# Interregional Accessibility and Firm Creation in the Fragmented Economy

Ruichi Xiong
Rotman School of Management
University of Toronto

*Job Market Paper*

October 15, 2022

**Abstract**

Does the reduction in international travel time lower the entry cost for firms, especially those accessing non-local suppliers in the globally fragmented economy? I study this question in the context of recent US-China aviation network expansions, documenting sharp and unevenly distributed travel time reductions between US cities and Chinese prefectures. Employing a novel instrument for travel time constructed from the gradual deregulation of the US-China flight market, I show that the reduction in travel time to China promotes the creation of firms in US cities, more in industries that use many different suppliers. To account for the heterogeneity in supplier presence within China, I estimate a quantitative spatial model featuring sourcing location choice, input-output structure, and a firm entry decision. The model illuminates that the 2004-2013 US-China aviation network expansion increases US firm creation by 1.7%. The heterogeneity in supplier presence across Chinese prefectures drives 42% of the increase because of assortative matching between supplier presence and time reductions in the sparse US-China flight network.

*Keywords*: International Travel Time, Entrepreneurship, Flight Network, Supplier Presence, Fragmented Economy.

*JEL Codes:* F15, M13, R12, R41.

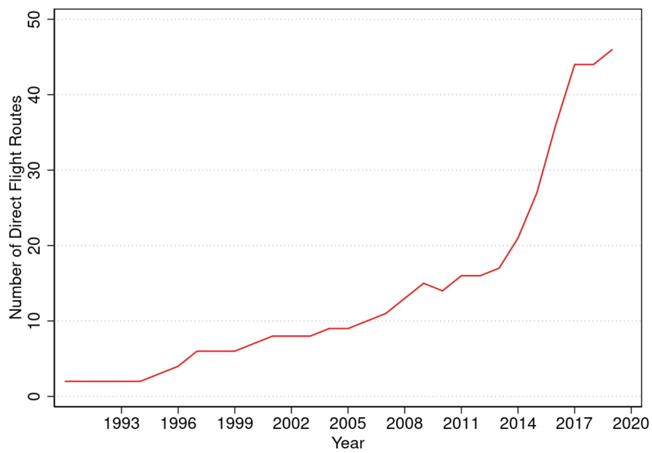
Ruichi Xiong: Ruichi.Xiong@Rotman.Utoronto.Ca, Rotman School of Management at the University of Toronto. I am grateful to Nathaniel Baum-Snow, April Franco, and William Strange for their continuous support and guidance. I thank Stephan Heblich, Yue Yu, Kevin Lim, and Pamela Medina Quispe for their valuable comments and suggestions. I appreciate the efforts spent by Leanne Trimble from the University of Toronto Library on getting the access to the Infogroup data. I also thank participants in the Urban Economics Working Group Meeting and the International Trade Seminar at the University of Toronto and the North American UEA Meeting for helpful discussions and comments. All errors remain my own.

# 1 Introduction

The explosion in long-haul flights in recent decades has sharply reduced international travel time, thereby enhancing entrepreneurs' accessibility to the global economy. In particular, accessing foreign locations with high supplier presence could lower entry costs for firms. This paper studies the effect of international travel time reductions on firm creation and estimates the importance of global supplier presence in determining the effect.

The importance of supplier presence for firm creation has been conceived as a local phenomenon since Vernon (1960) and Chinitz (1961), because of the necessity of face-to-face contacts with suppliers (Jacobs, 1969). In contrast with the past, entrepreneurs today start businesses globally by sourcing from foreign suppliers (Isenberg, 2008). Two recent changes could drive this shift. First, global aviation network expansions have drastically reduced international travel time. Second, production has become globally fragmented and critical suppliers are often located in foreign countries now (Johnson and Noguera, 2012).

**Figure 1:** Introduction of US-China Direct Flight Routes




Note: The figure plots the number of direct flight routes between US and China over time during 1990-2019. The data used here is the T100 segment data. Direct flight routes mean nonstop routes between US gateway airports and Chinese gateway airports. A new direct flight between two gateway airports is also new route only when there are no direct flights between the two airports before.

These two changes are pronounced in the US-China context. The two countries have become closely connected in the global supply chain and the US-China passenger flight network has witnessed gradual expansions over the last decades.[1] Figure 1 shows the

[1]The US-China cargo flight network is different than the passenger flight network. This paper focuses on the passenger flight network and uses an IV strategy introduced later to address concerns about omitted variable biases associated with the cargo network. Flights all refer to passenger flights later on.

expansion process since the 1981 restoration of the US-China flight market. To study the effects of US-China aviation network expansions on firm creation in US cities, I construct travel time networks, measure travel time between US CBSAs[2] and Chinese prefectures, and document its changes over time and across cities.[3]

I show that one standard deviation's reduction of travel time to China[4] causes about 3% increase in firm creation at both the CBSA level and the CBSA-industry level, with a novel IV strategy. Mechanism analyses indicate that the effects of travel time reductions are higher in industries that use many different suppliers. Using a quantitative spatial model to account for the importance of supplier presence in China, I demonstrate that the 2004-2013 US-China aviation network expansion increases US firm creation by 1.7% and welfare by 0.4%. The heterogeneity in supplier presence across Chinese prefectures drives 42% of this aggregate impact due to the fact that there is assortative matching between supplier presence and time reductions in the sparse US-China flight network.

The paper uses a long difference specification between 2004 and 2013 for estimating the effects of travel time reductions between US cities and China on changes in firm creation in US cities. However, there are potential concerns about omitted variables, the endogeneity of travel time reductions, and measurement errors in travel time. For identification, I propose a novel re-centered IV for travel time reductions, constructed using the institutional context of the US-China flight market deregulation.

The re-centered IV takes advantage of the uncertainty in route applications by US airlines. Since the 1980 agreement on restoration of the flight market between the US and China, there have been three amendments that expanded the quotas on US-China flights. On the US side, the Department of Transportation (DOT) held applications for allocating the quotas to US airlines. Given the high cost of an application[5], airlines won't participate if the chance of winning is low. In fact, there were about five large airlines proposing routes in applications and most of the losers filed rebuttals after the applications. Therefore, the results of applications were unanticipated shocks to participating airlines.

The construction of the re-centered IV follows the logic of traditional IVs in the

[2]A Core-Based Statistical Area is geographic area that consists of one or more counties (or equivalents). It is formed around a urban center of at least 10,000 people. It covers also adjacent counties that are socioeconomically inseparable to the urban center by commuting. 929 CBSAs of US and Puerto Rico include both 388 Metropolitan Statistical Areas and 541 Micropolitan Statistical Areas.

[3]Throughout this paper, cities are interchangeably used with prefectures in the context of China and CBSAs in the context of the US.

[4]Throughout the paper, travel time to China refers to the average travel time to Chinese prefectures.

[5]Airlines need to go through a complicated administrative process and lobby a great number of senators, congressmen, and entities. Moreover, they usually disclosed their participation in applications and losing the applications harmed the expectation of investors.

transportation literature. I first fix the US domestic flight network to the base year and recalculate the pseudo travel time reductions. I then focus on only the indirectly affected CBSAs that have never been connected to China by direct flights. However, the indirectly affected CBSAs would still be non-randomly exposed to the connecting of US gateway airports to China, due to the correlation between the fixed domestic network and the unobserved economic geography within the US.

Therefore, I further employ the application uncertainty by permuting winning routes and losing routes, digitalized manually from government files. Specifically, I recalculate the travel time with the base year US domestic flight network and the US-China flight network permutations. The average travel time reduction among all permutations captures the variation in the non-random exposure. By subtracting it from the pseudo travel time reductions, I get the re-centered IV, following the terminology of Borusyak and Hull (2020).

The identifying variation isolated by the re-centered IV on the *included indirectly affected cities* comes from the as-good-as-random winner-loser comparisons between the *excluded directly affected cities*. The re-centered IV therefore could correct the biases from the positive selection on connected cities on the demand side. Since the same fixed US domestic network is used with winning and losing routes, the winner-loser comparisons by definition re-center the domestic network's correlation with the unobserved economic geography. Therefore, the re-centered IV is orthogonal to the non-random exposure that causes the biases on the supply side. I provide various tests to show that the re-centered IV, compared to the un-centered IV, is more balanced between the treated and the control US CBSAs.

The reduced-form estimates cannot be directly used to extrapolate the aggregate impact of US-China aviation network expansions. The long difference specification requires averaging travel time reductions across Chinese prefectures for each US CBSA because firm creation is not a bilateral outcome between the US and China. Hence the reduced-form estimate implicitly assumes that the reductions in travel time to Chinese prefectures are homogeneous to US entrepreneurs. However, I show that travel time reductions facilitate firm creation in US cities by improving the interregional accessibility of potential entrepreneurs to non-local suppliers in Chinese prefectures. The heterogeneity in supplier presence across Chinese prefectures also contributes to the aggregate impact.

I show this mechanism by estimating the heterogeneous effects of travel time reductions across industries with different supplier or customer intensities. These intensities are measured based on the US input-output table following Levchenko (2007). The effects are larger in the industries with above median supplier intensities but homogeneous across industries with different customer intensities. Therefore, it is the upstream accessibility to non-local suppliers in China that encourages entrepreneurs in the US to found more firms. Furthermore, I find that travel time reductions have no effect on the quality of entrants or incumbent firms in terms of employment. Altogether, these suggest that the reductions in travel time between the US and China lower the entry cost of getting suppliers for entrepreneurs. This leaves the quality of entrants unaffected. On the other hand, the cost of switching suppliers and the competition pressure from more entrants lead to a null effect on incumbents.

Motivated by the reduced-form findings, the last part of the paper constructs a quantitative spatial model for understanding the aggregate impact of US-China aviation network expansions on firm creation in the US. It features sourcing location choice, input-output structure, and a firm entry decision. In the model, the reductions in travel time to Chinese prefectures affect firm creation in US CBSAs only through the specific channel of lowering the entry barriers identified in the reduced-form analyses. Lower travel time to China decreases the expected cost of getting suppliers faced by potential entrepreneurs in the US when making entry decisions, thereby facilitating firm creation. The model captures how interregional accessibility and input-output relationships between locations jointly determine the supply of entrepreneurship in a parsimonious way.

After estimating the model, I evaluate the aggregate impacts of the 2003-2014 US-China flight network expansion considering the heterogeneity in supplier presence across prefectures within China. Overall, this leads to a 1.7% increase in firm creation and a 0.4% gain in welfare. By decomposing, I find that 42% of the effect on firm creation is driven by the heterogeneity in supplier presence across Chinese prefectures, as the Chinese prefectures with higher supplier presence also receive larger time reductions. If all the international airports were connected, however, the same destination heterogeneity becomes unimportant because now the flight network is not sparse, the travel time reductions are flat, and there is no assortative matching between time reductions and supplier presence in the US-China travel time network.

This paper is closely related to both the literature on the effects of travel time reductions (e.g., Cristea (2011); Giroud (2013); Charnoz, Lelarge and Trevien (2018); Blonigen and Cristea (2015); Bernstein, Giroud and Townsend (2016); Chu, Tian and Wang (2019); Pauly and Stipanicic (2021); Bai, Jin and Zhou (2021); Da et al. (2021)) and the literature on how *local industrial conditions* affect entrepreneurship (e.g., Glaeser and Kerr (2009). See a summary in Chatterji, Glaeser and Kerr (2014)). Travel time reductions could improve entrepreneurs' accessibility to *non-local industrial conditions*, which may also be an important determinant for entrepreneurship.

I address this gap between the two literatures by estimating the effects of travel time reductions to China on entrepreneurship in US cities and show that accessibility to Chinese suppliers also matters. As a preparatory step of the estimation, this paper also contributes to the literature by constructing the first measure of travel durations between all cities in the two countries over time. In addition, this paper proposes a new theory for understanding the understudied supply side of entrepreneurship (Glaeser, Rosenthal and Strange, 2010), by modeling the entry cost of getting suppliers.

To achieve identification, this paper creates a novel re-centered IV that contributes to the transportation network literature (e.g., Baum-Snow (2007); Faber (2014); Duranton, Morrow and Turner (2014); Ghani, Goswami and Kerr (2016); Donaldson and Hornbeck (2016); Baum-Snow et al. (2017); Jedwab, Kerby and Moradi (2017); Lin (2017); Agrawal, Galasso and Oettl (2017); Martincus, Carballo and Cusolito (2017); Banerjee, Duflo and Qian (2020); Duranton and Turner (2012). See a summary in Baum-Snow and Ferreira (2015).), especially in the context of studying flight network expansions (e.g., Campante and Yanagizawa-Drott (2018)). The instrument, borrowing the insights in Borusyak and Hull (2020), isolates exogenous variation on part of the transportation network from uncertainty on the other part of the network. This IV strategy has the merit of being orthogonal to the non-random exposure problem that commonly exists in transportation network literature but is challenging to be solved.

This paper also contributes to the literature on understanding the importance of interregional and within-region supplier-customer proximity. Bernard, Moxnes and Saito (2019) and Startz (2016) show that face-to-face contact with non-local suppliers is important to firm performance in different contexts. Given the cost of switching suppliers, proximity to suppliers could exert larger impacts on new firms. I contribute to the literature by exploring whether interregional accessibility to global suppliers affects the entry of potential entrepreneurs in the US-China context, both empirically and quantitatively.

Notice that the two papers' findings are not in contradiction with the null effect found on incumbent firms in this paper because Bernard, Moxnes and Saito (2019) studies within-Japan supplier-customer proximity and Startz (2016) uses a sample of Nigerian wholesale and retail trading firms. The non-local industry conditions accessible to firms are vastly different in the two papers than in this paper. I also contribute to the literature on the relation between entrepreneurship and within-region local supplier-customer proximity which helps to understand the Vernon-Chinitz effect (e.g., Rosenthal and Strange (2010)). The contribution lies in considering instead the effects of interregional accessibility on firm creation in the globally fragmented economy.

Last, this paper contributes to the entrepreneurship literature and the international business literature on born-global firms, dating back to at least McDougall, Shane and Oviatt (1994), McDougall and Oviatt (2000), and Knight and Cavusgil (2004). Such firms are much more common today than when they were first documented in 1993 (Rennie, 1993) but the fundamental factors underpinning the formation of them remain understudied (Cavusgil and Knight, 2015). The classical theory in Oviatt and McDougall (2005) explicitly proposes easier travel to foreign locations as one of the potential factors. This paper provides quantitative evidence in support of the hypothesis, from the perspective of taking advantage of global suppliers.

The rest of the paper is organized as follows. Section 2 introduces the data, including a measurement of travel time between US CBSAs and Chinese prefectures. Section 3 proposes the empirical specification and the identification strategy. Section 4 presents the main empirical findings. Section 5 empirically explores the underlying mechanisms. Section 6 shows the quantitative spatial model and counterfactual simulations. Section 7 concludes.

# 2 Data Construction

This section introduces the construction of measures used in this paper. I first build a measure of travel time between US CBSAs and Chinese prefectures over time. Then I measure firm creation in US at both the city level and the city-industry level. Finally, I characterize the industry compositions across US and Chinese cities and the US input-output structure.

## 2.1 Geographic Units of Analysis in the US and China

As the geographic units of analysis, I choose CBSAs on the US side and prefectures on the Chinese side. Travel time refers to the duration of going from one US CBSA to one Chinese prefecture.

In the contexts of US and China, CBSAs and prefectures are natural choices for delineating both local urban markets [6] and the markets served by commercial passenger

[6] For example, Glaeser and Kerr (2009) in the context of US and Baum-Snow et al. (2017) in the context of China.

flights.[7] Specifically, I use the 2013 delineation of CBSAs in contiguous US states and the mostly updated delineation of prefectures excluding HK, Macau, Taiwan, minority provinces, and minority prefectures. In the end, I calculate the travel time between 909 US CBSAs and 255 Chinese prefectures.[8]

## 2.2 Travel Time Measure

**Data Sources.** I use three data sources for constructing the travel time networks: the US T100 segment data, the US Origin-Destination Survey (ODS) flight coupon data, and OpenStreetMap (OSM).

The T100 segment data has been widely used in literature for characterizing flight networks since Giroud (2013). However, small certified and commuter carriers were added into the data only starting in 2002.[9] Hence, using the T100 segment data before 2002 would miss part of the flight networks and generate a mechanical surge in 2002 in available nonstop flight routes in the US as seen in Figure A.4.

The ODS data has no duration information but does not suffer from the change in reporting standard problem as seen in Figure A.5. Therefore, I combine the duration information in the T100 segment data and the route information in the ODS data to construct flight networks, improving the common method used in literature. Besides, because tickets purchased by passengers represent real routes chosen by them while the T100 segment data summarizes routes reported by carriers, the route information in the ODS data is more reliable than the T100 segment data.

Measuring travel time from US CBSAs to Chinese prefectures needs also OSM, as not all US CBSAs are served directly by commercial passenger flights. OSM enables me to calculate current road driving time from any CBSA to any US airport by routing on the up-to-date US road network.[10]

**Travel Time Network Construction.** I measure the travel time from US CBSAs to Chinese prefectures as the travel time from CBSA centroids to prefecture centroids.

[7] The city markets used by the US Bureau of Transportation Statistics (BTS) are similar to CBSAs. Most prefectures in China have no more than one public airport served by commercial passenger flights.

[8] The 909 CBSAs are shown in Figure A.1 and the 255 prefectures are shown in Figure A.2.

[9] See the rule at https://www.federalregister.gov/documents/2001/08/28/01-21457/air-carrier-traffic-and-capacity-data-by-nonstop-segment-and-on-flight-market.

[10] I use the current road driving time as train is not commonly used in the US for traveling to airports and the US road system does not change much during the sampling period 1993-2019. In rare cases, flight time between locations could be longer than driving time. I always use minimal travel time in this paper.

This choice has three advantages. First, the travel time between CBSAs and prefectures as *points* has a precise definition while the travel time between *areas* does not. Second, using centroids allows the inclusion of the variation in travel time from centroids to airports. Third, by assuming passengers can drive from city centroids to any airport, the travel time from or to cities with no airports can be well defined.

On the travel time networks I construct, there are four types of points: US CBSA centroids, US airports, Chinese gateway airports, and Chinese prefecture centroids. So there are five types of connections: (US CBSA centroid, US-airport), (US airport, US-airport), (US gateway airport, Chinese gateway airport), (Chinese gateway airport, Chinese gateway airport), and (Chinese gateway airport, Chinese prefecture centroid). I get from the three data sources both the durations spent on the connections and the availabilities of the connections to construct the flight networks over time.

Flight time observed in the T100 segment data and flight routes observed in the ODS data are combined to obtain the durations and the availabilities of the (US airport, US-airport) connections. Figure A.6 shows that great-circle distances predict flight duration almost perfectly in linear regressions with the T100 segment data. Speeds of commercial flights remain stable over time since 1990 and do not vary much across routes[11]. The stability of flight speed is present among both domestic and international flights.

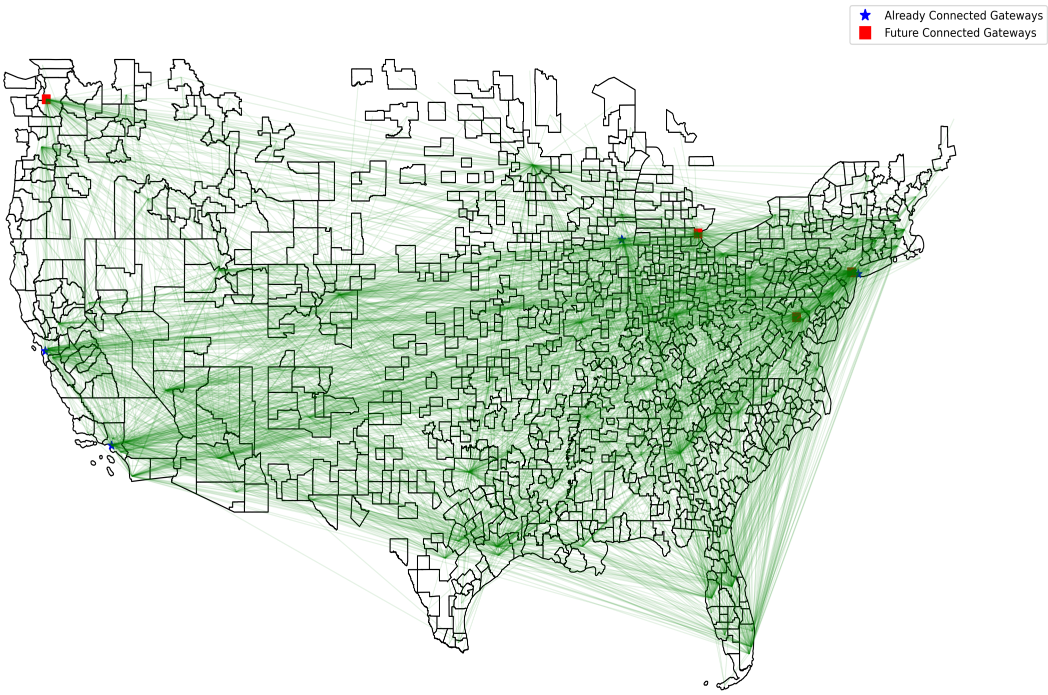
This pattern motivates me to impute the average speed of flights in the T100 segment data to the observed connections in the ODS data[12]. In this way, I get both the availabilities of connections from the ODS data and the durations of these connections from the T100 segment data. At the same time, I avoid the misreporting problem in the T100 segment data. In the end, I obtain a panel of domestic flight networks between US airports at the quarterly level since 1993. One cross-section at the third quarter of 2004 is visualized in Figure 2.

For connections of the (US gateway airport, Chinese gateway airport) type, I instead rely fully on the T100 segment data for two reasons. First, the ODS data does not report flights operated by non-US international carriers and therefore misses a large fraction of international flights. Second, the change in reporting standard targets small carriers while only large carriers are able to operate international flights between the US and China. Therefore I would not miss flight observations by using only the T100 segment data. I corroborate the quality of the data on the availabilities of (US gateway airport,

[11] Pauly and Stipanicic (2021) shows that the speeds of commercial flights changed significantly during 1950s and uses this event as the identifying variation for estimating the effects of travel time reduction on knowledge creation and diffusion. My focus is on periods after 1990s and the main analysis of this paper is actually done for the ten-year period between 2004 and 2013.

[12] The domestic flight duration distribution is shown in Figure A.9 in the Appendix.

**Figure 2:** Domestic Flight Network in 2004




Note: This figure shows the domestic flight network in US at the third quarter of 2004. I focus on 909 CBSAs in the contiguous US states. The points on the map are operating airports and the lines are available routes between them at the third quarter of 2004 according to the ODS data. In particular, I emphasize by blue stars the locations of gateway airports which are already connected by US-China international flights in 2004 and by red squares the locations of gateway airports which will be connected during the ten year period 2004-2013. These gateway airports are all located in large cities. The map also shows that not every CBSA has operating airports and some special CBSAs have airports which position centrally and work as hubs in the domestic flight network.

Chinese gateway airport) type of connections with the US DOT's decision files[13]. These files record the allocation of quotas on designated carriers and number of weekly flights between US and China. For consistency, I impute the durations of US-China nonstop international flights in the same way as US domestic flights[14].

Remaining connections always exist and I get their time-invariant durations from the OSM and the geographic distances. Durations spent on the (US CBSA centroid, US airport) connections are time-invariant and are obtained from the OSM by using the coordinates of US CBSA centroids and US airports. Travel time on the connections of type (Chinese gateway airport, Chinese gateway airport) is imputed consistently with the same speed used for US domestic flights and US-China international flights[15].

The road driving time on the connections of type (Chinese gateway airport, Chinese prefecture centroid) is calculated from the geographic distances between Chinese gateway airports and the centroids of Chinese prefectures. I use a constant driving speed 100km/h (about 60mi/h) following Bai, Jin and Zhou (2021). The connections of types (US CBSA centroid, US airport) and (Chinese gateway airport, Chinese prefecture centroid) by definition always exist.

Combining the above connections, I construct time-variant travel time networks among US CBSA centroids, US airports, Chinese gateway airports, and Chinese prefecture centroids. Each of the networks spans points including 909 CBSAs, 499 US airports, 17 Chinese gateway airports, and 255 prefectures at the quarterly level during 1993-2019. The travel time network changes over time in its availabilities of connections but not the duration on given connection.

**Travel Time Measure.** I measure the bilateral minimal travel time between CBSA centroids and Chinese prefecture centroids in each quarter by searching for the fastest routes on the constructed travel time networks with the Dijkstra Algorithm[16], assuming one hour spent at each stop, following Giroud (2013). This leads to a quarterly panel of CBSA-prefecture pairs with time-variant travel time observed. I average travel time across quarters to arrive at a yearly panel of bilateral travel time between CBSAs and prefectures.

Figure 3 summarizes the pattern. On average, one-way trip from US CBSAs to Chi-

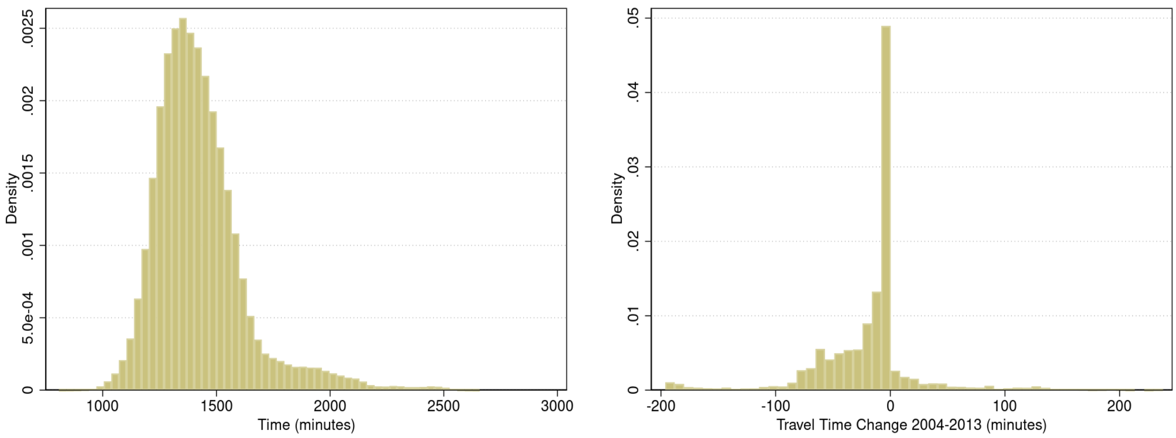
[13]Files can be downloaded from https://www.regulations.gov/.

[14]The flight duration distribution is shown in Figure A.10 in the Appendix.

[15]I confirm that these connections always exist during the sampling period from the archived websites of Chinese airlines on https://archive.org/web/.

[16]Detailed steps in constructing travel time networks and calculating minimal travel time spent on getting from US CBSAs centroids to Chinese prefecture centroids are shown in Appendix A.1.

**Figure 3:** Travel Time to China: Distribution and Change




Note: This figure shows the distribution of minimal travel time from each US CBSA to each Chinese prefecture and its changes over the ten-year period 2004-2013 studied by this paper. The data used for calculating the minimal travel time includes both the ODS data and the T100 segment data. On the left panel, I plot the travel time distribution in all years and across all CBSA-prefecture pairs. The mean is around 24 hours and the distribution is left skewed. On the right panel, I plot the distribution of changes of travel time during 2004-2013. Most of the CBSA-prefecture pairs experience travel time reductions. Few CBSA-prefecture pairs receive increases in travel time during this period because of the decline of the US domestic flight market, as shown in Figure A.5.

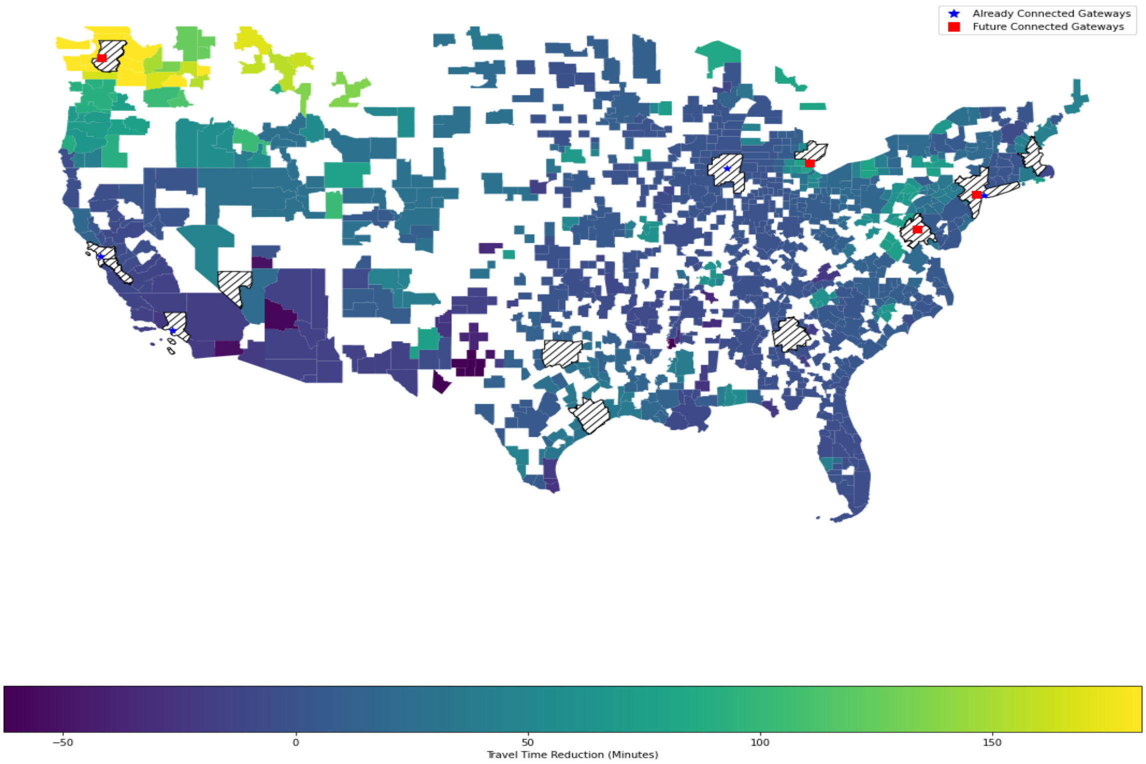
nese prefectures takes about 24 hours. The figure also depicts the changes in the distribution over time. The travel time from US CBSAs to Chinese prefectures decreases sharply during 2004-2013 and the reductions are unevenly distributed across CBSA-prefecture pairs.

Figure 4 visualizes the spatial distribution of the reductions in travel time from US CBSAs to China during 2004-2013 on a map. The travel time reductions between US CBSAs and Chinese prefectures have to be averaged at the CBSA level for drawing the map. It shows surprisingly that the travel time reductions are not necessarily large for the CBSAs closer to the gateway airports connected to China during 2004-2013.

On the other hand, the travel time reductions are always small or even negative for the CBSAs closer to the gateway airports already connected to China before 2004. For example, the CBSAs surrounding DC do not get large time reductions, though the Washington D.C. Dulles Airport (IAD) is connected by nonstop US-China international flights during 2004-2013. Since these CBSAs are also close to the already-connected John F. Kennedy International Airport (JFK), switching to flying from the IAD for traveling to China would not make much difference to travel time as the two airports are too close to each other.

This finding points out that the reductions in travel time to China are affected by

**Figure 4:** Travel Time Reductions across US CBSAs during 2004-2013




Note: This figure shows the reductions in travel time to China from US CBSAs over the ten-year period 2004-2013 studied by this paper. The data used for calculating the travel time reductions include both the ODS data and the T100 segment data. The measured travel time reductions between US CBSAs and Chinese prefectures are averaged at the CBSA level. On the map, the CBSAs with lighter colors experience smaller time reductions. White color with black lines are given to those special CBSAs which have ever had gateway airports with nonstop flights to China. Besides, I emphasize by blue stars the locations of gateway airports which are already connected by nonstop US-China international flights in 2004 and by red squares the locations of gateway airports which will be connected during the ten-year period 2004-2013. The CBSAs closer to the future-connected gateway airports not necessarily get higher travel time reductions. The CBSAs closer to the already-connected gateway airports, on the other hand, always receive very small or even no travel time reductions. For example, the travel time from the CBSAs surrounding Seattle to China declines a lot as they are not close to any already-connected gateway airport and close to the future-connected Seattle-Tacoma International Airport (SEA).

CBSAs' travel time to the gateway airports which are already connected by nonstop US-China international flights in 2004 and the gateway airports which will be connected during the ten-year period 2004-2013. The positions of CBSAs relative to the already-connected gateways and the future-connected gateways on flight network could also correlate with economic geography. Therefore the identification strategy for isolating exogenous variation in travel time reductions should remove the biases correlated with the positions of CBSAs relative to both the already-connected and the future-connected gateway airports.

## 2.3 Firm Creation in US CBSAs

The main data used for characterizing firm creation in US CBSAs is the Historical Business Database provided by Data Axle (previously known as Infogroup). This data has information on US businesses since 1997. Tens of millions of US firms are included in the data. The data has variables including name, exact location, detailed industry classification, and employment.

Importantly, the data records the founding years of firms since 2004. I therefore focus on years after 2004 for avoiding considering only the survived firm entrants. I exclude branch and subsidiary businesses to focus on only the creation of new firms.

I cross check the representativeness of the sample of new firms in the Infogroup data with the Business Dynamic Statistics (BDS) data provided by US Census at both the CBSA-year level and the CBSA-year-industry (2-digit) level. The results indicate that the Infogroup data only becomes representative when I compare it with the BDS data within industries. The fact that industries are sampled differently across cities in the Infogroup data causes the inconsistency between the Infogroup data and the BDS data at the CBSA level.

The Infogroup data is therefore suitable for the city-industry level analysis but not for the city-level analysis, though the estimate from the analysis would not capture variation in industries with less coverage in the Infogroup data. Besides, in the Infogroup data, I can observe detailed industry classifications which are warranted in my city-industry level analysis. In conclusion, I use the BDS data for the CBSA level analysis while use the Infogroup data for the city-industry level analysis.

## 2.4 Industry Composition and Input-Output Structure

I characterize the industry composition across cities and the input-output structure in the US with a set of datasets. The 2000 population censuses of US and China are used to measure the predetermined industry compositions across cities at the 2-digit NAICS level. This is the most detailed level I can achieve with the available data[17].

To characterize the input-output structure between industries in the US, I use the 2002 benchmark IO table provided by the Bureau of Economic Analysis (BEA). Unfortunately, the industry classifications are not the same across these datasets. I use the concordances provided by Liao et al. (2021) and the US Census Bureau to match industry classifications across these datasets.

## 2.5 Summary Statistics

The summary statistics at the CBSA level and the CBSA-industry level in Appendix A.8 preview the results expected in reduced-form estimation later with simple comparisons. The variables I am interested in are the long run differences of the travel time to China and firm creation during the ten-year period 2004-2013. I summarize the two variables for CBSAs with different exposure to the introduction of US-China nonstop international flight routes and industries with different supplier or customer intensities.

The exposure of CBSAs is defined as the travel time to the already-connected US gateways to China. For characterizing the two intensities, I choose to follow Levchenko (2007) and define the following intensity measure:

$$F_k = 1 - \sum_{s}^{S} (\alpha_{mk})^2 \tag{1}$$

where $\alpha_{mk}$ is the value share of industry $m$ as input (output) of industry $k$ for characterizing $k$'s supplier (customer) intensity[18].

I find that the reductions in travel time are associated with larger changes of firm creation during 2004-2013. The exposure of CBSAs to the introduction of nonstop US-China routes is positively correlated with the reductions in travel time to China while

[17] The 2000 population census of US does not provide CBSA delineation. I therefore match the observations in the census to CBSAs using the geographic correspondence engine provided by the Missouri Census Data Center. The web application locates at https://mcdc.missouri.edu/applications/geocorr2014.html.

[18] This measure is one minus the HHI index which evaluates the level of concentration in input-output table. It has the advantage of comprehensiveness, compared to others such as Rauch (1999) and Nunn (2007) which only cover internationally traded industries.

negatively correlated with the changes of firm creation, conditional on city size. This pattern holds at both the CBSA level with the BDS data and the CBSA-industry level with the Infogroup data. It indicates a negative selection problem and downward biases in reduced-form estimation which needs to be corrected for with a valid instrument.

I also find that the changes in firm creation during 2004-2013 are larger in the industries with high supplier intensities and the industries with high customer intensities. This result indicates that the industry heterogeneity in supplier intensities or customer intensities matters for firm creation. The effects of travel time reductions on firm creation could be heterogeneous across industries. I therefore need to compare CBSAs within the same industry to control for that heterogeneity in reduced-form estimation.

# 3 Empirical Strategy

## 3.1 Baseline Specification

I choose a long difference specification for estimating the effect of the reductions in travel time to China on firm creation for several reasons.

First, the US-China nonstop international routes are introduced gradually in the last two decades as seen in Figure 1. Therefore travel time from US CBSAs to China do not sharply decline in a short period. The long difference specification helps me define a clean treatment and enables me to interpret the estimates as the local average treatment effect. As emphasized by recent literature on the two-way fixed effects (TWFE) model[19], heterogeneous dynamic effects and the continuous changes of travel duration to China over time would bias my estimates and make the interpretation of estimates unclear if I employ instead a panel specification.

Second, this paper estimates the effect of travel time reductions on firm creation through evaluating the impact of the US-China flight network expansion over the last two decades. Then it would be sufficient to identify the average treatment effect by comparing the treated CBSAs and the control ones before and after the expansion of the network during 2004-2013. The separation of the immediate effects and the dynamic effects enabled by the TWFE model, with a considerable cost in the context of this paper, is unnecessary.

Last, as discussed in the next section, my calculation of travel time between US

[19] See a recent summary in De Chaisemartin and D'Haultfoeuille (2022).

and China cannot avoid measurement errors due to the ignored factors such as flight frequencies, the transfers outside US and China, and the departure and arrival time of flights. Focusing on evaluating the long run impacts over ten years can minimize the influences of these measurement errors.

Specifically, I choose to use a ten-year long difference regression between 2004 and 2013 as the baseline specification:

$$\Delta \log(Y_i) = \alpha + \beta \Delta T_i + \gamma X_i + \epsilon_i \tag{2}$$

where $\Delta \log(Y_i)$ is the change in outcome of interest between 2004 and 2013 for CBSA $i$ and $\Delta T_i$ is the reduction in average travel time from CBSA $i$ to all Chinese prefectures during the same period. I control for location characteristics in $X_i$. For accommodating industry heterogeneity, I expand this baseline specification by using outcomes at the CBSA-industry level and controlling for industry fixed effects in later analyses.

The long difference has 2004 as the baseline year because the Infogroup data starts to record the founding years of firms in 2004. I therefore only observe survived firms which are founded before 2004. Furthermore, focusing on the period after the WTO accession has two advantages: the potentially confounding influences of the WTO accession can be avoided in the estimation and the accessibility to Chinese suppliers matter more to the US entrepreneurs after China joins the WTO.

I choose 2013 as the ending year as there are no significant changes in travel time from US CBSAs to China after 2013 because most of the routes added after 2013 are between already-connected gateway airports in the two countries. These routes are mostly added by Chinese carriers who start to enter the international flight market on a massive scale in 2015. Their entry changes the flight frequency, the available seats, and the ticket prices but not the minimal travel time, which is the focus of this paper.

## 3.2 Measurement Errors, Endogeneity, and Interpretation

The first identification challenge to estimating Equation 2 is attenuation biases caused by the potential errors in my travel time measure. I calculate the minimal travel time from US CBSAs to Chinese prefectures and average them across prefectures for each CBSA. The T100 segment data and the ODS data used for constructing these networks might have errors in observations. A valid instrument can remove the attenuation bias associated with classical measurement errors.

The omitted factors results in the endogeneity of travel time reductions, which is the second identification challenge. First, there are omitted factors relevant to the travel between US and China other than my measure of minimal travel time. These factors could also affect firm creation.[20] It is however infeasible to precisely measure the travel costs between US and China over time. Moreover, after ten years, the changes of minimal travel time driven by the US-China aviation network expansion should be more important than the omitted factors. Therefore I rely on the instrument which isolates exogenous variation in my measure of minimal travel time for addressing the biases associated with these omitted factors.

Second, the heterogeneity in industry composition across Chinese prefectures is omitted in the reduced-from analysis because firm creation is a unilateral outcome and I have to average travel time across Chinese prefectures for each CBSA. But the firm creation in different industries within one CBSA could benefit differently from the reductions in travel time to different Chinese prefectures with different industry composition. For example, reducing the travel time to a Chinese prefecture with a large textile industry would facilitate the entry of firms producing sleepwear but not firms selling agricultural products.

On the demand side, the omitted heterogeneity in industry composition across Chinese prefectures could lead to a positive selection on locations with larger travel demand, driven by the fact that locations with different industry composition benefit differently from being closer to each other. For example, CBSAs could lobby for connections to China given that the reduced travel time to China could facilitate firm creation there. Airlines could also try to get their certain hubs connected to China because they expect to attract passengers to travel to China from the hubs.

On the supply side, there is the non-random exposure problem pointed out by Borusyak and Hull (2020). CBSAs which are already well-connected with US gateway airports to China expect to receive smaller reductions in travel time as travel time cannot be even lower than direct flight time. These CBSAs could also have higher trends in firm creation than other CBSAs which are badly connected with China, as shown in Table A.3. Traditional empirical strategies such as considering only places unintentionally treated or using fixed historical transportation network are unable to purge biases associated with the omitted and uncontrollable non-random exposure.

[20] These factors include but not limited to non-time considerations such as ticket prices, flight frequencies, departure time, airline alliances, transferring choices, and time-related considerations such as time spent on average trip to China instead of by taking fastest route and travel time changes caused by the rapid transportation infrastructure upgrading in last decades within China.

I propose a novel IV strategy to address the concerns about the measurement errors in travel time and the endogeneity in travel time reductions in the following sections. This strategy, however, could only capture the local average treatment effects induced by the changes of my travel time measure. Since I omit improvements in many aspects related to the travel between US and China and within China, my estimate cannot capture the effects of these improvements and is likely to be the lower bound of the true effect of the US-China flight network expansion on firm creation in the US.

Besides, my travel time measure treats the Chinese prefectures as being homogeneous to US entrepreneurs, ignoring the fact that the CBSA-prefecture pairs with more complementary industry composition could receive higher time reductions. My IV, though being orthogonal to the omitted factors, cannot capture the effects of the assortative matching between industry composition and time reductions. Therefore, the reduced-form estimates would underestimate the effects of the US-China aviation network expansion on US firm creation. I employ a quantitative spatial model in Section 6 to explicitly consider the heterogeneity in industry composition across US CBSAs and Chinese prefectures and the input-output structure. This model enables me to uncover the impact of the US-China aviation network expansion as a combination of effects induced by both travel time reductions and the assortative matching between time reductions and industry composition across CBSA-prefecture pairs.

## 3.3 Quota Applications and Counterfactual Routes

There were no direct flights between US and mainland China before the normalization of diplomatic relation between the United States of America and the People's Republic of China in 1979. Immediately after the normalization, the two countries reached a set of agreements in 1980 on various issues, including opening to each other's commercial passenger flights. The 1980 agreement allowed only two airlines and two weekly passenger flights to carriers in each of the two countries. During the 1980s and 1990s, though expanding, the US-China commercial passenger flight market was severely regulated. Only Los Angeles and San Francisco had direct connections to Beijing and Shanghai.

The two countries later reached three amendments of the 1980 agreement in 1999, 2004, and 2007. These amendments were for expanding quotas on designated airlines, route authorities, and number of weekly flights. An additional negotiation in 2010 was scheduled in 2007 to further liberalize the market. However, due to the financial crises and political factors, no agreements had been achieved since the 2007 amendment. Nevertheless, the US-China commercial passenger flight market had grown tremendously over the last two decades, as shown in Figure A.11.

Because of those amendments, the two countries had to assign quotas to their carriers. On the Chinese side, the quota constraint was not binding. The number of gateway airports in China connected directly to US only began to grow since 2015. This was expected, as the Chinese civil aviation industry became competitive only in recent years. On the US side, there was, however, fierce competition immediately since the first amendment was reached in 1999.

The US DOT hosted five centralized application cycles for allocating quotas to the routes proposed by airlines. Applications were very costly for airlines and only the largest carriers such as United, American, and Delta participated. Airlines had to propose routes, frequencies, and aircraft used for the routes. Moreover, they needed to lobby congressmen, senators, big firms, trade associations, and regional authorities to support them. Airlines proposing passenger flights sometimes also had to compete with airlines willing to use the quota for all-cargo flights[21].

The airline applicants in each of these cycles must have high enough prior on chances of winning. Otherwise, they would not pay the enormous cost of applications. The decisions of DOT therefore were as good as exogenous shocks to the winners and losers in every one of application cycles[22].

Following the insights of Borusyak and Hull (2020), I permute winners and losers in the five application cycles to get counterfactuals for constructing the re-centered instrument for my measure of travel time reduction (or control for the expected travel time reduction across counterfactuals). I am able to get winners and losers from regulations.gov for every one of the five application cycles between 2000 and 2010.

These application cycles allocated quotas assigned by the three amendments in 1999, 2004, and 2007 to US carriers. I then can draw applications randomly with equal probability for generating counterfactuals. Airlines, which propose these applications, are assumed to operate their routes conditional on not yet being selected in counterfactuals.

Different sequences of nonstop US-China international flights operated by US carriers then would be introduced over the ten-year period 2004-2013 in counterfactuals than the real one. Notice that the nonstop US-China international flights introduced by the Chinese carriers are not permuted and remain the same as the observed ones in my

[21] The US-China all-cargo flight market was fully liberalized before 2010.

[22] It was common that losing airlines filed rebuttals after the US DOT announced the decisions because the results were unexpected to them. For example, the rebuttal of Delta was reported at https://www.bizjournals.com/atlanta/stories/2007/07/23/daily62.html.

counterfactual sequences.

For making the counterfactual sequences as close as possible to the observed sequence, only the routes eventually selected would be permuted and the number of routes selected in each counterfactual is set to be the same as the observed sequences. Performing this random selection process for every one of the five application cycles generates 238 distinct counterfactual route introduction series in the end. The detailed procedures in generating the counterfactuals are described in Section A.10.

## 3.4 Re-Centered IV Construction

The variation in the minimal travel time $T_{it}$ from CBSA $i$ to China in year $t$ comes from two sources: the introduction of nonstop flight routes from US to China and the changes in US domestic travel network during 2004-2013. Before showing the re-centered IV construction, I calculate an unadjusted IV following the methods widely adopted in literature: fixing the domestic flight network in the baseline year 2004 and focus on the CBSAs which are indirectly affected by the introduction of nonstop US-China flight routes.

For calculating the unadjusted IV, I search for the fastest routes on the pseudo travel time networks where domestic flight networks are kept as the same as the baseline year 2004[23]. Then I obtain the pseudo travel time $\tilde{T}_{it}$ which is not affected by the endogenous changes of domestic flight network in the US.

Nonstop US-China flight routes, however, could still be introduced to selective CBSAs during 2004-2013. For addressing the endogeneity caused by this concern, I further exclude from the data the 13 CBSAs which have ever had gateways airports to China since 1980. The pseudo travel time of the remaining 896 CBSAs are only indirectly affected by the expansion of the US-China aviation network. I then can instrument the reductions in travel time $\Delta T_i$ by the reductions in pseudo travel time $\Delta \tilde{T}_i$ over the same period 2004-2013.

Using the unadjusted IV $\Delta \tilde{T}_i$ however is unlikely to purge biases caused by the non-random exposure problem. Denote the 2004 domestic network as $D$ and the change of international network between US and China as $\Delta F$. Then the unadjusted IV depends

[23] For avoiding the influences of seasonality, the domestic flight network used for calculating the pseudo travel time includes only the nonstop flight routes appearing in every season of 2004.

on both of them: $\Delta\tilde{T}_i = \Delta\tilde{T}_i(\Delta F, D)$. We have:

$$E[\Delta\tilde{T}_i\epsilon_i] = E[E[\Delta\tilde{T}_i(\Delta F, D)\epsilon_i|D]] = E[E[\Delta\tilde{T}_i(\Delta F, D)|D] \times E[\epsilon_i|D]] \neq 0 \tag{3}$$

even if international network change $\Delta F$ is exogenous as omitted economic geography in $\epsilon_i$ correlates with domestic network $D$. The non-random exposure $E[\epsilon_i|D] \neq 0$. Therefore identification cannot be achieved by the unadjusted IV as the exclusion restriction does not hold.

I address this identification challenge by re-centering the unadjusted IV to get the re-centered IV which is plausibly exogenous to the non-random exposure. By permuting the winners and losers in route applications, I obtain the counterfactual sequences of nonstop international routes between US and China: $\{\Delta F^c, c \in C\}$. Following the same procedure for calculating the unadjusted IV, I can calculate the pseudo travel time $\tilde{T}^c_{it}$ from each one of the 238 counterfactuals.

An average reduction of travel time to China $\Delta\overline{T}_i = \frac{1}{238}\sum_{c=1}^{238} \Delta\tilde{T}^c_{it}$ during 2004-2013 across all counterfactuals can then be obtained. It can be interpreted as the expected reduction of travel time to China before the application results as shocks are realized. In the end, I get re-centered IV by re-centering the unadjusted IV relative to the expected travel time: $RIV_{it} = \Delta\tilde{T}_{it} - \Delta\overline{T}_{it}$.

The average across all counterfactuals captures the variation in travel time reductions which can be expected from the observed fixed domestic network in the baseline year 2004 no matter which nonstop US-China route will be connected during 2004-2013. By subtracting it from the unadjusted IV, the residual time reduction $RIV_{it}$ becomes orthogonal to the non-random exposure. In Appendix A.11, I use two stylized counterfactuals to show that re-centering the unadjusted IV mitigates the concerns about non-random exposure.

The re-centered IV can also address the identification challenges on the demand side. Since I have restricted my analysis to the 896 US CBSAs which have never had gateway airports connected directly to China, the re-centered IV only isolates variation in *indirectly connected cities'* travel time reductions driven by the winner-loser comparisons between the *directly connected cities*. Notice also that those losers are comparable large cities and they are eventually connected to China by direct flights. The decision of the US DOT is as good as random for these indirectly connected cities. Hence the re-centered IV should be also orthogonal to the demand-side positive selection on which places to be connected to China.

Formally, the identification assumption for my empirical strategy is: shocks $\Delta F^c$ on *directly connected cities* are independent with the economic geography of *indirectly connected cities* conditional on the predetermined domestic flight network $D$. It implies that the re-centered time reduction is orthogonal to the omitted factors such as measurement errors and the non-random exposure caused by the high dimension geography.

With the identification assumption, we have the following result which guarantees identification:

$$\begin{aligned} & E\{[\Delta\tilde{T}_i(\Delta F^c, D) - E[\Delta\tilde{T}_i(\Delta F^c, D)|D]]\epsilon_i\} \\ =& E\{E[\Delta\tilde{T}_i(\Delta F^c, D)\epsilon_i|D] - E[\Delta\tilde{T}_i(\Delta F^c, D)|D]E[\epsilon_i|D]\} \\ =& 0 \end{aligned} \tag{4}$$

This paper contributes by proposing such novel identification strategy based on the insights of Borusyak and Hull (2020) and the historical institutional contexts of the building of US-China flight network. This identification strategy could be applied in other contexts and is plausibly more valid than the strategies commonly adopted in the literature, such as constructing instruments from hub openings and M&A between airlines or focusing on the unintentionally affected places.

## 3.5 IV Validity Tests

**Balancing Tests.** In Table 1, I compare the unadjusted IV and the re-centered IV by regressing them on geographic controls including the travel time to the already-connected and the future-connected gateway airports in the baseline year 2004.

In column (1), the correlation between the unadjusted IV and the geographic controls is very high. The coefficients of the geographic controls are large and precise. The R square is above 0.7. On the contrary, I get much smaller coefficients and R square below 0.1 when I regress the re-centered IV on the geographic controls in column (2). This suggests that the re-centered IV is orthogonal to the omitted variables including the non-random exposure, which is highly correlated with the geographic controls.

In columns (3) and (4), I further check the correlation between the re-centered IV and the expected travel time reduction across counterfactuals. In column (3), as expected, the correlation is not significant. In column (4), after adding the geographic controls, the correlation becomes significantly positive but is still small. I add these geographic

**Table 1:** Regression of Unadjusted IV and Re-Centered IV on Geography Variables

| | Unadjusted IV | Re-Centered IV | Re-Centered IV | Re-Centered IV |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Log City Size 2004 | -0.027 | 0.002 | | -0.000 |
| | (0.010) | (0.004) | | (0.004) |
| Time to China 2004 | -0.529 | 0.028 | | -0.018 |
| | (0.027) | (0.009) | | (0.013) |
| Time to Existing Gateways 2004 | 1.107 | 0.028 | | 0.118 |
| | (0.041) | (0.010) | | (0.023) |
| Time to Future Gateways 2004 | -0.589 | -0.059 | | 0.104 |
| | (0.022) | (0.009) | | (0.016) |
| Time to Other Airports 2004 | -0.088 | 0.006 | | -0.002 |
| | (0.019) | (0.009) | | (0.009) |
| Expected Time Reduction | | | 0.010 | 0.084 |
| | | | (0.006) | (0.017) |
| $R^2$ | 0.713 | 0.003 | 0.001 | 0.101 |
| Observations | 896 | 896 | 896 | 896 |

Note: This table presents the results of four regressions using the travel time data constructed from the T100 segment data and the ODS data. The unadjusted IV is the pseudo travel time reduction calculated from fixing the US domestic flight network in the baseline year 2004. The re-centered IV is obtained from counterfactuals permuting the winners and losers in US carriers' applications for nonstop US-China flight routes. The existing gateways in 2004 include SFO, LAX, JFK, and ORD. The future gateways include SEA, DTW, IAD, and EWR. The expected time reduction is the average time reduction across counterfactuals.

controls in the re-centered IV estimation and confirm that they indeed do not affect the results in Section 4.

**Network Centrality.** In Table A.6 in Appendix A.12, I perform further balancing tests by regressing travel time reduction, the unadjusted IV, and the re-centered IV on airports' network centrality. The centrality is measured on the domestic flight network in the baseline year 2004. I use three common network centrality measures in the three columns. Every cell in the table represents a correlation between the column variable and the row variable.

I find that travel time reductions are negatively correlated with all the three network centrality measures. This is also true for the unadjusted IV. The central locations are also close to the already-connected gateway airports. The non-random exposure therefore explains the negative correlation between network centralities and travel time reductions.

This negative correlation still persists for the unadjusted IV. So the unadjusted IV can not remove the non-random exposure and is not valid for correcting biases caused by the non-random exposure problem. On the contrary, the re-centered IV is not correlated with any one of the three network centrality measures as shown in the third row. Therefore the re-centered IV is able to correct the biases associated with the non-random exposure.

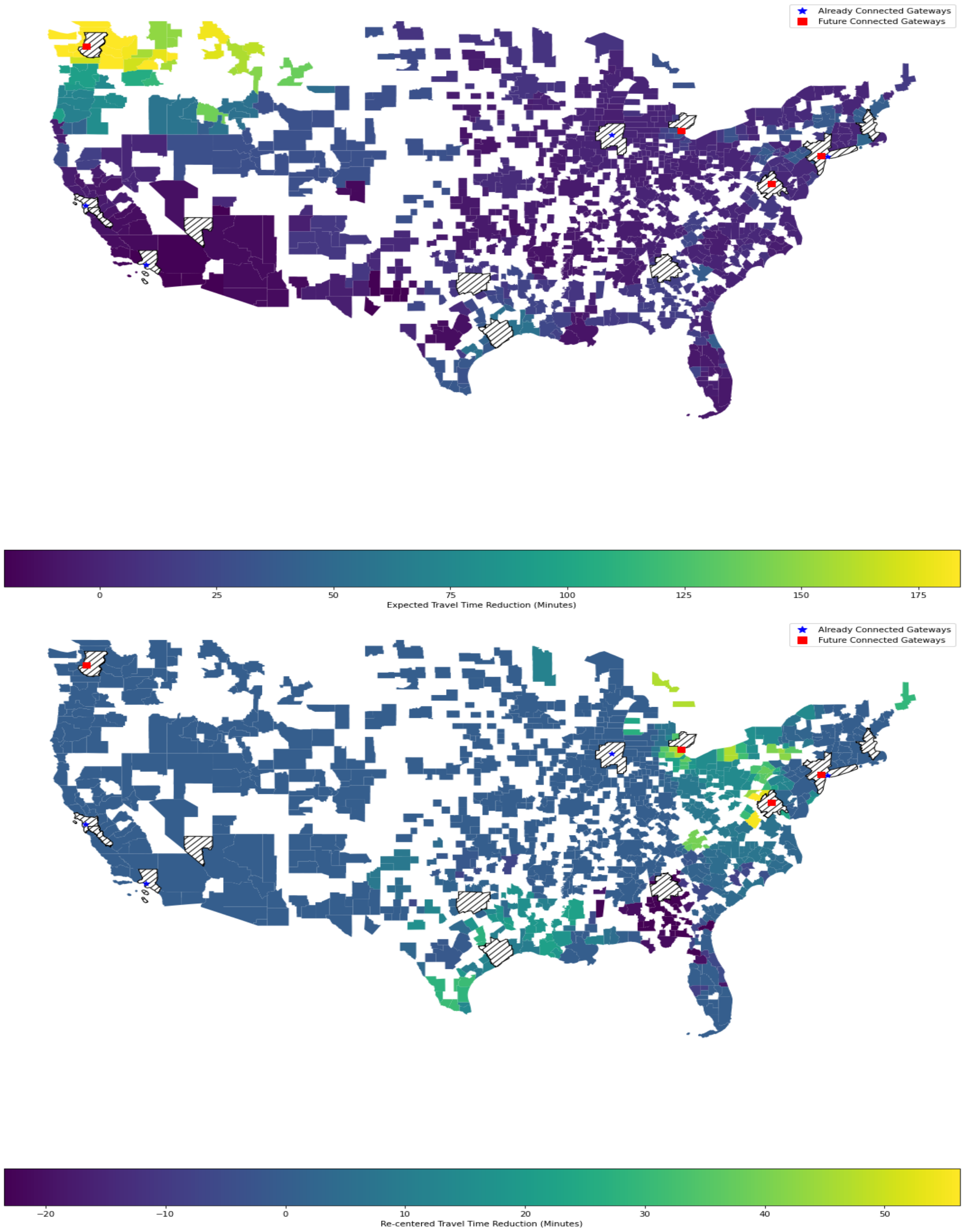
**Expected versus Re-Centered.** Figure 5 compares the re-centered IV and the expected time reduction across counterfactuals on a map. It shows that I indeed have removed the non-random exposure by re-centering. The top panel plots the expected travel time reduction across US CBSAs.

CBSAs which are close to the already-connected gateway airports expect to receive smaller time reductions. For example, CBSAs around DC have small expected time reductions because they are close to the already-connected JFK, though DC is connected during 2004-2013. On the other hand, CBSAs which are close to the future-connected gateways but not close to the already-connected gateways expect large time reductions. The most pronounced examples are the CBSAs around Seattle.

In the bottom panel, when examining the re-centered time reductions across US CBSAs, both the high exposure of CBSAs around Seattle and the low exposure of CBSAs around DC have been removed. CBSAs around Seattle now do not get large time reductions because Seattle's expected exposure across counterfactuals is high and the re-centering by subtracting the expected exposure keeps only the residual exposure which is low.

Similarly, the re-centering removes the influences of JFK on the travel time reductions

**Figure 5:** Expected versus Re-Centered Travel Time Reduction




Note: This figure shows the expected and the re-centered reductions in travel time to China from US CBSAs over the ten-year period 2004-2013 studied by this paper. The data used for calculating the travel time reductions include both the ODS data and the T100 segment data. The measured travel time reductions between US CBSAs and Chinese prefectures are averaged at the CBSA level. On the map, white color with black lines are given to those special CBSAs which have ever had gateway airports with nonstop flights to China. The top graph plots the expected reductions in travel time to China while the bottom the re-centered reductions.

of CBSAs around DC. After the re-centering, these CBSAs get higher travel time reductions from connecting DC and China during 2004-2013 because I remove the influences of the already-connected JFK by permutation. By re-centering, I therefore only keep the residual variation in travel time reduction originated from the as-good-as-random comparisons between the excluded CBSAs which ever had gateways to China.

**Parallel Trends.** This part compares the pre-trends of CBSA-industry pairs which are treated by the reductions in travel time to China and the other CBSA-industry pairs which are not during 2004-2013. Since I use the re-centered IV to get exogenous travel time reductions, the treatment here is defined as having re-centered IV positive: $\Delta RIV_i > 0$. Therefore the analysis is essentially a balance test of the treated CBSA-industry pairs and the control CBSA-industry pairs based on the exogenous treatment driven by the re-centered IV.

For achieving this goal, I employ a standard event-study specification below for years during 1997-2004:

$$\log(Y_{ikt}) = \alpha + \sum_{\tau=1997}^{2004} \beta_\tau \times treated_i \times d_{t,\tau} + \theta_i + \theta_{kt} + X'_{ikt} \times \gamma + \epsilon_{ijt} \tag{5}$$

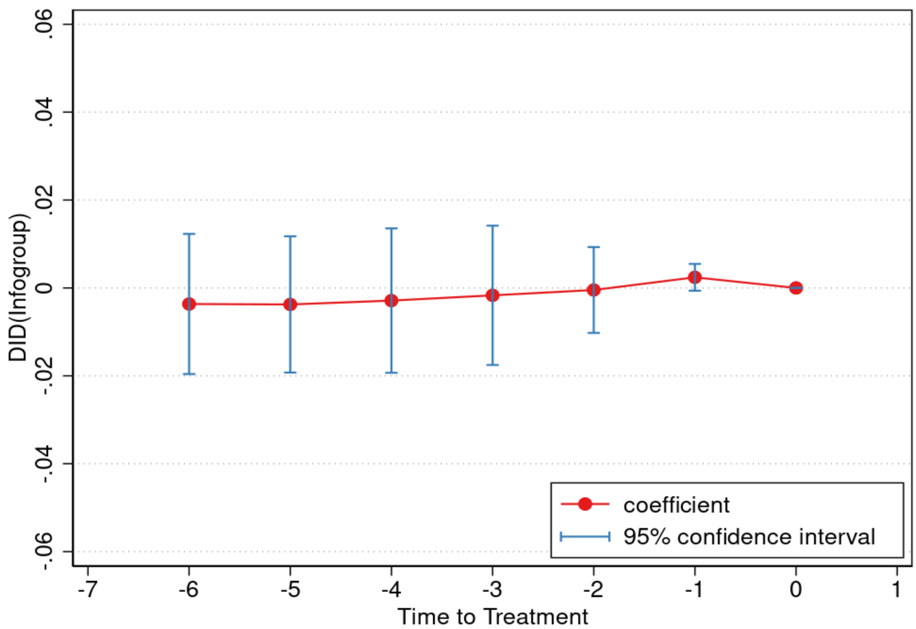
where $\log(Y)_{ijt}$ is our outcome of interest for CBSA $i$ in industry $j$ and year $t$. The time dummy $d_{t,\tau}$ indicates whether year $t$ is larger than the pseudo event time $\tau$ or not. I control for the CBSA fixed effects in $\theta_i$ and the industry-year fixed effects in $\theta_{kt}$. $X'_{it}$ includes all the controls at the CBSA-year level. Standard errors are clustered at the level of CBSA. Insignificant $\beta_\tau$ is expected if there are indeed no pre-trends between the treated CBSA-industry pairs and the control CBSA-industry pairs.

The results using the Infogroup data at the CBSA-year-industry (6 digit) level are shown in Figure 6. I plot the coefficients for each event year relative to the baseline year 2004. Reassuringly we do not see any difference in pre-trends between the treated CBSA-industry pairs and the control ones.

Notice that the Infogroup data only starts to record firms' founding years since 2004. Therefore here we are actually using survivors. This explains why the parallel trends are more precise for the years closer to 2004 than the years further away. Because of this data limitation, I complement the analysis using the BDS data at the more aggregated CBSA level.

The results are shown in Figure A.15 in the Appendix A.12. As expected, the estimates are noisier due to the aggregation. But the pattern does not reject the hypothesis

**Figure 6:** No Pre-Trends




Note: This figure shows the coefficients and the 95% confidence intervals from the event study specified in Equation 5 at the CBSA-year-industry (6-digit) level with the Infogroup data. 2004 is the baseline year and the coefficients represent differences in pre-trends relative to the baseline year 2004 between the treated CBSA-industry pairs and the control ones. The treatment here is defined as having positive re-centered travel time reduction. I control for both the industry-year fixed effects and the CBSA fixed effects. I also control for lagged city employment. Standard errors are clustered at the CBSA level. The results show that the trends of firm creation across the treated CBSA-industry pairs and the control ones are parallel to each other.

that there are no pre-trends between the treated and the control CBSAs before 2004. Besides, I compare the trends in CBSA employment with the same even-study specification and find no pre-trends either in Figure A.16 in the Appendix A.12.

# 4 Results

## 4.1 CBSA Level

Table 2 presents the effect of reduced travel time to China on firm creation over ten-year period 2003-2014 at the CBSA level with the BDS data. In panel A, I present the results from regressions without controlling for geographic factors and industry composition. There are no significant effects from the OLS regression as shown in column (1). As I use the unadjusted IV in column (2), the effect becomes positive but remains not significant.

In column (3), the re-centered IV is introduced and the effect becomes much larger. This is consistent with the previous findings which indicate that there could be downward bias because of the non-random exposure problem and the re-centered IV could correct such bias. This result also implies that the non-random exposure problem is only partially solved by the unadjusted IV used in column (2).

In column (4), the same unadjusted IV as column (2) is used, but with the expected travel time reduction being controlled for. The re-centered IV regression and the controlled IV regression essentially use the same identifying variation. The estimate in column (4), as expected, is indeed similar to the one in column (3).

In panel B, I control for geographic factors and industry composition. The coefficient from the OLS regression in column (1) of panel B becomes positive, though not significant. Controlling for geographic factors and industry composition therefore can only partially alleviate the biases associated with the omitted variables. In fact, the correlation between the structure of the domestic flight network and the economic geography cannot be controlled for, due to its high dimension nature.

This is further supported by the results in columns (2) - (4). Controlling for geographic factors and industry composition does correct the downward bias in the unadjusted IV regression in column (2). The estimate now is significantly positive compared to panel A because controlling for geographic factors and industry composition purges biases in the error term. But the use of the re-centered IV or the controlled IV in columns (3) and (4) still makes the estimate significantly larger. This result indicates that adding the

**Table 2:** Effects of Travel Time Reduction on Firm Creation: City-Level

| | 2004-2013 Difference in Log Number of New Firms | | | |
|---|---|---|---|---|
| | OLS | Unadjusted IV | Re-Centered IV | Controlled IV |
| | (1) | (2) | (3) | (4) |
| *Panel A: No Geo or Ind Controls* | | | | |
| Time Reduction (hours) | -0.002 | 0.014 | 0.056 | 0.062 |
| | (0.014) | (0.014) | (0.055) | (0.062) |
| F statistic | | 2732.263 | 129.801 | 120.728 |
| *Panel B: Add Geo and Ind Controls* | | | | |
| Time Reduction (hours) | 0.004 | 0.032 | 0.060 | 0.050 |
| | (0.016) | (0.019) | (0.057) | (0.035) |
| F statistic | | 724.887 | 26.045 | 77.277 |
| N | 895 | 895 | 895 | 895 |
| Expected Time Reduction | N | N | N | Y |
| Log City Size 2004 | Y | Y | Y | Y |

Note: This table reports coefficients from regressing the long difference in log number of new firms on the reductions in travel time to China at the city level with the BDS data. Geographic controls in baseline year 2004 include: the travel time to China, the minimum travel time to the existing US gateway airports to China, the minimum travel time to the future US gateway airports to China, and the minimum travel time to other airports. One CBSA is dropped because of the missing city size. The expected travel time reduction is calculated as the average travel time reduction across counterfactuals. The controlled IV regression uses the unadjusted IV and controls for the expected time reduction. I control for 3-digit industry compositions by adding the predicted log firm creation in regressions. I predict the change of log firm entry with industry composition controls in the base year 2004 using maximum likelihood regression with log normal density following propensity score matching literature (Hirano and Imbens, 2004). For 3-digit industry compositions, I have the average size of firms, the number of new firms of each industry in each CBSA, and the HHI index of firm employment for each CBSA as controls. I report bootstrap standard errors as I use two-step control function approach. Notice that the maximum likelihood estimation cannot converge if controlling for industry composition at a level more disaggregated than 3-digit. Here I do not use the inverse sine hyperbolic transformation as there are no log zero in the outcome variable at the city level. Kleibergen-Paap rk Wald F statistic is reported for IV regression.

specific geographic or industry controls is unable to purge all the biases originated from the omitted variables such as the non-random exposure.

Notice the re-centered IV and the controlled IV estimates in panel B are similar to in panel A in columns (3) and (4). This implies that the use of the re-centered IV does remove the biases correlated with the geographic factors and industry composition at the city level. These estimates themselves are also close to the one in CBSA-industry level analysis in Table 3. This result further supports the validity of the re-centered IV in purging biases associated with the omitted variables.

## 4.2 CBSA-Industry Level

**Table 3:** Effects of Travel Time Reduction on Firm Creation: City-Industry Level

| | 2004-2013 Difference in Log Number of New Firms | | | |
|---|---|---|---|---|
| | OLS | IV | Re-Centered IV | Controlled IV |
| | (1) | (2) | (3) | (4) |
| *Panel A: No Geo Controls* | | | | |
| Time Reduction (hours) | 0.011 | 0.008 | 0.049 | 0.056 |
| | (0.003) | (0.003) | (0.011) | (0.012) |
| F statistic | | 2735.326 | 35.596 | 120.999 |
| *Panel B: Add Geo Controls* | | | | |
| Time Reduction (hours) | 0.015 | 0.021 | 0.052 | 0.040 |
| | (0.005) | (0.008) | (0.024) | (0.016) |
| F statistic | | 737.519 | 28.040 | 83.141 |
| N | 826085 | 826085 | 826085 | 826085 |
| Expected Time Reduction | N | N | N | Y |
| Log City Size 2004 | Y | Y | Y | Y |

This table reports coefficients from regressing the long difference in log number of new firms on the reductions in travel time to China at the CBSA-industry level with the Infogroup data. Geographic controls in the baseline year 2004 include: the travel time to China, the minimum travel time to the existing US gateway airports to China, the minimum travel time to the future US gateway airports to China, and the minimum travel time to other airports. The expected travel time reduction is calculated as the average travel time reduction across counterfactuals. The controlled IV regression uses the unadjusted IV and controls for the expected time reduction. Here I use the inverse hyperbolic sine transformation for dealing with log zero in the outcome variable at the CBSA-industry level. Kleibergen-Paap rk Wald F statistic is reported for IV regression. I cluster standard errors at the CBSA level.

The best way to achieve reasonable comparisons is to conduct the analysis at the city-industry level, as the disaggregation allows me to control for industry fixed effects. Besides, the city-industry level analysis also enables me to compare industries with different supplier intensities or customer intensities in Section 5. For matching with the most disaggregated US input-output table to take full advantage of the variation in supplier and customer intensities across industries, the CBSA-industry level analysis is conducted at the 6-digit NAICS level.

Table 3 shows the results from the same regressions as Table 2 at the CBSA-industry level with the Infogroup data, controlling for 6-digit industry fixed effects. I use the inverse hyperbolic sine transformation to deal with zeros in the outcome variable. The estimates in both panel A and panel B are now considerably more precise. The effect of travel time reduction on firm creation in the OLS regression is significantly positive as shown in column (1) in both panels. Controlling for industry heterogeneity indeed corrects biases in estimation even in the OLS regression.

In column (2), the unadjusted IV estimates are still significantly different in the two panels as the results at the CBSA level in Table 2. On the contrary, in columns (3) and (4), the re-centered IV estimates and the controlled IV estimates are the same in panel A without geographic controls and in panel B with geographic controls. This pattern, similar to the results at the CBSA level, indicates that after controlling for industry heterogeneity the re-centered IV is orthogonal to the omitted variables, including the non-random exposure, which are correlated with the geographic controls.

## 4.3 Robustness and the OLS-IV Gap

In Appendix A.13, a set of robustness checks is conducted to ensure that the results remain significantly positive in various settings. To avoid the potential biases from the use of inverse hyperbolic sine transformation, I instead run the regression without log or using Pseudo Poisson Maximum Likelihood (PPML) estimation. I implicitly assume different industries have the same weight in baseline results. For correcting potential biases associated with this implicit assumption, I re-weight industries by the industry size measured by employment in 2004 or the inverse sampling probability.

In Appendix A.14, I explain why we have the downward biases found in the comparison between the OLS estimates and the re-centered IV estimates and provide supportive evidence. The OLS-IV gap comes from the fact that the cities close to the already-connected gateways have both smaller time reduction and higher trends in firm creation. I show that the re-centered IV only corrects the downward biases for the cities which have below-median travel time to the existing gateways to China.

## 4.4 Estimate Interpretation and Aggregation Problem

The estimates have to be interpreted with caution because the travel time reductions are averaged across Chinese prefectures. The coefficients in columns (3) and (4), at face value, inform us that a one standard deviation (0.611 hours) reduction, in average travel time to all Chinese prefectures, leads to about three percentage points in log firm creation for an average CBSA-industry pair.

However, the travel time from US CBSAs are only reduced for part of the prefectures. Then half an hours' time reduction from one CBSA to China could imply more than two hours' time reduction to some of the 255 Chinese prefectures. The estimates therefore imply smaller effects of travel time reductions on firm creation than the number obtained directly from Table 3.

We cannot extrapolate the estimates to evaluating the aggregate impact of the US-China aviation network expansion during 2004-2013 on firm creation by simply multiplying the average time reduction with the estimated semi-elasticity 0.052. The reduced-form estimation cannot accomplish this goal because the travel time reductions to different Chinese prefectures with various industry compositions could have different effects on the same industry in the same US CBSA.

For fulfilling the purpose of evaluating the aggregate impact, I first identify that the supplier presence across Chinese prefectures matters for the heterogeneous effects of travel time reductions to these prefectures on firm creation in US CBSAs in Section 5. Then I construct a quantitative spatial model aggregating the time reductions and the supplier presence across prefectures in Section 6. Finally I can evaluate the aggregate impact of the US-China aviation network expansion and decompose it into the effect of time reductions and the effect of supplier presence heterogeneity, with the help of the model.

# 5 Mechanisms

What accessibility has been improved along with the reduced travel time to China and how the better accessibility facilitates firm entry? For this purpose, I first distinguish industries by their supplier intensities and customer intensities. Then I show that the effects of travel time reductions differ by supplier intensity but not customer intensity. In particular, the effects of travel time reductions on firm creation are larger in the industries with above median levels of supplier intensities.

The second analysis I do is to examine the effects on the quality of entrants defined as entrants' future sizes of employment. No effects are found on the quality of entrants in various settings. In the end, I check whether there are effects of travel time reductions on the employment sizes of incumbent firms. It turns out that on average the growth of incumbent firms is not affected by the improved interregional accessibility to suppliers resulted from the reductions in travel time to China either.

## 5.1 Supplier Intensity and Customer Intensity

The reductions in travel time to China help firm enter by increasing the accessibility of potential entrepreneurs to resources in China through face-to-face interactions. For understanding the impacts of building flight network between US and China on the firm creation in US, I further distinguish what accessibility has been elevated for the potential entrepreneurs in US when they can travel to China with significantly less time. One likely explanation for the effects I identified in the previous analysis, is that the increasing proximity between US and China could facilitate the face-to-face contacts between suppliers and customers in the two countries, as suggested by the literature in other contexts (for example, Bernard, Moxnes and Saito (2019) and Startz (2016)).

Entrepreneurs could have better accessibility to the suppliers as well as the customers in China. For figuring out which explanation is supported by the empirical evidence, I take advantage of the detailed industry classification in the Infogroup data and distinguish industries by their supplier intensities and customer intensities. Travel time reductions should benefit entrepreneurs more in the industries with higher supplier (customer) intensities if increasing proximity facilitates the accessibility to suppliers (customers) because the higher supplier (customer) intensity implies the higher demand for getting many different suppliers (customers) in different industries.

Table 4 summarizes the results of estimating the effects of travel time reductions on the firm creation in industries with different supplier or customer intensities. In the top panel, I distinguish the industries with above median level and below median level of supplier intensities while in the bottom panel I distinguish the industries with above median level and below median level of customer intensities. Comparing columns (1)-(2) to columns (3)-(4) shows that the effects are significantly larger in the industries with high supplier intensities. On the other hand, estimates in columns (5)-(6) are not significantly different from the estimates in columns (7-8). Therefore the reductions in travel time to China are more likely to benefit firm creation through improving US entrepreneurs' accessibility to the potential suppliers in China.

**Table 4:** Heterogeneous Effects by Supplier Intensity and Customer Intensity

| | 2004-2013 Difference in Log Number of New Firms | | | |
|---|---|---|---|---|
| | High Supplier Intensity | | Low Supplier Intensity | |
| | Re-Centered IV | Controlled IV | Re-Centered IV | Controlled IV |
| | (1) | (2) | (3) | (4) |
| Time Reduction (hours) | 0.066 | 0.074 | 0.033 | 0.037 |
| | (0.016) | (0.017) | (0.007) | (0.008) |
| F Statistic | 129.946 | 120.999 | 129.946 | 120.999 |
| N | 422400 | 422400 | 397380 | 397380 |
| | High Customer Intensity | | Low Customer Intensity | |
| | Re-centered IV | Controlled IV | Re-centered IV | Controlled IV |
| | (5) | (6) | (7) | (8) |
| Time Reduction (hours) | 0.053 | 0.059 | 0.047 | 0.053 |
| | (0.013) | (0.014) | (0.010) | (0.011) |
| F Statistic | 129.946 | 120.999 | 129.946 | 120.999 |
| N | 409910 | 409910 | 409910 | 409910 |
| Log City Size 2004 | Y | Y | Y | Y |
| Industry FE | Y | Y | Y | Y |

Note: Compared to the re-centered IV regressions, the controlled IV regressions use the unadjusted IV while controlling for the expected travel time reduction across counterfactuals. High supplier (customer) intensity of industry means that the supplier (customer) intensity is above the median across all industries. These intensities are measured in Equation 1 with the BEA 2002 input-output table. I use the inverse hyperbolic sine transformation to deal with the problem of log zero. Kleibergen-Paap rk Wald F statistic is reported for the IV regressions. Standard errors are clustered at the CBSA level.

Though the 2-digit industry classification hides the heterogeneity across 6-digit industries in supplier intensities, I show the effects of travel time reductions by industry at the 2-digit level in Table A.11 in Appendix A.15 for supporting the supplier accessibility mechanism in a nonparametric and interpretable way[24]. Several industries, including agriculture industries and non-tradable local service industries, do not see significant effects. This result supports the supplier accessibility mechanism.

There are large effects in construction, retail trade, and information. These industries are very likely to source inputs from China. On the other hand, the effects are also large in finance, insurance, real estate, rental, and leasing. This could be driven by indirect general equilibrium effects or agglomeration spillovers and the improved accessibility to consumers in China.

In summary, the estimates by 2-digit industries support that the improved supplier accessibility induced by the reduced travel time to China is the main mechanism for accounting for the effect on firm creation in US, though might not be the only one at work. This paper focuses on this mechanism and leaves the other possible mechanism to future research.

## 5.2 Effects on the Quality of Entrants and Incumbent Firms

The improved accessibility to input suppliers could increase firm productivity and therefore encourage firm creation. It could also just remove entry barriers or fixed cost of production after entry for potential entrepreneurs. I distinguish these channels by identifying the effects of travel time reduction on the quality of entrants measured by the future employment of survived entrants[25] and the employment of incumbents.

**Quality of Entrants.** Specifically, I measure the quality of entrants founded in 2004 as their employment in 2009 while the quality of entrants founded in 2013 as their employment in 2018. As the treatment I use is at the CBSA level and I don't have any firm level controls, I compute the average quality of entrants for every CBSA-year-industry as the outcome. Then I employ the same long difference specification as before. I calculate the log changes of the quality of entrants founded in 2004 and in 2013. With these log changes, the long difference compares the trends in the quality of entrants in the

[24]Notice that I can show the effects by more detailed industries. But the results would be uninterpretable without adding parametric structure such as the supplier intensities and customer intensities I use.

[25]Firm exits are not well defined in the Infogroup data. Identifying firm exits is also admitted to be hard in administrative data such as the Longitude Business Database (LBD) (Chow et al., 2021).

treated CBSAs and the control CBSAs within industries. Notice that I use a balanced panel of CBSA-industry pairs which have new firms in both 2004 and 2013.

In Appendix A.16, Table A.12 shows that there are no effects of travel time reductions on the quality of entrants in the various settings in columns (1)-(4) where I either use the baseline long difference specification or do re-weighting by the inverse sampling probabilities or the industry sizes in 2004. From these results, I can rule out the possibility that the better accessibility to suppliers *only* facilitates entry through improving productivity. Because if that's the case, I should also observe the effects of travel time reductions on the quality of entrants.

However, with these results, we still cannot conclude that travel time reductions mainly help entrepreneurs create firms by removing entry barriers. It could be as well possible that travel time reductions work through two channels combined. Removing fixed production cost after entry could select firms with lower productivity in the spirit of Melitz (2003) while the improved accessibility to suppliers could increase new firm's productivity at the same time. There then could be ambiguous effect of travel time reductions on qualities of entrants if travel time reductions works at both the intensive and the extensive margin after entry.

**Incumbent Firms.** For further distinguishing the underlying mechanism, I take a step forward to estimate the effects of travel time reductions on the incumbent firms' employment. If the reduced travel time to China indeed improves firm productivity through the improved accessibility to suppliers, by no means the effects only present on new firms. The analysis now is at the firm level and I consider a balanced panel of firms observed in both 2004 and 2013.

The results are shown in Table A.13 in Appendix A.16. There is no significant effect on the employment in incumbent firms, as seen in columns (1) to (4). It is unlikely that the better accessibility to suppliers driven by the reductions in travel time to China improves firm productivity, at least on average[26]. Then given that there is no ex-post selection on the quality of entrants, the improved accessibility to suppliers should reduce the entry barriers instead of the fixed cost of production after entry.

Taking together the findings in this section, the most likely mechanism for explaining the effects of the reductions in travel time to China on firm creation is that the improved

[26]The reductions in travel time to China might have distributed effects on the productivity of firms. Some firms with low supplier switching cost are benefited while the other firms with high supplier switching cost are harmed, through general equilibrium effects. This is not the focus of this paper. The data does not allow a formal estimation firm productivity either. The distributive effects of interregional accessibility on firm productivity will be left as a promising avenue for future research.

interregional accessibility to China lowers entry barriers in accessing/establishing relationship with suppliers. There are no effects of the reductions in travel time to China, on the contrary, on either the quality of entrants or the sizes of employment of incumbents. This motivates a quantitative spatial model in next section. In the model, the reductions of travel time to China only affects firm creation through reducing the expected sourcing cost faced by potential entrepreneurs when they make the ex-ante entry decision.

# 6 Quantitative Spatial Model

The reductions in travel time to China facilitate firm creation in US cities through improving the interregional accessibility to Chinese suppliers. The effects of connecting US CBSAs and Chinese prefectures therefore must also depend on the presence of non-local suppliers across Chinese prefectures, in the same spirit as the research on local supplier presence and entrepreneurship (for example, Glaeser and Kerr (2009)). For example, air connections to Shanghai should be more valuable than the air connections to agricultural prefectures in China.

Interregional accessibility of industry $k$ in US location $i$ to the suppliers at Chinese location $j$ has two determinants: the travel time from the US location $i$ to that Chinese location $j$ and the presence of suppliers in that Chinese location $j$. The presence of suppliers in Chinese prefectures for industries in US, on the other hand, depends on the industry compositions across Chinese prefectures and the input-output structure. The heterogeneity of industry compositions across Chinese prefectures therefore could contribute as the quality of air connections to the aggregate and distributional impacts of building the US-China aviation network on US entrepreneurial activities.

However, reduced-form estimation is unable to identify the importance of the heterogeneity in supplier presence across Chinese prefectures, as firm creation is not a bilateral outcome between US CBSAs and Chinese prefectures. I have to aggregate Chinese prefectures ignoring their heterogeneity in reduced-form estimation. Therefore this section builds and estimates a quantitative spatial model to understand the aggregate impacts of the 2003-2014 US-China aviation network expansion. The model enables the decomposition of the aggregate impacts into two parts driven separately by: (1) the reductions in travel time from US CBSAs to Chinese prefectures; (2) the heterogeneity in supplier presence across Chinese prefectures.

## 6.1 Model

With the purposes illustrated above in mind, I construct a spatial model with input-output structure (Caliendo and Parro, 2015), sourcing location choice (Antras, Fort and Tintelnot, 2017), and a firm entry decision.

**Setup.** The economy has two countries with many locations within each of them. These locations are indexed by $i, j, d \in J = J^u \cup J^c$ where the set of locations $J$ is an union of locations in US $J^u$ and in China $J^c$. Notice that these locations $i, j, d$ can be within the same country when they are referred without indicating which country they are located in. I have multiple industries indexed by $k, m, h \in S$ where $S$ is the set of all industries. $S^k$ is used to indicate the set of industry $k$'s input industries.

Monopolistic competition is assumed for a continuum of final good producers at each location to enable firm entry decision and endogenous firm creation. I assume free entry. On the other hand, the intermediate input market is assumed to be perfectly competitive. Therefore equilibrium firm creation is solely determined by final good producers.

This modeling choice is motivated by the empirical finding in reduced-form estimation that firm creation is facilitated by the better air connections to China because of the improved accessibility to suppliers in China. Moreover, following Antras, Fort and Tintelnot (2017) for tractability, I assume final goods are non-tradable while intermediates can be freely traded.

The production of final goods follows:

$$y_i^k = z_i^k \prod_{m=1}^{S} \left(v_i^{mk}\right)^{\gamma^{mk}}, \text{ with } \sum_{m=1}^{S} \gamma^{mk} = 1 \tag{6}$$

where $z_i^k$ is the productivity in industry $k$ of location $i$. Assume it is determined by employment: $z_i^k = (L_i^k)^{\lambda}$.

If sourcing inputs from location $j$, intermediates are produced with labor in linear function:

$$v_{ji}^{mk} = \phi_{ji}^{mk} z_j^m l_{ji}^{mk} \tag{7}$$

where $\phi_{ji}^{mk}$ is the relationship-specific productivity shock. $\phi_{ji}^{mk}$ follows an independent Fréchet distribution with the same shape parameter $\theta$: $G_i^k(\phi) = \exp\{-T_i^k \phi^{-\theta}\}$. $l_{ji}^{mk}$ is the employment demanded in industry $m$ at location $j$ for producing intermediate inputs for industry $k$ of location $i$.

Therefore with local wage indicated by $w_j$, the unit cost of intermediates from industry

$m$ at location $j$ used in the production of industry $k$ at location $i$ is:

$$c_{ji}^{mk} = \frac{w_j}{\phi_{ji}^{mk} z_j^m} \tag{8}$$

Firms need to pay a relationship-specific fixed sourcing cost $f \prod_{m=1}^{S^k} (f_{j^m,i})^{\gamma^{mk}}$ for establishing a buyer-supplier relationship with suppliers. $f$ is a constant. $\gamma^{mk}$ is the share of industry $m$ as input in the production of industry $k$. It represents the importance of the buyer-supplier relationship at industry-pair level. $f_{ji}$ is the travel time between location $i$ and location $j$. It measures the difficulty in establishing the buyer-supplier relationship at location-pair level.

Firms face the trade-off between location productivity and travel time to that location when choosing the best location to source inputs for each of its input industry. They will take such trade-off into consideration when they make the ex-ante entry decision by comparing their expected profit and expected sourcing cost. The expected sourcing cost therefore will depend on the ex-post trade-off between travel time and location productivity in a complicated way as shown in Antras, Fort and Tintelnot (2017).

However, this paper focuses on firm creation and does not intend to explain the sourcing location choice of firms. Hence I delay such trade-off to firm entry decision to avoid the unnecessary complication by separating the sourcing location choice and the entry decision. Specifically, I assume a continuum of intermediaries in each industry of each location. These intermediaries draw relationship-specific productivity $\{\phi_{ji}^{mk}\}_{j=1}^{J}$ and buy samples of intermediate inputs from suppliers with least cost. Potential entrepreneurs, after entry, are matched with these intermediaries and the associated suppliers randomly.

Therefore, ex-ante, when potential entrepreneurs decide whether to enter or not, they compare the expected variable profit and the expected sourcing cost. The expected sourcing cost still depends on not only the bilateral travel time but also the productivity distributions across all locations. I therefore maintain the important trade-off between travel time and location productivity at the extensive margin this paper focuses on while maintain tractability with this assumption.

The representative consumer has utility:

$$U_i = (q_i^0)^{\alpha^0} \prod_{k=1}^{S} \left( \int_{\omega \in \Omega_i^k} [q_i^k(\omega)]^{\frac{\sigma-1}{\sigma}} d\omega \right)^{\frac{\sigma \alpha^k}{\sigma-1}}, \text{ with } \sum_{k=1}^{S} \alpha^k = 1 \tag{9}$$

where $q_i^k(\omega)$ is the demand for variety $\omega$ of industry $k$ at location $i$. The set of available varieties is $\Omega_i^k$ as I assume no final good trade. I have standard CES preferences combined with a Cobb-Douglas function to feature input-output structure. The CES parameter $\sigma$ is assumed to be the same across industries.

I assume an outside industry 0 here as the "sponge" industry of this economy. Products in this industry are freely traded and the production uses only labor. By assuming the share $\alpha^0$ is large enough, the wage of the economy can be determined outside of the model and treated as the numeraire. In another word, I have $w_i = w_j = w = 1$. Furthermore, assume $L_i$ measure of workers at location $i$ as endowment and no labor mobility is allowed. Each worker provides one unit of labor. Hence labor supply is inelastic.

The sponge sector assumption and no labor mobility assumption together rule out any GE effects or agglomeration spillovers. It therefore makes the model essentially a partial equilibrium model. This modeling choice makes sure that the change of travel time between locations could only affect firm creation through reducing expected sourcing costs. I then can be confident that the estimation of the model later does not have any upward biases introduced by the concurrent GE effects or agglomeration spillovers happening through labor market adjustments or the unobserved labor market shocks.

Notice the I ignore the potential competition effects of incumbent firms on firm creation here. It is possible that the incumbent firms could benefit from the reductions in travel time to China and preempt the entry of new firms. However, first I find in reduced-form analysis that on average travel time reductions do not affect the employment of incumbent firms. Second, the potential negative spillovers of incumbent firms on entrepreneurship should already be embodied in the reduced-form estimate of the effect of travel time reductions on firm creation. I am essentially identifying the net effect on firm creation conditional on potential preemptive effects from incumbent firms on firm creation. The fact that I still get positive effects on firm creation implies that entrepreneurs benefit more from travel time reductions than incumbent firms, probably because of the supplier switching cost. Therefore without less of generality, I only model the entry decision of entrepreneurs and put the incumbent firms in the background.

**Sourcing Location Choice.** Intermediaries make sourcing location choices by minimizing the unit cost in Equation 8 after drawing idiosyncratic relationship-specific productivity shocks. I therefore get the probability of sourcing intermediate inputs in industry $m$ from location $j$ for the production in industry $k$ of location $i$ as:

$$x_i^{mk}(j) = \frac{(z_j^m)^\theta}{\sum_d (z_d^m)^\theta} \tag{10}$$

Hence the model has the property that industries in locations with higher productivity are more likely to be sourced from. Furthermore, the probability of sourcing from a set of locations by industry $k$ in location $i$ is:

$$x_i^k(\{j^m\}_{m=1}^{S^k}) = \frac{\prod_{m=1}^{S^k} (z_{j^m}^m)^\theta}{\prod_{m=1}^{S^k} \left[ \sum_d (z_d^m)^\theta \right]} \tag{11}$$

where $\{j^m\}_{m=1}^{S^k}$ is a sourcing strategy which indicates the location choices for each input industry.

**Free Entry Equilibrium.** Denote $b = \lambda\theta$. Free entry implies:

$$N_i^k = \frac{\alpha^k L_i}{\sigma} \left\{ \underbrace{f \frac{\prod_{m=1}^{S^k} \left[ \sum_{d=1}^{J} (L_d^m)^b (f_{id})^{\gamma^{mk}} \right]}{\prod_{m=1}^{S^k} \left[ \sum_{d=1}^{J} (L_d^m)^b \right]}}_{\text{weighted average sourcing cost}} \right\}^{-1} \tag{12}$$

*Proposition1: Equilibrium industry labor allocation within location is independent of travel time.*

Changes in travel time would not shift the relative labor demand of industries because the shares of expenditures spent on labor in industries are fixed by the labor shares in production functions and consumption shares in utility functions. In addition, the total incomes of locations are fixed by the inelastic labor supply and the fixed wage. So there would be no sectoral reallocation because of the changes.

With this proposition, I can consider the comparative statics of changing travel time between locations with locations' productivity and employment held constant. My model, in this way, focuses on the specific channel through which travel time reductions facilitate firm creation only by improving interregional accessibility. The proof of Proposition 1 is presented in Appendix A.17.

*Proposition 2:* $-\frac{\partial \log(N_i^k)}{\partial f_{ij}} > 0$ *increases in* $b$ *and* $S^k$.

This model predicts that travel time reductions increase firm creation, especially in industries requiring many different supplying industries. This result confirms that the model could explain the reduced-form findings through the specific channel found by this paper.

Furthermore, the model also uses the parameter $b$ to indicate the key trade-off between location productivity and travel time in determining firm creation. It predicts that travel time reductions have larger effects on firm creation when $b$ is larger. $b$ is the multiplication of agglomeration elasticity $\lambda$ and inverse dispersion of idiosyncratic relation-specific shocks $\theta$. Intuitively, when firms make the ex-ante entry decision, industry composition heterogeneity is more important when agglomeration elasticity is larger and idiosyncratic productivity shocks are less dispersed. The proof of Proposition 2 is also presented in the Appendix A.17.

## 6.2 Link to Reduced-Form Estimates

Consider a change in travel time to Chinese locations from $\{f_{ji}\}_{j \in J^c}$ to $\{\tilde{f}_{ji}\}_{j \in J^c}$, the change of log firm entry is:

$$\Delta \log\left(N_i^k\right) = \sum_{m=1}^{S^k} \left[ \log\left( \left[ \sum_{d=1}^{J} (L_d^m)^b (f_{di})^{\gamma^{mk}} \right] \right) - \log\left( \left[ \sum_{d=1}^{J} (L_d^m)^b (\tilde{f}_{di})^{\gamma^{mk}} \right] \right) \right] \tag{13}$$

This change in log firm entry depends on not only changes of travel time to Chinese locations but also industry composition across these locations. There is an explicit complementarity between time reduction and supplier presence in Equation 13. The same time reduction to Chinese prefecture has larger effect on firm creation if the Chinese prefecture has high presence of suppliers. Both the input-output relationship and the industry composition across Chinese prefectures play important roles here.

For linking to the reduced-form estimates, I use first-order approximation and get:

$$\begin{aligned} \Delta \log\left(N_i^k\right) = & \underbrace{\left( \sum_{m=1}^{S^k} \gamma^{mk} \left[ \frac{\sum_{j=1}^{J^c} (L_j^m)^b (f_{ji})^{\gamma^{mk}} \left[ \frac{\Delta \log(f_{ji})}{\sum_{j=1}^{J^c} \Delta \log(f_{ji})} \right]}{\sum_{d=1}^{J} (L_d^m)^b (f_{di})^{\gamma^{mk}}} \right] \right)}_{\beta_i^k(b) = \text{location-industry heterogeneous marginal effects}} \underbrace{\left[ -\sum_{j=1}^{J^c} \Delta \log(f_{ji}) \right]}_{\text{shock}} \\ & + O\left( \{\Delta f_{ji}\}_{j=1}^{J^i} \right) \end{aligned} \tag{14}$$

The effects of travel time reductions are heterogeneous across both origin-destination pairs and input-output relationships. This result gives us an explicit way to aggregate the heterogeneous effects to arrive at the average treatment effect identified in reduced-form estimation. First, the travel time reduction to all Chinese destinations are summed in

$\Big[ -\sum_{j=1}^{J^c} \Delta \log(f_{ji}) \Big]$ as the treatment to location $i$ in the US. The heterogeneous effects across Chinese locations are then averaged with each Chinese location's share of time reductions as weight.

I therefore get $\beta_i^k$ as the average treatment effect at CBSA-industry level. This aggregation implied by the model enables an interpretation of the effect identified in reduced-form estimation as the local average treatment effect across industries in all origin-destination pairs with non-zero changes of travel time.

The model not only helps us understand and interpret the reduced-form estimates but also explicitly shows us possible biases and their directions. One potential bias comes from the endogeneity of travel time reductions as the paper discussed earlier. The already-connected places which are plausibly also more developed can have simultaneously less expected time reductions and higher trends in firm creation. This causes downward bias in reduced-form estimation.

The other possible bias comes from the correlation between heterogeneous marginal effects $\beta_i^k$ and the travel time reductions across locations. The effects of time reductions could be stronger for better locations with other complementary conditions in place. These locations also expect smaller time reductions. Therefore we would still have potentially downward bias from the endogeneity of heterogeneous effects.

## 6.3 Estimation and Model Fit

With the link between the model and the reduced-form estimates, I then can perform an indirect estimation of the parameter $b$. This is achieved by minimizing the distance between the reduced-form estimates and the impulse responses of the model from introducing the same exogenous time reductions. The impulse response is obtained from the previous comparative statics results while the exogenous time reductions are from the first-stage prediction of the re-centered IV in 2SLS estimation. Specifically, the minimization problem is:

$$\min_{b>0}\left\{\hat{\beta}-\sum_{i=1}^{J^u}\sum_{k=1}^{S}\beta_i^k(b)\left(\frac{\sum\limits_{j=1}^{J^c}|\Delta^e f_{ji}|}{\sum\limits_{d=1}^{J^u}\sum\limits_{k=1}^{S}\sum\limits_{j=1}^{J^c}|\Delta^e f_{jd}|}\right)\right\}^2$$

$$\text{s.t. } \beta_i^k(b)=\sum_{m=1}^{S^k}\gamma^{mk}\left[\frac{\sum_{j=1}^{J^c}(L_j^m)^b(f_{ji})^{\gamma^{mk}}\left[\frac{\Delta^e\log(f_{ji})}{\sum\limits_{j=1}^{J^c}\Delta^e\log(f_{ji})}\right]}{\sum\limits_{d=1}^{J}(L_d^m)^b(f_{di})^{\gamma^{mk}}}\right] \tag{15}$$

where $\hat{\beta}$ is the reduced-form estimate following specification in Equation 14 and $\Delta^e f_{ji}$ is the exogenous change of bilateral travel time.

Estimating $b$ in this way would incorporate variation in supplier presence across not only the Chinese prefectures but also the US CBSAs. For identification, I use the 2000 population censuses of China and US to get sectoral employment. Industry classifications in both censuses are not very detailed. Therefore I aggregate and map industries in both the two censuses and the IO table to NAICS1997 2-digit industries to achieve consistency. Objective function has only relative travel time. So it is scale free and I therefore directly use travel time in estimation.

Table 5 summarizes the estimation results. Notice that the model-implied log-log regression equation here is different than the reduced-form estimation in previous sections where semi-elasticity is estimated. In column (1), I show that the parameter $b$'s estimate is 1.539. With productivity dispersion 8.28 from Eaton and Kortum (2002), the implied elasticity of customer-supplier relationship productivity with respect to supplier presence is 0.186 which is about half of the Chinitz effect estimated in Glaeser and Kerr (2009). With the exogenous component of travel time reductions predicted by the re-centered iv, I can also obtain log change of firm entry predicted by the model. The simulated data can then be used to run the same regression as the reduced-form specification in Equation 14.

The estimation results using the data and the model-generated data are shown in columns (2) to (5). The model replicates the reduced-form estimates well. In addition, for testing model fit, I also check whether the model can replicate the reduced-form estimates by 2-digit industry. The comparison between the reduced-form estimates and the model-generated estimates by 2-digit industry is shown in Figure A.17. Reassuringly, these moments not targeted by the estimation do not deviate from the predictions of the model for most of the 2-digit industries.

**Table 5:** Estimation Results and Model Fit

| | Estimation Data | 2004-2013 Difference in Log Number of New Firms | | | |
|---|---|---|---|---|---|
| | | Data | | Model | |
| | | Re-Centered IV | Controlled IV | Re-Centered IV | Controlled IV |
| | (1) | (2) | (3) | (4) | (5) |
| b | 1.539 | | | | |
| | (0.172) | | | | |
| Shock | | 0.0055 | 0.0041 | 0.0049 | 0.0046 |
| | | (0.0024) | (0.0015) | (0.0008) | (0.0005) |
| F Statistic | | 25.963 | 89.144 | 26.137 | 89.257 |
| Industry FE | | Y | Y | Y | Y |
| Geo Controls | | Y | Y | Y | Y |

Note: Shock is $\left[-\sum_{j=1}^{J^c}\Delta\log(f_{ji})\right]$ which is the model-implied regressor of interest which represents the sum of log reductions in travel time across all prefectures. Regressions in column (2) and (3) employ the same specification and the same re-centered instrument with the observed data and the model-generated data. Geo controls include log city size, the mean travel time to China, minimal travel time to the already-connected gateway airports, minimal travel time to the future-connected gateway airports, and minimal travel time to all other airports in the baseline year 2004. I report bootstrap standard error in column (1). Standard errors are clustered at the CBSA level. I use the inverse hyperbolic sine transformation to deal with log zero in column (2). Kleibergen-Paap rk Wald F statistic is reported for the IV regressions.

## 6.4 Aggregate Effect and Supplier Presence Heterogeneity

With the estimated model, finally I can evaluate the aggregate impact of the US-China aviation network expansion on firm creation and welfare in US cities. This aggregate impact can then be decomposed into two parts: the part driven by the time reductions to China and the part driven by the supplier presence heterogeneity across Chinese prefectures, which is ignored in the reduced-form estimation.

In Table 6, I consider one baseline and three counterfactuals for understanding the aggregate impacts and the importance of supplier presence heterogeneity across Chinese prefectures in accounting for the aggregate impacts. The calculation of welfare gains is based on the following result:

$$\Delta \log(U_i) = \sum_{k=1}^{S} \Big(\frac{\alpha^k}{\sigma - 1}\Big) \sum_{m=1}^{S^k} \Bigg[ - \Delta \log\Bigg( \Big[ \sum_{d=1}^{J} (L_d^m)^b (f_{id})^{\gamma^{mk}} \Big] \Bigg) \Bigg] \tag{16}$$

In the baseline, I fix the domestic flight network to 2004 while introducing the observed new US-China international routes during 2004-2013. The model predicts a 1.7% increase in firm entry while 0.4% increases in aggregate welfare. I then decompose the aggregate effect into two parts: the change of travel time and the heterogeneity in supplier presence across Chinese prefectures. I achieve this decomposition in the following three counterfactuals in Table 6.

The second counterfactual shocks the model in the same way as the baseline except for assuming $b = 0$. This shuts down the importance of the heterogeneity in industry composition across Chinese prefectures and treats Chinese prefectures as being homogeneous to different industries in US CBSAs. Compared to the baseline, the effects on firm creation becomes 42% smaller while the welfare gains decrease by half. Then in the third and fourth counterfactuals, I simply add the heterogeneity in supplier presence by assuming $b$ goes from 0 to 1.539 before and after the route introduction.

Comparing the third counterfactual to the fourth one in last row of Table 6 indicates that the assortative matching between time reductions and the supplier presence across Chinese prefectures leads to 0.7% increases in firm creation and 0.2% increases in welfare after route introduction. Therefore, as shown in column (2) and column (4), the assortative matching induced by the heterogeneity in industry compositions across Chinese prefectures accounts for 42% of the effects of travel time reductions on firm creation while 50% on welfare gains.

**Table 6:** Introducing Observed Routes

| Counterfactuals | 04-13 Δ Log New Firms | | 04-13 Welfare Gain | |
|---|---|---|---|---|
| | CBSA-Industry | | CBSA | |
| | (1) | (2) | (3) | (4) |
| | Mean | Percentage | Mean | Percentage |
| Introduce Observed Routes (Baseline) | 0.017 | 100% | 0.004 | 100% |
| | (0.023) | | (0.006) | |
| Introduce Observed Routes (b=0) | 0.010 | 58% | 0.002 | 50% |
| | (0.021) | | (0.005) | |
| b = 0 to b = 1.539 before Introduction | 0.068 | | 0.020 | |
| | (0.139) | | (0.004) | |
| b = 0 to b = 1.539 after Introduction | 0.075 | | 0.022 | |
| | (0.138) | | (0.004) | |
| Cov(Time Reduction, Supplier Presence) | 0.007 | 42% | 0.002 | 50% |

Note: This table reports the predicted changes of log new firms and the welfare gains for four counterfactuals specified in the left column. The first row adds the observed US-China international flights during 2004-2013 to the flight network holding the 2004 domestic network unchanged. The second row introduces the same set of flights with $b = 0$. The third row changes $b$ before the flight introduction while the fourth row changes $b$ after the flight introduction. The fifth row calculates the difference between the third row and the fourth row. The changes in firm creation come from Equation 13 while the welfare gains come from Equation 16. In column (1), I report the mean and the standard error of changes of firm creation in each counterfactual across CBSA-industry pairs. In column (2), I further show the percentage of counterfactual compared to the baseline in the first row. In column (3), I report the mean and the standard error of welfare gains of each counterfactual across CBSAs. In column (4), I calculate the percentage of welfare gain of each counterfactual compared to the baseline. For counterfactual calculation, I use $\sigma = 5$ from Broda and Weinstein (2006) and $\{\alpha^k\}_{k=1}^S$ from the IO Table. The mean travel time reduction is 17.892 minutes with standard deviation 47.352.

## 6.5 Supplier Presence Heterogeneity in a Dense Flight Network

**Table 7:** Connecting All International Airports in US and China

| Counterfactuals | 04-13 $\Delta$ Log New Firms | | 04-13 Welfare Gain | | Welfare Gain per Minute |
|---|---|---|---|---|---|
| | CBSA-Sector(2 digit) | | CBSA | | CBSA |
| | (1) | (2) | (3) | (4) | (5) |
| | Mean | Percentage | Mean | Percentage | Ratio ($10^{-3}$) |
| Introduce All Routes (Baseline) | 0.092<br>(0.029) | 100% | 0.022<br>(0.006) | 100% | 0.14 |
| Introduce All Routes (b=0) | 0.091<br>(0.025) | 99% | 0.022<br>(0.006) | 100% | |
| b = 0 to b = 1.539 before Introduction | 0.068<br>(0.139) | | 0.002<br>(0.005) | | |
| b = 0 to b = 1.539 after Introduction | 0.069<br>(0.128) | | 0.021<br>(0.004) | | |
| Cov(Time Reduction, Supplier Presence) | 0.001 | 1% | 0 | 0% | |
| Introduce Observed Routes | 0.017<br>(0.023) | | 0.004<br>(0.006) | | 0.22 |

Note: This table reports the predicted changes of log new firms and the welfare gains for four counterfactuals specified in the left column. The first row adds all possible US-China international flights during 2004-2013 to the flight network holding the 2004 domestic network unchanged. The second row introduces the same set of flights with $b = 0$. The third row changes $b$ before the flight introduction while the fourth row changes $b$ after the flight introduction. The fifth row calculates the difference between the third row and the fourth row. The changes in firm creation come from Equation 13 while the welfare gains come from Equation 16. In column (1), I report the mean and the standard error of changes of firm creation in each counterfactual across CBSA-industry pairs. In column (2), I further show the percentage of counterfactual compared to the baseline in the first row. In column (3), I report the mean and the standard error of welfare gains in each counterfactual across CBSAs. In column (4), I calculate the percentage of welfare gain of each counterfactual compared to the baseline. In column (5), I show the welfare gain per minute. For counterfactual calculation, I use $\sigma = 5$ from Broda and Weinstein (2006) and $\{\alpha^k\}_{k=1}^S$ from the IO Table. The mean travel time reduction is 156.984 minutes with standard deviation 49.336.

The heterogeneity in supplier presence is as important as travel time reductions in accounting for the effects of introducing observed routes on firm creation and welfare. During the period 2004-2013, the flight network in the US-China market is very sparse. There are only about 50 direct routes connecting 13 gateways in US and 15 gateways in China even in 2019. With the help of the model, an useful counterfactual connecting all these gateways could be simulated. In this counterfactual, the flight network is much denser than the one considered in Table 6. Therefore the travel time reductions from US CBSAs to Chinese prefectures would be very flat.

As a comparison, the travel time reductions in Table 6 have mean 17.892 and standard deviation 47.532 while the travel time reductions in the counterfactual with full connection have a much larger mean of 156.336 and a similar standard deviation 49.366. When travel time reductions are flat across Chinese prefectures, the importance of destination heterogeneity could decline as there would be less assortative matching between time reductions and supplier presence.

This is indeed the case as seen in Table 7. In column (2) and column (4), I find that now the correlation between time reduction and supplier presence only accounts for one percent of the effects of route introduction on firms creation and less than one percent on welfare gains. Furthermore, I use welfare gain per minute to measure the efficiency of route introduction. In column (5), the efficiencies of connecting the observed routes during 2004-2013 and connecting all the gateways airports are compared.

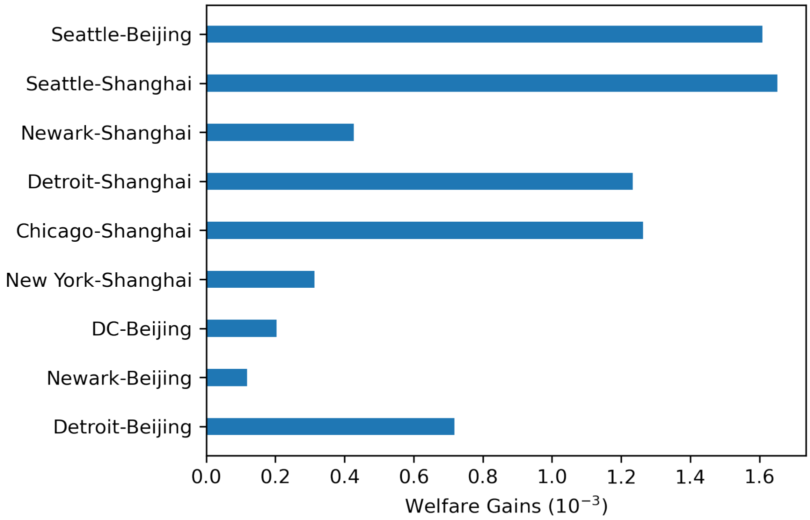
Though the counterfactual with full connections has much larger welfare gains, it is less efficient than connecting only the observed routes. This is intuitive, as the allocation of connections is now not based on supplier presence across Chinese prefectures. Therefore, there would be many lower value connections included. Given the constraints on number of routes, there should be assortative matching between time reductions and supplier presence as the observed routes introduction during 2004 to 2013 for achieving efficiency from the perspective of social planner.

## 6.6 Value of Each US-China Flight Route

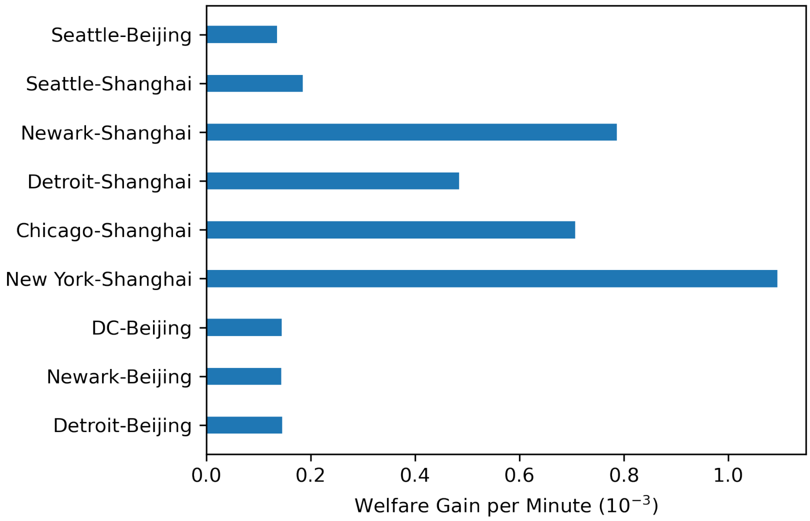
Different US-China routes have different values for different CBSAs and industries in the US, as seen from the previous counterfactuals, because of the heterogeneity in supplier presence across Chinese prefectures. The optimal allocation of a given number of routes should consider the efficiency of each route in terms of aggregate welfare gain generated by the introduction of these routes. There is, however, interdependence between the effects of adding different routes. For estimating the effects of introducing routes between US CBSAs and Chinese prefectures, I consider a list of counterfactuals connecting each observed US-China direct route during 2004-2013 one by one. In this way, I can back out the value of each route from the model ignoring the interdependence.

Several insights can be obtained from the results. First, the connections from Seattle have the highest impacts on travel time and therefore largest effects on firm entry and welfare. This is because Seattle is far away from any of the already-connected gateway airports. Places connected to Seattle expect high time reduction from the direct connection between Seattle and China. On the contrary, since New York has already been

**Figure 7:** Welfare Gain and Efficiency of Each Route




Welfare Gain of Each Route




Welfare Efficiency of Each Route

Note: This figure reports welfare gains and efficient of each route for nine counterfactuals respectively for in the top graph and the bottom one. In each counterfactual, I introduce the particular route specified in the x-axis only. The welfare gains come from Equation 16. The efficiency is measured as the welfare gain per minute. For counterfactual calculation, I use $\sigma = 5$ from Broda and Weinstein (2006) and $\{\alpha^k\}_{k=1}^{S}$ from the IO Table.

connected to Beijing, the New York - Shanghai route does not change the travel time much. This is also the case for Newark which is very close to New York.

However, when we look at the efficiency of each route by comparing welfare gains per minute of travel time reduction. The picture becomes very different. In the last column of Table A.14, I show that New York - Shanghai and Newark - Shanghai are actually the two most valuable routes with the highest welfare gains per minute. This pattern is better illustrated in Figure 7. If travel time reductions have to be allocated among routes, these two routes should have the highest priority for maximizing welfare gains. Furthermore, in fact, all connections to Shanghai are more valuable than connections to Beijing because Shanghai is much closer to the most productive Chinese prefectures in Yangtze River Delta.

# 7 Conclusions

This paper provides new theory and evidence on whether and how international travel time reductions affect firm creation. Using a measure of the travel time between US CBSAs and Chinese prefectures, I causally estimate the effects of travel time reductions to China on firm creation in US cities with a novel re-centered IV. The reduced-form estimation cannot account for the within-China supplier presence heterogeneity, which is shown, using empirical evidence, to be important. To address this problem, I construct a quantitative spatial model to show that there would be a 1.7% increase in firm creation if we were to introduce the observed US-China routes while keeping the US domestic travel time network unchanged. Supplier presence heterogeneity across Chinese prefectures drives 42% of the increase because of assortative matching between time reductions and supplier presence.

The findings of this paper have important real-world implications, especially following the global pandemic. The global COVID-19 crisis provoked concerns about possible de-globalization (Antràs, 2020) from the perspective of international trade. This paper suggests that the persistent disruption in international travel since 2019 could dampen entrepreneurs' incentives to found firms. Furthermore, the disruption could also widen the gap between entrepreneurs who can access the best suppliers locally and entrepreneurs who cannot. The presence of a local supply chain has become much more important for sustaining entrepreneurship than in the pre-pandemic era.

Based on this paper, future work along three directions would of particular interest. First, accessing global suppliers through international travel could also be important to firm innovation. Fort et al. (2020) show that the proximity between production and innovation leads to higher patenting rates of firms. But manufacturing is declining in the US over the last decades and production activities mostly happen in developing countries such as China. Therefore, US-China aviation network expansions could also benefit firm innovation activities in the US. Second, expansions of the US-China flight network could also benefit entrepreneurs in China, probably differently due to the distinct industry structure in China. Third, because of data limitations, this paper focuses on the direct effect of international travel time reductions and ignores any indirect spillovers. Estimating such potentially important spillovers could be a fruitful research avenue.

# References


**Agrawal, Ajay, Alberto Galasso, and Alexander Oettl.** 2017. "Roads and Innovation." *Review of Economics and Statistics*, 99(3): 417–434.

**Antràs, Pol.** 2020. "De-Globalisation? Global Value Chains in the Post-COVID-19 Age." National Bureau of Economic Research.

**Antras, Pol, Teresa C Fort, and Felix Tintelnot.** 2017. "The Margins of Global Sourcing: Theory and Evidence from US Firms." *American Economic Review*, 107(9): 2514–64.

**Bai, John Jianqiu, Wang Jin, and Sifan Zhou.** 2021. "Proximity and Knowledge Spillovers: Evidence from the Introduction of New Airline Routes." Working Paper.

**Banerjee, Abhijit, Esther Duflo, and Nancy Qian.** 2020. "On the Road: Access to Transportation Infrastructure and Economic Growth in China." *Journal of Development Economics*, 145: 102442.

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

**Baum-Snow, Nathaniel, and Fernando Ferreira.** 2015. "Causal Inference in Urban and Regional Economics." In *Handbook of regional and urban economics*. Vol. 5, 3–68. Elsevier.

**Baum-Snow, Nathaniel, Loren Brandt, J Vernon Henderson, Matthew A Turner, and Qinghua Zhang.** 2017. "Roads, Railroads, and Decentralization of Chinese Cities." *Review of Economics and Statistics*, 99(3): 435–448.

**Bernard, Andrew B, Andreas Moxnes, and Yukiko U Saito.** 2019. "Production Networks, Geography, and Firm Performance." *Journal of Political Economy*, 127(2): 639–688.

**Bernstein, Shai, Xavier Giroud, and Richard R Townsend.** 2016. "The Impact of Venture Capital Monitoring." *Journal of Finance*, 71(4): 1591–1622.

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

**Borusyak, Kirill, and Peter Hull.** 2020. "Non-Random Exposure to Exogenous Shocks: Theory and Applications." NBER Working Paper.

**Broda, Christian, and David E Weinstein.** 2006. "Globalization and the Gains from Variety." *Quarterly Journal of Economics*, 121(2): 541–585.

**Caliendo, Lorenzo, and Fernando Parro.** 2015. "Estimates of the Trade and Welfare Effects of NAFTA." *Review of Economic Studies*, 82(1): 1–44.

**Campante, Filipe, and David Yanagizawa-Drott.** 2018. "Long-Range Growth: Economic Development in the Global Network of Air Links." *Quarterly Journal of Economics*, 133(3): 1395–1458.

**Cavusgil, S Tamer, and Gary Knight.** 2015. "The Born Global Firm: An Entrepreneurial and Capabilities Perspective on Early and Rapid Internationalization." *Journal of International Business Studies*, 46(1): 3–16.

**Charnoz, Pauline, Claire Lelarge, and Corentin Trevien.** 2018. "Communication Costs and the Internal Organisation of Multi-Plant Businesses: Evidence from the Impact of the French High-Speed Rail." *Economic Journal*, 128(610): 949–994.

**Chatterji, Aaron, Edward Glaeser, and William Kerr.** 2014. "Clusters of Entrepreneurship and Innovation." *Innovation Policy and the Economy*, 14(1): 129–166.

**Chinitz, Benjamin.** 1961. "Contrasts in Agglomeration: New York and Pittsburgh." *American Economic Review*, 51(2): 279–289.

**Chow, Melissa C, Teresa C Fort, Christopher Goetz, Nathan Goldschlag, James Lawrence, Elisabeth Ruth Perlman, Martha Stinson, and T Kirk White.** 2021. "Redesigning the Longitudinal Business Database." NBER Working Paper.

**Chu, Yongqiang, Xuan Tian, and Wenyu Wang.** 2019. “Corporate Innovation along the Supply Chain.” *Management Science*, 65(6): 2445–2466.

**Cristea, Anca D.** 2011. “Buyer-Seller Relationships in International Trade: Evidence from US States’ Exports and Business-Class Travel.” *Journal of International Economics*, 84(2): 207–220.

**Da, Zhi, Umit G Gurun, Bin Li, and Mitch Warachka.** 2021. “Investment in a Smaller World: The Implications of Air Travel for Investors and Firms.” *Management Science*, 67(1): 417–435.

**De Chaisemartin, Clément, and Xavier D’Haultfoeuille.** 2022. “Two-Way Fixed Effects and Differences-in-Differences with Heterogeneous Treatment Effects: A Survey.” NBER Working Paper.

**Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda.** 2014. “The Role of Entrepreneurship in US Job Creation and Economic Dynamism.” *Journal of Economic Perspectives*, 28(3): 3–24.

**Donaldson, Dave, and Richard Hornbeck.** 2016. “Railroads and American Economic Growth: A “Market Access” Approach.” *Quarterly Journal of Economics*, 131(2): 799–858.

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

**Duranton, Gilles, Peter M Morrow, and Matthew A Turner.** 2014. “Roads and Trade: Evidence from the US.” *Review of Economic Studies*, 81(2): 681–724.

**Eaton, Jonathan, and Samuel Kortum.** 2002. “Technology, Geography, and Trade.” *Econometrica*, 70(5): 1741–1779.

**Faber, Benjamin.** 2014. “Trade Integration, Market Size, and Industrialization: Evidence from China’s National Trunk Highway System.” *Review of Economic Studies*, 81(3): 1046–1070.

**Fernández-Val, Iván, and Martin Weidner.** 2016. “Individual and Time Effects in Nonlinear Panel Models with Large N, T.” *Journal of Econometrics*, 192(1): 291–312.

**Fort, Teresa C, Wolfgang Keller, Peter K Schott, Stephen Yeaple, and Nikolas Zolas.** 2020. “Colocation of Production and Innovation: Evidence from the United States.”

**Ghani, Ejaz, Arti Grover Goswami, and William R Kerr.** 2016. "Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing." *Economic Journal*, 126(591): 317–357.

**Giroud, Xavier.** 2013. "Proximity and Investment: Evidence from Plant-Level Data." *Quarterly Journal of Economics*, 128(2): 861–915.

**Glaeser, Edward L, and William R Kerr.** 2009. "Local Industrial Conditions and Entrepreneurship: How Much of the Spatial Distribution Can We Explain?" *Journal of Economics & Management Strategy*, 18(3): 623–663.

**Glaeser, Edward L, Stuart S Rosenthal, and William C Strange.** 2010. "Urban Economics and Entrepreneurship." *Journal of Urban Economics*, 67(1): 1–14.

**Hirano, Keisuke, and Guido W Imbens.** 2004. "The Propensity Score with Continuous Treatments." *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives: An Essential Journey with Donald Rubin's Statistical Family*, 73–84.

**Isenberg, Daniel J.** 2008. "The Global Entrepreneur." *Harvard Business Review*, 86(12): 107–111.

**Jacobs, Jane.** 1969. *The Economy of Cities.* Random House.

**Jedwab, Remi, Edward Kerby, and Alexander Moradi.** 2017. "History, Path Dependence and Development: Evidence from Colonial Railways, Settlers and Cities in Kenya." *Economic Journal*, 127(603): 1467–1494.

**Johnson, Robert C, and Guillermo Noguera.** 2012. "Fragmentation and Trade in Value Added over Four Decades." NBER Working Paper.

**Knight, Gary, and S Tamar Cavusgil.** 2004. "Innovation, Organizational Capabilities, and the Born-Global Firm." *Journal of International Business Studies*, 35(2): 124–141.

**Levchenko, Andrei A.** 2007. "Institutional Quality and International Trade." *Review of Economic Studies*, 74(3): 791–819.

**Liao, Steven, In Song Kim, Sayumi Miyano, and Hao Zhang.** 2021. "Product Concordance." *available at the Comprehensive R Archive Network (CRAN)*, http://CRAN.R-project.org/package=concordance.

**Lin, Wei, and Jeffrey M Wooldridge.** 2019. "Testing and Correcting for Endogeneity in Nonlinear Unobserved Effects Models." In *Panel Data Econometrics*. 21–43. Elsevier.

**Lin, Yatang.** 2017. "Travel Costs and Urban Specialization Patterns: Evidence from China's High Speed Railway System." *Journal of Urban Economics*, 98: 98–123.

**Martincus, Christian Volpe, Jerónimo Carballo, and Ana Cusolito.** 2017. "Roads, Exports and Employment: Evidence from a Developing Country." *Journal of Development Economics*, 125: 21–39.

**McDougall, Patricia Phillips, and Benjamin M Oviatt.** 2000. "International Entrepreneurship: the Intersection of Two Research Paths." *Academy of Management Journal*, 43(5): 902–906.

**McDougall, Patricia Phillips, Scott Shane, and Benjamin M Oviatt.** 1994. "Explaining the Formation of International New Ventures: The Limits of Theories from International Business Research." *Journal of Business Venturing*, 9(6): 469–487.

**Melitz, Marc J.** 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica*, 71(6): 1695–1725.

**Nunn, Nathan.** 2007. "Relationship-Specificity, Incomplete Contracts, and the Pattern of Trade." *Quarterly Journal of Economics*, 122(2): 569–600.

**Oviatt, Benjamin M, and Patricia Phillips McDougall.** 2005. "Toward a Theory of International New Ventures." *Journal of International Business Studies*, 36(1): 29–41.

**Pauly, Stefan, and Fernando Stipanicic.** 2021. "The Creation and Diffusion of Knowledge: Evidence from the Jet Age." Working Paper.

**Rauch, James E.** 1999. "Networks versus Markets in International Trade." *Journal of International Economics*, 48(1): 7–35.

**Rennie, Michael W.** 1993. "Born Global." *McKinsey Quarterly*, 45–53.

**Rosenthal, Stuart S, and William C Strange.** 2010. "Small Establishments/Big Effects: Agglomeration, Industrial Organization, and Entrepreneurship." In *Agglomeration Economics*. 277–302. University of Chicago Press.

**Startz, Meredith.** 2016. "The Value of Face-to-Face: Search and Contracting Problems in Nigerian Trade." Working Paper.

**Vernon, Raymond.** 1960. "Metropolis 1985." In *Metropolis 1985*. Harvard University Press.

# Online Appendix

## Contents

## A.1 Detailed Procedure for Constructing Travel Time Networks

This section presents detailed steps in constructing the travel time networks on US CBSA centroids, US airports, Chinese gateway airports and Chinese prefecture centroids over time and calculating the bilateral minimal travel time between US CBSA centroids and Chinese prefecture centroids. I start from four data sources and one routing tool:

- US T100 segment data from the website of BTS[1]. I get a sample of observed durations on connections from this data which will be used later for duration imputations.
- US ODS coupon data also from the website of BTS[2]. I mainly get the availabilities of US domestic connections over time from this data.
- US airports data provided by BTS[3]. Coordinates of airports come from this data. Noticed that same airport could appear multiple times in this data with different but very close coordinates as airports are usually very spacious. I use the average latitudes and longitudes as coordinates for these airports.
- 2013 shapefile of US CBSAs[4]. I calculate coordinates of CBSA centroids from this shapefile. I restrict my analysis to the 48 contiguous states in this paper.
- The tool used in getting the road driving time from CBSA centroids to US airports is the OpenStreetMap Routing Machine (OSRM). The tool is the open-source version of Google Map and can process millions of requests for free. For using the tool, one need to install the local API of OSRM on computer and extract map data of US[5]. The calculation, as using the Google Map, depends not only on geographic distances but also on road systems recorded in the map.

With these data, I proceed in steps as following:

1. **Direct flight durations between airports.** I get all direct flights from 1990 to 2019 operated by all carriers (domestic and international) at month level from US

[1] The data can be downloaded from *All Carrier Statistics (Form 41 Traffic) - All Carriers* on page https://www.transtats.bts.gov/DataIndex.asp. The data start from 1990 at month level and I use data from 1990 to 2019 for avoiding the impact of the global covid pandemic starting from 2020.

[2] The data can be downloaded from *Airline Origin and Destination Survey (DB1B)* on the same page. This data start from 1993 at quarter level and I use data from 1993 to 2019.

[3] This data can be downloaded from the *Aviation Support Tables* on the same page

[4] The shapefile is retrieved from https://www2.census.gov/geo/tiger/TIGER2013/CBSA/.

[5] The map data is downloaded from https://download.geofabrik.de/north-america.html.

T100 segment data. Then I exclude those non-passenger direct flights by keeping only flights with class F, with passenger aircraft configuration and with aircraft group suitable for commercial flights (for example, flight operated by helicopters are unlikely to be commercial passenger flights). Then I impute missing duration for direct flights with zero ramp-to-ramp time and drop direct flights with zero departures in the month. For avoiding outliers, I also trim the bottom 0.05% and top 99.95% of the observed flight durations. Then I conclude from this dataset that durations do not deviate much over time for the same direct flight route and are perfectly predicted by great circle distances calculated from airport coordinates. Additionally, I got an estimate for the speed of commercial passenger flight. With that estimate, I obtain imputed durations for all available direct flights by multiplying the speed estimate[6] with great circle distance on each segment.

2. **Domestic direct flight availability between US airports.** T100 segment data provide me both domestic and international direct flights. However, there is a change in reporting standard in 2002. Before 2002, small regional airlines in US are not required to file Form 41. So I will only use international flight information in T100 segment data. The domestic direct flight availability in US can be extracted from the ODS. Random 10% of the flight tickets sold by US airlines are recorded and aggregated into quarterly data in ODS. And the coupon version of ODS lists all direct flights from the ticket information. Therefore I can get whether direct flight routes existed or not in particular quarter from the ODS coupon data. And there is no change in reporting standard for ODS which is compiled from different sources (DB1B Form) than the T100 segment data (Form 41). For including only meaningful commercial flights, I exclude those segments with origin and destination the same and with quarterly volume of passengers less than 600.

3. **International direct flight availability between US and China.** International direct flights are also obtained from US T100 segment data. Carriers in this market are not small regional airlines so I don't need to worry about the change in reporting standard problem. For getting relevant international flights between US and China, I drop temporary flights with monthly departures less than 4 (not even a weekly flight)[7] and short-lived flights with operating time length less than 6 months before 2019 which is the end year of my sampling period. Another thing I do is to match origin airports to destination airports to exclude one-way flights which are most likely also charter flights. Then I group data at quarter level to

[6] I in fact got one over speed as average duration divided by distances.

[7] Many charter flights are dropped here.

match with domestic direct flights obtained from ODS data. I also drop international flights with less than 600 quarterly passengers to achieve consistency with sampling restriction used in the ODS. Finally, appending these international direct flights with domestic direct flights and matching them with the imputed flight durations give us a panel of direct flights with both time-invariant duration and time-variant availabilities observed.

4. **Travel time between CBSA centroids and US Airports.** There are 499 US airports ever serving 909 CBSAs in the final data of direct flights. And the number of direct flight routes did change over time as seen in Figure A.4. For covering US CBSAs with no airports and arriving at a balanced panel for analysis to avoid selection bias, I need to also measure travel time between CBSA centroids and US airports. Passengers in CBSAs without airports can travel through airports in other CBSAs to China. Travel time from CBSA centroids to airports is calculated using OSRM. I map the coordinates of US airports in the airport data and the centroids of CBSAs in the shapefile of 2013 delineation of US CBSAs to OSRM and calculate then the current road driving time between any airport and any CBSA centroid accounting for existing road system. Since road network did not change much since 1970s, the road driving time I calculate from OSM would be a proper measure for travel time between CBSA centroids and US airports through my sampling period 1993-2019[8].

5. **Travel time within China.** As an initial step, I focus on 255 prefectures by excluding HK, Macau, Taiwan, minority provinces and minority prefectures. Then I calculate first the distances from prefecture centroids to gateway airports. Following Bai, Jin and Zhou (2021), due to data limitation, I measure travel time between prefecture centroids and gateway airports by distances over a constant speed of 100 km/h. The travel time between gateway airports are computed with great circle distances and same flight speed as before giving the bilateral air connections between these gateway airports already existed before 2004.

6. **Quarterly Panel of travel time networks over time.** By appending the direct flights and the CBSA-to-airport connections, I can construct the travel time networks on CBSA centroids, the US airports, the Chinese gateway airport and the Chinese prefecture centroids. An important feature of this procedure is that I get a panel of direct flight existences at quarter level while the durations of flights

[8]It is insufficient to have traveling time between one CBSA centroid and the CBSA's closet airport because for some CBSAs, traveling through airports other than the closet one might be faster for going to China.

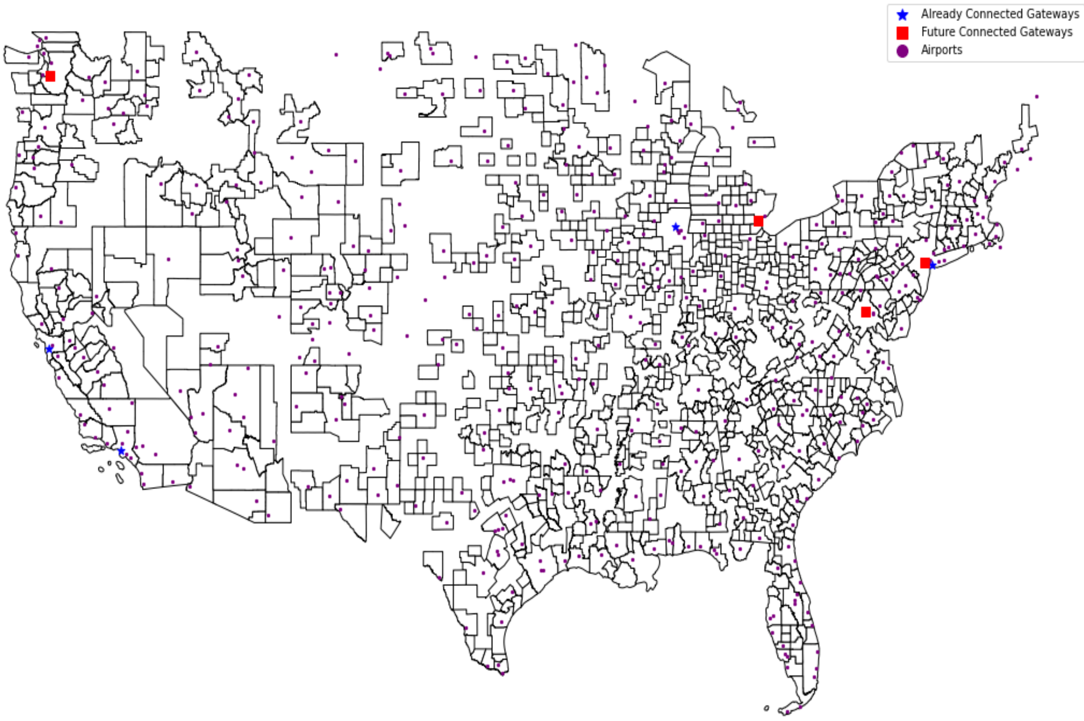
are time-invariant. Besides, the existences of CBSA-to-airport connections and prefecture-to-airport connections are also time-invariant by definition. Therefore the described procedure above gives me a sample of travel time networks at quarter level during 1993-2019 where the variation over time solely comes from the change of the network structure determined by the existences of the international and domestic direct flight routes.

7. **Search for the fastest route and get minimal travel time.** I can search for the fastest route on the constructed travel time networks from any CBSA centroid in US to any prefecture centroid in China by Dijkstra algorithm. Here I made two assumptions: (1) one hour is spend at any stop for taking flight following Giroud (2013); (2) passengers make optimal choice by searching for the fastest route. There exists one concern on this approach that the time spent on waiting for domestic flights and the international flights could be very different from each other. Noticed that this would not be relevant for the fastest route. Fastest routes would not be affected as every route for traveling from one US CBSA centroid to one Chinese prefecture has exactly one stop in a US gateway airport for waiting for international flights.

8. **US-China travel time network.** The final result for analysis is a panel of 231795 ($909 \times 255$) pairs of US CBSAs and Chinese prefectures with minimal travel time and optimal routes observed at quarter level during 1993-2019.

## A.2 US Airports and Chinese Prefectures

**Figure A.1:** Airports in the US



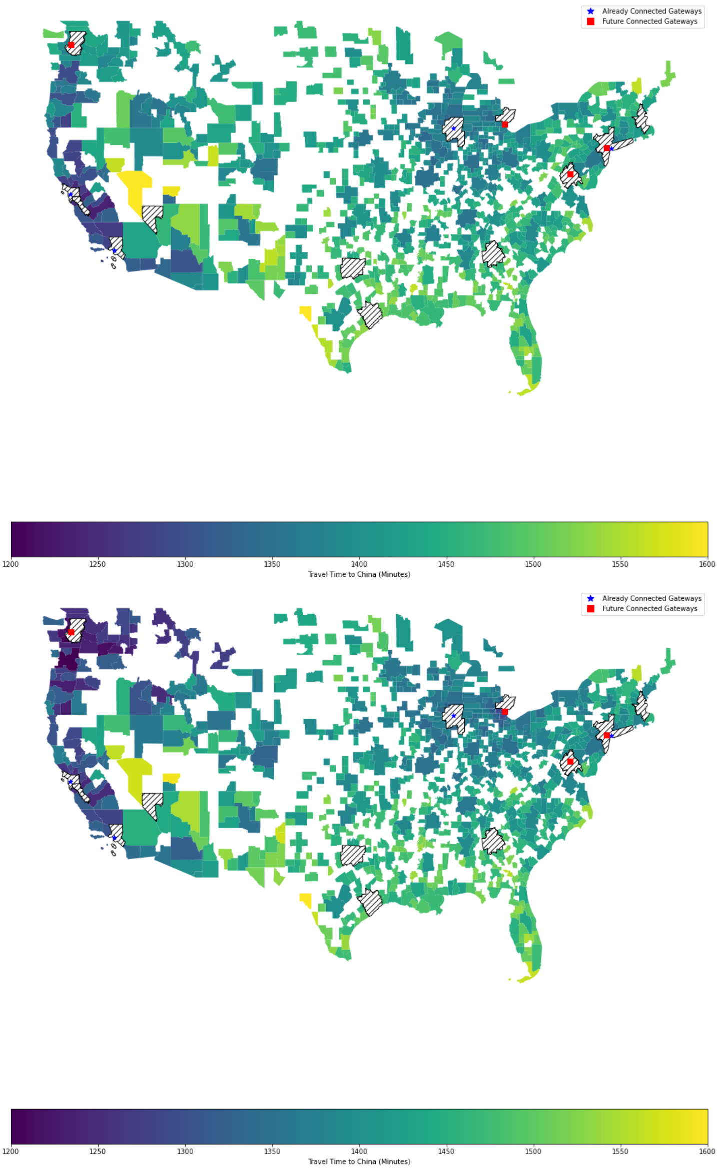

Note: This figure shows the locations of airports which in the data that is constructed for measuring travel time between US CBSAs and Chinese prefectures using the T100 segment data and the ODS data. The boundary delineates the CBSAs and every purple dot represents an airport. Besides, I specifically show the positions of the already-connected airports before 2004 by blue stars and the future-connected airports during 2004-2013 by red squares.

**Figure A.2:** The Included 255 Prefectures

Note: This figure shows the 255 prefectures included in the analysis of this paper in the shaded area of purple color on a map of China.

## A.3 Travel Time to China in 2004 and 2013

**Figure A.3:** Travel Time to China in 2004 and 2013




Note: This figure separately shows the average travel time from each US CBSA to all Chinese prefectures in 2004 in the top plot and in 2013 in the bottom plot.

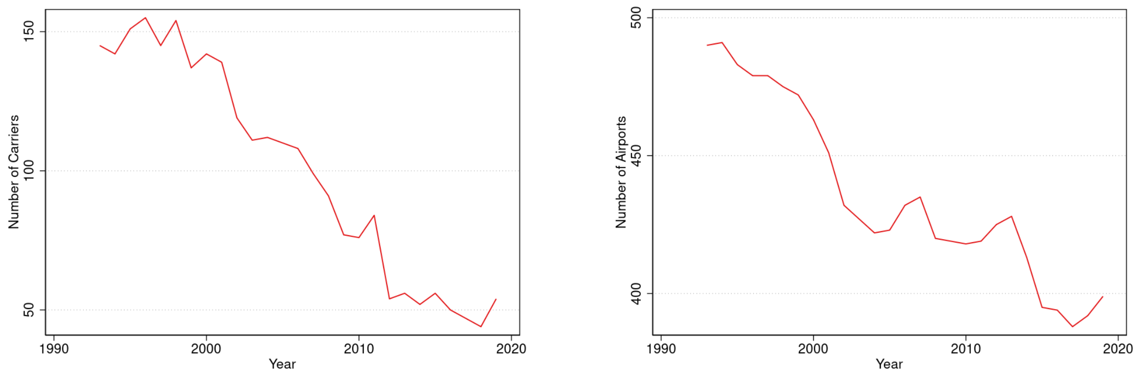
## A.4 Compare the T100 Segment Data and the ODS Data

**Figure A.4:** Spikes in the T100 Segment Data


T100 T100

Note: This figure shows the number of carriers and the number of airports in the T100 segment data over time. We see immediately that there are spikes around 2001 in both the left graph and the right graph.

**Figure A.5:** No Spikes in the ODS Data

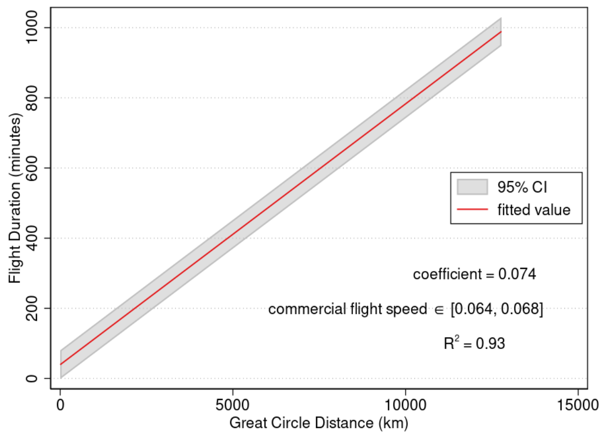


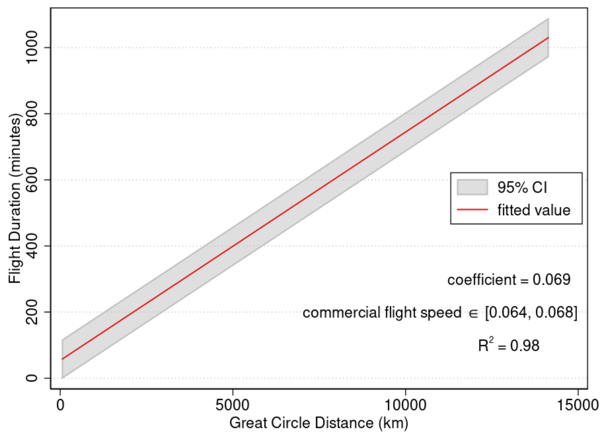

ODS ODS

Note: This figure shows the number of carriers and the number of airports in the ODS data over time. We see immediately that there are no spikes around 2001 in both the left graph and the right graph. In fact, the number of carriers and the number airports decline over time. This is consistent with the trends of the US domestic flight market.

## A.5 Predict Flight Durations by Distances

**Figure A.6:** Predict Durations by Distances: Domestic and International




Note: This figure shows that linear regressions on great circle distances fit the flight durations almost perfectly for both the domestic flights and the international flights in the T100 segment data. The implied flight speeds of domestic flights and international flights are indistinguishable and close to the common commercial flight speed.

## A.6 Compare the Infogroup Data and the BDS Data

**Table A.1:** Compare Data at the CBSA-Year Level

| | BDS Data | | | |
|---|---|---|---|---|
| | Number of New Firms | | Log Number of New Firms | |
| | (1) | (2) | (3) | (4) |
| Number of New Firms in Infogroup | 1.502 | 1.532 | | |
| | (0.066) | (0.138) | | |
| Log Number of New Firms in Infogroup | | | 0.610 | 0.829 |
| | | | (0.003) | (0.008) |
| $R^2$ | 0.578 | 0.588 | 0.644 | 0.862 |
| N | 20907 | 20907 | 20907 | 20907 |
| Year FE | N | Y | N | Y |
| Cluster at CBSA | N | Y | N | Y |

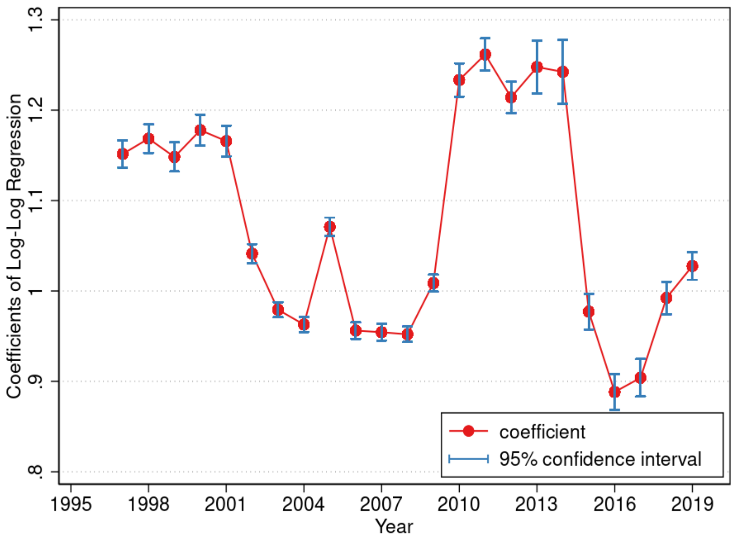
Note: This table reports regression coefficients and standard errors in parenthesis. I use the BDS data in these regressions. I use the inverse hyperbolic sine transformation to deal with zeros in log outcomes.

**Table A.2:** Compare Data at the CBSA-year-industry (2 digit) Level

| | BDS Data | | | |
|---|---|---|---|---|
| | Number of New Firms | | Log Number of New Firms | |
| | OLS | OLS | OLS | OLS |
| | (1) | (2) | (3) | (4) |
| Number of New Firms in Infogroup | 1.42 | 1.45 | | |
| | (0.03) | (0.14) | | |
| Log Number of New Firms in Infogroup | | | 0.91 | 1.03 |
| | | | (0.001) | (0.00) |
| $R^2$ | 0.52 | 0.52 | 0.58 | 0.66 |
| N | 312505 | 312505 | 312505 | 312505 |
| Year FE | N | Y | N | Y |
| Cluster at CBSA | N | Y | N | Y |

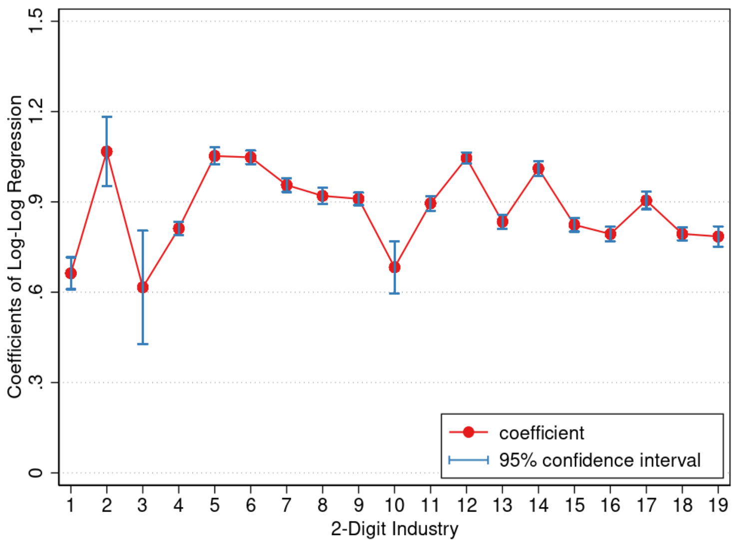
Note: This table reports regression coefficients and standard errors in parenthesis. I use the BDS data in these regressions. I use the inverse hyperbolic sine transformation to deal with zeros in log outcomes.

**Figure A.7:** Compare the Infogroup Data and the BDS Data over Time




Note: This figure shows the coefficients and their 95% confidence intervals from regressing log firm creation in the Infogroup data on the log firm creation in the BDS data at the CBSA-industry (2-digit) level year by year. The coefficients are close to one in all years. This indicates that the Infogroup data is a representative sample of the administrative BDS data at the CBSA-industry (2-digit) level.

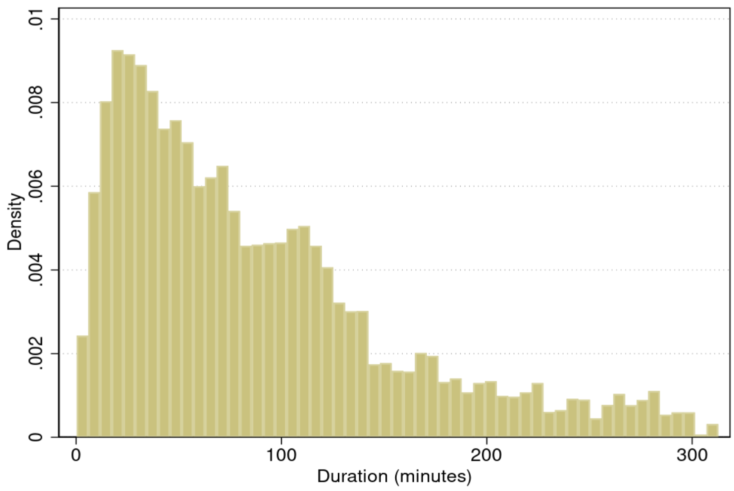
**Figure A.8:** Compare the Infogroup Data and the BDS Data across Industries




Note: This figure shows the coefficients and their confidence intervals from regressing log firm creation in the Infogroup data on the log firm creation in the BDS data at the CBSA-year level industry by industry (2-digit). The coefficients are close to one in most industries. This indicates that the Infogroup data is a representative sample compared of the administrative BDS data at the CBSA-year level for most industries. However, we do see that in certain industries the coefficients are far away from one. This concern however is mitigated in the estimation at the CBSA-industry (6-digit) level controlling for industry fixed effects as we compare same industries across CBSAs in this specification.

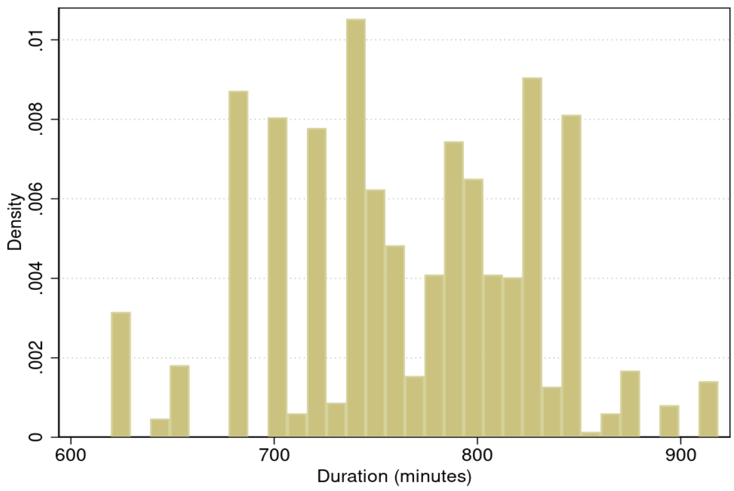
## A.7 Distributions of Direct Flight Durations

**Figure A.9:** Duration Distribution of US Domestic Direct Flights




Note: This figure plots a histogram of all US domestic flight durations observed in the flight data constructed from the T100 segment data and the ODS data. This histogram covers all flights in the data spanning years from 1993 to 2019.

**Figure A.10:** Duration Distribution of US-China International Direct Flights




Note: This figure plots a histogram of all US-China international flight durations observed in the flight data constructed from the T100 segment data. This histogram covers all flights in the data spanning years from 1993 to 2019.

## A.8 Summary Statistics

In Table A.3, I summarize long run differences of two key variables over the ten-year period 2004-2013 at the CBSA level with the BDS data. The first difference I look at is the reduction of travel time to China. The second one is the change of firm creation. Column (1) shows that on average the travel time to China reduces by 0.288 hours with a large 0.611 hours' standard deviation. The firm creation in US CBSAs declines on average by 82.013 with standard deviation 190.235. The documented decline of US entrepreneurship is consistent with the literature (e.g., Decker et al. (2014)).

The top panel of Table A.3 shows that higher travel time reduction is associated with lower entrepreneurship decline during 2004-2013. I split these CBSAs into two groups: the ones with positive time reductions and the ones without. Then I find that the 2004-2013 long difference in firm creation is larger for CBSAs with positive time reductions. This is still true when we control for city size through dividing the firm entry rate by city size. The pattern suggests that the reduced travel time to China over the decade could have positive impact on firm creation across US CBSAs.

The bottom panel of Table A.3 informs us that CBSAs' exposure to the introduction of US-China nonstop international flight routes are correlated with the travel time reductions and the changes of firm creation at the same time. I define cities with high exposure as those with high travel time to the already-connected US gateways to China in the baseline year 2004 because if CBSAs are far away to these gateways, the travel time from them to China should expect to decrease more when new US-China routes are introduced during 2004-2013. The 896 CBSAs are then split into two groups: the ones with above median exposure and the ones with below median exposure. In columns (2) and (3), the travel time reductions are indeed higher in the CBSAs with high exposure. The long differences of firm creation are also higher in the CBSAs with high exposure.

This observation, however, is counterintuitive. The places with high exposure are closer to the already-connected gateway airports. They should be a selective sample of CBSAs receiving positive spillovers from the nearby connected gateways. I turn to again control for city size and check whether the pattern changes. The firm creation changes divided by city sizes are indeed larger in the CBSAs with low exposure. In another word, once we control for city size, CBSAs' higher exposure implies both larger travel time reduction and *smaller* change of firm creation during the ten-year period.

This suggests the existence of negative selection. The introduction of new US-China nonstop international routes has smaller effect on the travel time to China from the CBSAs closer to the already-connected gateway airports such as the JFK in New York.

At the same time, these CBSAs are also more prosperous and entrepreneurial, exactly because they have already been well-connected to China. Due to this problem of negative selection, we could get downward bias in estimation without a valid instrument.

I then turn to use the Infogroup data for its detailed industry classifications which can be utilized to compare industries with different characteristics. The unit of analysis here is CBSA-industry where the industry is at the NAICS 6-digit level for matching with the IO table[9]. I distinguish industries by two dimensions: (1) supplier intensity which measures the need to sourcing inputs from suppliers in many industries (2) customer intensity which measures the need to selling outputs to customers in many industries.

In Table A.4, I classify industries as the ones with above and below median supplier intensities and the ones with above and below median customer intensities. In columns (2) and (3), I compare the industries with high versus low supplier intensities. In columns (5) and (6), I compare the industries with high versus low customer intensities.

The top panel suggests that the changes of firm creation over the ten-year period are larger in the industries with high supplier intensities than the industries with low supplier intensities. The bottom panel indicates that the long differences in firm creation are also larger in the industries with high customer intensities than the industries with low customer intensities. The high-low gap however is more pronounced when we compare the industries of different supplier intensities than when we compare the industries with different customer intensities.

These results indicate that it is more difficult for entrepreneurs to enter into the industries with high supplier intensities or customer intensities because of potential higher entry barriers. The expansion of the US-China aviation network could therefore help US entrepreneurs in the industries with high supplier intensities or customer intensities more, by reducing the entry costs of getting suppliers or customers.

These findings motivate the analyses in Section 3 at the disaggregated city-industry level for taking into account the industry heterogeneity in firm creation. Moreover, the results in Table A.4 ask for identifying the heterogeneous effects of the reductions in travel time to China on firm creation in the industries with different supplier/customer intensities in Section 5. Besides, I should consider the input-output structure between industries across locations in the quantitative spatial model in Section 6 for evaluating eventually the aggregate impact of US-China aviation network expansions.

[9]Table A.5 presents at the CBSA-industry level exactly the same pattern as Table A.3.

**Table A.3:** Compare CBSAs with the BDS Data

| | CBSA Panel (BDS) | | |
|---|---|---|---|
| | All | With Reduction | Without Reduction |
| 2004-2013 Long Difference | (1) | (2) | (3) |
| Travel Time (Hours) | 0.288 | 0.491 | -0.146 |
| | (0.611) | (0.632) | (0.206) |
| Firm Creation | -82.013 | -73.467 | -100.241 |
| | (190.2353) | (161.233) | (240.018) |
| Firm Creation/City Size (10,000) | -12.381 | -11.725 | -13.784 |
| | (13.384) | (13.037) | (14.019) |
| Observations | 896 | 286 | 610 |
| | CBSA Panel (BDS) | | |
| | All | High Exposure | Low Exposure |
| 2004-2013 Long Difference | (4) | (5) | (6) |
| Travel Time (Hours) | 0.288 | 0.402 | 0.174 |
| | (0.611) | (0.765) | (0.370) |
| Firm Creation | -82.013 | -54.862 | -109.165 |
| | (190.2353) | (144.585) | (223.780) |
| Firm Creation/City Size (10,000) | -12.381 | -13.628 | -11.136 |
| | (13.384) | (16,831) | (8.516) |
| Observations | 896 | 448 | 448 |

Note: The table summarizes the changes of two key variables studied in this paper during the period 2004-2013: the mean travel time from US CBSAs to China across all prefectures and the firm creation in US CBSAs. The CBSAs with reductions are defined as having positive time reductions during 2004-2013. The CBSAs with high exposure are those ones with above median travel time to the four already-connected US gateway airports to China: SFO (San Francisco International Airport), LAX (Los Angeles International Airport), ORD (Chicago O'Hare International Airport), and JFK in the baseline year 2004. The panel includes 896 US CBSAs and uses the BDS data.

**Table A.4:** Compare Industries with the Infogroup Data

| | CBSA-Industry Panel (Infogroup) | | |
|---|---|---|---|
| | All | High Supplier Intensity | Low Supplier Intensity |
| 2004-2013 Long Difference | (1) | (2) | (3) |
| Firm Creation | -0.258 | -0.305 | -0.176 |
| | (2.423) | (2.429) | (2.435) |
| Firm Creation/City Size (10,000) | -0.035 | -0.041 | -0.024 |
| | (0.212) | (0.224) | (0.189) |
| Observations | 820,736 | 533,120 | 287,616 |
| | CBSA-Industry Panel (Infogroup) | | |
| | All | High Customer Intensity | Low Customer Intensity |
| 2004-2013 Long Difference | (4) | (5) | (6) |
| Firm Creation | -0.258 | -0.290 | -0.211 |
| | (2.423) | (2.803) | (1.677) |
| Firm Creation/City Size (10,000) | -0.035 | -0.037 | -0.031 |
| | (0.212) | (0.222) | (0.196) |
| Observations | 820,736 | 503,885 | 315,935 |

Note: The table compares firm creation across US CBSA-industry pairs. The industry is at the 6-digit NAICS level. The supplier intensities and customer intensities are defined in Equation 1. Those high supplier (customer) intensities are above the median across all industries. The table uses the Infogroup data.

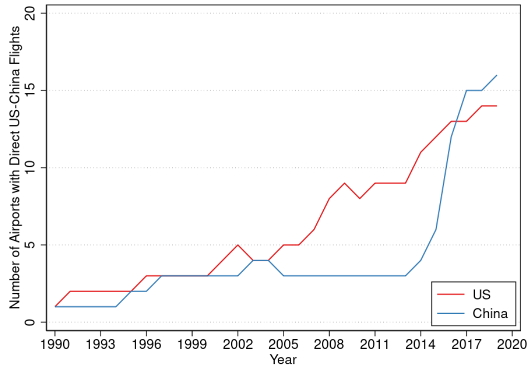
**Table A.5:** Summary Statistics at the CBSA-Industry Level with the Infogroup Data

| | CBSA-Industry Panel (Infogroup) | | |
|---|---|---|---|
| | All | With Reduction | Without Reduction |
| 2004-2013 Long Difference | (1) | (2) | (3) |
| Travel Time (Hours) | 0.288 | 0.491 | -0.146 |
| | (0.611) | (0.632) | (0.206) |
| Firm Creation | -0.258 | -0.220 | -0.340 |
| | (2.423) | (1.688) | (3.058) |
| Firm Creation/City Size (10,000) | -0.035 | -0.033 | -0.038 |
| | (0.212) | (0.204) | (0.228) |
| Observations | 827,800 | 263,978 | 563,030 |
| | CBSA-Industry Panel (Infogroup) | | |
| | All | High Exposure | Low Exposure |
| 2004-2013 Long Difference | (4) | (5) | (6) |
| Travel Time (Hours) | 0.288 | 0.491 | -0.146 |
| | (0.611) | (0.632) | (0.206) |
| Firm Creation | -0.258 | -0.169 | -0.348 |
| | (2.423) | (2.376) | (2.466) |
| Firm Creation/City Size (10,000) | -0.035 | -0.039 | -0.030 |
| | (0.212) | (0.249) | (0.166) |
| Observations | 827,800 | 413,504 | 413,504 |

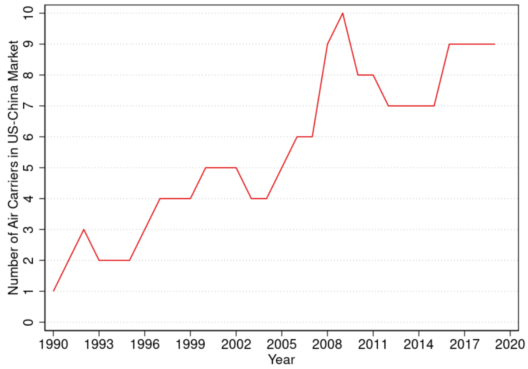
Note: The table summarizes the changes of two key variables studied in this paper during the period 2004-2013: the mean travel time from US CBSA to China across all prefectures and the firm creation in US CBSA. The CBSAs with reductions are defined as having positive time reductions during 2004-2013. CBSAs with high exposure are those ones with above median travel time to four already-connected US gateway airports to China: SFO, LAX, ORD, and JFK. The panel uses the BDS data.

## A.9 Trends in the US-China Flight Market

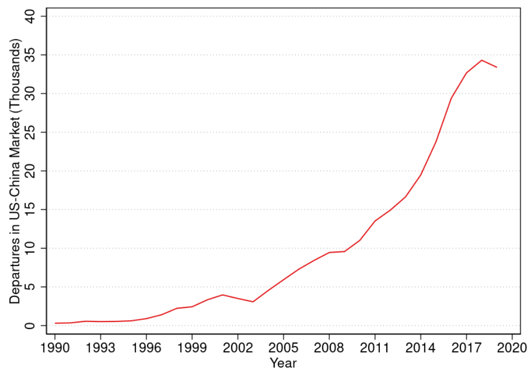
**Figure A.11:** Trends in US-China Flight Market



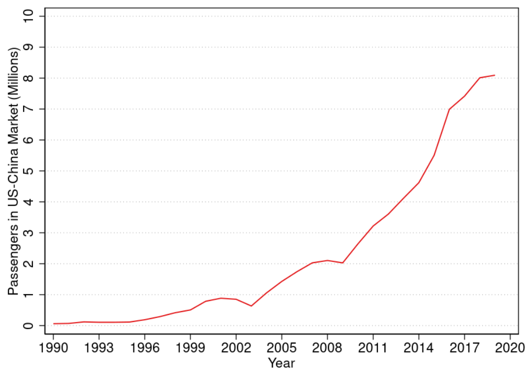

Gateways




Carriers




Departures (Thousands)




Passengers (Millions)

Note: This figure shows the trends in US-China flight market with four plots. Here I use the T100 segment data.

## A.10 Counterfactual Route Sequences Construction

The counterfactual route sequences used for constructing the re-centered IV come from the applications in five application cycles hosted by the Department of Transportation of the United States.

In 1979, US and China restored their diplomatic relation. Then in 1980, the two countries signed agreements on a set of issues regarding international trade and national security. One important part of the agreement was to restart the international flights between the two countries. The US-China passenger flight market, however, is highly regulated even now. The agreement in 1980 only allowed two airlines in US to operate two nonstop routes to China including both passenger flights and all-cargo flights. United Airline managed to get a route to operate nonstop passenger flights to China while FedEx got the other one to operate all-cargo flights. Northwest obtained one direct flight route to Beijing later in 1990s.

There have been three amendments to the 1980 agreement for expanding the quota on US-China international flight market: the 1999 amendment, the 2004 amendment, and the 2007 amendment. The two countries scheduled a 2010 amendment to further deregulate the US-China flight market. However, they never achieved the goal till today. The department of transportation hosted three application cycles for the quotas assigned in 2004 amendment because the quotas were allocated to frequencies airlines could use only since certain years. Therefore there have been five application cycles on the US side for allocating quotas to US airlines.

Notice I don't permute the routes operated by Chinese airlines as first I cannot find information on the application cycles on the China side and second the quotas on the China size are not binding at least before 2015 according to the observed number of direct routes operated by Chinese carriers in the data. These routes are kept as the same as the observed one in my counterfactual route sequences and won't affect the re-centered IV I construct at all.

I will show the procedure to construct my counterfactual route sequences with the 1999 application cycle as example[10]. The 1999 amendment allowed one more new airline to fly direct route to China. It can be either all-cargo or passenger. 17 frequencies starting in 2000 are allocated to the incumbent airlines while 10 additional frequencies starting in 2001 to all airlines. Therefore the Department of Transportation hosted two separate applications for the two types of frequencies.

[10]All the information on the amendments and the application cycles can be found on https://www.regulations.gov.

In the application for incumbent frequencies, FedEx applied for using all frequencies to operate all-cargo flights, United applied for operating SFO-PVG and SFO-PEK passenger nonstop routes, and Northwest applied for using part of the frequencies to operate DTW-PVG nonstop passenger route. The 17 frequencies are shared by all the three incumbent airlines in the decision by the Department of Transportation. Northwest was able to fly DTW-PVG while United was able to fly SFO-PEK and SFO-PVG in 2000.

I construct counterfactual based on the applications. There are four possible results: (1) FedEx got all the frequencies; (2) FedEx and Northwest shared the frequencies; (3) FedEx and United shared the frequencies; and (4) the three airlines shared the frequencies. New routes under these counterfactuals are: (1) none; (2) DTW-PVG in 2000; (3) SFO-PVG and SFO-PEK in 2000; and (4) DTW-PVG, SFO-PVG, and SFO-PEK in 2000.

In the application for frequencies starting in 2001, Northwest applied for 5 frequencies to operate DTW-PEK and DTW-PVG, Delta applied for 10 frequencies to operate JFK-PEK and PDX-PEK, American applied for 10 frequencies to operate ORD-PVG and ORD-PEK, United applied for 2 frequencies to operate SFO-PVG and ORD-PVG, Polar Air applied for 6 frequencies for all-cargo flights, FedEx applied for 8 frequencies for all-cargo flights, and UPS applied for 10 frequencies for all-cargo flights. UPS managed to become the fourth designated airline and got 6 frequencies. United got 2 frequencies. Northwest and FedEx each got one.

There are three possible results: (1) American became the new designated airline and shared the frequencies with incumbents; (2) Delta became the new designated airline and shared frequencies with incumbents; and (3) UPS/Polar Air became the new designated airline and share frequencies with incumbents. New routes under these counterfactuals are: (1) ORD-PEK and ORD-PVG in 2001; (2) ORD-PEK and JFK-PEK in 2001; and (3) ORD-PEK in 2001.

Notice that for generating close counterfactuals, I only consider gateways which ever existed as gateways to China. Therefore the PDX-PEK route which only appeared once in applications and was not selected is not permuted in counterfactuals. Similarly, in counterfactual, United would fly ORD-PEK instead of the proposed ORD-PVG because that's what it chose in reality.

I apply the same methodology as the one showcased above to the following application cycles. When the counterfactual routes already existed in previous application permutations, I exclude them from the possible counterfactuals. In the end, I got 238 distinct counterfactual route sequences for constructing the re-centered IV.

## A.11 Understand the Re-Centering

Figure A.12 illustrates this point with a simplified example. On the left panel, I plot a travel network in 2004 where points represent locations, lines represent connections, and the lengths of lines represent travel time. Only one gateway airport SFO has direct flights to China (CN). The CBSAs with high exposure (high), though being close to the gateway airport DC, have to transfer from SFO for traveling to China.

On the right panel, I plot the travel network in 2013 where I connect DC and China. The domestic network in 2013, in the spirit of the unadjusted IV, is the same as 2004. The CBSAs with high exposure then switch to travel from the nearby gateway airport DC to China. The introduction of direct flights between DC and China therefore saves travel time for the CBSAs of high exposure.

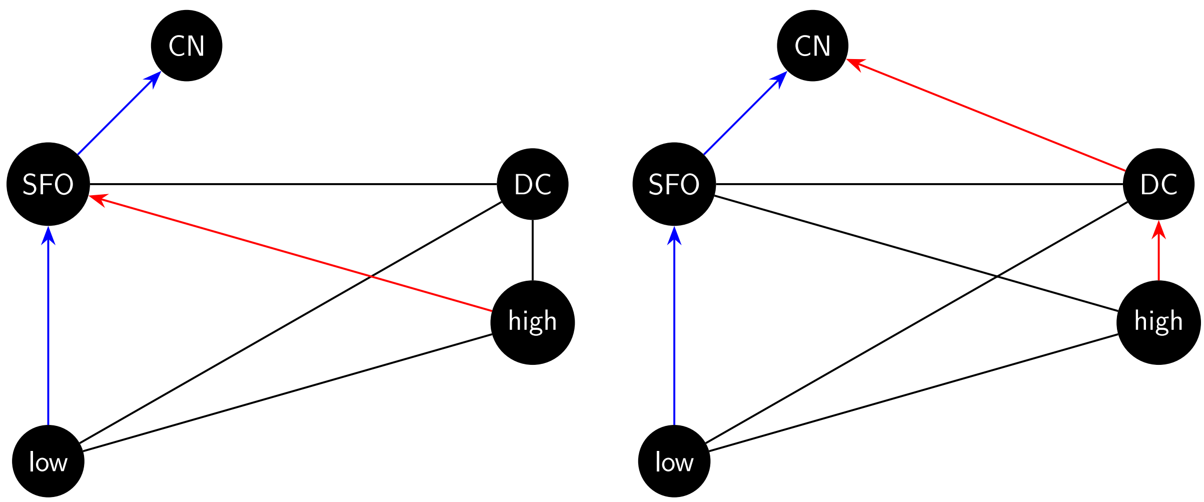
It however does not reduce travel time for the CBSAs with low exposure (low) because they are close to the already-connected gateway airports SFO and they won't switch to transfer through DC in 2013 for traveling to China. In another word, the CBSAs close to the already-connected gateway airports are less exposed while the CBSAs close to the future-connected gateway airports are more exposed to the US-China aviation network expansion.

The positions of these CBSAs, relative to the already-connected gateways and the future-connected gateways on the travel network, cannot be exogenous. Being close to the already-connected gateway airports implies more firm creation as seen in Table A.3 exactly because passengers' travel time to China from these CBSAs are smaller. In this example, *though we have fixed the domestic network and excluded the CBSAs with gateway airports*, the reductions in travel time to China are still positively correlated with the non-random exposure which in turn is negatively correlated with firm creation. This would lead to downward biases in estimation if we used the unadjusted IV for identification.

Figure A.13 and Figure A.14 illustrate the intuition of re-centering and why the re-centered IV can purge the biases associated with the omitted non-random exposure by two stylized counterfactuals. Following the same notations as Figure A.12, I plot 2004 network and 2013 network of the two counterfactuals in the two figures and compare them to the realized networks in Figure A.12.

Figure A.13 shows the first counterfactual relative to the realized networks in Figure A.12. I permute the future-connected gateway airport to be CF instead of DC in the 2013 network of counterfactual 1. In this counterfactual, the CBSAs with low exposure still won't switch to use the newly-connected gateway while the CBSAs with high expo-

**Figure A.12:** Example of Non-Random Exposure

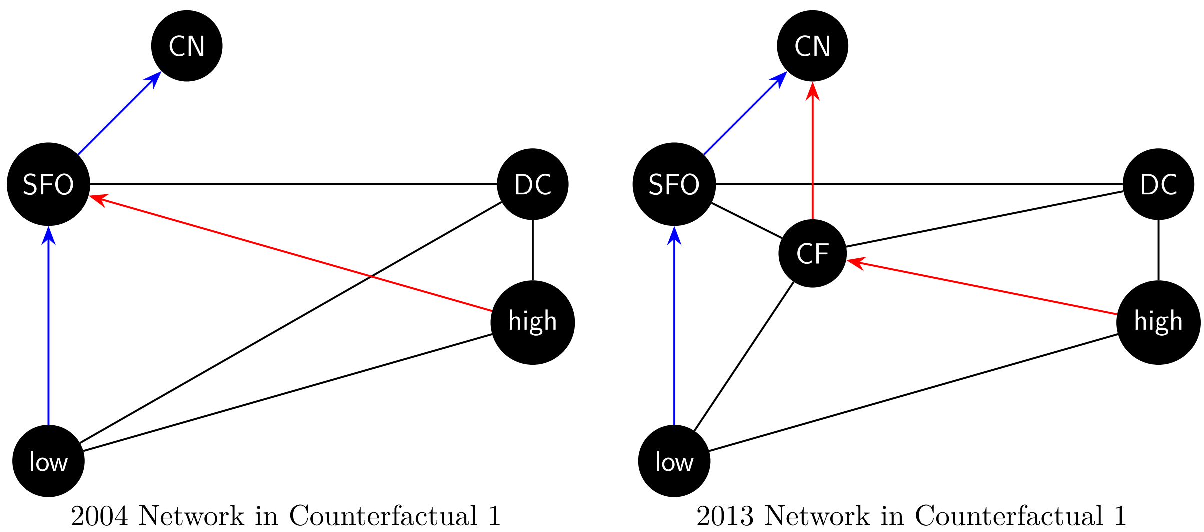



2004 Network

2013 Network

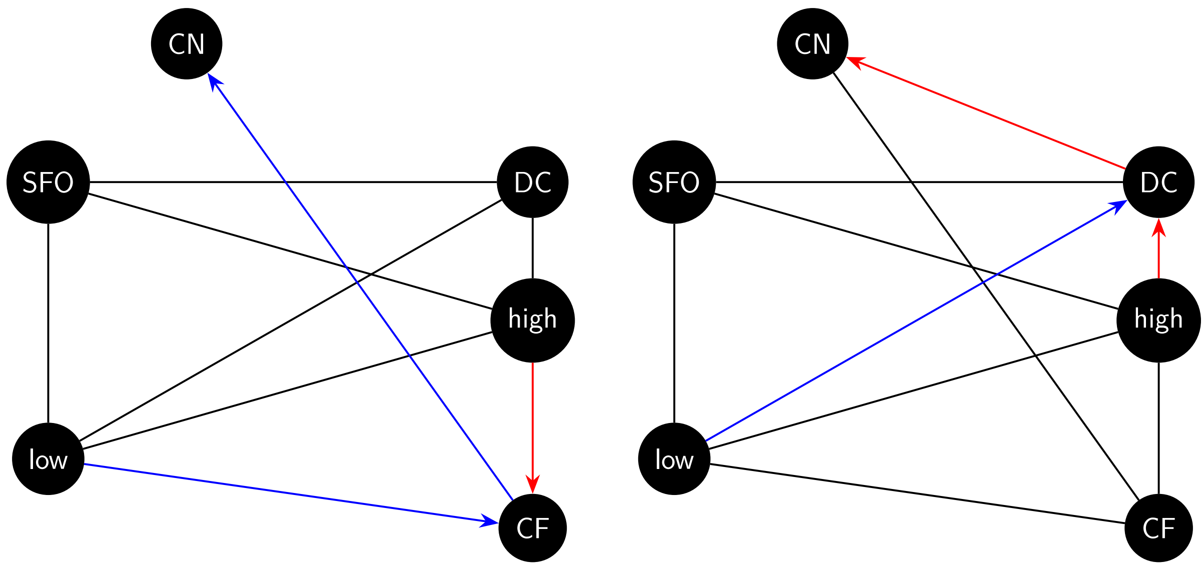
Note: This figure illustrates the intuition of non-random exposure problem in this paper's setting. Every node here represents a location. They are connected by lines which represent direct flights. The lengths of lines represent travel time. The colored arrow indicates the fastest route for traveling to CN which represents China. SFO and DC are two gateway airports in the US. SFO is in San Francisco and has been connected to China in 2004 while DC will be connected during 2004-2013. Low and high represent two types of CBSAs: the ones with low exposure and the ones with high exposure to the introduction of direct flights between DC and China. On the left panel, I show the network in 2004. Passengers in the CBSAs of both types have to transfer through the SFO airport for traveling to China. On the right panel, DC and China are connected on the network of 2013. Because the CBSAs of low type are close to the already-connected gateway airport SFO, they receive no time reductions during 2004-2013. At the same time, the CBSAs of high type are close to the newly connected gateway airports DC. They then receive positive time reductions from the introduction of direct flights between DC and China. This figure therefore shows that the exposure of CBSAs to the US-China aviation network expansion is not random. The exposure is determined by CBSAs' relative positions to the already-connected gateway airports and the future-connected gateway airports. The travel time reductions are hence correlated with the unobserved and uncontrollable complex economic geography which affects firm creation.

**Figure A.13:** Re-Centering Example 1




Note: This figure illustrates the intuition of re-centering. Every node here represents a location. They are connected by lines which represent direct flights. The lengths of lines represent travel time. The colored arrow indicates the fastest route for traveling to CN which represents China. SFO, DC, and CF are three gateway airports in the US. SFO is in San Francisco and has been connected to China in 2004 while *CF instead of DC* will be connected during 2004-2013. Low and high represent two types of CBSAs: the ones with low exposure and the ones with high exposure to the introduction of direct flights between DC and China. On the left panel, I show the network in 2004. Passengers in the CBSAs of both types have to transfer through the SFO airport for traveling to China. On the right panel, CF and China are connected on the network of 2013. Because the CBSAs of low type are close to the already-connected gateway airport SFO, they receive no time reductions during 2004-2013. At the same time, the CBSAs of high type are close to the newly connected gateway airports CF. They then receive positive time reductions from the introduction of direct flights between CF and China. Re-centering is to consider only the time reductions in Figure A.12 relative to this counterfactual. We then have zero travel time reductions for the CBSAs with low exposure and smaller travel time reductions for the CBSAs with high exposure. This figure therefore shows that re-centering can remove the non-random exposure correlated with CBSAs' positions relative to the future-connected gateway airports.

**Figure A.14:** Re-Centering Example 2




Note: This figure illustrates the intuition of re-centering. Every node here represents a location. They are connected by lines which represent direct flights. The length of lines represent travel time. The colored arrow indicates the fastest route for traveling to CN which represents China. SFO, DC, and CF are three gateway airports in the US. *CF instead of SFO* has been connected to China in 2004 while DC will be connected during 2004-2013. Low and high represent two types of CBSAs: the ones with low exposure and the ones with high exposure to the introduction of direct flights between DC and China. On the left panel, I show the network in 2004. Passengers in the CBSAs of both types have to transfer through the CF airport for traveling to China. On the right panel, DC and China are connected on the network of 2013. Because the CBSAs of both types are close to the newly-connected gateway airport DC relative to the already-connected gateway CF, they all receive positive time reductions during 2004-2013. Re-centering is to consider only the time reductions in Figure A.12 relative to this counterfactual. We then have positive instead of zero travel time reductions for the CBSAs with low exposure and smaller travel time reductions for the CBSAs with high exposure. This figure therefore shows that re-centering can remove the non-random exposure correlated with CBSAs' positions relative to the already-connected gateway airports.

sure switch to transfer from CF. By re-centering, I effectively only consider travel time reductions in Figure A.12 *relative* to the counterfactual situation in Figure A.13. The re-centered time reductions for the CBSAs with low exposure then would still be zero in this case.

On the other hand, the re-centered time reductions for the CBSAs with high exposure would be smaller. The gap in travel time reductions between the two types of CBSAs would then decrease. Re-centering therefore removes the non-random exposure originated from CBSAs' positions relative to the future-connected gateway airports when I permute the future-connected gateways in counterfactuals.

Similarly, Figure A.14 shows the second counterfactual where I permute the already-connected gateway airport to be CF instead of SFO in the 2004 network. The CBSAs with low exposure now switch to use the newly-connected gateway and the CBSAs with high exposure also switch to transfer from CF. After re-centering, the residual time reductions for the CBSAs with low exposure would become higher in this case.

On the other hand, the residual time reductions for the CBSAs with high exposure would become smaller. Re-centering then increases the gap in travel time reductions between the two types of CBSAs. Therefore I can remove the non-random exposure originated from CBSAs' positions relative to the already-connected gateway airports if I permute the already-connected gateways in counterfactuals.

These two examples show that re-centering the unadjusted IV mitigates the concerns about biases on the supply side. The average across all counterfactuals captures the variation in travel time reductions which can be expected from the observed fixed domestic network in the baseline year 2004 no matter which nonstop US-China route will be connected during 2004-2013. By subtracting it from the unadjusted IV, the residual time reduction $RIV_{it}$ becomes orthogonal to the non-random exposure.

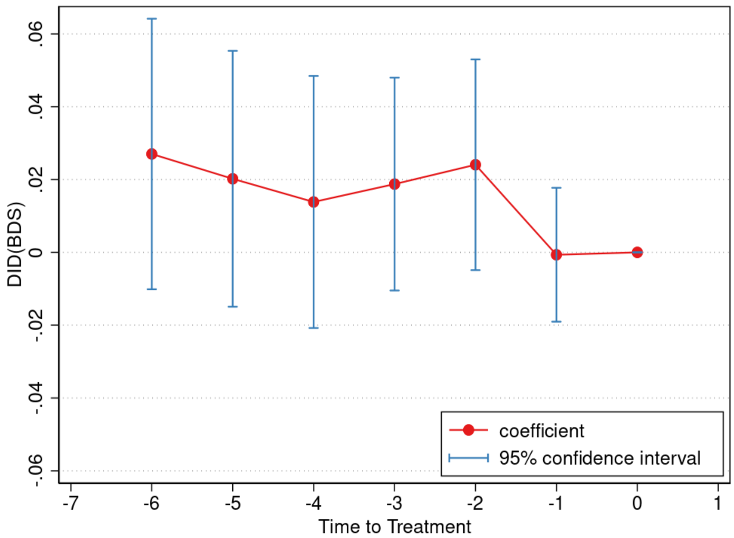
## A.12 Additional IV Validity Tests

**Table A.6:** Network Centrality

| | Degree | Eigenvector | Closeness |
|---|---|---|---|
| | (1) | (2) | (3) |
| Time Reduction | -0.009 | -0.007 | -0.011 |
| | (0.005) | (0.003) | (0.005) |
| IV | -0.009 | -0.008 | -0.011 |
| | (0.005) | (0.003) | (0.005) |
| Re-centered IV | 0.000 | 0.000 | -0.001 |
| | (0.019) | (0.011) | (0.017) |

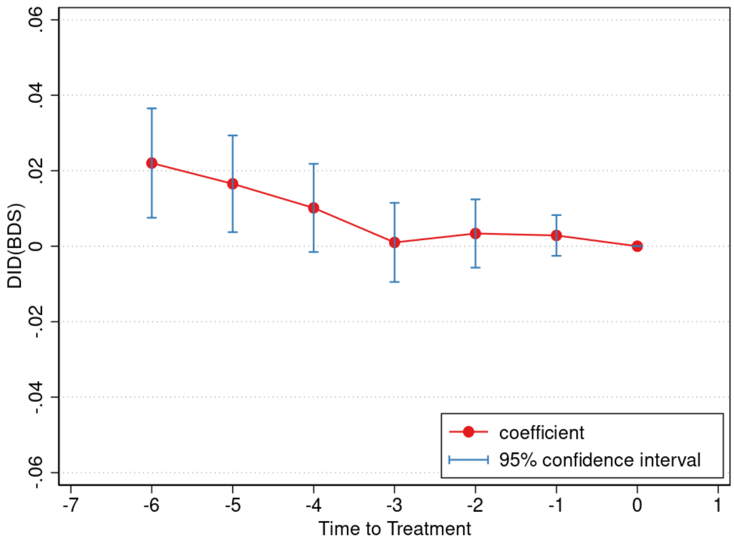
Note: This table presents the bilateral correlations between the three network centrality measures in columns and the three kinds of travel time reductions in rows. Every cell is a coefficient from a separate regression of the column variable on the row variable. The network centrality measures are calculated on the domestic flight network in the baseline year 2004. The sample therefore includes only the CBSAs with airports.

**Figure A.15:** No Pre-Trends: CBSA Firm Creation




Note: This figure shows the coefficients and confidence intervals of event study in Equation 5 at the CBSA level with the BDS data. 2004 is the baseline year and the coefficients represent differences in pre-trends relative to 2004 between treated CBSA-industry pairs and control ones. The treatment here is defined as having positive re-centered travel time reduction. I control for both the year fixed effects and the CBSA fixed effects. I also control for lagged city employment and use city size in 2004 as weights. Standard errors are clustered at the CBSA-level. The results show that the trends of firm creation across the treated and control are parallel to each other.

**Figure A.16:** No Pre-Trends: CBSA Employment




Note: This figure shows the coefficients and confidence intervals of event study in Equation 5 at the CBSA level with the BDS data. 2004 is the baseline year and the coefficients represent differences in pre-trends relative to 2004 between treated CBSA-industry pairs and control ones. The treatment here is defined as having positive re-centered travel time reduction. I control for both the year fixed effects and the CBSA fixed effects. Standard errors are clustered at the CBSA-level. The results show that the trends of city employment across the treated and control are parallel to each other.

## A.13 Robustness Checks

At the highly disaggregated 6-digit industry level, I have a lot of zeros in the outcome variable. Following the literature, I use the inverse hyperbolic sine transformation to deal with the log zero problem . For ensuring that my results do not come from the potential biases caused by such transformation, I first consider the changes of my outcome variable without log in columns (1) and (2) of Table A.7. Both columns show that the effects remain significantly positive. Then I use Pseudo Poisson Maximum Likelihood (PPML) estimation in columns (3) and (4) and show that the significant positive effects of travel time reductions on firm creation persist[11].

Since the unit of analysis is CBSA-industry pair and the industry classification used here is very disaggregated at the 6-digit NAICS level, the estimation treating industries as the same might overestimate the effects by putting too much weight on small industries. For mitigating this concern, I first weight industries at 4 digit level using employment data from the BDS data. The results are shown in the first two columns of Table A.8. The estimates actually become larger. This indicates that the effects concentrate in large industries and I probably underestimate the effect without weighting industries by sizes.

In column (3) and column (4) of A.8, I seek to check the robustness of the results to re-weighting observations by the inverse sampling probabilities. I calculate the sampling probabilities by dividing the number of new firms in the Infogroup data over the number of new firms in the BDS data. These probabilities refer to the chances of being drawn if the sampling is random. Then I weight the regressions by the inverses of the sampling probabilities. The results suggest that the estimates do not change much after the re-weighting and actually become more precise since we purge some noises by the re-weighting.

[11] However, the precision of the estimates decreases because the PPML method drops almost 20% of the observations in estimation. This comes from the well-known incidental parameter problem in estimating PPML with high dimension fixed effects. Fernández-Val and Weidner (2016) proves that this incidental parameter problem would disappear in asymptotic when the dimensions of fixed effects grow at the same rate. However, I have a fixed time dimension two and growing dimension of industry fixed effects in asymptotic in my long-difference setting. Then the PPML estimation becomes problematic and less reliable. The PPML estimate is also larger than the baseline estimate. The dropped 20% of the observations are the ones which have less variation in firm creation for identifying the fixed effects. These are therefore also the CBSA-industry pairs which have smaller treatment effects because the conditions which limits the firm creation also constraints the effects of the reductions in travel time to China. I therefore use the inverse hyperbolic sine transformation as my main method in estimation.

**Table A.7:** Counting Data

| | 2004-2013 Difference in Number of New Firms | | | |
|---|---|---|---|---|
| | Without Log | | IVPPML | |
| | Re-Centered IV | Controller IV | Re-Centered IV | Controller IV |
| | (1) | (2) | (3) | (4) |
| Time Reduction (hours) | 0.246 | 0.272 | 0.275 | 0.080 |
| | (0.069) | (0.078) | (0.025) | (0.044) |
| F statistic | 129.942 | 120.999 | | |
| N | 826085 | 826085 | 649322 | 649322 |
| Log City Size 2004 | Y | Y | Y | Y |
| Industry FE | Y | Y | Y | Y |

Note: IVPPML uses a panel of two years to identify the same effect as the baseline long-difference specification and uses the control function approach to conduct the IV regression in 2SLS (Lin and Wooldridge, 2019). The standard errors for IVPPML come from bootstrapping the sample. Number of observations becomes smaller in PPML because of being dropped as singletons or separated by fixed effects. I use inverse hyperbolic sine transformation to deal with the problem of log zero. Kleibergen-Paap rk Wald F statistic is reported for IV regression. Standard errors are clustered at the CBSA level.

**Table A.8:** Re-weighting

| | 2004-2013 Difference in Log Number of New Firms | | | |
|---|---|---|---|---|
| | By Industry Size in 2004 | | By Sampling Weight | |
| | Re-Centered IV | Controlled IV | Re-Centered IV | Controlled IV |
| | (1) | (2) | (3) | (4) |
| Time Reduction (hours) | 0.078 | 0.089 | 0.052 | 0.055 |
| | (0.018) | (0.020) | (0.010) | (0.011) |
| F statistic | 48.590 | 120.999 | 118.111 | 152.949 |
| N | 778650 | 778650 | 826085 | 826085 |
| Industry FE | Y | Y | Y | Y |
| Log City Size 2004 | Y | Y | Y | Y |

Note: This table reports coefficients from long difference regressions weighted by either industry size in 2004 or sampling probabilities at CBSA-industry (2-digit) level. Industry size is measured by employment at NAICS 4-digit industry level using the BDS data. Sample size becomes smaller when weighting by industry size as not every industry in the Infogroup data have employment observed from the BDS data. Sampling probability is obtained by comparing the Infogroup data with the administrative aggregated BDS data. I use inverse hyperbolic sine transformation to deal with the problem of log zero. Kleibergen-Paap rk Wald F statistic is reported for IV regression. Standard errors are clustered at the CBSA level.

## A.14 Understanding the Direction of Bias

**Table A.9:** First-Stage, Reduced Form and Direction of Bias

| | 2004-2013 Difference | | | | |
|---|---|---|---|---|---|
| | Time Reduction | Log Number of New Firms | Time Reduction | Log Number of New Firms | Re-Centered IV |
| | (1) | (2) | (3) | (4) | (5) |
| Re-centered IV | 0.403 | 0.021 | | | |
| | (0.076) | (0.009) | | | |
| Time to Existing Gateways | 0.972 | -0.008 | 0.983 | -0.008 | 0.028 |
| | (0.047) | (0.004) | (0.045) | (0.004) | (0.010) |
| Time to Future Gateways | -0.553 | -0.011 | -0.577 | -0.012 | -0.059 |
| | (0.028) | (0.003) | (0.026) | (0.003) | (0.009) |
| $R^2$ | 0.622 | 0.304 | 0.609 | 0.304 | 0.081 |
| N | 895 | 826085 | 895 | 826085 | 895 |
| Industry FE | N | Y | N | Y | N |
| Controls | Y | Y | Y | Y | Y |

Note: I use the inverse hyperbolic sine transformation to deal with the problem of log zero. Controls include log city size, the minimal time to other airports, and the average travel time to Chinese prefectures in the baseline year 2004. Standard errors are clustered at the CBSA level.

The estimates are downward biased because places which are already well connected to the gateway airports to China before 2004 are likely to have both higher trends in firm creation and less expected time reductions from the connecting of other gateway airports. Table A.9 illustrates this intuition. The first stage and reduced-form estimates for the 2SLS regression in Table 3 are shown in columns (1) and (2). We have strong first stage and reduced-form results to support the re-centered IV estimates. In column (1), as expected, I find that being far away from the existing gateways to China results in higher time reductions. Column (2), on the other hand, suggests that these CBSAs with higher time reductions also have lower increases in firm creation. The negative selection causes the downward bias observed in estimation. This kind of biases originated from the non-random exposure problem in high dimension transportation network, however, cannot be mitigated by controlling for the travel time to the existing gateways. The estimates on the other controls in columns (1) and (2) suggest that there could be other biases induced by the non-random exposure problem. Places with high travel time to the future gateways have both low time reductions and low firm creation. This positive selection therefore could cause upward bias in estimation. Columns (3) and (4) indicate that the two potential biases are orthogonal to our re-centered IV as controlling for the re-centered IV does not affect the coefficients of the time to existing gateways and the time to future gateways at all.

**Table A.10:** Effects by Time to Existing Gateways in US

| | 2004-2013 Difference in Log Number of New Firms | | |
|---|---|---|---|
| | Above Median | | |
| | Unadjusted IV | Re-Centered IV | Controlled IV |
| | (1) | (2) | (3) |
| Travel Time Reduction (hours) | 0.009 | 0.022 | 0.024 |
| | (0.004) | (0.022) | (0.022) |
| F Statistic | 1901.195 | 24.129 | 21.687 |
| N | 412581 | 412581 | 412581 |
| | Below Median | | |
| | Unadjusted IV | Re-centered IV | Controlled IV |
| | (4) | (5) | (6) |
| Travel Time Reduction (hours) | 0.026 | 0.050 | 0.064 |
| | (0.011) | (0.009) | (0.013) |
| F Statistic | 886.722 | 574.044 | 775.704 |
| N | 413540 | 413540 | 413540 |
| Log City Size 2004 | Y | Y | Y |
| Industry FE and Cluster at CBSA | Y | Y | Y |

Note: I use the inverse hyperbolic sine transformation to deal with the problem of log zero. Kleibergen-Paap rk Wald F statistic is reported for IV regression. Standard errors are clustered at the CBSA level. Columns (1) to (3) uses sample of CBSAs which have above-median travel time to the existing US gateways to China in 2004. Columns (4) to (6) uses sample of CBSAs which have below-median travel time to the existing US gateways to China in 2004. Some of those CBSAs which have above-median travel time to the existing gateways are remote CBSAs. They get small time reductions in all counterfactuals and the permutation of gateways cannot change their travel time reductions much. This explains the weaker first stage in the top panel. But the F-statistics are still higher than the cutoff value.

The results in Table 3 suggest that the downward biases are larger than the upward bias and the re-centered IV can correct for the biases. In column (5), I provide further evidence for this assertion. I regress the re-centered IV on the two controls which reflect different biases and find that the controls which cause downward biases have much smaller correlations with the re-centered IV. This indicates that the re-centered IV is much less correlated with the non-random exposure which could make the effects of travel time reductions underestimated. The results therefore explain why the re-centered IV estimate is larger than the unadjusted IV estimate. In Table A.10, additional evidence are provided. I separate the sample into two groups with above and below median levels of travel time to the existing gateways. Then the same set of regressions are conducted. I find that the positive gap between the re-centered IV estimate and the unadjusted IV estimate only exists in places close to the existing gateways. This is consistent with my understanding of the downward biases as only these places are expected to have smaller time reductions associated with more firm creation and consequently negative selection problem.

## A.15 Effects by Industry

**Table A.11:** Effects by 2-digit Industry

| 2-digit Industry | Estimate | 2-digit Industry | Estimate |
|---|---|---|---|
| Agriculture/Forestry/Fishing/Hunting | 0.003 | Finance/Insurance | 0.098 |
| | (0.014) | | (0.041) |
| Mining/Quarrying/Extraction | -0.039 | Real Estate/Rental/Leasing | 0.147 |
| | (0.029) | | (0.053) |
| Utilities | 0.018 | Professional/Scientific/Technical Services | 0.067 |
| | (0.027) | | (0.065) |
| Construction | 0.215 | Management of Companies/Enterprises | 0.165 |
| | (0.082) | | (0.093) |
| Manufacturing | 0.007 | Administrative/Support/Waste Management/ Remediation Services | 0.111 |
| | (0.006) | | (0.045) |
| Wholesale Trade | 0.047 | Educational Services | 0.041 |
| | (0.033) | | (0.054) |
| Retail Trade | 0.163 | Health Care/Social Assistance | -0.026 |
| | (0.065) | | (0.055) |
| Transportation/Warehousing | 0.053 | Arts/Entertainment/Recreation | 0.073 |
| | (0.025) | | (0.044) |
| Information/Cultural Industries | 0.110 | Accommodation/Food Services | 0.031 |
| | (0.049) | | (0.069) |
| | | Other Services | 0.076 |
| | | | (0.044) |

Note: There are 19 NAICS 2-digit industries in this table. I report the re-centered IV regression estimates separately for each of the 19 industries.

## A.16 Effects on the Quality of Entrants and Incumbent Firms

**Table A.12:** Effects on the Quality of Entrants

| | 2004-2013 Difference in Log Future Average Employment of New Firms | | | |
|---|---|---|---|---|
| | Re-Centered IV | | | |
| | Baseline | Add Geo Controls | Weight by Sampling Probability | Weight by Industry Size in 2004 |
| | (1) | (2) | (3) | (4) |
| Time Reduction (hours) | 0.057 | 0.086 | 0.064 | -0.144 |
| | (0.137) | (0.271) | (0.159) | (0.169) |
| F Statistic | 35.295 | 9.555 | 28.257 | 40.742 |
| | Controlled IV | | | |
| | Baseline | Add Geo Controls | Weight by Sampling Probability | Weight by Number of New Firms in 2004 |
| | (5) | (6) | (7) | (8) |
| Time Reduction (hours) | 0.065 | 0.043 | 0.059 | -0.152 |
| | (0.150) | (0.178) | (0.151) | (0.176) |
| F Statistic | 32.516 | 23.735 | 45.070 | 40.180 |
| N | 8116 | 8116 | 8116 | 8119 |
| Industry FE | Y | Y | Y | Y |
| Log City Size 2004 | Y | Y | Y | Y |

Note: I measure the quality of new firms founded in 2004 as their employment in 2009 and the quality of new firms founded in 2013 as their employment in 2018. As the travel time reductions are at the CBSA level and I don't have any firm-level controls, I compute the average quality of new firms founded in 2004 or 2013 for each CBSA-industry. I employ the same long difference specification in Equation 2 and the panel I use for estimation is balanced by keeping only CBSA-industry pairs which have new firms in both 2004 and 2013. Geo controls in baseline year 2004 includes: mean travel time to Chinese prefectures, minimum travel time to existing US gateways to China, minimum travel time to future US gateways to China, and minimum travel time to all other airports. Sampling probability is obtained by comparing the Infogroup data with the administrative aggregated BDS data. Industry size is measured by employment at NAICS 4-digit industry level using the BDS data. Sample size becomes smaller when weighting by industry size as not every industry in the Infogroup data have employment observed from the BDS data. I use inverse hyperbolic sine transformation to deal with the problem of log zero. Kleibergen-Paap rk Wald F statistic is reported for IV regression. Standard errors are clustered at the CBSA level.

**Table A.13:** Effects on Incumbent Firms

| | 2004-2013 Difference in Log Employment of incumbent Firms | | | |
|---|---|---|---|---|
| | Re-Centered IV | | | |
| | Baseline | Add Geo Controls | Weight by Sampling Probability | Weight by Industry Size in 2004 |
| | (1) | (2) | (3) | (4) |
| Time Reduction (hours) | -0.012 | -0.036 | -0.011 | -0.008 |
| | (0.019) | (0.036) | (0.020) | (0.031) |
| F Statistic | 74.715 | 23.842 | 61.819 | 64.347 |
| | Controlled IV | | | |
| | Baseline | Add Geo Controls | Weight by Sampling Probability | Weight by Industry Size |
| | (5) | (6) | (7) | (8) |
| Time Reduction (hours) | -0.015 | -0.025 | -0.011 | -0.012 |
| | (0.022) | (0.026) | (0.020) | (0.036) |
| F Statistic | 58.744 | 44.786 | 97.245 | 47.138 |
| N | 193666 | 193666 | 193666 | 190601 |
| Industry FE | Y | Y | Y | Y |
| Log City Size 2004 | Y | Y | Y | Y |

Note: This is firm-level estimation. I employ the same long difference specification in Equation 2 and the panel I use for estimation is balanced by keeping only incumbent firms which are founded before 2004 and are observed in both 2004 and 2013. Geo controls in baseline year 2004 includes: mean travel time to Chinese prefectures, minimum travel time to existing US gateways to China, minimum travel time to future US gateways to China, and minimum travel time to all other airports. Sampling probability is obtained by comparing the Infogroup data with the administrative aggregated BDS data. Industry size is measured by employment at NAICS 4-digit industry level using the BDS data. Sample size becomes smaller when weighting by industry size as not every industry in the Infogroup data have employment observed from the BDS data. I use inverse hyperbolic sine transformation to deal with the problem of log zero. Kleibergen-Paap rk Wald F statistic is reported for IV regression. Standard errors are clustered at the CBSA level.

## A.17 Model Appendix

### Deriving Free Entry Equilibrium

From Cobb-Douglas production, given sourcing strategy $\{j^m\}_{m=1}^{S^k}$, we get unit cost[12]:

$$c_i^k\left(\{j^m\}_{m=1}^{S^k}\right) = \eta^k \frac{1}{z_i^k} \prod_{m=1}^{S^k} (\phi_{ji}^{mk} z_{j^m}^m)^{-\gamma^{mk}} \tag{1}$$

With CES preference, we have fixed markup and price is simply

$$p_i^k\left(\{j^m\}_{m=1}^{S^k}\right) = \frac{\sigma}{\sigma-1} c_i^k\left(\{j^m\}_{m=1}^{S^k}\right) \tag{2}$$

while profit is:

$$\pi_i^k\left(\{j^m\}_{m=1}^{S^k}\right) = \underbrace{\frac{1}{\sigma}(\frac{\sigma-1}{\sigma\eta^k})^{\sigma-1} B_i^k \left[c_i^k\left(\{j^m\}_{m=1}^{S^k}\right)\right]^{1-\sigma}}_{\text{variable profit}} - \underbrace{f \prod_{m=1}^{S^k} (f_{j^m,i})^{\gamma^{mk}}}_{\text{sourcing cost}} \tag{3}$$

where $B_i^k = \alpha^k L_i (P_i^k)^{\sigma-1}$ and price index is:

$$P_i^k = \left(\int_{\omega\in\Omega_i^k} \left[p_i^k(\omega)\right]^{1-\sigma} d\omega\right)^{\frac{1}{1-\sigma}}$$

$$= \left(\int_{\{\phi_{ji}^{mk}\}_{m=1}^{S^k}} \sum_{\{j^m\}_{m=1}^{S^k}} \left[\frac{\sigma}{\sigma-1} c_i^k(\{j^m\}_{m=1}^{S^k})\right]^{1-\sigma} N_i^k x_i^k(\{j^m\}_{m=1}^{S^k}) d\left[G_i^k(\{\phi_{ji}^{mk}\}_{m=1}^{S^k})\right]\right)^{\frac{1}{1-\sigma}} \tag{4}$$

where $N_i^k$ is the equilibrium mass of firms. Since I assume that only intermediates are tradable, the price index does not have a general equilibrium component which depends on the whole economy.

Free entry implies that the expected variable profit must equal expected sourcing cost in equilibrium. I get equilibrium firm creation as the ratio of variable profit over a weighted average sourcing cost:

[12] $\eta^k = \prod_{m=1}^{S} (\gamma^{mk})^{-\gamma^{mk}}$ is a constant.

$$N_i^k = \frac{\alpha^k L_i/\sigma}{\sum_{\{j^m\}_{m=1}^{S^k}} \Big[\big(f \prod_{m=1}^{S^k} (f_{j^m,i})^{\gamma^{mk}}\big) x_i^k\big(\{j^m\}_{m=1}^{S^k}\big)\Big]} = \frac{\text{variable profit}}{\text{weighted average sourcing cost}} \tag{5}$$

By substituting the sourcing probability into the weighted average sourcing cost, the equilibrium firm creation can be written as:

$$N_i^k = \frac{\alpha^k L_i}{\sigma} \Bigg\{ \underbrace{\frac{\sum_{\{j^m\}_{m=1}^{S^k}} \Big[\prod_{m=1}^{S^k} (L_{j^m}^m)^b\Big] \Big[f \prod_{m=1}^{S^k} (f_{j^m,i})^{\gamma^{mk}}\Big]}{\prod_{m=1}^{S^k} \Big[\sum_{d=1}^{J} (L_d^m)^b\Big]}}_{\text{weighted average sourcing cost}} \Bigg\}^{-1} \tag{6}$$

I then rearrange the terms to represent the equilibrium firm creation in the most intuitive way as Equation 12:

$$N_i^k = \frac{\alpha^k L_i}{\sigma} \Bigg\{ \underbrace{f \frac{\sum_{\{j^m\}_{m=1}^{S^k}} \Big[\prod_{m=1}^{S^k} (L_{j^m}^m)^b (f_{j^m,i})^{\gamma^{mk}}\Big]}{\prod_{m=1}^{S^k} \Big[\sum_{d=1}^{J} (L_d^m)^b\Big]}}_{\text{weighted average sourcing cost}} \Bigg\}^{-1} = \frac{\alpha^k L_i}{\sigma} \Bigg\{ \underbrace{f \frac{\prod_{m=1}^{S^k} \Big[\sum_{d=1}^{J} (L_d^m)^b (f_{id})^{\gamma^{mk}}\Big]}{\prod_{m=1}^{S^k} \Big[\sum_{d=1}^{J} (L_d^m)^b\Big]}}_{\text{weighted average sourcing cost}} \Bigg\}^{-1} \tag{7}$$

**Proof of Proposition 1**

Employment of industry $k$ at location $i$ $L_i^k$ is determined by labor market clearing:

$$L_i^k = \sum_{j=1}^{J} \sum_{h=1}^{S^h} x_j^{kh}(i) \alpha^h L_j \gamma^{kh} = \sum_{j=1}^{J} \sum_{h=1}^{S^h} \frac{(L_i^k)^{\lambda\theta}}{\sum_d (L_d^k)^{\lambda\theta}} \alpha^h L_j \gamma^{kh} \tag{8}$$

We immediately see that the labor market clearing is independent with travel time. The change of travel time in the model would only affect the expected sourcing cost of firms because I assume the trade-off between travel time and location productivity only happens at the extensive margin. Therefore the only consequence is more firm creation after travel time is reduced.

In this economy with no labor mobility, the expenditure on each industry is fixed because the total income does not change and the expenditure shares are governed by the Cobb-Douglas function. The higher equilibrium firm creation increases the variety within each industry but not the expenditure on that industry. Similarly, the input-output table governs how these expenditures paid by consumers are spent on intermediates in each input industries. Therefore the labor demanded for producing intermediates in each industry will not change either. As the consequence, labor demand in each industry within one location would not be affected by travel time changes. Labor supply, on the other hand, is fixed. What we have is only more final good producers after travel time is reduced. Notice that this conclusion still holds without the fixed wage assumption as the wage would not be affected by the travel time change either.

**Comparative Statics**

Writing the equilibrium firm entry as:

$$\log\big(N_i^k\big) = \log\left(\frac{\alpha^k L_i}{f\sigma}\right) - \sum_{m=1}^{S^k} \log\left(\Big[\sum_{d=1}^{J} (L_d^m)^b (f_{id})^{\gamma^{mk}}\Big]\right) + \sum_{m=1}^{S^k} \log\left(\Big[\sum_{d=1}^{J} (L_d^m)^b\Big]\right) \tag{9}$$

Then when travel time from US locations to Chinese locations change from $\{f_{ij}\}_{j\in J^c}$ to $\{\tilde{f}_{ij}\}_{j\in J^c}$, the change in log firm entry is:

$$\Delta \log\big(N_i^k\big) = -\Delta\Big\{\sum_{m=1}^{S^k} \log\left(\Big[\sum_{d=1}^{J} (L_d^m)^b (f_{id})^{\gamma^{mk}}\Big]\right)\Big\} \tag{10}$$

By first-order approximation, I get:

$$\Delta \log\big(N_i^k\big) = \sum_{j=1}^{J^c} \left\{ \sum_{m=1}^{S^k} \gamma^{mk} \underbrace{\left[\frac{(L_j^m)^b (f_{ji})^{\gamma^{mk}}}{\sum_{d=1}^{J} (L_d^m)^b (f_{di})^{\gamma^{mk}}}\right]}_{\beta_{ji}^{mk}(b)=\text{relationship-specific effect}} \right\}\Big[-\Delta \log(f_{ji})\Big] + O\Big(\{\Delta f_{ji}\}_{j=1}^{J^c}\Big) \tag{11}$$

where we apply the conclusion in Proposition 1 that travel time change is independent with sectoral employment within and across locations in our model. I then can get the comparative statics result in Equation 14 by rearranging terms.

**Proof of Proposition 2**

By applying Proposition 1, I get the derivative of log firm entry as:

$$-\frac{\partial log(N_i^k)}{\partial f_{ij}} = \sum_{m=1}^{S^k} \frac{\partial \log\left(\left[\sum_{d=1}^{J} (L_d^m)^b (f_{id})^{\gamma^{mk}}\right]\right)}{\partial f_{ij}} = \sum_{m=1}^{S^k} \frac{(L_j^m)^b (f_{ij})^{\gamma^{mk}-1}\gamma^{mk}}{\left[\sum_{d=1}^{J} (L_d^m)^b (f_{id})^{\gamma^{mk}}\right]} \tag{12}$$

Taking derivative with respect to $b$, I get:

$$\begin{aligned}
\frac{\partial}{\partial b}\Big[-\frac{\partial log(N_i^k)}{\partial f_{ij}}\Big] &= \sum_{m=1}^{S^k} \frac{b(L_j^m)^{b-1}\log\big(L_j^m\big)(f_{ij})^{\gamma^{mk}-1}\gamma^{mk}\left[\sum_{d=1}^{J}(L_d^m)^b(f_{id})^{\gamma^{mk}}\right]}{\left[\sum_{d=1}^{J}(L_d^m)^b(f_{id})^{\gamma^{mk}}\right]^2} \\
&\quad - \sum_{m=1}^{S^k} \frac{(L_j^m)^b(f_{ij})^{\gamma^{mk}-1}\gamma^{mk} b(L_j^m)^{b-1}\log\big(L_j^m\big)(f_{ij})^{\gamma^{mk}}}{\left[\sum_{d=1}^{J}(L_d^m)^b(f_{id})^{\gamma^{mk}}\right]^2} \\
&= \sum_{m=1}^{S^k} \frac{b(L_j^m)^{b-1}\log\big(L_j^m\big)(f_{ij})^{\gamma^{mk}-1}\gamma^{mk}\left[\sum_{d\neq j}^{J}(L_d^m)^b(f_{id})^{\gamma^{mk}}\right]}{\left[\sum_{d=1}^{J}(L_d^m)^b(f_{id})^{\gamma^{mk}}\right]^2} > 0
\end{aligned} \tag{13}$$

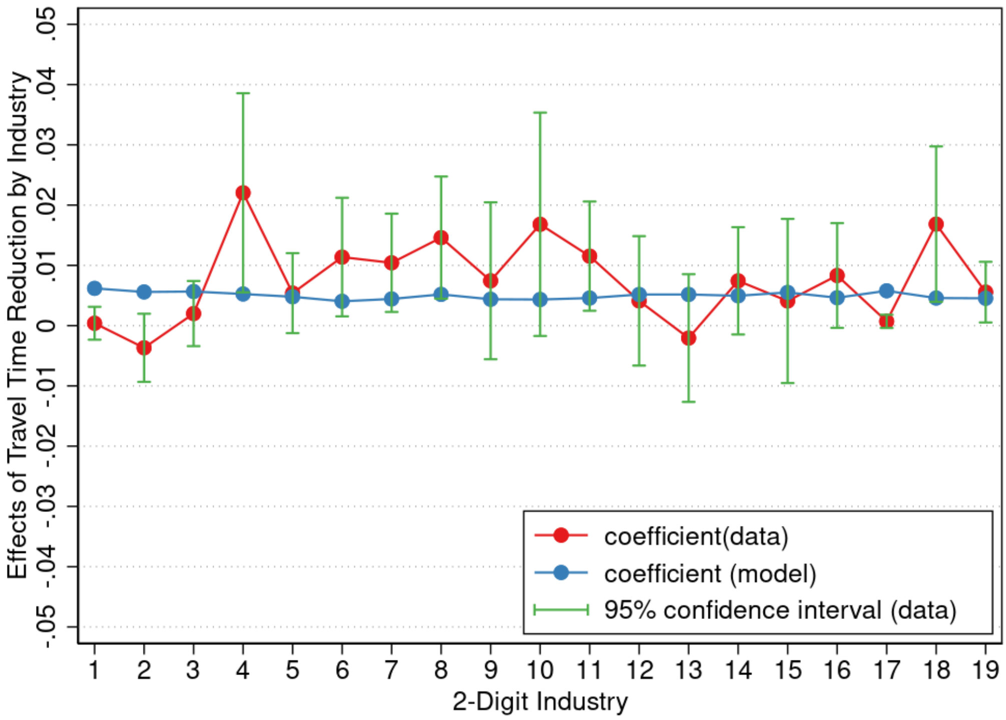
Taking derivative with respect to $S^k$, I get:

$$\frac{\partial}{\partial b}\Big[-\frac{\partial log(N_i^k)}{\partial f_{ij}}\Big] = \frac{\partial}{\partial S^k}\Big\{\sum_{m=1}^{S^k} \frac{(L_j^m)^b(f_{ij})^{\gamma^{mk}-1}\gamma^{mk}}{\left[\sum_{d=1}^{J}(L_d^m)^b(f_{id})^{\gamma^{mk}}\right]}\Big\} > 0 \tag{14}$$

as each one of the summed $\frac{(L_j^m)^b(f_{ij})^{\gamma^{mk}-1}\gamma^{mk}}{\left[\sum_{d=1}^{J}(L_d^m)^b(f_{id})^{\gamma^{mk}}\right]}$ is positive. Notice here I heuristically take $S^k$ as continuous variable, though it is actually discrete. But the underlying reasoning applies and we can conclude that the effect of time reduction on equilibrium firm creation is higher if the industry requires more input industries in terms of larger $S^k$.

## A.18 Model Fit by Industry

**Figure A.17:** Model Fit by Industry




Note: This figure compares the estimates of the effects of travel time reductions to China on firm creation by 2-digit industry from reduced-form estimation and model prediction. It shows the coefficients and confidence intervals of re-centered IV regressions for each 2-digit industry with the observational data or the model-generated data. The coefficients predicted by the model do not deviate from the estimates obtained by the reduced-form estimation for most of the 2-digit industries.

## A.19 Values of Routes

**Table A.14:** Values of Routes

| | Time Reduction | | Δ Log New Firm | | Welfare Gains | | Welfare Gains per Minute |
|---|---|---|---|---|---|---|---|
| | CBSA-Prefecture | | CBSA-Sector(2 digit) | | CBSA | | CBSA |
| Counterfactuals | Mean (1) | Std (2) | Mean (3) | Std (4) | Mean (5) | Std (6) | Value (7) |
| Detroit-Beijing | 4.9224 | 13.4954 | 0.0030 | 0.0008 | 0.0007 | 0.0019 | 0.0001 |
| Newark-Beijing | 0.8192 | 4.9210 | 0.0005 | 0.0030 | 0.0001 | 0.0007 | 0.0001 |
| DC-Beijing | 1.4050 | 7.4855 | 0.0008 | 0.0045 | 0.0002 | 0.0011 | 0.0001 |
| New York-Shanghai | 0.2860 | 3.7192 | 0.0012 | 0.0024 | 0.0003 | 0.0006 | 0.0011 |
| Chicago-Shanghai | 1.7861 | 8.6374 | 0.0050 | 0.0040 | 0.0013 | 0.0010 | 0.0007 |
| Detroit-Shanghai | 2.5461 | 11.4235 | 0.0049 | 0.0062 | 0.0012 | 0.0015 | 0.0005 |
| Newark-Shanghai | 0.5433 | 5.0188 | 0.0017 | 0.0032 | 0.0004 | 0.0008 | 0.0008 |
| Seattle-Shanghai | 8.9217 | 29.5728 | 0.0069 | 0.0157 | 0.0017 | 0.0036 | 0.0002 |
| Seattle-Beijing | 11.8897 | 34.7617 | 0.0068 | 0.0198 | 0.0016 | 0.0047 | 0.0001 |

Note: This table reports the predicted changes of log new firms and welfare gains for the nine counterfactuals specified in the left column. In each counterfactual, I introduce the particular route only. The changes in firm creation come from Equation 13 while the welfare gains come from Equation 16. In columns (1) and (2), I report the mean and the standard error of changes of travel time in each counterfactual across CBSAs in minutes . In columns (3) and (4), I report the mean and the standard error of changes of firm creation in each counterfactual across CBSA-industry pairs. In columns (5) and (6), I report the mean and the standard error of welfare gains in each counterfactual across CBSAs. In column (7), I show the welfare gain per minute. For counterfactual calculation, I use $\sigma = 5$ from Broda and Weinstein (2006) and $\{\alpha^k\}_{k=1}^S$ from the IO Table.