaltered, in columns (7) and (8), when we focus on the smaller subsample in which the richer country is classified as "high income." Although these results should be interpreted with caution—power suffers as we split the sample, and the first stage becomes relatively weak—this nevertheless suggests that the impact of air links on business connections operates as a force for convergence.

Things are rather different, however, when we turn to pairs that include one country classified as "low income": in this case, columns (9) and (10) show no first-stage relationship between distance below 6,000 miles and the existence of a direct connection between the pair (Panel A). This means that we cannot make any statement on whether connections would have had an effect for the poorest countries. Instead, it indicates that receiving a favorable shock to potential connections does not translate into actual connections for those countries, presumably because they are too poor for there to be a demand for connecting in the first place. Put simply, the capital flows in question are essentially taking place between "high income" and "middle income" countries, while countries classified as "low income" are essentially shut out of this process.

VI.D. In Sum

The evidence shows that air links matter for business links: when two cities get connected, there is a substantial increase

TABLE VIII

DIRECTION OF FDI: NUMBER OF COMPANIES OWNED BY CITY IN RICHER AND POORER COUNTRY, AIRPORT-PAIR LEVEL, RDD ESTIMATES

Sample countries:	Pairs: All				Pairs: Different income groups		Pairs: High, nonhigh income		Pairs: Low, nonlow income	
Outcome, direction of the FDI:	Richer → Poorer		Poorer → Richer		Richer → Poorer	Poorer → Richer	Richer → Poorer	Poorer → Richer	Richer → Poorer	Poorer → Richer
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: First stage										
Below 6,000 miles, dummy	0.0041***	0.0033***	0.0041***	0.0029***	0.0017**	0.0017**	0.0020**	0.0020**	0.0006	0.0006
	(0.0011)	(0.0009)	(0.0010)	(0.0011)	(0.0007)	(0.0007)	(0.0009)	(0.0009)	(0.0028)	(0.0028)
Panel B: Reduced form										
Below 6,000 miles, dummy	0.697***	0.625***	0.295***	0.264***	0.218***	0.074*	0.249**	0.088*	0.010	-0.0037*
	(0.251)	(0.187)	(0.067)	(0.083)	(0.077)	(0.039)	(0.101)	(0.050)	(0.017)	(0.0021)
Panel C: IV (fuzzy RD)										
Connected 2014, dummy	172.2***	191.7***	71.27***	90.07***	127.6**	42.69*	126.2**	44.08	N/A	N/A
	(61.36)	(65.50)	(18.13)	(33.83)	(53.77)	(25.33)	(55.86)	(27.51)	N/A	N/A
Polynomial order	1st	1st	1st	1st	1st	1st	1st	1st	1st	1st
Baseline covariates	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth selection, optimal	687	908	794	519	1,052	1,021	924	938	1,000	1,000
Observations, effective	50,374	66,145	57,863	38,244	62,143	60,484	44,457	45,029	17,954	17,954
Observations, total	289,924	289,924	289,924	289,924	192,747	192,747	144,720	144,720	26,135	26,135

Note. Local polynomial regression discontinuity estimates using rdrobust, using triangular kernel function, first-order polynomial in the running variable, and optimal bandwidth (selected using one common MSE-optimal bandwidth selector). The exception to bandwidth selection is columns (9) and (10), where the algoritm is unable to identify the optimal bandwidth; 1,000 miles is used in this case. Running variable is distance between airports, discontinuity at 6,000 miles. The dependent variable is number of cross-owned companies, the number of companies within a 100-mile radius of one of the airports in the pair that are owned by individuals/firms located within a 100-mile radius of the other airport in the pair. "Richer → poorer" refers to companies located in the poorer of the two countries in the pair (as measured by real PPP-adjusted GDP per capita in 1990, as per the Penn World Table 8.0) that are owned by the individuals/firms located in the richer country; and conversely for "Poorer → richer." The sample in columns (5)–(10) is restricted to airport pairs where the two countries are not in the same World Bank income classification: "high," "upper middle," "lower middle," "low." (Argentina is "not classified," but for our purposes we reclassify it as "upper middle.") The firm ownership data is measured in the Orbis database (2016) and the Connected 2014 is taken from the most recent air links data set (2014). Robust standard errors in parentheses. *** $p < .01$, ** $p < .05$, * $p < .1$.

in cross-ownership of companies and in the number of business events involving the two cities, as recorded by news accounts. Consistent with that, cities that are well-placed in terms of obtaining additional long-range air links end up with a greater number of business connections with distant places.

This suggests that the movement of people fosters the movement of capital: the ability to establish face-to-face contact between people is an important factor buttressing the ability to do business. This is in spite of the fact that there is no special technological reason why capital flows should rely on airplanes: one can easily transfer resources and set up businesses at the touch of a button, yet the ability to actually go somewhere induces the establishment of business links.

The evidence shows that this matters over long distances, and that it can translate into a broad economic impulse. This can work as a force for convergence, as the increase in business links is mostly driven by capital flowing from relatively rich countries to middle-income ones. However, this is predicated on the ability to actually connect: the poorest places are left out. As such, perhaps a dearth of connections can be a contributing factor magnifying the relative lack of capital flows from richer to poorer areas (Lucas 1990).

VII. Concluding Remarks

The world is now connected in a global network of air links, through which people can travel back and forth, and thus interact, across long distances as never before. We have found that having more and better connections within this network has a causal impact on economic development: it increases economic activity at the local level, and fosters business links and capital flows.

This naturally leads us to wonder about other possible effects beyond economic activity and the business environment. For instance, more connections could have a direct impact on cultural views and attitudes, which could in turn affect other relevant development outcomes, as well as the potential indirect impact to the extent that those views and attitudes might also be affected by the economic transformations. Would globalization, in this sense, affect political stability, or the prevalence of conflict, or the spread of democracy? These are issues left for future research.

Still on the economic side, the evidence provides a potential rationale for policy interventions designed to increase the number

of connections available from a given airport or city. It is important to keep in mind that while the empirical strategy allows us to make policy recommendations at the local level, substantial challenges remain when it comes to considering their broader impact. We have shown that the local impact cannot be entirely explained by relocation from the hinterland, but we cannot rule out the presence of negative effects on the latter, which in any case also gets left behind in relative terms. We discussed and provided some evidence on spillover effects across major airports, which are also a crucial consideration for policy making at the national level and beyond. The empirical setting is not suited for the full-scale aggregate analysis that would be required for that, which is also an important topic for future research.

Finally, another layer of inequality underlying these results is at the global level, as not all places get to benefit even if they get a lucky draw in terms of potential long-range connections. We have seen that for places that are too poor to begin with, there is no first-stage relationship between the share of potential quality connections just below the 6,000-mile threshold and the actual increase in the quality of connections: it does not matter if a place got lucky in terms of potential connections, if very few would want to fly there anyway. This means that poor places also miss out on the convergence potential induced by increased business links and the capital flows embedded in them. This suggests that while long-range connections can foster development, one has to be in a position to catch that figurative plane. In its aerial dimension, at least, globalization can help some places take off, but others seem to get left behind on the runway.

<div align="center">APPENDIX</div>

Details for Grid-Cell Analysis ([Section III.B](#))

Let $\delta > 0$ denote a relatively small number (compared with 6,000 miles). We start by defining the following sets of cities:

$$\Omega_i^{\delta-} = \{j|6,000 - \delta \leqslant d_{ij} \leqslant 6,000\},$$
$$\Omega_i^{\delta+} = \{j|6,000 < d_{ij} \leqslant 6,000 + \delta\},$$
$$\Omega_i^{\delta} = \Omega_i^{\delta+} \cup \Omega_i^{\delta-}.$$

In words, $\Omega_i^{\delta-}$ and $\Omega_i^{\delta+}$ define, respectively, the set of cities that are just under 6,000 miles away from city i, and the set of cities

just over 6,000 miles away from city i. Ω_i^δ thus defines the set of cities in a δ neighborhood of 6,000 miles away from city i.

With a slight abuse of notation, let $G_i(m; d)$ denote the conditional distribution describing the economic potential of all the pairs involving i. In consonance with the basic identification assumption underlying the regression discontinuity analysis, this function is assumed to be continuous in a neighborhood of the $d = 6,000$ threshold.

Now let N_{ik} denote the total number of cities in Ω_i^k, and K_{ik} denote the number of cities in Ω_i^k that are actually connected to city i via a direct air link. Using equations (2) and (3), we can write:

$$K_{i\delta} = \sum_{j \in \Omega_i^\delta} p_{ij} = N_{i\delta} - \sum_{j \in \Omega_i^{\delta-}} G_i(f_{low} + d_{ij} * c; d_{ij})$$

$$(11) \qquad - \sum_{j \in \Omega_i^{\delta+}} G_i(f_{high} + d_{ij} * c; d_{ij}).$$

Now define $G_i(f_{low} + 6{,}000 * c; 6{,}000) \equiv z_{low}^i$ and $G_i(f_{high} + 6{,}000 * c; 6{,}000) \equiv z_{high}^i$. Because $f_{high} > f_{low}$, and G_i is a probability distribution, it follows that $0 \leqslant z_{low}^i \leqslant z_{high}^i \leqslant 1$. If δ is small, and since $G_i(m; d)$ is continuous in a neighborhood of the $d = 6{,}000$ threshold, we can approximate $G_i(f_{low} + d_{ij}*c; d_{ij})$ with z_{low} in $\Omega_i^{\delta-}$, and approximate $G_i(f_{high} + d_{ij}*c; d_{ij})$ with z_{high} in $\Omega_i^{\delta+}$.

We can thus approximate equation (11) as:

$$(12) \qquad K_{i\delta} \approx N_{i\delta} \left(1 - z_{high}^i\right) + (z_{high}^i - z_{low}^i) * N_{i\delta-},$$

which yields equation (5).

If we instead focus on the economic potential associated with connections, we can define a weighted measure of connections as $\hat{K}_{i\delta} \equiv \sum_{j \in \Omega_i^\delta} p_{ij} * m_{ij}$. If we define $M_{ik} \equiv \sum_{j \in \Omega_i^k} m_{ij}$ to be the total economic potential associated with links to all the cities in Ω_i^k, it immediately follows that:

$$(13) \qquad \hat{K}_{i\delta} \approx M_{i\delta} \left(1 - z_{high}^i\right) + (z_{high}^i - z_{low}^i) * M_{i\delta-},$$

which yields equation (6).

JOHNS HOPKINS UNIVERSITY AND NATIONAL BUREAU OF ECONOMIC RESEARCH
UNIVERSITY OF ZÜRICH

Supplementary Material

An Online Appendix for this article can be found at *The Quarterly Journal of Economics* online. Code used to generate tables and figures in this article can be found in Campante and Yanagizawa-Drott, (2017), in the Harvard Dataverse, doi:10.7910/DVN/O0FPHS.

References

Anderson, John D., *The Airplane: A History of Its Technology* (Reston, VA: American Institute of Aeronautics, 2002).

Antràs, Pol, Luis Garicano, and Esteban Rossi-Hansberg, "Offshoring in a Knowledge Economy," *Quarterly Journal of Economics*, 121 (2006), 31–77.

Bacchetta, Marc, and Marion Janseneds, eds., *Making Globalization Socially Sustainable* (Geneva:ILO and WTO, 2011).

Bel, Germà, and Xavier Fageda, "Getting There Fast: Globalization, Intercontinental Flights and Location of Headquarters," *Journal of Economic Geography*, 8 (2008), 471–495.

Bernstein, Shai, Xavier Giroud, and Richard R. Townsend, "The Impact of Venture Capital Monitoring," *Journal of Finance*, 71 (2016), 1591–1622.

Bleakley, Hoyt, and Jeffrey Lin, "Portage and Path Dependence," *Quarterly Journal of Economics*, 127 (2012), 587–644.

Bloch, Francis, Matthew O. Jackson, and Pietro Tebaldi, "Centrality Measures in Networks," SSRN Working Paper, 2017.

Bonacich, Phillip, "Factoring and Weighting Approaches to Status Scores and Clique Identification," *Journal of Mathematical Sociology*, 2 (1972), 113–120.

Bourguignon, François, *The Globalization of Inequality* (Princeton, NJ: Princeton University Press, 2015).

Brueckner, Jan K., "Airline Traffic and Urban Economic Development," *Urban Studies*, 40 (2003), 1455–1469.

Burchardi, Konrad, Thomas Chaney, and Tarek A. Hassan, "Migrants, Ancestors, and Foreign Investments," NBER Working Paper 21847, 2016.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik, "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 82 (2014), 2295–2326.

Campante, Filipe, and David Yanigazawa-Drott, "Replication Code for: 'Long-Range Growth: Economic Development in the Global Network of Air Links'," *Harvard Dataverse* (2017), doi:10.7910/DVN/O0FPHS.

Catalini, Christian, Christian Fons-Rosen, and Patrick Gaule, "Did Cheaper Flights Change the Direction of Science?," *CEPR Discussion Paper 11252*, 2016.

Clark, Paul, *Buying the Big Jets* (Aldershot, UK: Ashgate, 2002).

Collier, Paul, *The Bottom Billion: Why the Poorest Countries Are Failing and What Can Be Done About It* (Oxford: Oxford University Press, 2007).

Conley, T.G., "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, 92 (1999), 1–45.

Cristea, Anca D., "Buyer-Seller Relationships in International Trade: Evidence from U.S. States' Exports and Business-Class Travel," *Journal of International Economics*, 84 (2011), 207–220.

Dollar, David, "Globalization, Poverty, and Inequality since 1980," *World Bank Research Observer*, 20 (2005), 145–175.

Dollar, David, and Aart Kraay, "Trade, Growth, and Poverty," *Economic Journal*, 114 (2004), F22–F49.

Donaldson, Dave, "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review* (forthcoming).

Donaldson, Dave, and Richard Hornbeck, "Railroads and American Economic Growth: A 'Market Access' Approach," *Quarterly Journal of Economics*, 131 (2016), 799–858.

Dreher, Axel, "Does Globalization Affect Growth? Evidence from a New Index of Globalization," *Applied Economics*, 38 (2006), 1091–1110.

Federal Aviation Administration, "Economic Values for FAA Investment and Regulatory Decisions, A Guide—Final Report," 2016, https://www.faa.gov/regulations_policies/policy_guidance/benefit_cost/.

Feyrer, James, "Trade and Income: Exploiting Time Series in Geography," NBER Working Paper no. 14910, 2009.

Frankel, Jeffrey, and David Romer, "Does Trade Cause Growth?," *American Economic Review*, 89 (1999), 379–399.

Giroud, Xavier, "Proximity and Investment: Evidence from Plant-Level Data," *Quarterly Journal of Economics*, 128 (2013), 861–915.

Glaeser, Edward L., *The Triumph of the City* (New York: Macmillan, 2011).

Green, Richard K., "Airports and Economic Development," *Real Estate Economics*, 35 (2007), 91–112.

Henderson, J. Vernon, Tim Squires, Adam Storeygard, and David N. Weil, "The Global Spatial Distribution of Economic Activity: Nature, History, and the Role of Trade," *Quarterly Journal of Economics*, 133 (2018), 357–406.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil, "Measuring Economic Growth from Outer Space," *American Economic Review*, 102 (2012), 994–1028.

Hummels, David, "Transportation Costs and International Trade in the Second Era of Globalization," *Journal of Economic Perspectives*, 21 (2007), 131–154.

Imbens, Guido, and Thomas Lemieux, "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, 142 (2008), 615–635.

Kasarda, John D., and Greg Lindsay, *Aerotropolis: The Way We'll Live Next* (New York: Farrar, Straus and Giroux, 2011).

Keller, Wolfgang, and Stephen R. Yeaple, "The Gravity of Knowledge," *American Economic Review*, 103 (2013), 1414–1444.

Kharina, Anastasia, and Daniel Rutherford, "Fuel Efficiency Trends for New Commercial Jet Aircraft: 1960 to 2014," *International Council on Clean Transportation White Paper*, 2015).

Leetaru, Kalev, and Philip A. Schrodt, "GDELT: Global Data on Events, Location and Tone, 1979–2012," International Studies Association Annual Conference, 2013.

Lowe, Matt, "Night Lights and ArcGIS: A Brief Guide," MIT (unpublished), 2014.

Lucas, Robert E., Jr, "Why Doesn't Capital Flow from Rich to Poor Countries?," *American Economic Review*, 80 (1990), 92–96.

Manacorda, Marco, and Andrea Tesei, "Liberation Technology: Mobile Phones and Political Mobilization in Africa," CEP Discussion Paper no. 1419, 2016.

McKenney, David, Ken Prusak, Lenny Robichaux, and Wally Tweeden, "United Airlines Crew Augmentation Study," 2000, https://www.eurocockpit.be/sites/default/files/Crew-Augmentation-Study-2000.pdf

Michalopoulos, Stelios, and Elias Papaioannou, "Pre-colonial Ethnic Institutions and Contemporary African Development," *Econometrica*, 81 (2013), 113–152.

Mukkala, Kirsi, and Hannu Tervo, "Air Transportation and Regional Growth: Which Way Does the Causality Run?," *Environment and Planning A*, 45 (2013), 1508–1520.

Ortega, Francesc, and Giovanni Peri, "Openness and Income: The Roles of Trade and Migration," *Journal of International Economics*, 92 (2014), 231–251.

Pascali, Luigi, "The Wind of Change: Maritime Technology, Trade and Economic Development," *American Economic Review*, 107 (2017), 2821–2854.

Poole, Jennifer P., "Business Travel as an Input to International Trade," UC Santa Cruz (unpublished), 2013.

Proctor, Jon, Mike Machat, and Craig Kodera, *From Props to Jets: Commercial Aviation's Transition to the Jet Age 1952–1962* (Forest Lake, MN: Specialty Pr Pub & Wholesalers, 2010).

Redding, Stephen, Daniel Sturm, and Nikolaus Wolf, "History and Industrial Location: Evidence from German Airports," *Review of Economics and Statistics*, 93 (2011), 814–831.

Stein, Ernesto, and Christian Daude, "Longitude Matters: Time Zones and the Location of Foreign Direct Investment," *Journal of International Economics*, 71 (2007), 96–112.

Stilwell, Justin D., and R. John Hansman, "The Importance of Air Transportation to the U.S. Economy: Analysis of Industry Use and Proximity to Airports," *MIT International Center for Air Transportation (ICAT) Report ICAT-2013-03*, 2013.

Storper, Michael, and Anthony J. Venables, "Buzz: Face-to-Face Contact and the Urban Economy," *Journal of Economic Geography*, 4 (2004), 351–370.

The Economist, "Business on the Fly: How Airports Act as a Hub for Business Development," November 24, 2005.

The Economist, "Do Hub Airports Boost Growth?," December 1, 2015.

Yilmazkuday, Demen, and Hakan Yilmazkuday, "The Role of Direct Flights in Trade Costs," Federal Reserve Bank of Dallas Globalization and Monetary Policy Institute Working Paper no. 179, 2014.