

Long-Range Growth: Economic Development in the Global Network of Air Links*

Filipe Campante[†] David Yanagizawa-Drott[‡]

September 2016

Abstract

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 give rise to a discontinuity in connectedness between cities at a distance of 6000 miles. We show that these air links have a positive effect on local economic activity, as captured by satellite-measured night lights. To shed light on how air links shape economic outcomes, we first present evidence of positive externalities in the global network of air links: connections induce further connections. We then 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, our results suggest that increasing interconnectedness generates economic activity at the local level by inducing links between businesses, but also gives rise to increased spatial inequality locally, and potentially globally.

Keywords: Globalization; Air Travel; Connections; Economic Activity; Local Development; Cities; Business Links; FDI; Convergence; Spatial Inequality.

JEL Codes: F15, F21, F23, F63, O11, O18, O19, O47, R11, R12, R40.

*We thank Abhijit Banerjee, Leo Bursztyn, Fred Finan, Ed Glaeser, Tarek Hassan, Ben Olken, Gautam Rao, Andrei Shleifer, and Noam Yuchtman for helpful conversations, seminar participants at Bocconi, Columbia, Harvard, the 2016 AEA meeting (San Francisco), the Harvard Cities Mini-Conference, and the Barcelona GSE Political Economy of Conflict and Development workshop for valuable comments, Laura Alfaro, Michele Coscia, and Irfan Mahmud for their generous help with data collection, and Jonah Rexer for truly outstanding research assistance. We gratefully acknowledge the financial support provided by the Weatherhead Center for International Affairs at Harvard, the Harvard Kennedy School Dean's Research Fund, and the Taubman Center for State and Local Government. All errors are our own.

[†]Harvard Kennedy School and NBER. Email: filipe_campante@harvard.edu

[‡]University of Zürich, UBS Center of Economics in Society. Email: david.yanagizawa-drott@econ.uzh.ch

1 Introduction

It is often said, to the point of being cliché, that we are living through an age of globalization. One key aspect of that is certainly a marked reduction in transportation costs of goods, of information, of people: a globalized world is an interconnected world, in all these dimensions. However, understood as such globalization is obviously far from new. Previous episodes of global integration, such as in the late 19th century, also witnessed a steep reduction in transportation costs over long distances, as steamships and railways sharply increased trade and migration across borders.

The current episode is unique, however, in that it is now far cheaper and faster than ever to transport people, and this has made it possible to travel back and forth between distant places. 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 of air travel, enabling a constant flow of people between countries and continents far apart.

This paper studies the impact of direct long-distance air links, to present the first evidence of causal effects of that transformation on economic development. A long intellectual tradition has posited that proximity – and 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. While people have been able to move from Shanghai to London or New York for a long time, and goods have been moving for just as long, now people can, unlike ever before, 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, which is a lot, but Shanghai added 34. Over that period, Jakarta grew substantially, but Shanghai grew substantially more.

Yet 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 in 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 6000 miles apart are distinctly more likely to have direct air links, as compared to 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 6000 miles – a little over what separates London from Shanghai, or Istanbul from Jakarta. On the technological side, the introduction of two landmark long-range airplane models (Boeing’s 747-400, in 1989, and Airbus’s A330 and A340, in 1993-94) 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. We implement this strategy on two levels: first, we compare pairs of cities that are just below 6000 miles apart (such as Shanghai and Milan, 5650 miles) to pairs that are just above (Shanghai and Madrid, 6350). Second, we 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, as Jakarta.¹ Since places and pairs arguably do not differ systematically across the 6000-mile threshold, we have a source of exogenous variation in the number of available air links. We show that this variation is not correlated with outcomes as of circa 1989, when the discontinuity was weaker, underscoring its plausible exogeneity.

Using this strategy, we first show a set of basic, “first-stage” results: pairs of places like Shanghai and Milan (connected by a non-stop flight since 2003) are indeed more likely to be connected than pairs like Shanghai and Madrid (no non-stop flight before 2016).² Similarly, places with a greater share of potential links just below the threshold indeed have a larger number of connections.

This sets the stage for our key question: do these connections matter for economic development? We show that they do. First of all, using granular data at the level of grid cells, we find that places close to airports with a larger share of potential connections just below the 6000-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: additional connections explain about 6% of the growth in night lights over the period, for the median cell in our sample.

We exploit the spatial richness of the data to shed additional light on the nature of that effect. First, we show that it is not driven by pure spatial reallocation of economic activity: while the

¹Specifically, over three out of four major international airports in our sample which are between 5500 and 6500 miles from Shanghai happen to be below 6000 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 6000 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.

²As we will discuss, there is reason to believe that the 6000-mile discontinuity is in the process of disappearing, after regulatory changes implemented starting in 2014.

positive effect dissipates with distance to the airport, as expected, we find no evidence of a negative net effect at longer distances. This means that connections induce spatial inequality, as the places that get connected grow more, but they create economic activity, rather than merely shifting it across space. On the other hand, the increase in economic activity is not matched by increased population: we find essentially no effect on population at any distance from the airport, suggesting that increased productivity may be translating into economic rents.

We then study how long-distance air links shape economic outcomes and development. Most directly, we first show that a shock to air links over a specific distance range gets magnified because connections induce further connections: there is a spillover to shorter distances and an increase in the overall quality of air links, as additional long-haul links increase a city's desirability for other connections, as well as an increase in the total flow of passengers. These spillovers also show up across airports in the network: we find a positive "medium-range" effect (2000 to 5500 miles) but no evidence of significant effects at shorter distances, as positive and negative spillovers seem to balance out.

We then focus 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 [sic] do not mind having to 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 we 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, our data allows us to find over 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, yet again we find a discontinuity right at the 6000-mile threshold – pairs of cities just below 6000 miles apart have substantially more business cross-ownership links. From this we estimate an impact that implies that a given increase in connections generates about a similar proportional increase in ownership links. In addition, the evidence suggests that most of this increase constitutes

³For 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 non-stop 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, Stilwell and Hansman 2013, Kasarda 2016). Last but not least, there is also growing empirical evidence of the business value of direct flight links (Giroud 2013, Bernstein, Giroud, and Townsend forth.).

capital flowing from relatively richer to relatively poorer countries: 3/4 of the increase in business connections could be attributed to companies in high-income countries owning companies in middle-income ones, and 1/4 in the opposite direction.

As a result, cities with a large share of potential connections just below the threshold end up with more foreign-owned companies and owning more companies abroad. We also show that, consistent with the pattern found for the network of flights, the increased cross-ownership links arise precisely over a range of distances just below 6000 miles, with some spillovers in shorter distances and no impact above that threshold. We also find similar patterns using alternative measures of business connections, based on geolocated business-related events mentioned in news reports worldwide, as well as evidence that connections increase the likelihood of presence of large-company headquarters.

In sum, our evidence suggests that increasing the number of direct connections between different places has a significant impact on development, as it increases economic activity at the local level. This seems to be driven by 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 why capital would need airplanes to move around. This suggests that policy interventions designed to increase the number of connections could potentially spur development, though at the price of increased spatial inequalities at the local level.

Yet our results also highlight the potential for affecting inequality on a larger scale. In particular, our first-stage relationship linking potential long-haul connections just below 6000 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 get shut out of the increase in business links and capital flows, suggesting 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, globalization, in its long-range air dimension, has helped the Shanghais of the world achieve convergence, but also increased the distance between them and the Vientianes. Whether the overall effect increases inequality or decreases it depends on which of these two forces is seen as more important from that perspective.

Our paper 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; Hummels 2007; Bacchetta and Jensen 2011; Ortega and Peri 2014). In particular, some of the work in that vein have looked at the effect of transportation technologies, such as steamships (Pascali 2015), railroads (Donaldson (forth.); Donaldson and Hornbeck 2016), 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 else 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. By doing so, we also 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 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 literature has looked into the connection (e.g. Brueckner 2003; Green 2007; Mikkala 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 post-war 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, Bersntein, Giroud, and Townsend 2016) or science (Catalini, Fons-Rosen, and Gaule 2016), using the introduction of air links in the US as a source of variation, and also on trade (Cristea 2011; Poole 2013; Yilmazkuday and Yilmazkuday 2014). We differ in that our approach allows for causal identification at a global level, and for studying both the impact on economic activity and the potential channels via business links.

The paper is organized as follows. Section 2 provides background on the recent evolution of long-haul air travel, to lay out the foundations of our identification strategy. Section 3 describes the data and specifications that implement that strategy, as well as establishing the presence and effect of our discontinuity. Section 4 contains the results establishing the key effects on economic activity. Sections 5 and 6 focus on mechanisms: network spillovers and the flow of people, and business links. Section 7 concludes.

2 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 Koderer 2010). The technological evolution of commercial airplanes (Anderson 2002, ch.7) enabled greater and greater distances to be covered: from the Boeing 707, which started flying transatlantic routes in 1958, to the Boeing 747 (aka “Jumbo Jet”), which enabled, for instance, the route between San Francisco and Sydney which, at just under 7500 miles, became in 1976 the longest regularly scheduled non-stop 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 definition singles out flights that take longer than 12 hours. Given customary speeds, a 12-hour flight translates into about 6000 miles, corresponding to the distance between London or Paris and Tokyo.⁴ The distinction is apparent in the range of modern commercial aircraft by Airbus and Boeing: there is a clear distinction between aircraft designed to fly up to 4000 nautical miles (about 4600 miles), and those that fly at least 6000 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.⁵ 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 US 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 three (or more) pilots, as well as an additional flight crew member, and “adequate sleeping quarters” on the plane (Code of Federal Regulations, §121.485). 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 post-flight 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).⁶ This type of regulation entails that ULH flights are discontinuously costly, and significantly so given that personnel constitutes a

⁴Specifically, the air distance between Heathrow and Narita airports is of 5966 miles (as per www.airmilescalculator.com), and non-stop flight hours are estimated at 11:40 (as per a simple Google flight search). By contrast, from Chicago O’Hare to Narita takes 6267 miles, and an estimated 13:15. More generally, while we do not have information on flight duration in our 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 6000 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.

⁵More broadly, the range of a given plane is not exactly a clear-cut number: there is a trade-off between so-called “payload” (namely, 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 in order to have more fuel. (For a discussion on this, see Clark (2012, ch.5).) In any case, given the capabilities of modern aircraft, range is no longer a dominant concern in airline fleet selection (Clark 2012, p.31): the vast majority of city-pairs that could sustain direct connections has been covered by existing aircraft since the 1970s.

⁶See <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, US regulations also have an impact elsewhere: for instance, in 1992 the Indian regulator introduced its own rules (AIC 28) essentially adopting US standards. 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

substantial share of the costs of a flight.⁷

How important is this discontinuity in practice? Using our sample of cities with major international airports from the International Civil Aviation Organization, which are shown in the map in Figure 1, we can calculate the distance between all city pairs as well as the number of non-stop flights between cities. (We describe the data in detail in Section 3.)

[FIGURE 1 HERE]

The discontinuity manifests itself clearly in Figure 2 (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 in the figure corresponds to the number of pairs within a 200-mile bin in terms of distance. We can see that, while the number of flights unsurprisingly declines with distance, there is a substantial drop right at 6000 miles: there are substantially more connected city pairs in which the two cities sit at a distance between 5800 and 6000 miles, say, than is the case for pairs situated between 6000 and 6200 miles apart.

[FIGURE 2 HERE]

This discontinuity, however, has not always been quite this pronounced. In fact, Panel B in Figure 2 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 is 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 above 4600 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) around the same distance.

What explains this pattern? As it turns out, the late 1980s and early 1990s witnessed important shifts in the long-haul civil aviation landscape, both in terms of technology and market structure. The Boeing 747-400, which started commercial operations in 1989, was able to fly about 1000 miles longer than previous commercial airliners (up to almost 9000 miles). By 1990 there had already been over 100 units delivered, and the model went on to become the best-selling subset in the 747 family, with nearly 700 units sold over 20 years.⁸ A few years later, in 1993-94, Airbus

⁷For instance, personnel salaries represents about 20% of the costs of a typical US domestic flight (Wall St. Journal 2012), second only to fuel. Unlike the latter, it is also a fixed cost, since a relatively empty plane can fly with less fuel, but still requires the same number of pilots and crew. On top of that, additional crew and sleeping quarters imply less space and weight available for carrying payload, thus reducing revenue potential.

⁸A full list is available at <https://www.planespotters.net/production-list/Boeing/747/747-400>.

introduced its A330 (up to 7300 miles) and A340 (up to 8400 miles) models, which made the company into a serious competitor for Boeing. The combination of these very successful models, coupled with the increased market competition, 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 2014.⁹

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 the Online Appendix (Figure A2). The number of long-haul flights (above 4500 miles) goes up sharply right after 1989, and this is largely pushed by the range below 6000 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-6000 range as well.¹⁰

There is reason to believe that this discontinuity may be in the process of disappearing. Both US and European regulators have adopted major revisions to flight time limit regulations in 2014 (known 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).¹¹

As it happens, the two-plus decades over which this discontinuity has existed provide us with a unique window to identify the causal impact of long-haul connections. To the extent that cities that, as of the late 1980s, happened to have many airports lying just below 6000 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.

3 Empirical Framework

Let us now describe how we implement this idea in practice. We start off with our key data sources, and then discuss the empirical specifications we will use. Finally, we will establish more formally the presence and impact of the 6000-mile discontinuity.

⁹Interestingly, 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. For instance, as of 2014, we see in the range between 5500 and 6000 miles (resp. 6000 to 6500): 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).

¹⁰The 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 upon availability.

¹¹See <https://www.law.cornell.edu/cfr/text/14/part-117> and https://www.eurocockpit.be/sites/default/files/ftl_commission_regulation_83_2014.pdf

3.1 Data

3.1.1 Air Links

Our key source of data comes from the International Civil Aviation Organization (ICAO). Specifically, we use the “Traffic by Flight Stage” (TFS) dataset, which gives us, 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, the number of flights operated, and the traffic (passengers, freight and mail) carried. This dataset contains information on city and country names, which we use to merge with information on the coordinates of major international airports, from the “Airport Traffic” dataset.¹²

We are left with 819 cities with major international airports, from 200 different countries, which are shown in the map in Figure 1. (Descriptive statistics can be found in Table A1 in the Online Appendix.) 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, as 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 our 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 6000-mile threshold, we would expect there to be a sharp discontinuity in the number of airport pairs at that point, which Panel D in Figure 2 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. Our baseline analysis defines two cities as being connected if they have at least weekly flights between them.¹³ We can either consider the snapshot of whether there is a connection in a specific year, or aggregate the information over a multi-year 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 will 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.¹⁴

¹²For cities with more than one airport, we use the average coordinates of all airports in question. While keeping that in mind, for simplicity we will use “airports” and “cities” interchangeably.

¹³We define a weekly connection as having at least 52 flights back and forth in a year. We will 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.

¹⁴Specifically, 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

We can also aggregate the information to the level of cities. For each one of the cities in our sample, we calculate the number of other cities to which it is connected in a given year, and also the aggregate number of connection-years over a longer period. Similarly, we also compute the total number of flights to/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 2000 miles, in order to concentrate on the range over which airplanes are essentially the only relevant means of transporting people.¹⁵

We can also summarize the quality of those links, with a measure of network centrality. We will focus on eigenvector centrality, which is constructed as follows: we first describe the structure of the network of air links with the “adjacency matrix” A , in which each entry a_{ij} takes a value of 1 if cities i and j are connected, and 0 otherwise. We take the eigenvector associated with the greatest eigenvalue of A , and the n th component of that eigenvector corresponds to the centrality measure of city n . Intuitively, this procedure 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.

3.1.2 Economic Activity

To capture economic activity at the local level, on a global scale, we use the by 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 papers (Henderson, Storeygard, and Weil 2012; Bleakley and Lin 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 x 0.25 degrees. We focus on growth over the two decades following the introduction of the Boeing 747-400 (and the start of our sample of air links), so we compute average nights lights in the cell for 1992 (first year available) and 2010. In the Online Appendix (Figures A3-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.

We will also use data on population at the local level. This comes from the Gridded Population of the World (GPW) version 4, which we obtain from the PRIO-GRID website. We use the first and last available years (1990 and 2010), and also grid-cell level variables to use as controls

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.

¹⁵Results are very similar if we use 3000 miles as the threshold instead.

(precipitation, temperature).

3.1.3 Business Links

In order to shed light on the potential effects on potential mechanisms behind the effect of air links of economic activity, we will look at their impact on business links over long distances. For that, we make use of three datasets with spatial information and global coverage.

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 over 195 million firms in 229 countries. Our sample consists of all of the one and two-way business ownership links between cities located in different countries and 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). Here we define the GUO of a given firm as any company that 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.¹⁶

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.¹⁷

Since the Orbis online database is continuously updated, our 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.

We again map the distribution of ownership links in space, as shown in the Online Appendix (Figures A5-A7). 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

¹⁶We 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.

¹⁷Specifically, 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 sub-national unit is used.

companies located abroad that are owned by individuals or firms located in that grid cell.¹⁸ 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.

Major Business Events. We use the GDELT dataset, which identifies, classifies, and geolocates events mentioned in broadcast, print, and web news reports worldwide. (For a more detailed description, see Leetaru and Schrodtt 2013, or Manacorda and Tesei 2016.) In particular, for each event there are two actors (“source” and “target”), with latitude and longitude coordinates for each of them. We restrict our analysis to events that capture “material cooperation,” where at least one actor is a business entity (i.e. classified as BUS, “business,” or MNC, “multinational corporation”), and where there were at least two separate articles reporting about the event. To match across datasets, we collapse the spatial data to cell level (0.25 x 0.25 degrees), calculating the sum of observations in each cell. The data are available between 1979 and 2014, which we will aggregate into pre-2000 (1979-1999) and post-2000 sub-periods. We map the post-2000 events 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 is consistent with the fact that they represent weaker links between two businesses, and thus capture a different dimension of business interaction across distances.

Headquarter Locations. We look at the location of corporate headquarters of companies in the Forbes “Global 2000” list. This list compiles the 2000 largest publicly traded companies, and the headquarters data is obtained from the World City Relational Data, made available by the Globalization and World Cities (GaWC) Research Network. Specifically, we use their Data Set 26 (Global Command and Control Centres), with information for 2006, 2009, and 2012, based on the 2010 edition of the Forbes list. The companies are spread over 386 cities in 2006, 416 cities in 2009, and 433 cities in 2012.¹⁹

3.2 Identification Strategy and Specifications

In order to identify a causal effect of air links, we will rely on the discontinuity in the likelihood of links as a function of the distance between two cities, at 6000 miles. We will now discuss how we will use it to implement our empirical strategy, which we will break down, depending on the

¹⁸We show pictures for the full set of foreign-owned firms, as well as for the subset of firms located at least 3000 miles away from their owners. and again the case of Asia for greater detail.

¹⁹The data excludes corporations with headquarters registered in the Cayman Islands, Bermuda and Liechtenstein. It also assigns companies located in small towns that belong to a metropolitan area to the largest city in that area.

nature of our outcomes of interest, into two distinct levels of analysis: city pairs and grid cells. We will discuss these in order.

3.2.1 City-Pair Analysis

As Figure 2 indicates, our identification strategy is based on the idea that the likelihood that two cities get connected depends crucially on whether their bilateral distance happens to be just above or just below 6000 miles. Since this variation is arguably as good as randomly assigned, this strategy allows us to test whether long-haul flights affect economic activity in, and between, cities. To test that more formally, we perform a regression discontinuity analysis at the city-pair level:

$$(1) \quad Y_{ij} = \alpha + \beta * Below6K_{ij} + f(Distance_{ij}) * \gamma + \varepsilon_{ij},$$

where ij denotes a city-pair, Y_{ij} is an outcome of interest, $Distance_{ij}$ is the distance in miles between the two cities (i.e. airports), and $Below6K_{ij}$ is a dummy equal to one if $Distance_{ij}$ is less than 6000 miles. It is well known that higher order polynomials in $f()$ can result in approximation errors due to over-fitting or biases at boundary points, so in the baseline specification we use a parsimonious specification allowing for different linear slopes above and below the 6000 mile threshold. We also 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).²⁰ 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, 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 6000-mile threshold is meaningful, we expect β to be positive. We then 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 will present scaled instrumental variable estimates – the marginal effect of getting connected on outcomes – using a ‘fuzzy’ regression discontinuity approach.

²⁰This is implemented in Stata using the *rdrobust* routine.

3.2.2 Grid-Cell Analysis

To test whether connections affect economic activity, we expand the analysis beyond city-pair outcomes and use data at the grid-cell level. The key identification assumption in this case boils down to the idea that there is no reason why airports that happen to have relatively many major airports sitting just under 6000 miles away should be systematically different from airports that happen to have many just above that threshold. If that is the case, for any given airport in our sample we can take the number of airports within a narrow window below 6000 miles – say, 500 miles – accounting for the number of airports within a similar window above the threshold, as a source of variation in the number of long-haul connections that this airport will actually have.²¹

To implement this logic, we define our baseline instrument, *ShareBelow6K_i*, as the number of airports 5500 to 6000 miles away from airport *i*, divided by the number of airports 5500 to 6500 miles away. (The list of top and bottom 50 cities, ranked by *ShareBelow6K* can be found in Table A2 in the Online Appendix.) To build intuition for how this instrument is constructed, Figure 3 provides the graphical example of San Francisco (SFO).

[FIGURE 3 HERE]

Note that we will control, in all specifications, for the total number of airports in the range of 5500 to 6500 miles, which means that we will account for factors related to the general isolation or broad location of the airport. The residual variation is what is arguably idiosyncratic. As an example, Philadelphia and Boston will naturally have a similar number of airports located between 5500 and 6500 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 6000-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.

We start with the following reduced-form specification:

$$(2) \quad Y_{ic} = \alpha + \beta * ShareBelow6K_{ic} + X_{ic}\gamma + \varepsilon_{ic},$$

where *c* denotes a grid cell, *i* denotes the closest airport (within the same country) in our sample, *Y_{ict}* is an outcome of interest (night lights, or population, 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, we expect β to be positive.

All regressions will include in the vector *X* the total number of airports between 5500 and 6500 miles away, as discussed above, as well as the log distance in miles from the grid cell *c* centroid to the airport *i*, and region fixed effects (as per the World Bank classification) to ensure that the

²¹We provide robustness tests showing that using alternative windows do not qualitatively alter the main results.

results are not driven by variation across regions. We will further control for grid-cell night lights as of 1992 (earliest data available), as well as population as of 1990, to reduce residual variation and increase the precision of our estimates, given persistence in the data over time.²² In addition, we will use various predetermined covariates to ensure that the results are robust.

To estimate the magnitude of the effects in ways that are more easily interpretable, we can then scale the reduced-form estimates with a first stage estimate using Two-Stage Least Squares (IV/2SLS). This will result in an estimate that captures the marginal effect of additional international connections on local economic growth. In this case, the first stage specification is simply:

$$(3) \quad \text{Connections}_{ic} = \alpha + \beta * \text{ShareBelow6K}_{ic} + X_{ic}\gamma + \varepsilon_{ic},$$

where Connections_{ic} is the number of cities the airport i is connected to (at least weekly flights), and all other variables are defined the same as in equation (2).

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. Our estimations of equation (2) include grid-cells within 100 miles of the airport, as it seems plausible *ex ante* that such cells are potentially affected.²³ However, since 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:

$$(4) \quad Y_{ic} = \alpha + \beta_1 * \text{ShareBelow6K}_{ic} + \beta_2 * \text{ShareBelow6K}_{ic} * \text{Distance}_{ic} + X_{ic}\gamma + \varepsilon_{ic},$$

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 – essentially cells in the city – since Distance_{ic} takes values around zero 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 for at which distance the effects are no longer positive. We also provide results using more flexible estimations of these spatial relationships, which will enable probing for whether there is a point where the effects turn

²²Note that these variables may have been themselves affected by our discontinuity, given the timing of the regulations and the entry of the Boeing 747-400, as we have discussed. However, we will show that our variation is essentially uncorrelated with them, consistent with the idea that important effects would have taken some years to be felt.

²³We will show robustness with respect to other thresholds.

negative, for example due to rural-urban reallocation of economic activity or population.

In all of our specifications, we will cluster the standard errors at the country level, to allow for the possibility of correlated shocks across cities in the same country. We will 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.

3.3 Establishing the Discontinuity

Figure 2 has provided graphical evidence for the existence and evolution of the discontinuity in the likelihood of connection between city pairs at a distance of 6000 miles. We now turn to the task of establishing this more systematically.

The key evidence is in Table 1, implementing versions of (1) with both robust and country-pair-clustered standard errors reported. We see a robust pattern where city pairs just over 6000 miles apart are about 0.3-0.4 percentage points less likely to be connected by at least weekly flights, as of 2014, as compared to those separated by slightly less than 6000 miles. Since the overall likelihood of a given pair in our sample being connected is around one percent, this entails a quantitatively substantial difference.

[TABLE 1 HERE]

The result holds with a first-order polynomial for $f(\text{Distance}_{ij})$ (Columns 1-4), as well as with a second-order polynomial (Columns 5-6). 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 1000 miles (Column 2), before presenting our optimal-bandwidth baseline (Column 3). Neither is it substantially changed when we control 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 (Column 4).

We further probe this result by running specifications where we arbitrarily impose “placebo” discontinuity thresholds other than 6000 – specifically, every 50 miles between 4500 and 7500, leaving aside the range between 5750 and 6250 miles as the match between the regulations and the specific distance is an approximation.²⁴ Figure 4 shows that the estimate at the 6000-mile threshold is much larger, in absolute value, than the placebo alternatives, falling far to the left of the distribution computed for the latter. This reassures us that the effect we pick up is unlikely to be spurious.

²⁴Notably, the absolute value of the coefficient is maximized precisely at 6000 (-0.0038), with the next-highest value at -0.0023 for the discontinuity set at 6050. 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.

[FIGURE 4 HERE]

We can also ask how the discontinuity evolved over time. Column (7) in Table 1 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=0.099), indicating that, in the year of the launch of the Boeing 747-400, the likelihood of connection just below the threshold was perhaps higher, but not strongly so. The remainder of the table then 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 6000-mile threshold (Column 10). In short, the discontinuity seems to have magnified rapidly upon the technological developments of the late 1980s and early 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 will use for our grid-cell analysis. Put simply, does the share of potential connections just below the threshold, *ShareBelow6K_i*, predict the total number of connections that are actually available in airport *i*? Table 2 answers in the affirmative. Columns 1-2 show the basic correlation for the total number of (at least) weekly connections as of 2014, first with no controls at all and then controlling for the total number of airports in the 5500-6500-mile range, as well as region fixed effects. The magnitude of the effect is largely unaffected, and precision greater, 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.

[TABLE 2 HERE]

Interestingly, Columns 4-5 show that there was no significant correlation between our variation and the total number of connections as of 1989. 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.

The remainder of Table 2 then considers the evolution of the pattern over time, by looking at the total number of connection-years for the city. We see that cities with a large share of potential connections just below the threshold already had a significantly higher number of connections in the 1990s (Column 6), and even more so in the 2000s (Column 7), adding up to a strong effect over the two decades.

The estimated magnitudes indicate a substantial effect. A coefficient of 4.88 (Column 2) entails that going from the 25th percentile to the 75th percentile in *ShareBelow6K* (0.488 and 0.640,

respectively) would translate into just under 0.75 additional long-haul connections. This compares to an average of 2.83 connections as of 2014, and 1.39 additional connections between 1989 and 2014. Alternatively, that change translates into about 4.3 additional connection-years over the two decades, again substantial as compared to the average value of 40.42 (Column 8).

It is also interesting to consider the sources of the variation being used in our estimation. The graphical representation in Figure 5 displays the residual variation in *ShareBelow6K* after controlling for the total number of airports in the 5500-6500-mile range and region fixed effects. We can see that the places with very high and very low draws in the “lottery” of potential connections just below the threshold are spread all over the world. Also notably, we can see that there are places with very positive and very negative shocks located very close to one another. This reassures us that the variation is essentially idiosyncratic, and not driven by specific parts of the world.

[FIGURE 5 HERE]

Lastly, we also check that our 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 5500 and 6000 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 (5700-6300, 5200-6800).

In sum, city pairs that are just under 6000 miles apart are indeed more likely to be connected than those just over the threshold, and the difference seems to have increased substantially starting in the 1990s. This translates into the fact that airports with a large share of potential connections just below the threshold have a larger total number of connections as of 2014, as well as of total connection-years over the intervening decades, again essentially driven by an increase in connections after 1989.

4 Flying Economy: Air Links and Economic Development

4.1 Baseline Results

We now turn to studying the impact of air links on economic activity. Table 3 shows results using grid-cell level night lights within a 100-mile radius of any of the 777 airports we can match to the night lights data. We start with reduced-form results linking night lights to the share of potential connections just below the 6000-mile threshold, controlling for the total number within 500 miles of that threshold. Column 1 first shows the correlation without any additional controls other than regional dummies, 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.

[TABLE 3 HERE]

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 then 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 then adds controls for the 1992 level of night lights, as well as population as of 1990, in order to account for possible convergence effects and to increase precision. Columns 5-6 further show that the result is essentially unaltered if we also control for baseline airport characteristics, geographical controls and initial GDP at the country level. (In Table A4 in the Online Appendix we show that *ShareBelow6K* is not correlated with the geographical characteristics either.) Note in particular that the coefficient increases in magnitude as we add covariates.

We then turn to 2SLS specifications, scaling our reduced-form results so as to interpret their implications in terms of the impact of long-haul connections on local economic activity. Columns 7-8 reproduce the sets of controls from Columns 5-6, showing a positive and statistically significant effect of more connections, as measured by total connection-years.²⁵ Columns 9-11 then show that the picture that emerges if we focus on growth rates instead is similar across the board.

To make sense of the magnitudes, consider the baseline coefficients from Column 5. For the reduced form, we find that going from the 25th to the 75th percentile in terms of the share of potential connections just below the threshold explains about one-sixth of a standard deviation of the distribution of the increase in night lights over the period, and about the same for the distribution of growth rates. In the 2SLS context, the estimates imply that one additional connection-year explains about 0.03 standard deviation of both distributions. Put another way, given that the airport closest to the median cell in our sample had one connection-year added over the period, we can explain about 3% of the growth rate of that median location.

Our 2SLS estimates mask considerable heterogeneity in the extent to which the link between potential connections just below the threshold and actual connections materializes for different places. This can be seen when we estimate the first stage separately for different subsamples, depending on how developed they were in 1992 (as measured by night lights over the 100-mile radius

²⁵Given the magnitudes of the first-stage F-statistics, we focus on Anderson-Rubin Wald test p-values, for weak-instrument-robust inference. Results are very similar with the Stock-Wright S-statistic, and with uncorrected p-values as well, as can be seen by comparing coefficients and the corresponding standard errors.

around the airport). In our baseline specification (Column 6), the first-stage coefficient implies that the impact of a unit change in *ShareBelow6K* on additional connection-years is 31.83. As it turns out, the first-stage coefficient for the top quartile of airports is 201.85 (p-value=0.006); in contrast, the one for the bottom half is a mere 2.26 (p=0.747).²⁶ In other words, a place like Vientiane (Laos) gets an excellent draw when it comes to the “lottery” around the 6000-mile threshold, but that does not translate into more connections.

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 our exogenous variation in potential long-haul connections. This indicates a causal impact of air links on economic activity at the local level, but one which is available only to places that were developed enough, to begin with, that they could indeed get connected.

We further check the robustness of our main findings in a number of different ways. (For brevity, all the results are shown in the Online Appendix, with our full set of controls.) We first experiment with the different implementation of our key source of variation: the results remain if we consider the number of cities between 5500 and 6000 miles, instead of the share (Table A6). We then 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 country at a time. We plot the resulting distribution of 2SLS coefficients in a histogram in the Online Appendix (Figure A9), and the coefficients are essentially all within a tight window of the baseline estimate of 0.143, and all very much on the positive (and sizable) side. Interestingly, the smallest coefficient comes when excluding China (0.081). This smaller magnitude suggests that the effect of connections might be particularly strong in a context of fast growth, in which a positive shock to potential links is more likely to materialize as others will be inclined to seize on those links. In any case, we also show that the results excluding China are still statistically significant (Table A7).

We further consider different subsamples, in the remainder of Table A7. We show that the results are unaltered if we leave out, for each airport, the cell whose centroid is closest to the airport coordinates, indicating that the effect on night lights is not being driven by the airport itself, as opposed to a broader increase in economic activity. We also experiment with alternative thresholds: our results still hold when we choose 50 or 150 miles instead.

Finally, we check robustness with respect to different ways of computing standard errors (Table A8). First, our results remain qualitatively unaltered if we cluster standard errors at the level of

²⁶The threshold for the top quartile is St Louis, MO, while the median place is Cape Town. The typical (median) city in that bottom half is Abidjan, and the subsample is disproportionately in Sub-Saharan Africa: the region has about 10% of airports in our sample, but 20% of those in the bottom half.

airports. This deals specifically with the fact that the variation we use is at that level. Second, we use the Conley (1999) approach allowing for spatial correlation across different cells, and again the results hold, using thresholds of 100 or 500 miles.

4.2 Spatial Patterns

This basic result begs an important question: does this effect constitute a genuine increase in the overall level of economic activity, or is it evidence of its spatial reallocation instead? In particular, it could be the case that more connections lead business or individuals to relocate closer to the airport, leaving other more distant locations that would then endure negative spillover effects. In other words, the growth around the airport could be at the expense of the hinterland, which would obviously have very different implications from a standpoint of policy or welfare.

We can study this question by exploiting the spatial richness and granularity in the available data. Specifically, we do not have to restrict our attention to the immediate vicinity of the airports, but rather 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 *ShareBelow6K* and its interaction with grid-cell distance to the nearest airport in the country. Based on our previous results, we would expect a positive coefficient for the main effect of *ShareBelow6K*, 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 4, and align with that intuition. Column 1 displays the increase in night lights between 1992 and 2010 as the dependent variable, and controlling for *Around6K*, regional dummies, the main effect of (log) distance to the airport, as well as 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 grid cells at most 500 miles away from the closest airport. Using the same sample, Columns 3 adds airport and geographic controls, confirming the same message.

[TABLE 4 HERE]

The point estimates from Columns 2-3 imply that the effect turns negative at a distance of just under 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 way. Specifically, we run our baseline regressions, as in Table 3, but restricting the sample to the ranges of 100-250 and 250-500 miles from the closest airport. (For comparison, Column 4 reproduces the result from Table 3, within a 100-mile range.) As it turns out, the results in Columns 6-7 show precisely estimated zeros. Columns 8-10 establish that the same is true when we look at growth rates instead.

This provides evidence against the hypothesis of pure spatial reallocation – at the very least, any negative spillovers for more remote places are canceled out by positive ones. In that sense, connections seem to generate new economic activity on net. However, it is still true that the effect is to exacerbate inequalities over space, as connected cities grow faster relative to their hinterlands.²⁷

That said, a more complete account of the spatial patterns requires us to take into account that people can move across space. For that we turn to grid-cell level population as the outcome of interest for our regression specifications, which we display in Table 5. Perhaps surprisingly, the results display no evidence of an increase in population in the immediate vicinity of the airport (Column 1). There do not seem to be significant changes in population farther away from the airport either (Columns 2-3). The remainder of the table show that the same message is true when considering the specification with the interaction with distance to the airport, both in terms of the increase or growth rates.

[TABLE 5 HERE]

This suggests that the increased economic activity is not being matched by population movements, at least over the time frame of the analysis. Of course, it might be that those movements take place over a longer horizon. Alternatively, to the extent that standard spatial equilibrium arguments would require individuals and businesses to be indifferent across space (Glaeser and Gottlieb 2009), this can be interpreted as evidence that higher incomes may be getting translated into economic rents.

In sum, we find robust evidence of a causal impact of long-haul air links on economic activity at the local level. This impact is not merely driven by spatial reallocation, though it does seem to fuel inequality across space, which does not appear to be compensated by population moves, at least over the time horizon we are able to study.

This begs the question of what underpins this impact. The key differential of air transportation relative to other alternatives, as we have argued, is its implications for the ease of transporting

²⁷Note 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 country. We will return to this question in the next section.

people over long distances – much more so than goods, as trade still flows mostly by sea. It stands to reason that the impact of air links comes from fostering connections between people.

To understand how this plays out, we will first map out how the variation in long-haul connections spills over into different types of connections, thereby affecting a given airport’s position in the network of air links, and how that translates into the actual flow of people. Then we will turn to how the flow of people can foster the flow of capital, as connections might influence links between businesses across long distances.

5 Connecting People: Network Spillovers and Passenger Flows

Our key source of variation affects directly the availability of air links over a specific range, yet we have found effects on the total number of connections available at a given airport, and that this in turn translates into an important impact on economic activity. 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 6000 miles. The result is in Column 1 of Table 6. (All specifications in Table 6 control for the number of airports in the 5500-6500-mile range, region fixed effects, and 1989 airport controls.) We see that places with more potential connections just below the threshold indeed add more connections over the 5500-6500-mile range.

[TABLE 6 HERE]

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 8 total long-haul connections overall. In fact, the contrast between Columns 3-4 and Columns 5-6 shows that the spillovers are essentially all coming from the shorter range between 2000 and 5500 miles. The effect is considerably smaller above 6500 miles, which is unsurprising in light of the relatively low number of ultra-long-haul flights. 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.

This pattern also manifests itself in spillovers across different airports in the network. We present in the Online Appendix (Table A9) the results from running specifications that, in addition to airport i ’s own *ShareBelow6K_i*, include the average value of *ShareBelow6K* for all airports within a given range from airport i – less than 1000 miles, 1000-2000 miles, 2000-5500 miles, and above 6500 miles. We find evidence of significant positive spillovers in the 2000-5500-mile range, and negative but small and statistically insignificant coefficients for the other ranges. In short, it seems that, when airports that are within a medium-to-long-range distance from airport i receive a positive shock to their number of long-haul connections, airport i increases its number

of connections: put simply, Milan’s connection to Shanghai can induce further connections in New York (4000 miles away) or Lagos (2700 miles), as it becomes more appealing to fly to Milan. While this could have had a negative diversion impact on closer airports such as Madrid, that seems to be balanced out, arguably by Madrid’s own extra incentive to connect to Milan as well.

We can also show that the more numerous connections actually translate into a better position in the network. This is what we see in Columns 7-8 of Table 6, in which the dependent variable is a measure of (eigenvector) centrality of airport i . This is designed to capture the influence of that specific node in the overall network: it assigns to each node a score, and connections to high-scoring airports contribute disproportionately to the score of airport i . We find that more potential connections below the threshold increases this measure, indicating that when airport i receives a favorable shock, it manages to add flights to better connected airports.

Still, what ultimately matters is whether those extra flights increase the flow of people. Of course, it would be rather surprising if that were not the case, and the last two columns in Table 6 confirm that intuition. The 2SLS estimate indicates that an extra (at least weekly) long-haul connection brings in 130 thousand additional passengers per year. For perspective on this magnitude, this corresponds roughly to a standard Boeing 747-400 flying twice a week in and out of the airport at full capacity (660 passengers).

Taken together, these results establish that the impact of a shock yielding more connections within a relatively narrow range of distance is magnified by the ripple effect that this has over shorter distances, with additional connections inducing yet more additional connections. This in turn translates into a substantial increase in the flow of people going through a city.

6 Flying Business: Air Links, Business Links, and Capital Flows

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.²⁸ 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.²⁹

²⁸Another possibility is that the impact we find is driven by leisure tourism flows, but this sector seems too small to justify the sizable impact we find. 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 responds 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.

²⁹The president of Alitalia captured the sentiment as he announced his airline’s new Italy-China ventures: “China

Then there is simply revealed preference: non-stop flights typically command a substantial premium over the alternatives.³⁰ Those flights save time, 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, etc. On the same vein, businesses tend to locate disproportionately close to airports.³¹ Last but not least, there is also growing empirical evidence of the business value of direct flight links (Giroud 2013, Bernstein, Giroud, and Townsend forth.).

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

6.1 City-Pair Evidence

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).³² 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, we 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 first compute, for each company in the data, the distance between the airport in our sample that is closest to its location and the one that is closest to the location of its owners. Figure 6 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 our sample. We see a substantial drop in the number of ownership links around the 6000-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

represents a fundamental market for our country, which must aim at growth of Chinese investments and tourism, [and] I believe that the Milan-Shanghai flight (...) will establish a very important bridge.” (*China Daily*, May 5, 2015)

³⁰For instance, a simple Google flight search for a business-type Monday-Friday roundtrip between London and Shanghai, for July 2016, yields \$704 for the cheapest one-stop flight, against \$1,365 for the cheapest non-stop alternative. The direct flight saves about 2 hours over the shortest connecting alternative – a difference of about one-sixth.

³¹Stilwell and Hansman (2013, p.69) show that, in the US, the headquarters of over 50% of Fortune 500 companies, as well as 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 2016) – and this in spite of obvious drawbacks (noise, traffic, height restrictions on buildings). See also Bel and Fageda (2008).

³²There is evidence that proximity matters for venture capital investors (Bernstein, Giroud and Townsend, forth.), and that migration links matter for FDI across locations in the US (Burchardi, Chaney and Hassan 2015).

of direct air links on the businesses connections between cities.

[FIGURE 6 HERE]

In order to assess more systematically whether that is the case, we match the Orbis data to all possible airport dyads in our data set. Specifically, focusing again on companies with both parties within 100 miles of one of the airports in our 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 then add all the companies for each of the nearly 335 thousand possible pairs.

This allows us to resort again to RD methods to estimate the reduced-form impact of distance around the 6000-mile threshold on the number of ownership links. The results are in Table 7, 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, a second-order polynomial, 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 4. As can be seen in the Online Appendix (Figure A10), it is once again the case that the estimate at the 6000-mile threshold is a left outlier in the distribution computed for the placebo estimates. Quantitatively, we estimate a drop of around 0.9 firms comparing the two sides of the discontinuity, which corresponds to about 65% of the average, or 0.05 standard deviations.

[TABLE 7 HERE]

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 2 shows a drop in the number of connected city pairs, around the 6000-mile threshold, by a factor of roughly 1/3. Figure 6, in turn, shows a drop in the number of ownership links by a factor of roughly 1/3. This suggests that a given increase in connections generates about a similar proportional increase in ownership links.³³ 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 our data, and an increase of that magnitude is comparable to taking this number to the level of links between London and Malmo (Sweden).

³³To use more precise numbers: there are 107 connected city pairs between 5500 and 6000 miles, and 34 between 6000 and 6500 (a factor of 0.32), against 27,964 and 10,229 ownership links (a factor of 0.36). Since $0.32/0.36 \approx 0.9$, this means that increasing the number of connected pairs by 10% leads to an increase in ownership links by about 9%.

This is confirmed by Columns 7-8 in Table 7, which exploit a fuzzy RD design where the independent variable of interest is a dummy that takes a value of 1 if the airport pair happened to be connected via (at least) weekly flights in 2014. (Results are essentially identical if we consider weekly connections at some point between 2005 and 2014.)

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 emergence of the 6000-mile discontinuity in air links. We start off by constructing a plot analogous to Figure 6: 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 7, 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 6000-mile threshold. Interestingly, the white dots depicting the pre-2000 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.

[FIGURE 7 HERE]

We can then pursue a similar RD-based exercise using the GDELT data. (Full results are left to Table A10 in the Online Appendix, in the interest of brevity.) Not surprisingly in light of Figure 7, the results mirror our findings using the ownership data, indicating a causal impact of air links. In particular, pairs of cities just below 6000 miles apart have more instances of business collaboration after 2000, and witnessed a larger increase relative to the pre-2000 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 of the connections between businesses located in each of them, consistent with the idea that an enhanced ability to engage in face-to-face interactions fosters those connections.

6.1.1 Catching the Convergence Plane: Where Does Capital Flow To?

We can dig deeper into the nature of these business links, 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 airport as “richer” or “poorer” according to the relative (PPP-adjusted) income per capita of the country they are in, as of 2011, measured by the Penn World Tables (version 8.0).³⁴ We then focus our analysis on all pairs of airports such that the countries in which they are located are 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 way we can avoid flagging a German firm opening a subsidiary in Luxembourg as an example of capital flowing from the poor to the rich.

The results are in Table 8. Columns 1-2 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. (The p-value for Column 2 is around 0.12.) In fact, if we compare the magnitudes of the two estimates, we can conclude that 3/4 of the effect on total cross-ownership links comes from capital flowing from rich to poor. Columns 3-4 break it down further by focusing on pairs such that the richer country is classified as “High income.” The results are very much the same, indicating that the flows are originating largely in that wealthiest tier.

[TABLE 8 HERE]

This suggests that the impact of air links on business connections operates as a force for convergence. We must qualify this statement, however, as Columns 5-6 show that the capital flows in question are essentially taking place between “High income” and “Middle income” countries – in other words, from rich countries to the Chinas and Indias and Brazils of the world. In contrast, countries classified as “Low income” are essentially shut out of this process (Columns 7-8). In fact, if we focus on the first-stage relationship between distance below 6000 miles and the existence of a direct connection between the pair (Panel A), we see that the poorest countries do not benefit from having potential connections. This again indicates that the payoff from that potential does not materialize when one is too poor for there to be a demand for connecting in the first place.

6.2 Grid-Cell Evidence

We are also interested in whether the enhanced business links underlie the effect of air links on economic activity at the local level. To shed further light on this question, we now ask whether the city-pair evidence we have just presented aggregates up to a substantial effect on the ability of businesses in a given location to connect with businesses elsewhere over long distances. For that we once again turn to the rich spatial information afforded by our data. Both in the case of ownership links and business events, we have detailed information that allows us to match companies and events to specific locations at the global grid-cell level.

³⁴We use 2011 to increase the sample size, as many countries do not have data for 1989 or before. Results are very similar if we use 1989 instead.

We again start off with the ownership data. We first aggregate the data to the level of specific locations as defined by latitude and longitude coordinates. This allows us to obtain, for each of these locations, the number of firms that are located there and owned by foreigners, as well as the number of foreign companies owned by individuals or firms located there. We match each of the locations to the closest grid cell in our full spatial data, and aggregate numbers at the grid-cell level. As a result, for each grid cell we have the information on the number of firms owned abroad and owned by foreigners.

We then consider whether our key airport-level variable, *ShareBelow6K*, predicts the level of connections established by businesses located close to a given airport, again focusing on locations that are within 100 miles from one of the airports in our sample. The results are in Table 9, using Poisson specifications.³⁵

[TABLE 9 HERE]

Column 1 shows that indeed it is the case that locations surrounding airports with more potential links just below the discontinuity host more foreign-owned companies. Column 2 presents the IV (GMM) estimate, using *ShareBelow6K* as an instrument for weekly connections in 2014. We see a positive effect of air links on the presence of such companies. Columns 3-4 then show that the effect is basically the same when it comes to companies owned abroad, showing that the impact of air links is felt in both directions. Quantitatively, an extra weekly connection added over the period is associated with a 4% increase in the number of firms in a grid cell.³⁶

But where exactly are those foreign companies and owners? Our variation induces connections just below the 6000-mile threshold, and we have seen that these spill over into shorter distances as the new connections enhance a city's position in the network. We can compute the measure of foreign-owned and owned-abroad companies separately for different ranges of distance to a given city, and check whether the pattern of ownership links matches what we would expect from the pattern of connections. Columns 5-8 in Table 8 show that this is precisely what happens in the case of foreign-owned companies. The strongest increase happens for owners located 5000 to 6000 miles away from the city (Column 7), with spillovers between 3000 and 5000 (Column 6) and 1000 and 3000 (Column 5). In contrast, there is no increase in the number of foreign owners located more than 6000 miles away, in line with the fact that, as we have seen, our variation does

³⁵Note that almost 87% of cells in our sample (and 94% of the total cells) have no foreign-owned or owned-abroad companies.

³⁶How does this magnitude compare with the extra 250 firms we estimated from an additional connection at the city-pair level? If we aggregate cells to the city level, we have an average of 636 companies owned abroad, and 630 foreign-owned companies. A 4% increase would thus correspond to an extra 51 companies in total. The previous estimate, however, was by definition a local estimate around the discontinuity. If we focus on the change in connections around the threshold, at the city level, the estimated coefficient would entail a 27% increase, which would correspond at the mean to an extra 342 companies.

not induce many connections over that range. Columns 9-12 then convey a similar message for companies owned abroad, with the strongest increase over the 5000-6000-mile range, albeit with weaker spillovers for shorter distances.

The evidence thus suggests a causal impact of increased air links, at the local level, and the number of business links established by local companies, in terms of ownership. How about interactions as measured by geolocated business collaboration events? We reproduce our basic analysis using the GDELT data, constructed analogously to what we have done with the Orbis data. (Full results again left to the Online Appendix, Table A11.) We find a positive impact, indicating that places close to airports that add more connections induced by our variation indeed record more cooperation events linking local companies to businesses across long distances over the post-2000 period. In contrast, no such relationship exists for events recorded in the pre-2000 period, again consistent with the increase in importance of our key discontinuity over time.

As a final piece of evidence on the effect of air links on the business environment at the city level, we look at the location of corporate headquarters. It is, after all, generally understood that multinational corporations appear to value being located in connected places.³⁷ Similarly, the enhanced business connections we have detected could also offer opportunities for the growth of local firms.

With that in mind, we adapt our basic empirical exercise using the presence of “Global 2000” corporations headquarters in a given location (grid cell) as the outcome variable of interest, in a probit specification. The results (in the remainder of Table A12 in the Online Appendix) indicate a positive impact of air links on that likelihood. We have data from 2006, 2009, and 2012, and for all three years we find a positive and significant coefficient for *ShareBelow6K* in the reduced form. The magnitude of the marginal effect is around one percentage point, implying that going from the 25th to the 75th percentile would increase the likelihood of presence of a corporate HQ by about 0.15 percentage points – for comparison, the average probability in our sample is 0.8%. We also detect an effect of similar magnitude on the probability of there being an increase in the number of corporate HQs in a given cell between 2006 and 2012.

It thus seems that connections also have an impact on the likelihood of presence of large corporate headquarters. That said, we cannot tell from the data the extent to which this would have been driven by location decisions specifically, or by connections enabling the growth of local companies that would then make their way into the global list.

³⁷See for instance Pred (1977, p.24): “There are tremendous savings in time, and hence costs, that accrue from the clustering of organizational head offices and ancillary business services in major metropolitan areas [...] compounded by the superior air-transport connections those places possess. [...] Centers which do not have a wide variety and great number of daily nonstop flights to the leading metropolitan complexes with a given system of cities are not particularly attractive [...] because they do not permit nonlocal personal contacts [...] to be carried out [...] efficiently.”

6.3 In Sum

The evidence shows that air links matter for business links: when two cities get connected, there is a substantial increase 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 abroad, and with an increased likelihood of presence of large corporate headquarters.

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.

Our 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).

7 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 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, presumably by enhancing the possibilities for face-to-face contact over long distances.

This naturally leads us to the question of 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 effect 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 that we investigate in ongoing research.

Still on the economic side, our evidence provides a potential rationale for policy interventions designed to increase the number of connections available from a given airport, city, or country

– and even, to some extent, the existence of airline subsidies, flag carriers, and the like.³⁸ This interpretation, however, would require a lot of caution. First, the welfare impact can be called into question, especially in light of the absence of an effect on the spatial distribution of the population: higher incomes may be getting translated into economic rents – though perhaps population movements could take place over longer time horizons.

In addition, our results indicate that, while the expansion around the airport represents genuinely increased levels of economic activity, as opposed to pure spatial reallocation to the detriment of more distant areas, the fact of the matter is that those distant areas still get left behind. In other words, more connections induce spatial inequality, which should be taken into account when assessing the desirability of interventions designed to increase them.

Finally, another layer of inequality underlying our 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 those places that are too poor to begin with, there is no first-stage relationship between the share of potential connections just below the 6000-mile threshold and the actual increase in the number 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 the increased business links and the capital flows that are 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 in the runway.

References

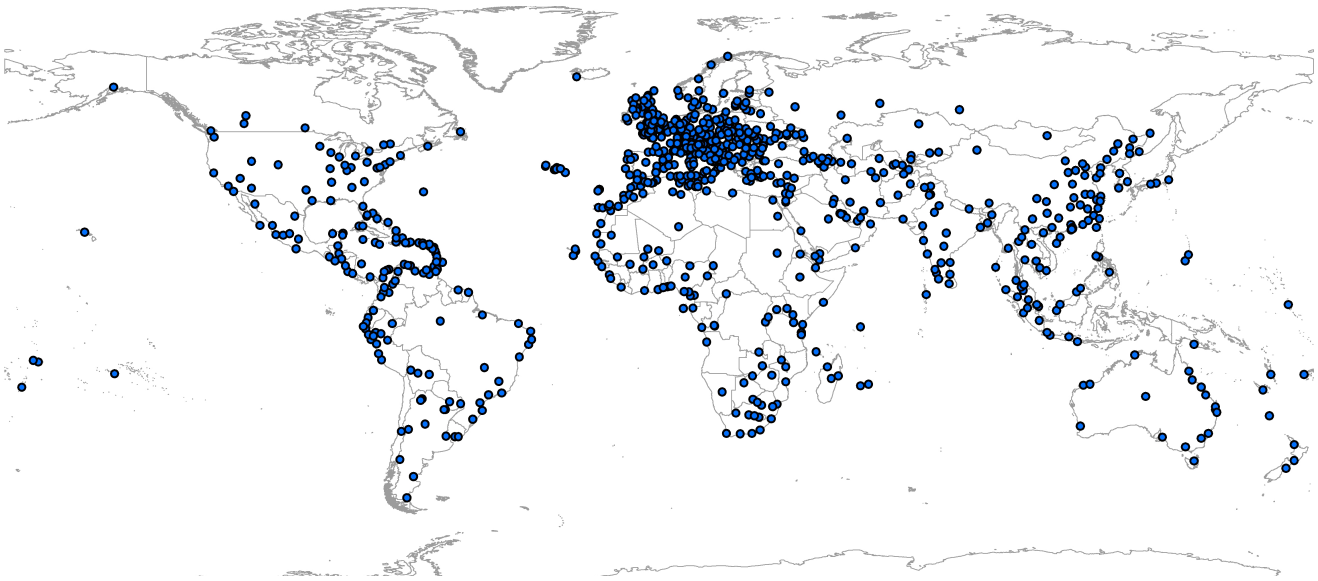
- Anderson, John D. (2002).** *The Airplane: A History of Its Technology*. American Institute of Aeronautics.
- Bacchetta, Marc and Marion Jansen (Eds.) (2011)** *Making Globalization Socially Sustainable*. ILO and WTO.
- Bel, Germà and Xavier Fageda (2008).** “Getting There Fast: Globalization, Intercontinental Flights and Location of Headquarters.” *Journal of Economic Geography* 8: 471-495.
- Bernstein, Shai, Xavier Giroud, and Richard R. Townsend (forth.).** “The Impact of Venture Capital Monitoring.” *Journal of Finance*.

³⁸See headlines such as “Africa’s ailing national airlines survive on USD2.5 billion of government subsidy” (<http://centreforaviation.com/analysis/africas-ailing-national-airlines-survive-on-usd25-billion-of-government-subsidy-not-sound>), “U.S. airline group: Subsidies give Gulf carriers an unfair edge” (<http://www.travelweekly.com/Travel-News/Airline-News/US-airline-group-Subsidies-give-Gulf-carriers-an-unfair-edge>), or “How much do taxpayers support airlines?” (<http://www.usatoday.com/story/travel/columnist/mcgee/2015/09/02/how-much-do-taxpayers-support-airlines/71568226/>)

- Bleakley, Hoyt and Jeffrey Lin (2012).** “Portage and Path Dependence.” *Quarterly Journal of Economics* 127(2): 587-644.
- Bourguignon, François (2015).** *The Globalization of Inequality*. Princeton University Press.
- Brueckner, Jan K. (2003).** “Airline Traffic and Urban Economic Development.” *Urban Studies* 40(8): 1455-1469.
- Burchardi, Konrad, Thomas Chaney, and Tarek A. Hassan (2015).** “Migrants, Ancestors, and Foreign Investments.” University of Chicago (unpublished).
- Catalini, Christian, Christian Fons-Rosen, and Patrick Gaule (2016).** “Did Cheaper Flights Change the Direction of Science?” CEPR Discussion Paper 11252.
- Clark, Paul (2002).** *Buying the Big Jets*. Ashgate Publishing Limited.
- Collier, Paul (2007).** *The Bottom Billion: Why the Poorest Countries are Failing and What Can Be Done About It*. Oxford University Press.
- Conley, T.G. (1999).** “GMM Estimation with Cross Sectional Dependence.” *Journal of Econometrics* 92: 1-45.
- Cristea, Anca D. (2011).** “Buyer-Seller Relationships in International Trade: Evidence from U.S. States’ Exports and Business-Class Travel.” *Journal of International Economics* 84(2): 207-220.
- Dollar, David (2005).** “Globalization, Poverty, and Inequality since 1980.” *World Bank Research Observer* 20(2): 145-175.
- Dollar, David and Aart Kraay (2004).** “Trade, Growth, and Poverty.” *Economic Journal* 114(493): F22-F49.
- Donaldson, Dave (forth.).** “Railroads of the Raj: Estimating the Impact of Transportation Infrastructure.” *American Economic Review*.
- Donaldson, Dave and Richard Hornbeck (2016).** “Railroads and American Economic Growth: A ‘Market Access’ Approach.” *Quarterly Journal of Economics*.
- Dreher, Axel (2006).** “Does Globalization Affect Growth? Evidence from a New Index of Globalization.” *Applied Economics* 38(10): .
- Feyrer, James (2009).** “Trade and Income: Exploiting Time Series in Geography.” NBER WP 14910.
- Frankel, Jeffrey and David Romer (1999).** “Does Trade Cause Growth?” *American Economic Review* 89(3): 379-399.
- Giroud, Xavier (2013).** “Proximity and Investment: Evidence from Plant-Level Data.” *Quarterly Journal of Economics* 128(2): 861-915.
- Glaeser, Edward L. (2011).** *The Triumph of the City*. Macmillan.
- Glaeser, Edward L. and Joshua D. Gottlieb (2009).** “The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States.” *Journal of Economic Literature* 47(4): 983-1028.
- Green, Richard K. (2007).** “Airports and Economic Development.” *Real Estate Economics* 35: 91-112.
- Henderson, J. Vernon, Adam Storeygard, and David N. Weil (2012).** “Measuring Economic Growth from Outer Space.” *American Economic Review* 102(2): 994-1028.

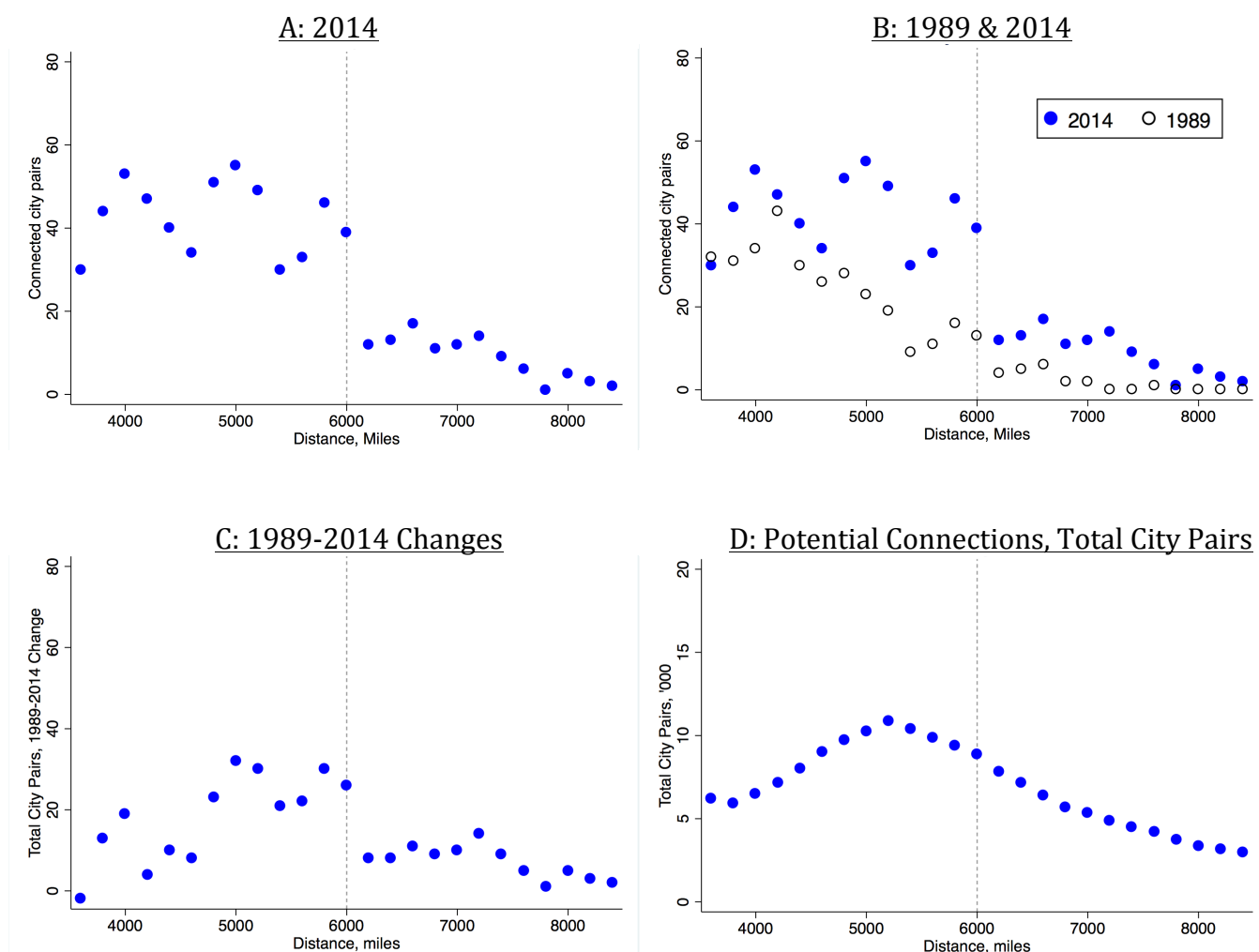
- Hummels, David (2007).** “Transportation Costs and International Trade in the Second Era of Globalization.” *Journal of Economic Perspectives* 21(3): 131-154.
- Imbens, Guido and Thomas Lemieux (2008).** “Regression Discontinuity Designs: A Guide to Practice.” *Journal of Econometrics* 142(2): 615-635.
- Kasarda, John D. (2016).** “Aerotropolis.” In: *Wiley-Blackwell Encyclopedia of Urban and Regional Studies*. West Sussex UK: John Wiley & Sons Press.
- Leetaru, Kalev and Philip A. Schrodt (2013).** “GDELT: Global Data on Events, Location and Tone, 1979-2012.” International Studies Association Annual Conference.
- Lowe, Matt (2014).** “Night Lights and ArcGIS: A Brief Guide.” MIT (unpublished).
- Lucas, Robert E., Jr. (1990).** “Why Doesn’t Capital Flow from Rich to Poor Countries?” *American Economic Review* 80(2):92-96.
- Manacorda, Marco and Andrea Tesei (2016).** “Liberation Technology: Mobile Phones and Political Mobilization in Africa.” Queen Mary University of London, unpublished.
- Michalopoulos, Stelios and Elias Papaioannou (2013).** “Pre-colonial Ethnic Institutions and Contemporary African Development.” *Econometrica* 81(1): 113-152.
- Mukkala, Kirsi and Hannu Tervo (2013).** “Air Transportation and Regional Growth: Which Way Does the Causality Run?” *Environment and Planning A* 45(6): 1508-1520.
- Ortega, Francesc and Giovanni Peri (2014).** “Openness and Income: The Roles Of Trade and Migration.” *Journal of International Economics* 92: 231-251.
- Pascali, Luigi (2015).** “The Wind of Change: Maritime Technology, Trade and Economic Development.” University of Warwick (unpublished).
- Poole, Jennifer P. (2013).** “Business Travel as an Input to International Trade.” UC Santa Cruz (unpublished).
- Pred, Allan (1977).** *City-Systems in Advanced Economies: Past Growth, Present Processes, and Future Development Options*. John Wiley and Sons.
- Proctor, Jon, Mike Machat, and Craig Kodera (2010).** *From Props to Jets: Commercial Aviation’s Transition to the Jet Age 1952-1962*. Specialty Pr Pub & Wholesalers.
- Redding, Stephen, Daniel Sturm, and Nikolaus Wolf (2011).** “History and Industrial Location: Evidence from German Airports.” *Review of Economics and Statistics* 93(3): 814-831.
- Stilwell, Justin D. and R. John Hansman (2013).** “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.
- Storper, Michael and Anthony J. Venable (2004).** “Buzz: Face-to-Face Contact and the Urban Economy.” *Journal of Economic Geography* 4(4): 351-370.
- The Economist (2005).** “Business on the Fly: How airports act as a hub for business development.” Nov 24, 2005.
- The Economist (2015).** “Do Hub Airports Boost Growth?” Dec 1, 2015, available at: <http://www.economist.com/blogs/gulliver/2015/12/economics-aviation-capacity>
- Wall St Journal (2012).** “How Airlines Spend Your Airfare.” Jun 6, 2012, available at: <http://www.wsj.com/articles/SB10001424052702303296604577450581396602106>
- Yilmazkuday, Dumen and Hakan Yilmazkuday (2014).** “The Role of Direct Flights in Trade Costs.” Federal Reserve Bank of Dallas Globalization and Monetary Policy Institute Working Paper No. 179.

Figure 1. Cities with Major International Airports



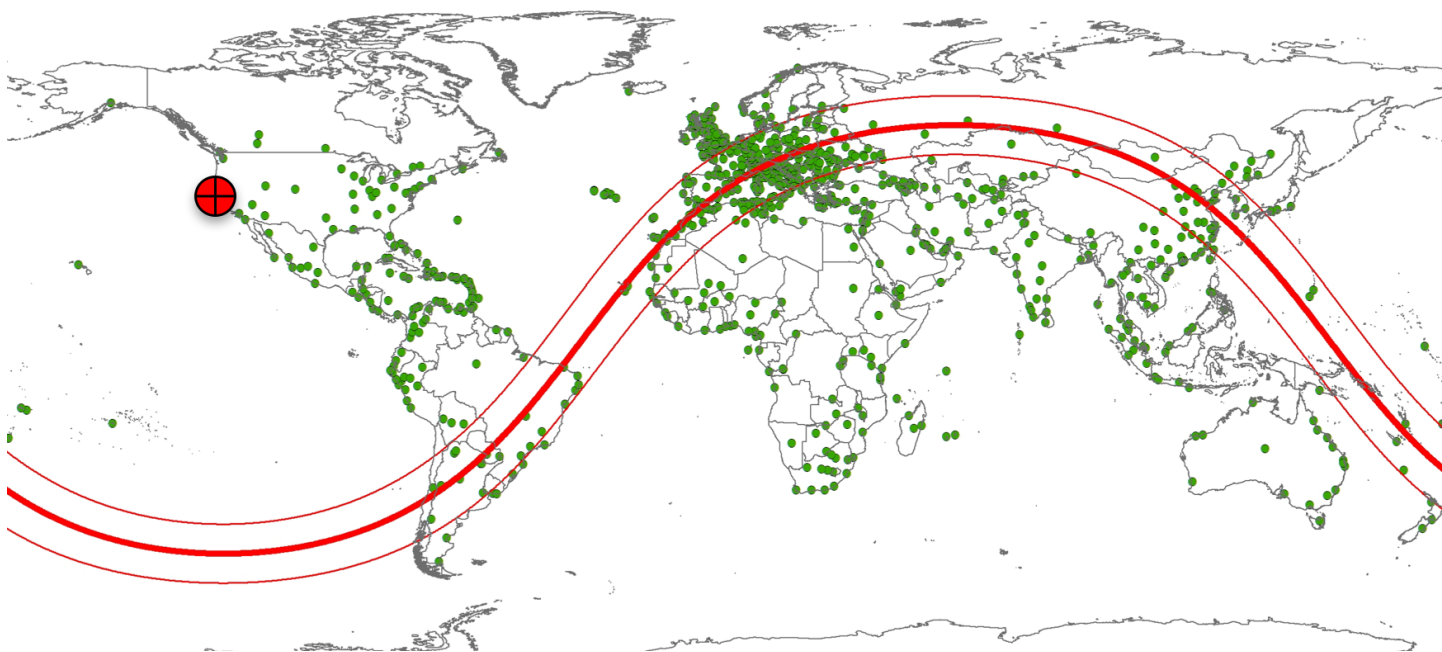
Notes: The map plots the locations of the 819 cities in the sample, all with major international airports as defined by International Civil Aviation Organization (ICAO).

Figure 2. Connections Between City Pairs, by Distance



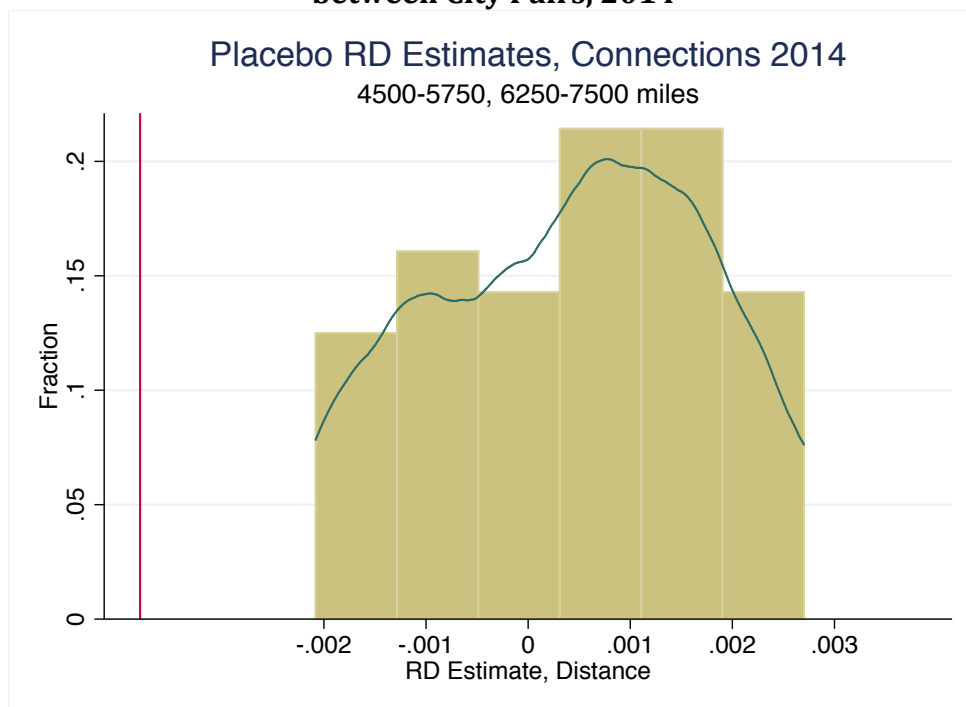
Notes: A connected city pair (airport pair) is defined as having at least weekly non-stop flights between the two cities. The data consists of the 819 cities in our baseline sample. Panel A displays the total number of connected city pairs in 2014 by distance. Panel B adds connected pairs in 1989. Panel 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 6000 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 6000 miles.

Figure 3. Constructing the Airport-Level Instrument, SFO Example



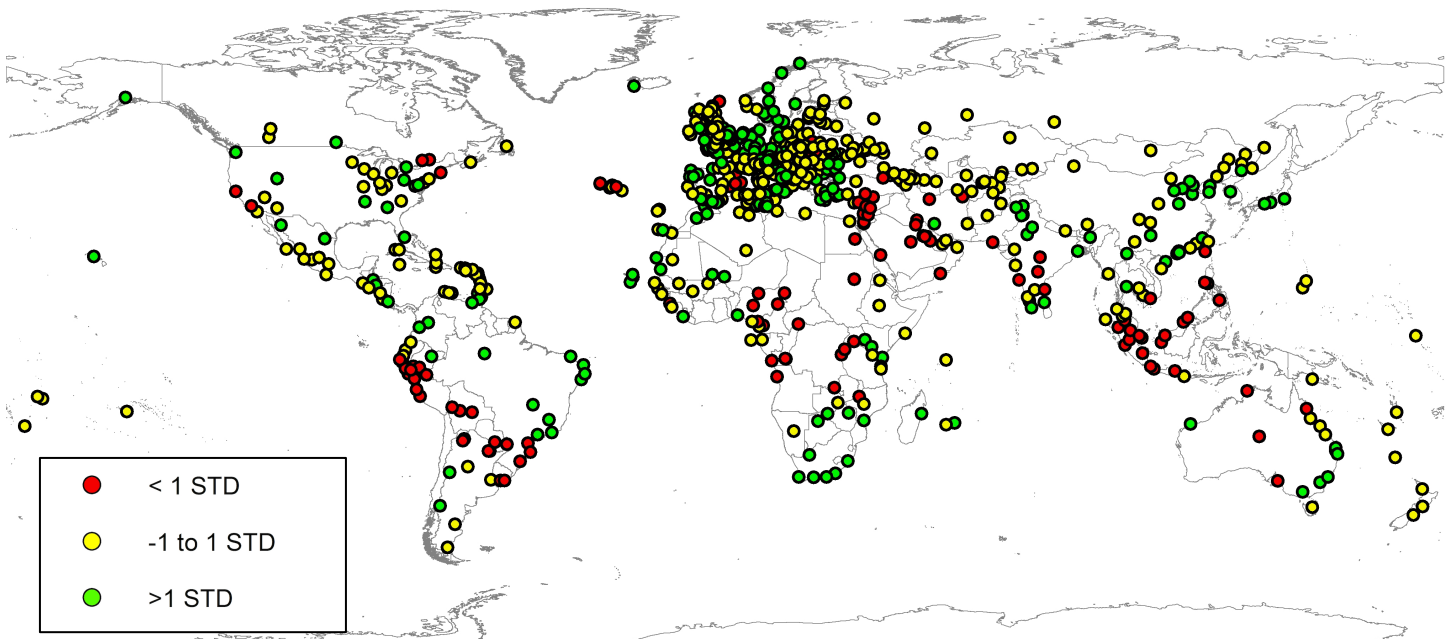
Notes: The thick red line is drawn 6000 miles from San Francisco International Airport (SFO). The buffer around the thick line indicates which other airports (cities) that are located within 5500-6500 miles from SFO. For each of the 819 observations, the airport-level instrument is the share of other cities within the buffer that are located below 6000 miles.

Figure 4. Placebo Regression Discontinuity Estimates for Presence of Connections between City Pairs, 2014



Notes: Histogram (and kernel approximation) for regression discontinuity estimates computed using each 50-mile point between 4500 and 5750 miles, and between 6250 and 7500 miles as distance thresholds. Specifications use first-order polynomial and optimal bandwidth. Vertical line depicts estimate with 6000-mile threshold. The plot shows that the estimate at the 6000-mile threshold is a clear outlier.

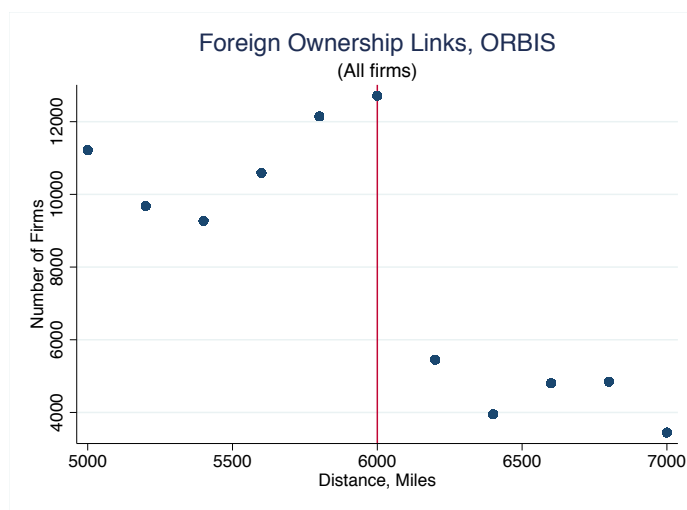
Figure 5. Identifying Variation, Airport-Level



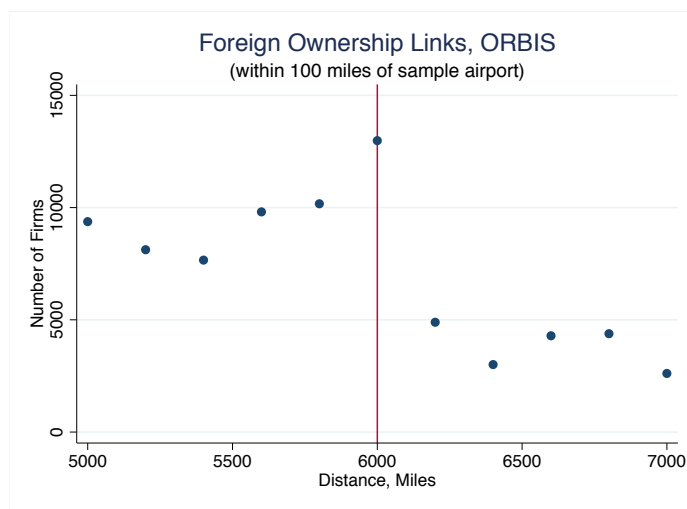
Notes: The map depicts the identifying variation across the 819 airports. The identifying variation for each airport/city is the OLS residual of the instrument, after controlling for region fixed effects and the number of other airports in the 5500-6500 miles buffer. The map shows that there is meaningful variation within regions.

Figure 6. Number of Firms with Cross-Ownership Links, by Distance between Closest Airports

A: All Firms

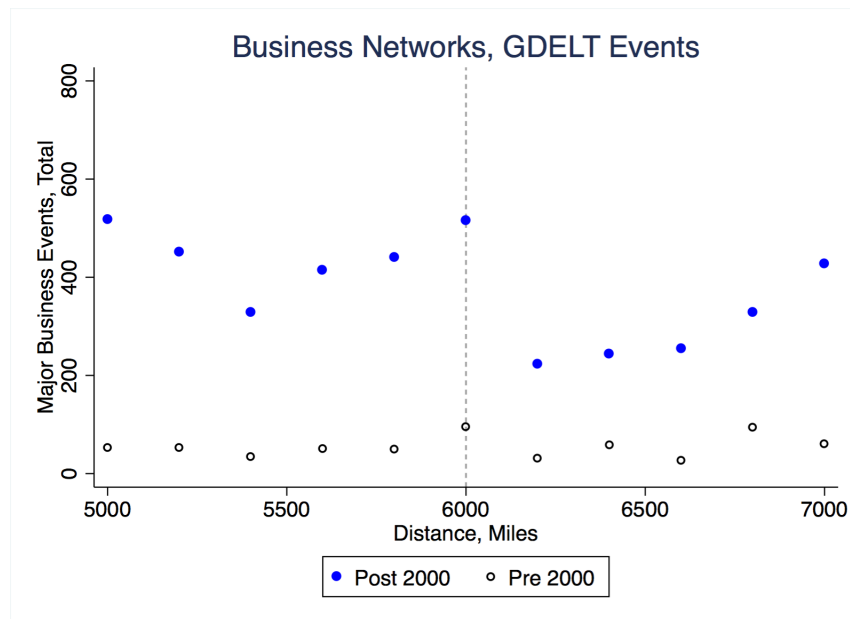


B: Firms within 100 miles of Airport



Notes: 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 our 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 our 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 6000 miles.

**Figure 7. Major Business Events (GDELT) Pre- and Post-2000,
by Distance between Closest Airports**



Notes: 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-2000” refers to events recorded after and including 2000, “Pre-2000” refers to events recorded before 2000. 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 6000 miles.

Table 1. Regression Discontinuity Regressions Around 6000 Miles, City-Pair Level

<i>Bandwidth, Miles</i>	Dep. Var.: Connected, Dummy							Dep. Var: Years Connected		
	2014	2014	2014	2014	2014	2014	1989	1990-2000	2000-2010	1990-2010
	500	1000	Optimal	Optimal	Optimal	Optimal	Optimal	Optimal	Optimal	Optimal
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Distance Below 6000 Miles	0.0031	0.0035	0.0038	0.0030	0.0040	0.0030	0.0010	0.0139	0.0190	0.0347
Robust S.E.	(0.0012)***	(0.0008)***	(0.0009)***	(0.0008)***	(0.0010)***	(0.0009)***	(0.0005)*	(0.0042)***	(0.0073)***	(0.0111)***
Cluster S.E.	(0.0012)***	(0.0009)***	(0.0010)***	(0.0008)***	(0.0011)***	(0.0009)***	(0.0006)*	(0.0044)***	(0.0076)**	(0.0116)***
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	72,537	113,565	120,260	60,595	76,646	73,974	75,619
Dep. Var. Mean	0.0103	0.0103	0.0103	0.0103	0.0103	0.0103	0.0057	0.0678	0.0938	0.1554

Local polynomial Regression Discontinuity estimates, using the *rdrobust* command in Stata (default options unless otherwise stated). Running variable is distance between airports, discontinuity at 6000 miles. Optimal bandwidth selected using one common MSE-optimal bandwidth selector. "Polynomial order" refers to 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 in parentheses. "Cluster S.E." refers to cluster-robust nearest neighbor variance estimation at the country-pair level, using a minimum of three nearest neighbors. The covariates are all measured in 1989 and consist of: a dummy for having weekly flights, 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<0.01, ** p<0.05, * p<0.1

Table 2. Effect on Air Links, Airport Level

Dependent Variable:	<i>Connected Cities, 2014</i>	<i>Connected Cities, 2014</i>	<i>Connected Cities, 2014</i>	<i>Connected Cities, 1989</i>	<i>Connected Cities, 1989</i>	<i>Total Connection-Years, 90-00</i>	<i>Total Connection-Years, 00-10</i>	<i>Total Connection-Years, 90-10</i>
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share of Cities Below 6000 Miles	5.15** (2.09)	4.88** (2.28)	5.46*** (1.59)	-0.37 (1.52)	-0.46 (1.56)	11.23** (4.88)	17.78** (7.46)	28.37*** (10.89)
Observations	819	819	819	819	819	819	819	819
R-squared	0.01	0.06	0.61	0.00	0.02	0.94	0.86	0.91
<u>Controls</u>								
Number of Cities Around 6000 Miles	No	Yes	Yes	No	Yes	Yes	Yes	Yes
Region FE	No	Yes	Yes	No	Yes	Yes	Yes	Yes
Airport Controls, 1989	No	No	Yes	No	No	Yes	Yes	Yes
Dep. Var. Mean	2.83	2.83	2.83	1.44	1.44	18.04	24.14	40.42

A unit of observation is an airport/city. A connection is defined as weekly flights (at least) to and from the city. The outcome "Connected Cities" is the number of long-distance connections, as of 2014 (columns 1-3) or 1989 (4-5). Total connection-years in columns 6-8 is the sum of connections during the specified time period. Airport Controls, 1989 includes: Numbers 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; Eigenvalue centrality. Robust standard errors in parentheses, clustered at the country level. *** p<0.01, ** p<0.05, * p<0.1

Table 3. Effect on Night Lights (100-mile radius), Grid-Cell Level

Dependent Variable: Night Lights	1992	2010	92-10 Change	92-10 Change	92-10 Change	92-10 Change	92-10 Change	92-10 Change	92-10 Growth	92-10 Growth	92-10 Growth
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	RF	RF	RF	RF	RF	RF	2SLS	2SLS	RF	RF	2SLS
Share of Cities Below 6000 Miles	1.91	6.76**	4.85***	4.32***	4.45***	4.68***			0.430***	0.452***	
	(1.92)	(3.00)	(1.55)	(1.36)	(1.27)	(1.35)			(0.157)	(0.167)	
Total Connection-Years, 1990-2010							0.140**	0.143**			0.014*
							(0.069)	(0.073)			(0.008)
Anderson-Rubin p-value							0.001	0.001			0.007
Stock-Wright p-value							0.004	0.004			0.019
Angrist-Pischke first stage F-stat							6.10	5.65			5.65
Airports	777	777	777	777	777	734	777	734	777	734	734
Observations	39,487	39,487	39,487	39,487	39,355	37,766	39,487	37,766	39,487	37,766	37,766
R-squared	0.27	0.30	0.22	0.30	0.32	0.32	NA	NA	0.170	0.203	NA
<u>Controls</u>											
Number of Cities Around 6000 Miles	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	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	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic Controls	No	No	No	No	No	Yes	No	Yes	No	Yes	Yes
Country Real GDPpc, 1990 logs	No	No	No	No	No	Yes	No	Yes	No	Yes	Yes

A unit of observation is a grid cell, within 100 miles of the closest airport in the country. The independent variables refer to the nearest airport. A connection is defined as weekly flights (at least) to and from the city. Total connection-years is the sum of long-distance connections during 1990-2010. All regressions include the main effect of Log Distance to the closest airport in the country. Airport Controls, 1989 includes: Numbers of daily, twice-weekly, and weekly flights, 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, Eigenvalue centrality. Geographic controls: Average yearly precipitation 1980-2014, Average yearly temperature, 1980-2014. Robust standard errors in parentheses, clustered at the country level. Anderson-Rubin p-value refers to the weak instrument robust inference using the Anderson-Rubin Wald test (F-stat version), and the Stock-Wright p-value comes from the LM S statistic. *** p<0.01, ** p<0.05, * p<0.1.

Table 4. Effect on Night Lights, Spatial Patterns, Grid-Cell Level

Dependent Variable: Night Lights	92-10 Change	92-10 Change	92-10 Change	92-10 Change	92-10 Change	92-10 Change	92-10 Growth	92-10 Growth	92-10 Growth
Sample (miles from closest airport)	All	< 500	< 500	<100	100-250	250-500	<100	250-500	250-500
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Share of Cities Below 6000 Miles	9.94*** (3.78)	13.36*** (4.26)	13.31*** (4.40)	4.68*** (1.35)	0.53 (0.32)	-0.36 (0.30)	0.45*** (0.17)	0.14 (0.09)	-0.06 (0.08)
Share of Cities Below 6000 Miles X Log Distance	-1.65** (0.65)	-2.36*** (0.76)	-2.29*** (0.79)						
Airports	777	777	734	734	734	734	734	734	734
Observations	229,382	166,012	157,223	37,766	61,513	57,944	37,766	61,513	57,944
R-squared	0.34	0.35	0.36	0.85	0.84	0.73	0.20	0.18	0.11
<u>Controls</u>									
Number of Cities Around 6000 Miles	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Night Light in 1992	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Population in 1990	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Airport Controls, 1989	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic Controls	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country Real GDPpc, 1990 logs	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

A unit of observation is a grid cell. The independent variables refer to the nearest airport in the same country. The samples vary depending on the distance from the centroid of the cell to the nearest airport, as denoted above. All regressions include the main effect of Log Distance to the closest airport in the country. All other variables are defined the same as in in Table 3. Robust standard errors in parentheses, clustered at the country level. *** p<0.01, ** p<0.05, * p<0.1.

Table 5. Effect on Population, Spatial Pattern, Grid-Cell Level

Dependent Variable	<i>Pop. Change 1990-10</i>	<i>Pop. Change 1990-10</i>	<i>Pop. Change 1990-10</i>	<i>Pop. Change 1990-10</i>	<i>Pop. Change 1990-10</i>	<i>Pop. Growth 1990-10</i>	<i>Pop. Growth 1990-10</i>
Sample (miles from closest airport)	<100	100-250	250-500	<500	All	<500	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Share of Cities Below 6000 Miles	-32.39 (22.57)	-13.30 (9.19)	0.08 (1.91)	16.36 (77.73)	-16.08 (53.38)	-0.57 (0.39)	0.10 (0.35)
Share of Cities Below 6000 Miles X Log Distance				-5.95 (15.24)	1.13 (9.56)	0.06 (0.08)	-0.08 (0.07)
Airports	734	734	734	734	734	734	734
Observations	37,760	61,468	57,897	157,125	218,713	157,125	218,713
R-squared	0.54	0.46	0.71	0.54	0.54	0.18	0.20
<u>Controls</u>							
Number of Cities Around 6000 Miles	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Night Light in 1992	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Population in 1990	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline Controls, 1989	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes

A unit of observation is a grid cell. The independent variables refer to the nearest airport. Population change is measured in thousands. All regressions include the main effect of Log Distance to the closest airport in the country. All other controls are defined the same as in in Table 3. Robust standard errors in parentheses, clustered at the country level. *** p<0.01, ** p<0.05, * p<0.1.

Table 6. Network Spillovers, 2014, Airport Level

Dependent Variable:	Dep. Var. Number of Connected Cities, by Distance						Network Centrality, Eigenvector		Total Passengers, Millions	
	5500-6500 miles		2000-5500 miles		>6500 miles		RF	2SLS	RF	2SLS
	RF	2SLS	RF	2SLS	RF	2SLS				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Share of Cities Below 6000 Miles	0.70*** (0.25)		4.05*** (1.15)		0.45** (0.20)		0.022*** (0.005)		0.719*** (0.246)	
Number of Connected Cities, 5500-6500 miles		7.75*** (2.07)		5.82*** (1.80)		0.65** (0.29)		0.031*** (0.008)		
Number of Connected Cities										0.133*** (0.019)
Anderson-Rubin p-value		0.0004		0.0005		0.0292		0.0001		0.0038
Stock-Wright p-value		0.0009		0.0009		0.0311		0.0004		0.0047
Angrist-Pischke first stage F-stat		7.78		7.78		7.78		7.78		7.78
Observations	819	819	819	819	819	819	819	819	819	819
R-squared	0.69	NA	0.58	NA	0.51	NA	0.66	NA	0.75	NA

A unit of observation is an airport/city. All regressions control for Number of Cities Around 6000 Miles, Region fixed effects, and Airport Controls, 1989. Airport Controls, 1989 includes: Numbers of daily, twice-weekly, and weekly flights (total and 2000-5500, 5500-6500, and above 6500 ranges), 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, Eigenvector centrality. Robust standard errors in parentheses, clustered at the country level. Anderson-Rubin p-value refers to the weak instrument robust inference using the Anderson-Rubin Wald test (F-stat version). *** p<0.01, ** p<0.05, * p<0.1 (using weak instrument robust inference when applicable).

Table 7. Effect on Number of Cross-Owned Companies, Airport-Pair Level

Independent variable:	Dep. Var.: Number of Cross-Owned Companies							
	Below 6000 Miles	Below 6000 Miles	Below 6000 Miles	Below 6000 Miles	Below 6000 Miles	Below 6000 Miles	Weekly Flight, 2014	Weekly Flight, 2014
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
RD Estimate	-0.604	-0.922	-0.899	-0.890	-0.926	-0.987	257.3	291.8
Robust S.E.	(0.123)***	(0.197)***	(0.244)***	(0.330)***	(0.221)***	(0.220)***	(69.04)***	(78.83)***
Cluster S.E.	(0.212)***	(0.301)***	(0.353)**	(0.432)**	(0.326)***	(0.339)***	(82.70)***	(89.76)***
First Stage							-0.0036*** (0.0010)	-0.0034*** (0.0009)
RD Design	Sharp	Sharp	Sharp	Sharp	Sharp	Sharp	Fuzzy	Fuzzy
Bandwidth (miles)	2000	1000	750	500	Optimal	Optimal	Optimal	Optimal
Polynomial order	1st	1st	1st	1st	1st	2nd	1st	2nd
Observations, Total	334954	334954	334954	334954	334954	334954	334954	334954
Observations, Effective	146,682	81,842	61,688	41,477	70,366	134,799	67,986	135,505

Local polynomial Regression Discontinuity estimates, using rdrobust command in Stata (default options unless otherwise stated). Running variable is distance between airports, discontinuity at 6000 miles. Optimal bandwidth selected using one common MSE-optimal bandwidth selector. "Polynomial order" refers to order of the polynomial in distance. The dependent variable is "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. "Weekly Flight" is a dummy equal to one if there was an at least weekly flight between the two airports in 2014. Standard errors in parentheses. "Cluster S.E." refers to cluster-robust nearest neighbor variance estimation at the country-pair level, using a minimum of three nearest neighbors. The covariates are all measured in 1989 and consist of: a dummy for having weekly flights, 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<0.01, ** p<0.05, * p<0.1

Table 8. Effect on Number of Companies Owned by City in Richer and Poorer Country, Airport-Pair Level

Sample Countries:	Dep. Var.: Number of Cross-Owned Companies							
	Pairs: All		Pairs: High, Non-High		Pairs: High, Middle		Pairs: Low, Non-Low	
<i>Outcome, Direction of the FDI:</i>	<i>Richer -> Poorer</i>	<i>Poorer -> Richer</i>	<i>Richer -> Poorer</i>	<i>Poorer -> Richer</i>	<i>Richer -> Poorer</i>	<i>Poorer -> Richer</i>	<i>Richer -> Poorer</i>	<i>Poorer -> Richer</i>
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: First Stage								
Below 6000 Miles, Dummy	0.0017** (0.0008)	0.0017** (0.0008)	0.0021** (0.0010)	0.0021** (0.0010)	0.0019** (0.0009)	0.0019** (0.0009)	0.0008 (0.0034)	0.0008 (0.0034)
Panel B: 2SLS								
Connected 2014, Dummy	138.0** (63.68)	45.78 (28.58)	131.1** (60.10)	43.19 (28.17)	134.3** (60.73)	43.11 (27.06)	17.26 (71.05)	-3.234 (8.244)
Observations	179,355	179,355	138,240	138,240	153,868	153,868	25,487	25,487

Local polynomial Regression Discontinuity estimates, using triangular kernel function, optimal bandwidth (selected using one common MSE-optimal bandwidth selector), and 1st order polynomial in the running variable. Running variable is distance between airports, discontinuity at 6000 miles. The sample in all columns is restricted to airport pairs where the two countries are not in the same World Bank income classification ("High," "Upper Middle," "Lower Middle," "Low"), as of 2016. (Argentina is "Not classified," but for our purposes we re-classify it as "Upper Middle.") The dependent variable is "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. "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 2011, 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". Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 9. Effect on Number of Companies Owned Abroad and by Foreigners, Grid-Cell Level

Dependent Variable:	Owned by Foreigners		Owned by Foreigners, IV		Owned Abroad		Owned Abroad, IV		Owned by Foreigners		Owned by Foreigners		Owned by Foreigners		Owned by Foreigners		Owned Abroad		Owned Abroad		Owned Abroad		Owned Abroad	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
Share of Cities Below 6000 Miles	3.543*** (1.263)		3.685** (1.863)		2.169** (0.965)	3.001** (1.179)	5.059*** (1.537)	-0.234 (0.982)	0.193 (1.768)	2.452 (2.257)	4.510** (1.978)	0.209 (1.602)												
Number of Connected Cities, 2014		0.024*** (0.003)		0.025*** (0.005)																				
First-Stage F-statistic		8.76		8.76																				
Distance (miles)	All	All	All	All	1000-3000	3000-5000	5000-6000	> 6000	1000-3000	3000-5000	5000-6000	> 6000												
Airports	777	777	777	777	777	777	777	777	777	777	777	777												
Observations	39,487	39,487	39,487	39,487	39,487	39,487	39,487	39,487	39,487	39,487	39,487	39,487												

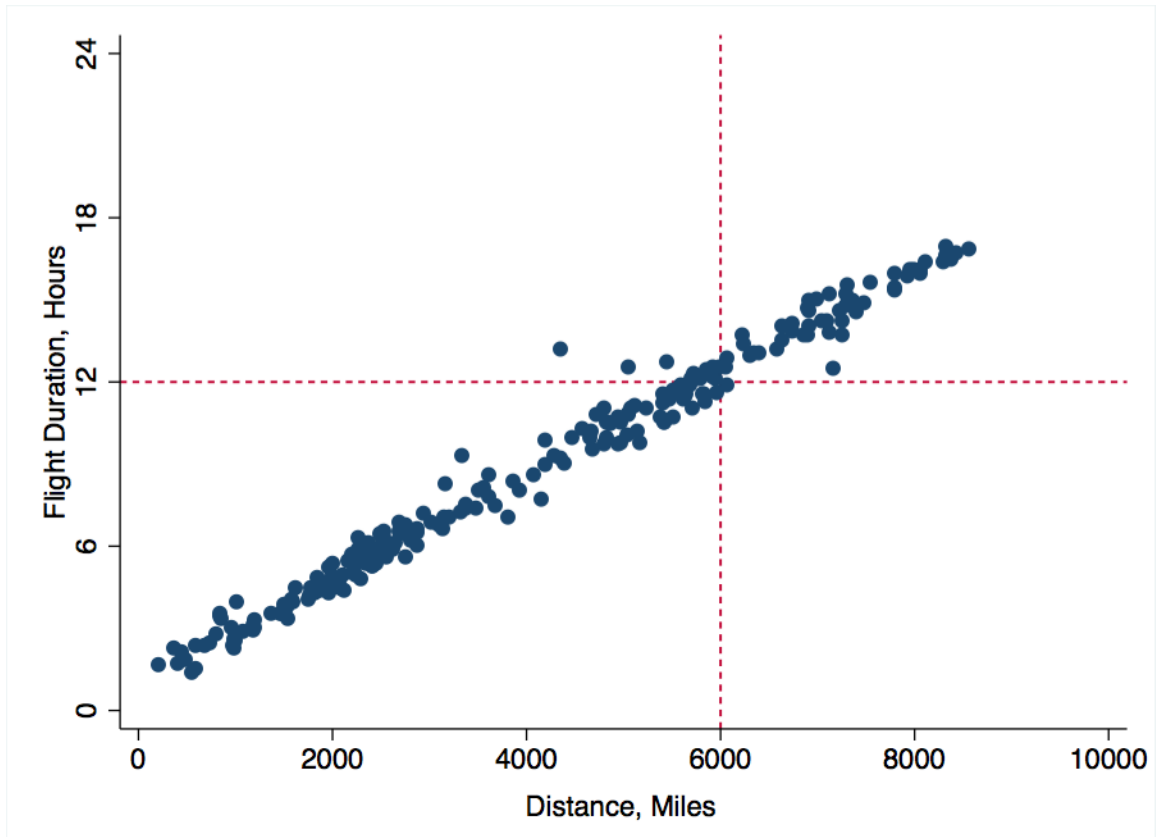
Poisson estimates (IV estimated using GMM). A unit of observation is a grid cell, within 100 miles of the closest airport in the country. All regressions include the main effect of Log Distance to the closest airport in the country, Number of Cities Around 6000 miles, Region fixed effects, Night Lights in 1992, Population in 1990. "Owned by Foreigners" is the number of companies in the cell that are owned by individual/firm located in another country within a 100-mile radius of an airport. "Owned Abroad" is the number of companies within 100 miles of some other airport in our sample that are owned by individual/firm located in the cell. "Distance (miles)" refers to the sample used in calculating the number of firms: for instance, "Owned Abroad, 1000-3000" counts the number of companies within 100 miles of some other airport in our sample that is between 1000 and 3000 miles away from the airport in question. Robust standard errors in parentheses, clustered at the country level. *** p<0.01, ** p<0.05, * p<0.1

**“Long-Range Growth:
Economic Development in the Global Network of Air Links”**

Filipe Campante and David Yanagizawa-Drott

ONLINE APPENDIX

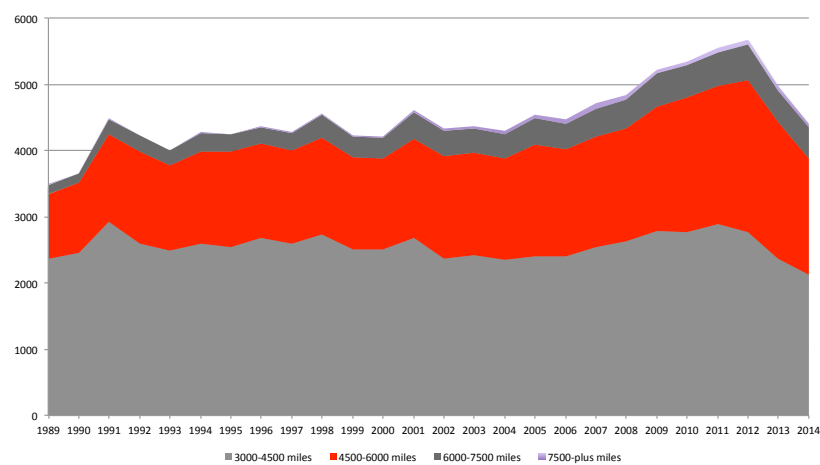
Figure A1. Distance and Scheduled Flight Duration



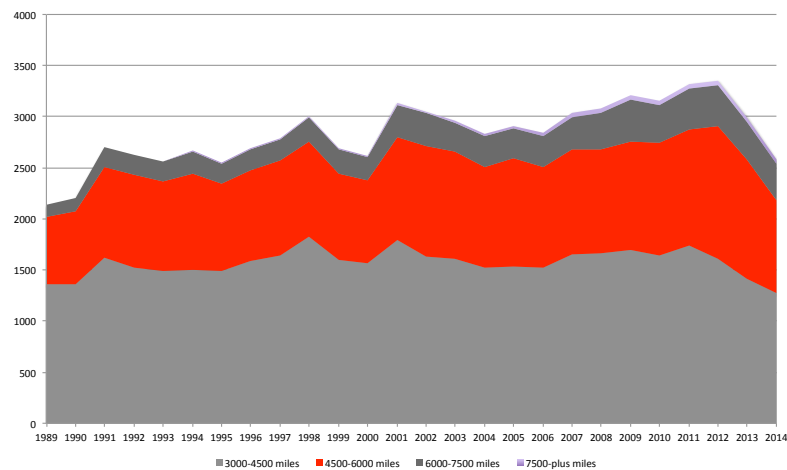
Notes: Data for longest flight for each airline, as of 2016, available from Wikipedia https://en.wikipedia.org/wiki/Non-stop_flight#By_airline_companies_28currently_scheduled.29)

Figure A2. Total Number of Long-Haul Connections, by Year and Distance

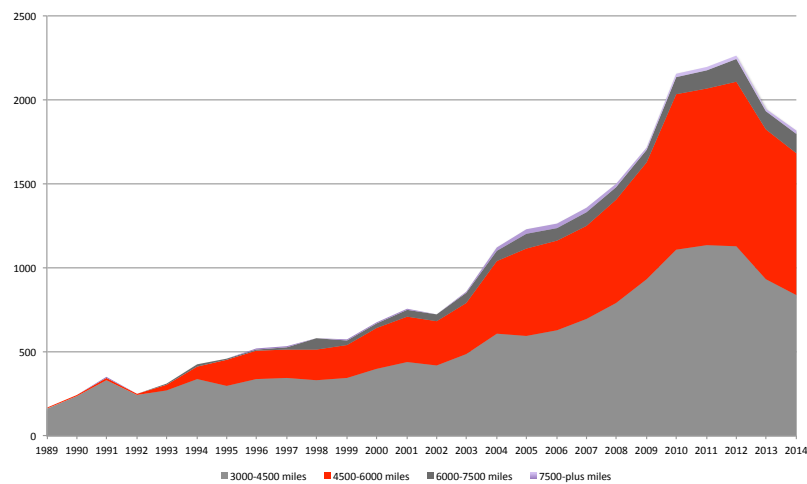
A: Total



B: Boeing



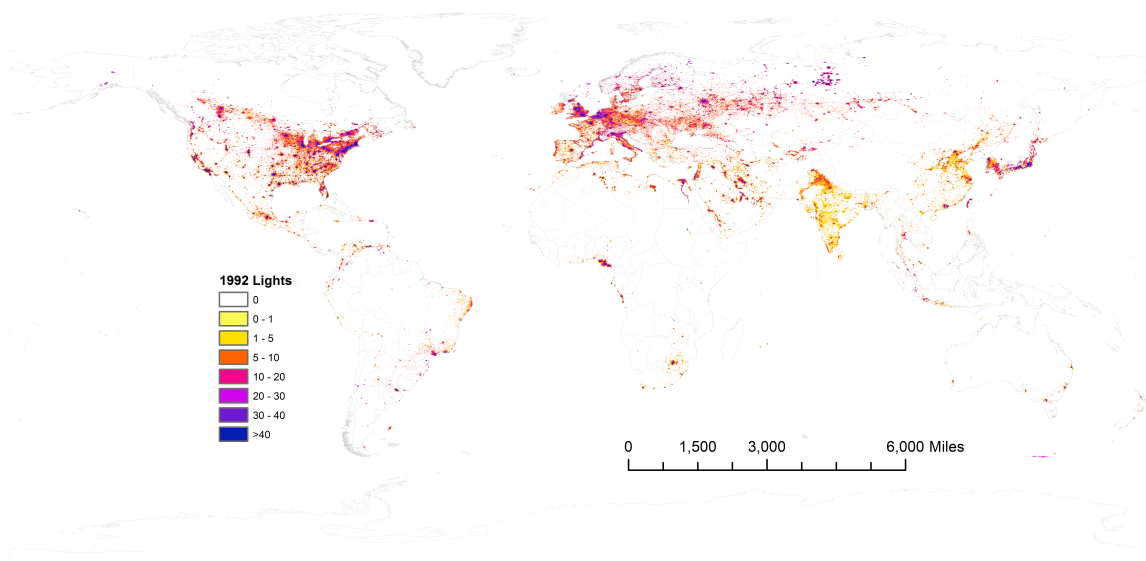
C: Airbus



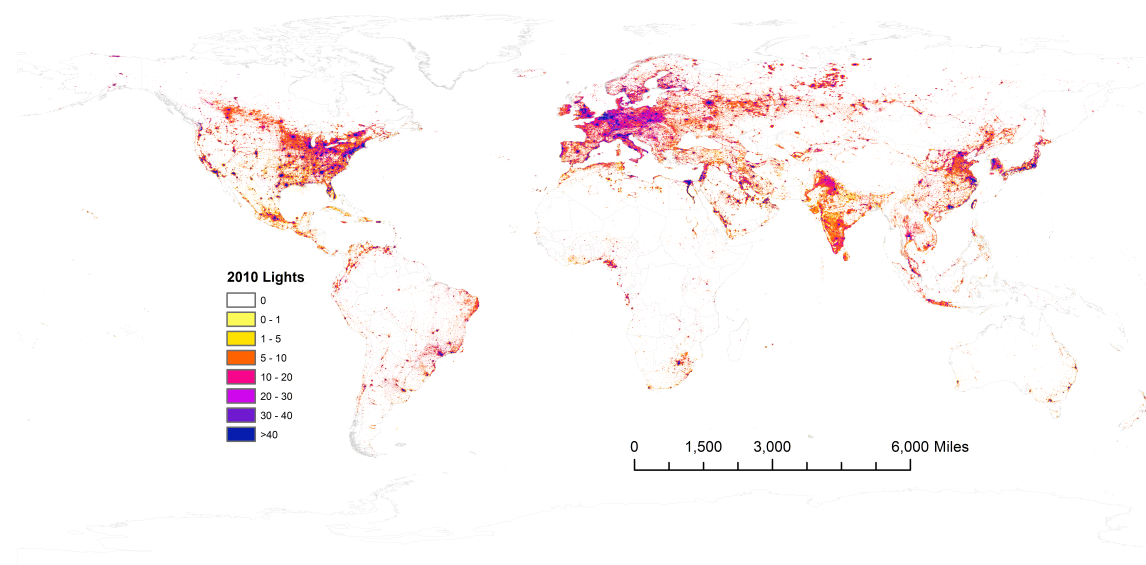
Notes: Data from ICAO, see data description in main text.

Figure A3. Geographical Distribution of Satellite Night Lights, 1992 and 2010

A: 1992



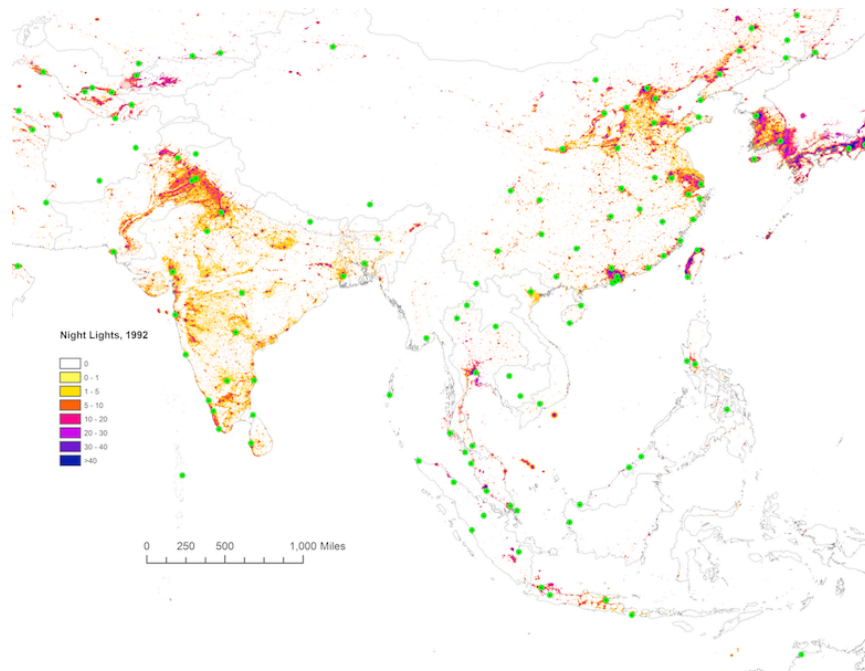
B: 2010



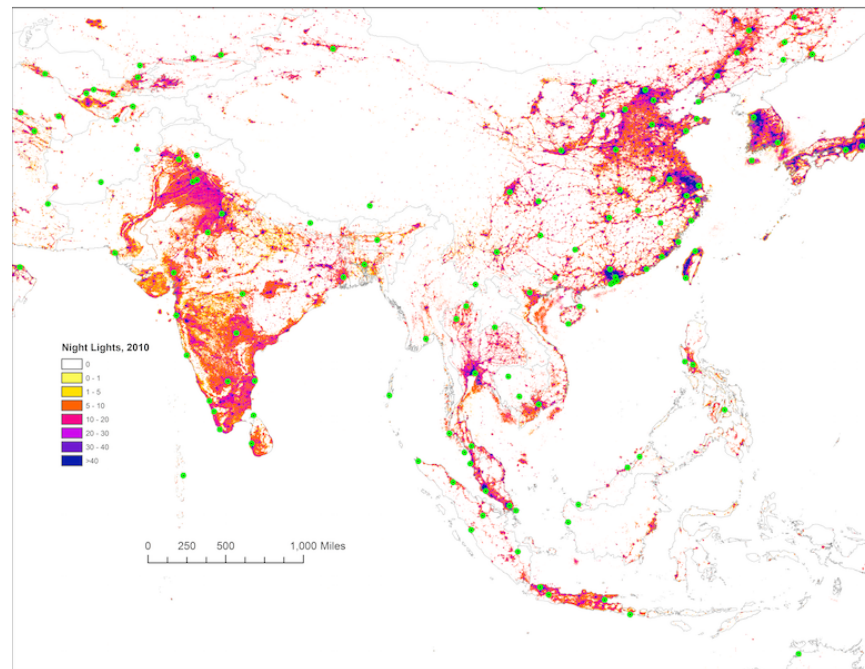
Notes: Average satellite night lights at the 0.25x0.25-degree level, 1992 and 2010.

Figure A4. Geographical Distribution of Satellite Night Lights and Major Airports in Asia, 1992 and 2010

A: 1992



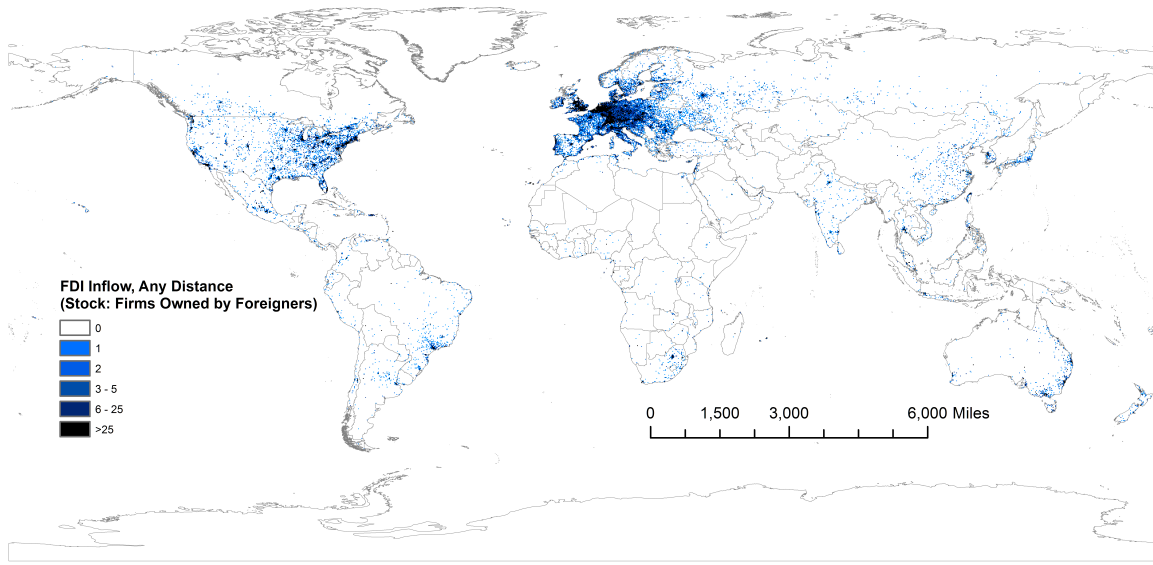
B: 2010



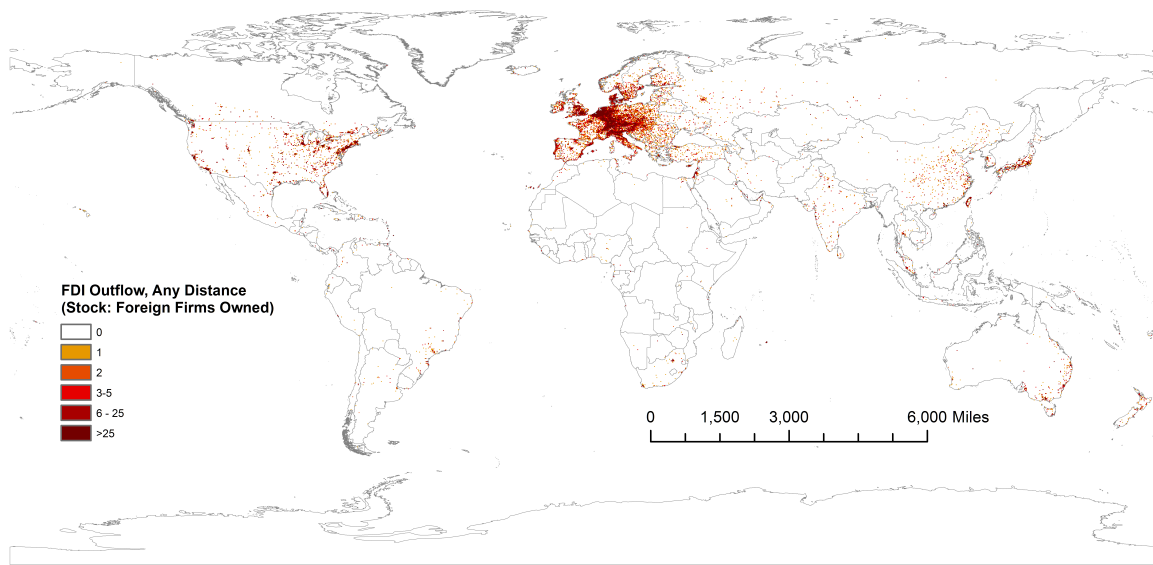
Notes: Average satellite night lights at the 0.25x0.25-degree level in Asia, 1992 and 2010. We also add the airports from our sample of 819 ICAO major airports.

Figure A5. Geographical Distribution of Firms Owned by Foreigners and Owned Abroad (Orbis)

A: Firms Owned by Foreigners



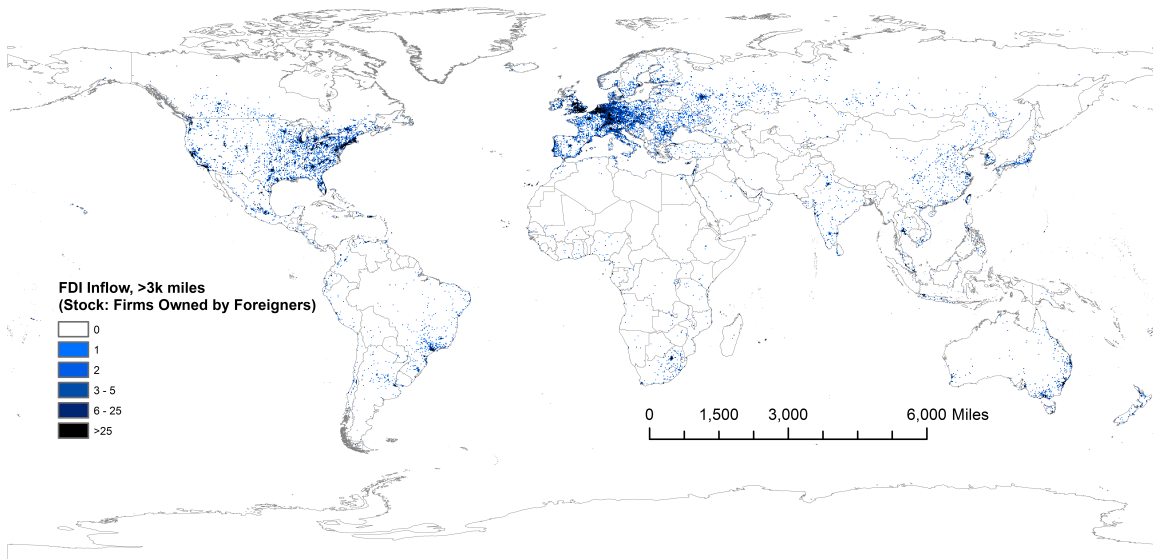
B: Firms Owned Abroad



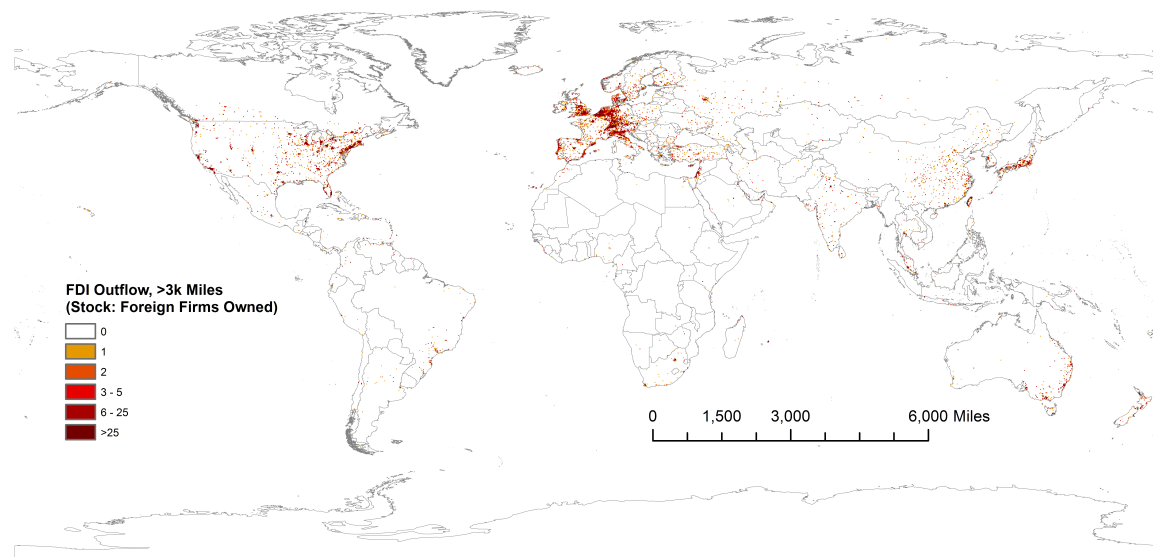
Notes: Panel A: For each 0.25x0.25-degree cell, we depict the number of firms located in that cell that are reported in the Orbis dataset as having ultimate owners located in a different country. Panel B: For each 0.25x0.25-degree cell, we depict the number of firms with foreign owners, anywhere in the world, that are reported in the Orbis dataset as having ultimate owners located in that cell.

Figure A6. Geographical Distribution of Firms Owned by Foreigners and Owned Abroad (Orbis), at least 3000 miles of distance

A: Firms Owned by Foreigners



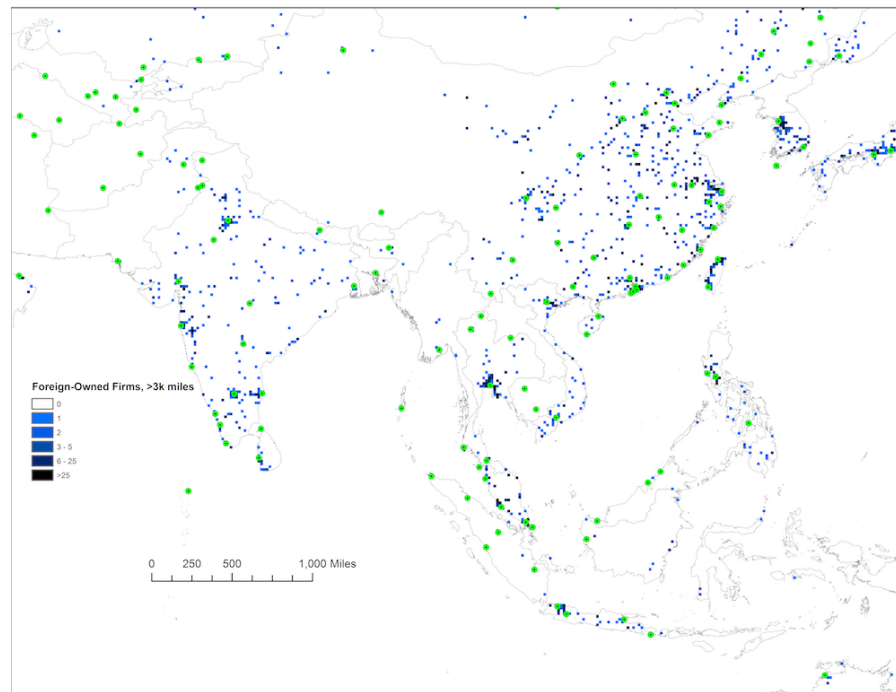
B: Firms Owned Abroad



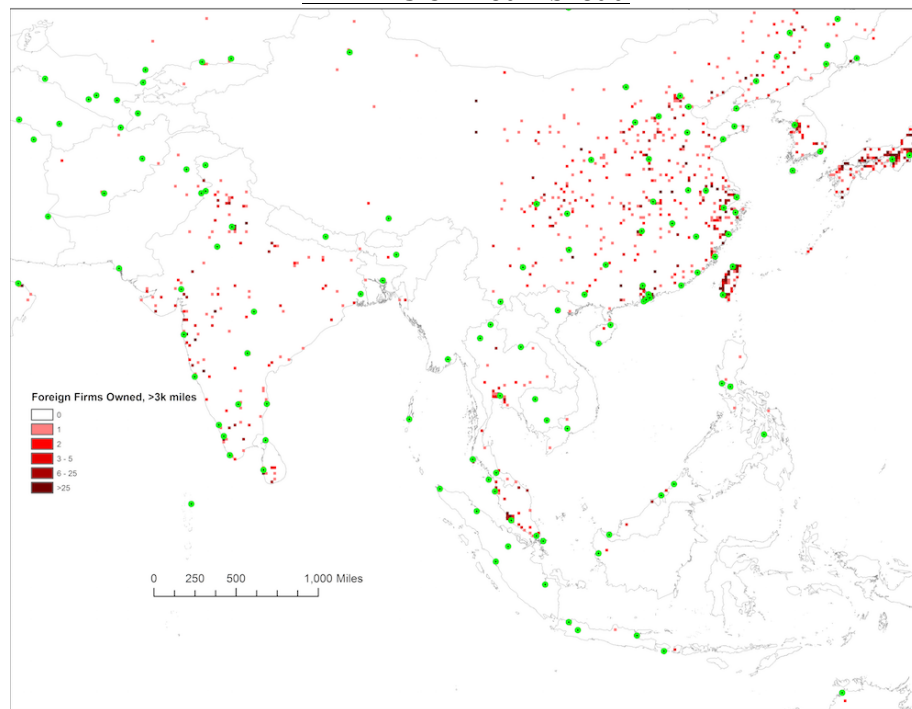
Notes: Panel A: For each 0.25x0.25-degree cell, we depict the number of firms located in that cell that are reported in the Orbis dataset as having ultimate owners located in a different country, and at least 3000 miles away. Panel B: For each 0.25x0.25-degree cell, we depict the number of firms with foreign owners that are located at least 3000 miles away from that cell, and which are reported in the Orbis dataset as having ultimate owners located in that cell.

Figure A7. Geographical Distribution of Firms Owned by Foreigners and Owned Abroad (Orbis) in Asia, at least 3000 miles of distance

A: Firms Owned by Foreigners



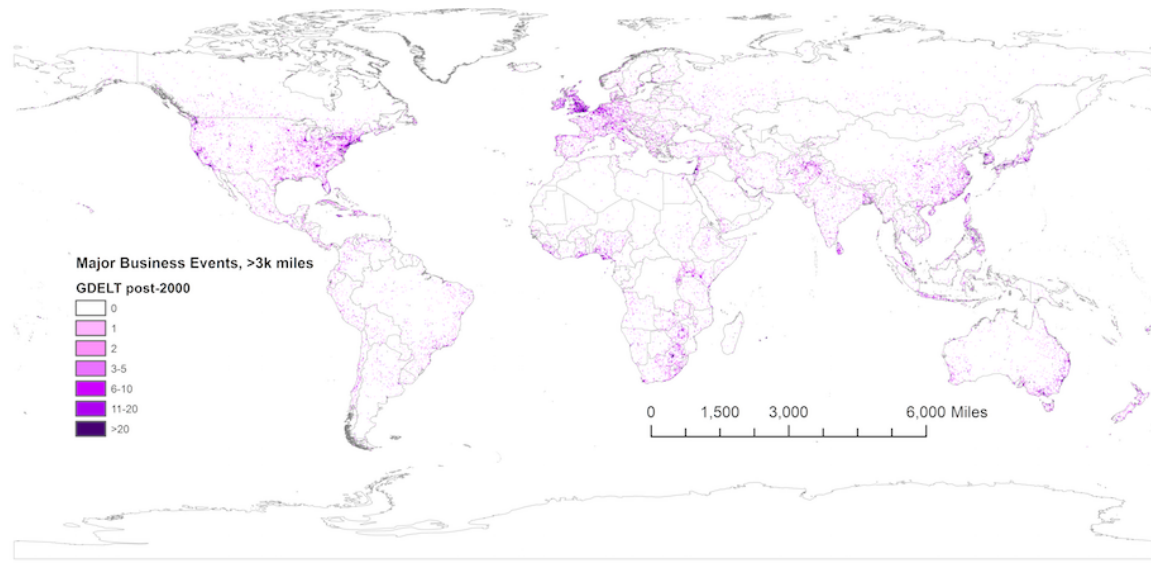
B: Firms Owned Abroad



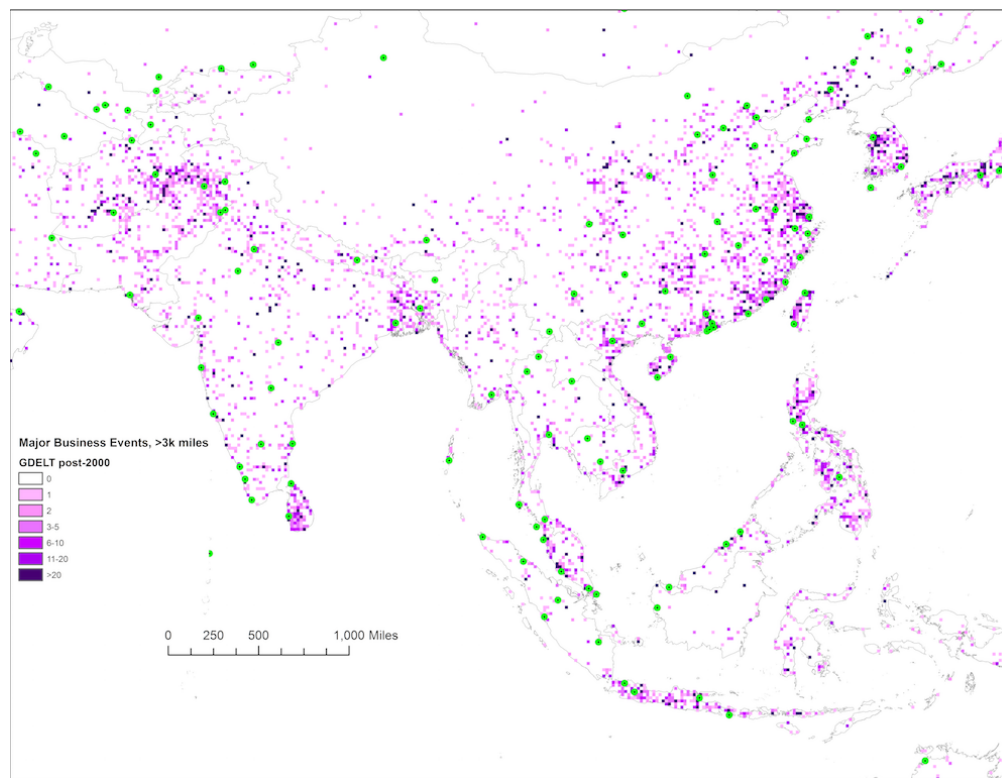
Notes: Same as in Figure A6, but restricted to Asia and adding the airports from our sample of 819 ICAO major airports.

**Figure A8. Geographical Distribution of Business Cooperation Events (GDELT),
at least 3000 miles of distance**

A: World

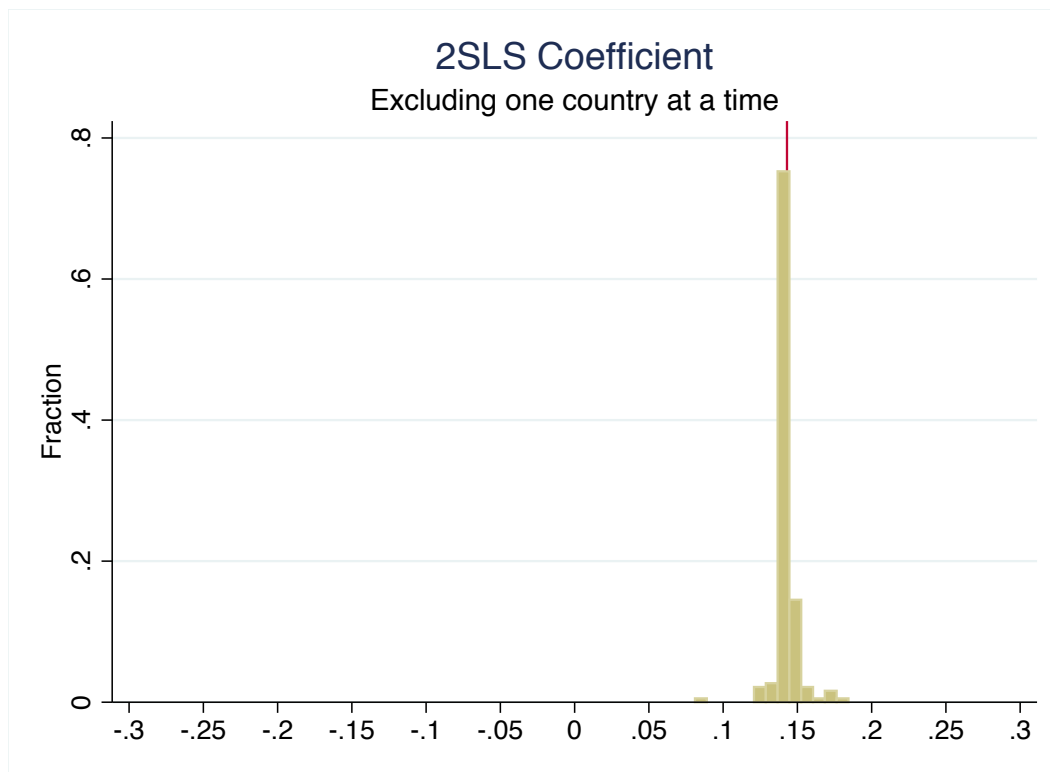


B: Asia



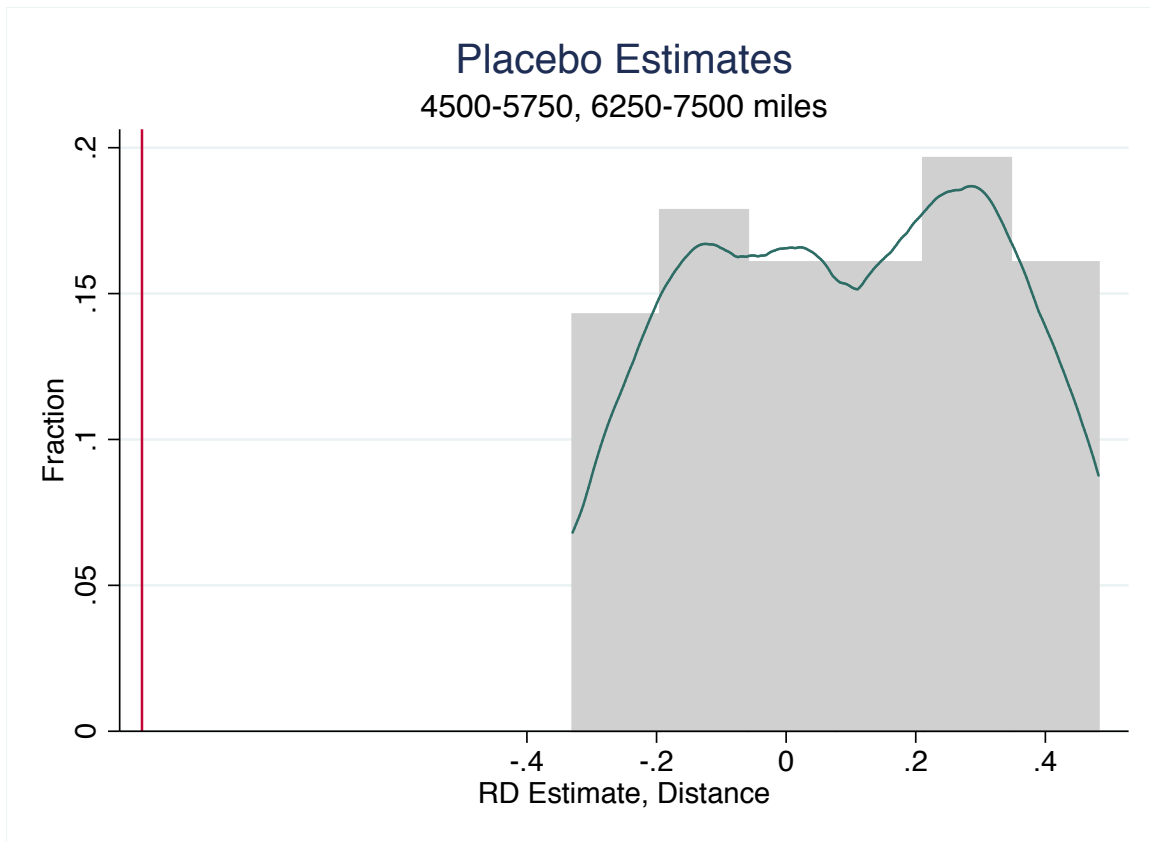
Notes: For each 0.25x0.25-degree cell, we depict the number of major business cooperation events involving one party located in that cell and another party located more than 3000 miles away, as reported in the GDELT dataset. Panel B presents Asia, and adds the airports from our sample of 819 ICAO major airports.

**Figure A9. 2SLS Coefficient on Connection-Years,
Excluding One Country at a Time**



Notes: The picture depicts estimates obtained running the specification in Table 3 (Column 8), excluding each country in the sample at a time. Dependent variable is the change in night lights 1992-2010, independent variable is total connection-years 1990-2010, full set of control variables. The vertical red line depicts the full-sample coefficient. The plot shows that estimates are not very sensitive, with the possible exception of China (left outlier). We present the estimates excluding China in Table A7 below.

Figure A10. Placebo Regression Discontinuity Estimates for Presence of Ownership Links between City Pairs (Orbis)



Notes: Histogram (and kernel approximation) for regression discontinuity estimates computed using each 50-mile point between 4500 and 5750 miles, and between 6250 and 7500 miles as distance thresholds. Specifications use first-order polynomial and optimal bandwidth. Vertical line depicts estimate with 6000-mile threshold. The plot shows that the estimate at the 6000-mile threshold is a clear outlier.

Table A1. Descriptive Statistics

Variable	Observations	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
		<u>2014</u>				<u>1989</u>			
<u>City Level</u>									
Share of Cities Below 6000 Miles	819	0.556	0.118	0.136	0.79				
Number of Cities 5500-6500 Miles	819	102.15	58.15	7	301				
Total Connection-Years, 1990-2010	819	40.4	134.3	0	1627				
East Asia & Pacific	819	0.145	0.353	0	1				
Latin America & Caribbean	819	0.149	0.357	0	1				
Middle East & North Africa	819	0.088	0.283	0	1				
North America	819	0.05	0.218	0	1				
South Asia	819	0.033	0.179	0	1				
Sub-Saharan Africa	819	0.101	0.302	0	1				
Europe & Central Asia	819	0.433	0.496	0	1				
Number of Weekly Connections	819	8.76	18.87	0	186	5.02	12.72	0	145
Number of Twice-Weekly Connections	819	7.59	17.25	0	177	3.86	10.62	0	128
Number of Daily Connections	819	3.83	10.15	0	100	1.75	5.93	0	84
Weekly Connections, 5500-6000 Miles	819	0.355	1.31	0	15	0.099	0.691	0	10
Weekly Connections, 3000-5500 Miles	819	1.35	4.47	0	52	0.911	3.39	0	45
Weekly Connections, 6000-10000 Miles	819	0.264	1.22	0	12	0.06	0.402	0	6
Eigenvector Centrality (standardized)	819	-0.014	0.03	-0.03	0.197	-0.018	0.032	-0.032	0.222
Total Number of Connections (Cities)	819	12.5	24.01	0	249	7.45	16.71	0	177
Total Number of Connections (Countries)	819	9.97	21.93	0	223	4.84	13.34	0	147
Total Number of Seats (millions), year	819	0.851	2.54	0	27.63	0.432	1.67	0	28.29
Total Number of Flights, year	819	5071.45	14210.75	0	151583	2497.45	9495.67	0	162771
Total Number of Passengers (millions), year	819	0.647	1.96	0	21.23	0.273	1.09	0	18.53
<u>City-Pair Level</u>									
Weekly Connection (Dummy)	334954	0.01	0.101	0	1	0.006	0.075	0	1
Distance between Airports (miles)	334954	4463.27	2802.86	0.045	12388				
Number of Cross-Owned Companies (2016)	334954	1.39	18.89	0	3182				
Total City Pairs within Country Pair	334954	248.54	400.13	1	1980				
Total Connections within Country Pair	334954	1.92	6.89	0	102	1.29	4.51	0	51
Total Passengers between Country Pair (millions), year	334954	0.315	1.22	0	16.4	0.194	0.834	0	8.26

Notes: All variables computed from ICAO data. Data on connections, flights, seats, passengers refer to international flights.

Table A1. Descriptive Statistics (Continued)

Variable	Observations	Mean	S.D.	Min	Max
<u>Grid-Cell Level</u>					
Night Lights, 2010	229382	2.11	5.95	0	63
Night Lights, 1992	229382	1.08	4.04	0	63
Night Lights Growth, 1992-2010	229382	0.198	0.398	-2.53	3.87
Companies Owned by Foreigners, 2016	229382	2.28	75.67	0	19760
Companies Owned Abroad, 2016	229382	2.28	102.75	0	20179
Distance to Closest Airport (miles)	229382	418.45	400.37	0.012	2329.28
Population, 2010 (thousands)	229139	46.24	235.48	0	25299.13
Population, 1990 (thousands)	229382	21.84	98.7	0	8899.5
Distance to Border (miles)	229151	251.82	257.67	6.71E-06	1319.89
Distance to Capital (miles)	229151	1801.39	1570.57	1.77	7884.23
Average Precipitation, 1980-2014 (yearly, mm)	229151	787.27	655.96	4.14	4962.12
Average Temperature, 1980-2014 (yearly, Celsius)	229029	10.45	13.56	-18.12	45.12
Business Events, 2000s	229382	6.22	335.11	0	94442
Business Events, 1980s	229382	0.183	12.4	0	3361
Business Connections, 2000s	229382	0.904	22.42	0	4454
Business Connections, 1980s	229382	0.061	2.58	0	633
Large Company Headquarters, 2006	233110	0.006	0.54	0	189
Large Company Headquarters, 2012	233110	0.009	0.517	0	154

Notes: See sources and description in the main text.

Table A2. Cities in Airport Sample (Top and Bottom 50, by ShareBelow6K)

City Name	City Code (ICAO)	ShareBelow6K	Total Number of Cities, 5500-6500
ABADAN (IRAN (ISLAMIC REPUBLIC OF))	ABD	0.1363636	22
NOUMEA (NEW CALEDONIA)	NOU	0.1428571	21
KUWAIT (KUWAIT)	KWI	0.15	20
BAHRAIN (BAHRAIN)	BAH	0.2	15
PISCO (PERU)	PIO	0.2105263	152
ALICE SPRINGS (AUSTRALIA)	ASP	0.2156863	51
SALALAH (OMAN)	SLL	0.2222222	9
ANTA (PERU)	ATA	0.2299465	187
DOHA (QATAR)	DOH	0.2307692	13
LIMA (PERU)	LIM	0.2322581	155
TRUJILLO (PERU)	TRU	0.2417582	182
ASUNCION (PARAGUAY)	ASU	0.2435233	193
IGUAZU (ARGENTINA)	IGR	0.2441315	213
PORTO ALEGRE (BRAZIL)	POA	0.2485876	177
ASHKHABAD (TURKMENISTAN)	ASB	0.25	24
YEKATERINBURG (RUSSIAN FEDERATION)	SVX	0.2542373	59
BUKHARA (UZBEKISTAN)	BHK	0.2592593	27
LA PAZ (BOLIVIA)	LPB	0.265	200
YEREVAN (ARMENIA)	EVN	0.2653061	49
GYANDZHA (AZERBAIJAN)	KVD	0.2666667	45
DHAHRAN (SAUDI ARABIA)	DHA	0.2666667	15
DAMMAM (SAUDI ARABIA)	DMM	0.2666667	15
CHICLAYO (PERU)	CIX	0.2741936	186
PORT VILA (VANUATU)	VLI	0.28	25
ALEPPO (SYRIAN ARAB REPUBLIC)	ALP	0.28125	64
SAIPAN (SAIPAN (MARIANA ISLANDS))	SPN	0.2857143	63
EILAT (ISRAEL)	ETH	0.2903226	62
FLORIANOPOLIS (BRAZIL)	FLN	0.2913043	230
NAKHICHEVAN (AZERBAIJAN)	NAJ	0.2916667	48
MASHAD (IRAN (ISLAMIC REPUBLIC OF))	MHD	0.2941177	17
RIYADH (SAUDI ARABIA)	RUH	0.2941177	17
MARY (TURKMENISTAN)	MYP	0.2941177	17
TBILISI (GEORGIA)	TBS	0.2962963	54
AMMAN (JORDAN)	AMM	0.296875	64
COCHABAMBA (BOLIVIA)	CBB	0.2970297	202
KUTAI SI (GEORGIA)	KUT	0.2985075	67
CAJAMARCA (PERU)	CJA	0.3010204	196
CORRIENTES (ARGENTINA)	CNQ	0.3037037	135
SAMARA (RUSSIAN FEDERATION)	KUF	0.3055556	72
OVDA (ISRAEL)	VDA	0.3064516	62
BATUMI (GEORGIA)	BUS	0.3076923	65
RESISTENCIA (ARGENTINA)	RES	0.3100775	129
TEHRAN (IRAN (ISLAMIC REPUBLIC OF))	THR	0.3103448	29
JUJUY (ARGENTINA)	JUJ	0.3109244	119
DAMASCUS (SYRIAN ARAB REPUBLIC)	DAM	0.3114754	61
GYUMRI (ARMENIA)	LWN	0.3148148	54
GUAM ISLAND (GUAM)	GUM	0.3157895	57
TRABZON (TURKEY)	TZX	0.3181818	66
BEIRUT (LEBANON)	BEY	0.328125	64
BANDUNG (INDONESIA)	BDO	0.3298969	97

Table A2. Cities in Airport Sample (Top and Bottom 50, by ShareBelow6K), Continued

City Name	City Code (ICAO)	ShareBelow6K	Total Number of Cities, 5500-6500
KIMBERLEY (SOUTH AFRICA)	KIM	0.7112299	187
BILLUND (DENMARK)	BLL	0.7125	80
MANZINI (SWAZILAND)	MTS	0.7125748	167
EINDHOVEN (NETHERLANDS)	EIN	0.7128713	101
BUSAN (REPUBLIC OF KOREA)	PUS	0.7142857	224
NUREMBERG (GERMANY)	NUE	0.7142857	91
NADI (FIJI)	NAN	0.7142857	49
SALVADOR (BRAZIL)	SSA	0.7142857	133
REIMS (FRANCE)	RHE	0.7156863	102
MAASTRICHT (NETHERLANDS)	MST	0.7156863	102
BLOEMFONTEIN (SOUTH AFRICA)	BFN	0.715847	183
MAPUTO (MOZAMBIQUE)	MPM	0.7169811	159
TAN TAN (MOROCCO)	TTA	0.7179487	39
LEIPZIG (GERMANY)	LEJ	0.7209302	86
BONAIRE (NETHERLANDS ANTILLES)	BON	0.7222222	90
SANA'A (YEMEN)	SAH	0.7222222	18
AGADIR (MOROCCO)	AGA	0.7234042	47
CHIANG RAI (THAILAND)	CEI	0.725	160
HANOVER (GERMANY)	HAI	0.7252747	91
COLOMBO (SRI LANKA)	CMB	0.728395	81
LAHORE (PAKISTAN)	LHE	0.7307692	26
CHIANG MAI (THAILAND)	CNX	0.73125	160
HAMBURG (GERMANY)	HAM	0.7325581	86
ACCRA (GHANA)	ACC	0.7333333	45
VIENTIANE (LAO PEOPLE'S DEM. REP.)	VTE	0.7339901	203
CURACAO (NETHERLANDS ANTILLES)	CUR	0.7346939	98
ESSAOUIRA (MOROCCO)	ESU	0.7346939	49
WUYISHAN (CHINA)	WUS	0.7348837	215
LANZAROTE (SPAIN)	ACE	0.7368421	38
AMRITSAR (INDIA)	ATQ	0.7407407	27
WUHAN (CHINA)	WUH	0.7410072	139
NINGBO (CHINA)	NGB	0.7442923	219
OUAGADOUGOU (BURKINA FASO)	OUA	0.75	44
DHAKA (BANGLADESH)	DAC	0.75	68
HANOI (VIET NAM)	HAN	0.7526316	190
SHANGHAI (CHINA)	SHA	0.7605634	213
PORT BLAIR (INDIA)	IXZ	0.7612903	155
CHANGSHA (CHINA)	CSX	0.7654321	162
NANNING (CHINA)	NNG	0.7668394	193
HANGZHOU (CHINA)	HGH	0.7685185	216
NANCHANG (CHINA)	KHN	0.7712766	188
JOHANNESBURG (SOUTH AFRICA)	JNB	0.7727273	154
SUN CITY (SOUTH AFRICA)	NTY	0.7744361	133
GUILIN (CHINA)	KWL	0.7758621	174
HEFEI (CHINA)	HFE	0.77707	157
MALE (MALDIVES)	MLE	0.7777778	54
PERTH (AUSTRALIA)	PER	0.7777778	54
NANJING (CHINA)	NKG	0.779661	177
YANGON (BURMA)	RGN	0.7814569	151
GABORONE (BOTSWANA)	GBE	0.7903226	124

Notes: *ShareBelow6K* is the share of cities between 5500-6000 miles of the city in question, out of the total number of cities between 5500-6500 miles. Sample is the 819 major airports as defined by ICAO.

Table A3. Correlation between *ShareBelow6K* and “Pre-Determined” Outcomes (City Level, 1989)

Dependent Variable:	<i>Connected Cities, Weekly</i> (1)	<i>Connected Cities, Twice-Weekly</i> (2)	<i>Connected Cities, Daily</i> (3)	<i>Connected Countries</i> (4)	<i>Log Total Seats</i> (5)	<i>Log Total Flights</i> (6)	<i>Log Passengers (millions)</i> (7)	<i>Network Centrality, Eigenvector</i> (8)
Share of Cities Below 6000 Miles	3.415 (4.411)	3.576 (3.345)	2.374 (1.835)	4.680 (4.140)	3.028 (3.162)	1.996 (1.896)	2.860 (3.031)	0.011 (0.011)
Observations	819	819	819	819	819	819	819	819
R-squared	0.018	0.024	0.031	0.026	0.080	0.074	0.079	0.045
Controls								
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Cities Around 6000 Miles	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Standardized Coefficient	0.035	0.04	0.047	0.041	0.059	0.064	0.058	0.041

A unit of observation is a city. All regressions control for Number of Cities around 6000 Miles, region fixed effects. Robust standard errors in parentheses, clustered at the country level. All logged variables add a one to deal with cases of zero values. Robust standard errors in parentheses, clustered at the country level. *** p<0.01, ** p<0.05, * p<0.1

Table A4. Correlation between *ShareBelow6K* and “Pre-Determined” Outcomes (Grid-Cell Level)

Dependent Variable:	<i>Night Light, 1992</i> (1)	<i>Population, 1990 (millions)</i> (2)	<i>Log Business Events, 1980s</i> (3)	<i>Log Business Connections, 1980s</i> (4)	<i>Average Yearly Precipitation, 1980-2014 (mm)</i> (5)	<i>Average Yearly Temperature 1980-2014 (Celsius)</i> (6)
Share of Cities Below 6000 Miles	2.009 (1.951)	0.008* (0.004)	-0.016 (0.026)	-0.013 (0.021)	87.542 (353.173)	-1.330 (3.311)
Observations	39,487	39,487	39,487	39,487	39,388	39,355
R-squared	0.267	0.156	0.035	0.035	0.300	0.499
Controls						
Region FE	Yes	Yes	Yes	Yes	Yes	Yes
Number of Cities Around 6000 Miles	Yes	Yes	Yes	Yes	Yes	Yes
Standardized Coefficient	0.032	0.048	-0.006	-0.006	0.017	-0.02

A unit of observation is a grid cell, within 100 miles of the closest airport in the country. All regressions control for Number of Cities around 6000 Miles, region fixed effects, and all the (log of the) distance between grid cell and the closest airport in the same country. Robust standard errors in parentheses, clustered at the country level. All logged variables add a one to deal with cases of zero values. Robust standard errors in parentheses, clustered at the country level. *** p<0.01, ** p<0.05, * p<0.1

Table A5. Alternative Instrument and Definitions of *ShareBelow6K*

Dependent Variable:	<i>Connected Cities, 2014</i>	<i>Connected Cities, 2014</i>	<i>Connected Cities, 2014</i>	<i>Connected Cities, 1989</i>	<i>Connected Cities, 1989</i>	<i>Total Connection- Years, 90-00</i>	<i>Total Connection- Years, 00-10</i>	<i>Total Connection- Years, 90-10</i>	<i>Total Connection- Years, 90-10</i>	<i>Total Connection- Years, 90-10</i>
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Number of Cities Below 6000 Miles	0.041*** (0.012)	0.034*** (0.011)	0.011* (0.006)	0.005 (0.007)	0.001 (0.014)	0.073* (0.044)	0.149** (0.058)	0.214** (0.085)		
Share of Cities Below 6000 Miles									32.0** (13.0)	25.9** (11.6)
Definition of Connection	Weekly	Twice-Week	Daily	Weekly	Weekly	Weekly	Weekly	Weekly	Weekly	Weekly
Bandwidth around 6000 miles	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline	5700-6300	5200-6800
Observations	819	819	819	819	819	819	819	819	819	819
R-squared	0.612	0.603	0.678	0.001	0.045	0.940	0.861	0.906	0.906	0.906
Controls										
Number of Cities Around 6000 Miles	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Airport Controls, 1989	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes
Dep. Var. Mean	2.83	2.48	0.84	1.44	1.44	18.04	24.14	40.42	40.42	40.42

A unit of observation is an airport/city. The independent variable in columns 1-8 is the number of cities below 6000 miles in the bandwidth of 5500-6500 miles, as compared to the share used in our baseline instrument. In columns 5 and 6 connections are defined as twice-weekly and daily scheduled flights, respectively. In columns 9 and 10 the bandwidth around 6000 miles of the baseline instrument is modified, made narrower or wider, as noted. All other variables are defined as in Table 2. Robust standard errors in parentheses, clustered at the country level. *** p<0.01, ** p<0.05, * p<0.1

**Table A6. Robustness I: Effect on Night Lights Using Alternative Instrument
(Number of Cities 5500-6500 miles)**

Dependent Variable: Night Lights	1992	2010	92-10 Change	92-10 Change	92-10 Change	92-10 Change	92-10 Change
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	RF	RF	RF	RF	RF	RF	2SLS
Number of Cities Below 6000 Miles	0.017	0.050**	0.033***	0.029***	0.028***	0.028***	
	(0.016)	(0.023)	(0.011)	(0.009)	(0.009)	(0.010)	
Total Connection-Years, 1990-2010							0.122**
							(0.072)
Anderson-Rubin p-value							0.004
Stock-Wright p-value							0.017
Angrist-Pischke first stage F-stat							4.43
Airports	777	777	777	777	777	734	734
Observations	39,487	39,487	39,487	39,487	39,355	37,766	37,766
R-squared	0.267	0.298	0.213	0.296	0.316	0.316	NA
<u>Controls</u>							
Number of Cities Around 6000 Miles	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Night Light in 1992	No	No	No	Yes	Yes	Yes	Yes
Population in 1990	No	No	No	Yes	Yes	Yes	Yes
Airport Controls, 1989	No	No	No	No	Yes	Yes	Yes
Geographic Controls	No	No	No	No	No	Yes	Yes
Country Real GDPpc, 1990 logs	No	No	No	No	No	Yes	Yes

A unit of observation is a grid cell, within 100 miles of the closest airport in the country. The independent variable is the number of cities below 6000 miles, as compared to the share used in our baseline instrument. All other variables are defined as before. Anderson-Rubin p-value refers to the weak instrument robust inference using the Anderson-Rubin Wald test (F-stat version), and the Stock-Wright p-value comes from the LM S statistic. *** p<0.01, ** p<0.05, * p<0.1 (using weak-instrument robust inference for 2SLS, indicating the highest p-value of Anderson-Rubin and Stock-Wright).

Table A7. Robustness II: Effects on Light Lights with Alternative Samples

Dependent Variable	92-10 Change	92-10 Change	92-10 Change	92-10 Change	92-10 Change	92-10 Change	92-10 Change	92-10 Change
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	RF	2SLS	RF	2SLS	RF	2SLS	RF	2SLS
	<u>Excluding Airport Cell</u>		<u>Excluding China</u>		<u>Different Distance Thresholds</u>			
Share of Cities Below 6000 Miles	4.478*** (1.313)		3.228** (1.431)		6.213*** (1.923)		3.695*** (0.991)	
Total Connection-Years, 1990-2010		0.139* (0.072)		0.081* (0.045)		0.223* (0.124)		0.110** (0.055)
First-Stage F-statistic		5.41		6.85		5.08		5.06
Anderson-Rubin p-value		0.001		0.026		0.002		0.000
Airports	719	719	690	690	732	732	734	734
Sample (miles)	<100	< 100	<100	< 100	< 50	< 50	<150	< 150
Observations	37,032	37,032	33,946	33,946	14,471	14,471	60,961	60,961
R-squared	0.331	NA	0.336	NA	0.270	NA	0.339	NA

A unit of observation is a grid cell. All regressions include the full set of controls as in Table 3. "Excluding Airport Cell" excludes from the sample, for each airport, the cell located at the smallest distance to the airport coordinates. Robust standard errors in parentheses, clustered at the country level. Anderson-Rubin p-value refers to the weak instrument robust inference using the Anderson-Rubin Wald test (F-stat version). *** p<0.01, ** p<0.05, * p<0.1.

Table A8. Robustness III: Alternative Standard Errors

Dependent Variable	92-10 Change	92-10 Change
	(1)	(2)
	<i>RF</i>	<i>2SLS</i>
Share of Cities Below 6000 Miles	4.683	
<i>Clustered, Airport Level</i>	(0.919)***	
<i>Spatial Correlation (100-mile threshold)</i>	(1.086)***	
<i>Spatial Correlation (50-mile threshold)</i>	(0.931)***	
Total Connection-Years, 1990-2010		0.143
<i>Clustered, Airport Level</i>		(0.069)***
<i>Spatial Correlation (100-mile threshold)</i>		(0.039)***
<i>Spatial Correlation (50-mile threshold)</i>		(0.033)***
Airports	734	734
Sample (miles)	<100	< 100
Observations	37,766	37,766
R-squared	0.320	NA

Same specifications as in Columns 6 and 8 (Table 3). *** p<0.01, ** p<0.05, * p<0.1

Table A9. Spillovers from Other Airports across Different Ranges, Airport Level

Dependent Variable:	<i>Connected Cities,</i> 2014	<i>Connected Cities,</i> 2014	<i>Connected Cities,</i> 2014	<i>Connected Cities,</i> 2014	<i>Connected Cities,</i> 2014	<i>Connected Cities,</i> 2014	<i>Connected Cities,</i> 2014	<i>Connected Cities,</i> 2014	<i>Connected Cities,</i> 2014
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(8)
Share of Cities Below 6000 Miles (Own)	1.27** (0.58)	1.04*** (0.38)	1.15*** (0.41)	1.17*** (0.36)	1.05* (0.54)	0.83** (0.35)	0.95** (0.38)	0.94*** (0.33)	
Share of Cities Below 6000 Miles (0-1000 Miles)	-0.20 (0.58)				-0.19 (0.53)				
Share of Cities Below 6000 Miles (1000-2000 Miles)		-0.40 (0.52)				-0.51 (0.46)			
Share of Cities Below 6000 Miles (2000-5500 Miles)			1.68** (0.69)				1.48** (0.61)		
Share of Cities Below 6000 Miles (Above 6500 Miles)				-0.57 (0.88)				-0.57 (0.76)	
Observations	815	818	819	819	815	818	819	819	819
R-squared	0.21	0.21	0.21	0.23	0.18	0.18	0.18	0.20	

A unit of observation is an airport/city. Connected Cities defined as (at least) weekly connections. Share of Cities Below 6000 Miles (X-Y Miles) is the average Share of Cities Below 6000 Miles for all airports located between X and Y miles away from the airport in question. These variables are all standardized, in order to facilitate the comparison between different coefficients. All regressions include the average Number of Cities around 6000 Miles (X-Y Miles), as well as own Number of Cities around 6000 Miles, Region fixed effects, and Airport Controls (1989): Numbers of daily, twice-weekly, and weekly flights, 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, Eigenvalue centrality. Robust standard errors in parentheses, clustered at the country level. *** p<0.01, ** p<0.05, * p<0.1

Table A10. Effect on Number of Major Business Events, GDELT Data, Airport-Pair Level

Independent variable:	Dep. Var.: Major Business Events, post-2000				Dep. Var.: Growth in Major Business Events, pre-2000 to post-2000			
	Below 6000 Miles	Below 6000 Miles	Below 6000 Miles	Below 6000 Miles	Below 6000 Miles	Below 6000 Miles	Below 6000 Miles	Below 6000 Miles
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
RD Estimate	0.0453	0.042	0.0434	0.0474	0.0053	0.007	0.0069	0.0072
Robust S.E.	(0.0166)***	(0.0119)***	(0.0130)***	(0.0137)***	(0.0030)*	(0.0020)***	(0.0020)***	(0.0026)***
Cluster S.E.	(0.0176)***	(0.0129)***	(0.0141)***	(0.0150)***	(0.0035)	(0.0025)***	(0.0025)***	(0.0031)**
Bandwidth (miles)	500	1000	Optimal	Optimal	500	1000	Optimal	Optimal
Polynomial order	1st	1st	1st	2nd	1st	1st	1st	2nd
Observations, Total	334954	334954	334954	334954	334954	334954	334954	334954
Observations, Effective	41,477	81,842	66,771	128,497	41,477	81,842	66,935	123,245
Dep. Var. Mean	0.083	0.083	0.083	0.083	0.025	0.025	0.025	0.025

Local polynomial Regression Discontinuity estimates, using rdrobust command in Stata (default options unless otherwise stated). Running variable is distance between airports, discontinuity at 6000 miles. Optimal bandwidth selected using one common MSE-optimal bandwidth selector. "Polynomial order" refers to order of the polynomial in distance. The dependent variable "Major Business Events" is the total number of observations of GDELT business collaboration events, where the "source" and "target" actors are both within a 100-mile radius of one of the airports in the pair, and at least one actor is a business entity. It is measured 2000-2014 in "post-2000", and 1979-1999 in "pre-2000". "Cluster S.E." refers to cluster-robust nearest neighbor variance estimation at the country-pair level, using a minimum of three nearest neighbors. *** p<0.01, ** p<0.05, * p<0.1

Table A11. Effect on Business Events and HQ Locations, Grid-Cell Level

Dependent Variable:	Major Business Events, Logs				Large Corporation HQ in Cell, 2006-12, Dummy					
	pre-2000	post-2000	post-2000	post-2000	2006-2012	2006-2012	2006-2012	2006	2009	2012
	OLS	OLS	OLS	OLS	Probit	Probit	Probit	Probit	Probit	Probit
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Share of Cities Below 6000 Miles	-0.031 (0.041)	0.260** (0.106)	0.192** (0.084)	0.167* (0.093)	0.017** (0.008)	0.013*** (0.004)	0.014*** (0.004)	0.013*** (0.004)	0.010** (0.004)	0.010** (0.004)
Airports	777	777	777	734	777	777	734	734	734	734
Observations	39,487	39,487	39,487	37,766	39,487	39,487	37,766	37,766	37,766	37,766
R-squared	0.057	0.521	0.570	0.580	NA	NA	NA	NA	NA	NA
<u>Controls</u>										
Number of Cities Around 6000 Miles	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Major Business Events, Pre-2000, Logs	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Night Light in 1992	No	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Population in 1990	No	No	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Airport Controls, 1989	No	No	No	Yes	No	No	Yes	Yes	Yes	Yes
Geographic Controls	No	No	No	Yes	No	No	Yes	Yes	Yes	Yes
Country Real GDPpc, 1990 logs	No	No	No	Yes	No	No	Yes	Yes	Yes	Yes

A unit of observation is a grid cell, within 100 miles of the closest airport. The dependent variable "Major Business Events" in column 1-3 is the total number of observations of GDELT business collaboration events, where the "source" and "target" actors are both within a 100-mile radius of one of the airports in the pair, and at least one actor is a business entity. The cell-level outcome is the sum of all such events where the cell is either the "source" or the "target", and where there is at least 3000 miles between the two airports associated with the cells. The dependent variable in column 4-6 is a dummy for whether there was a large corporation head quarter in the cell at any point between 2006-2012, measured by the Forbes "Global 2000" list. It is measured 2000-2014 in "post-2000", and 1979-1999 in "pre-2000". The marginal effects at the sample mean are reported for Probit regressions in columns 5-10. The other variables are defined as before (see Table 3). Robust standard errors in parentheses, clustered at the country level. *** p<0.01, ** p<0.05, * p<0.1