The Role of Aviation Networks for Urban Development*

Anca D. Cristea[†]
University of Oregon

Liliana Danila[‡]
British Retail Consortium

October, 2017

Abstract

City officials are continuously working to attract airlines willing to fly to new destinations. The inherent expectation is that a more extensive aviation network stimulates economic growth. This paper investigates empirically the causal implication of this hypothesis. Using data on non-stop flights by origin and destination over the period 1984-2001, we propose a new measure for a city's connectivity to the national aviation network. We then use this measure to investigate its contribution to local economic development, as reflected by the growth in population, in total employment, in per-capita income and new firm entry. To ensure causality, we use instrumental variable methods that exploit geography and weather patterns at destination cities as a way to capture the exogenous variation in the likelihood to add new travel routes. Our results suggest that a city's air connectivity, resulting from an extensive local aviation network, has a positive effect on population size, employment level and on the number of businesses established in that location.

JEL: O18, R1, R4
Keywords: air transport, aviation network, urban growth, regional development, connectivity

*We thank Bruce Blonigen, Jan Brueckner, Andrew Cassey, Nicholas Sheard, Katsumi Tanabe, as well as participants at the 62nd North American Meetings of the Regional Science Association International (Portland, OR, 2015), the 42nd Annual Eastern Economic Association Conference (Washington, D.C., 2016) and the 51st Pacific Northwest Regional Economic Conference (Bend, OR, 2017) for helpful discussions and suggestions. We are grateful to Xavier Giroud for graciously sharing the aviation data, and to Anna Bezner and Kymberly Teoh for excellent research assistance. Any remaining errors are our own.

This paper should not be reported as representing the views of the British Retail Consortium. The views expressed in this paper are those of the authors and do not necessarily represent those of the British Retail Consortium.

[†]Corresponding Author: Department of Economics, University of Oregon, 1285 University of Oregon, Eugene, OR 97403, USA. E-mail: cristea@uoregon.edu.

[‡]British Retail Consortium, 2 London Bridge, London SE1 9RA. E-mail: Liliana.Danila@brc.org.uk.

1 Introduction

Announcements of new flights and new travel destinations are received with much enthusiasm by city residents and local officials alike, being viewed as a sign of economic vitality and future economic growth. There is a general belief that the expansion of air services will provide a boost to the local economy that goes beyond the immediate jobs created as a result of increased airport activity. In fact, the prospects of new opportunities and future economic growth are what motivates city officials to offer subsidies and spend substantial efforts in pursuing strategic partnerships with commercial airlines.

A plethora of examples exists to support the above statements. Some very recent ones, as of October 2017, involve the city of Missoula and San Antonio. In Missoula, a coalition of business groups launched an initiative to attract more direct flights to the Missoula International Airport, stating that such a move would boost the economy and would lower air fares.[1] The city of San Antonio is currently examining whether to expand its airport, citing that direct flights to major cities "are an obstacle [to the development of the local economy]" and that "[the lack of nonstop flights is a] very big challenge for us to overcome." A case in point: San Antonio's poor connectivity is considered as one of the major drawbacks in the citys bid to attract Amazon's second headquarters.[2] Another relevant example is AT&T moving its headquarters from San Antonio to Dallas in 2008 due to the poor connectivity of the former city. The AT&T senior executive vice president of executive operations stated: "San Antonio is a wonderful city, but it can be a difficult place to get to and fly out of."[3]

The contribution of commercial aviation to urban growth – while never doubted by city officials – has been supported by recent empirical research.[4] Perhaps less understood is the

[1]See Erickson (2017): http://missoulian.com/news/local/missoula-airport-trying-to-attract-direct-flight-service-to-texas/article_6efe19e3-e40f-5f42-9c90-da802c23304f.html

[2]See Joshua Fechter and Druzin (2017): http://www.expressnews.com/business/local/article/New-San-Antonio-airport-expanded-facilities-on-12268129.php

[3]See Poling and Pack (2008): http://www.mysanantonio.com/business/article/AT-T-leaving-San-Antonio-for-Dallas-1758611.php

[4]See, for example, Brueckner (2003), Green (2007), Blonigen and Cristea (2015), Sheard (2015) and McGraw (2017).

importance that a city's aviation network structure plays for local economic development. How valuable is it for a city to add more direct routes and connect to new destination markets? Or is it enough to simply reinforce the number of flights on existing destinations? Does the reach of a city's aviation network matter for economic growth and development?

At first glance, these questions seem to have straightforward answers. A wider aviation network should benefit a city as it reduces the time costs incurred by agents when reaching markets nationwide. However, a wide aviation network can increase the monetary cost of travel and become difficult to sustain in the long run. Serving multiple destinations by non-stop air service may decrease the level of airline competition within a city-pair market, and also eliminate the benefits from the economies of scale generated by the use of large capacity aircrafts. Adding all the costs and benefits, it becomes less obvious whether expanding a city's aviation network invariably brings economic gains to consumers and local businesses.

In this paper we examine empirically the contribution of a city's local aviation network to its economic development. We propose a novel measure to capture the air connectivity of a city to the wider national aviation network. This measure is constructed as a weighted count of all non-stop flights per destination, where the weights take into account the importance of each destination airport within the national aviation network. Our intention is to assess how important aviation connectivity is for urban development, where the latter is measured by economic indicators such as population size, per-capita income and employment levels, as well as by the number of new businesses that choose to locate in that city.

Our main econometric challenge comes from the interdependence between a city's economic development and the size of its aviation network. Large communities demand more air travel and are able to sustain non-stop service to more destinations in comparison to small communities. This reverse causality channel raises endogeneity concerns. A solution is to employ instrumental variable methods.

In this paper we develop several novel exogenous variables to instrument for a city's connectivity to the national aviation network. We exploit information on geography (i.e., distance between cities) and weather patterns (i.e., average maximum temperature and total snowfall) at destination to predict the extent to which a city will offer direct flight services to a given destination. Furthermore, holding a city's aviation network constant at pre-sample levels, we also exploit information on the growth of air traffic at destination cities to infer the resulting gains to an origin city from the improved connectivity to the broader national aviation network.

A related concern and potential source of endogeneity comes from the fact that a city's air connectivity is likely related to the more encompassing measure of market access (also known in the economic geography literature as market potential). Proximity to large metro areas benefits a community directly as it facilitates the exchange of goods, services and ideas, resulting in higher per-capita income and in employment growth. At the same time, proximity to large markets may increase the likelihood of nonstop service to those markets, increasing a city's air connectivity. To ensure that our estimation results do not suffer from omitted variable bias, we include in all specifications a control for the market access of a given city.

Using historical data on business activity and air passenger traffic for 147 core-based statistical areas (CBSAs) spanning the period 1984-2001, we provide direct evidence of the causal effect of air connectivity on local economic development. To estimate our econometric model, we exploit time series variation within CBSAs. Based on our instrumental variables estimations, we find that a one standard deviation increase in a city's connectivity to the national aviation network leads to a 3 percent growth in urban population, on average. We find a larger effect on total employment at city level, equal to 6.4 percent.

The estimated increases in economic size and employment levels are driven to a significant degree by the entry of new firms in the local market. In support of this claim, we find that a one standard deviation increase in air connectivity leads to a 5.3 percent increase in the number of local businesses. This growth is driven primarily by small and medium size firms. Our results are robust to using different subsets of exogenous variables as instruments for air connectivity.

Our findings contribute to a growing literature on the role of transportation infrastructure for regional development. Brueckner (2003) and Green (2007) are among the first to investigate the role of air passenger transport for urban growth, finding large positive effects. More recently, Blonigen and Cristea (2015), Sheard (2014, 2015) and McGraw (2017) exploit novel sources of exogenous variation in the provision of local air services to bring further evidence on the direction of causality. Several studies have focused on other modes of transport such as road or rail transport, examining their respective roles for urban growth and industry specialization.[5]

A study that is perhaps the closest to ours is LeFors (2015). It proposes a city-level measure of air accessibility, constructed following a market access approach, which is then used to investigate its contribution to urban development. Employing a cross-sectional sample of about 200 MSAs, LeFors (2015) finds robust evidence that air accessibility impacts the growth rate of employment in service industries, though he finds no effect on total employment or on productivity growth. Our paper is similar to LeFors (2015) in that both studies propose measures of air transport services that go beyond the total volume of air traffic, by exploiting information about the actual network structure of a city's aviation services. The implementation techniques of the two studies, however, are different. First, the construction of the air connectivity, respectively, the air accessibility measure differs across the two papers in that we exploit an extensive margin measure that only uses information on a city's flight services to non-stop destinations, while LeFors (2015) exploits a market access measure that accounts for *all* – direct and indirect – connections to *all* the cities in the sample. Second, the time period and variation exploited for model identification also differ

[5]Duranton and Turner (2012) examine the causal effect of the U.S. interstate highway system on urban growth, finding positive effects on population growth and employment. Faber (n.d.) focuses on a rapidly developing country such as China to investigate how improved highway infrastructure affects negatively small markets due to the reinforced concentration of economic activity in large markets. Donaldson (n.d.) and Donaldson and Hornbeck (n.d.) quantify the importance of railroads in improving a location's market access, generating income growth and welfare.

across the two studies. Our sample is a panel dataset covering almost two decades of data, a feature that enables us to control for unobservable city-specific effects in our regression analyses. Meanwhile LeFors (2015) uses a cross-sectional dataset of long-run urban growth rates that is combined with a cross-section of air accessibility measures and other control variables at city level. Finally, while both papers rely on instrumental variables methods, the proposed instruments depart from each other, exploiting different sources of exogenous variation in the aviation network of a city. For all these reasons, we consider our paper complementary to LeFors (2015) and to all the existing evidence on the role of aviation services and airport infrastructure for urban development.

The remainder of our paper proceeds as follows. Section 2 presents a decomposition of the total air traffic in a city along various dimensions. This helps us define our proposed air connectivity measure, intended to capture the reach of a city's local aviation network. The econometric model and estimation strategy are laid out in section 3. Section 4 discusses the data sources and variable construction, while section 5 presents the estimation results. Section 6 discusses the main policy implications and concludes.

2 Growth in Passenger Aviation

Airlines respond to the growth over time in air passenger travel by increasing capacity utilization on a given origin-destination route (i.e., intensive margin), as well as by increasing the frequency of flights and the number of destinations reached by direct service (i.e., extensive margin). Existing studies that investigate the impact of aviation services on urban growth typically exploit information on total air passenger traffic in a city, making little effort to distinguish between the various channels of air traffic growth. Whichever channel at play, the primary goal has been to identify the direction of causality running from the availability of air services to the overall economic growth of cities.

Given the efforts spent by local communities to expand the reach of their aviation network, we think it is worth focusing on a more refined measure of a city's aviation network – which we label *air connectivity* – and examine its contribution to economic development and urban growth. In what follows, we define the air connectivity measure and then propose some further refinements for data implementation.

Defining Air Connectivity

We start from the volume of air passengers handled by city i at time t, denoted a_{it}, which is computed as the sum of air passengers on all non-stop flights f from origin city i to destination city d. We decompose the total air traffic a_{it} into: (1) the number of destinations d reached by non-stop flight service from city i at time t, which we denote by N_{it}, (2) the average number of flights per destination, which we calculate as the total number of flights across all destinations divided by the total number of non-stop destinations and we denote by $\overline{F_{it}}$, and (3) the average number of passengers per flight per destination, which we denote by $\overline{a_{it}}$. Formally, we can derive this decomposition as follows:

$$a_{it} \equiv \sum_{d} \sum_{f \in \{idt\}} a_{idft} = F_{it} \cdot \sum_{d} \left(\frac{F_{idt}}{F_{it}} \cdot \frac{\sum_{f \in \{idt\}} a_{idft}}{F_{idt}} \right) = \tag{1}$$

$$= N_{it} \cdot \frac{F_{it}}{N_{it}} \cdot \sum_{d} \left(\frac{F_{idt}}{F_{it}} \cdot \frac{\sum_{f \in \{idt\}} a_{idft}}{F_{idt}} \right) \equiv \tag{2}$$

$$\equiv N_{it} \cdot \overline{F_{it}} \cdot \overline{a_{it}} \tag{3}$$

where i indexes the origin city, d indexes the destination city, f indexes a non-stop flight connecting two cities $\{id\}$, and t indexes the time period. In equation (1), the total number of flights out of origin city i is defined as $F_{it} \equiv \sum_d F_{idt}$, where F_{idt} denotes the annual number of non-stop flights offered from origin city i to destination d at time t.

We define the air connectivity of city i to the national aviation network at a given point in time t as the expansion of traffic due to the introduction of new flights to existing or to new destinations. That is:

$$AirConnect_{it} \equiv \sum_{d} F_{idt} = N_{it} \cdot \overline{F_{it}} \tag{4}$$

As defined in equation (4), the air connectivity measure corresponds to the first two terms of the air traffic decomposition in equation (3).

Variations in the total demand for air travel get reflected either in changes in air connectivity (i.e., extensive margin), or in changes in aircraft capacity as captured by the number of passengers per flight per destination, $\overline{a_{it}}$ (i.e., intensive margin). The latter happens whenever airlines respond to demand shocks by optimizing their fleet usage as well as the capacity utilization per non-stop flight. As an example, airlines could maximize the passenger load factor, or they could adjust the existing route capacity by changing aircrafts size.

Further Refinements of the Air Connectivity Measure

One can view the air connectivity measure defined by equation (4) as a count of distinct destinations d that city i is able to reach via direct air service, where the importance or "weight" of each destination d is given by the frequency of service between city i and that destination, i.e., F_{id}. That is:

$$AirConnect_{it} \equiv \sum_{d} I(Direct_{idt} = 1) \cdot F_{idt} \tag{5}$$

where $I(\cdot)$ denotes a 1/0 indicator function, and $Direct_{idt}$ denotes whether an origin city i has a direct flight service to destination city d at time t.

Equation (5) highlights the ability of the air connectivity measure to capture the extensive margin of a city's aviation network. However, one limitation of the current specification of the air connectivity measure is that it provides no information on how important each destination d is overall, within the national aviation network. Therefore, we augment the air connectivity measure in equation (5) as follows:

$$AirConnect_{it} = \sum_{d \neq i} I\Big(Direct_{idt} = 1\Big) \cdot F_{idt} \cdot \Big(\frac{F_{dt}}{F_t}\Big) \tag{6}$$

where F_{dt} represents the total number of departures offered out of destination d at time t, and F_t denotes the overall number of flights operated nationwide at a given point in time (i.e., $F_t = \sum_i \sum_d F_{idt}$). Thus F_{dt}/F_t captures the fraction of all U.S. departures accounted for by city i at time t.

As equation (6) shows, we compute the measure of air connectivity as the number of destinations d connected by non-stop flights to city i (identified by the indicator function in the first bracket term), where we weight the importance of each destination d based on: 1) the number of flights that city i offers to destination d (given by the second term), and based on 2) the "hub-ness" of city d as captured by the share of destination d in total U.S. departures (given by the third bracket term). We think that this weighted count of routes is superior to the simple count approach as we do not think that every aviation route brings the same value to consumers and businesses in a location.

The air connectivity measure in equation (6) essentially captures the degree to which origin city i is connected to the national aviation network. A large value of the air connectivity measure indicates that a city has frequent access to multiple large hub airports. Our econometric analysis examines the degree to which air connectivity contributes to local economic growth.

3 Econometric Strategy

Baseline Model

One empirical challenge in estimating the impact of air connectivity on local economic development is the presence of many unobservable factors that could affect both variables of interest. So, a first step towards mitigating the omitted variable bias problem is to take advantage of the panel structure of our dataset and estimate a model using three-year differences in variables. In particular, the regression model that we propose takes the following form:

$$\Delta_3 lnY_{it} = \beta_1 \Delta_3 ln(AirConnect)_{it} + \Delta_3 X_{it}{}'\gamma + \alpha_t + \epsilon_{it} \tag{7}$$

where $\Delta_3 Z \equiv Z_t - Z_{t-3}$ for any variable Z, and i and t index a city and year, respectively. Y denotes an economic outcome of interest such as urban population, total employment, average per-capita income, or the number of business establishments active in a given city. X represents a vector of control variables and captures any observable time-varying determinants of urban growth that could be correlated with a city's air connectivity to the national aviation network. α_t denotes period fixed effects. Any unobservable macroeconomic shock that affects city growth in the same way for all cities in the U.S. will be directly accounted for by α_t.

The decision to estimate the regression model in equation (7) using three-year differences in variables is motivated by the need to eliminate city-specific effects and reduce serial correlation. Thus, any unobservable city characteristics that are time invariant or highly persistent over time will be eliminated through differencing. This ensures the removal of any factors that may influence both urban growth and air connectivity, thus mitigating the omitted variables bias problem.

Although city fixed effects would also address the problem of unobservable location-specific factors, one additional benefit of using long differences is the ability to minimize any attenuation biases coming from errors-in-variables (Griliches and Hausman, 1986). In deciding the length of time over which to difference the data, Griliches and Hausman (1986) argue that the tradeoff is between removing too much signal and removing some of the noise in variables. Given the nature of our data, some of the economic variables of interest evolve sufficiently slowly over time to warrant opting for a longer difference estimation instead of first differencing.

Our goal is to estimate the causal effect of air connectivity on regional growth. Thus, the coefficient of interest in equation (7) is β_1. The main challenge in obtaining an accurate estimate of β_1 is the endogeneity problem involving a city's air connectivity, which is caused by its simultaneity with urban growth.

Solving the Endogeneity Problem: Instrumental Variables

The connectivity of a city's airport to the domestic aviation network is directly related to the city's economic growth. Locations that do well and grow fast are going to attract more airlines that are willing to provide non-stop service to new destinations. This means that the first two terms in equation (6) – the existence of a direct connection $I(Direct_{idt} = 1)$, and the number of departures per destination, F_{idt} – evolve over time based on city i's rate of economic growth. This raises endogeneity concerns for the empirical analysis.

To instrument for city i's connectivity to the national aviation network we propose two exogenous variables. In defining them, we stay close to the format of the air connectivity measure defined in equation (6). The first instrumental variable takes the following form:

$$AirConnect\_IV_{it} = \sum_{d \neq i} I\Big(Direct_{id,T_0} = 1\Big)\Big(\frac{1}{Distance_{id}}\Big)\Big(\frac{F_{dt}}{F_t}\Big) \tag{8}$$

where T_0 denotes the base year of our sample period.

The first bracket term in equation (8) identifies the set of destinations reached from city i by non-stop flight service at the beginning of our sample period, T_0. We keep the set of non-stop destinations constant at the base year level in order to condition the instrument's variation over time to sources exogenous to city i. Our strategy is to exploit only variation coming from changes in total departures at destination cities, weighted by how close those destination cities are to origin city i. As our regression model in equation (7) is identified from within city changes over time, this approach allows us to isolate a source of time variation that is orthogonal to the growth path of city i.

Importantly, the aviation network of a city at the beginning of our sample period should overlap with the set of non-stop destinations served at any time t during our sample period. This would ensure the necessary correlation between the excluded instrument and the endogenous air connectivity variable.[6] Yet, by being held constant at its initial level, the reach of the aviation network is going to be orthogonal to city i's subsequent economic growth by construction.

The second bracket term in equation (8), i.e., the inverse of the geographical distance between the origin city i and destination d, is an exogenous determinant of the number of departures offered between cities i and d. All else equal, further apart locations are usually connected by less frequent flights. The benefit of using the inverse of bilateral distance in constructing the excluded instrument comes again from its orthogonality with respect to the economic development of city i.

Finally, the third bracket term in equation (8) is the same as the corresponding term in equation (6). This is the only component that drives the instrument's variation over time. In exploiting this source of variation for the model identification, an implicit assumption is that the economic development of destination cities d is independent of the growth trajectory of city i. As it will become clear later on, the selection of the control variables to be included in the regression estimation has been guided by the need to ensure that this assumption holds in the data.

A second instrument that we exploit in our analysis follows a similar construction strategy as for the instrument we just described, except that it employs information on climatic conditions at the destination city d.[7] We use weather as an exogenous shock to the number of flights F_{idt} operated between origin city i and destination d at time t. More specifically, we propose the following weather-related instrumental variable:

[6] In the data, we ensure that an overlap exists between the routes offered in the base year and in the subsequent periods by restricting the estimation sample to MSA locations that offer at least one route throughout the sample period. We provide more details on the construction of the estimation sample in the data section.

[7] The motivation to use weather data to generate exogenous variation in air travel patterns comes from Forbes and Lederman (2009). This study investigates vertical integration decisions in the U.S. airline industry and tests whether major airlines are more likely to own regional carriers on city pairs with adverse weather because of the frequency of adaptation decisions (and implicit renegotiation costs) resulting from air traffic disruptions.

$$AirConnect\_IV_{it} = \sum_{d \neq i} \Big(Direct_{id,T_0} = 1\Big)\Big(Weather_{dt}\Big)\Big(\frac{F_{dt}}{F_t}\Big) \tag{9}$$

where the variable $Weather$ stands for: (1) the monthly maximum temperature in city d, averaged across all months in year t; and (2) the total amount of snowfall in city d in year t. All else equal, colder and snowier locations are expected to have more frequent weather-related disruptions to air traffic. By multiplying the $Weather$ variable by the share F_{dt}/F_t of all U.S. departures accounted for by destination d, we are essentially weighting the weather conditions at a given destination by the importance of that destination within the domestic aviation network. Bad weather is more problematic if it affects a large hub, as it dramatically reduces the air connectivity of origin cities who rely primarily on that large hub to access further locations.

Control Variables

A possible concern with the instrumental variables defined by equations (8) and (9) is that proximity to large cities benefits a given city i through channels other than improved aviation connectivity. For example, proximity to a large and fast-growing urban area could lead to positive spillover effects, which would impact the local labor market and the overall economic growth of the city. Without controlling for such potential spillover effects, we may mistakenly interpret any positive impact of a city's air connectivity as a causal effect. To address this concern, in the regression model we are going to control for city i's market access, which we construct as:

$$MktAccess_{it} = \sum_{d \neq i} \frac{GDP_{dt}}{Distance_{id}} \tag{10}$$

This measure is also known in the economic geography literature as the market potential of a location.[8] It is important to note that, unlike the air connectivity expression from equation

[8] To the best of our knowledge, the concept of market potential goes back to Harris (1954). Subsequent work in economic geography, such as Krugman (1991) or Fujita et al. (1999), provides the theoretical background to structurally derive a location's market potential from general equilibrium spatial models.

(4), the market access variable sums up information about *all* locations d in our sample. It essentially captures the access that a given city i has to markets nationwide, where the importance of each market is measured by the GDP of that location weighted by the inverse of the geographic distance to that location. Several studies provide both theoretical and empirical evidence of faster employment and income growth in regions with better market access.[9]

In addition to the market potential variable, we also account for differences in secular trends across cities using information on historical population levels. In particular, we use city population data lagged by 50 years to control for city-specific long-run growth trends. Thus, in the regression model in equation (7), the vector X of control variables consists of market potential and 50-year population lags, both differenced over 3-year periods.

4 Data Sources

To estimate our regression model, we need to combine three main sources of information: data on airline flight schedules, airport weather data, and economic data at city level.

Air service data.

The data on air services is collected from two datasets: the T100 Domestic Segment Database provided by the Department of Transportation (DOT), and the ER-586 Service Segment Database provided by the National Archive and Records Administration (NARA).[10] Both databases are compiled from Form 41, a form which contains data on the domestic operations of all certified airlines who are legally required to fill it.

Prior to 1990, the aviation data compiled from Form 41 was stored and publicly disseminated in the form of magnetic tapes by NARA. Starting in 1990, the same data have been provided online by the DOT under the name of T100 Segment Database. Each data set

[9] See, among others, Redding and Venables (2004) for a cross-country analysis, or Hanson (2005) and Head and Mayer (2006) for regional analyses.

[10] We thank Xavier Giroud for graciously providing us with the ER-586 Service Segment Data.

covers all flights operated between any two airports in the U.S. in a given year; it contains monthly level information on the number of departures scheduled and performed, on the number of available seats and passengers carried by each airline on each origin-destination route they operate at a given point in time. We aggregate the original aviation data across all the months within a year, and then across all airports within a city to obtain annual data on air services at CBSA level.

Our estimation sample covers the time period 1984–2001. In choosing the sample starting year, we wanted to avoid the aftermath of the 1978 Airline Deregulation Act, which represents a major policy change.[11] The end-point of the sample period was decided due to a modification in the data collection process of the T100 Segment Database (which introduces a structural break in the time series variation past year 2001).

Weather data.

Weather data for all the main U.S. airports are collected from the National Oceanic and Atmospheric Administration (NOAA) over the sample period 1984 - 2001. These data include, among others, information on the maximum temperature levels in a given month and year, and on the total amount of snowfall per calendar year.[12] We aggregate the monthly airport records across all the months in a year to get annual values, and then we average the annual values across all the airports within a city to get weather data at CBSA level.

City-level economic data.

The economic indicators of interest are the population size of a city, the employment level, the average personal income and the number of active business establishments (total and by size category). The main sources for these data are the Bureau of Economic Analysis (for population and per-capita income) and the U.S. Census County Business Patterns (for

[11]Blonigen and Cristea (2015) provide reasons to believe that by 1983 the aviation market nationwide completed its transition to a free market regime, settling into its new growth trajectory.

[12]The maximum monthly temperature levels are constructed by averaging the daily maximum temperatures over all the days in the month.

employment and establishment data).[13] Since most economic data going back to 1984 are only reported at county level, we construct a mapping of counties into the corresponding CBSAs they belong to, and use it to aggregate the county level data up to the CBSA level.[14]

Main variable construction.

After combining the three main sources of data, we obtain an unbalanced panel at city level. We then restrict the resulting sample to include only CBSAs that offer at least one aviation route continuously throughout our sample period. This restriction is necessary given the approach we take in constructing the instrumental variables in equations (8) and (9). By holding fixed a city's network of non-stop destinations at the base year level, one way to ensure that the excluded instruments are going to be correlated with the endogenous air connectivity variable is if there are (at least one) aviation routes in common between the base period and all the subsequent sample years. Once we impose the "continuously offered route" restriction, the resulting estimation sample consists of 147 CBSAs observed over the period 1984–2001.

We then employ the resulting city-level panel dataset to construct our variables of interest. In particular, we follow equation (6) to construct our preferred air connectivity measure. Furthermore, we use the geographical coordinates of the airports within cities to calculate the bilateral distance between any two CBSAs in our sample. Combined with information on total income at city level (obtained by multiplying population size with the average per-capita income level), we apply equation (10) to construct the market potential of a city in our sample. We also use the bilateral distance to implement equations (8) and (9), and obtain the exogenous instruments for air connectivity.

[13] County level population data going back to 1930s are available only at decennial level. To construct the 50-year lag control variable, the decennial population data were first linearly interpolated to obtain annual population levels for periods pre-1950. Then the interpolated values were differenced at 3-year intervals to generate 3-year population changes lagged by 50 years.

[14] We use the current mapping of counties into the core based statistical areas available from the U.S. Census through their U.S Gazetteer Place data from 2006. Even if the current delineation does not correspond to the one implemented several decades ago, its application throughout the entire sample period ensures the consistency of statistical areas throughout the panel period.

Finally, once we have all our data ready, we construct three-year changes in the variables of interest. Note that because of differencing the data, the first three sample years are going to be omitted from all the regression estimations. Thus, the estimation sample will include observations for 147 CBSAs observed over 15 time intervals. Table 1 reports summary statistics for all the variables used in the econometric analysis, expressed in three-year differences.

5 Estimation Results

In this section we report the coefficients obtained from estimating the regression model given by equation (7), and discuss the results in light of their possible policy implications.

A crucial part of our econometric analysis is the ability to overcome the endogeneity problem using instrumental variables. Before turning to the main estimation results, we discuss some preliminary descriptive evidence that motivates our choice of instruments and builds confidence in their expected performance during the regression estimations.

5.1 Preliminary Data Analysis

In finding valid instruments for a city's connectivity to the national aviation network, one approach we take is to exploit information about the weather conditions at each destination city, weighted by the importance of that destination (as a hub) within the national aviation network. Non-stop service to a destination with unfavorable weather conditions should contribute less to the overall aviation network of an origin city because of all the uncertainties and flight disruptions caused by bad weather. One way to provide direct evidence to support this claim is to collect data on flight delays and flight cancellations at CBSA level and then inspect the degree to which weather conditions contribute to such disruptions in the scheduled departures.

Defining the difference between scheduled and operated departures as the number of cancelled flights in a city within a year, Table 2 reports the estimates from a panel regression in three-year differences. We regress flight cancellations on weather conditions, controlling for market potential and for the population size of the city, as well as for period fixed effects. The estimation sample covers the same set of CBSAs and the same time period (1984–2001) as used for the main econometric analysis of the paper. Column 1 reports the results when using as weather indicator three-year changes in the average monthly maximum temperature (in Celsius) recorded in a city in a given year. Column 2 reports the same estimation exercise but using as weather indicator three-year changes in total annual snowfall (in cm). Finally, column 3 reports the results when adding both weather variables to the estimation. All three specifications indicate statistically significant effects of weather conditions on the number of flight cancellations at a given location. In particular, a fall in the average maximum temperature leads to an increase in flight cancellations. Not surprisingly, colder areas witness more disruptions in their flight schedules, all else equal (i.e., conditional on macroeconomic factors, on unobserved location specific effects and on city size growth). Cancellations also rise from an increase in the total snowfall recorded at a location.

Similar findings are obtained when examining the effect of weather conditions on the frequency of flights that are delayed more than 15 minutes.[15] Since 1988, the DOT has been collecting statistics on airlines' flight performance and causes of delay, which are publicly available in the On-Time Performance Database. We use this dataset to construct for each city and time period the fraction of all scheduled flights that are delayed more than 15 minutes. We then regress the three-year changes in the share of delayed flights on the same weather indicators as previously described, controlling again for population size and market potential, as well as for period fixed effects. The results are reported in Table 3. The sign and statistical significance of the estimates are consistent with expectations and follow the same

[15]The Bureau of Transportation Statistics uses the 15 minute period to classify a flight as delayed.

patterns as observed before. Essentially, cities with low average maximum temperatures and with high records of snowfall have a higher share of delayed departures.

In both sets of exercises, the F-statistics for the joint significance of the three-year weather variables has a p-value of less that 1 percent, which confirms the expectation that a worsening of weather-related conditions has direct implications for the frequency and reliability of aviation services.[16] The strength of these effects is important to mitigate the weak instrumental variables problem.

5.2 Estimation Results

Our goal is to estimate the causal effect of air connectivity on urban development. We estimate equation (7) using as dependent variables a city's population size, total employment, average per-capita income and the number of business establishments, respectively. Since this is a panel regression estimated using three-year differences and period fixed effects, the identification of the model parameters comes from comparing changes in the variables of interest across the 147 sample CBSAs over time. Using three-year differences helps eliminate any time invariant factors that affect city growth, many of which may be hard to measure or are unobservable. However, this still leaves open the possibility of omitted variable bias coming from city-specific secular trends. To overcome this concern, we construct a 50-year lag series of city level population values, which we use as a control variable in all the estimations along with the market access control variable. Adding these variables to the regression model is intended to mitigate the endogeneity problem. Furthermore, to ensure a causal interpretation of our estimates, we are also going to correct for any remaining endogeneity using the standard instrumental variables (IV) approach.

The first set of estimates are reported in Table 4 and employ city population size as the economic outcome to be explained. Column 1 reports the OLS estimates, while the remaining three columns report the IV estimates obtained by limited information maximum

[16]The joint F-statistics from the regression on the number of cancelled flights is 9.72, respectively it is 7.89 in the regression explaining the fraction of delayed departures.

likelihood (LIML).[17] The excluded instruments consist of the distance-based air connectivity variable from equation (8) in column 2, as well as the two weather-based air connectivity variables from equation (9) in column 3. All three excluded instruments are pooled together in the estimation reported in column 4.

The OLS effect of air connectivity on city population is positive, which is expected, but it is not statistically significant. This result may seem surprising in light of the likely upward bias coming from endogeneity concerns. However, it is possible that large and fast-growing cities witness a relatively slower growth in direct service over our sample period, in part because many of the major direct connections had already been established prior to the start of the sample period.[18] Importantly, the IV estimates are able to correct for any endogeneity biases (whether positive or negative) as long as the proposed excluded instruments are valid. This seems to be the case judging from their statistical test performance. The F-statistics for the excluded instruments reported at the bottom of Table 4 meet conventional critical levels in all of the three IV specifications. This is suggestive of a significant correlation between the air connectivity measure and the excluded instruments. Furthermore, the test for overidentifying restrictions (when possible to be computed) fails to reject the null hypothesis that the excluded instruments are orthogonal to the regression residual.

Examining the IV estimates from Table 4, one thing to notice is their consistency across specifications (especially columns 2 and 4), but also their substantially larger magnitude in comparison to the OLS estimate. Although only marginally significant, the estimate in column 2 suggests that a 10 percent increase in air connectivity causes a 0.13 percent change in population size. This effect is substantial in magnitude. To see this, note that a one standard deviation increase in air connectivity corresponds to a 230 percent change relative

[17]The choice to use LIML over two-stage least squared (2SLS) is dictated by the better performance of LIML estimation methods in the presence of small samples and potentially weak instruments (Anderson et al., 1982). However, the estimation results are almost identical when using the 2SLS method. These results are available upon request.

[18]For a city that starts out with a large number of non-stop destinations and many flights per route, the variation over time in the air connectivity index is going to come from adding smaller, less frequent destinations to an already dense aviation network. This may lead to a downward bias in the estimate.

to the mean, which implies a 3 percent growth in city population. The estimate in column 4 is similar in magnitude, though it is also only significant at 10 percent confidence level.

To gain confidence in the LIML estimations and in the causal interpretation of the coefficients, we report in Table 5 the results from the first stage IV regressions. Columns 1–3 report the first stage estimates corresponding to the LIML specifications in columns 2–4 from Table 4. Focusing on the behavior of the excluded instrumental variables, we note that the distance-based instrument in columns 1, respectively in column 3, has the expected sign and is highly significant as a determinant of air connectivity. Being in the proximity of destination cities that operate a large share of the total departures nationwide implies a large value for the distance-based connectivity instrument. The result in column 1 of Table 5 shows that a 1 percent increase in this distance-based instrument leads to a 1.2 percent increase in a city's air connectivity, which is consistent with expectations. We get a similar finding when adding the other excluded instruments to the estimation, as reported in column 3.

The air connectivity instruments based on weather-related disruptions to air traffic provide mixed results. Recall that the weather-based instruments represent a weighted average of the weather conditions across all destination cities connected by non-stop flights to a given origin city. Focusing on column 2 of Table 5, the estimates indicate that a 1 percent increase in the average maximum temperature across all the destinations reached by non-stop service from a given city leads to a 0.77 percent increase in air connectivity. This result is statistically significant and consistent with prior expectations that flying to warmer areas, which are less prone to flight cancellations, improves a city's connectivity to the national aviation network. However, this effect goes away in column 3 when the other excluded instruments are added in (possibly because of a negative correlation between geographic distances and low temperatures). We also fail to identify any systematic relationship between the excluded instrument based on total snowfall and air connectivity. The small and insignificant coefficient in column 2, respectively in column 3, does not confirm the expectation that flying to snowier locations, which are more prone to flight cancellations, negatively impacts a city's connectivity to the national aviation network. This outcome may seem surprising given the evidence from Table 3 on the disruptive effect of snowfalls on flight schedules. One explanation that may reconcile these findings is that by differencing the data over 3-year intervals, we might remove important variation from the air connectivity instrument based on snowfall data. Nevertheless, one important thing in all three IV specifications is that the F-statistics for the excluded instruments are large enough to mitigate the problem of weak instruments (this is especially true when the distance-based instrument is used alone). For this reason, we view the proposed excluded instruments as valid in drawing causal inferences about the effect of air connectivity on urban development.

We next examine the performance of another economic outcome of great policy interest, i.e., employment growth. Table 6 reports the results from estimating the regression in equation (7) using total employment as dependent variable. The layout of the table follows the same sequence of model specifications as in Table 4. Column 1 presents the OLS estimates, while columns 2–4 report the LIML estimates using the distance-based IV, respectively the weather-based IVs, and then all of the IVs pooled together. Once again we find that the OLS effect of air connectivity on employment growth is small although, this time, marginally significant at conventional levels. On the other hand the IV estimates are much larger and generally significant. It is worth emphasizing the similar magnitude and significance level for the coefficient of interest in columns (2) and (4), the two IV specifications with the strongest set of excluded instruments. Focusing on the more comprehensive instrumental variables estimation in column 4, we find that a 10-percent increase in air connectivity leads to a 0.28 percent growth in total employment. To put results in perspective, note that a one standard deviation increase in air connectivity corresponds to a 230 percent change relative to the mean, which implies a 6.4 percent growth in employment.

Does the growth in employment translate into an increase in income? This could be the case if, for example, the growth in employment is associated with a compositional shift among firms, occupations or industries towards higher paid jobs. To investigate this question, we regress the 3-year changes in average per-capita income on a city's growth in air connectivity. We report the estimation results in Table 7. Following the same model specifications and the same set of excluded instruments as in the previous tables, we find no evidence of any changes in per-capita income as a result of a city's improvements in air connectivity. While the coefficient of interest has a positive sign across all specifications, the magnitude is very close to zero and an order of magnitude smaller than any prior estimates.

Finally, one last yet important economic outcome of interest is the number of establishments that enter and operate profitably in a city. Anecdotal evidence suggests that a city's aviation network is essential for its ability to attract entrepreneurial talent, thus having a direct effect on the business environment of that location. Using data on the total number of active establishments in a CBSA, we use our regression model to examine this hypothesis. We report the estimation results in Table 8. Both OLS and LIML estimates have the expected sign and are statistically significant, though much larger in magnitude in the instrumental variables specifications. Focusing on the results from the last column, which are estimated using all three excluded instruments, we find that a 10 percent increase in air connectivity over a three-year period leads to a 0.23 percent increase in the number of business establishments in a city. Given that a one standard deviation increase in air connectivity corresponds to a 230 percent change relative to the mean, this implies a 5.3 percent growth in the number of local business. This is an important result not only because new establishments provide more job opportunities and help raise total employment (consistent with our previous findings), but they also contribute to agglomeration effects responsible for positive spillover benefits.

Using information on the size category of establishments, we further examine which types of businesses benefit more from an improved aviation network. We define establishments with less than 20 employees as small size, and establishments with more than 500 employees as large size, with the remaining establishments classified as medium size. Table

9 reports the estimation results from the LIML specification based on all three exogenous instruments included in the estimation. For reference purposes, column 1 reproduces the results from column 4 in Table 8, using the total number of establishments as dependent variable. Comparing the estimated effects from columns 2–4, it seems that small and especially medium size firms benefit the most from better air connectivity. While it may seem surprising at first that large enterprises do not seem to respond to improvements in the network structure of a city's aviation services, it is possible that large firms stand to gain more when basing their location decisions on other location-specific characteristics such as the availability of production factors or of strategic input suppliers. It is also possible that large firms rely more extensively on private aviation services than small and medium firms do.

6 Conclusions

It is generally believed that good air connectivity is essential for conducting business activities and for enhancing firms' productivity, contributing towards overall regional economic growth. Rigorous empirical evidence to support these insights is relatively scarce, however.

Using a panel dataset on annual non-stop flights for approximately 150 CBSAs over the period 1984 – 2001, this study provides quantitative evidence on the gains from improved air connectivity for local economic growth. Our results show positive effects on population size, on employment levels and on the number of business establishments. Specifically, we find that a one standard deviation increase in a CBSA's air connectivity measure is associated with a 3 percent growth in city population, a 6.4 percent growth in total employment, and a 5.3 percent growth in the number of business establishments located in that CBSA. In light of the statistical strength of our exogenous instruments and of the long time period analyzed using a 3-year differences regression model, we think that these results capture the causal impact of a city's aviation network on its economic development. These findings have important economic significance as they support public initiatives that incentivize airlines to operate non-stop flights to new destinations and/or more flights to existing destinations.

This paper makes three contributions to the existing literature on the link between air connectivity and urban growth. First, our air connectivity measure is constructed to take into account the changes in the number of destinations and flights per route offered from a specific city. This measure exploits the extensive margin of passenger aviation, paying close attention to the actual structure of a city's aviation network. This aspect is important because many urban communities spend a lot of effort in expanding their aviation network with little understanding of the benefits expected in return. To our knowledge, no prior study in the field has focused on the extensive margin angle in order to characterize a city's provision of aviation services.

A second contribution of our study is the approach we take in dealing with the potential endogeneity issue. We propose two novel instruments that exploit information on the geography (i.e., distance) and weather patterns at destination cities. These are exogenous shocks to the number of flights operated between two cities. Finally, our econometric analysis employs a comprehensive dataset covering approximately 150 CBSAs observed over a span of eighteen years. The richness of our data allows us to employ panel techniques and exploit sources of variation that are not always feasible in this area of research.

Overall, the econometric analysis in this paper provides robust evidence that the size and the economic vitality of an urban location depends on the quality of its aviation network. These findings have important policy implications. They encourage city officials to use their efforts and resources to secure direct air service from their community to important gateway cities. Expanding the reach of their local aviation network has direct economic returns in the form of new business establishments and increased overall employment.

References

Anderson, T. W., Naoto Kunitomo, and Takamitsu Sawa, "Evaluation of the Distribution Function of the Limited Information Maximum Likelihood Estimator," *Econometrica*, 1982, *50*, 1009–1027.

Blonigen, Bruce A. and Anca D. Cristea, "Air Service and Urban Growth: Evidence from a Quasi-Natural Policy Experiment," *Journal of Urban Economics*, 2015, *86*, 128–146.

Brueckner, Jan K., "Airline traffic and urban economic development," *Urban Studies*, 2003, *40* (8), 1455–1469.

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

_ **and Richard Hornbeck**, "Railroads and American Economic Growth: A Market Access Approach," *Quarterly Journal of Economics.* forthcoming.

Duranton, Gilles and Matthew A. Turner, "Urban Growth and Transportation," *The Review of Economic Studies*, 2012, *79* (4), 1407–1440.

Erickson, David, "Missoula Airport Trying to Attract Direct Flight Service to Texas Cities," October 17, 2017. http://missoulian.com/news/local/missoula-airport-trying-to-attract-direct-flight-service-to-texas/article_6efe19e3-e40f-5f42-9c90-da802c23304f.html.

Faber, Benjamin, "Trade Integration, Market Size, and Industrialization: Evidence from Chinas National Trunk Highway," *Review of Economic Studies.* forthcoming.

Fechter, Samantha Ehlinger Joshua and Rye Druzin, "New San Antonio Airport, Expanded Facilities on the Table under New Nirenberg Initiative," October 10, 2017. http://www.expressnews.com/business/local/article/New-San-Antonio-airport-expanded-facilities-on-12268129.php.

Forbes, Silke J. and Mara Lederman, "Adaptation and Vertical Integration in the Airline Industry," *American Economic Review*, 2009, *99*, 1831–1849.

Fujita, M., P. Krugman, and A. Venables, *The Spatial Economy: Cities, Regions, and International Trade*, MIT Press, Cambridge, 1999.

Green, Richard K., "Airports and Economic Development," *Real Estate Economics*, 2007, *35* (1), 91–112.

Griliches, Zvi and Jerry A. Hausman, "Errors in Variables in Panel Data," *Journal of Econometrics*, 1986, *31*, 93–118.

Hanson, Gordon, "Market Potential, Increasing Returns, and Geographic Concentration," *Journal of International Economics*, 2005, *67*, 1–24.

Harris, Chauncy D., "The Market as a Factor in the Localization of Industry in the United States," *Annals of the Association of American Geographers*, 1954, *44*, 315–348.

Head, Keith and Thierry Mayer, "Regional Wage and Employment Responses to Market Potential in the EU," *Regional Science and Urban Economics*, 2006, *36*, 573–594.

Krugman, Paul, "Increasing Returns and Economic Geography," *Journal of Political Economy*, 1991, *99*, 481–499.

LeFors, Michael, "The Role of Air Accessibility in Urban Development," 2015. University of California Irvine, mimeo.

McGraw, Marquise J., "Perhaps the Sky's the Limit? Airports and Employment in Local Economies," 2017. Working paper, www.marquisemcgraw.com/research.

Poling, Travis E. and William Pack, "AT&T Leaving San Antonio for Dallas," Jun 28, 2008. http://www.mysanantonio.com/business/article/AT-T-leaving-San-Antonio-for-Dallas-1758611.php.

Redding, Stephen and Anthony Venables, "Economic Geography and International Inequality," *Journal of International Economics*, 2004, *62*, 53–82.

Sheard, Nicholas, "Airports and Urban Sectoral Employment," *Journal of Urban Economics*, 2014, *80*, 133–152.

— , "Airport Size and Urban Growth," 2015. University of Liverpool, mimeo.

A Tables

Table 1: Summary Statistics

	Obs	Mean	St. Dev.	Min	Max
Aviation Network Variables:					
Δ_3 Ln Air Connectivity	2205	0.130	0.428	-1.996	3.519
Δ_3 Ln Air Connect IV (w/ Distance)	2205	0.006	0.115	-0.407	1.301
Δ_3 Ln Air Connect IV (w/ Max. Temp)	2205	0.019	0.121	-0.369	1.125
Δ_3 Ln Air Connect IV (w/ Snowfall)	2204	-0.085	0.859	-7.435	6.933
Δ_3 Ln Number of Flight Cancellations	1840	0.156	0.722	-3.709	3.117
Δ_3 Share of Delayed Departures	1340	0.509	3.118	-9.751	9.964
Urban Economic Indicators:					
Δ_3 Ln Population	2205	0.039	0.036	-0.082	0.232
Δ_3 Ln Population Lag (50 years)	2205	0.062	0.055	-0.041	0.566
Δ_3 Ln Total Employment	2205	0.079	0.061	-0.248	0.417
Δ_3 Ln Personal Income	2205	0.049	0.041	-0.149	0.315
Δ_3 Ln Number of Establishments	2205	0.049	0.045	-0.150	0.226
Δ_3 Ln Number of Small Size Establishments	2205	0.044	0.045	-0.143	0.226
Δ_3 Ln Number of Medium Establishments	2205	0.083	0.066	-0.242	0.643
Δ_3 Ln Number of Large Establishments	2188	0.095	0.210	-1.099	1.792
Δ_3 Ln Market Access	2205	0.089	0.035	-0.001	0.201
Weather Variables:					
Δ_3 Avg. Monthly Max. Temperature (^{o}C)	1884	0.114	1.171	-4.325	4.083
Δ_3 Total Annual Snowfall (cm)	1781	0.444	451.9	-2802	2366

Notes: The estimation sample covers 147 CBSAs observed over the period 1984–2001. The data sources and the construction of variables are described in the paper. Δ_3 denotes 3-year differences in variables. So, for any variable X we have $\Delta_3 X_t \equiv X_t - X_{t-3}$ with $t \geq 1997$.

Table 2: Effect of Weather Conditions on Flight Cancellations

	Dependent variable: Δ_3 Ln Number of Flight Cancellations		
	(1)	(2)	(3)
Δ_3 Avg. Monthly Max Temp (^{o}C)	-0.075***		-0.056**
	[0.022]		[0.022]
Δ_3 Total Annual Snowfall (cm)		0.002***	0.002***
		[0.001]	[0.001]
Δ_3 Ln Market Access	4.805**	6.628***	5.804***
	[1.882]	[1.875]	[1.925]
Δ_3 Ln Population	0.117	0.050	0.051
	[0.378]	[0.401]	[0.398]
Period fixed effects	yes	yes	yes
Observations	1,568	1,489	1,489
R-squared	0.134	0.138	0.143

*** p<0.01, ** p<0.05, * p<0.10. Robust standard errors clustered at CBSA level in brackets

Notes: The regression model is estimated in 3-year differences, so $\Delta_3 X \equiv X_t - X_{t-3}$ for any variable X. The sample period consists of three-year intervals between 1984 – 2001. Flight cancellations are calculated as the difference between the number of scheduled and performed departures. To remove outliers, the top and bottom 5 percent of flight cancellation observations have been dropped. The F-statistic for the joint significance of all the weather variables from the specification reported in column 3 is 9.72 with a p-value of 0.000.

Table 3: Effect of Weather Conditions on Departure Delays

	Dependent variable: Δ_3 Share of Delayed Departures		
	(1)	(2)	(3)
Δ_3 Avg. Monthly Max Temp (^{o}C)	-0.327***		-0.257*
	[0.125]		[0.133]
Δ_3 Total Annual Snowfall (cm)		0.006***	0.005***
		[0.002]	[0.002]
Δ_3 Ln Market Access	9.076	11.153	8.402
	[7.391]	[8.392]	[7.953]
Δ_3 Ln Population	3.287	3.689	3.843
	[2.913]	[2.971]	[2.911]
Period fixed effects	yes	yes	yes
Observations	1,173	1,088	1,088
R-squared	0.355	0.365	0.370

*** p<0.01, ** p<0.05, * p<0.10. Robust standard errors clustered at CBSA level in brackets

Notes: The regression model is estimated in 3-year differences, so $\Delta_3 X \equiv X_t - X_{t-3}$ for any variable X. The sample period consists of three-year intervals between 1988 – 2001 (the shorter time period compared to 1984–2001 is dictated by the availability of flight cancelation data). A flight departure is considered delayed if it departs from the gate 15 or more minutes later than the scheduled departure time. To remove outliers, the top and bottom 5 percent of flight cancellation observations have been dropped. The F-statistic for the joint significance of all the weather variables from the specification reported in column 3 is 7.89 with a p-value of 0.001.

Table 4: Effect of Aviation Network on Urban Population

	Dependent variable: Δ_3 Ln Population			
	(1) OLS	(2) IV-LIML (Distance)	(3) IV-LIML (Weather)	(4) IV-LIML (All)
Δ_3 Ln Air Connectivity	0.003	0.013*	0.019	0.012*
	[0.002]	[0.008]	[0.017]	[0.007]
Δ_3 Ln Market Access	1.040***	1.003***	0.981***	1.004***
	[0.145]	[0.147]	[0.140]	[0.148]
Δ_3 Ln Population Lag (50 years)	0.161**	0.162**	0.162**	0.161**
	[0.081]	[0.081]	[0.081]	[0.081]
Period fixed effects	yes	yes	yes	yes
Observations	2,205	2,205	2,204	2,204
R-squared	0.220	0.206	0.184	0.207
First Stage Statistics				
F-stat		31.98	9.03	12.20
Hansen J stat		n.a.	1.612	2.368
Hansen J p-val		n.a.	0.204	0.306

*** p<0.01, ** p<0.05, * p<0.10. Robust standard errors clustered at CBSA level in brackets.

Notes: The reported results correspond to the regression model in equation (7) estimated by ordinary least squares (OLS) and by limited information maximum likelihood (LIML) methods. The unit of observation is a CBSA observed in a given year. The panel dataset includes 147 CBSAs and covers the period 1984–2001. All the variables are expressed in 3-year differences (hence, by construction, the first three sample years are dropped). The dependent variable is the CBSA population size, and the variable of interest is the air connectivity measure. The variables construction is described in equations (6), (8) and (9), respectively. The first stage coefficients from the estimations reported in columns 2–4 are provided in Table 5. The 50-year lag of city population and the market access index serve as control variables. All regression specifications include period fixed effects, and the reported standard errors are clustered at the city level.

Table 5: First Stage Regressions

	Dependent var: Δ_3 Ln Air Connectivity		
	(1)	(2)	(3)
	IV-LIML	IV-LIML	IV-LIML
	(Distance)	(Weather)	(All)
Δ_3 Ln Market Access	0.957	2.743**	0.900
	[1.049]	[1.111]	[1.060]
Δ_3 Ln Population Lag (50 years)	-0.184	-0.144	-0.176
	[0.294]	[0.298]	[0.291]
Excluded Instruments			
Δ_3 Ln Air Connect IV (w/ Distance)	1.204***		1.306***
	[0.213]		[0.365]
Δ_3 Ln Air Connect IV (w/ Ave. Max Temp)		0.768***	-0.145
		[0.187]	[0.232]
Δ_3 Ln Air Connect IV (w/ Snowfall)		0.005	-0.002
		[0.017]	[0.018]
Period fixed effects	yes	yes	yes
Observations	2,205	2,204	2,204
R-squared	0.129	0.075	0.129
First Stage Statistics			
F stat	31.98	9.03	12.20
F p-val	0.000	0.000	0.000

*** p<0.01, ** p<0.05, * p<0.10. Robust standard errors clustered at CBSA level in brackets

Notes: The reported results correspond to the first stage coefficients obtained from estimating the regression model in equation (7) by limited information maximum likelihood (LIML). The set of excluded instruments is described by equations (8) and (9). The partial F statistics for the excluded instruments are reported at the bottom of the table. All the variables are expressed in 3-year differences. The dependent variable represents the air connectivity measure, which is the endogenous variable in our main regression model. The reported first stage estimates are independent of the actual dependent variable employed in the second stage estimation. All the reported regression specifications include period fixed effects, and the reported standard errors are clustered at the city level.

Table 6: Effect of Aviation Network on Total Employment at CBSA Level

	Dependent variable: Δ_3 Ln Employment			
	(1) OLS	(2) IV-LIML (Distance)	(3) IV-LIML (Weather)	(4) IV-LIML (All)
Δ_3 Ln Air Connectivity	0.007*	0.026*	0.010	0.028**
	[0.004]	[0.013]	[0.025]	[0.014]
Δ_3 Ln Market Access	2.419***	2.350***	2.408***	2.343***
	[0.230]	[0.230]	[0.231]	[0.231]
Δ_3 Ln Population Lag (50 years)	0.007	0.008	0.007	0.008
	[0.096]	[0.096]	[0.096]	[0.096]
Period fixed effects	yes	yes	yes	yes
Observations	2,205	2,205	2,204	2,204
R-squared	0.268	0.251	0.268	0.248
First Stage Statistics				
F-stat		31.99	9.030	12.20
Hansen J stat		n.a.	5.048	4.677
Hansen J p-val		n.a.	0.025	0.097

*** p<0.01, ** p<0.05, * p<0.10. Robust standard errors clustered at CBSA level in brackets.

Notes: The reported results correspond to the regression model in equation (7) estimated by ordinary least squares (OLS) and by limited information maximum likelihood (LIML). The unit of observation is a CBSA observed in a given year. The panel dataset includes 147 CBSAs and covers the period 1984–2001. All the variables are expressed in 3-year differences (hence, by construction, the first three sample years are dropped). The dependent variable represents the total employment level in a CBSA, and the variable of interest is the air connectivity measure. The variables construction is described in equations (6), (8) and (9), respectively. The first stage coefficients from the estimations reported in columns 2–4 are provided in Table 5. The 50-year lag of city population and the market access index serve as control variables. All regression specifications include period fixed effects, and the reported standard errors are clustered at the city level.

Table 7: Effect of Aviation Network on average Personal Income at CBSA Level

	Dependent variable: Δ_3 Ln Personal Income			
	(1) OLS	(2) IV-LIML (Distance)	(3) IV-LIML (Weather)	(4) IV-LIML (All)
Δ_3 Ln Air Connectivity	0.002	0.005	-0.003	0.006
	[0.002]	[0.007]	[0.013]	[0.006]
Δ_3 Ln Market Access	1.272***	1.261***	1.290***	1.259***
	[0.138]	[0.140]	[0.139]	[0.141]
Δ_3 Ln Population Lag (50 years)	-0.130***	-0.130***	-0.130***	-0.130***
	[0.032]	[0.032]	[0.032]	[0.032]
Period fixed effects	yes	yes	yes	yes
Observations	2,205	2,205	2,204	2,204
R-squared	0.459	0.458	0.456	0.458
First Stage Statistics				
F-stat		31.99	9.030	12.20
Hansen J stat		n.a.	2.020	2.244
Hansen J p-val		n.a.	0.155	0.326

*** p<0.01, ** p<0.05, * p<0.10. Robust standard errors clustered at CBSA level in brackets

Notes: The reported results correspond to the regression model in equation (7) estimated by ordinary least squares (OLS) and by limited information maximum likelihood (LIML) methods. The unit of observation is a CBSA observed in a given year. The panel dataset includes 147 CBSAs and covers the period 1984–2001. All the variables are expressed in 3-year differences (by construction the first three sample years are dropped). The dependent variable represents the average personal income in a CBSA, and the variable of interest is the air connectivity measure. The variables construction is described in equations (6), (8) and (9), respectively. The first stage coefficients from the estimations reported in columns 2–4 are provided in Table 5. The 50-year lag of city population and the market access index serve as control variables. All regression specifications include period fixed effects, and the reported standard errors are clustered at the city level.

Table 8: **Effect of Aviation Network on the Number of Local Businesses**

	Δ_3 Dependent variable: Ln No. Establishments			
	(1) OLS	(2) IV-LIML (Distance)	(3) IV-LIML (Weather)	(4) IV-LIML (All)
Δ_3 Ln Air Connectivity	0.009***	0.022***	0.019	0.023***
	[0.003]	[0.008]	[0.015]	[0.008]
Δ_3 Ln Market Access	1.891***	1.845***	1.857***	1.843***
	[0.178]	[0.177]	[0.174]	[0.178]
Δ_3 Ln Population Lag (50 years)	-0.014	-0.013	-0.014	-0.013
	[0.088]	[0.087]	[0.087]	[0.087]
Period fixed effects	yes	yes	yes	yes
Observations	2,205	2,205	2,204	2,204
R-squared	0.298	0.284	0.290	0.283
First Stage Statistics				
F-stat		31.99	9.030	12.20
Hansen J stat		n.a.	3.468	3.366
Hansen J p-val		n.a.	0.0626	0.186

*** p<0.01, ** p<0.05, * p<0.10. Robust standard errors clustered at CBSA level in brackets

Notes: The reported results correspond to the regression model in equation (7) estimated by ordinary least squares (OLS) and by limited information maximum likelihood (LIML) methods. The unit of observation is a CBSA observed in a given year. The panel dataset includes 147 CBSAs and covers the period 1984–2001. All the variables are expressed in 3-year differences (by construction the first three sample years are dropped). The dependent variable represents the number of business establishments operating in a CBSA, and the variable of interest is the air connectivity measure. The variables construction is described in equations (6), (8) and (9), respectively. The first stage coefficients from the estimations reported in columns 2–4 are provided in Table 5. The 50-year lag of city population and the market access index serve as control variables. All regression specifications include period fixed effects, and the reported standard errors are clustered at the city level.

Table 9: Effect of Aviation Network on the Number of Local Businesses

	Dependent variable: Δ_3 Ln No. Establishments			
	Total	***Small Size***	***Medium Size***	***Large Size***
	IV-LIML	IV-LIML	IV-LIML	IV-LIML
	(1)	(2)	(3)	(4)
Δ_3 Ln Air Connectivity	0.023***	0.018**	0.064***	-0.073
	[0.008]	[0.007]	[0.019]	[0.054]
Δ_3 Ln Market Access	1.843***	1.786***	2.305***	2.923***
	[0.178]	[0.177]	[0.234]	[0.620]
Δ_3 Ln Population Lag (50 years)	-0.013	-0.009	-0.036	0.340***
	[0.087]	[0.085]	[0.108]	[0.091]
Period fixed effects	yes	yes	yes	yes
Observations	2,204	2,204	2,204	2,187
R-sq	0.283	0.294	0.153	0.067
First Stage Statistics				
F-stat	12.20	12.20	12.20	12.17
Hansen J stat	3.366	2.243	7.892	0.528
Hansen J p-val	0.186	0.326	0.019	0.768

*** p<0.01, ** p<0.05, * p<0.10. Robust standard errors clustered at CBSA level in brackets

Notes: The reported results correspond to the regression model in equation (7) estimated by limited information maximum likelihood (LIML) methods using the full set of excluded instruments. The same explanations as included in the footnote for Table 8 apply. The first stage coefficients for the estimations reported here correspond to those provided in column 3 of Table 5. The dependent variable in each of the reported specifications is the number of business establishments by size category. Small size establishments are defined as having less than 20 employees, while large size establishments are defined as having 500 employees and above. The remaining establishments are classified as medium size.