The Role of Aviation Networks for Urban Development*

Anca D. Cristea†
University of Oregon

December, 2022

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 1984-2013, we propose a new measure for a metropolitan area's connectivity to the national aviation network. We then use this measure to investigate its contribution to local economic development, as captured 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 destination airports growth as a way to capture the exogenous variation in the likelihood to add new travel routes. Our results suggest that a metropolitan area's air connectivity, resulting from an expansive local aviation network, has a positive effect on population, on employment and on the number of businesses established in that location.

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

*I am grateful to the support and contributions of Liliana Danila throughout this project. I thank the editor and the two anonymous referees for their helpful comments and suggestions. I also thank Bruce Blonigen, Jan Brueckner, Andrew Cassey, Mark Colas, 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. I am 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 my own.

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

1 Introduction

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

A plethora of examples exists to support these claims. In October 2017, a coalition of business groups in the city of Missoula, MT launched an initiative to attract more direct flights to the Missoula International Airport, stating that such a move would boost the economy.[1] Around the same time, the city of San Antonio was examining whether to expand its airport, citing that direct flights to major cities "are an obstacle [to the development of the local economy]". A case in point: San Antonio's poor air connectivity was considered a major drawback in the city's bid to attract Amazon's second headquarters.[2] Back in 2008, AT&T moved its headquarters from San Antonio to Dallas for a similar reason. The AT&T senior 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 importance of a city's aviation *network* for local economic development.[5] How valuable is it for a city to add more non-stop routes and to connect to new destination markets? Or is it enough to simply reinforce seat capacity for existing destinations? Does the expansion and reach of a city's aviation network matter for economic growth and development? We try to address some of these questions in our paper.

At first glance, these questions seem to have straightforward answers. A wider aviation network should benefit a metropolitan area as it reduces the time cost incurred by economic agents to reach markets nationwide. However, an expansive aviation network can increase the monetary cost of travel as serving multiple destinations by direct flights may decrease airline competition within each city-pair market, and erode any economies of scale from employing large aircrafts. Adding all the costs and benefits, it becomes less obvious whether

[1] See Erickson (2017): http://missoulian.com/news

[2] See Joshua Fechter and Druzin (2017): http://www.expressnews.com/business/local

[3] See Poling and Pack (2008): http://www.mysanantonio.com/business

[4] See, for example, Brueckner (2003), Green (2007), Blonigen and Cristea (2015), Sheard (2019), McGraw (2020), Gibbons and Wu (2020), Sheard (2021).

[5] Throughout this paper we use the term "city" to refer to a core-based statistical area (CBSA).

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 metropolitan area's aviation network to its economic development. We propose a novel measure to capture the air connectivity of a location to the wider national aviation network. Motivated by the product differentiation literature, this measure is constructed as a weighted count of all destinations reached by non-stop flight service from a given origin location, where the weights take into account: 1) the importance of each destination airport within the national aviation network, 2) the frequency of flight service to these destinations, and 3) the degree of substitution between aviation routes from the perspective of consumers. 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, employment level, or the number of new businesses that choose to locate in that metropolitan area.

The main econometric challenge in answering our research question comes from the interdependence between a location's economic development and the expansion 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 propose two related exogenous variables to instrument for a metropolitan area's connectivity to the national aviation network. Inspired by network centrality indexes, we exploit information on physical geography (i.e., distance between locations) and on destination airport characteristics (i.e., network nodes) to predict the extent to which a metropolitan area will offer non-stop flight services to a given destination. By design, our excluded instruments only exploit time-variation coming from the growth of air traffic at destination cities. This variation is then used to infer the expected gains for an origin location from improvements in air connectivity.

A related concern and potential source of endogeneity comes from the fact that the connectivity of a metropolitan area to the national aviation network 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 (Redding and Venables, 2004). At the same time, proximity to large markets may increase the likelihood of nonstop service to those markets, increasing a location's air connectivity. To ensure that our estimation results do not suffer from omitted variable bias, we include in our preferred model specification multiple controls for the market access of a given location.

Using historical data on business activity and air passenger traffic for 262 consolidated core-based statistical areas (CBSAs) spanning the period 1984-2013, 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 an increase in a city's air connectivity equal in size to the observed sample mean leads to 0.1 percent growth in urban population, on average. We find a larger effect on total employment at the local level, equal to 0.4 percent growth. The estimated increases in population 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, evaluated at the sample mean, an increase in air connectivity leads to 0.3 percent growth in the number of local businesses. This growth is driven primarily by small and medium size firms. Our results are robust to using 3-year differences in variables to remove unobservable time-invariant location effects, in addition to using a long list of control variables and fixed effects.

Our findings contribute to a growing literature on the role of air transportation infrastructure for regional development. Brueckner (2003) and Green (2007) are among the first to investigate the role of air passenger transport for population and employment growth. Using distance to the US population center of gravity as instrument for an airport's volume of air traffic (i.e., likelihood of being a hub), both studies find large positive effects of airport infrastructure on local employment and population growth. More recently, Blonigen and Cristea (2015) exploit the 1978 Airline Deregulation Act to identify exogenous variation in the provision of air services across US metropolitan areas and estimate the contribution of passenger aviation to urban growth. Sheard (2014) and McGraw (2020) use historical information from the 1944 National Airport Plan and the 1920s location of emergency air field (as defined by the Army Air Service) to develop novel instruments for airport infrastructure, respectively a novel identification strategy, and bring further evidence on the causal effect of air services on industrial composition and employment growth. Sheard (2019, 2021) revisits the evidence and proposes several Bartik (1991) style instruments to predict changes in airport traffic independent of local employment effects.

All the evidence discussed so far uses data for metropolitan areas in the United States and local economic development measures such as employment or population growth. Campante and Yanagizawa-Drott (2018) expand the analysis to the global level and bring information from a sample of 819 cities worldwide whose economic activity is captured by satellite-measured night lights. Using a discontinuity in the provision of non-stop flights beyond 6000 miles determined by technological or regulatory constraints, Campante and Yanagizawa-Drott (2018) provide convincing evidence that non-stop flights have a causal effect on local economic activity. Increased foreign capital, as measured by the number of foreign owned firms attracted to highly connected cities, represents one of the channels of economic growth. Using Chinese data on firm performance and on new airport construction, Gibbons and Wu (2020) bring additional micro-level evidence for the effect of airports on local economic performance. They use as exogenous variable of interest an airport centrality index that has similarities to our excluded instruments.

In addition to research focused on air transportation, several studies have examined other modes of transport such as road or rail transport to undertand the role of transport infrastructure for urban growth and industry specialization. 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 (2014) 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 (2018) and Donaldson and Hornbeck (2016) quantify the importance of railroads in improving a location's market access, generating income growth and welfare.

Most of the literature investigating the economic role of transport infrastructure, and of airport services in particular, makes no direct attempt at distinguishing between the effects of scale (e.g., volume of air traffic) versus scope (e.g., network of non-stop destinations). We think our paper is one of the first studies to emphasize this distinction by focusing on the contribution of the local aviation network for economic growth. The papers that are perhaps the closest to ours are Sheard (2019) and Campante and Yanagizawa-Drott (2018). They employ measures that arguably capture the extensive margin of a city's aviation network. For example, Sheard (2019) defines the "air access" of a metropolitan area as the number of flights operated to each non-stop destination, weighted by the population size of that destination. This measure is similar to our proposed measure of air connectivity, the main differences being that our weights consist of the shares in total departures accounted by each destination airport (i.e., the "hub-ness" of the destination airport), in addition to factoring in information on the elasticity of substitution between destination routes when evaluating flight frequency. While "air access" is not the preferred airport infrastructure measure used in Sheard (2019), its estimated effect on local employment growth is similar to our estimate.[6] Campante and Yanagizawa-Drott (2018) construct a measure of a location's

[6]It is important to note that while the "air access" measure used by Sheard (2019) and the "air connectivity" measure defined in this study are fairly similar in construction, the estimation models employed by the two studies are different, which demands caution in making direct comparisons between the magnitude of the estimated coefficients. First, Sheard (2019) estimates the model in first differences while our preferred specification employs three-year differences in variables (we perform robustness checks using year-on-year changes for a more direct comparison). Second, Sheard (2019) instruments for "air access" using a set of Bartik-style variables that leverage very different sources of exogenous variation. Third, the CBSA coverage

aviation network centrality, which captures the idea that a direct link to a well-connected destination airport is more economically meaningful than a link to a poorly-connected one. This airport centrality measure is similar to our proposed "air connectivity" measure in that they both use information on the air links of destination airports (i.e., the "hub-ness" of the destination airports) to assess their importance for the origin city. One key difference is that our measure also accounts for the frequency of non-stop flights to those destinations (subject to the elasticity of substitution between them). This may matter since infrequent connections from an origin city to a very large airport hub may hinder access to the services provided by that hub no matter how big that hub might be. Nevertheless, the airport centrality measure used by Campante and Yanagizawa-Drott (2018) provides qualitatively similar findings in that it contributes directly to local economic growth (measured by night lights). The mechanism emphasized in the paper is the increase in the number of foreign-owned businesses attracted to that location. Overall, we think that our proposed measure of "air connectivity" complements the existing studies, reinforcing their positive findings.

The remainder of our paper proceeds as follows. Section 2 introduces our proposed measure of air connectivity, which is meant to capture the extensive margin of a city's local aviation network. The econometric model and estimation strategy are described in section 3. Section 4 presents the data sources and variable construction. Section 5 discusses the estimation results, while section 6 examines the robustness of the main findings. Section 7 discusses the main policy implications and concludes.

2 Air Connectivity: An Extensive Margin Measure

Airlines respond to the growing demand 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 non-stop 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 metropolitan area, making little effort to distinguish between the various dimesions 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 service to the overall economic growth of cities.

Given the efforts spent by local communities to expand their aviation network, we think it is worth focusing on a more refined measure of a city's aviation network that better

and the sample periods differ between the two studies, with Sheard (2019) covering a shorter time period (1991– 2015) compared to our study (1984–2013). Lastly, our model specification uses a more extensive set of control variables, in addition to time and location fixed effects. This might explain the slightly smaller coefficient magnitude that we find.

captures the extensive margin dimension of air traffic. We label this proposed measure "*air connectivity*" and examine its contribution to economic development and urban growth.

To define the air connectivity measure of a metropolitan area, we use insights from the product differentiation literature where the representative consumer has "love of variety" preferences over a differentiated good (Dixit and Stiglitz, 1977; Spence, 1976). For our specific purposes, the differentiated good will be air travel. Thinking of non-stop flight destinations (i.e., aviation routes) as varieties of the differentiated good of interest (i.e., air travel), we define the utility of a representative consumer in city i over their consumption of aviation flights using a constant elasticity of substitution (CES) function:

$$U_i = \left(\sum_{d=1}^{N_i} \alpha_d f_{id}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \tag{1}$$

where d indexes the destination cities to which origin city i is connected to via non-stop air service, N_i is the total number of non-stop destinations reached from origin i, α_d denotes a preference parameter for destination d, f_{id} denotes the flight frequency of air service between origin i and destination d, and $\sigma > 1$ denotes the elasticity of substitution between flight departures.

The motivation to use a CES framework to model the demand side of aviation markets comes from Berry (1990), who first introduced the idea that the services provided by airlines at hub airports can be modeled as a form of product differentiation. Differentiated products generally command a higher price mark-up over the marginal cost without impacting negatively the quantities sold, a feature that is consistent with air traffic observed at hub airports. To rationalize this empirical regularity of incumbent airlines charging a price premium at hub airports while also capturing large market shares, Berry (1990) argues that consumers must place a high value on flight frequency and on the number of direct routes offered by airlines from hub airports (hence the product differentiation set-up). Other studies followed up on this idea, resorting either to random coefficients discrete choice utility or to CES utility frameworks to model the demand side of aviation markets (see Berry and Jia, 2010; Berry et al., 1996; Mitra et al., 2018, among others).

The appeal of equation (1) is that it maps the extensive margin of aviation services offered out of location i – as captured by the number of non-stop destinations (N_i) and the frequency of flight departures per destination (f_{id}) – into consumer utility units. The value that the consumers place on the spatial and temporal diversity of air service depends crucially on the elasticity of substitution parameter σ. A key property of the CES utility function is that for finite values of σ, U_i in equation (1) grows faster when adding new routes (i.e., increasing N_i) compared to adding new departures on existing routes (i.e., increasing

f_{id}). That is, the representative consumer has a stronger preference for consuming a new route variety (i.e., flying to a new destination) instead of increasing its consumption of an existing route variety (i.e., flying more frequently to an existing destination). Hence the "love of variety" terminology.

A quick and easy way to illustrate the "love of variety" property of the CES utility function is to simplify equation (1) – for the sake of example only – by assuming demand symmetry. When flight frequency is the same across all destinations (i.e., $f_{id} = f \ \forall i, \forall d$), and all destinations are equally important to the representative consumer (i.e., $\alpha_d = \alpha = 1 \ \forall d$), then consumer utility becomes $U_i = f N_i^{\frac{\sigma}{\sigma-1}}$. If the total number of flights currently consumed (i.e., Nf) were to double, it becomes clear that doubling the flight frequency on all the existing N routes to a level $2f$ would generate a smaller increase in utility level than doubling the number of non-stop destinations to reach a level of $2N$. That is: $(2f)(N_i)^{\frac{\sigma}{\sigma-1}} < f(2N_i)^{\frac{\sigma}{\sigma-1}}$ for $\sigma > 1$ (even if both cases involve an increase by Nf in flights consumed). Furthermore, the smaller the value of $\sigma > 1$, the stronger is the "love for variety".

Building on this framework and intuition, we define the air connectivity of city i to the national aviation network as the expansion of air traffic due to the introduction of new flights to new and existing destinations. From equation (1), we focus on the term in brackets for simplicity[7], and define the air connectivity of city i as:

$$AirConnect_i \equiv \sum_{d=1}^{N_i} \alpha_d {f_{id}}^{\frac{\sigma-1}{\sigma}} \tag{2}$$

To construct the air connectivity index from observable data, we need to define the preference parameter α_d, and we need to set a value for σ. Berry and Jia (2010) argue that, conditional on airfares, travelers have a "preference for convenience". So, one possibility is to measure the value that consumers place on a destination d by the characteristics of that airport "node" within the national aviation network. Airport characteristics such as the number of daily departures or the number of nonstop destinations represent airport features that generate direct utility to consumers.[8] Moreover, the idea of valuing air service to a given destination by the"quality" of the air link being established through that flight service is an idea that has been encountered in transportation literature as well when assessing the air connectivity of a location (see, for example, Allroggen et al., 2015).

[7]In our empirical analysis we will use a log transformation of the air connectivity measure, so leaving out the exponent $\sigma/(\sigma-1)$ over the summation term in equation (1) is without loss of generality.

[8]Berry (1990) and Berry et al. (1996) bring direct evidence to suggest that consumers value airlines offering a large number of routes out of a destination airport, all else equal.

Motivated by these ideas, we assume that the value placed by the representative consumer from city i on nonstop flights to a given destination d comes from the "hub-ness" of that destination. So, we measure α_d by the fraction of all (nationwide) flight departures operated out of destination airport d:

$$\alpha_d \equiv \frac{f_d}{f} = \frac{\sum_j f_{dj}}{\sum_d \sum_j f_{dj}} \tag{3}$$

where $f \equiv \sum_d \sum_j f_{dj}$ denotes the total number of flight departures operated out of all locations, and $f_d \equiv \sum_j f_{dj}$ denotes the total number of flight departures operated out of city d.

Substituting equation (3) into equation (2), and adding a time subscript t to capture network changes over time, the air connectivity measure becomes:

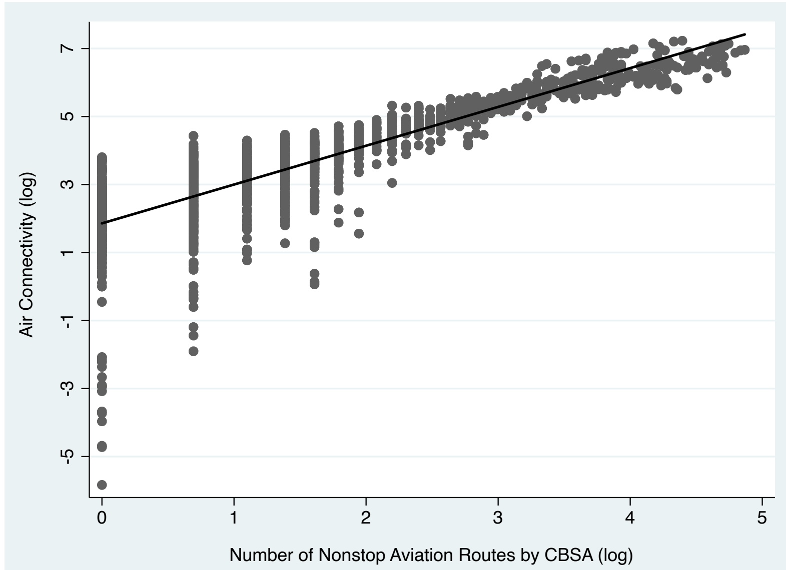
$$AirConnect_{it} \equiv \sum_{d=1}^{N_{it}} \left(\frac{f_{dt}}{f_t}\right) f_{idt}^{\frac{\sigma-1}{\sigma}} \tag{4}$$

One way to think of the proposed air connectivity measure is as a *weighted count* of all destinations d that city i is able to reach via non-stop air service, where we weight the importance of each destination d based on how well destination d is connected to the national aviation network (i.e., the "hub-ness" of destination d), and based on the frequency of service between city i and each destination d. A large value of the air connectivity measure indicates that a city has frequent access to multiple large hub airports.

In constructing the air connectivity index from available data, we have to take a stand on the size of the elasticity of substitution σ. Unfortunately, we are not aware of any study that estimates how substitutable aviation routes are from the perspective of consumers. Absent such an industry-specific parameter estimate, we rely on the international trade literature where this elasticity plays an important role, and borrow the modal value corresponding to tradeable goods. So, we set σ equal to 5.[9] In the data section, we further motivate our parameter choice using our own back-of-the-envelope calculations. We also verify the sensitivity of our results to alternative σ values in robustness exercises.

Figure 1 displays the correlation between the proposed air connectivity measure and a simple count of non-stop aviation routes observed among the sample metropolitan areas during our sample period, 1984-2013. While the two measures are related by construction,

[9] Anderson and van Wincoop (2004) survey the trade literature that estimates the elasticity of substitution parameter and find values ranging between 5 and 10. More recently, Simonovska and Waugh (2014) identify parameter values between 2 and 5 estimated using the latest methodological developments. Thus, $\sigma = 5$ has become the modal value for the elasticity of substitution in the trade literature.



Source: Authors' Calculations

Figure 1: Air Connectivity and Number of Non-stop Routes among CBSAs

Note: The figure displays the correlation between a city's air connectivity and the number of non-stop aviation routes operated out of that city, in log terms. Each observation corresponds to one of the 262 consolidated CBSAs in our sample observed every 4 years over the period 1984-2013.

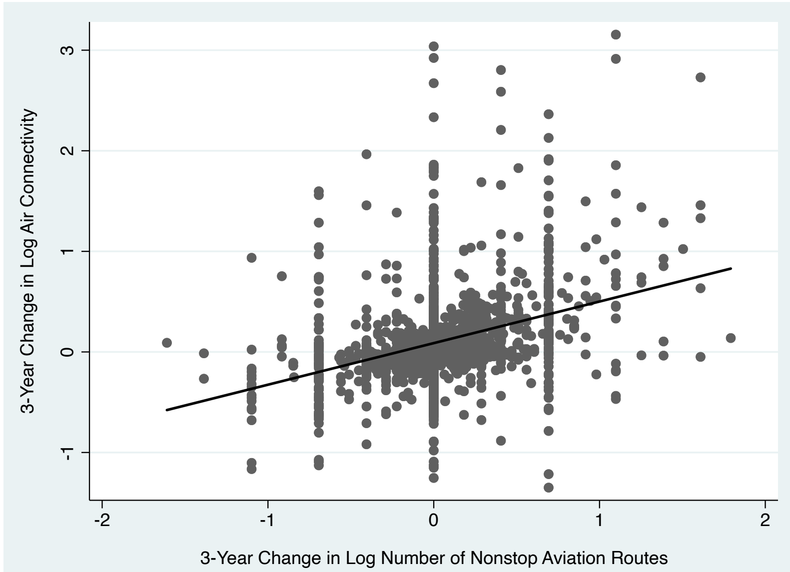
the tight correlation between them is illustrative of the strong driving force represented by the extensive margin in determining the magnitude of the air connectivity index.

Figure 2 provides a similar scatterplot but instead of reporting the variables in levels, it plots 3-year differences in air connectivity, respectively in the number of non-stop aviation routes, across the sample metropolitan areas. The fact that the correlation between the two variables is maintained after removing location-specific effects through data differentiation speaks to the significant time variation available in the data. The econometric analysis will exploit this source of temporal variation. Our goal is to examine the extent to which our proposed measure of 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 endogeneity, which could be caused by omitted variables and by reverse causality. Since several unobservable factors may affect both variables of interest (i.e., air connectivity and local development), a first step towards mitigating the omitted variable problem is to take advantage of the panel structure of our dataset and estimate a model using three-



Source: Authors' Calculations

Figure 2: 3-Year Changes in Air Connectivity and Number of Non-stop Routes among CBSAs

Note: The figure displays the correlation between three-year changes in log air connectivity and in the log number of non-stop aviation routes over the period 1984-2013 among the 262 CBSAs in our sample.

year differences in variables. In particular, the regression model that we propose takes the following form:

$$\Delta_{\{t+3,t\}}\mathrm{lnY}_i = \beta\,\Delta_{\{t+3,t\}}\mathrm{lnAir}_i + X_{i,t}{}'\gamma + u_i + u_t + \epsilon_{i,t} \tag{5}$$

where i indexes a metropolitan area, t indexes a sample year, and $\Delta_{\{t+3,t\}}\mathrm{lnY}_i \equiv \mathrm{lnY}_{i,t+3} - \mathrm{lnY}_{i,t}$ denotes a 3-year log-change in the variables of interest over the period [t, t+3]. The variable of interest $\Delta_{\{t+3,t\}}\mathrm{lnAir}_i$ is lagged by one year, following the approach in Gibbons and Wu (2020), but we suppress the one-year lag in the subscript notation for simplicity.[10] The dependent variable Y denotes an economic outcome of interest such as urban population, total employment, average per-capita income, or the number of business establishments that are active in a given city, respectively. u_i denotes the metropolitan area fixed effects, while u_t denotes the three-year period fixed effects. Any unobservable factor that is location specific and that affects the growth of a metropolitan area in the same way each period (e.g., secular growth trends) will be captures by u_i. Similarly, any unobservable macroeconomic shock that affects the growth of all metropolitan areas to the same extent will be directly accounted for by u_t.

[10]Our results remain qualitatively the same when using contemporaneous changes in air connectivity. The results are available upon request.

The vector X consists of an extensive set of control variables observed in the initial period t of the three-year time interval [t, t+3]. The goal is to control for pre-existing differences in the rate of economic growth across metropolitan areas, as well as to capture any observable time-varying determinants of urban growth. Towards that goal, our estimation equation (5) will control for: *i*) base-year levels of the dependent variable (e.g., initial year population, employment, number of establishments); *ii*) base-year levels of air connectivity as well as airport size (captured by the total number of passengers); *iii*) pre-existing growth trends as captured by historical population levels (e.g. 15-year population lag, as well as population lags going from 3- to 8-decades back); *iv*) socio-economic characteristics of the local labor market (e.g., employment rate, average wages, manufacturing employment share, female population share, non-white population share, fraction of population with a high school degree or less), and finally *v*) economic geography forces as captured by measures of market access. Most of these initial-year control variables will not only influence urban growth as measured by the dependent variable in equation (5), but also air connectivity potentially. So, by directly controlling for initial period conditions, we hope to overcome the omitted variable bias problem.

The decision to use differences in variables is also motivated by the need to mitigate endogeneity concerns while still retaining useful data variation. Any unobservable city characteristic that is time invariant will be eliminated through differencing. This ensures the removal of any factors that may influence both urban growth and air connectivity (e.g., geographic location, area, topography, etc.), once again mitigating the omitted variables bias problem. Of course, one drawback of data differentiation is that it can remove too much information, especially in the case of persistent economic outcome variables that evolve slowly over time such as, for example, population or employment. In choosing three-year (rather than annual) differences in variables, our intention is to overcome this concern and be able to retain useful time variation in our variables of interest.[11]

The time length over which the differences in variables are constructed necessitates further discussion. The literature on the effects of transportation on urban growth is quite heterogeneous on this dimension, with examples of studies that use year-on-year changes in variables (Sheard, 2019, 2021), or mid-range periods such as four-year intervals (Gibbons and Wu, 2020), or decades-long differences in variables (Baum-Snow, 2007; Duranton and Turner, 2012). In making our decision to use three-year intervals, we factored in several opposing motivations. A larger time window would allow us to fully capture the impact of

[11]One additional benefit of using longer differences in variables 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.

air connectivity on urban growth, particularly if it takes time for new aviation routes to form and have a lasting effect on urban growth. At the same time, a larger time window introduces the risk of other events that happen during that time window to impact urban growth, raising concerns about endogeneity and omitted variable bias. A narrow time window, on the other hand, closes down on the possibilities for other factors to impact urban growth. The controls for initial conditions (i.e., base year variables) should also have more explanatory power and more precision in predicting expected trajectories of urban growth over shorter time spans. Importantly for us, a narrower time window allows us to observe a metropolitan area multiple times during our sample period, which makes it possible for us to employ panel methods and estimate equation (5) using location fixed effects in addition to differencing the variables of interest. Ultimately, in trying to accommodate all these considerations, we chose a mid-range period of three years as the preferred time interval over which to examine how changes in air connectivity influence urban development. We further provide some motivating evidence that a three-year time window captures most of the effect of aviation routes on urban growth. In addition, in robustness exercises, we experiment with alternative time intervals ranging from one year to six year periods.

The sizeable time dimension of our dataset allows us to estimate equation (5) using non-overlapping three-year differences across observations (e.g., 1985 -1988, 1989-1992, 1993-1996, etc.).[12] This implies a maximum number of seven (three-year period) observations per metropolitan area during the sample period 1984-2013. To correct for autocorrelation, we cluster the standard errors at city level.

Our goal is to estimate the causal effect of air connectivity on regional growth. Thus, the coefficient of interest in equation (5) is β. The main challenge in obtaining an accurate estimate of β is the endogeneity problem involving a city's air connectivity. Even after differencing the data, adding control variables and adding fixed effects as ways to mitigate the omitted variable bias, endogeneity may still be an issue due to reverse causality running from urban growth to air connectivity. To address this concern, we rely on instrumental variable methods.

Solving the Endogeneity Problem: Instrumental Variables

The connectivity of a city's airport to the national 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 number of non-stop destinations N_{it} in equation (4), as well as the number of departures

[12] We opted to avoid an overlap between the start and end years of each three-year period to minimize the issue of correlated errors, and any other similar concerns coming from the mis-measurement of variables.

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 that are very similar in construction. They follow closely the functional form of the endogenous air connectivity measure proposed in equation (4). The excluded instruments are defined as follows:

$$AirConnect\_IV1_{it} = \sum_{d \neq i} \left(f_{dt}\right)\left(\frac{1}{Dist_{id}}\right)^{\frac{\sigma-1}{\sigma}} \tag{6}$$

where d denotes *all* the destination cities available in our sample at time t excluding the origin city i, f_{dt} stands for the total number of departures operated out of destination city d at time t, and $Dist$ denotes the geographic distance between origin i and destination d.

The second instrument matches the one just defined but uses the total number of non-stop aviation routes operated out of destination city d as a weight for the importance of destination d within the national aviation network:

$$AirConnect\_IV2_{it} = \sum_{d \neq i} \left(routes_{dt}\right)\left(\frac{1}{Dist_{id}}\right)^{\frac{\sigma-1}{\sigma}} \tag{7}$$

For each instrument we construct three-year log differences over the period $[t, t+3]$, lagged by one year to match the construction of air connectivity, before employing them as excluded instruments for $\Delta_{\{t+3,t\}} lnAir_i$ in equation (5).

To ensure exogeneity, we exclude from the construction of both instruments any information involving origin city i (other than distance information). Specifically, we exclude the number of flights from destination d to origin city i when calculating the variable f_{dt}, respectively we exclude the air link from d to i when calculating the variable $routes_{dt}$ if such a direct route existed at time t. Also, for exogeneity reasons, in all our calculations we consider *all* destination cities in our sample to ensure that the set of *potential* air links to which origin city i can connect to is not influenced by economic factors specific to origin city i. Because the potential set of air links overlaps with the *actual* set of non-stop destinations served from origin city i at each point in time t, this ensures the necessary correlation between the excluded instrument and the endogenous air connectivity variable.

The proposed instruments from equations (6) and (7) have some similarities with aviation instruments used in other recent studies. As in Gibbons and Wu (2020), our proposed instruments represent weighted centrality indexes whose goal is to capture the "closeness" of an origin city to the national aviation network (measured either by the number of flights or by the number of air links provided out of each destination). Furthermore, as in Sheard

(2019, 2021), our proposed instruments rely exclusively on time variation coming from the continuous evolution and development of the *national* aviation network over time (as captured through changes happening at destination "network nodes"). Similar to the motivation behind shift-share instruments (Bartik, 1991), the exogeneity of the instruments relies on the argument that activity in origin city i does not influence national level trends in aviation growth and network development. In other words, the development of air traffic at destination cities d is assumed to be independent of the growth trajectory of origin city i. As it will become clear shortly, some of the control variables included in the estimating equation (5) – e.g., the market access variables – are motivated by the need to ensure that this assumption holds in the data.

Note that the first bracket term in equation (6), respectively (7), is the only component that drives the instruments' variation over time. The second bracket term – i.e., the inverse of the geographic distance between the origin city i and destination d raised to the power $\frac{\sigma-1}{\sigma}$ – is an exogenous time-invariant determinant of the number of departures offered between cities i and d (thus, correlated with f_{idt} in equation (4)). 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 from its orthogonality to the economic development of origin city i.

Accounting for Market Access

A possible concern with the instrumental variables defined by equations (6) and (7) is that proximity to large metropolitan areas benefits a given origin 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 measured impact of the air connectivity of a metropolitan area 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:

$$MktAccess_{it} = \sum_{d \neq i} \frac{Pop_{dt} \times Income_{dt}}{Dist_{id}} \tag{8}$$

This measure is also known in the economic geography literature as the market potential of a location.[13] It essentially captures the access that a given city i has to markets nationwide,

[13]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.

where the importance of each market is measured by the GDP of that location (proxied by the product between population size and average per-capita income), 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.[14]

It is important to emphasize that, similar to the air connectivity instruments from equations (6) and (7), the market access variable sums up information about *all* locations d in our sample. By directly controlling for market access in our empirical analysis, we eliminate the possibility that our proposed instruments of air connectivity are simply picking up economic geography linkages. The goal for our proposed instruments is to capture a city's "closeness" to airport destinations that, conditional on their economic size, happen to be airport hubs responsible for a large number of flight departures and nonstop routes. Of course, the location of airport hubs is endogenous to the economic conditions in those selected locations, but once a hub is established in a particular region of the country, it provides a substantial exogenous change in the air connectivity index of other cities located not too far from the hub. This rapid change in air connectivity for collateral cities is probably not matched by a simultaneous change in their market access, since it may take some time for hub cities to accumulate large-enough economic benefits from becoming hubs, and for these benefits to dissipate spatially to influence market access.[15]

The market access expression in (8) uses the functional form that is most commonly used in the literature, where the distance decay parameter takes the value 1. In robustness exercises, we also experiment with distance decay parameters that match our excluded instruments, i.e., $\frac{\sigma-1}{\sigma}$. We also consider polynomial functions of market access (i.e., squared term) as well as three-year differences in market access over the period [t, t+3] (lagged by 1 year to match the format of the air connectivity variable of interest). The goal is to control in the most flexible way possible for unobserved economic geography linkages between locations.

Lastly, to further remove concerns about the possibility that the impact of air connectivity on urban growth is simply capturing market access effects, we also consider alternative instrumental variables where we exclude from equations (6) and (7) destination cities located within a 250 kilometer radius from origin city i. To the extent that market access or agglomeration spillovers are most impactful over short distances, then by excluding such linkages from our proposed instruments of air connectivity, we further minimize the possibility of un-

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

[15]Consistent with this claim, the correlation coefficient between three-year changes in air connectivity and 3-year changes in market access is zero across the metropolitan areas in our sample.

intended correlations. It is worth noting that a potential drawback of excluding destination cities from equations (6) and (7) could be weaker predictive power for our excluded instruments if cities in our data sample establish most of their direct aviation routes to nearby airports. However, this is not what the aviation data seems to suggest.[16]

4 Data Sources

To estimate our regression model, we need to combine two main sources of information: data on airline flight schedules, 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).[17] 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 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 metropolitan area to obtain annual data on air services at CBSA level.

Our estimation sample covers the time period 1984–2013.[18] We restrict the aviation data to origin-destination routes with positive passenger traffic, with a minimum of 52 departures scheduled per year (e.g., one scheduled flight per week), and at least 26 departures performed per year (i.e., minimum bi-weekly route operations).[19] We add data on the geographical coordinates of all airports in the U.S. from the Bureau of Transportation Statistics (BTS) and construct geodesic distances between airports offering non-stop service.

[16]In our sample period 1984-2013, the average flight distance for non-stop air service between two domestic airports is 1043 kilometers, and the median distance is 797 kilometers.

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

[18]In choosing the sample starting year, we wanted to avoid the aftermath of the 1978 Airline Deregulation Act. 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. The end-point of the sample period was determined by data availability at the moment of collection.

[19]Our main estimation results are not influenced by the removal of infrequent aviation routes. Estimation results using fewer conditions on the original air traffic data are available upon request.

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 employment and establishment data). 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.[20]

We employ multiple population lag variables to account in the most flexible way for pre-existing growth trends. County level population data going back to 1900s are available from the U.S. Census at decennial frequency. The BEA started collecting population data at annual frequency only in 1969. To take advantage of population data at annual frequency, we construct a 15-year lag population variable to use as control. For all other lagged population variables, we use decennial data lagged for a specific number of decades relative to the decade in question. So, for example, for the sample year 1987, a 40-year population lag corresponds to population levels in year 1940 since we cannot observe population levels in 1947.

We construct socio-economic variables at the level of a metropolitan area by aggregating county-level information on working age population (15-65 year olds), female population, non-white population, and fraction of population with a high school degree or less. County-level data on population by age, race and gender from 1969 onwards come from the NBER, which collects this information from the Survey of Epidemiology and End Results (SEER). County-level data on educational attainment comes from the US Department of Agriculture, the Economic Research Service, which sources the data from the U.S. Census. For our analysis, we construct the following demographic variables: working age population (15-65 year olds), employment share, manufacturing employment share, female population share, non-white population share, fraction of population with a high school degree or less.

Main variable construction. After combining the aviation and urban data, we obtain an unbalanced panel of 293 unique CBSAs grouped into 262 consolidated metropolitan

[20]We use a mapping of counties into CBSAs available from the U.S. Census through their U.S Gazetteer Place data from 2006. We apply this mapping throughout the sample period to ensure consistency of statistical areas over time.

areas, which are observed over the period 1984–2013.[21] Figure 3 provides a map of the CBSAs included in our estimation sample.

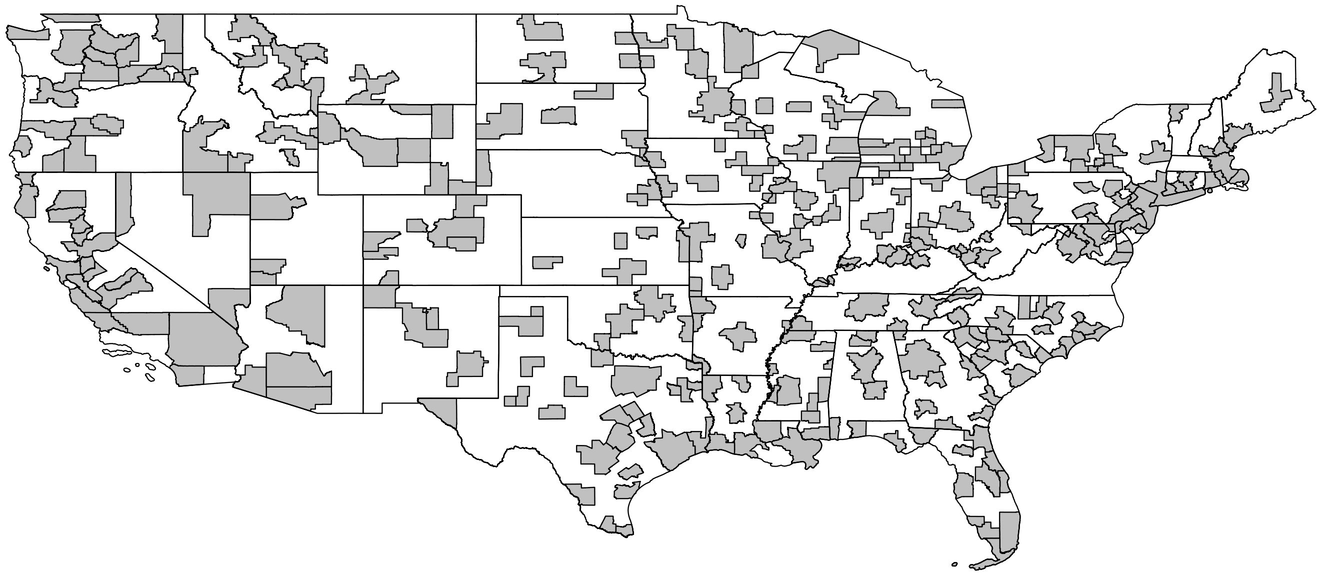


Figure 3: Map of the CBSAs in our sample.

Note: The shaded areas represent the 293 unique CBSAs which are combined into 262 consolidated CBSAs for our econometric analysis. They account for 99.3% of all air passenger travel, and 98.8% of all operated flight departures.

We then employ the resulting city-level panel dataset to construct our variables of interest. We follow equation (4) to construct our air connectivity measure. For that, we have to take a stand on the size of the elasticity of substitution σ. Unfortunately, we are not aware of any study that estimates how substitutable flight destinations are in the eyes of consumers. Instead, we rely on the international trade literature and set σ equal to 5, which is the corresponding modal value for tradeable goods (Anderson and van Wincoop, 2004; Simonovska and Waugh, 2014). We also experiment with lower ($\sigma = 2$) and higher ($\sigma = 10$) values in our robustness exercises. In choosing these parameter values, we have been encouraged by our own calculations of σ. Using our airport-level dataset on flight departures and a gravity equation framework, we infer possible values of σ from the estimated distance elasticity of flight frequency. Our imputations suggest a range for σ between 2.2 and 7.9, which further supports our preferred choice of $\sigma = 5$. Appendix A provides the details of our calculations.

[21] Some CBSAs have been combined into a consolidated metropolitan area due to proximity to the same airport. For example, CBSA 24860 (Greenville-Mauldin-Easley, SC) and CBSA 43900 (Spartanburg, SC) are both serviced by the Greenfield-Spartanburg International Airport (GSP), which is located within 21 miles of either business district. So, the two CBSAs have been combined into one consolidated area before being matched with the GSP airport. In total, there are 59 CBSAs that have been grouped into 29 conglomerated locations because of their joint proximity to a particular airport.

For the remaining regression variables, we use the geographic 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 (8) to construct the market access of a city in our sample. We also use the bilateral distance to implement equations (6) and (7), and obtain the exogenous instruments for air connectivity.

Table 1: Summary Statistics

	Obs	Mean	St. Dev.	Min	Max
Aviation Network Variables					
$\Delta_{\{t+3,t\}}$ Ln Air Connectivity ($\sigma = 5$)	1,482	0.122	0.704	-8.172	6.577
$\Delta_{\{t+3,t\}}$ Ln Air Connectivity ($\sigma = 2$)	1,482	0.105	0.640	-8.262	6.224
$\Delta_{\{t+3,t\}}$ Ln Air Connectivity ($\sigma = 10$)	1,482	0.127	0.731	-8.141	6.699
$\Delta_{\{t+3,t\}}$ Ln Air Connect IV1 ($\sigma = 5$)	1,482	0.065	0.109	-0.135	0.354
$\Delta_{\{t+3,t\}}$ Ln Air Connect IV1 ($\sigma = 2$)	1,482	0.066	0.106	-0.111	0.283
$\Delta_{\{t+3,t\}}$ Ln Air Connect IV1 ($\sigma = 10$)	1,482	0.065	0.110	-0.147	0.390
$\Delta_{\{t+3,t\}}$ Ln Air Connect IV2 ($\sigma = 5$)	1,482	0.048	0.101	-0.181	0.255
$\Delta_{\{t+3,t\}}$ Ln Air Connect IV2 ($\sigma = 2$)	1,482	0.049	0.100	-0.144	0.213
$\Delta_{\{t+3,t\}}$ Ln Air Connect IV2 ($\sigma = 10$)	1,482	0.048	0.103	-0.197	0.275
Δ Urban Economic Indicators					
$\Delta_{\{t+3,t\}}$ Ln Population	1,482	0.032	0.033	-0.203	0.215
$\Delta_{\{t+3,t\}}$ Ln Total Employment	1,482	0.046	0.062	-0.234	0.461
$\Delta_{\{t+3,t\}}$ Ln Personal Income	1,482	0.045	0.046	-0.102	0.379
$\Delta_{\{t+3,t\}}$ Ln Number Establishments	1,482	0.029	0.045	-0.150	0.365
$\Delta_{\{t+3,t\}}$ Ln Number of Small Establishments	1,482	0.026	0.045	-0.143	0.355
$\Delta_{\{t+3,t\}}$ Ln Number of Medium Establishments	1,482	0.052	0.068	-0.277	0.416
$\Delta_{\{t+3,t\}}$ Ln Number of Large Establishments	1,451	0.036	0.238	-1.099	1.386
$\Delta_{\{t+3,t\}}$ Ln Market Access	1,482	0.064	0.043	-0.005	0.159
Base Year Urban Economic Indicators					
Ln Population	1,482	12.832	1.263	9.512	16.749
Ln Total Passengers	1,482	12.624	2.058	5.911	17.479
Ln Avg. Wage	1,482	9.145	0.784	6.546	10.826
Ln Total Employment	1,482	11.833	1.334	8.327	15.856
Ln Number Establishments	1,482	9.163	1.214	6.140	13.193
Ln Air Connectivity ($\sigma = 5$)	1,482	3.742	1.719	-5.835	7.228
Ln Population Lag (15 years)	1,482	12.654	1.255	8.887	16.672
Ln Market Access	1,482	21.928	0.386	20.874	23.084
Ln Market Access Squared	1,482	480.984	16.948	435.741	532.875
Employment Share	1,482	0.567	0.116	0.179	1.255
Manufacturing Share of Employment	1,482	0.145	0.076	0.005	0.435
Fraction Female Population	1,482	0.502	0.014	0.364	0.536
Fraction Non-White Population	1,482	0.139	0.109	0.007	0.726
Fraction High School Educ or Less	1,482	0.481	0.090	0.221	0.750

Notes: The estimation sample covers 262 consolidated CBSAs observed every 4 years over the period 1984–2013. The data sources and the construction of variables are described in the paper. $\Delta_{\{t+3,t\}}$ denotes 3-year differences in variables.

Finally, once we have all our data ready, we construct three-year changes in the variables of interest. We keep only non-overlapping time periods in our estimation sample to minimize any issues arising from correlated residuals and from any measurement error in variables. Thus, the estimation sample includes observations for 262 consolidated CBSAs observed over seven non-overlaping time intervals, at most.[22] Table 1 reports summary statistics for all the variables used in the econometric analysis, expressed in three-year differences.

5 Estimation Results

Our goal is to estimate the causal effect of air connectivity on urban development. We estimate equation (5) using as dependent variable 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 in variables combined with location and period fixed effects, the identification of the model parameters comes from comparing deviation from average trends across the 262 sample CBSAs.

Using three-year differences in the variables of interest helps eliminate any time invariant factors that affect city growth, many of which may be hard to measure. However, this still leaves open the possibility of omitted variable bias coming from city-specific secular trends. This is where adding location-specific fixed effects plays a huge role. We also overcome any remaining concerns about differences in growth trends across cities over time by including a long list of control variables ranging from initial economic conditions, socio-demographic indicators, a long series of population lags, as well as a flexible way to capture market access forces. Adding all these control variables to the regression model in equation (5) is intended to mitigate the endogeneity problem between air connectivity and urban development. To further ensure a causal interpretation of our estimates, we also correct for any remaining endogeneity concerns using the standard instrumental variables (IV) approach.

The first set of estimates are reported in Table 2 and focus on city population size as the economic outcome of interest. For now, all the estimations are done by ordinary least squares method (OLS). They differ in the set of control variables being considered, going from the most basic specification to the preferred model that includes all the controls and fixed effects. The main idea behind reporting sequential model specifications is to showcase the extent of data variation available in our sample, and the sensitivity of our coefficient of interest to adding increasingly more controls. Column 1 reports the OLS coefficients

[22] The three-year non-overlaping time intervals consist of 1985-1988, 1989-1992, 1993-1996, 1997-2000, 2001-2004, 2005-2008, and 2009-2012. Recall that because $\Delta_{\{t+3,t\}} \ln \text{Air}_i$ in equation (5) is lagged by one-year, we had to start with 1985-1988 as our first period.

Table 2: OLS Estimates for the Effect of Air Connectivity on Urban Population

	(1) Basic	(2) +Pop	(3) +Demogr	(4) +MkAccess	(5) +CityFE
$\Delta_{\{t+3,t\}}$ Ln Air Connectivity	0.004***	0.003**	0.003**	0.003**	0.002**
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
Ln Population$_t$	-0.015	0.119***	0.104***	0.107***	-0.163***
	[0.012]	[0.010]	[0.013]	[0.014]	[0.031]
Ln Passengers$_t$	0.001	0.001	0.000	0.001	-0.003*
	[0.002]	[0.001]	[0.001]	[0.001]	[0.002]
Ln Air Connectivity$_t$	0.005**	0.002*	0.002**	0.001	0.001
	[0.002]	[0.001]	[0.001]	[0.001]	[0.001]
Initial values	yes	yes	yes	yes	yes
Population lags	no	yes	yes	yes	yes
Demographic characteristics	no	no	yes	yes	yes
Market access	no	no	no	yes	yes
Consolid. CBSA fixed effects	no	no	no	no	yes
Period fixed effects	yes	yes	yes	yes	yes
Observations	1,482	1,482	1,482	1,482	1,482
R-squared	0.294	0.565	0.569	0.595	0.325

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

Notes: The reported results correspond to the regression model in equation (5) estimated by ordinary least squares (OLS). The sample is an unbalanced panel of 262 consolidated CBSAs observed every 4 years over the period 1984-2013. The dependent variable is the 3-year change in log population. The variable of interest is the 3-year change in log air connectivity (lagged by one year), defined in equation (4) with $\sigma = 5$. The control variables included in each column are summarized in the table and described in the main text. Robust standard errors clustered at CBSA level are reported in brackets.

when controlling only for initial conditions, as captured by the base-year level of population, employment, number of establishments, average wages, airport size, and air connectivity. In addition, we also include time fixed effects. Note that because of space considerations, column 1 only reports a subset of the coefficients that capture initial conditions, but all the variables listed above are considered. Column 2 builds on the specification in column 1 and adds a long series of population lags (i.e., a 15-year lag, and decadal population lags going back from 3 to 8 decades). Column 3 subsequently adds socio-demographic information in base year t such as the city employment rate, the manufacturing share of employment, and the population shares for female, non-white and low education individuals. Column 4 further adds three market access variables to capture in the most flexible way economic geography forces. The three variables are the level of market access in the base year t, its squared term to account for non-linearities, and the three-year change in market access over [t, t+3] (lagged by 1 year to match the air connectivity measure). Lastly, column 5 adds city fixed effects to the specification in column 4. Because of completeness, this represents our preferred model specification.

The coefficients for air connectivity reported in Table 2 are quite stable across specifications. They are all positive and statistically significant, and even though the preferred coefficient in Column 5 is about half the size of the one reported for the most basic specification in column 1, no particular group of controls or fixed effects seems to break the positive impact of air connectivity on population growth. Of course, it remains to be seen whether this is true once accounting for any residual endogeneity between the two variables.

Table 3: IV Estimates for the Effect of Air Connectivity on Urban Population

	(1) Basic	(2) +Pop	(3) +Demogr	(4) +MkAccess	(5) +CityFE	(6) RD250
$\Delta_{\{t+3,t\}}$ Ln Air Connectivity	0.020***	0.018***	0.019***	0.010***	0.009***	0.016***
	[0.004]	[0.004]	[0.004]	[0.003]	[0.003]	[0.005]
Ln Population$_t$	-0.016	0.113***	0.094***	0.101***	-0.158***	-0.153***
	[0.012]	[0.011]	[0.015]	[0.014]	[0.031]	[0.031]
Ln Passengers$_t$	0.001	0.000	-0.000	0.001	-0.002	-0.002
	[0.002]	[0.001]	[0.001]	[0.001]	[0.002]	[0.002]
Ln Air Connectivity$_t$	0.008***	0.005***	0.005***	0.003**	0.004**	0.006***
	[0.002]	[0.002]	[0.002]	[0.001]	[0.002]	[0.002]
Initial values	yes	yes	yes	yes	yes	yes
Population lags	no	yes	yes	yes	yes	yes
Demographic characteristics	no	no	yes	yes	yes	yes
Market access	no	no	no	yes	yes	yes
Consolid. CBSA fixed effects	no	no	no	no	yes	yes
Period fixed effects	yes	yes	yes	yes	yes	yes
Observations	1,482	1,482	1,482	1,482	1,482	1,482
R-sq	0.190	0.474	0.473	0.576	0.281	0.156
First Stage Statistics						
F-stat	28.63	27.56	27.40	24.32	20.32	11.75
Hansen J stat	5.121	0.517	0.728	2.310	0.390	0.441
Hansen J p-val	0.024	0.472	0.394	0.129	0.532	0.507

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

Notes: The reported results correspond to the regression model in equation (5) estimated by limited information maximum likelihood (LIML). The sample is an unbalanced panel of 262 consolidated CBSAs observed every 4 years over 1984-2013. The dependent variable is the 3-year change in log population. The variable of interest is the 3-year change in log air connectivity (lagged by one year), defined in equation (4) with $\sigma = 5$. The control variables included in each column are summarized in the table and described in the main text. The excluded instruments for changes in air connectivity are defined in equations (6) and (7), and are transformed in 3-year changes before use. Robust standard errors clustered at CBSA level in brackets.

Table 3 replicates the five specifications in Table 2 but using the limited information maximum likelihood (LIML) method and the two excluded instruments defined in equations (6) and (7), transformed in three-year differences.[23] The last column of Table 3 re-estimates

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

our preferred specification in column 5 but using two alternative instrumental variables that exclude from their construction destination cities within a 250 kilometer radius of the origin city. As discussed in the econometric strategy section, our main motivation for excluding nearby metropolitan areas from the construction of the instruments was to weaken the correlation between market access (driven in large part by nearby urban agglomerations), and our proposed air connectivity instruments.

Across the board, the IV estimates in Table 3 are positive and statistically significant but substantially larger in magnitude compared to their OLS counterparts. This may seem surprising at first given the expected positive connection between unobservable factors driving population growth and positive changes in air connectivity. However, the opposite direction of bias has been encountered more often than not in the transportation and urban growth literature (e.g., Duranton and Turner, 2012; Sheard, 2014; Blonigen and Cristea, 2015; Sheard, 2021, among others). It suggests that cities with slower population growth are the ones to witness larger increases in air connectivity, on average. This would be consistent with anecdotal evidence about city officials being proactive about attracting airlines to their cities in hopes of revitalizing and spurring economic growth. It would also be consistent with the scenario of large and fast-growing cities witnessing a relatively slower growth in non-stop air services over our sample period, in part because many of the major non-stop connections had already been established prior to the start of the sample period.[24]

Importantly, the IV estimates are able to correct any endogeneity biases (whether positive or negative) as long as the proposed instrumental variables are valid. This seems to be the case judging from their statistical performance. The F-statistics for the excluded instruments reported at the bottom of Table 3 exceed conventional critical levels in almost all of the specifications. This is suggestive of a significant correlation between the air connectivity measure and the excluded instruments. Furthermore, the test for overidentifying restrictions also reported at the bottom of Table 3 generally fails to reject the null hypothesis that the excluded instruments are orthogonal to the regression residual.

Focusing on the magnitude of the coefficient of interest from our preferred specification, the estimate in column 5 suggests that a 10 percent increase in air connectivity over a three-year period causes a 0.09 percent change in population size over the same period. The average increase in air connectivity observed across all time periods and all consolidated CBSAs in our sample is 12.2 percent, implying a 0.11 percent change in population size. The effect is relatively small in size, explaining about 3.4 percent of the average growth in

[24]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 OLS estimate.

population size observed during our sample period among consolidated CBSAs. However, the small but statistically significant effect is consistent with other findings in the literature on the impact of air passenger transport on urban population (e.g., Sheard, 2019).

Table 4: First Stage Regressions

	(1) Basic	(2) +Pop	(3) +Demogr	(4) +MkAccess	(5) +CityFE	(6) RD250
$\Delta_{\{t+3,t\}}$ Ln AirConnect IV1	4.021***	4.162***	4.155***	4.575***	4.127***	
	[0.852]	[0.864]	[0.857]	[0.933]	[0.883]	
$\Delta_{\{t+3,t\}}$ Ln AirConnect IV2	2.746***	2.584**	2.469**	2.238**	2.312**	
	[1.039]	[1.023]	[1.025]	[1.038]	[1.053]	
$\Delta_{\{t+3,t\}}$ Ln AirConnect IV1 (r>250)						-0.629
						[1.574]
$\Delta_{\{t+3,t\}}$ Ln AirConnect IV2 (r>250)						6.810***
						[2.225]
Ln Population$_t$	0.081	0.319	0.589*	0.653*	-0.990	-1.026
	[0.108]	[0.226]	[0.318]	[0.336]	[0.671]	[0.683]
Ln Employment$_t$	-0.115	-0.045	-0.187	-0.251	-0.267	-0.271
	[0.139]	[0.131]	[0.331]	[0.332]	[0.530]	[0.542]
Ln Establishments$_t$	0.166	0.146	0.088	0.112	1.425**	1.459**
	[0.122]	[0.131]	[0.161]	[0.157]	[0.579]	[0.578]
Ln Passengers$_t$	0.065*	0.051	0.039	0.050	-0.045	-0.069
	[0.035]	[0.035]	[0.040]	[0.039]	[0.085]	[0.089]
Ln Air Connectivity$_t$	-0.225***	-0.223***	-0.223***	-0.233***	-0.340***	-0.344***
	[0.052]	[0.052]	[0.051]	[0.052]	[0.073]	[0.076]
Initial values	yes	yes	yes	yes	yes	yes
Population lags	no	yes	yes	yes	yes	yes
Demographic characteristics	no	no	yes	yes	yes	yes
Market access	no	no	no	yes	yes	yes
Consolid. CBSA fixed effects	no	no	no	no	yes	yes
Period fixed effects	yes	yes	yes	yes	yes	yes
Observations	1,482	1,482	1,482	1,482	1,482	1,482

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

Notes: The reported results correspond to the first stage coefficients obtained from estimating the regression model in equation (5) by limited information maximum likelihood (LIML). All the specifications use the same two excluded instruments described by equations (6) and (7) using $\sigma = 5$, transformed in 3-year log differences. The dependent variable represents the 3-year change in log air connectivity, defined in equation (4) with $\sigma = 5$. The control variables included in each column are summarized in the table and described in the main text. Robust standard errors clustered at CBSA level in brackets.

To gain confidence in the LIML estimations and in the causal interpretation of our estimated coefficients, we report in Table 4 the results from the first stage regressions. Columns 1 to 6 report the first stage estimates corresponding to the respective LIML specifications in Table 3. All the specifications employ the two excluded instruments defined in equations (6) and (7), transformed into three-year log differences. The excluded instruments are highly correlated as they are constructed in a very similar way (0.80 correlation coefficient), however using them jointly has advantages as they bring more information to be used in predicting our endogenous air connectivity variable. Focusing on their behavior in the first stage regressions, we note that both variables have the expected sign and are highly significant as determinants of air connectivity. Multicollinearity becomes most visible in the last column of Table 4, where it seems that by removing air links to the nearby airports (on a radius less than 250 kilometers), we are increasing even more the correlation between the excluded instruments (0.97 correlation coefficient). While multicollinearity affects the individual coefficients of the excluded instruments, their joint significance is not really impacted. So, we continue to report results using these alternative instruments (column 6) because of the additional insights that we may get from removing any potential agglomeration spillovers.

Table 5: Effect of Air Connectivity on Urban Employment (IV)

	(1) **Basic**	**(2)** **+Pop**	**(3)** **+Demogr**	**(4)** **+MkAccess**	**(5)** **+CityFE**	**(6)** **RD250**
$\Delta_{\{t+3,t\}}$ Ln Air Connectivity	0.055***	0.054***	0.055***	0.029***	0.032***	0.045***
	[0.012]	[0.012]	[0.012]	[0.009]	[0.009]	[0.013]
Ln Population$_t$	0.015	0.094***	0.013	0.038	-0.009	0.001
	[0.015]	[0.023]	[0.035]	[0.030]	[0.059]	[0.064]
Ln Employment$_t$	-0.037***	-0.038***	0.039	0.030	-0.327***	-0.321***
	[0.014]	[0.014]	[0.035]	[0.029]	[0.054]	[0.058]
Ln Passengers$_t$	0.002	0.003	0.002	0.004	-0.008*	-0.006
	[0.003]	[0.003]	[0.003]	[0.003]	[0.004]	[0.006]
Ln Air Connectivity$_t$	0.013***	0.010**	0.012**	0.005	0.010**	0.014**
	[0.004]	[0.004]	[0.005]	[0.003]	[0.005]	[0.006]
Initial values	yes	yes	yes	yes	yes	yes
Population lags	no	yes	yes	yes	yes	yes
Demographic characteristics	no	no	yes	yes	yes	yes
Market access	no	no	no	yes	yes	yes
Consolid. CBSA fixed effects	no	no	no	no	yes	yes
Period fixed effects	yes	yes	yes	yes	yes	yes
Observations	1,482	1,482	1,482	1,482	1,482	1,482
R-sq	-0.007	0.047	0.045	0.332	0.431	0.324
First Stage Statistics						
F-stat	28.63	27.56	27.40	24.32	20.32	11.75
Hansen J stat	0.795	0.298	0.489	1.621	5.907	1.174
Hansen J p-val	0.373	0.585	0.484	0.203	0.015	0.278

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

Notes: The reported results correspond to the regression model in equation (5) estimated by limited information maximum likelihood (LIML). The sample is an unbalanced panel of 262 consolidated CBSAs observed every 4 years over the period 1984-2013. The dependent variable is the 3-year change in log city employment. The variable of interest is the 3-year change in log air connectivity (lagged by one year), constructed based on equation (4) using $\sigma = 5$. The control variables included in each column are summarized in the table and described in the main text. The excluded instruments for changes in air connectivity are defined in equations (6) and (7), and are transformed in 3-year changes before use. Robust standard errors clustered at CBSA level in brackets.

We next examine the performance of another economic outcome of great policy interest – employment growth. Table 5 reports the IV results from estimating equation (5) by LIML using total employment as dependent variable.[25] The layout of the table follows the same sequence of model specifications as reported in Table 3, with column 1 reporting the most basic model specification and columns 5-6 reporting the preferred specification with all the control variables and fixed effects. Once again, across the board, we find that air connectivity has a positive and statistically significant impact on employment growth, with the magnitude of the effect dropping in size as we add more controls. Using the most comprehensive specification in column 5, we find that a 10-percent increase in air connectivity leads to a 0.32 percent growth in total employment. To put the results in perspective, note that the average three-year growth in employment among the consolidated CBSAs in our sample is 4.6 percent. Evaluated at the sample mean, air connectivity leads to 0.4 percent increase in employment, which explains 8.4 percent of the observed employment growth.

Does the growth in employment translate into an increase in personal 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 three-year changes in average per-capita income on a city's growth in air connectivity. We report the estimation results in Table 6. 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 improvement in air connectivity. Once all the control variables and fixed effects are considered, the magnitude of the coefficient of interest becomes very small and staistically indistinguishable from zero.

Finally, one last yet important economic outcome of interest is the number of business establishments that open up and operate profitably in a city. Anecdotal evidence suggests that a city's aviation network is essential for its ability to attract entrepreneurial talent, having a direct effect on the business environment of that location. Using data on the total number of active establishments in a consolidated CBSA, we use our regression model to examine this hypothesis. We report the IV estimation results obtained via LIML method in Table 7. Across all specifications the coefficient of interest has the expected positive sign and is statistically significant. Focusing on the estimate from the preferred model in column 5, which corresponds to the most comprehensive regression specification, we find that a 10 percent increase in air connectivity over a three-year period leads to a 0.24 percent increase in the number of business establishments in a city. The average three-year growth in the total number of establishments over our sample period is 2.9 percent. Evaluated

[25] In the interest of space, we omit the OLS estimates for the effect of air connectivity on total employment at consolidated CBSA level. The results are available upon request.

Table 6: Effect of Air Connectivity on Average Personal Income (IV)

	(1) Basic	(2) +Pop	(3) +Demogr	(4) +MkAccess	(5) +CityFE	(6) RD250
$\Delta_{\{t+3,t\}}$ Ln Air Connectivity	0.015**	0.016**	0.016**	-0.002	-0.004	-0.002
	[0.007]	[0.007]	[0.007]	[0.006]	[0.006]	[0.011]
Ln Population$_t$	-0.046***	-0.085***	-0.110***	-0.082***	0.037	0.039
	[0.015]	[0.017]	[0.022]	[0.021]	[0.045]	[0.045]
Ln Avg Wage$_t$	-0.004**	-0.001	-0.001	0.000	-0.084***	-0.086***
	[0.002]	[0.002]	[0.002]	[0.002]	[0.032]	[0.033]
Ln Passengers$_t$	0.001	0.002	0.001	0.004*	-0.003	-0.002
	[0.002]	[0.002]	[0.002]	[0.002]	[0.004]	[0.004]
Ln Air Connectivity$_t$	0.003	0.003	0.004	-0.003	-0.003	-0.002
	[0.002]	[0.003]	[0.003]	[0.002]	[0.003]	[0.004]
Initial values	yes	yes	yes	yes	yes	yes
Population lags	no	yes	yes	yes	yes	yes
Demographic characteristics	no	no	yes	yes	yes	yes
Market access	no	no	no	yes	yes	yes
Consolid. CBSA fixed effects	no	no	no	no	yes	yes
Period fixed effects	yes	yes	yes	yes	yes	yes
Observations	1,482	1,482	1,482	1,482	1,482	1,482
R-sq	0.178	0.185	0.194	0.318	0.396	0.399
First Stage Statistics						
F-stat	28.63	27.56	27.40	24.32	20.32	11.75
Hansen J stat	1.345	0.978	1.434	3.113	5.003	15.53
Hansen J p-val	0.246	0.323	0.231	0.078	0.025	0.000

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

Notes: The reported results correspond to the regression equation (5) estimated by limited information maximum likelihood (LIML). The sample is an unbalanced panel of 262 consolidated CBSAs observed every 4 years over the period 1984-2013. The dependent variable is the 3-year change in log personal income. The variable of interest is the 3-year change in log air connectivity (lagged by one year), defined in equation (4) with $\sigma = 5$. The control variables included in each column are summarized in the table and described in the main text. The excluded instruments for changes in air connectivity are defined in equations (6) and (7), and are transformed in 3-year changes before use. Robust standard errors clustered at CBSA level in brackets.

at the sample mean, the increase in air connectivity explains 10.1 percent of the observed growth in establishments. 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

Table 7: Effect of Air Connectivity on the Number of Local Businesses (IV)

	(1) Basic	(2) +Pop	(3) +Demogr	(4) +MkAccess	(5) +CityFE	(6) RD250
$\Delta_{\{t+3,t\}}$ Ln Air Connectivity	0.045***	0.045***	0.046***	0.025***	0.024***	0.029***
	[0.008]	[0.009]	[0.009]	[0.006]	[0.006]	[0.008]
Ln Population$_t$	-0.015	0.062***	0.016	0.038	0.028	0.032
	[0.013]	[0.019]	[0.027]	[0.023]	[0.042]	[0.042]
Ln Establishments$_t$	-0.010	-0.024**	-0.022*	-0.019*	-0.251***	-0.258***
	[0.013]	[0.010]	[0.012]	[0.011]	[0.037]	[0.039]
Ln Passengers$_t$	0.003	0.004*	0.003	0.005***	-0.002	-0.002
	[0.002]	[0.002]	[0.002]	[0.002]	[0.003]	[0.004]
Ln Air Connectivity$_t$	0.010***	0.007**	0.008**	0.002	0.008***	0.009***
	[0.003]	[0.003]	[0.003]	[0.002]	[0.003]	[0.003]
Initial values	yes	yes	yes	yes	yes	yes
Population lags	no	yes	yes	yes	yes	yes
Demographic characteristics	no	no	yes	yes	yes	yes
Market access	no	no	no	yes	yes	yes
Consolid. CBSA fixed effects	no	no	no	no	yes	yes
Period fixed effects	yes	yes	yes	yes	yes	yes
Observations	1,482	1,482	1,482	1,482	1,482	1,482
R-sq	-0.035	0.034	0.029	0.404	0.429	0.369
First Stage Statistics						
F-stat	28.63	27.56	27.40	24.32	20.32	11.75
Hansen J stat	0.390	0.000	0.001	0.199	1.370	1.655
Hansen J p-val	0.532	0.992	0.972	0.656	0.242	0.198

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

Notes: The reported results correspond to the regression equation (5) estimated by limited information maximum likelihood (LIML). The sample is an unbalanced panel of 262 consolidated CBSAs observed every 4 years over the period 1984-2013. The dependent variable is the 3-year change in log number of establishments. The variable of interest is the 3-year change in log air connectivity (lagged by one year), defined in equation (4) with $\sigma = 5$. The control variables included in each column are summarized in the table and described in the main text. The excluded instruments for changes in air connectivity are defined in equations (6) and (7), and are transformed in 3-year changes before use. Robust standard errors clustered at CBSA level in brackets.

8 reports the estimation results from our preferred LIML specification that includes all the control variables and fixed effects. Comparing the estimated coefficients from columns 1, 3 and 5, 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 appear 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 factors of production 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.

Table 8: **Effect of Air Connectivity on the Number of Local Businesses by Size (IV)**

	Small Estabs		**Medium Estabs**		**Large Estabs**	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	Main IV	**RD250**	**Main IV**	**RD250**	**Main IV**	**RD250**
$\Delta_{\{t+3,t\}}$ Ln Air Connectivity	0.020***	0.023***	0.049***	0.071***	-0.071	-0.031
	[0.005]	[0.007]	[0.012]	[0.022]	[0.043]	[0.055]
Ln Population$_t$	0.032	0.034	0.011	0.027	-0.233	-0.205
	[0.041]	[0.042]	[0.067]	[0.076]	[0.293]	[0.292]
Ln Establishments$_t$	-0.269***	-0.273***	-0.115*	-0.142*	0.838***	0.785***
	[0.037]	[0.038]	[0.062]	[0.073]	[0.239]	[0.239]
Ln Passengers$_t$	-0.001	-0.001	-0.011*	-0.009	-0.041	-0.036
	[0.003]	[0.004]	[0.006]	[0.008]	[0.026]	[0.026]
Ln Air Connectivity$_t$	0.006**	0.007**	0.020***	0.027***	-0.026	-0.015
	[0.003]	[0.003]	[0.006]	[0.009]	[0.020]	[0.022]
Initial values	yes	yes	yes	yes	yes	yes
Population lags	no	yes	yes	yes	yes	yes
Demographic characteristics	no	no	yes	yes	yes	yes
Market access	no	no	no	yes	yes	yes
Consolid. CBSA fixed effects	no	no	no	no	yes	yes
Period fixed effects	yes	yes	yes	yes	yes	yes
Observations	1,482	1,482	1,482	1,482	1,449	1,449
R-sq	0.462	0.434	0.297	0.084	0.096	0.125
First Stage Statistics						
F-stat	20.32	11.75	20.32	11.75	21.52	12.55
Hansen J stat	1.537	0.833	0.128	6.838	8.977	2.073
Hansen J p-val	0.215	0.361	0.720	0.009	0.003	0.150

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

Notes: The reported results correspond to the regression equation (5) estimated by limited information maximum likelihood (LIML) using the full set of control variables and fixed effects. The explanations included in the footnote of Table 7 apply. The dependent variable is the 3-year change in the log number of business establishments by size category. Small size establishments are defined as having less than 20 employees. Large size establishments employ 500 people and above. The remaining establishments are classified as medium size. Robust standard errors clustered at CBSA level in brackets.

6 Robustness Checks

In this section we describe three estimation exercises that we perform to test the robustness of our results as well as their sensitivity to various modeling choices.

First, we conduct a placebo exercise that regresses changes in air connectivity on *past* changes in urban growth. The idea here is to verify whether our model specification removes any pre-existing differences across locations in growth trends. The main identifying assumption in our econometric analysis is that by differencing the data, then removing average trends via fixed effects, and then by adding to the model time-varying control variables, we are able to account for location-specific factors that could influence growth trends in both

Table 9: Placebo Exercise

	(1) Population	(2) Employment	(3) Establishments
$\Delta_{\{t+3,t\}}$ Ln Air Connectivity	-0.015	0.007	-0.014
	[0.009]	[0.017]	[0.014]
Ln Population$_{t-4}$	-0.223***	-0.140*	-0.084
	[0.044]	[0.082]	[0.064]
Ln Employment$_{t-4}$	-0.004	-0.445***	-0.005
	[0.026]	[0.055]	[0.033]
Ln Establishments$_{t-4}$	0.084***	0.191***	-0.164***
	[0.029]	[0.067]	[0.055]
Ln Passengers$_t$	0.001	0.009*	0.004
	[0.002]	[0.005]	[0.003]
Ln Air Connectivity$_t$	-0.008	0.009	-0.006
	[0.007]	[0.012]	[0.010]
Initial values$_{t-4}$	yes	yes	yes
Population lags	yes	yes	yes
Demographic characteristics$_{t-4}$	yes	yes	yes
Market access$_{t-4}$	yes	yes	yes
Consolid. CBSA fixed effects	yes	yes	yes
Period fixed effects	yes	yes	yes
Observations	1,174	1,174	1,174
R-sq	0.201	0.520	0.480
First Stage Statistics			
F-stat	13.42	13.42	13.42
Hansen J stat	12.53	1.663	8.174
Hansen J p-val	0.000	0.197	0.004

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

Notes: The reported results correspond to equation (5) estimated by limited information maximum likelihood (LIML) using an unbalanced panel of 262 CBSAs observed every 4 years during 1984-2013. The dependent variable is the 3-year change in log population (col. 1), in log employment (col. 2), respectively in log number of establishments (col. 3), calculated over the period [t-4, t-1]. The variable of interest is the 3-year change in log air connectivity over the period [t, t+3]. All columns include the complete set of fixed effects and control variables, lagged to the base year $t-4$ except for the aviation controls. The instruments for air connectivity are defined in equations (6) and (7), and are transformed in 3-year changes over the period [t, t+3]. Robust standard errors clustered at CBSA level in brackets.

economic outcomes and in the aviation network. For any remaining model endogeneity, the instrumental variable method would be able to eliminate such concerns. To test whether our econometric strategy fully removes pre-existing trends, we regress *current* changes in air connectivity (over the period [t, t+3]) on *past* changes in urban growth measured over the interval [t-4, t-1], while conditioning on fixed effects and on previous period observable characteristics (i.e., $t-4$ initial conditions). The IV results obtained via LIML method are reported in Table 9. Column 1 uses population changes as dependent variable, column 2 uses employment changes as dependent variable, and column 3 uses changes in the total number of establishments as dependent variable. Across all three models, the coefficient for air connectivity is insignificant, confirming the absence of pre-existing location-specific growth trends.

The second robustness exercise that we conduct examines the sensitivity of our estimates to the choice of the elasticity of substitution, σ, which is used in the construction of our variable of interest and the two excluded instruments. The air connectivity index in equation (4) requires information on how substitutable aviation routes are from the perspective of the representative consumer. Absent a parameter estimate for σ specific to the airline industry, we have used in our analysis a value of $\sigma = 5$ borrowed from the international trade literature (Anderson and van Wincoop, 2004). This value also matches fairly well our own set of back-of-the-envelope calculations (see Appendix A). However, in reality there is a range of estimated values for σ, which generally spans the interval between 2 and 10 (Broda and Weinstein, 2006; Simonovska and Waugh, 2014). Low values for the elasticity of substitution indicate a stronger preference for distinct route "varieties", thus placing more importance on the route extensive margin. On the contrary, large values for the elasticity of substitution indicate significant substitution patterns between (non-stop) travel destinations in the eyes of consumers, implying smaller utility gains from adding new route "varieties".

Given the importance of the σ parameter for evaluating the gains from expanding the aviation network of a metropolitan area through the addition of new non-stop destinations, we experiment in Table 10 with air connectivity measures constructed using low ($\sigma = 2$) and high ($\sigma = 10$) values of the elasticity of substitution. Both the air connectivity variable of interest and the two excluded instruments are affected by the choice of σ value. In addition to these three variable, we also modify the market access variable defined in equation (8) to allow the distance decay parameter to depend on σ in the same way as the excluded instruments. Specifically, we add the exponent $(\sigma - 1)/\sigma$ to the distance term in equation (8). This would allow us to check if the effect of air connectivity on urban growth is affected in any way when we allow more distant locations to have the same influence on market access as they have on the excluded instruments. So, the coefficients reported in Table 10 using the default value of $\sigma = 5$ will not match prior estimates to the extent that the distance decay parameter used in constructing the market access variables plays a significant role.

Table 10: Sensitivity Analysis to Different Values for the Elasticity of Substitution σ

	Population			Employment			Establishments		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$\sigma = 2$	$\sigma = 5$	$\sigma = 10$	$\sigma = 2$	$\sigma = 5$	$\sigma = 10$	$\sigma = 2$	$\sigma = 5$	$\sigma = 10$
$\Delta_{\{t+3,t\}}$ Ln Air Connectivity ($\sigma = 2$)	0.013***			0.049***			0.034***		
	[0.004]			[0.015]			[0.009]		
$\Delta_{\{t+3,t\}}$ Ln Air Connectivity ($\sigma = 5$)		0.009***			0.032***			0.025***	
		[0.003]			[0.009]			[0.006]	
$\Delta_{\{t+3,t\}}$ Ln Air Connectivity ($\sigma = 10$)			0.008***			0.029***			0.022***
			[0.003]			[0.008]			[0.005]
Ln Population$_t$	-0.159***	-0.159***	-0.159***	-0.009	-0.012	-0.013	0.027	0.026	0.026
	[0.031]	[0.031]	[0.031]	[0.063]	[0.060]	[0.059]	[0.045]	[0.042]	[0.042]
Ln Passengers$_t$	-0.003	-0.002	-0.002	-0.010*	-0.008*	-0.007	-0.004	-0.002	-0.002
	[0.002]	[0.002]	[0.002]	[0.005]	[0.005]	[0.004]	[0.004]	[0.004]	[0.003]
Ln Air Connectivity$_t$ ($\sigma = 2$)	0.006***			0.017**			0.012***		
	[0.002]			[0.007]			[0.004]		
Ln Air Connectivity$_t$ ($\sigma = 5$)		0.004**			0.010**			0.008***	
		[0.002]			[0.005]			[0.003]	
Ln Air Connectivity$_t$ ($\sigma = 10$)			0.003**			0.009*			0.007**
			[0.001]			[0.004]			[0.003]
Initial values	yes	yes	yes	yes	yes	yes	yes	yes	yes
Population lags	yes	yes	yes	yes	yes	yes	yes	yes	yes
Demographic characteristics	yes	yes	yes	yes	yes	yes	yes	yes	yes
Market access	yes	yes	yes	yes	yes	yes	yes	yes	yes
Consolid. CBSA fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes
Period fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	1,482	1,482	1,482	1,482	1,482	1,482	1,482	1,482	1,482
R-sq	0.215	0.270	0.281	0.319	0.422	0.440	0.324	0.418	0.436
First Stage Statistics									
F-stat	13.06	20.08	22.24	13.06	20.08	22.24	13.06	20.08	22.24
Hansen J stat	0.200	0.335	0.441	5.395	5.936	6.033	0.996	1.360	1.545
Hansen J p-val	0.654	0.563	0.507	0.020	0.015	0.014	0.318	0.244	0.214

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

Notes: The reported results correspond to the regression equation (5) estimated by limited information maximum likelihood (LIML) methods using the full set of control variables and fixed effects described in the text. The same explanations as included in the footnote for Tables 3, 5 and 7 apply, except that here we experiment with different values for σ when constructing the air connectivity measure and the associated excluded instruments. The market access variables are modified to use the same distance decay parameter as the excluded instruments (i.e., $(\sigma - 1)/\sigma$). Robust standard errors clustered at CBSA level in brackets.

The first three columns of Table 10 report the results using three-year changes in CBSA population as dependent variable, the next three columns report the results with employment changes as dependent variable, and the last three columns report the results for three-year changes in the number of business establishments. All the reported coefficients are based on LIML estimations using the complete set of control variables and fixed effects. Importantly, the main findings of the paper seem to be robust to the choice of the elasticity of substitution, σ, used in the construction of the air connectivity variable. Both the sign and the statistical significance of the estimated effects are maintained across specifications. More importantly, the estimated effect of air connectivity on urban development is larger the lower the σ value is – a pattern that is consistent across all three dependent variables. So, by assigning more weight to the extensive margin of a city's aviation route structure, we are able to identify a larger effect of air connectivity on employment and business establishment growth. This provides further evidence that adding new non-stop destinations is important for metropolitan areas when considering the expansion of their aviation networks.

The last robustness exercise that we consider examines the sensitivity of our results to the time interval over which we construct changes in the variables of interest. As discussed in the econometric strategy section, the baseline model in equation (5) considers three-year changes in variables. The choice over the interval length was determined in part by the need to consider long-enough periods to capture the discrete nature of changes in aviation routes over time, but also by the need to consider the lag with which urban development indicators may respond to improvements in aviation networks. On the other hand, by extending the time interval by "too long", we face the drawback of a drastic reduction in the number of observations per CBSA while also raising the possibilities of omitted variables and alternative location-specific factors to simultaneously affect urban growth and air connectivity. While we do not have a precise method of assessing the optimal time interval length to consider in our analysis, in Appendix Table B.1 we have experimented with a model specification with multiple year-on-year changes in air connectivity variables going from contemporaneous periods to several years back. Although the exercise only reports OLS estimates for population regressions, the suggestive evidence seems to support our choice of a three-year interval.

Nevertheless, to ensure the robustness of our findings to alternative time horizons, in Table 11 we report results from estimating the regression equation (5) using changes in variables ranging from 1-year differences to 6-year differences. Panel A reports the results using changes in population as dependent variable, Panel B reports the results using changes in employment as dependent variable and Panel C reports the results using changes in the number of establishments. Because of indivisibility issues with our 30-year sample period (1984-2013), we sometimes have residual data at the start or end of the sample period. So, in certain cases, we have some flexibility in deciding the starting year for constructing the change in variables. This is why Table 11 reports different sets of estimates for three-year differences, for five-year differences, etc.

Despite signifiant changes in sample size across specifications, and despite coefficient sensitivities to starting periods or to CBSA sample composition, overall we find that our main findings do not depend on the three-year interval that we considered throughout the paper. Furthermore, consistent with expectation, specifications using shorter time differences such as year-on-year changes in variables lead to lower coefficient estimates compared to longer time differences (even though the latter suffer from higher imprecisions caused by the drastic reduction in sample size).

It is worth mentioning that the estimates reported in column 1 of Table 11, which are obtained from year-on-year differences in variables, can be compared to the results in Sheard (2019, 2021). While there are differences between studies in the estimation samples used and in the regression model specifications, a direct comparison of coefficient sizes indicates that our instrumental variables coefficients are close in range but smaller in magnitude (by 20 to 55 percent smaller) compared to Sheard (2019, 2021). We interpret this observation as consistent with the fact that our air connectivity measure is intended (by design) to only capture the extensive margin growth of a city's aviation network (leaving out intensive margin benefits). So, while the spread and reach of the air connectivity index proposed in this paper is very important to a city's urban growth, it does not capture *all* the effect or benefits of air passenger transportation.

Table 11: Sensitivity Analysis to Changes in Variables over Different Time Intervals

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Period Length (Δ)	**1Yr**	**2Yr**	**3Yr**	**3Yr**	**4Yr**	**5Yr**	**5Yr**	**6Yr**	**6Yr**
Starting Year	1984	1985	1985	1986	1985	1985	1988	1985	1989
Observations	6,561	3,040	1,482	1,484	1,476	982	1,042	755	817
					Panel A: Population				
$\Delta_{\{t+3,t\}}$ Ln Air Connectivity	0.004***	0.006***	0.009***	0.022***	0.018***	0.041***	0.005	0.016***	0.055
	[0.001]	[0.002]	[0.003]	[0.006]	[0.005]	[0.015]	[0.004]	[0.006]	[0.035]
R-sq	0.192	0.267	0.281	0.0829	0.248	-0.116	0.408	0.404	-0.148
F-stat	40.45	23.33	20.32	15.94	20.70	14.84	25.19	20.42	4.843
Hansen J stat	1.413	3.617	0.390	0.121	3.622	6.198	6.013	0.00281	2.338
					Panel B: Employment				
$\Delta_{\{t+3,t\}}$ Ln Air Connectivity	0.020***	0.023***	0.032***	0.033***	0.029***	0.029	0.012	0.018	0.052
	[0.004]	[0.006]	[0.009]	[0.012]	[0.009]	[0.022]	[0.008]	[0.012]	[0.039]
R-sq	0.373	0.501	0.431	0.578	0.601	0.748	0.733	0.735	0.709
F-stat	40.45	23.33	20.32	15.94	20.70	14.84	25.19	20.42	4.843
Hansen J stat	4.341	0.738	5.907	0.893	1.116	8.431	0.0189	2.856	1.141
					Panel C: Establishments				
$\Delta_{\{t+3,t\}}$ Ln Air Connectivity	0.012***	0.017***	0.024***	0.033***	0.031***	0.052**	0.012***	0.029***	0.129**
	[0.003]	[0.004]	[0.006]	[0.009]	[0.007]	[0.021]	[0.005]	[0.009]	[0.058]
R-sq	0.430	0.536	0.429	0.456	0.538	0.495	0.711	0.683	0.211
F-stat	40.45	23.33	20.32	15.94	20.70	14.84	25.19	20.42	4.843
Hansen J stat	5.709	3.725	1.370	0.962	2.462	6.585	0.264	2.856	1.406

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

Notes: The reported results correspond to the regression equation (5) estimated using differences in variables calculated over various length of time and using different starting years (see top of the table). The reported coefficients are estimated by limited information maximum likelihood (LIML) methods. All regressions include the full set of control variables and fixed effects described in the text. Robust standard errors clustered at CBSA level in brackets.

7 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 262 consolidated CBSAs over the period 1984 – 2013, 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 an increase in air connectivity equivalent in size to the average growth observed in our sample is associated with a 0.1 percent growth in city population over a 3-year period (explaining 3.4 percent of the actual observed population change). It leads to a 0.4 percent growth in total employment (explaining 8.4 percent of the actual observed employment change), and to a 0.3 percent growth in the number of business establishments located in that CBSA (explaining 10.1 percent of the actual observed change in establishments). In light of the statistical strength of our exogenous instruments and of the long time period analyzed using a three-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 as well as to offer more flights towards 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 (related) instruments that exploit information on the geography and network characteristics of destination cities. These are factors that, in combination, vary exogenously over time and influence to a heterogeneous degree the number of flights operated out of a given origin city.

Finally, our econometric analysis employs a comprehensive dataset covering approximately 260 consolidated CBSAs observed over a span of thirty years. The richness of our data allows us to employ panel techniques and to 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 depend on the quality and the reach of its aviation network. These findings have important policy implications. They encourage city officials to use their efforts and resources to secure non-stop 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

Allroggen, Florian, Michael D Wittman, and Robert Malina, "How Air Transport Connects the World–A New Metric of Air Connectivity and Its Evolution Between 1990 and 2012," *Transportation Research Part E: Logistics and Transportation Review*, 2015, *80*, 184–201.

Anderson, James E. and Eric van Wincoop, "Trade Costs," *Journal of Economic Literature*, 2004, *42*, 691–751.

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.

Bartik, Timothy J., "Who Benefits from State and Local Economic Development Policies?," 1991.

Baum-Snow, Nathaniel, "Did Highways Cause Suburbanization?," *The Quarterly Journal of Economics*, 2007, *122* (2), 775–805.

Berry, Steven and Panle Jia, "Tracing the Woes: An Empirical Analysis of the Airline Industry," *American Economic Journal: Microeconomics*, 2010, *2*, 1–43.

Berry, Steven T, "Airport Presence as Product Differentiation," *The American Economic Review*, 1990, *80*, 394–399.

_ , Michael Carnall, and Pablo T Spiller, "Airline Hubs: Costs, Markups and the Implications of Customer Heterogeneity," 1996.

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.

Borenstein, Severin, "Hubs and High Fares: Dominance and Market Power in the US Airline Industry," *The RAND Journal of Economics*, 1989, pp. 344–365.

Broda, Christian and David E. Weinstein, "Globalization and the Gains from Variety," *Quarterly Journal of Economics*, 2006, *121*, 541–585.

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

Brueckner, Jan K, Darin Lee, and Ethan S Singer, "Airline Competition and Domestic US Airfares: A Comprehensive Reappraisal," *Economics of Transportation*, 2013, *2*, 1–17.

Campante, Filipe and David Yanagizawa-Drott, "Long-Range Growth: Economic Development in the Global Network of Air Links," *Quarterly Journal or Economics*, 2018, *133*, 1395–1458.

Dixit, Avinash K. and Joseph E. Stiglitz, "Monopolistic Competition and Optimum Product Diversity," *The American Economic Review*, 1977, *67* (3), 297–308.

Donaldson, Dave, "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*, 2018, *108*, 899–934.

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

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 ChinaÕs National Trunk Highway," *Review of Economic Studies*, 2014, *81*, 1046–1070.

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.

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

Gibbons, Stephen and Wenjie Wu, "Airports, Access and Local Economic Performance: Evidence from China," *Journal of Economic Geography*, 2020, *20* (4), 903–937.

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.

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

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

McGraw, Marquise J, "The Role of Airports in City Employment Growth, 1950–2010," *Journal of Urban Economics*, 2020, *116*, 103240.

Mitra, Devashish, Cong S Pham, and Subhayu Bandyopadhyay, "Terrorism and International Air Travel: A Gravity Approach," *The World Economy*, 2018, *41*, 2852–2882.

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," *Economica*, 2019, *86*, 300–335.

— , "The Network of US Airports and Its Effects on Employment," *Journal of Regional Science*, 2021, *61* (3), 623–648.

Simonovska, Ina and Michael E. Waugh, "The Elasticity of Trade: Estimates and Evidence," *Journal of International Economics*, 2014, *92*, 34–50.

Spence, Michael, "Product Differentiation and Welfare," *The American Economic Review*, 1976, *66* (2), 407–414.

Winston, Clifford and Jia Yan, "Open Skies: estimating travelers' benefits from free trade in airline services," *American Economic Journal: Economic Policy*, 2015, *7*, 370–414.

A Inferring the Elasticiy of Substitution σ

The CES utility function in equation (1) delivers the following demand equation for flight frequency f_{id} between the origin i and destination d:

$$f_{id} = \left(\frac{p_{id}}{\alpha_d}\right)^{-\sigma} \frac{E_i}{\sum_{d=1}^{N_i}(\alpha_d)^{\sigma}(p_{id})^{1-\sigma}} \tag{A.1}$$

where E_i denotes the total expenditures of the representative consumer on air travel services.

Taking logs and substituting the location-specific variables, we can re-write equation (A.1) as a gravity model as follows:

$$\ln f_{id} = (-\sigma)\ln p_{id} + \kappa_i + \kappa_d \tag{A.2}$$

where κ_i denotes the origin-specific demand shifters, accounting for the real expenditure term $\frac{E_i}{\sum_{d=1}^{N_i}(\alpha_d)^{\sigma}(p_{id})^{1-\sigma}}$; κ_d denotes the destination-specific demand shifters, accounting for the taste parameter term $(\alpha_d)^{\sigma}$. By estimating the CES demand equation in (A.2), we could uncover the empirical value of $\hat{\sigma}$ from the price elasticity of demand.

Unfortunately, price information for a flight operated on the route from origin i to destination d, i.e., p_{id}, is not available in our dataset. We overcome this limitation by making functional form assumptions about prices. In particular, we consider that the average price p_{id} can be defined as the product between the marginal cost and the price mark-up over the marginal cost, where the marginal cost is proportional to the distance flown, $Dist_{id}$, and the mark-up is a function of market competition at origin airport, respectively market competition at destination airport. Under these assumptions, we can express the average price p_{id} as follows:

$$p_{id} = \mu_i \mu_d (Dist_{id})^{\gamma} \tag{A.3}$$

Taking logs and substituting $\ln p_{id}$ into equation (A.2), we obtain the following gravity equation regression model:

$$\ln f_{id} = \beta \ln Dist_{id} + u_i + u_d + \epsilon_{id} \tag{A.4}$$

where u_i and u_d denote origin, respectively destination fixed effects. Using the estimated coefficient for the distance elasticity $\hat{\beta}$, and estimated values of $\hat{\gamma}$ from the literature, we can then infer the value of $\hat{\sigma}$ based on the underlying structure of the gravity equation:

$$\hat{\sigma} = -\hat{\beta}/\hat{\gamma} \tag{A.5}$$

Using our aviation dataset for the period 1984-2013, Table A.1 reports the estimation results from the gravity equation in (A.4). Column 1 uses all the origin-destination routes in the sample, while Column 2 drops the shortest distance routes, defined as the routes in the lowest quartile distance bin. Column 3 continues to drop the routes in the shortest distance bin, but weights the regression observations by the average number of passengers per departure (i.e., route density).

The last three columns report estimates from subsets of origin-destination routes classified by the quartile distance bin they fall into, with Bin 2 including the second to bottom category and Bin 4 including the longest flights.

In the absence of airfare data, we rely on the literature to get estimates for the price elasticity of distance, γ, in equation (A.3). Brueckner et al. (2013) find a distance elasticities of airfare ranging between 0.28 to 0.32. Borenstein (1989) finds a distance elasticity ranging between 0.33 and 0.41. Using air cargo cost data instead of passenger fare data, Hummels (2007) finds a similar range of distance elasticities between 0.27 and 0.44. Winston and Yan (2015) find a distance elasticity ranging from 0.64 for business travelers to 0.32 for economy travelers. Broadly, we find estimates for $\hat{\gamma}$ that range between -0.3 and -0.6. Combining these $\hat{\gamma}$ values with the $\hat{\beta}$ estimates from columns 2-6 of Table A.1, and applying equation (A.5), we impute estimates for $\hat{\sigma}$ ranging between 2.15 and 7.92.[26]

Table A.1: Gravity Equation Estimation

VARIABLES	(1) All	(2) NoShortDist	(3) Weighted	(4) DistBin2	(5) DistBin3	(6) DistBin4
ln_distance	-0.784***	-1.287***	-1.313***	-1.307***	-1.392***	-2.377***
	[0.028]	[0.037]	[0.040]	[0.112]	[0.158]	[0.422]
Constant	12.476***	16.647***	17.053***	16.362***	17.541***	26.088***
	[0.185]	[0.269]	[0.295]	[0.778]	[1.169]	[3.356]
Observations	82,594	40,295	40,295	16,102	13,197	8,902
R-squared	0.636	0.719	0.729	0.805	0.797	0.749

Robust standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.10, + p<0.15

Notes: The reported coefficients come from estimating equation (A.4) using origin-destination airport data for 1984-2013. The dependent variable is the log number of departures performed on a route. All routes in the sample are split into 4 quartiles based on distance. Column 1 uses all the origin-destination routes in the sample. Column 2 drops the shortest distance routes, defined as the lowest quartile routes by distance (below 800 kilometers). Column 3 uses only routes in distance bins 2-4, but weights the observations by the average volume of passengers per flight (i.e., route density). Column 4 estimates the gravity model using only routes in Distance Bin 2, which ranges between 800km and 1312km. Column 5 reports the estimates based on the subset of routes in Distance Bin 3, which ranges between 1312km and 2076km. Column 6 reports the estimates based on the subset of routes from Distance Bin 4, which ranges between 2076km and 4321km. All estimations include origin-year and destination-year fixed effects. Reported standard errors are clustered by origin airport.

[26] We disregard the distance elasticity from Column 1, which include short distance routes, because of any possible nonlinearities in prices over shorter distances. In fact, many of the studies providing $\hat{\gamma}$ estimates for the price elasticity of distance involve predominantly longer-haul routes.

B Appendix Tables

Table B.1: OLS Estimates Using Lagged One-year Changes in Air Connectivity

	(1) Population	(2) Employment	(3) Establishments
$\Delta_{\{t+3,t+2\}}$ Ln Air Connectivity	0.007***	0.020***	0.009***
	[0.002]	[0.004]	[0.003]
$\Delta_{\{t+2,t+1\}}$ Ln Air Connectivity	0.004**	0.010***	0.008***
	[0.002]	[0.003]	[0.003]
$\Delta_{\{t+1,t\}}$ Ln Air Connectivity	0.003**	0.010***	0.004*
	[0.002]	[0.003]	[0.002]
$\Delta_{\{t,t-1\}}$ Ln Air Connectivity	0.002	-0.001	0.002
	[0.001]	[0.003]	[0.002]
Ln Population	-0.164***	-0.033	0.011
	[0.031]	[0.051]	[0.039]
Ln Employment	0.012	-0.349***	0.023
	[0.022]	[0.046]	[0.027]
Ln Establishments	0.041**	0.103**	-0.218***
	[0.020]	[0.045]	[0.034]
Ln Passengers	-0.003+	-0.009***	-0.004
	[0.002]	[0.003]	[0.003]
Ln Air Connectivity	0.003**	0.008**	0.003
	[0.001]	[0.003]	[0.002]
Initial values	yes	yes	yes
Population lags	yes	yes	yes
Demographic characteristics	yes	yes	yes
Market access	yes	yes	yes
Consolid. CBSA fixed effects	yes	yes	yes
Period fixed effects	yes	yes	yes
Observations	1,482	1,482	1,482
R-squared	0.331	0.531	0.547

*** p<0.01, ** p<0.05, * p<0.10, + p<0.15

Notes: The reported results correspond to equation (5) estimated by limited information maximum likelihood (LIML). The sample is an unbalanced panel of 262 consolidated CBSAs observed every 4 years over the period 1984-2013. The dependent variable is the 3-year change in log population (column 1), in log employment (column 2), respectively in log number of establishments (column 3), calculated over the period [t, t+3]. The variable of interest are year-on-year log changes in air connectivity (as defined in equation (4) using $\sigma = 5$). All columns include the complete set of control variables and fixed effects described in the main text. Robust standard errors clustered at CBSA level in brackets.