Air Traffic and Urban Growth: Evidence from Airline Networks

Grace Wang*

November 15, 2019

Job Market Paper

Abstract

This paper studies the causal effects of air traffic growth on local economic outcomes in the US metropolitan areas from 1993 to 2016. For the interdependence concern between air traffic and local economic growth, I apply Bartik shift-share (1991) on the airline networks to construct predicted air traffic as an instrument. Instruments for small airports and large hubs are constructed differently because of their different operating systems. For small airports, I multiply the air traffic growth in each hub by their shares in the small airport in the initial year and then sum the products across hubs as an instrument for air traffic. To choose hubs under the hub-and-spoke system, I apply the K-Means clustering unsupervised machine learning technique. The instrument of air traffic for large hubs are constructed similarly based on operation airlines. Results for cities with small airports reveal that a 10% increase in air traffic growth leads to a 0.64% increase in local employment growth, implying 3,200 new jobs created in a typical city. No statistically significant effects exist in cities with large hubs. The mechanism of these effects is mainly through labor demand. Financial service sectors are the main beneficiary of air traffic growth.

Keywords: air traffic, local economic outcomes, labor market, causal effect, machine learning

JEL codes: C26, C63, J01, L93, O18, R41

* Department of Economics, The University of Arizona, chaowang0827@email.arizona.edu. I thank Price V. Fishback, Mauricio Varela, Tiemen Woutersen and all participants in empirical seminars in the University of Arizona for useful comments. All errors are mine.

1 Introduction

Policymakers at all levels of governments have spent considerable resources to promote air traffic, motivated by the underlying belief that improved air traffic will increase their regional economic growth. In the United States, government spending on airports is substantial: the annual budget of the Federal Aviation Administration (FAA) is around $16 billion to fund airport operations, construction and maintenance, and research and development. Also, early in 2018, the Trump administration introduced a plan calling for $200 billion in federal funds and requiring a 4-to-1 match from state and local governments to stimulate about $1.5 trillion in spending to upgrade and improve the country's roads, bridges, airports, and other key infrastructure. The mechanism for air traffic to promote regional economic development is that new and improved airports provide better transportation facilities and potentially attract new investment to establish firms and increase employment. However, there is little empirical research justifying this positive effect, especially in modern times. The purpose of this paper is to estimate the effects of air traffic growth on local economic activities and how it differs between small airports and large hubs.

In general, it is challenging to estimate the causal effect of air traffic growth on local economic outcomes because of the interdependence between air traffic and local economic growth. A rapidly growing local economy is likely to demand more air traffic and tend to invest more in airport infrastructure. In turn, improved air transportation stimulates regional development. Moreover, both the local economy and air traffic can be affected by some external factors or events, such as government spending. Therefore, a direct regression between air traffic and local economic outcomes more likely reflects factors other than the causal effects we are interested in.

To identify the causal effect of air traffic on local economic growth, we need to find some variation in air traffic that is not correlated with factors affecting local economic outcomes. Most of the literature takes advantage of the 1978 airline deregulation as an exogenous shock to the air traffic and estimates the effect of air traffic growth on the local economic outcomes. However, the time frame is limited to the era between the 1970s and the 1980s. In this paper, I apply novel techniques to construct instruments for air traffic that are plausibly unrelated to the local economic outcomes. Then, I estimate the effect of air traffic changes associated with these instruments on the local economic outcomes to identify the causal effects of air traffic on local economic outcomes.

The instruments are constructed based on the shift-share technique that is proposed by Bartik (1991), but the construction differs between small airports and large hubs. For large hubs, the instrument for air traffic growth is constructed as the sum of shares of air traffic of each airline in the initial year multiplied by corresponding airline growth rates at the national level. The underlying assumption is that airline growth at the national level affects the air traffic growth at the hub but is uncorrelated with local economic growth at the hub. This assumption holds for large hubs because hubs are centers through which airline flights are routed, thus the air traffic at hubs are responsive to the nationwide airline growth. However, this assumption does not hold for small airports. Small airports have a much lower weight in the nationwide air traffic and thus there is no clear clue that they respond to the entire airline development. For small airports, the instruments are constructed based on the hub-and-spoke operation system. A majority of airlines in the United States have been operating under the hub-and-spoke system since airline deregulation in the late 1970s. Under this system, a hub operates as a center and connects all outlying airport points as "spokes." Thus, non-hub airports connect with other airports through a route to and out of a hub. Therefore, for small airports, instruments for air traffic are constructed as the sum of shares of air traffic at each connected hubs in the initial year 1993 multiplied by the air traffic growth rates of the corresponding hubs. The identification assumption is that the air traffic growth at the connected hubs has effects on the air traffic in the small airports through the hub-and-spoke system, while it is not directly correlated with the local economic growth where the small airports are located. This assumption is supported by the hub-and-spoke system because the more air traffic growth in the connected hub, the more air traffic can be directed to the "spoke" (small airport). More restrictions are applied to validate that connected hubs do not directly affect the local economic growth with small airports.

The results show that air traffic growth has a statistically significant positive effect on local employment, with an elasticity of 0.064. This estimate implies that in a typical core-based standard area (CBSA) with 1,000,000 residents and a 50% employment rate, a 10% increase in the air traffic will create 3,200 new jobs. Moreover, particular industry sectors are the main drivers for the effect based on the industry-level regressions. The industry sectors that have statistically significant positive relationships with air traffic growth are mining, construction, manufacturing, finance, insurance, and real estate.

The effects on alternative economic measures are estimated, including numbers of establishments, population, income, and total income growth, aggregate payroll, and wages. The results show that air service growth has statistically significant positive effects on population and aggregate payroll. These results imply that an increase in air traffic causes an increase in labor demand. There is no statistically significant effect on the wage, implying that either the labor supply is sufficiently elastic so that employment adjusts without a substantial increase in wages, or that the increased air traffic causes the labor supply to shift outwards. The magnitude of effects on population is much smaller than the one on employment and aggregate payroll, which suggests that the shift in labor demand dominates the shift in labor supply.

The results do not reveal any statistically significant effects of air traffic on economic outcomes at CBSAs with large hubs. Only in the finance, insurance, and real estate sector does there exist a significant positive effect of air traffic growth on employment at the hubs. The elasticity of 0.504 is statistically significant at a 10% level.

The contribution of this paper to the existing academic literature is threefold. First, it fills the gap in the study of the effects of air service on local economic outcomes in modern times, when airport infrastructure still involves substantial government spending every year. Despite a large literature on roads and railways, there exist few empirical research papers on the air traffic effects, especially in current times. Brueckner (2003) uses hub status and its proximity to the nearest cities as instruments and estimates the effect of air traffic in particular industries in 1996. His results show that a 10 percentage-points increase in enplaned passengers significantly leads to a 1 percentage point increase in local employment, with magnitudes close to the estimates in this paper. The service sector counts for most of the effects in his results. However, geographic factors and regional economics are dependent since the literature shows that spatial location influences regional economic growth through market access effect ((Redding and Venables (2004), Hanson (2005), Head and Mayer (2006), and Donaldson and Hornbeck (2016)). Green (2007) estimates the effects of air traffic growth on population and employment in the decennial census from 1990 to 2000 and finds statistically significant positive effects. However, his technique still does not solve the endogeneity problem since air traffic and economic outcomes are determined simultaneously. Blonigen and Cristea (2015) take advantage of the 1978 airline deregulation as a quasi-natural experiment and use a difference-in-difference method to estimate the effects of air services on regional growth. They find that for a given city, a 50% increase in the enplaned passenger growth accounts for an increase in annual population growth of between 1.55% and 4.2%. McGraw (2015) estimates the effects of small and mid-size commercial airports on local employment from 1950 to 2010 using three aviation policy changes as instruments. His findings reveal that the presence of an airport caused roughly 3.2% employment growth per decade in the business and service sectors. McGraw (2017) uses hub closure as a plausibly exogenous change in the airline network structure and estimates the marginal effect of a hub presence on the economic outcomes in the metropolitan areas during 1978-2012. The results show that a hub closure causes personal income to decrease by 2.3% and the number of establishments to fall by 1.6%. Campante and Yanagizawa-Drott (2017) exploit variation in long-range flights due to regulatory and technological restrictions and find a statistically significant positive effect of air traffic on local economic activities.

Second, it constructs instruments based on the air network. Specifically, it applies the Bartik shift-share technique on small airports under the hub-and-spoke system; for large hubs, it categorizes the shift-share by airline companies. Prior literature uses various techniques, either applying geographic or historical variables as instruments or exploiting variation in air traffic caused by policy changes or random discontinuity. Brueckner (2003) uses hub status and distance to the population center as instruments since the geographic centrality increases the likelihood of a city to be a hub. Green (2007) uses decennial air traffic from an earlier period as an instrument for current air traffic. Sheard (2014) uses the 1944 National Airport Plan as an instrument to address the potential endogeneity problem. Blonigen and Cristea (2015) use air traffic changes caused by the 1978 Airline Deregulation as an instrument. McGraw (2016, 2017) uses hub closures and three aviation policy changes as instruments for air traffic changes. Campante and Yanagizawa-Drott (2017) exploit a technical cutoff of long-distance flight at 6,000 miles and use this discontinuity as an instrument.

My instrument construction has several advantages over alternative methods, at least for the study of air traffic. First, the data required to implement my method are publicly available and the identification method is relatively simple. It does not require detailed historical data or the existence of a particular policy change. Second, this method can be implemented without time constraints or strict identification assumptions. It does not rely on the assumption that a specific policy change is exogenous to the outcomes of interest. Also, this method is relatively powerful and informative for policy analysis to estimate the short-term effects.

Third, it estimates the effects of air traffic by airport category and constructs instruments separately for small airports and large hubs. Sheard (2019) constructs instruments for air traffic based on the Bartik (1991) shift-share for all selected commercial airports and finds significantly positive effects on local economic outcomes. However, the shift-share construction relies on the assumption that national-level changes in divided categories correlate with changes in local air traffic, this assumption does not hold for small airports. In this paper, I apply the hub-and-spoke network system on the shift-share instrument construction for small airports. The results show statistically significant positive effects of air traffic on economic outcomes at CBSAs with small airports and statistically insignificant effects of air traffic on local economic outcomes at CBSAs with large hubs. These contrasting results imply that policymakers who seek to promote local economic growth should invest more in small airports, as they are more responsive to increased air traffic while air traffic in large hubs though attracts more attention has insignificant effects on local economic outcomes.

The remaining sections of the paper are structured as follows. In section 2, I present the theoretical model that formalizes the empirical estimation. Section 3 describes the data used in the analysis. In section 4, I present the IV empirical strategies that guide the empirical analysis. Section 5 shows the main findings of the study. I present results from a series of robustness checks in section 6. Finally, conclusions and remarks are presented in section 7.

2 Theoretical Model

This section presents a simple framework that formalizes the basis for the empirical estimation. In this model, it allows air traffic to function as a shifter for productivity and local amenity to affect local economic outcomes. I follow the theoretical setup in Glaeser, Scheinkman, and Shleifer (1995) and derived the equilibrium for empirical estimation.

The model treats each area as a separate open economy that has common national stocks of capital and labor endowments as well as free factor mobility. In equilibrium, capital and labor will be distributed across areas such that the rental rate and per-capita income, adjusted for local amenities, are equalized. This equilibrium outcome implies that neither exogenous changes in labor supply nor saving rates can be used as explanations for urban growth variation. Instead, factors of local fundamentals should be considered. Without loss of generalization, I assume areas differ only in the level of productivity and the quality of life determined by local amenities.

Suppose the total output in a metropolitan area as a Cobb-Douglas production function: $\mathrm{Y_{it}} = \mathrm{A_{it}f(L_{it})} = \mathrm{A_{it}L_{it}^{\alpha}}$, where $\mathrm{A_{it}}$ represents the level of productivity in area i at time t, L_{it} measures the employment-population of area i at time t, and α is the production parameter. Individuals gain utility from the earned labor income and the quality of life. Workers get paid the value of their marginal productivity by normalizing the output price to one. Thus, the labor income can be represented as: $\mathrm{W_{it}} = \alpha \mathrm{A_{it}L_{it}^{\alpha-1}}$. The quality of life $\mathrm{Q_{it}}$ captures location-specific factors. Without loss of generalization, I assume $\mathrm{Q_{it}}$ decreases as the employed population rise in area i at time t, through the impact of population size on housing prices, traffic congestion, criminality and other factors (Diamond, 2016). Also, the life quality term varies with local amenities that are exogenous to the production technology, denote as $\mathrm{M_{it}}$. Thus, $\mathrm{Q_{it}} = \mathrm{L_{it}^{-\delta}M_{it}}$, where $\delta > 0$. Therefore, the individual's utility function is given as:

$$\mathrm{U_{it}} = \gamma_1 \log(\mathrm{W_{it}}) + \gamma_2 log(\mathrm{Q_{it}}) = \gamma_1 log(\mathrm{A_{it}}) + \gamma_2 log(\mathrm{M_{it}}) - \gamma_3 log(\mathrm{L_{it}}) + c \quad (1)$$

where $\gamma_3 = \gamma_2\delta + \alpha - 1, c = \gamma_1 loga$.

Assume air traffic affects individual utility through its effects on productivity and local amenities:

$$\begin{cases} log(\mathrm{A_{it}}) \sim log(Air_{it}) + \epsilon_t^1 + \epsilon_i^1 + \epsilon_{it}^1 \\ log(\mathrm{M_{it}}) \sim log(Air_{it}) + \epsilon_t^2 + \epsilon_i^2 + \epsilon_{it}^2 \end{cases} \quad (2)$$

Therefore, an individual's utility can be expressed as:

$$\mathrm{U_{it}} = \theta_1 log(\mathrm{Air_{it}}) - \theta_2 log(\mathrm{L_{it}}) + \epsilon_i + \epsilon_t + \epsilon_{it} \quad (3)$$

In the equilibrium, free mobility of factors guarantees that individual utility is constant across areas at a given time. Denote the constant equilibrium utility as $\overline{U}_t$. Replace equation (3) by moving $\mathrm{L_{it}}$ to the left and $\overline{U}_t$ to the right, we can get:

$$log(\mathrm{L_{it}}) = \beta_0 + \beta_1 log(Air_{it}) + v_i + v_t + v_{it} \quad (4)$$

where v_i captures the area fixed effect for employment population, v_t captures the time fixed effect for employment population, v_{it} captures residuals that affect the employment population.

Substituting the equation (4) at time $t+1$ and subtracting it at time t yields the following equation:

$$log\left(\mathrm{L_{i(t+1)}}\right) - log(\mathrm{L_{it}}) = \beta_1 \left(log\left(Air_{i(t+1)}\right) - log(Air_{it})\right) + v_{t+1} - v_t + v_{i(t+1)} - v_{it} \quad (5)$$

I divide the residual term $v_{i(t+1)} - v_{it}$ as: $v_{i(t+1)} - v_{it} \sim log(L_{it}) + log(Air_{it}) + \varepsilon_i + \varepsilon_{it}$, where ε_{it} is uncorrelated with $log(L_{it})$ and $log(Air_{it})$. $log(L_{it})$ and $log(Air_{it})$ captures factors that affect the employment growth rate $log(\mathrm{L_{i(t+1)}}) - log(\mathrm{L_{it}})$ in levels of employed population and air traffic, ε_i captures the area fixed effect and the rest is captured in ε_{it}.

Therefore, the main empirical estimation equation is:

$$\mathrm{l_{it}} = \alpha_0 + \alpha_1 air_{it} + \alpha_2 log(L_{it}) + \alpha_3 log(Air_{it}) + \varepsilon_t + \varepsilon_i + \varepsilon_{it} \tag{6}$$

where $\mathrm{l_{it}} = log(\mathrm{L_{i(t+1)}}) - log(\mathrm{L_{it}})$ represents the growth rate of the employed population, $air_{it} = log(Air_{i(t+1)}) - log(Air_{it})$ captures the growth rate of air traffic. $log(L_{it})$ and $log(Air_{it})$ captures the factors in levels of employment and air traffic at time t, which are uncorrelated with the residual ε_{it}. ε_{it} captures all the remaining terms that affect employment growth, which may also affect air traffic growth through productivity and local amenity growth. For example, increases in local government spending on infrastructure will affect both the local air traffic and employment number. Thus, an instrument that correlates with changes in the air traffic growth and does not correlate with local economic outcomes, are needed.

3 Data

There are three main datasets used for the analysis: annual panel datasets of US air traffic, coupon data, and economic outcomes. The timeframe covers the period from 1993 to 2016. Air traffic and economic outcomes are assembled from several data sources and aggregated at the Core Based Standard Area (CBSA) defined based on CBSA to FIPS County Crosswalk from the National Bureau of Economic Research in 2013. A Core Based Statistical Area (CBSA) is defined by the Office of Management and Budget (OMB) based around an urban center and adjacent counties or county-equivalents; the urban center must have at least 10,000 people. The reason to analyze at the CBSA level is based on the results of Brueckner, Lee, and Singer (2014) that city pairs are the appropriate level of aggregation for analyses in passenger air traffic, rather than airport pairs.

3.1 Air Traffic

The air traffic data are from the T-100 Domestic segment data in the Bureau of Transportation Statistics (BTS). This dataset contains domestic non-stop segment data by U.S. air carriers, including carrier, origin (airport location ID, state), destination (airport location ID, state), transported passengers, airline, service class. Only scheduled and non- scheduled service classes are included. In order to get the exact airport address, I use the Airport Data and Contact Information dataset from the Federal Aviation Administration, which includes airport longitude and latitude[1], airport location ID[2], county and state.

The sample is restricted to contiguous states of the United States, excluding Alaska, Hawaii, the U.S. Virgin Islands, Puerto Rico, and U.S. Pacific Trust Territories and Possessions. The main measure of air traffic is the number of passengers. In this paper, air traffic is bidirectional, which means air traffic is the sum of both the boarding and departing passengers at the CBSA level. Based on the FAA commercial airport definition, the sample is limited to airports having at least 2,500 total passengers boarding per year from 1993 to 2016 and receiving scheduled passenger service.[3] Only CBSAs with such airports are included in the sample.

3.2 Coupon Data

The coupon data come from the Origin and Destination Survey (DB1B) Coupon Data from the US Department of Transportation. This dataset is available since 1993. It contains a 10% sample of airline tickets from all the reporting carriers, including origin, destination, ticketing airline, passengers, break identifying whether a passenger is assumed to have stopped for a reason other than changing planes.

The dataset is mainly used in the K-Means clustering to get the hub list that operates under the hub-and-spoke system, which is discussed in detail in the empirical strategy section 4.2. Based on the “break” variable, I construct ratios of routes and passengers into the airports only to transfer to the next stops. This analysis categorizes airports into large hubs and small airports. In the sample, there are 269 small airports located in 243 CBSAs and 38 selected large hubs located in 31 CBSAs. For CBSA with multiple airports of the same category, I sum up all the passengers of these airports as the air traffic at the CBSA level.

[1] FAA reports airport longitude and latitude in sexagesimal degree with degrees, minutes, and seconds, and they are transformed into decimal degrees, which is a common geographic location format.

[2] Airport location ID in T-100 is in IATA format while the airport address dataset is in FAA format. Most IATA and FAA airport location ID are the same, and 15 airports in the selected airports have different location ID in FAA from their IATA formats. All are transformed into FAA format.

[3] Therefore, a commercial airport is the one with at least 5,000 total passengers, boarding on or departing from the CBSA level.

3.3 Economic Outcomes

The main economic outcome measure is employment. Alternative economic measures include population, total income, income per capita, number of establishments, total payroll and wage. These variables are all at the county level and are aggregated into the CBSA level based on the OMB definition in 2013.

Data on employment, total payroll, wage and the number of establishments are from the County Business Patterns (CBP) from 1993 to 2016. All these economic measures in the industrial sectors are categorized based on NAICS definitions. In the CBP dataset, the industry category is based on SIC before 1997. Details about incorporating industry definition from SIC to NAICS can be found in Appendix Table A1. Data on Population, total income and income per capita are from the Bureau of Economic Analysis (BEA) at the county level.[4] All nominal variables are converted into real ones using annual CPI in the United States from 1993 to 2016. The CPI dataset is from the Bureau of Labor Statistics.

3.4 Summary Statistics

Table 1 shows the summary statistics of the main variables for small airports and large hubs. On average, CBSAs with small airports have fewer population, employment numbers, and numbers of firms; lower wages, personal income per capita, total wage (total payroll), and total personal income. The mean number of total passengers per year of CBSAs with small airports are 1,390,000 while the ones of CBSAs with large hubs are 28,000,000. CBSAs with small airports connect to 12.76 large hubs on average, with 1 minimum connection hub and 38 maximum connection hubs. CBSAs with large hubs have a mean of 23.74 operation airlines, with a minimum of 6 and a maximum of 46.

4 Empirical Strategy

To deal with the interdependence between air traffic and local economic outcomes, I construct instruments that reflect changes in air traffic without directly impacting the local economic

[4] Most BEA counties and county equivalents are identical to Federal Information Processing Standard (FIPS) codes, modified FIPS in BEA are converted to standard FIPS and more details can be found here: https://apps.bea.gov/regional/docs/statelist.cfm.

outcomes. I apply a Bartik (1991) shift share to construct the instrument for small airports and large hubs separately based on their different roles under the airline operation system. The Bartik (1991) shift-share uses the sum of the national employment growth rate in different industry sectors multiplied by the local industry shares in the initial year as an instrument for local employment growth rate. I adapt the Bartik (1991) shift-share to the airline networks. For small airports, I first apply the K-Means clustering to get a list of large hubs that operate mainly under the hub-and-spoke system and then construct the instrument for the air traffic growth in the small airports. To get an estimate of the traffic growth in the airport, I multiply the hub traffic growth by airport shares in each hub in the initial year and then sum across hubs (see Section 4.3 for details). Namely, I fix shares of the airport traffic at each hub in the initial year and then move the instrument across time using the growth in hubs (leaving the traffic in the small airport out). For large hubs, I multiply the air traffic growth of each major airline by the share of the corresponding airline in the initial year at the hub and then sum across airlines (see Section 4.4 for details). This adapted shift-share only works for large hubs since large hubs hold large enough market shares of major airlines to ensure a significant correlation between airline growth rate in the national level and air traffic growth in large hubs. For small airports, this correlation is insignificant and the underlying economic explanation is not well-established. The pervasive existence of the hub-and-spoke operation system provides another method to construct the shift-share for the small airports.

4.1 Background Content

There are two airline operation systems in the United States: hub-and-spoke and point-to-point systems. A hub is a central airport through which flights from or out of outlying points are connected and spokes are the routes that connect outlying points to the central hub. The hub-and-spoke system transports passengers from the origin to a destination through a central hub. For example, you are in Tucson, Arizona, and want to go to Portland, Oregon by American Airlines. The demand for the direct route of Tucson-to-Portland is not large enough for American Airlines to offer a direct flight, so there is no non-stop flight from Tucson to Portland by American Airlines. Los Angeles International Airport is a central hub for American Airlines. Instead, American Airlines flies you from Tucson to Los Angeles and then transfers you to Portland via a connecting flight out of Los Angeles. In contrast, a point-to-point system transports passengers directly to a destination with no stops, rather than going through a central hub.

The hub-and-spoke system has several advantages over the point-to-point system. It requires fewer routes to connect to all other airports. For each route, the plane carries more passengers because it assembles the demand of passengers from more than one airport. High capacity lowers the average operation cost for each passenger and therefore, it saves airline money. After the airline deregulation in 1978, it became the norm for most major airlines. Until 2001, most of the 12 major US passenger airlines operate under the hub-and-spoke system.

Therefore, for small airports, air traffic in the connected hubs provides wider airline network for the small airports and thus promote air traffic growth in the small airports. At the same time, the air traffic in the connected hubs does not directly affect the local economic growth in the areas with small airports if the connected hubs are located far away. I restrict the minimum distance to the connected hub at 100 miles for the baseline model and extend the restriction to farther distance.

4.2 Choosing Hubs using K-Means Clustering

There is no unique definition of hubs that people agree upon and hubs operating mainly under the hub-and-spoke system are not available. In order to construct the instrument for traffic growth in small airports, I use a machine learning technique, K-Means clustering, as a way to identify hubs. In the baseline model, I first use the Origin and Destination Survey (DB1B) Coupon Data from the US Department of Transportation to construct the route and passenger ratios of going through the airport as a stop. The DB1B coupon data provide a "break" variable that identifies whether a passenger stops for a reason rather than changing planes. Then, I apply the K-Means clustering on the route and passenger ratios for all medium and large airports defined by FAA each year from 1993 to 2016. Because there exist thousands of airports, I restrict clustering work to just medium and large airports as defined by the FAA. FAA defines a "medium" airport as having at least 0.25% of the nationwide total annual boarding passengers and a "large" airport as having at least 1% of nationwide total annual boarding passengers[5]. The medium and large airport lists are obtained based on the T100 dataset and a totally of 70 airports are on the list.

K-Means clustering is one unsupervised machine learning method that assigns a cluster label to examples based on the input variables. In the baseline model, the input variables $x^{(i)} \in \mathbb{R}^2$ are the

[5] FAA defines publicly owned commercial airports with at least 2500 annual boarding passengers; small hub has at least 0.05% but less than 0.25% of overall boarding passengers nationwide; median hubs has at least 0.25% but less than 1% of overall boarding passengers nationwide; large hubs has at least 1% of overall boarding passengers nationwide. Please find more details here: https://www.faa.gov/airports/planning_capacity/passenger_allcargo_stats/categories/.

route and passenger ratios of stopping at the airport just for changing planes. The algorithm works in four steps:

1. Randomly initialize cluster centroids $\mu_1, \mu_2, \ldots, \mu_K \in \mathbb{R}^2$;
2. For each $x^{(i)} \in \mathbb{R}^2$ in the median and large hub list, calculate its distance to all the cluster centroid and assign it to the index of the closest cluster centroid $c^{(i)}$:
$$\min_{c^{(i)}} \left\| x^{(i)} - \mu_{c^{(i)}} \right\|^2$$
3. For each $k = 1\ to\ K$, update the cluster centroids μ_k using the average (mean) of points assigned to cluster k;
4. Repeat the above steps until the cluster centroids do not change.

Figure 1 shows all the clustering results from 1993 to 2016. The highest cluster centroid is around 0.7 in passenger ratio and 0.6 in the route ratio, implying that most of the air traffic goes through the hub as a temporary stop rather than a destination. The second highest cluster centroid stays around 0.4 in the passenger ratio and 0.2 in the route ratio. The last cluster centroid stays around 0.1 in passenger ratio and 0.05 in the route ratio, implying that most of the air traffic goes through the airport as a destination. Higher ratios further imply that the airport carries a higher percentage of passengers or routes under the operation of the hub-and-spoke system. I select the airports in the highest two clusters to be in the hub list. All other commercial airports are included in the small airport list in the sample. Any airport that has ever been clustered into the hub list in any year is selected as a large hub in the baseline model.

To choose the number of clusters K, for each number of clusters K, I calculate the cost function which is the mean squared distance between each example and their assigned cluster centroid:

$$J\left(c^{(1)}, \ldots, c^{(m)}, \mu_1, \ldots, \mu_K\right) = \frac{1}{m} \sum_{i=1}^{m} \left\| x^{(i)} - \mu_{c^{(i)}} \right\|^2$$

Figure 2 shows the cost function of the number of clusters K ranging from 1 to 11 for K-Means clustering on median and large airports across years. Most graphs tend to have a sharp decline in the cost function between $K = 1$ and $K = 2$. The slope becomes moderate from $K = 2$ to $K = 3$ and comparatively, flat when $K > 3$. These tendencies resemble an elbow and thus I apply the elbow method to choose the number of clusters $K = 3$ for all the K-Means clustering across years.

Based on the K-Means clustering results, I get 245 CBSAs with 269 small airports and 31 CBSAs with 38 large hubs in the baseline analysis. Figure 3 shows the kernel nonparametric density of passengers transferred through these 38 hubs. From the figure, we can see that the density distribution stays around 0 before 0.6 and the highest density is around 0.9, implying that most small airports transfer at least 60% of the annual passengers through their connected hubs. Figure 4 shows the distribution of selected small airports and large hubs as well as their CBSAs. The flight represents one selected large hub. CBSAs with these large hubs are solid filled. Small airports are represented by a small circular dot, and their CBSAs are shaded by stripes. From the distribution map, we can see that most small airports concentrate in the southeastern and northwestern areas and surrounds around large hubs. Correspondingly, large hubs are in large cities, and small airports are located in the median and small cities. While most small airports and large hubs are in different CBSAs and have noticeable distance between them, there exist some exceptions that small airports and large hubs are located in the same CBSAs or close to each other. I exclude CBSAs within 100 miles to the large hubs from the sample of analyzing air traffic in CBSAs with small airports on local economic outcomes in the robust test.

4.3 IV Construction for Small Airports

To break the potential interdependence between unobserved factors in local economic outcomes and air traffic growth, for small airports, I use variation in air traffic within each small airport, generated by local exposure to connected hub-specific shocks. Specifically, I apply Bartik shift-share in the connected hubs that are clustered as operating mostly under the hub-and-spoke system. Here are steps to construct the instrument at the CBSA level:

1. For each small airport i, I calculate the passenger share s_{ih93} at each connected hub h for airport i at baseline time t_0 (in year 1993):

$$s_{ih93} = \frac{Air_{ih93}}{\sum_{h \in H} Air_{ih93}} \tag{7}$$

where Air_{ih93} is the air traffic (total boarding and departing passengers) in airport i to connected hub h at the initial year 1993; H is the hub list obtained from K-Means clustering; if airport i does not connect to hub h in the initial year, $Air_{ih93} = 0$.

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2018) suggest using the share in the initial year rather than the lagged year because share in the lagged year suffers from potential interdependence across years. Davila (2018) also uses shares in the initial year when constructing the Bartik shift-share.

2. Construct the instrument $\widetilde{air}_{it}$ for the air traffic growth air_{it} at airport i at time t:

$$\widetilde{air}_{it} = \sum_{h \in H} air_{h_{-i}t} \times s_{ih93} \tag{8}$$

where $air_{h_{-i}t}$ is the air traffic growth in hub h at time t, excluding air traffic originating or terminating at airport i at time t.[6] Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2018) call it the "leave-one-out" method. This prevents air traffic growth at airport i from contaminating the air traffic in connected hubs, which may potentially violate the exclusion condition of the instrument. Air traffic growth at a given airport interdepends with the local economic outcomes of interest, since local air traffic is a part of the air traffic in the connected hubs, there is a potential interdependence between the local economic outcomes and air traffic growth in the connected hubs. Excluding local air traffic at airport i from air_{ht} reduces this concern. Also, I restrict the distance between connected hubs and the small airport to be at least 100 miles and extend this restriction further in the robust test.

A valid instrument should satisfy the relevance and exclusion conditions. The relevance condition is straightforward to test. The underlying economic intuition is that an increase in the air traffic of connected hubs is expected to lead to more passengers from given airports and wider connection networks to final destinations, under the operation of the hub-and-spoke system. The underlying rationale for the exclusion assumption is that hub traffic growth after leaving out the small airport is unlikely to be correlated with unmeasured activity at the local area where the small airport is located, particularly controlling for CBSA and year fixed effects in the analysis.

3. Aggregate instrument for air traffic at the airport level into the CBSA level:

$$\widetilde{air}_{ct} = \sum_{i \in c} \widetilde{air}_{it} \times r_{ci93} \tag{9}$$

where $r_{ci93} = \frac{\sum_{h \in H} Air_{ih93}}{\sum_{i \in c} \sum_{h \in H} Air_{ih93}}$. c is the CBSA holding small airports. From Figure 4, there exist multiple small airports in one CBSA. Aggregation from the airport level to the CBSA level ensures the air traffic growth is from the connected hubs rather than the connected CBSAs, which ensures that the instrument construction adheres to the hub-and-spoke system.

[6] I appreciate Nicholas Sheard for the suggestion on excluding air traffic originating or terminating at the small airports when calculating the hub air traffic growth rate.

Figure 5 shows an airline network between Tucson and all connected hubs from 1993 to 2016. Tucson has a dense network and has connected to 37 out of all the 38 hubs across the years. Most connected hubs are in the northeastern and southeastern regions. The top 5 hubs that have carried the most passengers are Dallas/Fort Worth International, Phoenix Sky Harbor International, Los Angeles International, McCarran International, and Denver International Airport, marked by blue larger airplane shape. Figure 6 shows an airline network of the Flagstaff Pulliam Airport. Comparatively, Flagstaff Pulliam Airport has a less dense network, connecting to only 13 hubs. The top 5 hubs that have carried the most passengers are Phoenix Sky Harbor International, Los Angeles International, Denver International, San Francisco International, and Portland International Airport. All connected hubs are at least 100 miles from the small airports, which ensures that large connected hubs only affect small airports through air traffic.

4.4 IV Construction for Large Hubs

For large hubs, instead of constructing the instrument at the hub level and then aggregating into the CBSA level, I directly construct the instrument at the CBSA level for large hubs. Both constructions return the same results. First, I keep the first 99 largest airline based on the mean of annual passengers from 1993 to 2016 and categorize the remaining airlines as "Other" airline. For each airline a, I calculate their annual growth rate at the national level across the years and their shares at each CBSA with large hubs in the initial year 1993. Thus, the instrument is the sum of each airline growth rate at the national level multiplied by their shares at each CBSA with large hubs. This instrument captures the changes in the air traffic growth that are driven by overall changes in the airline, instead of local changes that can influence the air traffic.

Formally, the instrument for air traffic growth at CBSA i with large hubs at time t is:

$$\widetilde{air}_{it} = \sum_{a \in A} air_{a_{-i}t} \times s_{ia93} \tag{10}$$

where a represents the operation airline, A is the set of all airlines. s_{ia93} is the share of airline a at CBSA i with large hubs in the initial year 1993, where $s_{ia93} = \frac{Air_{ia93}}{\sum_{a \in A} Air_{ia93}}$. Air_{ia93} is the air traffic (boarding and departing passengers) of airline a at CBSA i in the initial year 1993. $air_{a_{-i}t}$ represents air traffic growth of airline a nationwide at time t, excluding air traffic originating or terminating at CBSA i. Therefore, the overall growth rate $air_{a_{-i}t}$ of each airline is calculated separately for each CBSA with large hubs. This helps to prevent the air traffic at CBSA i from contaminating the overall airline growth and thus ensures the stratification of the exclusion condition. For airline mergers during this period, I treat the merge and acquisition airlines as one airline across the years.[7]

Thus, the main empirical estimation equations at CBSA i using instruments are:

$$\mathrm{l_{it}} = \alpha_0 + \alpha_1 \widehat{air}_{\{it\}} + \alpha_2 log(L_{it}) + \alpha_3 log(Air_{it}) + \varepsilon_t + \varepsilon_i + \varepsilon_{it} \quad (11)$$

where $\widehat{air}_{\{it\}}$ is estimated from the first-stage regression:

$$air_{it} = \beta_0 + \beta_1 \widehat{\widetilde{air}}_{it} + \beta_2 log(L_{it}) + \beta_3 log(Air_{it}) + \epsilon_t + \epsilon_i + \epsilon_{it} \quad (12)$$

5 Main Results

In this section, I present the effect of air traffic at CBSAs with small airports on local employment as well as the effect when extending the distance restriction between small airports and large hubs. Next, I report the effects of air traffic on alternative economic measures. Then, I examine the effects of air traffic on local economic outcomes by industry. Finally, I investigate the effect of air traffic at CBSAs with large hubs on local economic outcomes.

5.1 Local Employment

Table 2 reports the effect of air traffic at CBSAs with small airports on local employment using the ordinary least squares (OLS) estimation of Equation (6). Columns 1 to 4 show the OLS results when adding the year and CBSA fixed effects. The elasticity of air traffic on local employment decreases with the addition of the year and CBSA fixed effects. Column (4) of the table presents a statistically significant positive OLS estimation of 0.003 with both year and CBSA fixed effects. It implies that a 10% increase in air traffic growth is correlated with a 0.03% increase in the local employment growth in the CBSAs with small airports.

It is not surprising that the OLS estimate is statistically positive. Higher employment means more potential travelers, thus a higher demand for air travel. We would expect the demand for air travel to have a positive effect on air traffic growth. Moreover, the OLS estimate is likely to be a

[7] There are 6 main airline mergers from 1994 to 2016: US Airways acquires America West Airlines, Delta Airlines acquires Northwest Airlines, Continental Airlines acquires Continental Micronesia Airlines, United Airlines acquires Continental Airlines, Southwest Airlines acquires AirTran Airlines, and American Airlines acquires US Airways.

biased estimator because unobservable factors in the residual that affect the local employment may also affect the air traffic in the local areas. For example, higher employment growth area tends to have a higher demand for air traffic and thus has more motivation to lobby to have more government spending in the fundamental infrastructure, which in turn improves the air traffic in the areas. Therefore, it is not possible to refer a causal effect of air traffic growth on employment from the OLS estimation.

In order to measure the causal effect of air traffic on local employment, I apply a Bartik shift-share to construct instruments to deal with the potential interdependence. Table 3 shows IV results from Column 1 to Column 4 by adding year and CBSA fixed effects. Both the first and second stage results decrease when adding year and CBSA fixed effects. Column (4) presents the estimates of the IV model with both year and CBSA fixed effects. The first-stage coefficient on the Bartik instrument on the connected hubs is about 0.5, statistically significant at a 1% level. It implies that an exogenous 10% change in the air traffic of connected hubs leads to a 5% increase in the air traffic at the CBSAs with small airports on average. Also, the first-stage is strong, with an F-statistic of 14.855 for the instrument, which rules out concerns related to weak instrument bias.

Turning to the second-stage results, the IV estimate of 0.064 is significant at the 5 percent level. It implies that a percentage point increase in the air traffic growth at CBSAs with small airports leads to a significant increase in the overall employment growth of 0.064 percentage points. Namely, in a typical CBSA with 1,000,000 residents and a 50% employment rate, a 10% increase in the air traffic will create 3,200 new jobs.

The magnitude of the IV coefficients is larger than the one of the OLS coefficients. It suggests a negatively-biased OLS estimate which is opposite to what we would expect as higher employment leads to an increase in air traffic growth. There are two potential explanations. First, local amenities which positively correlate with the local employment are negatively affected by air traffic growth. Infrastructure produces noise and temporarily causes inconvenience. Second, there exists a reserve causality that employment negatively affects air traffic growth. For example, in response to negative shocks to employment, investments in airport infrastructure may be made as a way to stimulate the local economy.

Furthermore, proximity to the connected hubs potentially violates the validation of the instrument since economic outcomes in the large hubs may spillover to the local areas. Figure 7 presents the nonparametric density of distance between small airports and their connected hubs.

We can see that despite the baseline restriction on at least 100 miles, the distance most likely lies between 100 and 500 miles, with a peak of around 200 miles. There are few samples with a distance above 10000 miles. In Table 4, I test the IV results by extending the distance restriction between small airports and large hubs from at least 100 miles to at least 200, 300, 400, and 500 miles. Column (1) replicates the results in Column (4) of Table 3. Column (2) presents the IV results when the distance between small airports and large hubs is at least 200 miles. The coefficient of the results is 0.056, statistically significant at the 10% level, which is close to the coefficient of 0.064 in Column (1). While a 100-miles distance between small airports and large hubs may still cast concerns of potential spillover effects from the large hubs, a 200-miles distance of around 4-hours driving is a more convincing and valid restriction to dispute the concerns. Longer distance restriction suffers from a much smaller sample size which causes the results insignificant.

5.2 Alternative Economic Measures

Next, I move to examine the effects of air traffic on alternative economic measures, including numbers of establishments, population, personal income per capita, total income, aggregate payroll, and wage. I replace the dependent variable with these measures in Equations (7) and (8). Each dependent variable is in log difference, control variables of total passengers and related economic outcomes are in log level. Table 5 reports the effects of air traffic on these alternative economic outcomes. The first-stage results reveal that on average, a percentage point increase in the air traffic of connected hubs leads to a significant increase in the air traffic at the CBSAs with small airports by about 0.5 percentage points. The instrument is not weak with an F-statistic around 15. Overall, coefficients of IV estimation are larger than the ones of OLS estimation, implying an intercorrelation between air traffic and local economic outcomes.

Column (2) reveals that a 10% increase in air traffic contributes to a 0.17 percentage point significant increase in the population growth at the CBSA level, while the OLS estimation shows only a 0.01 percentage point significant increase. In Column (6), I investigate the effects of air traffic on aggregate payroll at the CBSA level. It shows that air traffic has significantly positive effects on the aggregate payroll, and a 10% increase in air traffic growth leads to a 0.64% significant increase in aggregate payroll. The magnitude of the effect of air traffic on aggregate payroll is similar to the one on employment but much larger than the one on population. These results confirm that air traffic affects the labor market by increasing the labor demand, result in higher employment numbers and employment rates, but with much less effect on the overall population.

Column (7) reveals no significant effect of air traffic on wages in either the OLS or IV estimation. There are two explanations for the significant effects on aggregate payroll and employment and insignificant effects on wages. First, increased air traffic causes labor demand to increase, but labor supply is so elastic that changes in wage is omittable. Another explanation is that increased air traffic causes both labor demand and labor supply to increase since an increase in the air traffic serves as an improvement in the living amenity which makes the corresponding areas more desirable to live.

There are no significant effects of air traffic on personal income per capita and total income in either the OLS and IV estimation. Besides, in Column (1), both OLS and IV estimations reveal the insignificant effect of air traffic on the numbers of establishments. Though insignificant, the coefficient of IV estimation in Column (1) is 0.017 with t-statistics around 0.012, which is close to being significant. One possible explanation is that only specific industrial sectors respond to changes in air traffic growth.

5.3 Industry-level Results

Literature about the road infrastructure effects shows robust evidence that retail and wholesale industries benefit the most from improved market access and lower transportation costs (Michaels, 2008; Donaldson, D., and Hornbeck, R., 2016; Jaworski and Kitchens, 2016). Similarly, studies on air traffic have shown that service sectors have been the most responsive to changes in air services (Brueckner, 2003; Sheard, 2019). In this section, I exploit the effect of air traffic growth on employment, numbers of establishments and aggregate payroll by industry sectors, classified by NAICS code[8].

Table 6 reports the effects of air traffic on local employment growth by industry sector using OLS and IV estimation. The first-stage regression shows that one percentage point increase in the instrument significantly leads to about 0.5 percentage points increase in air traffic. The instrument is not weak since most F-statistic for the instrument is above 10. Comparing the OLS and IV results, we can see that one percentage point increase in air traffic at CBSAs with small airports

[8] From 1993 to 2016, industrial classification code changes from SIC to NAICS. Details about incorporating two-digit SIC and NAICS codes are attached in Appendix Table A1.

significantly leads to 0.717 percentage points decrease in the employment growth of the agriculture sector, 1.425 percentage points increase in the mining sector, 0.212 percentage points increase in the construction sector, 0.199 percentage points increase in manufacturing sector and 0.124 percentage points increase in the finance, insurance, and real estate sectors. These magnitudes of coefficients are much larger than the one of the overall employment growth, implying that industrial sectors in mining, construction, manufacturing and finance, insurance and real estate benefit the most from the increased air traffic growth. These results are consistent with findings in Brueckner (2003) and Sheard (2019)[9].

Table 7 reports the effects of air traffic growth on numbers of establishment growth by industry sector using OLS and IV estimation. The first-stage regression reveals that the instrument is strong and one percentage point increase in the instrument leads to about 0.5 percentage points increase in the air traffic growth, statistically significant at 1% level. While the IV results in the numbers of establishments overall are insignificant, results in Table 7 reveal that air traffic changes have significant effects on the numbers of establishments in agriculture, wholesale and finance, insurance and real estate sectors. One percentage point increase in air traffic growth contributes to 0.435 percentage points decrease in the numbers of establishments in the agriculture sector, 0.047 increase in the wholesale sector and 0.062 increase in the finance, insurance, and real estate sector.

Table 8 shows the effects of air traffic growth on aggregate payroll growth by industry sector using OLS and IV estimation. The coefficients of the instruments in the first-stage regression is around 0.5, significant at 1% level, and strong with an F-statistic around 15 for all industrial sectors except for the mining sector. IV results reveal that air traffic changes have significant effects on the aggregate payroll in agriculture, construction, manufacturing, retail and finance, insurance and real estate sectors. A one percentage point increase in air traffic growth leads to 1.336 percentage points decrease in the aggregate payroll growth of the agriculture sector, 0.213 percentage points increase in the construction sector, 0.425 percentage points increase in the manufacturing sector, 0.067 percentage point increase in the retail sector and 0.162 percentage points increase in the finance, insurance, and real estate sector. The results imply that the construction and manufacturing benefit the most from the increased air traffic growth, which is similar to the effects of air traffic on employment by industry.

[9] In both papers, they aggregate finance, insurance and real estate sector and service sector together as the service sector. Also, most literature tends to combine wholesale and retail sectors together.

5.4 Effects on Large Hubs

Table 9 reports the effects of air traffic growth at CBSAs with large hubs on different economic measures, including the numbers of establishments, population, personal income per capita, employment, total income, aggregate payroll, and wage. The first-stage regression reveals that on average, a percentage point increase in the instrument leads to around 0.1 percentage points increase in the air traffic growth. The instrument is a little weak with an F-statistic around 7. From the OLS results, it shows that air traffic growth has significantly positive effects on almost all the economic measures, except for personal income per capita. IV results show no significant effects of air traffic growth on these economic outcomes. One possible explanation is that only specific industrial sectors are responsive to air traffic growth at CBSAs with large hubs.

Table 10 presents the effects of air traffic on employment growth by industry at CBSAs with large hubs. The first-stage regression reveals that a one percentage point increase in the instrument leads to around 0.1 percentage points increase in air traffic growth. However, instruments are a little weak with an F-statistic around 7. Coefficients in IV regressions are larger than the ones in OLS regressions in general, confirming that OLS estimators are biased toward zero. Only in finance, insurance, and real estate sectors, the IV estimate is significantly positive at the 10% level. It reveals that a one percentage point increase in air traffic growth leads to a 0.504 percentage-points increase in employment growth.

There are no significant results for air traffic growth on numbers of establishments and aggregate payroll at CBSAs with large hubs in IV estimation. The first-stage results still reveal the existence of the weak instrument problem. The results are attached in Appendix Table A2 and A3.

6 Robustness Checks

In this section, I present the results of a series of robustness checks that I conduct based on the baseline IV estimates of the effect of air traffic growth on local economic outcomes. First, I test whether slight modifications to the sample restriction will drastically change the magnitude or significance of the critical estimates. Second, I conduct different methods to obtain the hub list and select small airports by other definitions to address concerns about estimation validity. Namely, I collect the hub list manually from the Wikipedia website and also combine the hub list with the ones from K-Means clustering. In the baseline estimation, I use all non-hub airports except for the ones in the hub list as small airports, I test the results using small airports defined by FAA airport category. I also tried to conduct the novel robustness checks proposed recently by Goldsmith-Pinkham, Sorkin, and Swift (2018), but the instrument construction is different from the standard Bartik shift-share, thus until now I have not gotten non-missing Rotemberg weights.

6.1 Sensitivity of Results to Sample Restrictions

Table 11 presents the effects of air traffic growth on all economic measures for CBSAs with small airports only. Namely, CBSAs with small airports only have no large hubs that are within 100 miles from any counties in the CBSAs. This restriction ensures that the air traffic in the CBSAs is exclusively from small airports. The first-stage regressions reveal that instruments are strong with an F-statistic around 12. Estimates of the effects of air traffic on these economics outcomes remain insignificant as they are in the baseline estimation: numbers of establishments, population, personal income per capita, total income, and wage. Column (2) reports the results of the effects of air traffic on population growth. The IV estimate reveals that a one percentage point increase in the air traffic leads to a 0.015 percentage-points increase in population growth at CBSAs with small airports only, significant at a 10% level. The results keep significant as in the baseline estimation, and the estimate of air traffic on population growth at CBSAs with small airports only decreases slightly. Column (4) presents the effects of air traffic on employment at CBSAs with small airports only. The results remain significant as in the baseline model and decrease in magnitude slightly. For both population and employment, both the significance and the magnitude of the estimates remain almost unaltered. The exception is the aggregate payroll. Column (6) shows the estimate of air traffic on aggregate payroll at CBSAs with small airports only is insignificant and the magnitude decreases by 0.009 percentage points.

6.2 Validity of Results to Sample Selections

Table 12 reports the effects of air traffic growth on employment using an alternative definition of a small airport and hub list obtained from manual collection. Column (1) shows the results when using the small airports defined by the FAA airport category. I keep all the non-hub airports and merge them with the small airport list which has at least 0.05% but less than 0.25% of overall boarding passengers nationwide. The sample size is slightly smaller, but the significance remains unchanged and the and magnitude of the estimate increases slightly from the baseline estimation. Column (2) further restricts the sample on CBSAs with small airports only, as discussed in the section above. The magnitude of the estimate decreases a little bit but it is still significant at a 10% level. Column (3) represents the results of using the hub list collected manually from the Wikipedia website as well as the small airport list defined by the FAA. The sample size is larger than the one in Column (1), but still much smaller than the baseline model. The estimate is 0.068, significant at a 5% level. The magnitude increases by 0.004 percentage points, which is omittable, from the baseline estimation. Column (4) further restricts CBSAs with small airports only. Comparative to Column (2) and baseline estimation, the magnitude overall stays mostly unaltered. Column (5) combines the hub list from K-Means clustering and manual collection from Wikipedia. From the results, the sample size stays the same as the one in Column (3). The significance and magnitude of the estimate stay almost the same.

Overall, estimates of the effects of air traffic growth on population and population stay almost unaltered in the significance and magnitude under various sample restrictions and selection methods.

7 Conclusion

This paper studies the effects of air traffic growth on local economic outcomes in the US metropolitan areas between 1993 and 2016. In order to deal with the interdependence between air traffic and local economic outcomes, I apply Bartik shift-share to construct an instrument for the air traffic for small and large airports respectively. For small airports, I use the air traffic growth of connected hubs to construct the instrument based on the hub-and-spoke operation system, where small airports connect other airports through transferring stops at large hubs. For large hubs, the instruments are constructed using passenger growth of each operation airline.

The main findings show that increased air traffic growth at CBSAs with small airports has a significantly positive effect on local employment growth. The results reveal that a one percentage point increase in air traffic growth leads to 0.064 percentage points increase in the local employment growth. Namely, a 10% increase in air traffic growth creates 3,200 new jobs in a typical city. Alternative economic outcomes are estimated to further understand the air traffic effects. A comparison of all the estimates suggests that increased air traffic affects local economic outcomes mainly through labor demand. Moreover, mining, construction, manufacturing, and financial service sectors benefit the most from air traffic growth. The effects of air traffic growth on CBSAs with large hubs are insignificant. These results inform policy discussions about the distribution of government spendings.

8 Tables

Table 1: Summary Statistics for Main Variables

CBSA Characteristics with Small Airports					
	N	Mean	St.Dev	Min	Max
Population	5651	601592.4	1620000	13139	2.02e+07
Number of employees	5651	231745.3	647749	3831	8286695
Number of firms	5651	15013.99	43706.15	401	576580
Mean Wage ($thousand)	5651	16.55	3.06	9.6	47.72
Personal income per capita ($thousand)	5651	16.89	4.18	7.62	63.15
Total wage ($million)	5651	4763.79	16872.37	44.89	244945.4
Total personal income ($million)	5651	11455.74	37207.4	255.64	566925.4
Total passengers	5538	1390000	3130000	21425	2.95e+07
Numbers of connected hubs	5651	12.76	11.79	1	38

CBSA Characteristics with Large Hubs					
	N	Mean	St.Dev	Min	Max
Population	744	4000000	3830000	281288	2.02e+07
Number of employees	744	1640000	1500000	129919	8286695
Number of firms	744	101042.4	102447.4	9125	576580
Mean Wage ($thousand)	744	21.06	3.12	15.42	36.03
Personal income per capita ($thousand)	744	19.98	3.16	12.99	36.17
Total wage ($million)	744	37365.31	40685.19	2145.8	244945.4
Total personal income ($million)	744	84788.93	92409.94	5197.34	566925.4
Total passengers	744	2.80e+07	1.96e+07	3251108	9.06e+07
Numbers of operation airlines	744	23.74	8.45	6	46

Note: There are 243 CBSAs with small airports and 31 CBSAs holding with large hubs. Dataset for CBSAs with small airports is not balanced, thus total passengers are based on the mean of total passengers across years in the CBSA level, and numbers of connected hubs are based on numbers of connected hubs in the airport level.

Table 2: OLS results of air traffic on employment for CBSAs with small airports

	(1)	(2)	(3)	(4)
Passenger growth rate in CBSA	0.005***	0.004***	0.005***	0.003**
	(0.002)	(0.001)	(0.001)	(0.001)
Total passengers	0.001**	0.003***	0.001**	0.002
	(0.001)	(0.001)	(0.001)	(0.001)
Employment	-0.002**	-0.115***	-0.002**	-0.090***
	(0.001)	(0.008)	(0.001)	(0.010)
Obs.	5253	5253	5253	5253
R-squared	0.005	0.106	0.259	0.312
Year FE	No	No	Yes	Yes
CBSA FE	No	Yes	No	Yes

Note: each regression contains 243 CBSAs with 269 small airports in total. Standard errors are in parenthesis, clustered by CBSA. ***, **, * denote significant level at 1%, 5%, 10% respectively.

Table 3: IV results of air traffic on employment for CBSAs with small airports

	(1)	(2)	(3)	(4)
Passenger growth rate in CBSA	0.234***	0.172***	0.077***	0.064**
	(0.028)	(0.026)	(0.029)	(0.027)
Total passengers	0.012***	0.043***	0.005***	0.016**
	(0.002)	(0.008)	(0.001)	(0.006)
Employment	-0.012***	-0.167***	-0.005***	-0.109***
	(0.003)	(0.015)	(0.001)	(0.013)
Obs.	4676	4676	4676	4676
R-squared	-4.800	-2.295	-0.187	0.030
Year FE	No	No	Yes	Yes
CBSA FE	No	Yes	No	Yes
First-stage regression:				
Instrument for passenger growth rate	0.669***	0.657***	0.539***	0.500***
	(0.087)	(0.098)	(0.128)	(0.130)
Obs.	4676	4676	4676	4676
R-squared	0.037	0.137	0.068	0.166
F test for instrument	59.080	45.092	17.635	14.855

Note: each regression contains 243 CBSAs with 269 small airports in total. Standard errors are in parenthesis, clustered by CBSA. ***, **, * denote significant level at 1%, 5%, 10% respectively.

Table 4: IV results of air traffic on employment by distance restriction

	(1) $\geq 100\ miles$	(2) $\geq 200\ miles$	(3) $\geq 300\ miles$	(4) $\geq 400\ miles$	(5) $\geq 500\ miles$
Passenger growth rate in CBSA	0.064**	0.056*	0.065	-0.052	2.030
	(0.027)	(0.033)	(0.114)	(0.147)	(81.151)
Total passengers	0.016**	0.013*	0.016	-0.010	0.442
	(0.006)	(0.007)	(0.025)	(0.033)	(17.597)
Employment	-0.109***	-0.103***	-0.107***	-0.087***	-0.342
	(0.013)	(0.012)	(0.021)	(0.024)	(9.828)
Obs.	4676	4322	3841	3619	3277
R-squared	0.030	0.175	0.151	0.184	-208.391
Year FE	Yes	Yes	Yes	Yes	Yes
CBSA FE	Yes	Yes	Yes	Yes	Yes
First-stage regression:					
Instrument	0.500***	0.418***	0.136	0.121	-0.003
	(0.130)	(0.122)	(0.113)	(0.119)	(0.111)
Obs.	4676	4322	3841	3619	3277
R-squared	0.166	0.170	0.169	0.166	0.168
F test for instrument	14.855	11.691	1.463	1.043	0.001

Note: this table contains IV regression results by distance restriction on the connected hubs. The first column is the original IV result in Table 3, which sets the distance between small airports and their connected hubs at least 100 miles. The following columns set distance of at least 200, 300, 400, and 500 miles. Standard errors are in parenthesis, clustered by CBSA. ***, **, * denote significant level at 1%, 5%, 10% respectively.

Table 5: Air traffic on alternative economic measures

	(1)	(2)	(3)	(5)	(6)	(7)
	Number of establishments	Population	Personal income per capita	Total income	Aggregate payroll	Wage
OLS results:						
Passenger growth rate in CBSA	0.001	0.001**	-0.000	0.001	0.002	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Total passengers	0.000	0.001	-0.000	0.000	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Employment	-0.051***	-0.044***	-0.118***	-0.073***	-0.084***	-0.176***
	(0.007)	(0.010)	(0.008)	(0.006)	(0.010)	(0.018)
Obs.	5253	5253	5253	5253	5253	5253
R-squared	0.367	0.075	0.256	0.223	0.252	0.204
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
CBSA FE	Yes	Yes	Yes	Yes	Yes	Yes
IV results:						
Passenger growth rate in CBSA	0.017	0.017**	-0.012	-0.008	0.064*	0.006
	(0.012)	(0.008)	(0.017)	(0.018)	(0.033)	(0.021)
Total passengers	0.004	0.004**	-0.003	-0.002	0.015*	0.002
	(0.003)	(0.002)	(0.004)	(0.004)	(0.008)	(0.005)
Employment	-0.054***	-0.050***	-0.120***	-0.072***	-0.100***	-0.194***
	(0.008)	(0.010)	(0.010)	(0.007)	(0.011)	(0.019)
Obs.	4676	4676	4676	4676	4676	4676
R-squared	0.318	0.061	0.271	0.246	0.096	0.218
First-stage regression:						
Instrument	0.522***	0.525***	0.493***	0.499***	0.505***	0.520***
	(0.131)	(0.132)	(0.131)	(0.130)	(0.130)	(0.131)
Obs.	4676	4676	4676	4676	4676	4676
R-squared	0.164	0.164	0.164	0.165	0.165	0.163
F test for instrument	15.798	15.912	14.136	14.644	14.979	15.823

Note: each regression contains 243 CBSAs with 269 small airports in total. Standard errors are in parenthesis, clustered by CBSA. ***, **, * denote significant level at 1%, 5%, 10% respectively.

Table 6: Air traffic on employment by industry

	(1) Transportation and utilities	(2) Agriculture	(3) Mining	(4) Construction	(5) Manufacturing	(6) Wholesale	(7) Retail	(8) Finance, insurance , and real estate	(9) Service
OLS results:									
Passenger growth rate in CBSA level	0.002	0.001	-0.021	0.008	0.004	0.001	0.004**	0.003	0.003
	(0.008)	(0.025)	(0.013)	(0.005)	(0.007)	(0.004)	(0.002)	(0.003)	(0.002)
Obs.	5253	4421	4335	5253	5253	5253	5253	5253	5253
R-squared	0.245	0.472	0.231	0.315	0.217	0.274	0.862	0.201	0.665
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CBSA FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
IV results:									
Passenger growth rate in CBSA level	-0.006	-0.717**	1.425**	0.212**	0.199**	0.063	0.035	0.124*	0.023
	(0.104)	(0.357)	(0.620)	(0.095)	(0.097)	(0.070)	(0.027)	(0.074)	(0.026)
Obs.	4676	4026	3983	4676	4676	4676	4676	4676	4676
R-squared	0.274	0.397	0.649	0.035	0.152	0.274	0.868	0.012	0.681
First-stage regression:									
Instrument	0.523***	0.612***	0.504***	0.496***	0.519***	0.514***	0.518***	0.519***	0.518***
	(0.130)	(0.141)	(0.152)	(0.132)	(0.130)	(0.130)	(0.131)	(0.130)	(0.131)
Obs.	4676	4205	4084	4676	4676	4676	4676	4676	4676
R-squared	0.163	0.158	0.170	0.164	0.163	0.164	0.166	0.164	0.167
F test for instrument	16.086	18.490	9.132	14.103	16.038	15.577	15.522	15.841	15.542

Note: each regression contains 243 CBSAs with 269 small airports in total. Standard errors are in parenthesis, clustered by CBSA. ***, **, * denote significant level at 1%, 5%, 10% respectively.

Table 7: Air traffic on numbers of establishment by industry

	(1) Transportation and utilities	(2) Agriculture	(3) Mining	(4) Construction	(5) Manufacturing	(6) Wholesale	(7) Retail	(8) Finance, insurance , and real estate	(9) Service
OLS results:									
Passenger growth rate in CBSA level	0.003	0.013	-0.005	-0.000	0.001	0.001	0.001	0.003	0.001*
	(0.003)	(0.012)	(0.006)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)
Obs.	5253	5218	5076	5253	5253	5253	5253	5253	5253
R-squared	0.452	0.692	0.163	0.417	0.257	0.535	0.899	0.364	0.867
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CBSA FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
IV results:									
Passenger growth rate in CBSA level	0.043	-0.435**	-0.045	0.056	0.006	0.047*	0.015	0.062**	-0.014
	(0.044)	(0.193)	(0.185)	(0.035)	(0.031)	(0.028)	(0.018)	(0.029)	(0.013)
Obs.	4676	4646	4544	4676	4676	4676	4676	4676	4676
R-squared	0.470	0.634	0.156	0.328	0.256	0.529	0.907	0.190	0.877
First-stage regression:									
Instrument	0.531***	0.536***	0.492***	0.504***	0.505***	0.520***	0.530***	0.523***	0.525***
	(0.131)	(0.133)	(0.129)	(0.131)	(0.128)	(0.131)	(0.132)	(0.131)	(0.131)
Obs.	4676	4653	4565	4676	4676	4676	4676	4676	4676
R-squared	0.164	0.164	0.163	0.165	0.164	0.163	0.163	0.164	0.164
F test for instrument	16.335	16.736	15.086	14.886	15.473	15.836	16.056	15.901	15.966

Note: each regression contains 243 CBSAs with 269 small airports in total. Standard errors are in parenthesis, clustered by CBSA. ***, **, * denote significant level at 1%, 5%, 10% respectively.

Table 8: Air traffic on aggregate payroll by industry

	(1) Transportation and utilities	(2) Agriculture	(3) Mining	(4) Construction	(5) Manufacturing	(6) Wholesale	(7) Retail	(8) Finance, insurance , and real estate	(9) Service
OLS results:									
Passenger growth rate in CBSA level	0.008	0.013	-0.004	0.003	0.001	-0.002	0.005**	0.007	0.006**
	(0.008)	(0.025)	(0.012)	(0.004)	(0.006)	(0.004)	(0.002)	(0.005)	(0.003)
Obs.	5228	3905	3160	5244	5178	5232	5253	5251	5253
R-squared	0.253	0.378	0.194	0.327	0.187	0.189	0.606	0.185	0.331
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CBSA FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
IV results:									
Passenger growth rate in CBSA level	0.203	-1.336***	1.600	0.213**	0.425*	0.140	0.067*	0.162*	0.028
	(0.156)	(0.445)	(1.082)	(0.083)	(0.257)	(0.093)	(0.035)	(0.093)	(0.044)
Obs.	4658	3552	2890	4673	4647	4668	4676	4674	4676
R-squared	0.190	0.135	0.924	0.004	0.355	0.066	0.556	0.096	0.341
First-stage regression:									
Instrument	0.524***	0.661***	0.425**	0.500***	0.522***	0.496***	0.518***	0.516***	0.513***
	(0.130)	(0.162)	(0.183)	(0.131)	(0.130)	(0.130)	(0.131)	(0.130)	(0.131)
Obs.	4664	3865	3172	4674	4655	4669	4676	4675	4676
R-squared	0.163	0.184	0.162	0.164	0.164	0.164	0.165	0.164	0.165
F test for instrument	16.420	16.237	3.292	14.541	15.882	14.495	15.597	15.846	15.270

Note: each regression contains 243 CBSAs with 269 small airports in total. Standard errors are in parenthesis, clustered by CBSA. ***, **, * denote significant level at 1%, 5%, 10% respectively.

Table 9: Air traffic on 7 economic measures for CBSAs with large hubs

	(1) Number of establishments	(2) Population	(3) Personal income per capita	(4) Employment	(5) Total income	(6) Aggregate payroll	(7) Wage
OLS results:							
Passenger growth rate in CBSA	0.031**	0.023**	0.014	0.052**	0.040**	0.064**	0.017*
	(0.014)	(0.010)	(0.010)	(0.022)	(0.018)	(0.029)	(0.010)
Obs.	713	713	713	713	713	713	713
R-squared	0.305	0.113	0.651	0.465	0.482	0.528	0.546
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CBSA FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
IV results:							
Passenger growth rate in CBSA	0.353	0.345	-0.052	0.375	0.317	0.301	-0.063
	(0.334)	(0.335)	(0.071)	(0.332)	(0.338)	(0.348)	(0.065)
Obs.	713	713	713	713	713	713	713
R-squared	0.171	0.557	0.633	0.177	0.281	0.424	0.496
First-stage regression:							
Instrument	0.119**	0.119**	0.122***	0.119**	0.113**	0.118**	0.123***
	(0.045)	(0.046)	(0.043)	(0.044)	(0.046)	(0.044)	(0.042)
Obs.	713	713	713	713	713	713	713
R-squared	0.350	0.345	0.338	0.342	0.354	0.346	0.345
F test for instrument	6.912	6.729	7.978	7.212	6.060	7.221	8.430

Note: each regression contains 31 CBSAs with 38 large hubs in total. Standard errors are in parenthesis, clustered by CBSA. ***, **, * denote significant level at 1%, 5%, 10% respectively.

Table 10: Air traffic on employment by industry for CBSAs with large hubs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Transportation and utilities	Agriculture	Mining	Construction	Manufacturing	Wholesale	Retail	Finance, insurance , and real estate	Service
OLS results:									
Passenger growth rate in CBSA level	0.083*	0.064	0.326**	0.086**	0.023	0.027	0.019	0.080***	0.055*
	(0.045)	(0.328)	(0.136)	(0.036)	(0.015)	(0.023)	(0.018)	(0.027)	(0.027)
Obs.	713	707	713	713	713	713	713	713	713
R-squared	0.452	0.781	0.302	0.660	0.564	0.493	0.914	0.376	0.799
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CBSA FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
IV results:									
Passenger growth rate in CBSA level	0.574	-0.323	0.460	0.375	0.475	0.543	0.235	0.504*	0.373
	(0.775)	(1.747)	(1.969)	(0.356)	(0.399)	(0.452)	(0.322)	(0.300)	(0.319)
Obs.	713	707	713	713	713	713	713	713	713
R-squared	0.344	0.780	0.302	0.619	0.363	0.215	0.899	0.146	0.718
First-stage regression:									
Instrument	0.126***	0.113**	0.125***	0.119**	0.125***	0.123**	0.121**	0.124***	0.113**
	(0.043)	(0.042)	(0.043)	(0.045)	(0.043)	(0.045)	(0.045)	(0.044)	(0.044)
Obs.	713	710	713	713	713	713	713	713	713
R-squared	0.336	0.338	0.337	0.343	0.335	0.340	0.342	0.337	0.349
F test for instrument	8.416	7.299	8.455	6.986	8.266	7.496	7.204	7.892	6.704

Note: each regression contains 31 CBSAs with 38 large hubs in total. Standard errors are in parenthesis, clustered by CBSA. ***, **, * denote significant level at 1%, 5%, 10% respectively.

Table 11: Air traffic on economic measures for CBSAs with small airports only

	(1) Number of establishments	(2) Population	(3) Personal income per capita	(4) Employment	(5) Total income	(6) Aggregate payroll	(7) Wage
OLS results:							
Passenger growth rate in CBSA	0.002***	0.002**	0.000	0.005***	0.002	0.004**	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Obs.	3946	3946	3946	3946	3946	3946	3946
R-squared	0.410	0.108	0.238	0.299	0.214	0.233	0.188
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CBSA FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
IV results:							
Passenger growth rate in CBSA	0.013	0.015*	-0.023	0.062**	-0.019	0.055	0.002
	(0.013)	(0.009)	(0.019)	(0.031)	(0.019)	(0.037)	(0.023)
Obs.	3437	3437	3437	3437	3437	3437	3437
R-squared	0.427	0.026	0.223	0.079	0.212	0.149	0.209
First-stage regression:							
Instrument	0.556***	0.556***	0.527***	0.520***	0.519***	0.524***	0.554***
	(0.151)	(0.150)	(0.151)	(0.149)	(0.150)	(0.150)	(0.150)
Obs.	3437	3437	3437	3437	3437	3437	3437
R-squared	0.160	0.160	0.159	0.164	0.162	0.163	0.158
F test for instrument	13.612	13.653	12.198	12.127	11.998	12.190	13.592

Note: each regression contains 182 CBSAs with small airports only in the surrounding areas. Standard errors are in parenthesis, clustered by CBSA. ***, **, * denote significant level at 1%, 5%, 10% respectively.

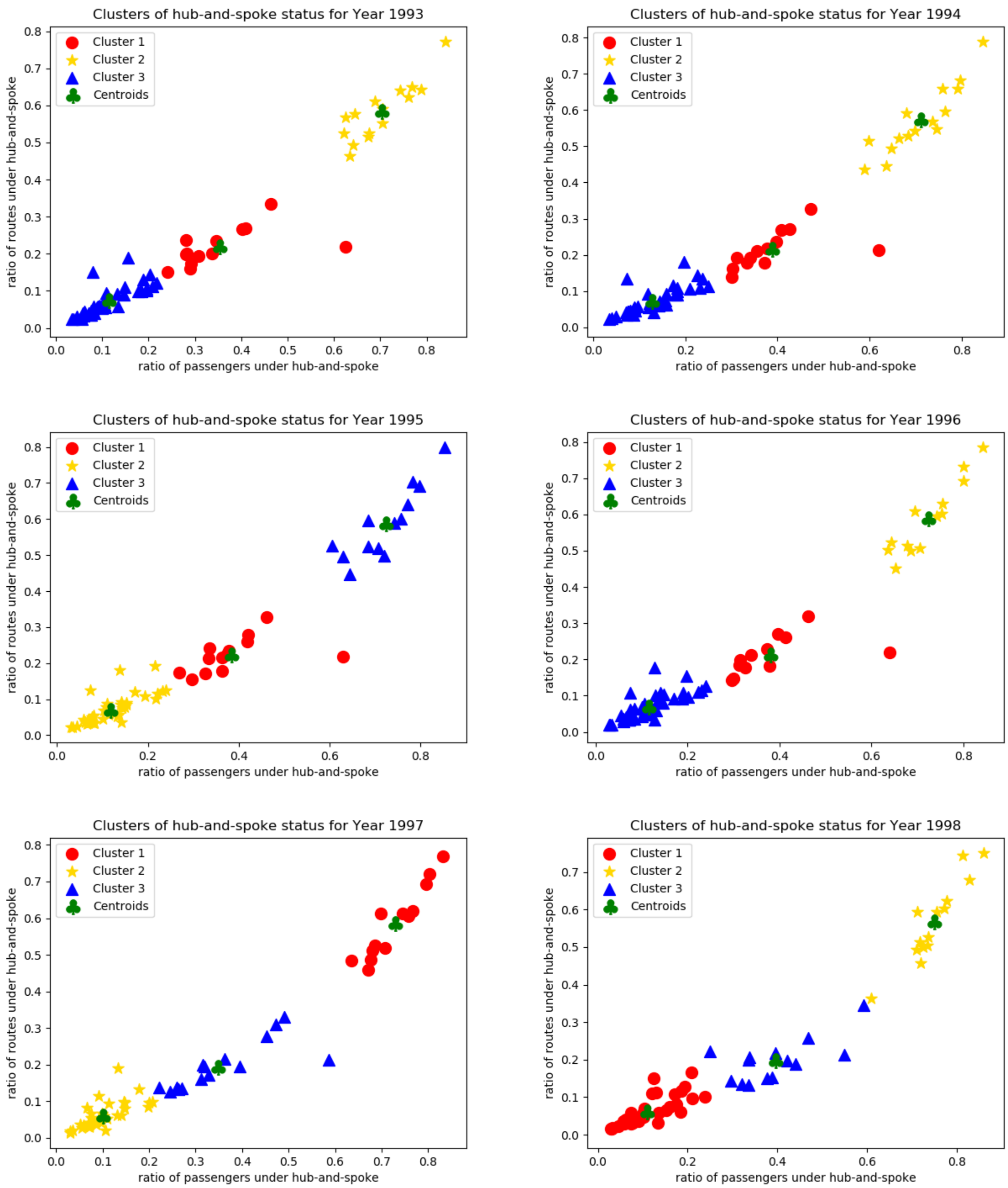
Table 12: Air traffic on employment by sample selection

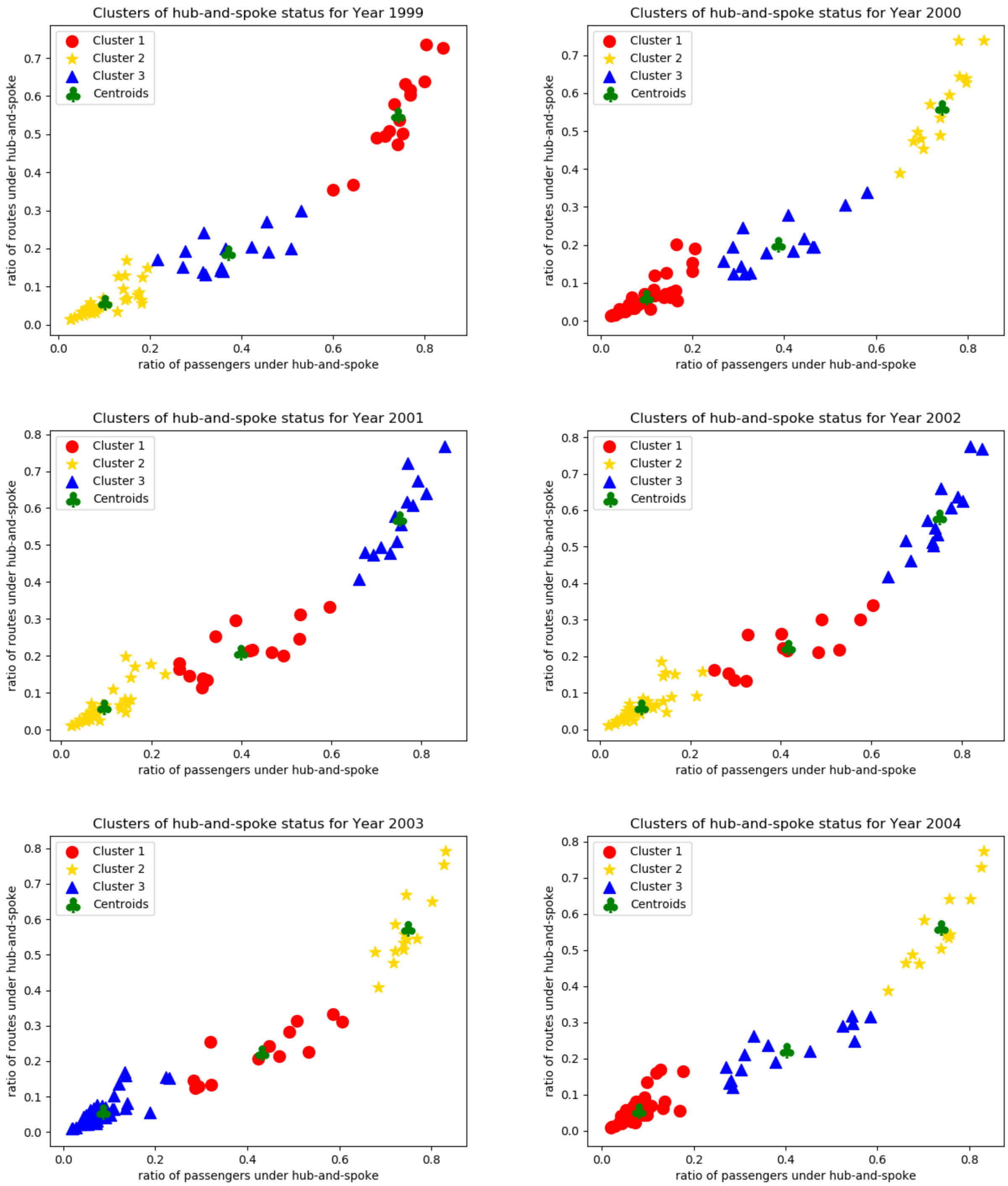
	(1)	(2)	(3)	(4)	(5)	(6)
	K-Means Small by FAA	K-Means Small by FAA Only	Wiki Small by FAA	Wiki Small by FAA Only	K-Means+Wiki Small by FAA	K-Means+Wiki Small by FAA Only
OLS results:						
Passenger growth rate in CBSA	0.003**	0.004***	0.003**	0.004***	0.003***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Obs.	4739	3534	4748	3494	4752	3429
R-squared	0.317	0.291	0.315	0.290	0.315	0.287
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
CBSA FE	Yes	Yes	Yes	Yes	Yes	Yes
IV results:						
Passenger growth rate in CBSA	0.066**	0.061*	0.068**	0.058*	0.067**	0.065*
	(0.030)	(0.032)	(0.031)	(0.032)	(0.031)	(0.034)
Obs.	4147	3023	4151	3009	4151	2940
R-squared	0.043	0.059	0.079	0.049	0.062	0.001
First-stage regression:						
Instrument for passenger growth rate	0.450***	0.515***	0.445***	0.517***	0.448***	0.491***
	(0.127)	(0.152)	(0.130)	(0.161)	(0.130)	(0.157)
Obs.	4147	3023	4151	3009	4151	2940
R-squared	0.171	0.169	0.186	0.197	0.186	0.192
F test for instrument	12.523	11.438	11.671	10.373	11.794	9.826

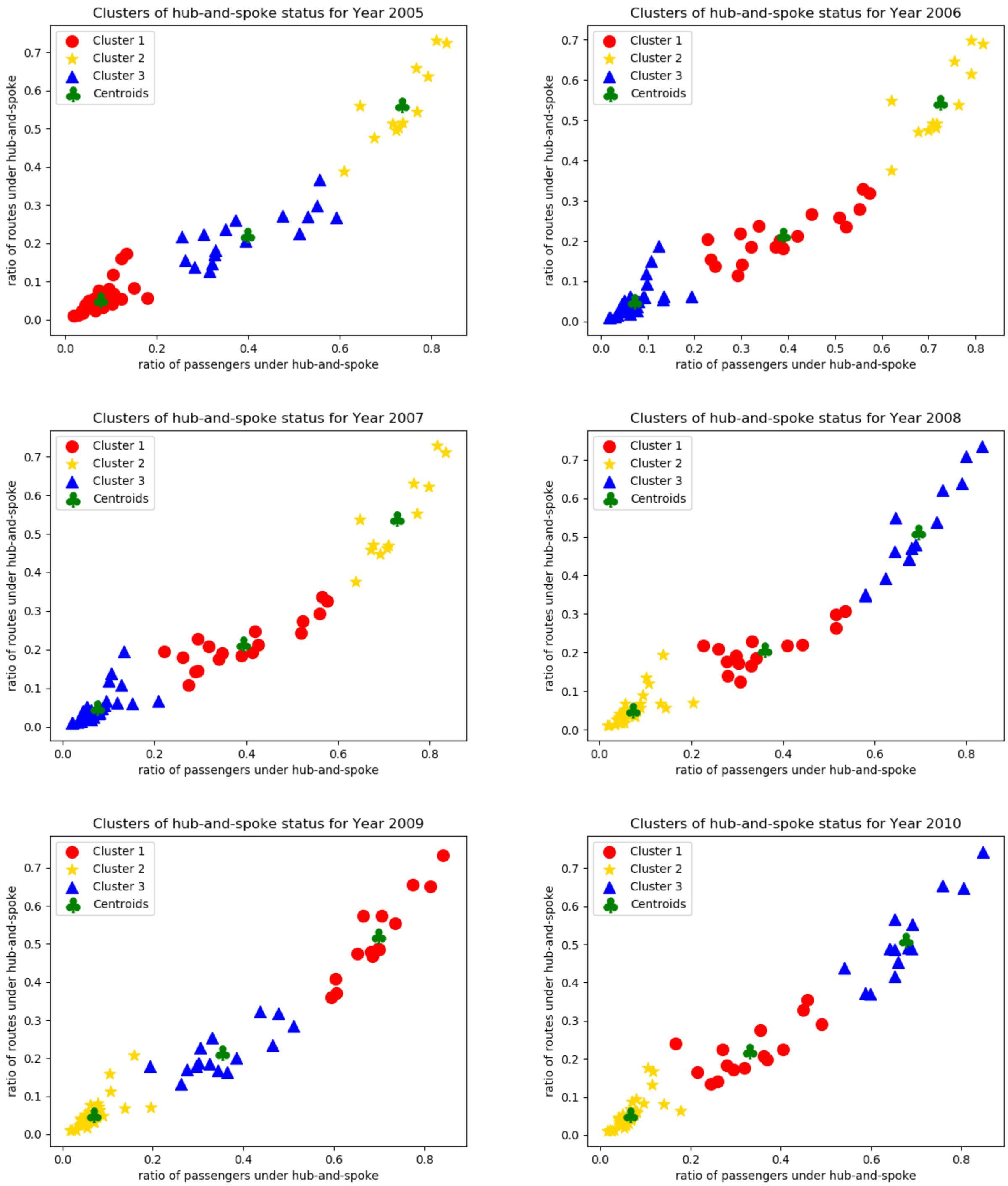
Note: This table shows results by different hub selection and small-airport definition methods. The results of the first column are based on the hub list using KMean clustering and small airport list based on the FAA airport category definition. Column 2 uses the same sample and also restricts CBSAs whose air traffic comes from small airports within 100 miles surrounding. Hus list in Column 3 is manually collected from Wiki for each main airline in the United States while the small airport list is based on the FAA definition. Column 4 also restricts on CBSAs with only small airports. The hub list in Column 5 combines the hub list from Wiki and KMean clustering method and the small airport definition is still based on the FAA. Column 6 further restricts on CBSAs with only small airports surrounding.

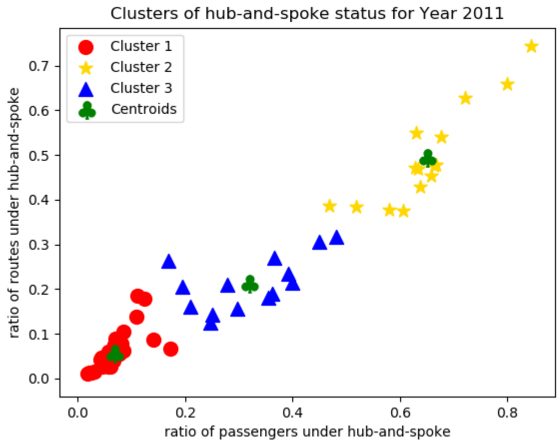
9 Figures

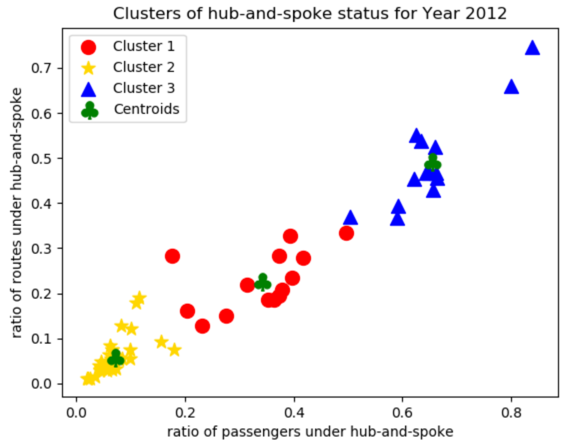
Figure 1: K-Means Clustering of Hub Results, 1993-2016

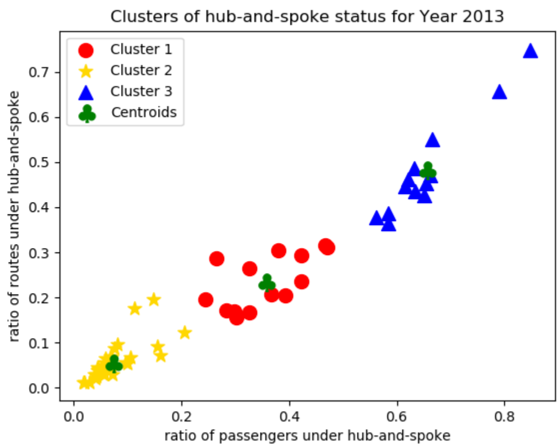


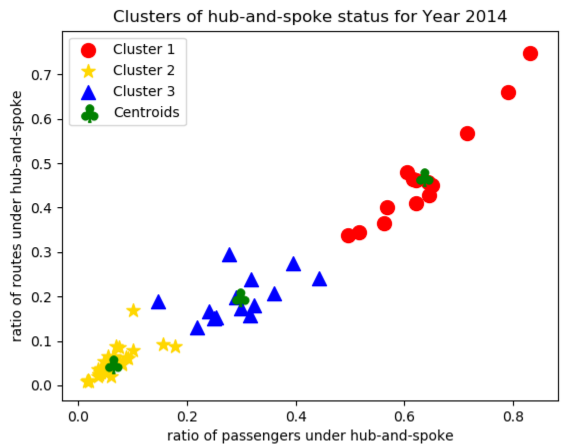

Clusters of hub-and-spoke status for Year 1999
Clusters of hub-and-spoke status for Year 2000
Clusters of hub-and-spoke status for Year 2001
Clusters of hub-and-spoke status for Year 2002
Clusters of hub-and-spoke status for Year 2003
Clusters of hub-and-spoke status for Year 2004
Cluster 1
Cluster 2
Cluster 3
Centroids
ratio of routes under hub-and-spoke
ratio of passengers under hub-and-spoke

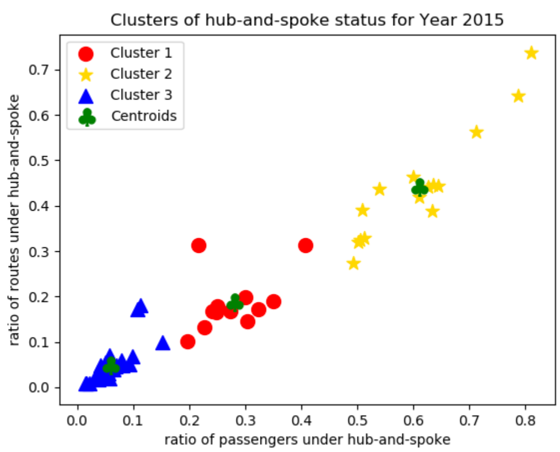

Clusters of hub-and-spoke status for Year 2005
Cluster 1
Cluster 2
Cluster 3
Centroids
ratio of routes under hub-and-spoke
ratio of passengers under hub-and-spoke
Clusters of hub-and-spoke status for Year 2006
Cluster 1
Cluster 2
Cluster 3
Centroids
ratio of routes under hub-and-spoke
ratio of passengers under hub-and-spoke
Clusters of hub-and-spoke status for Year 2007
Cluster 1
Cluster 2
Cluster 3
Centroids
ratio of routes under hub-and-spoke
ratio of passengers under hub-and-spoke
Clusters of hub-and-spoke status for Year 2008
Cluster 1
Cluster 2
Cluster 3
Centroids
ratio of routes under hub-and-spoke
ratio of passengers under hub-and-spoke
Clusters of hub-and-spoke status for Year 2009
Cluster 1
Cluster 2
Cluster 3
Centroids
ratio of routes under hub-and-spoke
ratio of passengers under hub-and-spoke
Clusters of hub-and-spoke status for Year 2010
Cluster 1
Cluster 2
Cluster 3
Centroids
ratio of routes under hub-and-spoke
ratio of passengers under hub-and-spoke


Clusters of hub-and-spoke status for Year 2011
Cluster 1
Cluster 2
Cluster 3
Centroids
ratio of routes under hub-and-spoke
ratio of passengers under hub-and-spoke


Clusters of hub-and-spoke status for Year 2012
Cluster 1
Cluster 2
Cluster 3
Centroids
ratio of routes under hub-and-spoke
ratio of passengers under hub-and-spoke


Clusters of hub-and-spoke status for Year 2013
Cluster 1
Cluster 2
Cluster 3
Centroids
ratio of routes under hub-and-spoke
ratio of passengers under hub-and-spoke


Clusters of hub-and-spoke status for Year 2014
Cluster 1
Cluster 2
Cluster 3
Centroids
ratio of routes under hub-and-spoke
ratio of passengers under hub-and-spoke


Clusters of hub-and-spoke status for Year 2015
Cluster 1
Cluster 2
Cluster 3
Centroids
ratio of routes under hub-and-spoke
ratio of passengers under hub-and-spoke

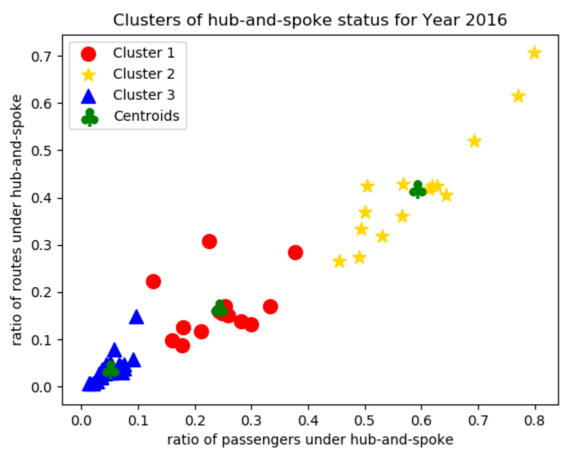
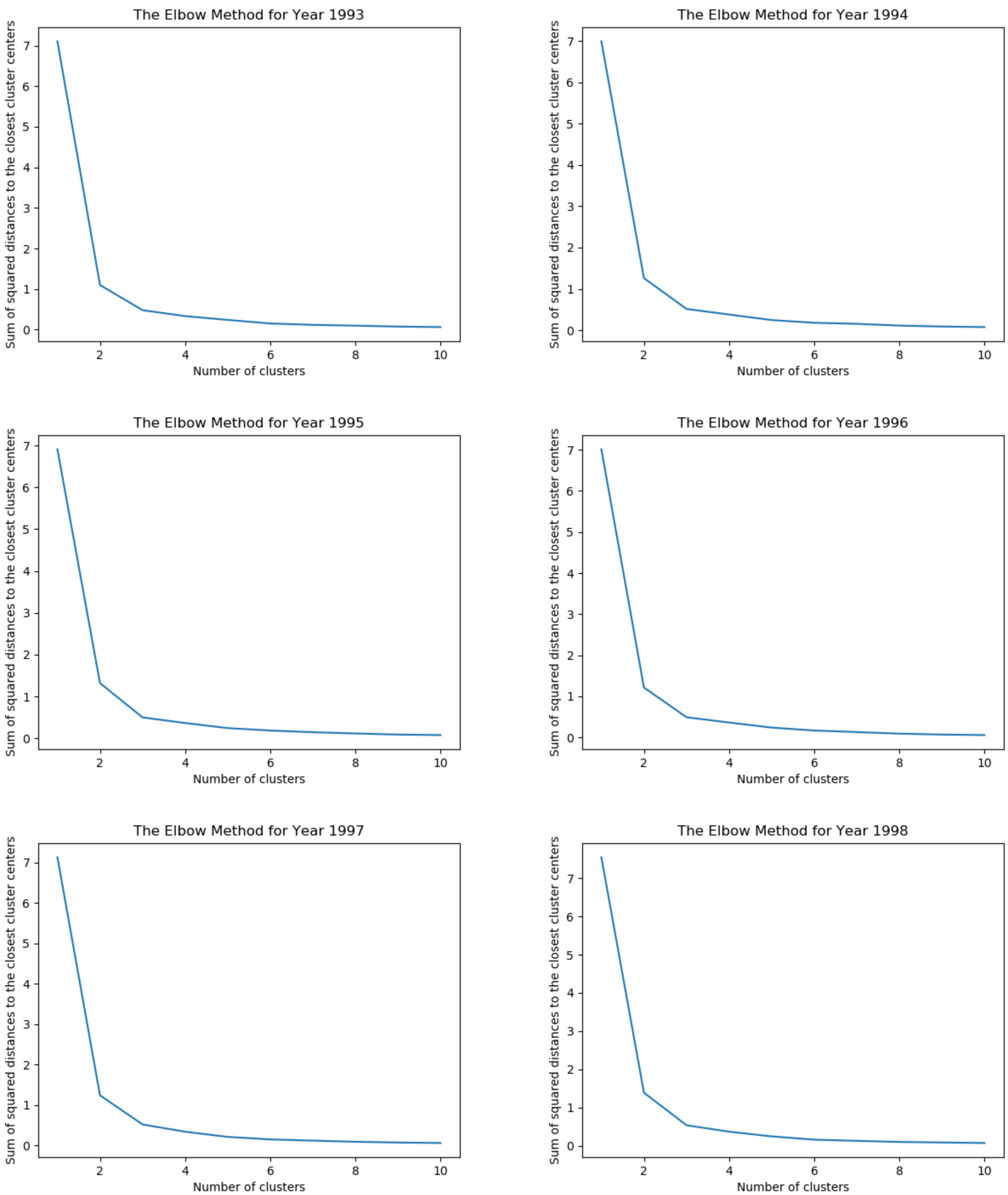
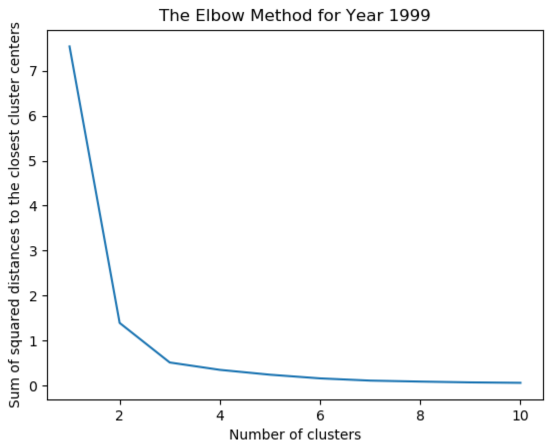
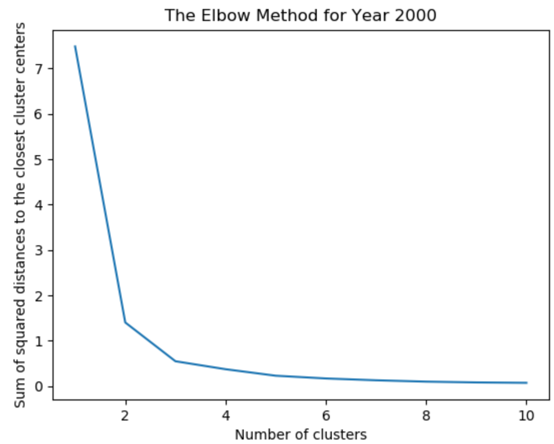
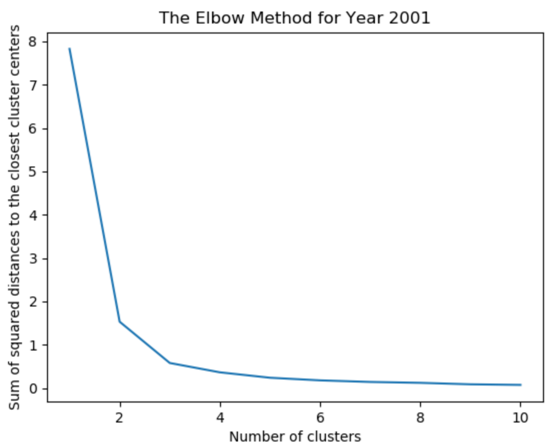

Clusters of hub-and-spoke status for Year 2016
Cluster 1
Cluster 2
Cluster 3
Centroids
ratio of routes under hub-and-spoke
ratio of passengers under hub-and-spoke

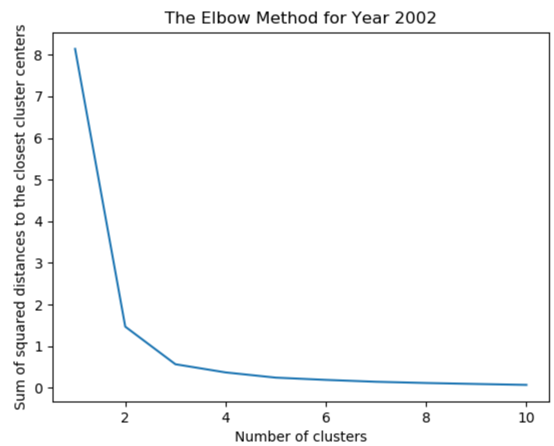
FIGURE 2: ELBOW METHOD OF KMEAN CLUSTERING, 1993-2016

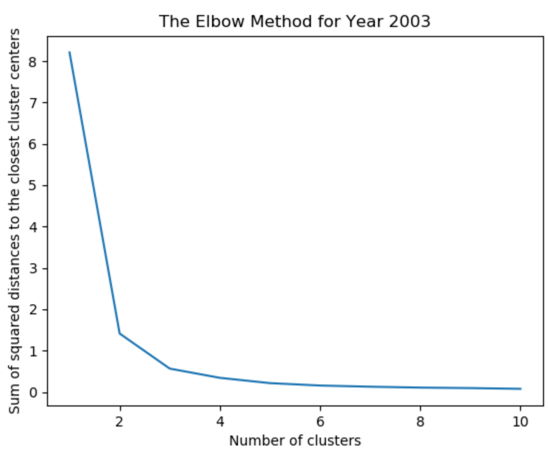


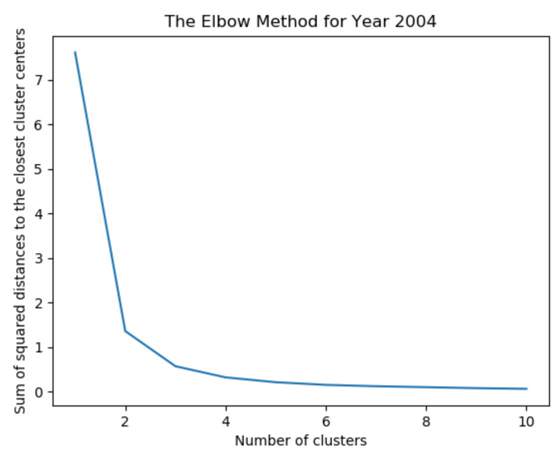

The Elbow Method for Year 1999
Sum of squared distances to the closest cluster centers
Number of clusters

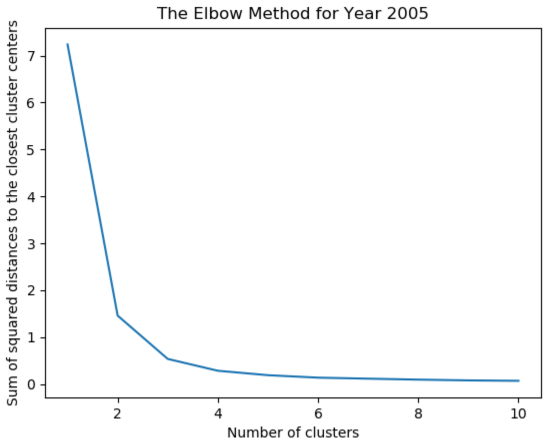

The Elbow Method for Year 2000
Sum of squared distances to the closest cluster centers
Number of clusters

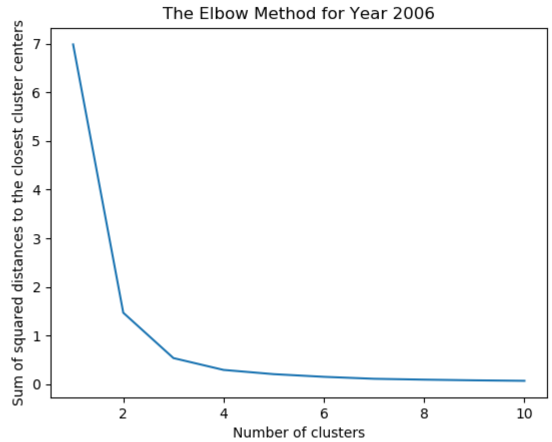

The Elbow Method for Year 2001
Sum of squared distances to the closest cluster centers
Number of clusters

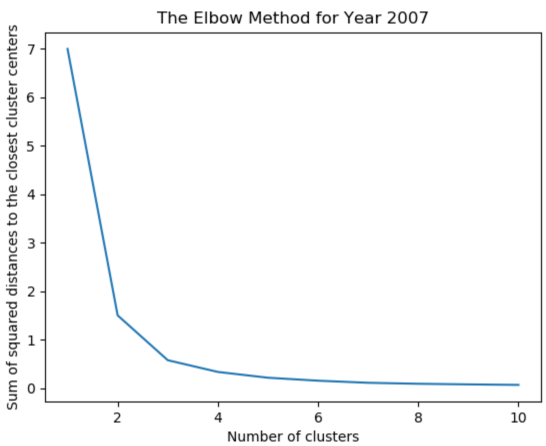

The Elbow Method for Year 2002
Sum of squared distances to the closest cluster centers
Number of clusters

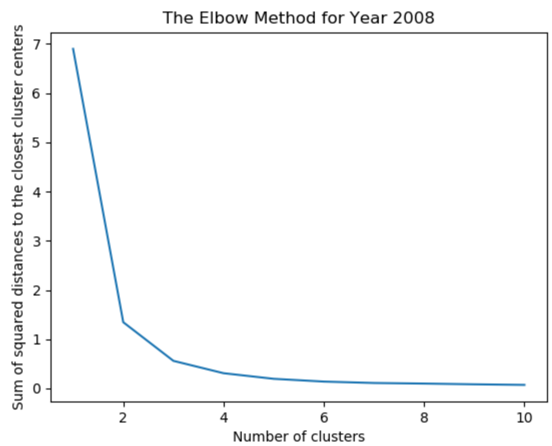

The Elbow Method for Year 2003
Sum of squared distances to the closest cluster centers
Number of clusters

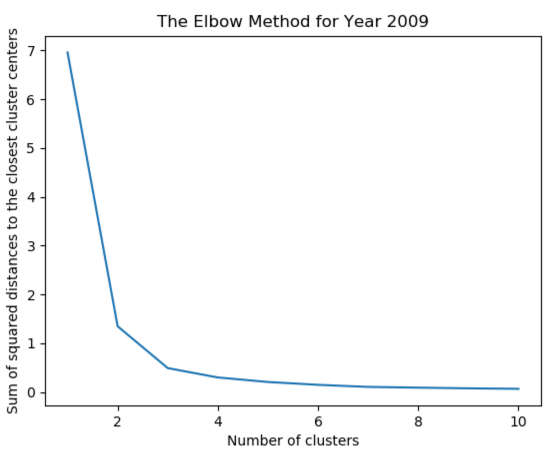

The Elbow Method for Year 2004
Sum of squared distances to the closest cluster centers
Number of clusters

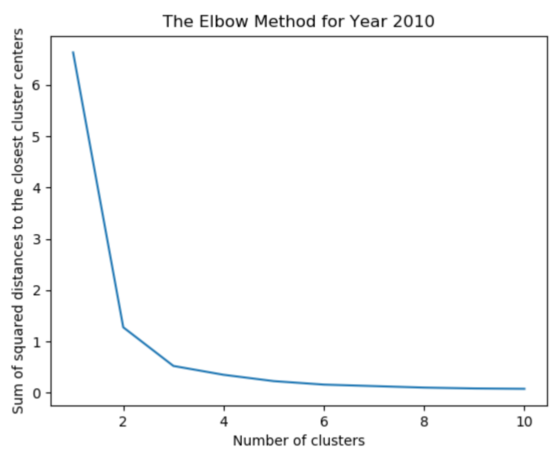

The Elbow Method for Year 2005
Sum of squared distances to the closest cluster centers
0 1 2 3 4 5 6 7
2 4 6 8 10
Number of clusters

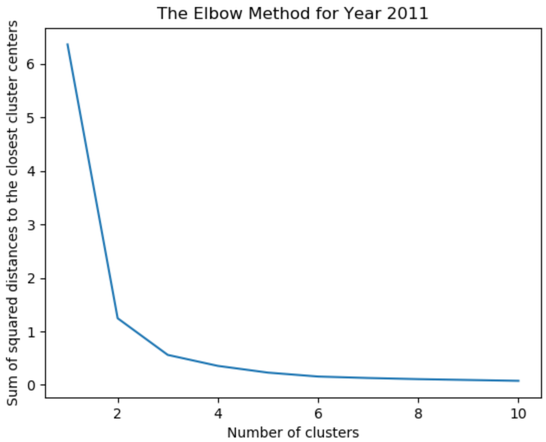

The Elbow Method for Year 2006
Sum of squared distances to the closest cluster centers
0 1 2 3 4 5 6 7
2 4 6 8 10
Number of clusters

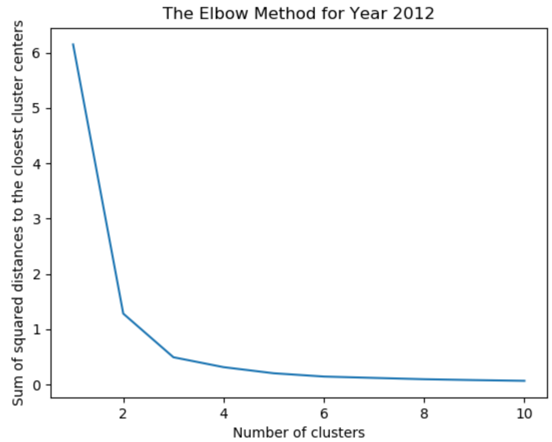

The Elbow Method for Year 2007
Sum of squared distances to the closest cluster centers
0 1 2 3 4 5 6 7
2 4 6 8 10
Number of clusters

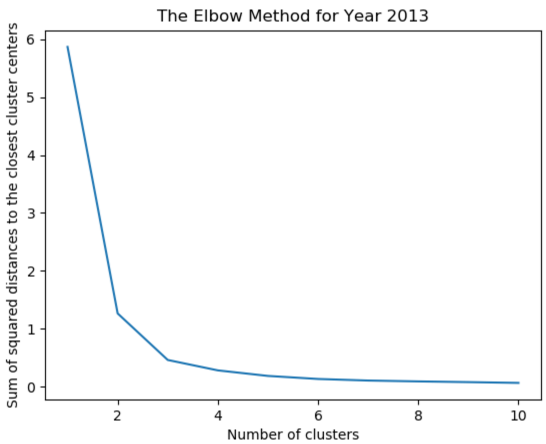

The Elbow Method for Year 2008
Sum of squared distances to the closest cluster centers
0 1 2 3 4 5 6 7
2 4 6 8 10
Number of clusters

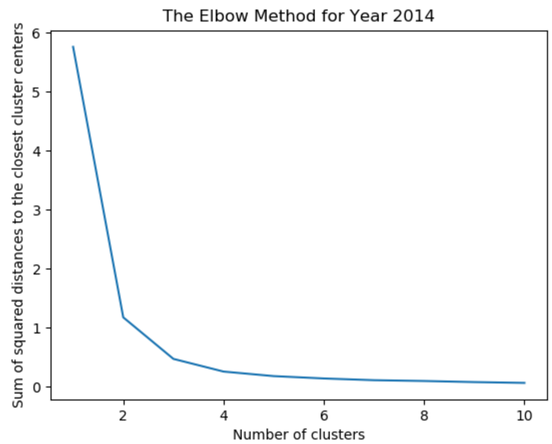

The Elbow Method for Year 2009
Sum of squared distances to the closest cluster centers
0 1 2 3 4 5 6 7
2 4 6 8 10
Number of clusters

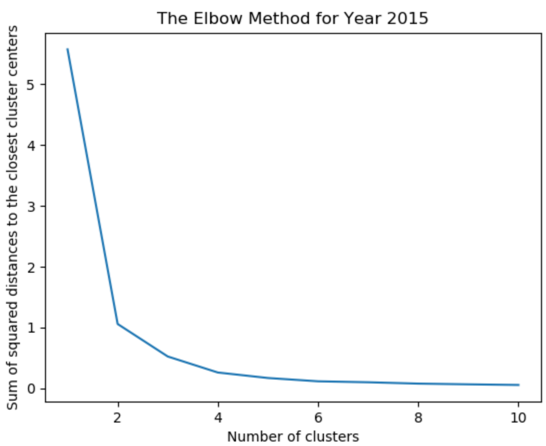

The Elbow Method for Year 2010
Sum of squared distances to the closest cluster centers
0 1 2 3 4 5 6
2 4 6 8 10
Number of clusters


The Elbow Method for Year 2011
Sum of squared distances to the closest cluster centers
0
1
2
3
4
5
6
2
4
6
8
10
Number of clusters


The Elbow Method for Year 2012
Sum of squared distances to the closest cluster centers
0
1
2
3
4
5
6
2
4
6
8
10
Number of clusters


The Elbow Method for Year 2013
Sum of squared distances to the closest cluster centers
0
1
2
3
4
5
6
2
4
6
8
10
Number of clusters


The Elbow Method for Year 2014
Sum of squared distances to the closest cluster centers
0
1
2
3
4
5
6
2
4
6
8
10
Number of clusters


The Elbow Method for Year 2015
Sum of squared distances to the closest cluster centers
0
1
2
3
4
5
2
4
6
8
10
Number of clusters

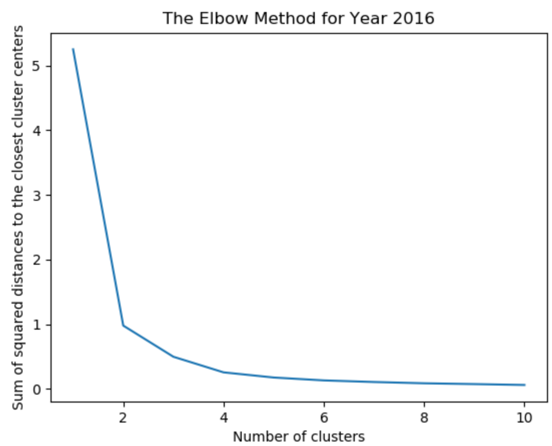

The Elbow Method for Year 2016
Sum of squared distances to the closest cluster centers
0
1
2
3
4
5
2
4
6
8
10
Number of clusters

FIGURE 3: KERNEL NONPARAMETRIC DENSITY OF PASSENGERS TRANSPORTED UNDER HUB-AND-SPOKE SYSTEM

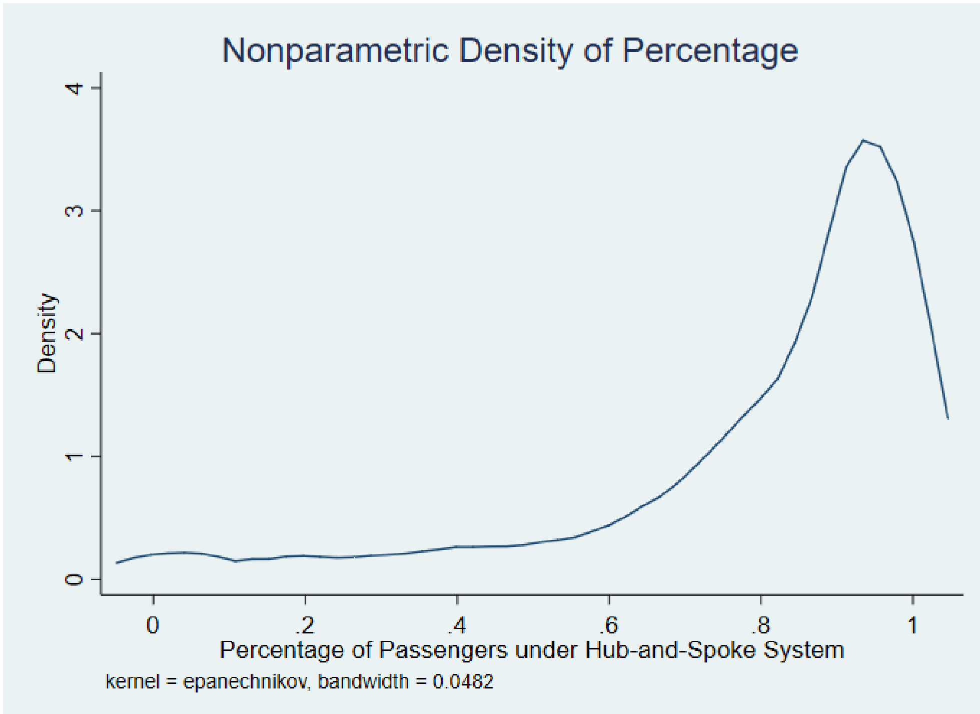
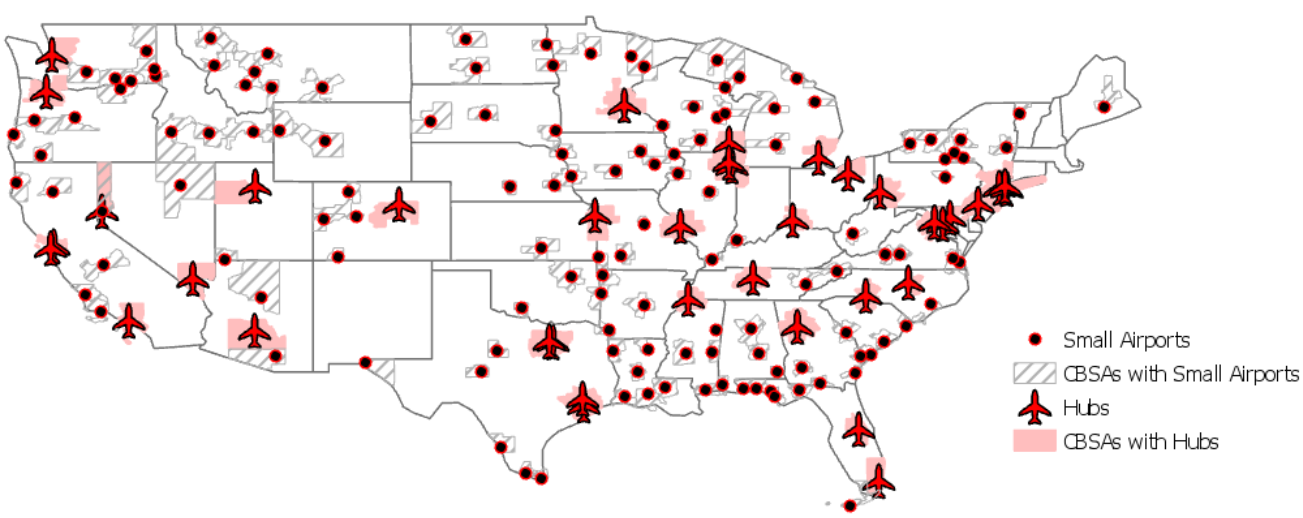
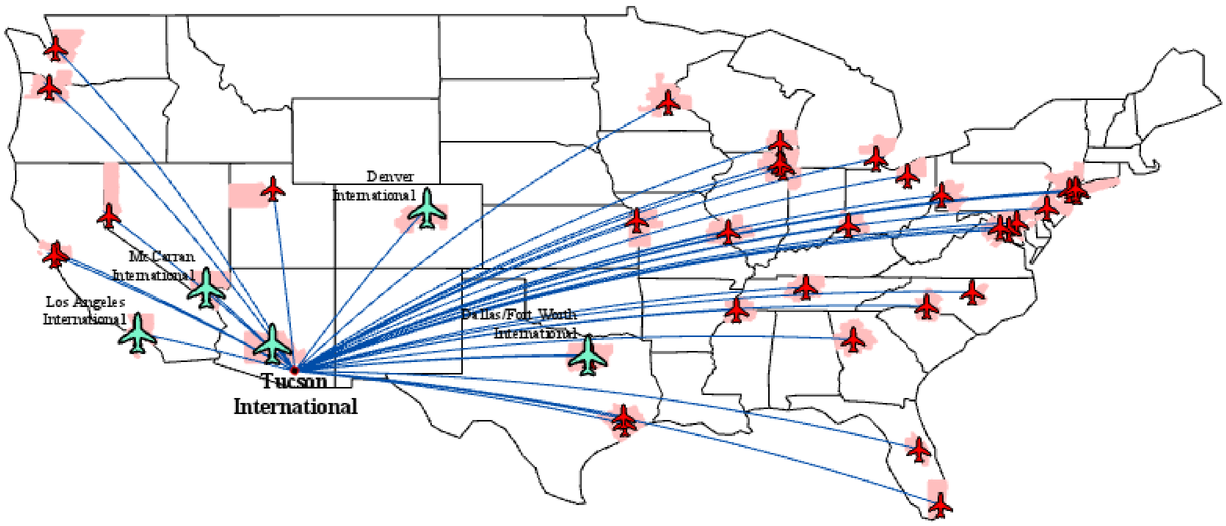


FIGURE 4: MAP OF THE SELECTED AIRPORTS AND CBSAS, 1993-2016



This map shows all the selected airports and their corresponding CBSAs from 1993 and 2016. There are 269 small airports and 243 CBSAs holding these small airports; while there are 38 selected hubs and 31 CBSAs holding these hubs. There can be more than one small airports or hubs in the corresponding CBSAs.

FIGURE 5: TUCSON AIRLINE NETWORK, 1993-2016

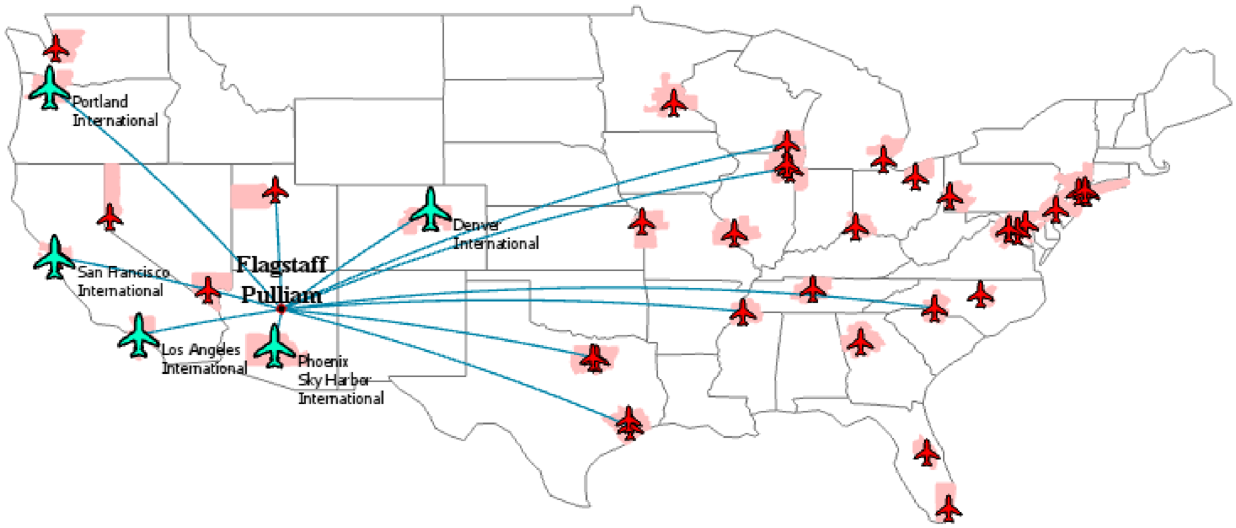


Top 5 Airports Connecting to Tucson Airport

Rank	Airport Name	CBSA Name	Passengers	Distance
1	Dallas/Fort Worth International	Dallas-Fort Worth-Arlington, TX	576832	813
2	Phoenix Sky Harbor International	Phoenix-Mesa-Scottsdale, AZ	536399	110
3	Los Angeles International	Los Angeles-Long Beach-Anaheim, CA	516995	451
4	McCarran International	Las Vegas-Henderson-Paradise, NV	329967	365
5	Denver International	Denver-Aurora-Lakewood, CO	292499	639

This network shows all connected hubs to Tucson International Airport from 1993 to 2016. There are 37 hubs in total connecting to Tucson airport across the years. The airport symbols represent all the hubs operating mainly under the hub-and-spoke system and the pink filled areas represent CBSAs holding these hubs. The larger airport symbol represents the top 5 hubs that transfer the most passengers from/to Tucson airport. The rank in the list is calculated based on the mean of passengers carried by the hub through Tucson airport across the years. Passengers in the list are the mean passengers who are transferred through the corresponding hub and Tucson airport from 1993 to 2016. Distance in the list is the flight distance in miles from Tucson airport to the corresponding hub.

FIGURE 6: FLAGSTAFF AIRLINE NETWORK, 1993-2016

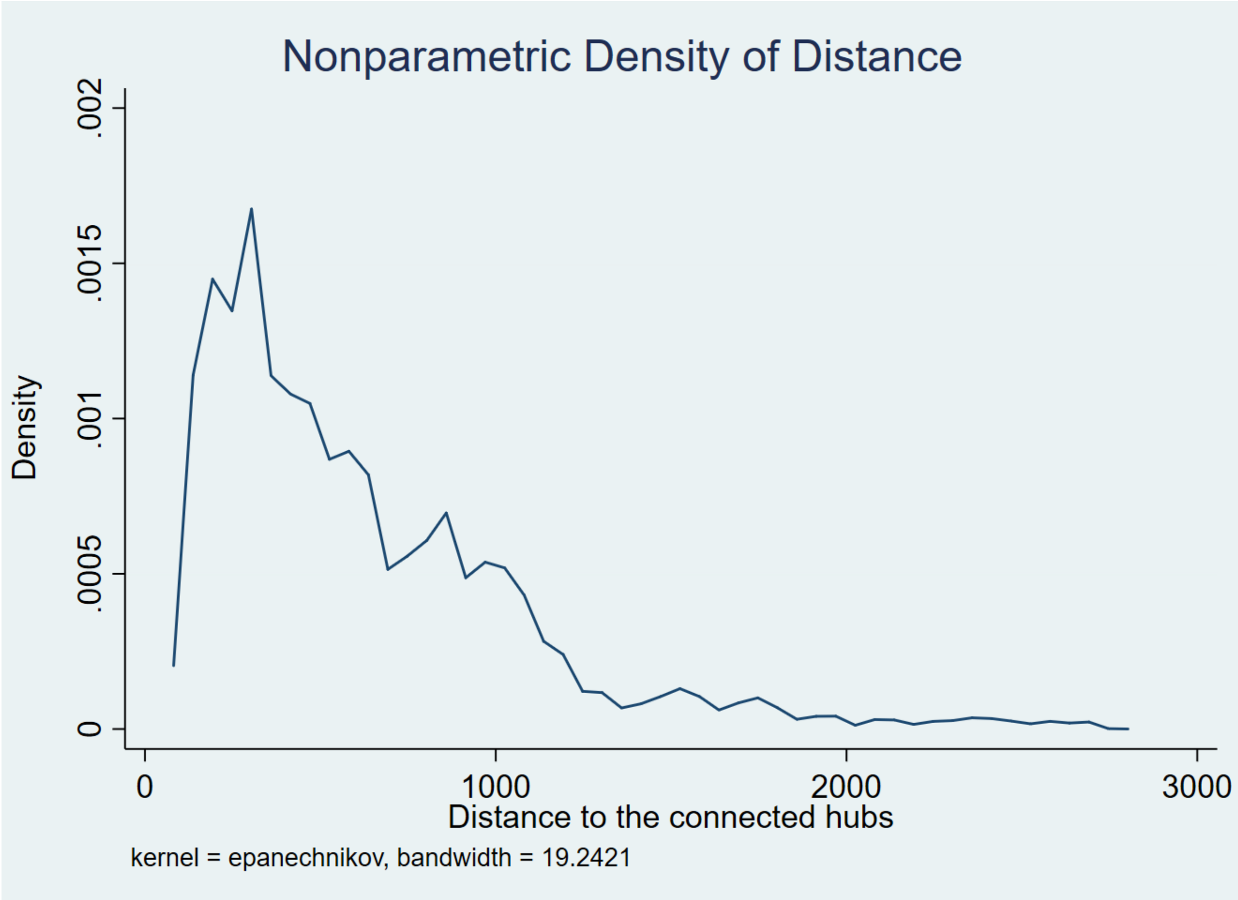


Top 5 Airports Connecting to Flagstaff Airport

Rank	Airport Name	CBSA Name	Passengers	Distance
1	Phoenix Sky Harbor International	Phoenix-Mesa-Scottsdale, AZ	91776	119
2	Los Angeles International	Los Angeles-Long Beach-Anaheim, CA	11116	393
3	Denver International	Denver-Aurora-Lakewood, CO	207	504
4	San Francisco International	San Francisco-Oakland-Hayward, CA	202	620
5	Portland International	Portland-Vancouver-Hillsboro, OR-WA	180	921

This network shows all connected hubs to Flagstaff International Airport from 1993 to 2016. There are 13 hubs in total connecting to Flagstaff airport across the years. The airport symbols represent all the hubs operating mainly under the hub-and-spoke system and the pink filled areas represent CBSAs holding these hubs. The larger airport symbol represents the top 5 hubs that transfer the most passengers from/to Flagstaff airport. The rank in the list is calculated based on the mean of passengers carried by the hub through Flagstaff airport across the years. Passengers in the list are the mean passengers who are transferred through the corresponding hub and Flagstaff airport from 1993 to 2016. Distance in the list is the flight distance in miles from Flagstaff airport to the corresponding hub.

FIGURE 7: KERNEL NONPARAMETRIC DENSITY OF DISTANCE TO THE CONNECTED HUBS



This figure shows the kernel nonparametric density of distance to the connected hubs of all selected small airports. The minimum distance to the connected hubs is restricted at 100 miles to avoid potential confound problems.

10 Appendix

TABLE A1: INDUSTRY DEFINITIONS FROM SIC AND NAICS CLASSIFICATIONS

Industry	SIC code	NAICS code
Agriculture	7	11
Mining	10	21
Construction	15	23
Manufacture	20	31
Wholesale	50	42
Retail	52	44
Finance, Insurance and Real Estate	60	52, 53
Transportation and Utilities	40	22, 48
Service	70	54, 55, 56, 61, 62, 71, 72, 81

TABLE A2: AIR TRAFFIC ON NUMBERS OF ESTABLISHMENT BY INDUSTRY FOR CBSAS WITH LARGE HUBS

	(1) Transportation and utilities	(2) Agriculture	(3) Mining	(4) Construction	(5) Manufacturing	(6) Wholesale	(7) Retail	(8) Finance, insurance , and real estate	(9) Service
OLS results:									
Passenger growth rate in CBSA level	0.042**	0.052	0.139***	0.030	0.008	0.033**	0.022	0.052***	0.031*
	(0.019)	(0.085)	(0.048)	(0.019)	(0.011)	(0.013)	(0.014)	(0.018)	(0.016)
Obs.	713	713	713	713	713	713	713	713	713
R-squared	0.731	0.953	0.251	0.626	0.364	0.490	0.910	0.470	0.811
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CBSA FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
IV results:									
Passenger growth rate in CBSA level	0.455	-0.371	-0.031	0.431	0.443	0.492	0.348	0.491	0.286
	(0.511)	(0.531)	(0.538)	(0.298)	(0.356)	(0.466)	(0.352)	(0.332)	(0.307)
Obs.	713	713	713	713	713	713	713	713	713
R-squared	0.639	0.951	0.242	0.342	-0.257	0.141	0.858	0.108	0.727
First-stage regression:									
Instrument	0.121**	0.122***	0.126***	0.122**	0.125***	0.122**	0.119**	0.122**	0.117**
	(0.044)	(0.044)	(0.043)	(0.045)	(0.044)	(0.045)	(0.046)	(0.045)	(0.045)
Obs.	713	713	713	713	713	713	713	713	713
R-squared	0.339	0.343	0.335	0.359	0.336	0.341	0.346	0.342	0.353
F test for instrument	7.433	7.682	8.424	7.330	8.074	7.449	6.737	7.461	6.800

Note: each regression contains 31 CBSAs with 38 large hubs in total. Standard errors are in parenthesis, clustered by CBSA. ***, **, * denote significant level at 1%, 5%, 10% respectively.

TABLE A3: AIR TRAFFIC ON AGGREGATE PAYROLL BY INDUSTRY FOR CBSAS WITH LARGE HUBS

	(1) Transportation and utilities	(2) Agriculture	(3) Mining	(4) Construction	(5) Manufacturing	(6) Wholesale	(7) Retail	(8) Finance, insurance, and real estate	(9) Service
OLS results:									
Passenger growth rate in CBSA level	0.114*	-0.115	0.594*	0.102**	0.009	0.063**	0.027	0.064*	0.064*
	(0.059)	(0.402)	(0.324)	(0.038)	(0.030)	(0.029)	(0.027)	(0.034)	(0.036)
Obs.	713	656	673	713	713	713	713	713	713
R-squared	0.445	0.696	0.330	0.674	0.461	0.388	0.803	0.419	0.734
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CBSA FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
IV results:									
Passenger growth rate in CBSA level	1.197	3.652	-4.627	0.506	0.383	0.335	0.346	0.384	0.251
	(1.011)	(2.339)	(3.275)	(0.387)	(0.401)	(0.403)	(0.326)	(0.356)	(0.365)
Obs.	713	656	673	713	713	713	713	713	713
R-squared	0.219	0.651	-0.016	0.603	0.392	0.325	0.743	0.334	0.711
First-stage regression:									
Instrument	0.125***	0.121**	0.134***	0.118**	0.125***	0.121**	0.122***	0.123***	0.115**
	(0.044)	(0.044)	(0.042)	(0.044)	(0.044)	(0.045)	(0.044)	(0.043)	(0.043)
Obs.	713	673	689	713	713	713	713	713	713
R-squared	0.335	0.316	0.353	0.344	0.335	0.340	0.341	0.342	0.346
F test for instrument	8.103	7.298	10.428	7.093	8.188	7.150	7.719	8.315	7.228

Note: each regression contains 31 CBSAs with 38 large hubs in total. Standard errors are in parenthesis, clustered by CBSA. ***, **, * denote significant level at 1%, 5%, 10% respectively.

References

Bartik, T. J. (1991). Who benefits from state and local economic development policies? *Kalamazoo, MI: W.E. Upjohn Institute for Employment Research*.

Blonigen, B. A., & Cristea, A. D. (2015). Air service and urban growth: Evidence from a quasi-natural policy experiment. *Journal of Urban Economics*, 86, 128-146.

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

Brueckner, J. K., Lee, D., & Singer, E. (2014). City-pairs versus airport-pairs: a market-definition methodology for the airline industry. *Review of Industrial Organization*, 44(1), 1-25.

Campante, F., & Yanagizawa-Drott, D. (2017). Long-range growth: economic development in the global network of air links. *The Quarterly Journal of Economics*, 133(3), 1395-1458.

Davila, M. A. (2018). Female Labour Market Access, Divorce, and Intra-Domestic Violence: Evidence from Mexico. *Working paper*.

Diamond, R. (2016). The determinants and welfare implications of us workers' diverging location choices by skill: 1980-2000. *American Economic Review,* 106(3), 479-524.

Donaldson, D., & Hornbeck, R. (2016). Railroads and American economic growth: A "market access" approach. *The Quarterly Journal of Economics*, 131(2), 799-858.

Glaeser, E. L., Scheinkman, J., & Shleifer, A. (1995). Economic growth in a cross-section of cities. *Journal of monetary economics*, 36(1), 117-143.

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2018). Bartik Instruments: What, when, why, and how. *National Bureau of Economic Research*, No. w24408.

Green, R. K. (2007). Airports and economic development. *Real estate economics*, 35(1), 91-112.

Hanson, G. H. (2005). Market potential, increasing returns and geographic concentration. *Journal of international economics*, 67(1), 1-24.

Head, K., & Mayer, T. (2006). Regional wage and employment responses to market potential in the EU. *Regional Science and Urban Economics*, 36(5), 573-594.

Jaworski, T., & Kitchens, C. T.. (2016). National policy for regional development: Evidence from Appalachian highways. *National Bureau of Economic Research*, No. w22073.

McGraw, M. J. (2015). Perhaps the sky's the limit? Airports and employment in local economies.

McGraw, M. J. (2017). Does airport size matter? hub airports and local economic outcomes.

Michaels, G. (2008). The effect of trade on the demand for skill: Evidence from the interstate highway system. *The Review of Economics and Statistics*, 90(4), 683-701.

Redding, S., & Venables, A. J. (2004). Economic geography and international inequality. *Journal of International Economics*, 62(1), 53-82.

Sheard, N. (2014). Airports and urban sectoral employment. *Journal of Urban Economics*, 80, 133-152.

Sheard, N. (2019). Airport size and urban growth. *Economica*, 86(342), 300-335.