The Creation and Diffusion of Knowledge: Evidence from the Jet Age*

Stefan Pauly† Fernando Stipanicic‡

March 19, 2024

This paper provides new causal evidence of the impact of air travel time on the creation and diffusion of knowledge. We exploit the beginning of the *Jet Age* as a quasi-natural experiment. We digitize airlines' historical flight schedules and construct a novel data set of the flight network in the United States. Between 1951 and 1966, travel time between locations more than 2,000 km apart decreased on average by 41%. The reduction in travel time explains 33% of the increase in knowledge diffusion as measured by patent citations. The increase in knowledge diffusion further caused an increase in the creation of new knowledge. The results provide evidence that jet airplanes led to innovation convergence across locations and contributed to the shift in innovation activity towards the South and the West of the United States.

JEL Classification: O31, O33, R41, N72, F60

*First version: February 2021. We are grateful for the continuous support of Christian Hellwig since the beginning of this project. We are indebted to Björn Larsson who has made this project possible by sharing with us his collection of historical flight schedules. We thank Taylor Jaworski and Carl Kitchens for sharing highway data with us. This paper has benefited from comments of Antonin Bergeaud, Davide Cantoni, Thomas Chaney, Fabrice Collard, Ben Faber, Ruben Gaetani, Victor Gay, Ulrich Hege, Enrico Moretti, Luigi Pascali, Mohamed Saleh, Mark Schankerman, Claudia Steinwender, Nicolas Werquin, Alex Whalley, Miguel Zerecero, and numerous participants in seminars and conferences in Berkeley, FREIT, IMF, INSEAD/Collège de France, RIDGE Growth, RIEF Paris, SED, Toulouse, Urban Economic Association, YES Princeton, and others.

†Email: *stefan.pauly@sciencespo.fr*

‡University of Oslo. Corresponding author, email: *fernando.stipanicic@econ.uio.no*

1. Introduction

The geography of innovation in the United States underwent a radical shift after World War II. In 1950, the number of patents per capita in the South and the West was less than half as high as in the Northeast and the Midwest. By 2010, this difference had disappeared completely with the emergence of prominent technology clusters in the South and the West. Developing technology clusters is of great interest to policy makers as this may have strong effects on the local economy (Moretti (2010), Moretti and Thulin (2013)). While the literature has emphasized knowledge spillovers as one of the drivers of innovation (Storper and Venables (2004), Furman and Stern (2011), Acemoglu et al. (2016)), highlighting the role of physical proximity for facilitating face to face interactions (Glaeser (2011), Carlino and Kerr (2015)), there is scarce evidence on the effect of long distance transport infrastructure on knowledge spillovers and the geography of innovation (Agrawal et al. (2017), Chatterji et al. (2014)).

This paper provides new causal evidence on this question by exploiting as a quasi-natural experiment the beginning of the *Jet Age* in the United States. During the 1950s the introduction of jet engines into civil aviation led to a large nationwide reduction in travel time. For example, travel time between New York City and San Francisco went from 11 hours in 1951 to 5 hours and 35 minutes in 1966. This large reduction in travel time potentially facilitated long distance face to face interactions.

We find that the decrease in travel time led to an increase in the diffusion of knowledge, which we convert into an increase in access to knowledge. Next, we find that the increase in access to knowledge spurred an increase in the creation of new knowledge. The results provide evidence that jet airplanes drove innovation convergence across locations and contributed to the shift in innovation activity towards the South and the West of the United States.

We start by constructing a new dataset of the flight network in the United States during the 1950s and 1960s. We digitize historical flight schedules of the major interstate airlines operating in the period and obtain the fastest route between every two airports in the network.[1] We document that between 1951 and 1966 travel time decreased on average by 29%, and the decrease is on average of 41% for airports located more than 2,000km apart.[2]

This nationwide shock was arguably exogenous as it happened in a strictly regulated environment. We decompose the change in travel time and find that 90% of the change is due to the improvement in aircrafts' speed, while 10% is due to a change in the flight routes. This is consistent with the fact that during this period the Civil Aeronautics Board (CAB) was imposing strong regulation in the interstate airline market. With the objective to promote a *stable* airline industry, the CAB determined ticket prices and restricted entry of airlines into new or existing routes.

Additionally, during the 1950s and 1960s airplanes were predominantly used to transport people and not goods. Hence, the change in travel time represented a shock to the mobility of people while not significantly affecting the shipment of goods.

To study knowledge creation and diffusion we use patent data. We follow Jaffe et al. (1993) and use patent citations as our observable measure of knowledge flow. We assemble one dataset with all corporate patents granted by the United States Patent and Trademark Office (USPTO) with filing year between 1949 and 1968, which includes for each patent: filing year, technology classification, location (Metropolitan Statistical Area, MSA) of the inventors when they applied for the patent, owner of the patent and citations to other patents which were granted by the USPTO.

We document three facts of patenting activity during our sample period. First, patent

[1] The 6 domestic airlines in our data accounted for 75% of total air passenger transport.

[2] New York and Boston are about 300km apart, while New York and San Francisco are located about 4,130 km apart. Between 1951 and 1966 we observe a reduction of travel time of 23% (13 minutes reduction) between New York and Boston, while the reduction is of 49% (5 hours 25 minutes reduction) between New York and San Francisco.

growth was stronger in initially less innovative MSAs. Second, it was also stronger in the South and the West of the US. Third, the mass of citations shifted towards longer distances. Our results show that the decrease in travel time contributed to all three facts.

We do our analysis in two steps. In the first step, we estimate a gravity equation to obtain the elasticity of citations to travel time. We identify the elasticity exploiting only within establishment-pair across-time variation in citations and travel time. The estimated elasticity implies that citations increased on average 2.4% due to the decrease in travel time between 1951 and 1966. We find that the absolute value of the elasticity is increasing with the distance between the citing and cited establishments. At a distance of more than 2,000km, the change in travel time implies an increase in citations of 6.9%. This accounts for 32.7% of the observed increase in citations in this distance range.

In order to rule out the possibility that the opening of new routes or the timing of adoption of jets at the route level was driven by variables that also affected knowledge flows, we perform an instrumental variables estimation. We instrument the observed travel time with a fictitious travel time computed by fixing routes to the initial time period and assuming in each year all routes are operated with the year's average airplane. Hence, changes in fictitious travel time are only due to the nationwide roll out of jets and is thus independent of decisions at the route level. The key source of variation in the instrument is the time-variation of in-flight speed which affects the relative importance of the amount of layovers relative to the distance between two MSAs. The results do not change significantly, reflecting the reduced scope for endogeneity of travel time. In addition, the results are robust to controlling for potential confounding factors such as changes in highway travel time, telephone connectivity and flight ticket prices. Finally, the results also remain after restricting the sample to contain only establishments that existed in the initial time period.

In the second step, using the estimated elasticity of diffusion of knowledge, we compute a measure of knowledge access that is specific to each location-technology. The measure captures changes in knowledge access that are only consequence of the change in travel time. We use knowledge access as an externality that affects the production of new patents and estimate the elasticity of new patents to knowledge access. We identify the elasticity at the establishment level comparing only across time variation in patents and knowledge access across establishments within a location, conditional on aggregate technological trends. Thus, the identification is independent of location specific changes in local population or R&D subsidies. The estimated elasticity implies that the amount of new patents filed increased at a yearly growth rate of 3.5% due to the increase in knowledge access, which accounts for 79.5% of the observed yearly growth rate.

Given the reduction in travel time was larger for longer distances, the increase in knowledge access was stronger in locations geographically far from the initial innovation centers located in the Midwest and the Northeast. Hence, by increasing access to knowledge, the reduction in travel time led to a shift in the distribution of innovative activity towards the South and the West of the US. The South and the West had an average yearly growth rate of patenting 2.1 percentage points higher than the Northeast and the Midwest during our sample period. The change in travel time explains 35% of the observed differential growth.

We find that the value of the elasticity of patents to knowledge access is bigger in magnitude for establishments located in initially less innovative locations. Within each technology class, we rank locations according to the amount of patents in the initial time period and split them into four quartiles. We find that the increase in knowledge access predicts a 4.5% yearly growth rate of patenting in locations in the lowest quartile of initial innovativeness, while it predicts a 3.4% yearly growth rate in the highest quartile. The difference in growth rates indicates that the increase in knowledge access acted as a convergence force between locations, and it can explain 21% of the convergence observed in the data. Results go in the same direction if we rank locations in terms of patents per capita.

Our results are robust to controlling for changes in market access by highway, changes in market access by airplanes and time changing telephone connectivity. Results do not change if we compute knowledge access using only knowledge located at long distances. Additionally, we present suggestive evidence that the results are not driven by a decrease in financial frictions.

We also estimate the elasticity by instrumental variables, constructing an instrumental knowledge access with the instrumental travel time. Following Borusyak and Hull (2023) we may be concerned that the instrumental knowledge access may contain a non-random component which is consequence of the underlying geography of the US: given that most innovation was initially located in the Northeast and Midwest, the South and the West of the US might have seen larger increases in knowledge access under any potential flight network due to their larger distance to initial knowledge centers. We estimate the elasticity with the instrumental knowledge access and with a recentered version by subtracting the expected instrumental knowledge access as suggested by Borusyak and Hull (2023). In both cases results go in the same direction as in the baseline analysis.

This paper contributes to multiple branches of literature. First, it contributes to the literature on agglomeration and knowledge spillovers. Agglomeration forces are usually understood as happening in a geographically localized manner (Glaeser (2011), Arzaghi and Henderson (2008)). The literature on technology clusters also documents this fact (Duranton et al. (2009), Kerr and Robert-Nicoud (2020), Moretti (2021)). The seminal paper Jaffe et al. (1993) finds that patent citations decay rapidly with distance. Our results show that jet airplanes allowed long distance knowledge spillovers, facilitating the development of technology clusters in other regions. The literature that provides evidence of knowledge spillovers usually focuses on changes in the supply of knowledge (Bloom et al. (2013), Acemoglu et al. (2016)). In our case we fix the supply of knowledge and focus on changes in the degree of accessibility.

We contribute to the literature on transportation by constructing a novel data set and studying a new quasi-natural experiment that isolates a shock to the mobility of people. To the best of our knowledge, this is the first quantitative analysis of the change in air travel time due to the roll out of jet airplanes in commercial aviation. Other papers have studied the impact of transportation improvements on innovation. Agrawal et al. (2017) study the impact on innovation of a region's stock of highways, while Perlman (2016) uses 19th century data on locations' density of railroads. Andersson et al. (2017) and Tsiachtsiras (2021) do so using the historical railroad expansion in Sweden and France. Relative to them, we contribute by exploiting a quasi-natural experiment that allows us to isolate a channel of face to face interactions, with little scope for a trade channel. In contemporaneous work Bai et al. (2021) estimate the elasticity of patent citations to air travel time using the introduction of new airline routes in a more recent period, post deregulation of the airline market. Relative to them, we contribute by exploiting a nationwide shock that creates heterogeneous changes in travel time across routes and in which the risk of endogeneity is limited. Our work is related to other literature which found that business travel affects innovation (Hovhannisyan and Keller (2015)), trade (Söderlund (2020)) and industrial activity (Coscia et al. (2020)). Also, air travel shapes collaboration between researchers (Catalini et al. (2020)).

The impact of transportation improvements in economic outcomes has long been a subject of study (Fogel (1963), Baum-Snow (2007), Michaels (2008), Donaldson and Hornbeck (2016), Campante and Yanagizawa-Drott (2017), Jaworski and Kitchens (2019) and Herzog (2021)). Our convergence result contrasts with previous studies on improvements in other means of transport. Pascali (2017) finds that the introduction of steam engine vessels in the second half of the 19th century led to an increase in international trade which contributed to economic divergence between countries. Faber (2014) finds that the expansion of the highway system in China led to a reduction of GDP growth of peripheral counties, with evidence suggesting a trade channel. While both papers emphasize a trade channel, in our set up the trade channel would not be of first order. Hence, we uncover a new effect of improved connectivity. Our paper is related to Campante and Yanagizawa-Drott (2017) who study changes in international airplane connectivity, finding that it affects capital flows and the spatial distribution of economic activity.

Finally, we contribute to the contemporaneous literature on innovation on the post WW2 period. Gross and Sampat (2023) study the long-term effects of the public R&D funding by the Office of Scientific Research and Development (OSRD) during WW2. They find that this R&D shock enlarged pre-existing patenting gaps across locations. Kantor and Whalley (2023) study the effects of NASA spending and the race to the moon during the 1960s, finding stronger growth in county-industries that were more space-relevant before the Space Race. We contribute to them by studying a shock that it is different in its nature: improved connectivity, rather than increased expenditure. We provide evidence that this shock contributed to the post WW2 shift in innovative activity towards the South and the West. Additionally, we show that improved connectivity had a differential effect on innovation which contributed to closing the patenting gap between locations. This is in line with the decline in innovation concentration documented in Andrews and Whalley (2021).

The paper is structured as follows. First, we present a simple theoretical framework which lays the foundations of how to think about the creation and diffusion of knowledge. The framework shows the two key parameters to estimate. Second, we describe the historical context in which jet airplanes were introduced. Third, we present the two datasets that we use: travel times and patents. Fourth, we perform the analysis to estimate the impact of travel time on the diffusion of knowledge, and convert it into changes of access to knowledge to study the creation of knowledge. Fifth, we conclude.

2. Conceptual framework

This section lays out a simple theoretical framework to think about the creation of knowledge. The framework clearly shows the two key parameters to estimate empirically: the elasticity of knowledge diffusion to travel time and the elasticity of knowledge creation to knowledge access.

Following Carlino and Kerr (2015) we consider a production function of knowledge which includes external returns in the form of knowledge spillovers. Knowledge output of a firm depends not only on firm's specific characteristics as its idiosyncratic productivity and input decisions, but also on an externality due to knowledge spillovers. We consider a production function of knowledge of the following form:

$$\text{New Knowledge}_{Fi} = f(z_{Fi}, inputs_{Fi}) \times \text{Knowledge Access}_i^{\rho} \tag{1}$$

where New Knowledge$_{Fi}$ is the knowledge created by firm F located in i. The output of Fi depends on an *internal* component and on an *external* component. The *internal* component is the firm's idiosyncratic productivity z_{Fi} and choice of inputs $inputs_{Fi}$. The *external* component represents the externality to which all firms F in location i are exposed to: Knowledge Access$_i$. This externality, *Knowledge Access*, represents the total amount of knowledge spillovers that the firm is exposed to. The degree to which the externality affects the production of knowledge is governed by the parameter ρ. If ρ is zero then knowledge spillovers have no effect on the creation of new knowledge. On the other hand, a positive ρ implies that, keeping productivity and inputs constant, an increase in the level of knowledge spillovers leads to an increase in firm F's creation of new knowledge.

A long standing literature studies the importance of knowledge spillovers for the creation of new knowledge.[3] The concept of knowledge spillovers goes back at least to

[3]The chapters of Audretsch and Feldman (2004) and Carlino and Kerr (2015) in the Handbook of Regional and Urban Economics provide an excellent review on the literature on knowledge spillovers,

Marshall (1890) who explains it as one of the agglomeration forces. Krugman (1991) refers to knowledge spillovers as one of the justifications for external increasing returns, and that the degree of spillovers are dependent on physical distance. The geographic decay of spillovers is grounded in the fact that not all knowledge is easy to codify, usually referred to as *tacit knowledge*, and geographic proximity increases the degree of knowledge spillovers by facilitating face to face interactions (Storper and Venables (2004), Glaeser (2011)). Hence, we consider the total amount of knowledge spillovers to which the firm F in location i is exposed to has the following functional form:

$$\text{Knowledge Access}_i = \sum_j \text{Knowledge stock}_j \times \text{distance}_{ij}^{\beta} \tag{2}$$

where Knowledge stock_j is the total amount of knowledge in location j (which is non-negative) that could potentially spill over to location i and distance_{ij} is a measure of distance from j to i. The amount of knowledge that spills over from j to i depends on distance and the degree with which distance impedes spillovers, governed by the parameter β. If β is zero, then distance does not affect knowledge spillovers from j to i and all locations perfectly share the same level of *Knowledge Access*. On the contrary, a negative β implies a decay in knowledge spillovers when distance increases. In other words, a negative β implies that if we reduce the distance from j to i while keeping every other distance constant, the amount of spillovers from j to i will weakly increase.

This theoretical framework bears resemblance to the concept of *Market Access* presented in Donaldson and Hornbeck (2016) and Redding and Venables (2004). If we interpret *Knowledge Access* as one of the inputs in the production function of knowledge, then $\text{Knowledge Access}_i$ could be interpreted as a measure of *Input Market Access*. This measure captures how cheaply firms in location i can access pre-existing knowledge, where the cost of accessing knowledge depends on distance between i and j. Also, *Knowledge Access* is similar to a measure of network centrality. The centrality of each location i (node) is the weighted sum of distance (edges) to every location, where the

their geographic decay and how they affect the creation of knowledge.

weight of each location is given by its knowledge stock.

One assumption of the theoretical framework is that New Knowledge$_{Fi}$ is multiplicative-separable on Knowledge Access$_i$.[4] To the extent that firm's productivity z_{Fi} and choice of inputs $inputs_{Fi}$ are relatively time invariant, this assumption is not restrictive.[5] However, if for example $inputs_{Fi}$ changes with Knowledge Access$_i$, then the estimated value of the elasticity would be the sum of the direct effect of Knowledge Access$_i$ on New Knowledge$_{Fi}$ (ρ) and the indirect effect through changes in $f(\cdot)$.

The theoretical framework highlights the two parameters to estimate: ρ and β. Empirically, we use travel time as a measure of distance to first estimate β and then conditional on β we estimate ρ. Changes in travel time due to improvements in commercial aviation allow us to estimate both parameters. First, we use citations between patents as a proxy for the diffusion of knowledge. We estimate β by relating changes in travel time between research establishments to changes in citations between them. Second, we use the stock of patents filed by inventors in each location as proxy for each location's stock of knowledge. We construct a measure of knowledge access using the patent stock, travel times and the value of β. New patents in each location proxy for new knowledge. Changes in travel time lead to changes in knowledge access which allow us to estimate ρ.

[4]The implicit assumption is that $\frac{\partial log(\text{New Knowledge}_{Fi})}{\partial log(\text{Knowledge Access}_i)} = \frac{\partial log(f(z_{Fi}, inputs_{Fi}))}{\partial log(\text{Knowledge Access}_i)} + \rho = \rho$, meaning that $\frac{\partial log(f(z_{Fi}, inputs_{Fi}))}{\partial log(\text{Knowledge Access}_i)} = 0$.

[5]In the empirical analysis we will include a firm-location fixed effect Fi that would absorb time-invariant characteristics.

3. Historical context

3.1. Air transport: jet arrival

The jet aircraft was first invented in 1939 for military use, with the German Heinkel He 178 being the first jet aircraft to fly. The first commercial flight by a jet aircraft was in 1952 by the British Overseas Airways Corporation (BOAC) from London, UK to Johannesburg, South Africa with a Havilland Comet 1. Nonetheless, given the amount of accidents of the Havilland Comet 1 due to metal fatigue, jet commercial aviation did not truly take off until the Boeing 707 entered commercial service in late 1958. The 24th of January of 1959 represented a major milestone in the jet era: American Airlines Flight 2 flew with a Boeing 707 jet aircraft from Los Angeles to New York, the first non-stop transcontinental commercial jet flight.[6]

In 1951 New York City and Los Angeles were connected with a one-stop flight in 10 hours and 20 minutes. The flight had a forced stop in Chicago and was operated with the propeller aircraft Douglas DC-6, which had a cruise speed of 500 kmh. By 1956, New York City and Los Angeles were connected with a non-stop flight in 8 hours and 30 minutes. This was accomplished due to the introduction of the propeller aircraft Douglas DC-7 which had a cruise speed of 550kmh, and a change in regulation which increased maximum flight time of a crew from 8 to 10 hours within a 24-hour window.[7] In 1961, the route was covered with the jet aircraft Boeing 707 in a non-stop flight in 5 hours 15 minutes, reaching 5 hours 10 minutes in 1966. The Boeing 707 had a cruise speed of 1000kmh, cutting travel time from New York City to Los Angeles in half

[6] The reader passionate of aviation history would enjoy reading the following New York Times article which tells the experience of the first transcontinental jet flight: https://www.nytimes.com/2009/01/26/nyregion/26american.html

[7] AA and TWA had transcontinental non-stop propeller flights scheduled since at least 1954. These flights were scheduled to take 7 hours 55 minutes, just under the maximum flight time allowed by regulation in domestic flights: regulation impeded pilots from being on duty more than 8 hours within a 24 hours window. Nonetheless, the propeller aircrafts used in these flights, the Douglas DC-7 and the Lockheed Super Constellation, overheated their engines due to excessive demand to cover the route in less than 8 hours. AA fought intensely until the CAB approved a waiver that allowed non-stop transcontinental flights to take up to 10 hours to accomplish the non-stop transcontinental flight. See page 16 of the edition of the 21st of June 1954 of the Aviation Week magazine https://archive.org/details/Aviation_Week_1954-06-21/page/n7/mode/2up

between 1951 and 1966.

3.2. Air transport: moving people, not goods

During the 1950s and 1960s, air transportation served to transport people but not goods. Figures 1 and 2 are images (edited for better readability) from annual reports of the Interstate Commerce Commission of 1967 and 1965 respectively. Figure 1 displays the amount of passenger-miles for Air, Motor and Rail transportation from 1949 to 1966.[8] We observe that, while transport of people by rail decreased and by motor remained relatively constant, transport of people by air multiplied by 6 in a 16-year period, which translates to around 12% compound annual growth. In 1966, air transport accounted for more passenger-miles than both rail and motor transportation together, reflecting the growing importance of this means of transportation.

Figure 2 shows shipments in ton-miles for the period 1939 to 1964 by means of transportation: Airways, Pipelines, Inland Waterways, Motor, Railroads. Interestingly, we observe that air transport of goods, even if it increased, it accounted for less than 0.1% of transport of goods in 1964.

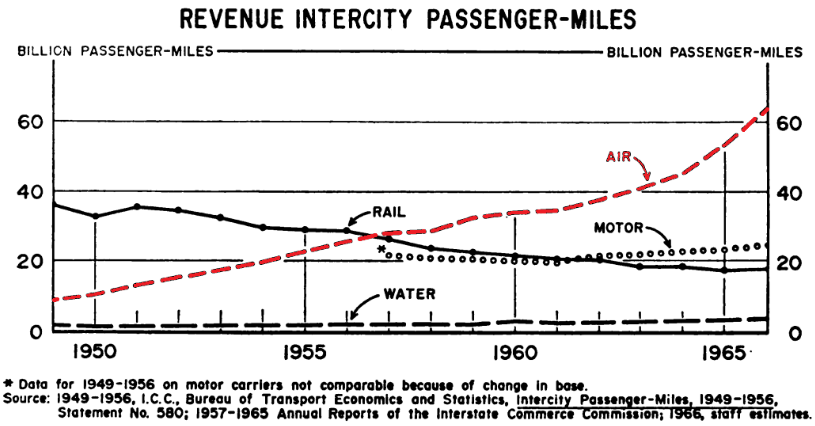


Figure 1: Passenger Miles
Source: Interstate Commerce Commission, Annual Report 1967. Edited by the authors

[8]Passenger-miles is a standard unit of measurement in transport, where one passenger-mile accounts for one person traveling one mile. The reasoning is the same for ton-miles, with one ton of goods traveling one mile.

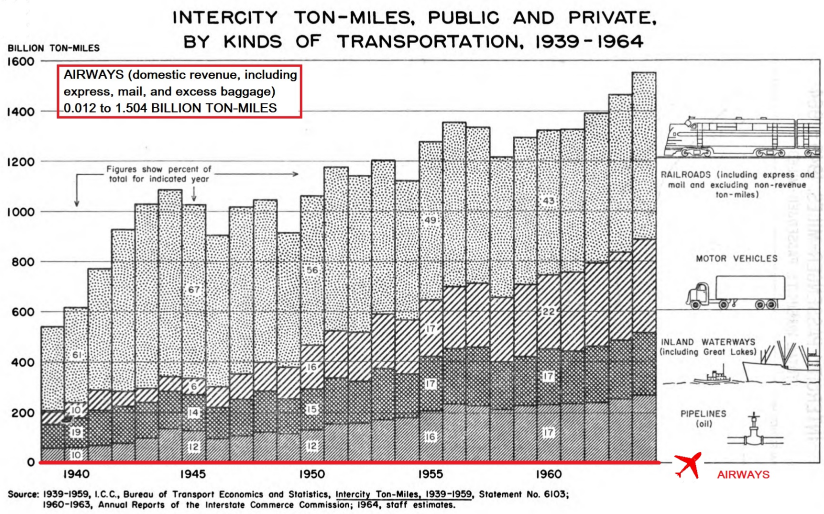


Figure 2: Ton Miles
Source: Interstate Commerce Commission, Annual Report 1965. Edited by the authors

3.3. Regulation

As explained in Borenstein and Rose (2014), in the 1930s the airline industry was seen as suffering from coordination issues, destructive competition and entry. Additionally, the industry was developing in a context of financial instability and increasing military concerns post Great Depression. A strong domestic airline industry was perceived as an interest of national defense. As consequence, the Civil Aeronautics Board (CAB) was created in 1938 with the objective to promote, encourage and develop civil aeronautics.[9] It was empowered to control entry, fares, subsidies and mergers.[10] In other words, the CAB regulated the market by deciding which airlines could fly, in which routes they could operate, the price that they charged in each route, the structure of subsidies and merger decisions. The CAB regulated the airline industry in a barely unchanged manner until it ceased to exist in 1985.

When the CAB was created, it conceived special rights to the existing airlines over the connections they were operating. The CAB did not permit entry of new airlines on

[9]The CAB was a federal agency hence, in principle, would not have control over intrastate routes. Nonetheless, according to Borenstein and Rose (2014) the CAB managed to have some intrastate markets under its control using legal arguments.

[10]Safety regulation was under the control of the Federal Aviation Administration.

interstate routes and gradually allowed current airlines to expand their routes. The CAB controlled both the system and each airline's network. The network was designed to maintain industry stability and minimize subsidies, leading to a system where each route was mainly operated by one or two airlines.[11] Importantly, Borenstein and Rose (2014) in pages 68-69 explain that *"the regulatory route award process largely prevented airlines from reoptimizing their networks to reduce operation costs or improve service as technology and travel patterns changed."* As a consequence, any technological improvement such as increases in aircraft speed, capacity or range would not affect each airline's flight network in the short term.

By regulating fares, the CAB explicitly encouraged airlines to adopt new aircraft. Airlines, when operating an older aircraft, would apply for a fare reduction arguing that it is needed in order to preserve demand for low quality service. The CAB would refuse this application, hence airlines would have to adopt new aircraft or risk losing consumers who would choose another airline which flies newer aircrafts.

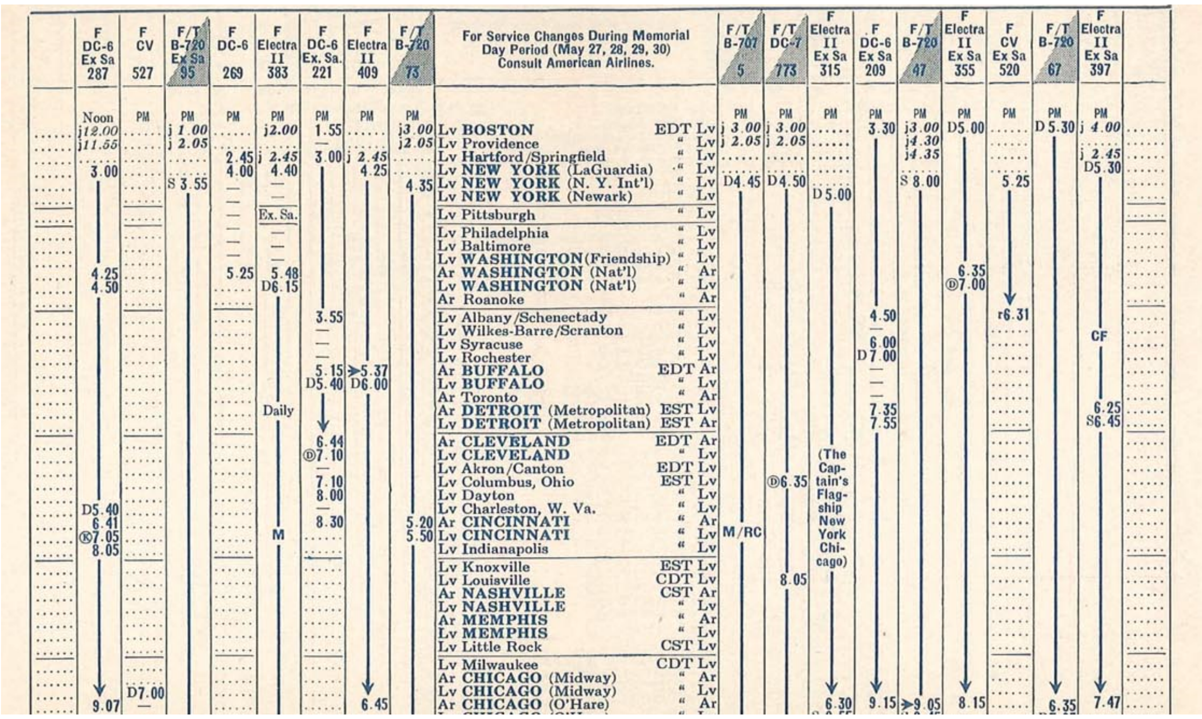
4. Air travel data

We construct a new data set of the flight network in the United States during the 1950s and 1960s. We collected and digitized information of all the flights operated by the main airlines and obtained the fastest route and travel time between every two airports in the network.

To construct the flight network we use historical flight schedules of the main airlines operating in 1950s and 1960s. Figure 3 is a fragment from an example page of the 1961 flight schedule of American Airlines. In the flight schedule we observe in the center column the name of departure and arrival cities (which we match to airports using airlines' historical ticket office geographical location), while the small columns

[11] Borenstein and Rose (2014) in page 68, based on Caves (1962), expose *"In 1958, for example, twenty-three of the hundred largest city-pair markets were effectively monopolies; another fifty-seven were effectively duopolies; and in only two did the three largest carriers have less than a 90 percent share."*

on the sides depict flights. In the top of the small columns we observe the type of service provided (first class, coach or both), aircraft operated, days operated (daily if information is missing) and flight number. The content of the small columns displays the departure and arrival time (local time, bold numbers represent PM) at each city, including all intermediate stops.


For Service Changes During Memorial Day Period (May 27, 28, 29, 30) Consult American Airlines.

Lv BOSTON EDT Lv
Lv Providence " Lv
Lv Hartford/Springfield " Lv
Lv NEW YORK (LaGuardia) " Lv
Lv NEW YORK (N. Y. Int'l) " Lv
Lv NEW YORK (Newark) " Lv
Lv Pittsburgh " Lv
Lv Philadelphia " Lv
Lv Baltimore " Lv
Lv WASHINGTON (Friendship) " Lv
Ar WASHINGTON (Nat'l) " Ar
Lv WASHINGTON (Nat'l) " Lv
Ar Roanoke " Ar
Lv Albany/Schenectady " Lv
Lv Wilkes-Barre/Scranton " Lv
Lv Syracuse " Lv
Lv Rochester " Lv
Ar BUFFALO EDT Ar
Lv BUFFALO " Lv
Ar Toronto " Ar
Ar DETROIT (Metropolitan) EST Lv
Lv DETROIT (Metropolitan) EST Ar
Ar CLEVELAND EDT Ar
Lv CLEVELAND " Lv
Lv Akron/Canton EDT Lv
Lv Columbus, Ohio EST Lv
Lv Dayton " Lv
Lv Charleston, W. Va. " Lv
Ar CINCINNATI " Ar
Lv CINCINNATI " Lv
Lv Indianapolis " Lv
Lv Knoxville EST Lv
Lv Louisville CDT Lv
Ar NASHVILLE CST Ar
Lv NASHVILLE " Lv
Ar MEMPHIS " Ar
Lv MEMPHIS " Lv
Lv Little Rock CST Lv
Lv Milwaukee CDT Lv
Ar CHICAGO (Midway) " Ar
Lv CHICAGO (Midway) " Lv
Ar CHICAGO (O'Hare) " Ar

Figure 3: Fragment of flight schedule American Airlines 1961

The center column displays the name of departure and arrival cities. The small columns on the sides display flights with departure and arrival time (local time, bold numbers represent PM). The top of the small columns shows the type of service provided (first class, coach or both), aircraft operated, days operated (daily if information is missing) and flight number.

We digitize flight schedules for the years 1951, 1956, 1961 and 1966 of six domestic airlines: American Airlines (AA), Eastern Airlines (EA), United Airlines (UA), Trans World Airlines (TWA), Braniff International Airways (BN), Northwest Airlines (NW).[12] This group of airlines includes the *Big 4*: AA, EA, UA and TWA, which accounted for between 69% and 74% of interstate air revenue passenger miles in the US in the years collected. BN and NW were digitized in order to have a wide geographical coverage,

[12]The selection of years was done based on data availability and with a criteria to be equally spaced. Patent data will be used aggregating in 5-year periods.

while PA provides international flights. Based on C.A.B. (1966), in the years collected, the six domestic airlines together accounted for between 77% and 81% of interstate air revenue passenger miles.

In total we have digitized 5,910 flights (unique combinations of flight number-year). However, flights often have multiple stops. If we count each non-stop part (*leg*) of these flights separately, our sample contains 17,469 legs. Our data connects 275 US airports creating 2,541 unique origin-destination (directional) airport links. Figure 4 displays the flight network in continental United States pooling all years together. In Appendix A.3 we show the US flight network by year, around 80% of the non-stop flights remain year-on-year.

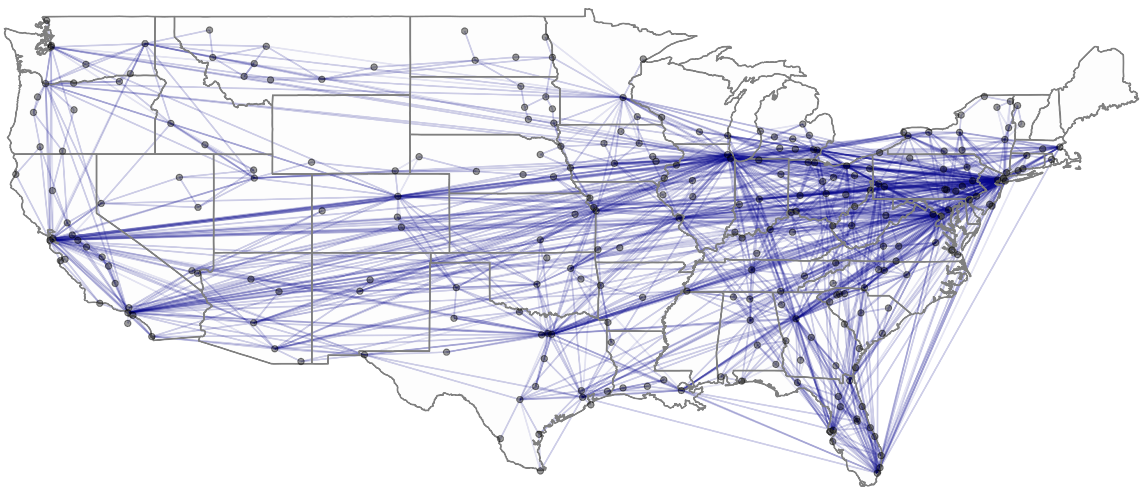


Figure 4: United States flight network 1951-1966

Using departure and arrival time of each flight at each airport, we obtain the fastest route and corresponding travel time between every two airports in our data. To obtain the fastest route and travel time we modify the Dijkstra algorithm to account for layover time in case the fastest route includes connecting flights.

Once the fastest route between every two airports is computed, we match every airport to 1950 Metropolitan Statistical Areas (MSA) using the shape file from Manson et al. (2020). We consider only MSAs in contiguous United States. We use MSAs as the geographical unit of analysis because they are constructed taking into account commuting flows. We assume that people in an MSA would use, for each desired route, the most appropriate airport lying inside or nearby the MSA. We match each airport to all MSAs for which it lies inside the MSA or is at most 15km away from its boundary.[13] 176 out of 275 US airports are matched to at least one MSA. Meanwhile, 142 out of 168 MSAs are matched to one or more airports in at least one year, and 108 MSAs are matched to one or more airports in the four years. We use the sample of 108 MSAs that are matched to at least one airport in the four years as our baseline travel time data.[14]

4.1. Descriptive statistics: Air travel

To understand the changes in travel time we will first study travel time of non-stop flights and then of all routes including connecting flights. Figure 5 displays the non-stop fastest flight within each MSA pair that was operating in each year. In 1951 the longest non-stop flight across MSAs was between Chicago and San Francisco using the Douglas DC-6, covering a distance of 2,960 km in 7 hours 40 minutes. This travel time was just under 8 hours, the maximum flight time allowed for a crew in a 24-hour period.[15] In 1956, new regulation allowed up to 10 hour flights for transcontinental flights, the longest non-stop flight between MSAs was New York to San Francisco with the Douglas DC-7, covering a distance of 4,151 km in 9 hours. Between 1951 and 1956 the main change observed is that longer non-stop routes were possible.

In 1961, the first year in which we have jet aircrafts in the travel time data, there is a reduction in travel time between MSA-pairs, especially for those far apart from each other. In 1966, there is a further decrease in travel time due to a widespread adoption of jet aircrafts in shorter distances. In Appendix Figure 21 we show the jet adoption

[13]The 15km distance was chosen after inspecting airports near the border that should arguably be matched, as for example, Atlanta ATL airport.

[14]In Appendix A.3 we include a table with the 168 MSAs, those connected at least once and those connected in the four years.

[15]Honolulu was not concerned by the regulation. Honolulu was connected with non-stop flights to San Francisco (9 hours 40 minutes), Los Angeles (11 hours) and Portland (12 hours 55 minutes).

rate by distance for MSAs connected with a non-stop flight. All MSA-pairs more than 3,000km apart connected with a non-stop flight operated at least one jet flight in 1961, and this expanded to all those more than 2,000km apart in 1966. The speed gain of jets relative to propeller aircrafts is increasing with the amount of time that the jet can fly at its cruise speed, arguing in favor of an adoption that is increasing with the distance between origin and destination.[16]

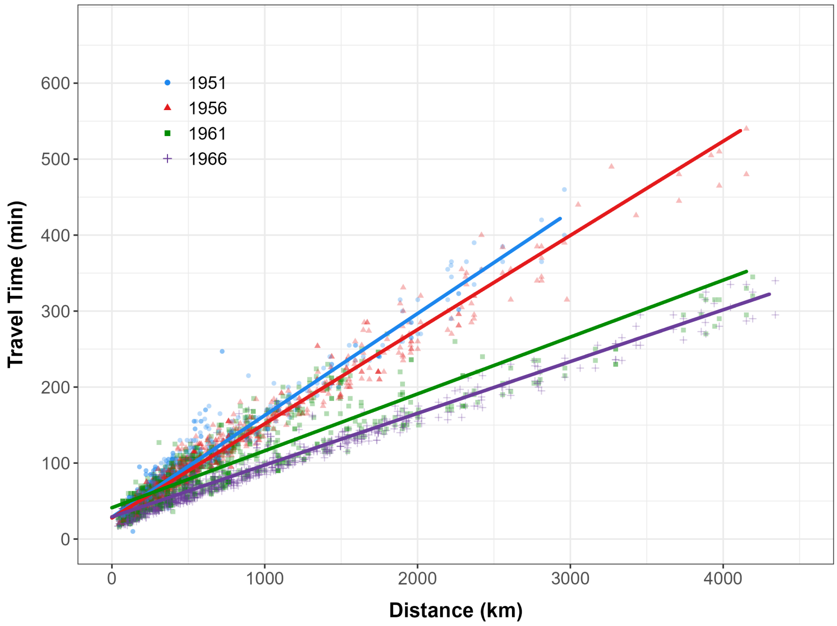


Figure 5: Non-stop fastest flights United States MSAs

The change in travel time in non-stop flights is also reflected in the travel time for connecting flights. Figure 6 shows, relative to 1951, the average and standard deviation change in travel time for all MSA-pairs, including non-stop and connecting flights.[17] Between 1951 and 1956, there is an average reduction in travel time of 9.2% which

[16]We are currently exploring the differential timing of jet adoption across airlines. Differences in (pre-existing) route distance and past contractual relationships with aircraft suppliers potentially led to different adoption rates at each time period. For example, Eastern Airlines' routes were particularly shorter than for other airlines. Also, those committed to buy Douglas airplanes (the leader US commercial aircraft supplier pre-jet era) would have adopted jets later, as Douglas launched jet airplanes later than Boeing.

[17]The plot includes only MSA-pairs with travel time in all time periods. The standard deviation for MSA-pairs less than 250km apart is not displayed because it distorts visualization is large relative to the ones at other distances.

is roughly constant for all distances over 500km. Between 1951 and 1961, there is a reduction in travel time that is increasing with distance. The average decrease in travel time is of 16.8%, while the reduction is of 29.4% for a distance of more than 2,000km and 39.2% for a distance of 4,250-4,500km. Between 1951 and 1966, there is an even stronger decrease in travel time at all distances. The average reduction in travel time is 28.7% across all distances, 40.8% for a distance of more than 2,000km and 48.4% for a distance of 4,250-4,500km. The increased adoption of jets for short distance flights implied that both non-stop flights at short distance and connecting flights at farther distance had a decrease in travel time.

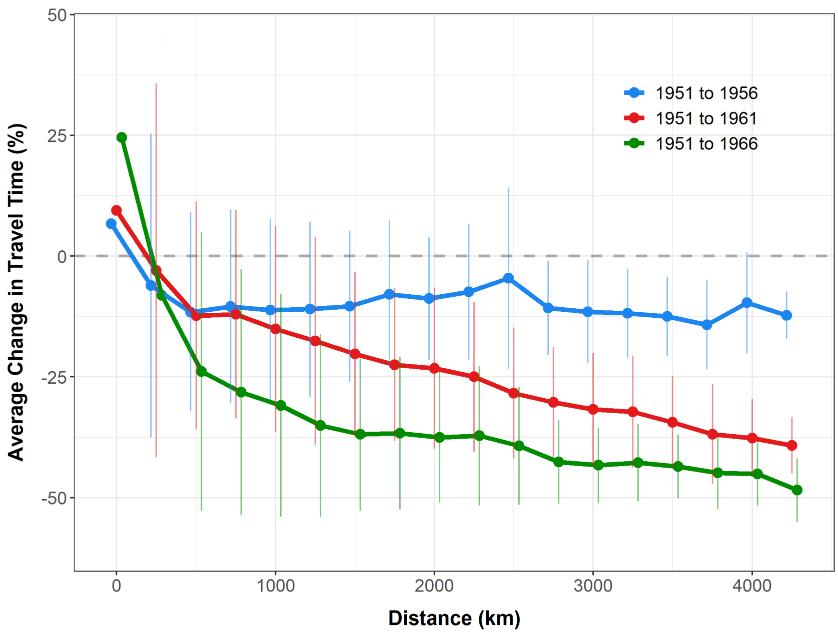


Figure 6: Change in MSAs travel time

Figure 24 in Appendix A.3 shows that the change in travel time is accompanied by a reduction of the amount of legs needed to connect two MSAs at every distance. This reduction is especially marked between 1951 and 1956, and 1961 and 1966. Between 1956 and 1961, we do not observe a big reduction in the amount of legs, implying that the decrease in travel time observed in Figure 6 between 1956 and 1961 comes from a source other than the amount of legs. In Appendix Figure 25 we open up the change in travel time by the way an MSA pair was connected in 1951 and 1966: either directly (non-stop flight) or indirectly (connecting flight). We observe that much of the increase in travel time for MSA pairs less than 250km apart comes from routes that in 1951 were operated non-stop while in 1966 were operated with connecting flights.[18] Interestingly, for MSA-pairs more than 2,000km apart travel time reduced on average 42% for those pairs that were connected indirectly in both periods, and 51% for those that switched from indirect to direct. This fact shows the relevance of improvements in flight technology even for MSAs that were not directly connected.

It could be the case that a reduction in the amount of legs or an increase in frequency of flights reduces layover time, which then translates into a reduction of travel time. In Appendix Figure 27 we compare the change in travel time from 1951 to 1966 with a counterfactual change in travel time in which we eliminate layover time in both time periods. We observe that the average change in travel time is stronger at every distance in the counterfactual scenario without layover time. This implies that the relative importance of layover time to total travel time within a route increased between 1951 and 1966, so total travel time did not decrease proportionally to the change of in-flight travel time. In short, layover time attenuated the reduction in travel time.

4.2. Constructing an instrument

In this section we construct an instrumental travel time that is based on the pre-existing flight routes and the time-varying nationwide roll out of jets. In this way, the instrument abstracts from the endogenous decisions of two agents: First, regulator's decision on the opening/closure of routes. Second, airlines' decision about to which routes allocate jet vs propeller airplanes and scheduling (frequency of flights and layover time). The key source of variation in the instrument is that as the speed of airplanes increases,

[18] Appendix Figure 26 repeats the exercise discarding layover time in all time periods. By comparing Figure 25 and Figure 26 we can disentangle the effect of layover time and the change in in-flight time. For MSA pairs less than 250km that changed from direct to indirect connection, 80% of the increase in travel time is due to the increase in layover time (which was previously zero as it was a non-stop flight), and 20% is due to the increase of in-flight time.

the importance of the number of stops relative to the distance flown changes, and this provides a decrease in travel time that is larger for MSA-pairs located farther apart. We first explain the idea and identifying assumptions of the instrument, and then we detail how it is constructed.

In Borenstein and Rose (2014) it is argued that, due to strict regulation, it was difficult for airlines to adapt their flight network when technology to fly changed. However, we may be concerned that the decision of the regulator to grant new routes could be targeted to specific pairs or correlated with unobservable variables that also affect the creation and diffusion of knowledge.[19] Hence, as the first step in the construction of our instrument, we *fix routes* to the ones we observe in 1951. In this way the instrumental travel time is computed only using non-stop flights present in 1951, and does not consider appearance or disappearance of non-stop flights in the data. The identifying assumption is that the network of flight routes in 1951 did not yet include the changes that would be optimal to operate with jet airplanes. In other words, we require that the regulator did not change routes already by 1951 in anticipation of the arrival of jet airplanes.[20,21]

Airlines could decide on two factors that affect travel time: the type of airplane (jet vs. propeller) operated in each route and scheduling, which consists on the frequency of flights and layover time in case of connecting flights.[22] We may be concerned that, as with the regulator, airlines' decisions could be correlated with unobservables that also affect the creation and diffusion of knowledge.[23] The second step in the construction of

[19]For example, the regulator could have targeted the opening of new routes between places in order to boost their economic activity.

[20]For example, in the instrument there are no non-stop transcontinental routes.

[21]In our estimations we exploit time variation for identification. Hence, if pre-existing routes affect the levels at the origin-destination level, this does not drive our identification. However, we may be concerned that pre-existing routes could affect future growth and not only levels. To address this concern, in robustness analysis we estimate the elasticity of citations to travel time using only MSA-pairs that are always indirectly connected. Results are consistent with baseline estimation.

[22]In 1961, all non-stop flights of more than 3,000km had at least one jet operating within them, while in 1966 it was the case in all non-stop flights of more than 2,000km. Therefore the endogeneity of jet adoption is a smaller concern for long distance flights.

[23]For example, airlines may have decided to prioritize the allocation of jets to routes which had a higher

our instrument is to discard layover time (hence discarding all scheduling decisions) in all time periods, and assume that in each year all routes are operated with a *fictitious average airplane* of the year. Hence, the change in instrumental travel time in a route is independent of the type of airplane used in the route and it only depends on the nationwide roll out of jets. The identifying assumption is that no single route had the power to shift the average speed of the year.

To construct the instrumental travel time we first estimate, separately for each year, a linear regression of travel time on flight distance using only the fastest non-stop flight in each origin-destination airport pairs.[24] These yearly regressions provide us with the fictitious average airplane of each year: the intercept gives the take-off and landing time of the airplane while the slope provides the (inverse) speed. Second, we fit these regressions to obtain predicted travel time in each non-stop flight and year. Third, for each year, we compute the fastest travel time using the Dijkstra algorithm. The Dijkstra algorithm looks for the fastest path using only 1951 non-stop flights, while the travel time in each non-stop flight in each year is given by the predicted travel time from the previous step. Layover time is set to zero in all years.

The key source of variation in the instrument is the time varying relative importance of in-flight travel time relative to the number of stops required to go from one MSA to another. In shorter flights, the amount of stops have a larger share of the total instrumental travel time and changes in flight speed have less of an influence. In long distance flights the flight speed becomes more relevant. As estimated flight speed more than doubles over the time period, longer flights have a larger reduction in travel time in the instrument. However, differences in the amount of stops required also leads to variation in changes of travel time among long distance routes. Long distance

share of business travel, which may be correlated with the diffusion of knowledge.

[24]The use of a linear regression is motivated by the linearity between travel time and distance displayed in Figure 5. To estimate these regressions we use all routes appearing in each year. Results of these estimations presented in Appendix Table 11 show that the implied average flight speed increases from 412 kmh in 1951, to 453 kmh in 1956, 758 kmh in 1961 and 876 kmh in 1966. On the other hand the intercept fluctuates, going from 25.3 minutes in 1951 to 29.9 minutes in 1966.

routes and routes with less amount of stops have a larger reduction of travel time in the instrument.[25]

Figure 7 shows the percentage change in observed and instrumental travel time relative to 1951. We compute the percentage change within each MSA-pair for each year and then take averages within 250km bins. We observe that the instrumental travel time follows pretty closely the observed change in travel time in each year. Especially, it replicates the pattern of a stronger decrease in travel time for MSAs located farther apart. It is only for MSAs less than 500km apart that the change in the instrumental travel time departs from the observed change.[26] This finding shows that most of the change in travel time that we observe is due to the change in speed of airplanes, and that the endogeneity concern is limited for MSAs located far away from each other.

[25]Using the coefficients in Appendix Table 11 the instrumental travel time for a pair of airports located 300km apart connected non-stop in 1951 and 1966 would be 69.1 minutes and 50.3 minutes in each respective year, implying a 27.2% reduction in travel time. In the case of a pair of airports located 2,000km apart connected non-stop the instrumental travel time would be 317.3 minutes and 165.9 minutes, implying a 47.7% reduction in travel time. Now lets assume that in both years both pair of airports had 2 intermediary stops that layed in a straight line in between the origin and destination airports (such that origin-destination distance and travel distance are the same). In the case with two stops, the reduction in travel time would have been 8% for the pair of airports 300km apart and 38.6% for those 2,000km apart.

[26]We observe an increase in travel time for short distances in 1961 relative to 1951. Given that non-stop routes are fixed and that for longer distances there is a decrease in travel time, the increase in travel time in short distances comes from an increase in the value of the intercept relative to the slope in 1961, relative to 1951.

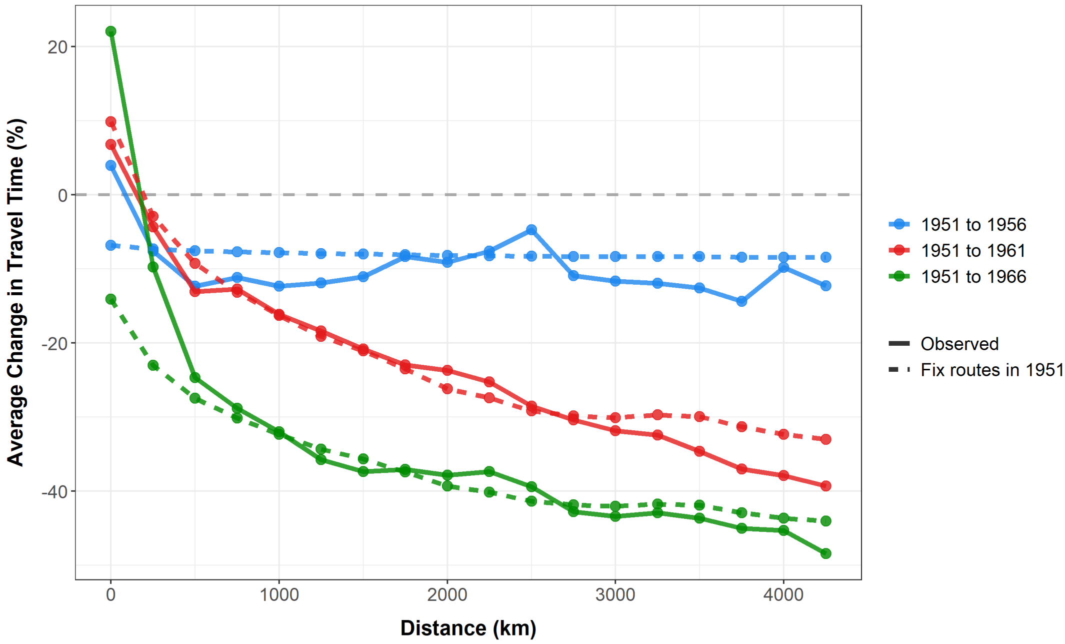


Figure 7: Instrumental Travel Time between US MSAs.

In Appendix A.3 we present other two counterfactual travel times: one in which we fix airplanes to be the average airplane of 1951 and allow routes to evolve, and another in which both the average airplane and routes are varying. These two counterfactuals together with the one presented in this section allow us to decompose the change in travel time by the change in routes and the change in speed of airplanes. We obtain that around 90% of the change in travel time is due to the change in speed of airplanes, while around 10% of the change is due to the change in the flight routes. Appendix Figure 29 shows that the share is roughly constant for all distances. This finding confirms that most of the observed changes in travel time are due to improvements in flight technology.

5. Patent data

We use patent data as our source of innovation information. We construct a dataset of all patents granted by the United States Patent and Trademark Office (USPTO) with filing year between 1949 and 1968, which includes for each patent: filing year, technology classification, location of the inventors when they applied for the patent, owner of the patent and citations to other patents also granted in the United States.[27] This dataset provides the distribution of patents and citations over the geographic space.

To construct the patent dataset we downloaded from Google Patents all patents granted by the USPTO with filing year between 1949 and 1968. This dataset contains patent number, filing year and citations.[28,29] Based on the patent number we merge it with multiple datasets. First, we obtained technology class from the USPTO Master Classification File and we aggregated them to the six technology categories of Hall et al. (2001).[30] Second, we obtained geographic location of inventors from three datasets: HistPat (Petralia et al. (2016)) and HistPat International (Petralia (2019)) for patents published until 1975, Fung Institute (Balsmeier et al. (2018)) for patents published after 1975.[31] We match all inventors' locations to 1950 Metropolitan Statistical Areas (MSAs) in contiguous United States. To do the match we obtain geographical coordinates from the GeoNames US Gazetteer file and Open Street Maps, and use the MSAs shape file from Manson et al. (2020). Third, we obtain ownership of patents from two sources: Kogan et al. (2017) for patents owned by firms listed in the US stock market and Patstat (Magerman et al. (2006)) for the remaining unmatched patents.[32]

[27] Filing year, also called application year, is the closest date to the date of invention that is present in the data and it represents the date of the first administrative event in order to obtain a patent. In the other hand, the publishing (also called granting year) is a later year in which the patent is granted. The difference between filing and publishing year depends on diverse non-innovation related factors (as capacity of the patent office to revise applications) and changes over time. Hence filing year is the date in our data that approximates the best to the date of invention.

[28] Very few patents had missing information on filing year. We complemented both missing filing year and citations with the OCR USPTO dataset.

[29] The patent citation record starts in 1947, year in which the USPTO made it compulsory to have front page citations of prior art. Gross (2019)

[30] USPTO Master Classification File: `https://www.google.com/googlebooks/uspto-patents-class.html`

[31] Due to the gap between the filing year and publishing year we also do the matching to patents published after 1968. Our underlying patent data actually covers a longer time period of filing years, which we need for example to construct forward and backward citation lags. However, there are limitations to use the geographic data in filing years 1971-1972. In Appendix B we show that during filing years 1971-1972 the rate of unmatched patents to inventors' location increases. This is probably due to Histpat and Fung data not being a perfect continuation of one another.

[32] Patent ownership in both datasets comes from the patent text, which is self declared by the patent applicant. Particularly, Kogan et al. (2017) does not explicitly state if it takes into account firm-ownership structure to determine the ultimate owner of a patent, neither does Patstat.

For the descriptives presented below and the posterior analysis we truncate and aggregate the data in the following way.[33] We drop patents that are owned by universities or government organizations. To count patents that are classified into multiple technology categories, we do a fractional count by assigning proportionally a part of the patent to each category. Citations are counted as the multiplication of the technology weight of the citing and cited patents. We drop patents (and their citations) that have inventors in multiple MSAs and citations in which the citing owner is the same as the cited owner.[34,35]

We aggregate the patent data to 4 time periods of 5 years each, with the center of each period being the year of travel time data collected. The periods are: 1951 (which contains the years 1949-1953), 1956 (1954-1958), 1961 (1959-1963) and 1966 (1964-1968). We consider only patents in MSAs that are matched to an airport in the four periods.[36] The final dataset contains 108 MSAs with patents and travel time.

5.1. Descriptive statistics: Patents

This section presents three facts about US patents over our sample period: First, initially less innovative locations had a higher patenting growth rate. The average yearly growth rate of locations in the lowest quartile of initial innovativeness was 7.2% while it was 1.9% for those in the highest quartile. Second, high growth locations were also primarily in the South and the West of the US. The South and the West grew three times as fast as the Midwest and the Northeast. Third, the mass of citations shifted towards longer distances. While the first quartile of citation distance remained relative stable over the time period, the third quartile increased its distance by 39%. At the same time,

[33] Details on sample selection are presented in Appendix.

[34] 3% of patents have inventors in more than one MSA. Working with multi-MSA patents requires an assumption on how to compute distance and travel time between the citing and cited patents, as they do not have a single origin-destination location pair. We hence prefer to abstract from multi-MSA patents.

[35] Incentives to self-cite may be different than to cite patents of other owners.

[36] We drop around 9% of patents that are in MSAs which are not matched to an airport in the four time periods. Descriptive statistics including those patents are similar to the ones presented here.

the share of citations at more than 2,000km increased by 30%.

We compute descriptive statistics by technology category. In here we present descriptives of averages across technologies. Technology specific descriptives are included in Appendix B.

Fact 1: Initially less innovative locations had a higher patenting growth rate

In the period 1951 to 1966 we observe that the highest growth of patenting takes place in locations that were initially less innovative. The differential growth rate implies a convergence rate of 5.3% per year.

Figure 8 shows the geographic distribution of patenting in 1951. Darker colors refer to a higher level of *initial innovativeness*, which is defined as the amount of patents filed by inventors in the MSA in 1951.[37] We observe that MSAs in the top quartile of patenting are concentrated in the Northeast (which includes New York) and the Midwest (which includes Chicago), with few additional MSAs in the West.[38,39]

Figure 9 shows the geographic distribution of patenting growth in 1951-1966.[40] We observe a striking pattern relative to Figure 8: high growth MSAs were those that were initially less innovative. High growth happens in initially less innovative locations in the South and the West but also in the Northeast. We confirm this pattern in Figure 10, which shows the MSA's ranking of innovativeness in 1951 and its subsequent

[37] To compute the level of initial innovativeness we only use patents filed in 1951 (years 1949-1953). We aggregate patents to the MSA-technology level and then compute the quantile-position of each MSA in the technology. Lower values of quantile-position refers to lower amount of patents in the technology (relative to other MSAs). Each MSA has a different value of quantile-position in each of the 6 technology categories. To obtain the MSA level quantile we take the average quantile across technologies within the MSA. Finally we classify MSAs into quartiles depending on whether the average quantile is higher or lower than the thresholds 0.25, 0.50, 0.75.

[38] In Appendix B we show that the 1951 geographic distribution of patents looks similar across technology categories.

[39] The top 5 patenting MSAs in 1951 were: New York City (25% of all patents), Chicago (11%), Los Angeles (8%), Philadelphia (6%) and Boston (4%).

[40] We compute the growth rate of patenting in each technology within a MSA and then take the average across technologies within the MSA.

patenting growth rate in 1951-1966. Figure 10 shows that MSAs that were initially more innovative (lower values in the ranking) are those that saw lower values of subsequent patenting growth.[41,42] We estimate a linear regression with an intercept and a slope, and find that the slope is positive and statistically different from zero. At the mean, lowering initial innovativeness by 10 positions in the ranking was associated with a subsequent 0.42 percentage points higher yearly growth rate of patenting.

Figure 10 presents average growth rates across technologies within a MSA.[43] The average yearly growth rate of MSA-technologies in the lowest quartile of initial innovativeness is 7.2% while it is 1.9% in the highest quartile.[44] The percentage point difference between the two growth rates implies that locations in the lowest quartile converged towards locations in the highest quartile at a speed of 5.3% per year.[45] The convergence in patenting across MSAs is consistent with *The Postwar Decline in Concentration, 1945-1990* described in Andrews and Whalley (2021).

[41]Each dot in Figure 10 is an MSA. To compute the MSA ranking we need to double-rank MSAs. First we rank all MSAs in each technology. Second we take the across-technology average ranking of each MSA. Third we rank all MSA's averages. To compute the MSA's yearly growth rate we first take the 1951-1966 growth rate for each technology in the MSA. We then take the average across technology. Finally we obtain the MSA's yearly growth rate by computing: $yearly\_growth\_rate = (1 + 19\_year\_growth\_rate)^{(1/19)} - 1$ (the 1951 to 1966 period is a 20 year window, we take growth rates as being from the first year 1949 to the last one 1968, which is 19 year growth).

[42]In Appendix B we replicate the plot differentiating geographic regions. MSAs that were initially less innovative and had high subsequent growth were located in all four regions, although they were primarily located in the South and the West.

[43]There are multiple ways to compute averages when there are multiple dimensions: MSAs, technologies, quartiles of initial innovativeness. We obtain a result that goes in the same direction if we compute the average growth rates across MSAs within a technology and quartile of initial innovativeness, and then take the average across technologies.

[44]We first compute the 1951-1966 growth rate (19-year growth rate) for each MSA-technology. We then take averages across MSAs within a quartile-technology, and after take averages across technologies within a quartile. Finally, we convert the 19-year growth rate into an average yearly growth rate.

[45]We note that the aggregate growth of patents is much smaller than the across MSAs unweighted average, and this is exactly because initially less innovative MSAs grew faster. If we compute the growth rate in nationwide amount of patents in each of the technologies and then average across technologies we obtain a yearly growth rate of 1.5%.

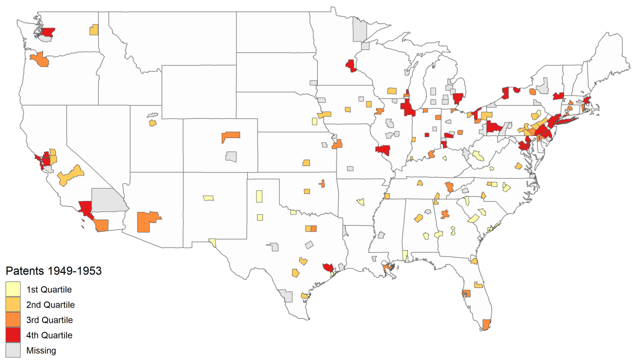


Figure 8: Geography of Patenting 1951

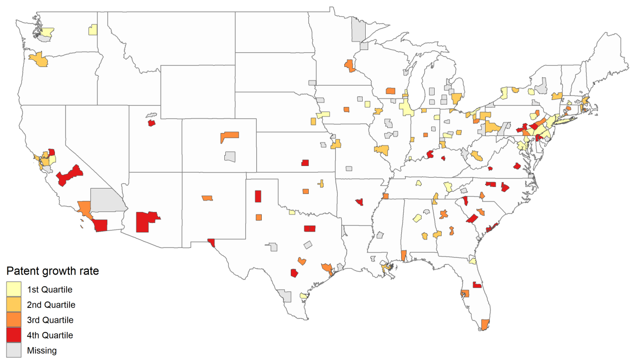


Figure 9: Patent growth 1951-1966

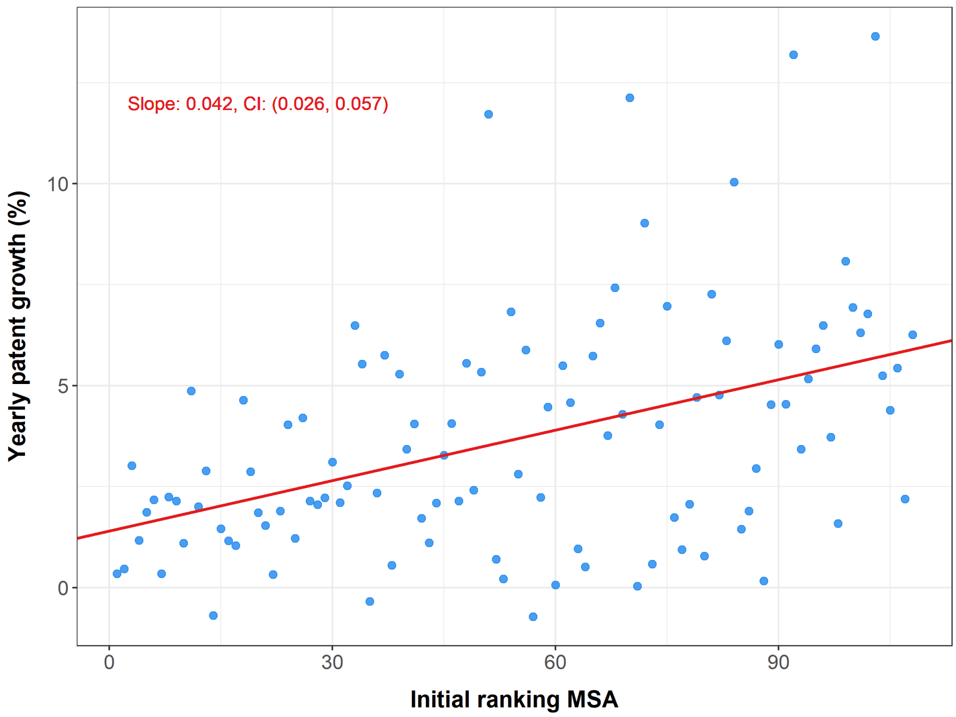


Figure 10: Patent growth rate by initial innovativeness ranking of MSA

Fact 2: The South and the West of the US had a higher patenting growth rate

Figure 9 shows that MSAs located in the South and the West of the US had a higher patenting growth rate in 1951-1966. We classify MSAs using Census Regions of the US (Midwest, Northeast, South and West) and aggregate patents within each region-technology-year.[46] Figures 11 and 12 present averages across technologies within a region-year. Figure 11 shows that the share of patents filed by inventors located in the Midwest and the Northeast decreased from 75% in 1951 to 68% in 1966, while the share of patents filed in the South and the West increased from 25% to 32%. The change in the shares implies a higher growth rate of patenting in the South and the West relative to the Midwest and the Northeast.

Figure 12 shows that in the period 1951-1966 the South and the West increased their amount of patenting by 80%, while the Midwest and the Northeast had a 22% growth.[47] Translated into yearly growth rates, the South and the West grew three times as fast as the Midwest and the Northeast (3.14% vs. 1.05% per year).[48]

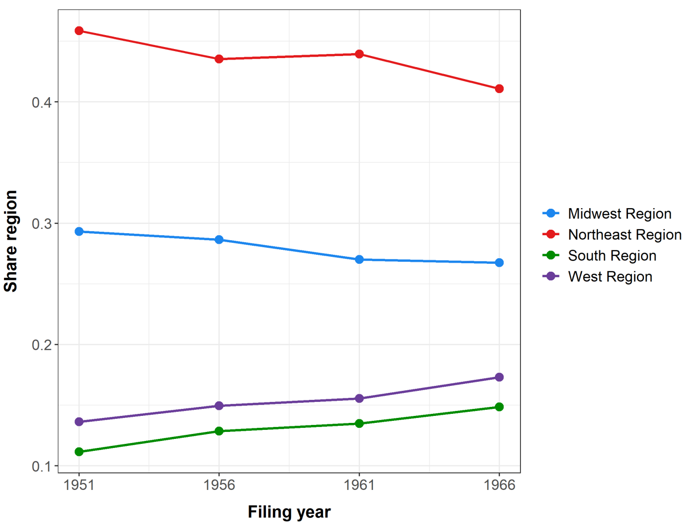


Figure 11: Share of patents by region

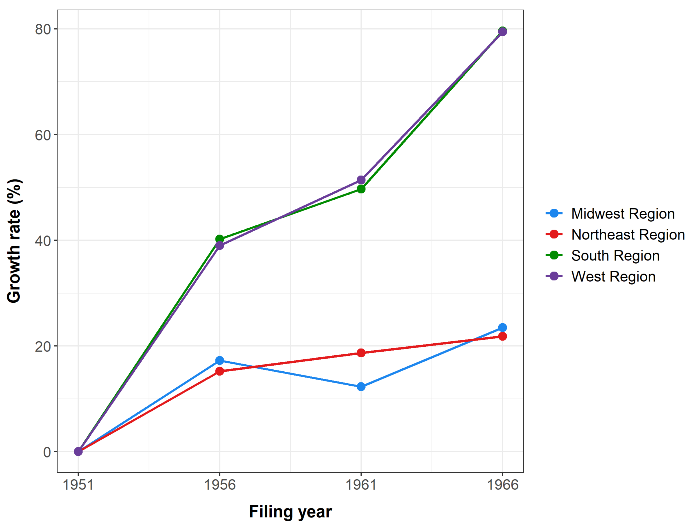


Figure 12: Patent growth by region

Fact 3: Distance of citations increased

In our analysis we use citations as a proxy for knowledge diffusion. According to Jaffe et al. (1993) *"a citation of Patent X by Patent Y means that X represents a piece of previously*

[46] In Appendix C we present a map with the four Census Regions. Some MSAs belong to multiple Census Regions. In here we present descriptive statistics duplicating such MSAs (assigning the MSA to both Census Regions). Statistics dropping such MSAs are quantitatively similar.

[47] Growth rates are computed by region-technology and then averaged across technologies within region.

[48] $3.14\% = 1.80^{(1/19)} \times 100$, $1.05\% = 1.22^{(1/19)} \times 100$

existing knowledge upon which Y builds." (page 580).[49] We compute the distance between the citing inventor and the cited inventor. Figure 13 shows the evolution over time of the first, second and third quartile of citation distance.[50] We observe that 25% of citations happened between inventors located less than 300km apart throughout our sample period. For the middle 50% of citations we observe that over time inventors cited other inventors located farther away. The third quartile of citation distance increased from 1,642km in 1951 to 2,284km in 1961, a 39% increase in the distance.[51] In other words, the mass of citations shifted towards longer distances.

In Figure 14 we present the share of citations by distance range between the citing and cited inventors.[52] The distance cutoffs where chosen in order to have a balanced share of citations in the initial time period, and considering the changes in travel time presented in Section 4.1. The share of citations that happen between inventors located more than 2,000km apart grew from 21.5% in 1951 to 27.9% in 1966. The 6.4 percentage points increase represents an increase of 30% of the share of citations at more than 2,000km.

[49] Jaffe et al. (1993) discusses the reasons why to cite and why not to cite. Using a survey of inventors, Jaffe et al. (2000) find that there is communication among inventors and citations are a *"noisy signal of the presence of spillovers."*

[50] We compute distance between MSA centroids.

[51] As a reference, the distance from New York City NY to other places is: Boston MA 300km, Chicago IL 1,140km, Dallas TX 2,200km, San Francisco CA 4,130km. The quantile 0.10 of was at 0km in every period, implying that 10% of citations took place within MSA. The quantile 0.90 was between 3,611km and 3,789km over the sample period.

[52] While Figure 13 shows how the distance of each quantile changes over time, Figure 14 shows the mass of citations (and hence the quantile to which belongs) in a certain distance cutoff. For example, in 1951 the share of citations in the 0-300km range was 31.6%, which is equal to saying that the quantile 0.316 in 1951 was 300km.

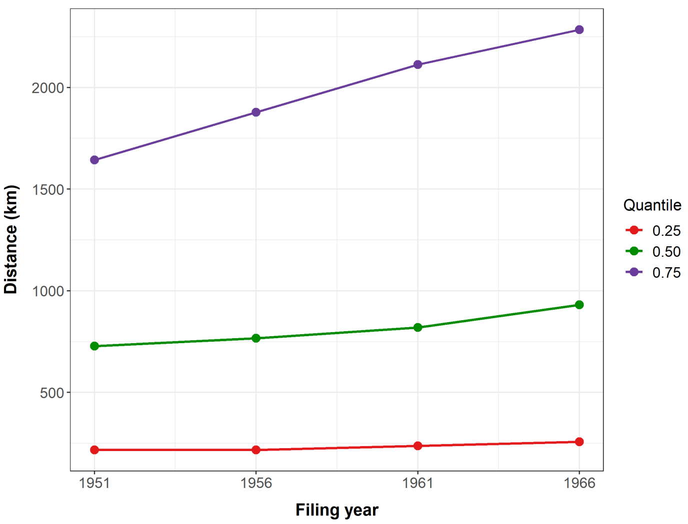


Figure 13: Quantiles of citation distance

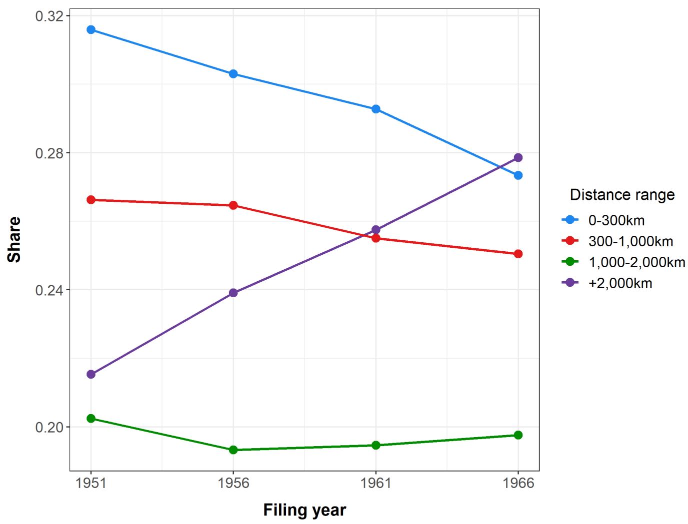


Figure 14: Share of citations by distance

6. Diffusion of knowledge

In this section we show that the reduction in travel time led to an increase in knowledge diffusion, especially over long distances. In doing so we estimate the parameter β highlighted in equation (2): the elasticity of knowledge diffusion to travel time.

To perform the analysis we merge the Air Travel and Patent datasets to obtain a final dataset that contains for each patent owner-location, the amount of patents filed in a certain 5-year period and technology class, the amount of citations to other patents with their respective owner identifier, location and technology class, and the travel time to every location. We label a patent owner a *firm* and call *research establishment* a firm-MSA pair for MSAs in which it has inventors applying for patents. We aggregate citations to the citing-cited establishment-technology within each period. We assume that passengers take a return flight, hence we make travel times symmetric.[53]

[53] $travel\ time_{ijt} = \frac{travel\ time_{ijt}^{original} + travel\ time_{jit}^{original}}{2}$ where $travel\ time_{ijt}^{original}$ stands for the travel time between MSA i and j at time period t.

6.1. Diffusion of knowledge: Baseline estimation

We estimate a gravity equation which relates citations between two establishments-technologies with their pairwise travel time.[54] We estimate the following regression:

$$citations_{FiGjhkt} = \exp\left[\boldsymbol{\beta}\, log(\text{travel time}_{ijt}) + FE_{FiGjhk} + FE_{Fiht} + FE_{Gjkt}\right] \times \varepsilon_{FiGjhkt} \quad (3)$$

where $citations_{FiGjhkt}$ is the amount of citations from patents filed by the establishment of firm F in location i, technology h and time period t, to patents filed by establishment of firm G in location j and technology k. We call Fi the research establishment of firm F in location i. travel time$_{ijt}$ is the air travel time (in minutes) between location i and j at time period t. The parameter of interest in the regression is β, which represents the elasticity of citations to travel time.[55] If citations are affected negatively by travel time we would expect a negative value of β.

Given the panel structure of our data, we can include the fixed effect FE_{FiGjhk} that absorbs any time invariant citation behavior within the *citing establishment-technology and cited establishment-technology*. This fixed effect flexibly controls for persistent relationships within an establishment pair that would lead to relatively more (or less) citations. That includes characteristics like physical distance, but also pre-existing commercial relationships between establishments. The fixed effects FE_{Fiht} and FE_{Gjkt} control for the time changing general level of citations specific to each establishment and technology. For example FE_{Fiht} controls for the fact that if Fih files more patents in a given period, it would mechanically make more citations to every establishment. On the other hand, FE_{Gjkt} controls for Gjk filing more patents or higher quality patents that would receive more citations from every establishment.[56]

[54] For explanation and micro foundations of the gravity equation see Head and Mayer (2014) and references thereof.

[55] A 1 percent increase in travel time has an effect of β percent increase (or decrease in the case of a negative β) in citations.

[56] In the International Trade literature, the parallel of the fixed effects (simplified for exposition) would be: FE_{ij} country-pair fixed effect, FE_{jt} origin-time fixed effect and FE_{it} destination-time fixed effect.

The inclusion of FE_{FiGjhk} implies that only variation across time within an establishment-pair is used for identification. By additionally including the fixed effect FE_{Fiht}, the across-time variation is compared only between citing-cited establishment-technology pairs $FiGjhk$ within a citing establishment-technology Fih in period t. As we also include FE_{Gjkt}, the comparison is done while controlling for the size of the cited establishment-technology Gjk in period t. Put differently and simplifying slightly, the identification of β relies on changes in citations and travel time within an establishment-pair, relative to another establishment-pair with the same citing establishment, conditional on the two cited establishments' sizes.

Following Silva and Tenreyro (2006), we estimate the gravity equation by Poisson Pseudo Maximum Likelihood (PPML).[57] This estimation methodology has two advantages over a multiplicative model that is then log-linearized to obtain a log-log specification. First, it only requires the conditional mean of the dependent variable to be correctly specified, while the OLS estimation of the log-linearized model would lead to biased estimates in the presence of heteroskedascity. Second, it allows to include zeros in the dependent variable, which is especially relevant when using disaggregated data. One downside of estimating PPML with the fixed effects that we include is that both coefficients and standard errors have to be corrected due to the incidental parameter problem (Weidner and Zylkin (2021)). We follow Weidner and Zylkin (2021) to use split-panel jackknife bias-correction on the coefficients and Dhaene and Jochmans (2015) to bootstrap standard errors which we also bias-correct with split-panel jackknife.[58]

Whenever $FiGjhk$ has positive citations in at least one period and missing value in another, we impute zero citations in the missing period.[59] Travel time is set to one minute whenever $i = j$.[60]

[57]We use the package *fixest* (Bergé (2018)) in R to estimate high dimensional fixed effects generalized linear models *feglm* with Poisson link function.

[58]Details on the bias correction and bootstrap procedures are provided in Appendix D.

[59]We do not impute zeros in $FiGjhk$ that are always zero, as those observations would be dropped due to not being able to identify FE_{FiGjhk}.

[60]We measure air travel time in minutes. In our sample 13% of citations happen within the same MSA.

	PPML		IV PPML	
Dep. variable:		*citatitions*		
	(1)	(2)	(3)	(4)
log(travel time)	-0.083^{***} (0.019)		-0.152^{***} (0.029)	
log(travel time) $\times$ 0-300km		0.019 (0.036)		-0.076 (0.221)
log(travel time) $\times$ 300-1,000km		-0.089^{***} (0.023)		-0.134^{***} (0.044)
log(travel time) $\times$ 1,000-2,000km		-0.094^{***} (0.033)		-0.112^{**} (0.047)
log(travel time) $\times$ +2,000km		-0.169^{***} (0.039)		-0.203^{***} (0.043)
Control residuals 1st stage	-	-	Yes	Yes
N obs. effective	$4,703,010$	$4,703,010$	$4,703,010$	$4,703,010$
R2	0.88	0.88	0.88	0.88

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$

Table 1: Elasticity of citations to travel time

Column (1) shows the result of Poisson Pseudo Maximum Likelihood (PPML) estimation of $citations_{FiGjhkt} = \exp\left[\boldsymbol{\beta}\, log(\text{travel time}_{ijt}) + FE_{FiGjhk} + FE_{Fiht} + FE_{Gjkt}\right] \times \varepsilon_{FiGjhkt}$, for citations of patents filed by establishment of firm F in location i, technology h and time period t, to patents filed by establishment of firm G in location j and technology k. travel time$_{ijt}$ is the travel time in minutes between location i and j at time period t, and it is set to 1 when $i = j$. When $FiGjhk$ has positive citations in at least one period and no citations in another, we attribute zero citations in the missing period. Column (2) includes the interaction of travel time$_{ijt}$ with a dummy for distance bin between the citing establishment Fi and the cited establishment Gj. Column (3) and (4) show the result of two step instrumental variables estimation, where $log(\text{travel time}_{ijt})$ is instrumented with $log(\text{travel time}_{ijt}^{\text{fix routes}})$, the travel time that would have taken place if routes were fixed to the ones observed in 1951 and in each year routes were operated with the average airplane of the year. Columns (3) and (4) include as controls residuals of first stage. Bootstrap standard errors are presented in parentheses. The coefficients and standard errors in columns (1) and (2) are jackknife bias-corrected. R2 is computed as the squared correlation between observed and fitted values.

Column (1) in Table 1 presents the results of estimating equation (3). The value of the elasticity of citations to travel time is estimated to be -0.083, statistically significant at the 1% level. Given the average reduction in travel time of 31.4% in the full estimating sample, the elasticity implies that citations increased on average 2.6% as consequence

The inclusion of those citations in the estimation increases the amount of observations available to identify of FE_{Fiht} and FE_{Gjkt}, and hence keeping them increases the amount of $FiGjhkt$ that remain in the effective sample to identify β. In order to include them we then need to impute a within-location travel time. We assume that within-location (air) travel time is not changing across time periods. Nonetheless, the identification of β is not affected by the value chosen for the within-location (time invariant) travel time, as β is identified by across time variation. In the appendix we show results using other values of (time invariant) within MSA travel time and the coefficients remain equal.

of the reduction in travel time. If we consider the average decrease in travel time across all MSAs in the baseline travel time data, the implied increase is 2.4%.[61]

The importance of air transport relative to other means of transport potentially depends on the distance to travel. Also, we observed in Section 4.1 that the improvements in air travel time depended on the distance to travel, with a difference in jet adoption for travel distances under and over 2,000km. Taking these two characteristics into account, we estimate a variation of equation (3) in which we allow the elasticity of citations to travel time to vary by distance interval between the locations of citing and cited establishments.[62] Column (2) in Table 1 shows the result of this estimation.[63] The estimated value of the elasticity in absolute terms increases with distance, reaching -0.169 for distances of more than 2,000km. Between 1951 and 1966 the average change in travel time in the full estimating sample is 47.7% for a distance of more than 2,000km. The estimated elasticity implies that citations between establishments at more than 2,000km apart increased by 8.1% due to the decrease in travel time. In total citations at more than 2,000km increased by 21%, implying that the change in travel time can account for 38.2% of the observed increase. If instead we consider the 40.8% average reduction in travel time across MSAs in the baseline data, the elasticity implies an increase in citations of 6.9%, accounting for 32.7% of the total citation increase.

In Appendix B we investigate different heterogeneous effects. We estimate an heterogeneous elasticity depending on the level of spatial concentration of the citing technology and the cited technology, we do not find a statistical difference. We also look at whether it is older patents or younger patents that get diffused, finding some slight evidence that it is technologies that take longer time to diffuse that increase more their diffusion with the reduction in travel time. We study citations to and from government

[61] These values come from the multiplication of the elasticity of citations to travel time 0.083 and the average decrease in travel time between 1951 and 1966: 31.4% in the full estimating sample and 28.7% in the raw data of travel time across MSAs.

[62] We compute distance between the geographical center of each MSA.

[63] The share of observations (citations) in each distance interval is: 0-300km 26.1% (28.5%), 300-1,000km 30.7% (28.5%), 1,000-2,000km 19.7% (23.4%), +2,000km 23.4% (19.6%).

patents, and self citations, on the whole we do not find a different pattern from the baseline. We also do not find a particular pattern of the elasticity depending on the citing *firm's size* as measured by the amount of patents filed in 1949-1953. Finally, we estimate the elasticity by citing and cited technology and most of the effect seems to come when the citing and cited technologies are the same.

There are two types of threats to identification in estimating equation (3): (i) the potentially targeted changes in travel time, which could be due to the opening of new routes, the allocation of jets across routes, or changes in scheduling, and (ii) time changes in other variables at the MSA-pair level which also drive the diffusion of knowledge and are correlated with the changes in travel time. In the remaining of this section we address the first type of threat by estimating the model by instrumental variables. In the following subsection we address the second type of threat by adding multiple controls. In both cases we show that results do not qualitatively change.

6.2. Diffusion of knowledge: Instrumental variables estimation

As mentioned in Section 4.2, we may be concerned that the timing and allocation of jets to routes and that the opening/closure of routes were not random. In case there is an omitted variable that drives both the change in travel time at the MSA-pair level and the change in citations across establishments within the same MSA-pair, we would estimate biased coefficients. In order to tackle the endogeneity concern due to omitted variable we do an instrumental variables estimation using the instrument proposed in Section 4.2. To implement the instrumental variables estimation we follow a control function approach described in Wooldridge (2014). We proceed in two steps estimating the following two equations:

$$\begin{aligned} log(\text{travel time})_{FiGjhkt} \; &= \lambda_2 \, log(\text{instrumental travel time}_{FiGjhkt}) \\ &\quad + FE_{FiGjhk} + FE_{Fiht} + FE_{Gjkt} + \, u_{FiGjhkt} \end{aligned} \tag{4}$$

$$
\begin{aligned}
citations_{FiGjhkt} =& \exp\left[\beta\, log(\text{travel time}_{ijt}) \;+\; \lambda\, \hat{u}_{FiGjhkt}\right. \\
&\left. + FE_{FiGjhk} + FE_{Fiht} + FE_{Gjkt}\right] \;\times\; v_{FiGjhkt}
\end{aligned} \tag{5}
$$

In a first step we estimate equation (4) and obtain estimated residuals $\hat{u}_{FiGjhkt}$. In a second step we use the estimated residuals as a regressor in equation (5) which *controls* for the endogenous component of travel time. To perform inference we bootstrap standard errors.[64]

Columns (3) and (4) of Table 1 show the results of the instrumental variables estimation. If airlines were allocating jet airplanes to routes that would have witnessed a higher degree of exchange of knowledge even in the absence of jets, then we would expect the instrumental variables estimate to be smaller in absolute terms relative to the baseline coefficient. On the other hand, if the regulator targeted the opening of new routes between places that were in a lower trend of exchange of knowledge, we would expect the instrumented coefficient to be larger in absolute terms. Column (3) estimates the elasticity to be -0.152, bigger in absolute value compared to the non-instrumented estimate. The instrumental variables corrects for a downward bias in absolute terms, which represents evidence in favor of the regulator targeting the opening of new routes between places that had a lower degree of exchange of knowledge.[65,66]

In column (4) of Table 1 we see the coefficients of the instrumental variable estimation by distance between the citing and cited establishments. We observe the presence of a bias in the same direction as in column (3), however the magnitude of the bias is smaller

[64] Appendix D includes details on the bootstrap procedure.

[65] The incidental parameter problem is potential present also in the instrumental variables estimation (IV PPML). However, we are not aware of any bias-correction procedure for IV-PPML. Hence, columns (3) and (4) in Table 1 are not bias-corrected. In column (2) of Table 2 we present the PPML estimation not bias-corrected.

[66] The literature on weak instruments for non-linear instrumental variables is scarce. The rule of thumb of Staiger and Stock (1997) based on the F statistic is constructed using the bias that a *weak instrument* generates in a linear second stage (see Staiger and Stock (1997), Stock and Yogo (2005) and Sanderson and Windmeijer (2016) for testing for weak instruments in linear IV regression). For informative purposes, in the first stage of the model estimated in column (3) in Table 1 we obtain $\hat{\lambda}_2 = 0.95$ with a standard error 0.039 (clustered at the non-directional location pair level, *ij* is the same location pair as *ji*), and a within R2 of 0.38 (the share of residual variation explained by the instrument, after projecting out fixed effects).

except for the distance bin 0-300km, which is not precisely estimated. In particular, at more than 2,000km, the coefficient is relatively similar to the baseline estimation. Appendix E Tables 18 and 19 present the regression results including coefficients on the residual *controls*. According to Wooldridge (2014), there would be evidence of endogeneity if the parameter λ on controls in equation (5) is estimated to be statistically different from zero. While the control is statistically significant when using only one coefficient for all distances, none of them is statistically significant when opening the coefficient by distance range. In other words, we do not find evidence of endogeneity at long distances, especially at +2,000km.

The instrument used in the instrumental variables estimation is constructed using the 1951 flight network. We may be concerned that the 1951 flight network is correlated with future changes of citations.[67] In order to address this concern in Appendix E we estimate equation (3) by restricting the sample to establishments in MSA-pairs that are always indirectly connected. Results go in the same direction.

6.3. Diffusion of knowledge: Robustness

We may be concerned that there are other variables that could drive the diffusion of knowledge and at the same time be correlated with the change in travel time. In order to bias the coefficients, such omitted variables should be time-changing at the origin-destination MSA pair and be systematically correlated with the change in MSA-pair air travel time.[68] We consider three potential variables that could bias our estimates: improvements in highways, improvements in telephone communication and changes in flight ticket prices. In Table 2 we show the results controlling for this variables separately, while in Appendix E we include them simultaneously. Estimates are robust to including these controls.

[67] We include a establishment pair fixed effect in the regressions, so a potential correlation between the 1951 flight network and the level of citations between research establishments does not affect our estimation.

[68] Variables that are not time changing or that are time changing at the MSA or establishment level do not represent a threat to identification, as they are flexibly controlled for with the fixed effects.

Columns (1) and (2) in Table 2 present the elasticity of citations to travel time by distance bin. In column (1) the elasticity is bias-corrected while in column (2) it is not. We observe that not doing the bias correction does not qualitatively affect the results. Columns (3) to (6) include the additional controls and should be compared to column (2).[69]

First, in 1947 the Congress published the official plan for the Interstate Highway System, a nationwide infrastructure plan to improve existing highways and build new ones (see Baum-Snow (2007), Michaels (2008), Jaworski and Kitchens (2019) and Herzog (2021)). In case the change in travel time by air is correlated with the change in travel time by highway, we would have an omitted variable bias if we include only one of them in the estimation. Taylor Jaworski and Carl Kitchens have graciously shared with us data on county-to-county highway travel time and travel costs for 1950, 1960 and 1970, which we converted to MSA-to-MSA and linearly interpolated to convert to the same years of our air travel data. Hence we have a MSA-to-MSA time-varying measure of highway travel time which we include as control.[70]

Second, other means of communication like telephone lines may have expanded or changed their price during the period of analysis. Haines et al. (2010) contains information on the share of households within each city with telephone lines in 1960. We aggregate the variable to the MSA level. For each MSA-pair, we take the log of the mean share of households with telephone lines.[71] To include the variable as control we interact it with a time dummy to make the measure time variant. The assumption

[69]The jackknife bias-correction due to the incidental parameter problem is computationally intensive. Due to the computational burden and given that the bias correction does not substantially change the results in the baseline analysis, we have not bias-corrected estimations of robustness analysis in columns (2) to (6) of Table 2

[70]In Appendix E we show the correlation of MSA-to-MSA change in air travel time and highway travel time.

[71]Data from the 1962 City Data Book which comes from the US Bureau of the Census. log(mean telephone share$_{ij}$ = log((telephone share$_i$+telephone share$_j$)/2). Using as control the multiplied share = telephone share$_i$ $\times$ telephone share$_j$ gives similar results.

behind the interaction is that, if telephone lines expanded or changed their price over the time period, this time-change specific to each year was proportional to the 1960 log mean share of the MSA-pair.

Third, during the period of analysis ticket prices were set by the Civil Aeronautics Board, so airlines could not set prices of their own tickets. Some airlines included a sample of prices in the last page of their booklet of flight schedules, which we digitized. In Appendix E we document multiple facts about prices. The relevant fact for this section is that during 1962-1963 we observe a drop in prices of around 20% for routes of more than 1,000km distance. We may be concerned that the change in flow of knowledge is actually consequence of the change in prices, which happens to be correlated with the change in travel time. Given that we do not have ticket prices for each route and year, we use an estimated route price which is time varying. We obtain estimated prices by using the sample of prices that we digitized and fitting, for each year, price on a third degree polynomial of distance between origin and destination. We use log of estimated prices as control.[72]

Column (3) to (5) of Table 2 include the described controls. All of the coefficients are in the ball park of the baseline coefficients in both columns (1) and (2).[73]

Fourth, we control for a time varying effect of distance on citations. We may believe that other variables may have an effect on the diffusion of knowledge, and those variables are related to the distance between the citing and cited establishments. In column (6) we include as control $log(distance)$ interacted with a time dummy. We observe that the coefficients reduce in magnitude, potentially due to the fact that the change in travel time is also correlated with distance, hence controlling for a time-varying effect of distance absorbs part of the effect. In spite of that, the coefficient for distance of more

[72] Standard errors presented are not adjusted by the fact that the regression includes a predicted regressor as control variable.

[73] Assuming the covariance across coefficients of different regressions is zero, none of the coefficients is statistically different from the baseline coefficients either in column (1) or (2).

than 2,000km remains statistically significant at the 5% level. This result highlights the importance of the origin-destination time varying travel time data when studying the impact of face to face interactions, pointing that travel time and distance are not equivalent measures. This result differentiates our analysis from the one of Feyrer (2019) who uses two types of time-invariant distance (sea distance and geographical distance) interacted with time dummies to study changes in international trade.

In Appendix E we present additional robustness analysis. We may be concerned that the change in diffusion of knowledge is only consequence of the change in the geographic location of innovation. Hence, we re-estimate equation (3) with different samples: first, using only citing establishments that were present in 1949-1953, and second using only citing and cited establishments that were present in 1949-1953. We find that across sub-samples the coefficient at more than 2,000km remains stable across samples and statistically significant at the 1% level. Next, we estimate a variation of equation (3) in the form of log-log and obtain results that are in the ballpark of the baseline estimation.[74]

7. Creation of knowledge

In this section we interpret the results on increased diffusion of knowledge through the lens of a model of knowledge spillovers. We show that the reduction in travel time to innovative locations led to an increase in knowledge creation. The effect on the creation of knowledge was stronger in initially less innovative locations, leading to convesrgence across locations in terms of innovation. Additionally, the reduction in travel time contributed to a change in the geographic distribution of knowledge creation, increasing the relative importance of locations in the South and the West of the United States.

[74]See Appendix E Table 20. Elasticity at +2.000km is estimated to be -0.161 by OLS.

	PPML bias-corrected	PPML not bias-corrected				
Dep. variable: *citations*	$citations_{FiGjhkt}$					
	(1)	(2)	(3)	(4)	(5)	(6)
log(travel time) × 0-300km	0.019	0.021	0.023	0.0198	0.025	0.032
	(0.036)	(0.039)	(0.039)	(0.039)	(0.038)	(0.040)
log(travel time) × 300-1,000km	-0.089^{***}	-0.099^{***}	-0.096^{***}	-0.094^{***}	-0.102^{***}	-0.075^{**}
	(0.023)	(0.027)	(0.028)	(0.027)	(0.027)	(0.030)
log(travel time) × 1,000-2,000km	-0.094^{***}	-0.093^{**}	-0.089^{**}	-0.071^{*}	-0.104^{**}	-0.040
	(0.033)	(0.042)	(0.044)	(0.042)	(0.042)	(0.052)
log(travel time) × +2,000km	-0.169^{***}	-0.185^{***}	-0.180^{***}	-0.172^{***}	-0.196^{***}	-0.124^{**}
	(0.039)	(0.049)	(0.050)	(0.050)	(0.049)	(0.059)
N obs. effective	4,703,010	4,703,010	4,703,010	4,703,010	4,703,010	4,703,010
R2	0.88	0.88	0.88	0.88	0.88	0.88
Controls:						
log(highway time)	-	-	Yes	-	-	-
log(telephone share) × time	-	-	-	Yes	-	-
log(price)	-	-	-	-	Yes	-
log(distance) × time	-	-	-	-	-	Yes

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$

Table 2: Robustness: Elasticity of citations to travel time

Column (1) shows the result of Poisson Pseudo Maximum Likelihood (PPML) estimation of $citations_{FiGjhkt} = \exp\left[\sum_d \boldsymbol{\beta}_d \, \mathbb{1}\{distance_{ij} \in d\} \, log(\text{travel time}_{ijt}) + \sum_d \alpha_d \, \mathbb{1}\{distance_{ij} \in d\} \, \mathbb{1}\{X_{FiGjhkt}\} \, log(\text{travel time}_{ijt}) + FE_{FiGjhk} + FE_{Fiht} + FE_{Gjkt}\right] \times \varepsilon_{FiGjhkt}$, for citations of patents filed by establishment of firm F in location i, technology h and time period t, to patents filed by establishment of firm G in location j and technology k. travel time$_{ijt}$ is the travel time in minutes between location i and j at time period t, and it is set to 1 when $i = j$. d are distance intervals: $[0 - 300km]$, $(300km - 1000km]$, $(1000km - 2000km]$, $(2000km - max]$. Column (1) presents jackknife bias-corrected coefficients and bias-corrected bootstrap standard errors. Column (2) repeats column (1) without bias-correction. Relative to (2), columns (3) through (6) contain additional controls. Column (3) controls for log highway time between i and j at period t. Column (4) controls for the log of the mean share of households with telephone line in 1960 in ij pair interacted with a time dummy. Column (5) controls for log flight ticket price between i and j at period t. Column (6) controls for log distance ij interacted with a time dummy. When $FiGjhk$ has positive citations in at least one period and no citations in another, we attribute zero citations in the missing period. Columns (2) through (6) present standard errors clustered at the non-directional location in parentheses (ij is the same non-directional location pair as ji). R2 is computed as the squared correlation between observed and fitted values.

We construct a measure of *Knowledge Access* by adapting equation (2) to an empirical set up with multiple technology categories and time periods. The measure of *Knowledge Access* (KA_{iht}) shows how *easy* it is in time period t for research establishments in location i and technology h to access knowledge created in other locations. We compute *Knowledge Access* as follows:

$$KA_{iht} = \sum_{k} \omega_{hk} \sum_{j, j \neq i} \text{Patent stock}_{jk,t=1953} \times \text{travel time}_{ijt}^{\beta} \tag{6}$$

where, from right to left, travel time$_{ijt}^{\beta}$ is the travel time between locations i and j at time period t, to the power of the elasticity of diffusion of knowledge to travel time. Patent stock$_{jk,t=1953}$ is the discounted sum of patents produced in location j and technology k between 1941 and 1953.[75,76] ω_{hk} is the share of citations of technology h that go to technology k at the aggregate level in 1949-1953, similar to an input-output weight.[77] Then, KA_{iht} is a weighted sum of the patent stock in each other location and technology, where the weights are how easy it is to access that patent stock (travel time$_{ijt}^{\beta}$) multiplied by how relevant that knowledge is (ω_{hk}).

In order to reduce concerns of potential endogeneity of accessing knowledge and creating knowledge, we exclude the patent stock in the location itself from the sum (we only use $j \neq i$).[78]

The measure of *Knowledge Access* contains across-time variation within a location-

[75] Patent stock$_{jk,t=1953} = \sum_{y \in [1941,1953]}$ Patents$_{jky} \times (1 - \text{depreciation rate})^{1953-y}$. We use a depreciation rate of 5%, which is in the range of average depreciation rates of R&D found by De Rassenfosse and Jaffe (2017).

[76] Location j and technology k would be the *source* location and technology, while i and h would be the *destination* location and technology.

[77] $\omega_{hk} = citations_{hk,t=[1949,1953]} / citations_{h,t=[1949,1953]}$ is included to weight each *source* technology category k by how important it is for the *destination* technology category h.

[78] The theory makes no distinction on whether the knowledge stock is in i or j, so in principle we would like to include the patent stock of i in the knowledge access of i. However, this could lead to econometric problems. First, we do not have exogenous variation of travel time within i. Second, if knowledge creation in i is a persistent process, by including the patent stock of i we would introduce a mechanical relationship between knowledge access and knowledge creation. Hence, our baseline measure of knowledge access of i does not consider the patent stock of i. This is similar to what Donaldson and Hornbeck (2016) in the case of the empirical approximation of their Market Access measure. In Appendix E we show that the inclusion of i's patent stock does not affect the results.

technology ih, and cross-sectional variation across technologies h within a location i. The across-time variation is only due to the change in travel time between locations, every other component of the measure is fixed to its 1949-1953 level. The cross-sectional variation comes from a distribution of Patent stock$_{jk,t=1953}$ within source technologies k that is not uniform across source locations j, and from the input-output weights ω_{hk}. The joint across-time and cross-sectional variation means that if travel time for ij reduces, there will be a differential change in *Knowledge Access* across technologies h within location i which depends on the initial patent stock and input-output weights.

The degree with which changes in travel time are reflected in access to knowledge depend on how *important* travel time is to get knowledge to diffuse, which is the elasticity of knowledge diffusion to travel time that we estimated in Section 6. As the baseline we use $\beta = 0.185$, which is the elasticity of citations to travel time at more than 2,000 km not bias corrected. In robustness we use distance-specific β and in Appendix E we do sensitivity analysis of the results to changing the value of β.

The measure of *Knowledge Access* allows us to translate changes in travel time between pairs of MSAs into a single location-technology specific characteristic, and to represent it on the same scale as patent growth in Figure 9. Figure 15 depicts the time change in log *Knowledge Access* from 1951 to 1966, averaged across technologies within each MSA. Darker colors represent higher growth in *Knowledge Access*. As with patent growth, we observe that MSAs that had the strongest growth are generally located in the South and the West of the United States, far from the knowledge centers of New York and Chicago. The reduction in travel time was larger between locations far apart, implying that locations which happened to be far from knowledge centers increased relatively more their *Knowledge Access*.

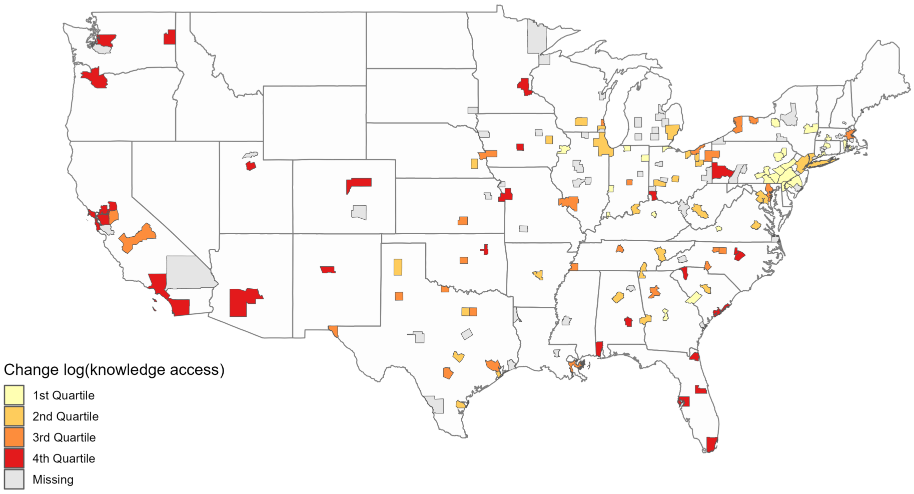


Figure 15: Change in log Knowledge Access 1951 - 1966

7.1. Creation of knowledge: Baseline estimation

With the measure of *Knowledge Access* we then adapt equation (1) to estimate:

$$\text{Patents}_{Fiht} = \exp\left[\rho \, log(KA_{iht}) + FE_{Fih} + FE_{it} + FE_{ht}\right] \times \xi_{Fiht} \tag{7}$$

where Patents_{Fiht} are patents applied by establishment of firm F in location i and technology h at time period t. The measure of knowledge access KA_{iht} is at the iht location-technology-time level, meaning that all establishments within an iht share the same level of knowledge access. The parameter of interest ρ is the elasticity of (the creation of new) patents to knowledge access. In the presence of knowledge spillovers as suggested in Section 2, we would expect ρ to be positive and statistically significant.

The fixed effect FE_{Fih} absorbs time invariant characteristics at the firm-location-technology level, as for example the productivity of the establishment-technology. This fixed effect is more fine grained than just a location-technology, which would absorb the comparative advantage of a location in a certain technology. The fixed effect FE_{it} absorbs characteristics that are time variant at the location level. For example, changes in R&D subsidies that are location specific and common across all technologies would be absorbed by this fixed effect. Also, better flight connectivity could spur economic activity as shown in Campante and Yanagizawa-Drott (2017), leading to an increase in patenting activity in the location. If that increase is general across technologies within the location, then FE_{it} would absorb it. Finally, the fixed effect FE_{ht} absorbs characteristics that are time variant at the technology level. If technologies had different time-trends at the national level, then the fixed effect would control for these trends in a flexible way.

The inclusion of FE_{Fih} implies that only across-time variation within an establishment-technology is used to identify ρ. The inclusion of FE_{it} implies that only variation across-technologies within a location-time is exploited, so across-time variation is compared across establishments within a location, and not across locations. The inclusion of FE_{ht} implies that the identifying across-time variation is conditional on aggregate trends of the technology. In short, identification of ρ relies on across-time changes in the amount of patents and knowledge access of an establishment, relative to other establishments in the same location, conditional on aggregate technological trends.

Column (1) in Table 3 shows the result of estimating equation (7). The elasticity of patents to knowledge access is estimated to be 10.14, significant at the one percent level. The average change in knowledge access at the location-technology level[79] is 9%, implying that on average the change in travel time predicts a 3.5% average yearly growth rate of patents.[80] The observed average yearly growth rate of new patents at the location-technology is 4.4%.[81] Comparing the predicted and observed growth rates,

[79]Due to entry, we cannot compute the growth rate at the establishment-technology level for 70% of establishment-technology, given that they had 0 patents in the initial time period. In the case of location-technology, 5% did not have patents in the initial period. We the prefer to interpret coefficients using location-technology growth rates, which we compute using the remaining 95% of location-technologies that had positive patents in the initial time period.

[80]The elasticity of 10.14 predicts an increase of 91.3% over the time period of 19 years ($10.14 \times 0.09 = 0.913$), which translates into a 3.5% average yearly growth rate ($(1{+}0.913)^{1/19}{-}1 \approx 0.035$).

[81]From the first time period (1949-1953) to the last time period (1964-1968) we observe an average growth rate of new patents of 127%. We obtain $0.044 \approx ((1+1.27)^{1/19}-1$

	PPML		IV PPML		IV PPML centered	
Dependent Variable:			*Patents*			
	(1)	(2)	(3)	(4)	(5)	(6)
log(knowledge access)	10.14^{***} (3.66)	9.36^{**} (3.69)	11.24^{*} (6.35)	10.26 (6.38)	9.86^{*} (5.73)	7.01 (5.83)
log(knowledge access) × 3rd quartile		2.05^{***} (0.58)		2.32^{***} (0.66)		3.99^{**} (1.25)
log(knowledge access) × 2nd quartile		3.80^{***} (0.90)		4.21^{***} (0.84)		7.57^{***} (2.30)
log(knowledge access) × 1st quartile		5.00^{***} (1.30)		5.77^{***} (1.11)		9.03^{***} (2.46)
R2	0.85	0.85	0.85	0.85	0.85	0.85
N obs. effective	991,480	991,480	991,480	991,480	991,480	991,480

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$

Table 3: Effect of knowledge access on patents, by MSA innovativeness quartile

Column (1) shows the result of Poisson Pseudo Maximum Likelihood (PPML) estimation of $\text{Patents}_{Fiht} = \exp\left[\boldsymbol{\rho}\, log(KA_{iht}) + FE_{Fih} + FE_{it} + FE_{ht}\right] \times \xi_{Fiht}$, for patents filed by establishment of firm F in location i, technology h and time period t. KA_{iht} is knowledge access of establishments in location i technology h and time period t. Column (2) opens the coefficient ρ by the quartile of innovativeness of location i within technology h, computed using patents filed in 1949-1953. Higher quartile indicates higher initial level of innovativeness. The fourth quartile is used as reference category. Columns (3) and (4) show the result of two step instrumental variables estimation, where KA_{iht} is instrumented with $\widetilde{KA}_{iht}$, knowledge access computed using the counterfactual travel time that would have taken place if routes were fixed to the ones in 1951 and each year routes were operated at the average aggregate flying speed of the year. Columns (5) and (6) use a centered version of $\widetilde{KA}_{iht}$ following Borusyak and Hull (2023) by subtracting the expected instrument which is computed using random flight networks. Standard errors are presented in parentheses. Column (1) and (2) present clustered at the location-technology ih. Column (3) and (4) present bootstrap standard errors. R2 is computed as the squared correlation between observed and fitted values.

the improvement in air travel time has the power to account for 79.5% of the observed average yearly patent growth rate.[82]

We aggregate predicted changes in patent growth at the Census Region level. The change in travel time predicts a yearly growth rate 0.74 percentage points higher in the South and the West relative to the Midwest and Northeast. In the data we observe 2.1 percentage points difference in the growth rate, implying that the change in travel time can account for 35% of the observed differential growth rate.[83]

Section 5.1 showed that in the data, initially less innovative MSAs had a larger growth rate of patenting. In column (2) in Table 3 we investigate if the increase in knowledge access had an heterogeneous effect on the amount of new patents created depending on the initial innovativeness of the location i in technology h. We compute the quartile of innovativeness of location i in technology h in the time period 1949-1953 and interact it with $\log(KA_{iht})$.[84] We use as reference category the highest quartile of initial innovativeness, hence the coefficient on $\log(KA_{iht})$ without interaction is the elasticity for the highest quartile. Coefficients on other quartiles should be interpreted relative to the highest quartile.

We find that the coefficients on lower quartiles of initial innovativeness are positive and statistically different from the coefficient in the highest quartile. Thus, knowledge access had a greater effect on patenting for establishments that were located in initially

[82] $79.5 = 3.5/4.4 \times 100$

[83] Using the coefficient of column (1) in Table 3, we compute the MSA-technology predicted level of patents for 1966 and aggregate it at the Census region - technology level. Then, we compute yearly growth rates within each region-technology and take averages across technologies. Next, we take the average between S and W, and MW and NE, and finally compute the differential predicted growth. If we use the quartile-specific coefficients of column (2) in Table 3 we obtain a predicted differential growth rate of 0.86 percentage points, which implies that the change in travel time can account for 41% of the observed differential growth rate.

[84] We use the quartiles of innovativeness defined in section 5.1, computed using the amount of patents of location i in technology h filed in the time period 1949-1953. Each location i has (potentially) a different value quartile in each technology h. The 1st quartile refers to the 25% initially least innovative MSAs in technology h.

less innovative locations.[85] Given the difference in the coefficients, the increase in knowledge access predicts an average yearly growth of new patents of 4.5% for the initially lowest quartile of innovativeness, while it predicts 3.4% for the highest quartile.[86] The change in knowledge access predicts differential growth rate of 1.1 percentage points. In the data we observe that the average yearly growth rate of patents in the lowest quartile is 5.3 percentage points higher than in the highest quartile. Comparing the predicted and observed differential growth rates, the improvement in knowledge access as consequence of the reduction in travel time explains 21% of the difference in growth rates of new patents between locations in the lowest and highest quartile of innovativeness.[87]

In Appendix Table 23 we present results estimating equation 7 weighting patents by quality using the breakthroughness level computed by Kelly et al. (2021).[88] We find a larger coefficient in magnitude, providing evidence that the results are not driven by the granting of lower quality patents. Results also go in the same direction with the quality-weighted knowledge access measure.

7.2. Creation of knowledge: Instrumental variables estimation

As in Section 6, we may be concerned that decisions of the regulator or airlines which affect travel time are endogenous to the diffusion of knowledge and consequently to knowledge access. Therefore, we construct an instrument for knowledge access in

[85] A given percentage change in knowledge access led to a stronger increase in patenting in initially less innovative locations.

[86] The change in knowledge access for the lowest quartile is on average 9.1%, which multiplied by the coefficient 14.36 (obtained by doing 9.36+5.00=14.36) gives a predicted growth of 131% over 19 years. Translated into average yearly growth it is 4.5% = $[(1 + 1.31)^{(1/19)} - 1] \times 100$. For the highest quartile, knowledge access changed on average 9.5%, which multiplied by the coefficient 9.36 predicts 89% growth rate, which is 3.4% yearly growth rate.

[87] $21\% \approx 1.2/5.1 \times 100$

[88] We use the patent's 5-year percentile of breakthroughness after demeaning by year fixed effects computed by Kelly et al. (2021). The measure of breakthroughness is computed by comparing the patent text of the focal patent with previous and future patents in a 5-year window to find whether the patent introduces new concepts that were not common before but became common after, making a *breakthrough*. Using the measure computed with 10-year data give similar results. Importantly, the computation of the measure does not use citation data.

which instead of using observed travel time, we use the fictitious travel time presented in section 4.2 in which routes are fixed to the ones in 1951 and each route is operated with the average airplane of the year:

$$\widetilde{KA}_{iht} = \sum_{k} \omega_{hk} \sum_{j, j\neq i} \text{Patent stock}_{jk,t=1953} \times \left(\text{instrumental travel time}_{ijt}\right)^{\beta} \tag{8}$$

Recently, Borusyak and Hull (2023) have pointed out that when multiple sources of variation are combined to define treatment according to a known formula, treatment exposure can be non-random. Failing to account for this difference in expected treatment can create omitted variable bias. Our instrument combines cross-sectional variation, the 1951 flight network and the spatial distribution of the knowledge stock in the early 1950s, with variation across time, the national rollout of jets. A MSA like San Francisco that is connected to knowledge hubs like Chicago via non-stop, long-distance flights will benefit greatly from the increase in flight speeds brought about by the jet engine and see a large increase in knowledge access. On the other hand, a MSA like Boston, already close to major innovation hubs like New York, benefits less.

While the initial flight network thus matters greatly for the variation captured by the instrument, this variation is in part driven by geography. Considering all possible connections, MSAs located far away from every other location have more possible long-distance connections and are thus more prone to benefit from faster airplanes. It is this non-random exposure to the national rollout of jets that might create omitted variable bias if not accounted for. Following Borusyak and Hull (2023), we recenter our instrument by subtracting the expected value of the instrument. To construct the *expected instrument* we draw a set of random counterfactual networks, compute travel time and the value of a counterfactual instrumental knowledge access under each of the networks. Counterfactual networks contain the underlying observed geography and hence locations farther apart from innovation centers see a larger increase in knowledge access even in random networks. We then take the average across counterfactual networks to obtain the expected instrument.[89] By recentering the instrument we purge it from the non-randomness that might be introduced by geography. The recentered instrument is:

$$log(\widetilde{KA}_{iht})_{centered} = log(\widetilde{KA}_{iht}) - \mathbb{E}[log(\widetilde{KA}_{iht})] \tag{9}$$

We implement the instrumental variables estimation by control function as in Section 6. Results are presented in Table 3. Columns (3) and (4) show results using the non-centered instrument while columns (5) and (6) use the centered version. The coefficients do not show an important change and the convergence prediction obtained using non-instrumented PPML remains valid.[90,91]

Figure 16 shows in the left panel the patent growth observed in the data (it replicates Figure 9), while in the right panel it is the predicted patent growth. We compute the prediction using the observed change in travel time and quartile specific elasticities of column (2) in Table 3. Similarly to what is observed in the data, the change in travel time predicts a larger patenting growth rate in the South and the West relative to the Northeast and Midwest.

The result in column (2) implies that a given change in *Knowledge Access* had a stronger effect on patenting growth in less innovative locations. In other words, knowledge spillovers as an externality had a more predominant role in the production of knowledge in locations that initially produced relatively fewer patents. Theoretically, this result implies that the parameter ρ in equation (1) varies depending on the level of previous production of knowledge of location i. Empirically the implication is that a given increase in knowledge spillovers leads to innovation convergence across lo-

[89] Details on the construction of the centered instrument are presented in Appendix Section E.2.2.

[90] The first stage of the model estimated in column (3) of Table 3 gives a $\hat{\lambda}_2 = 1.01$ with standard error 0.03 (clustered at the location-technology level *ih*), and a within R2 of 0.53.

[91] Using non-centered IV estimates, the predicted yearly patent growth rate in the lowest quartile is 4.9% while it is 3.7% in the highest quartile. The predicted differential growth rate is then 1.2 percentage points, meaning that the change in knowledge access can explain $(1.2/5.3) \times 100 \approx 23\%$ of the observed differential growth rate.

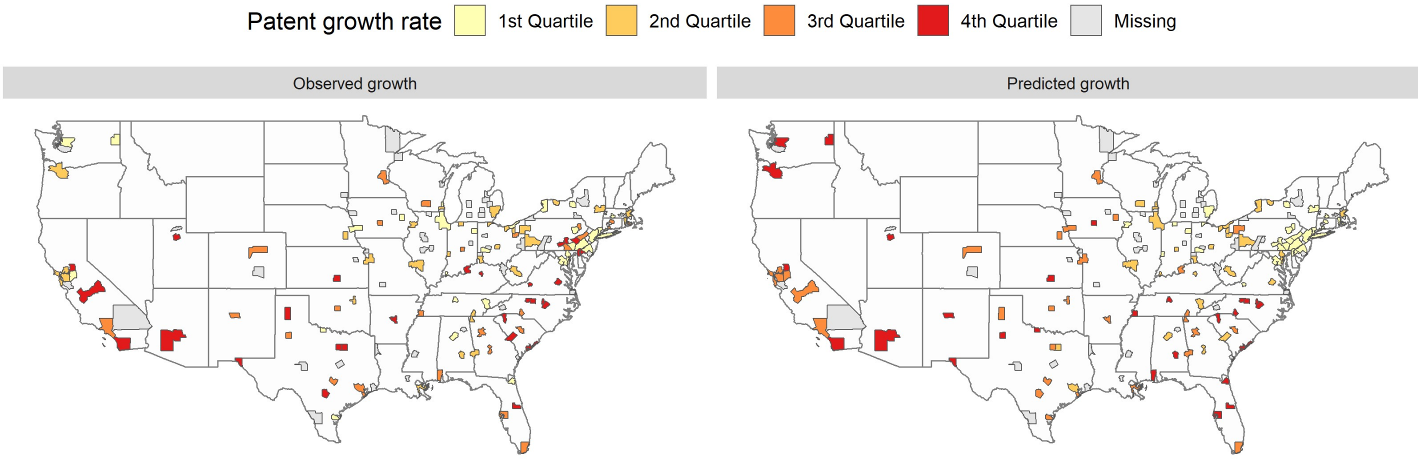


Figure 16: Observed vs. predicted patent growth 1951 - 1966

cations. As seen in section 5.1, during 1949-1968 we observe innovation-convergence across locations and that is exactly what the estimated coefficients predict following a reduction in travel time.

In order to understand the convergence result and compare it with other findings in the literature it is important to remember that commercial airplanes during 1950s and 1960s were a means of transportation mainly for people. On the other hand, other transportation improvements as those in water transport, railroads or highways also contain another ingredient: they were used to carry goods. Hence, other means of transportation have a simultaneous impact on face to face interactions and trade. Pascali (2017) finds that the introduction of the steam engine vessels in the second half of the 19th century had an impact on international trade that led to economic divergence between countries. Faber (2014) finds that the expansion of the highway system in China led to a reduction of GDP growth in peripheral counties, with evidence suggesting a trade channel due to reduction in trade costs. In our setup, the introduction of jet airplanes represented a big shock to the mobility of people while not affecting significantly the transport of merchandise. Therefore, studying the introduction of jet airplanes allows us to focus on improved face to face interactions, while the trade channel would be a second order effect.

7.3. Creation of knowledge: Robustness

In this section we show that the effect of *Knowledge Access* on the creation of new patents and the convergence effect remains after including different controls. Table 4 shows the results.

Jaworski and Kitchens (2019) show that improvements in the Interstate Highway System led to local increases in income through an increased market access. In our set up, if the effect of market access affects innovation in the same way across technologies, then it would be absorbed by the MSA-time fixed effect FE_{it} in equation (7). However, if the effect of market access on innovation varies across technologies, then it would be a confounder. To control for this potential confounder, we compute market access by highway and interact it with a technology dummy. We compute market access as:

$$\text{Market Access}_{it} = \sum_j \text{Population}_{j,t=1950} \times \tau_{ijt}^{\theta} \tag{10}$$

where $\text{Population}_{j,t=1950}$ is population in MSA j in 1950, τ_{ijt} are the shipping costs provided in the data of Taylor Jaworski and Carl Kitchens computed using each year's highway driving distance, highway travel time, petrol cost and truck driver's wage. θ is the elasticity of trade to trade costs which we set to -8.28, the preferred value of Eaton and Kortum (2002) and in the range of many other estimates in the literature (Head and Mayer (2014), Caliendo and Parro (2015), Donaldson and Hornbeck (2016)). Columns (3) and (4) of Table 4 show the results, we do not observe an important difference with the baseline estimates.

Campante and Yanagizawa-Drott (2017) shows that better connectivity by airplane leads to an increase in economic activity as measured by satellite-measured night light. Söderlund (2020) shows that an increase in business travel in the late 1980s and early 1990s led to an increase in trade between countries. In a similar way to knowledge access, we could think that better connectivity by airplane could have led to an increase in market access due to a reduction in information frictions, with goods being shipped by land. Similarly to highway market access, if the effect of market access by airplane is common to all technology categories then it would be absorbed by the MSA-time fixed effect FE_{it}. In order to account for a technology-specific effect, we construct a measure of airplane market access and interact it with a technology dummy. The measure of airplane market access is similar to equation (10) where τ is the travel time by airplane and θ is set to -1,22, the elasticity of trade to travel time from Söderlund (2020). The results are shown in columns (5) and (6) of Table 4. While the coefficients in all quartiles are reduced, the estimated value of ρ is positive and significant and the result on convergence remains.

Potential contemporaneous improvements in other means of communication, like telephones, could have spurred the creation of new patents. In columns (7) and (8) we include the log of the MSA's share of households with telephones in 1960 and double-interact it with a technology dummy and a time dummy. The results remain invariant with respect to the baseline.

Another potential explanation for the increase of patenting could be that better connectivity decreased technology-specific financial frictions. The potential reduction in financial frictions, rather than a confounder, would be a mechanism through which airplanes increased innovation. However, according to Jayaratne and Strahan (1996) during 1950s and 1960s interstate lending and bank branching were limited. Prior to the 1970s, banks and holdings were restricted in their geographic expansion within and across state borders. Additionally, the Douglas Amendment to the Bank Holding Company Act prevented holding companies from acquiring banks in other states. Therefore, it is unlikely that interstate bank financing would be a driving force. Nonetheless, if other sector-specific modes of financing like venture capital were active, they could be driving the results. In Appendix E we construct multiple measures of access to capital by using market capitalization of patenting firms listed in the stock market. The results present suggestive evidence that access to capital is not driving the results.

	PPML							
Dependent Variable:	*Patents*							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(knowledge access)	10.14*** (3.66)	9.36** (3.69)	9.28** (3.68)	8.23** (3.69)	6.22* (3.58)	5.84 (3.60)	10.34*** (3.44)	9.25*** (3.43)
log(knowledge access) × 3rd quartile		2.05*** (0.58)		2.16*** (0.57)		2.06*** (0.59)		2.23*** (0.57)
log(knowledge access) × 2nd quartile		3.80*** (0.90)		3.89*** (0.89)		3.75*** (0.88)		3.93*** (0.91)
log(knowledge access) × 1st quartile		5.00*** (1.30)		5.13*** (1.30)		5.08*** (1.29)		5.18*** (1.32)
N obs. effective	991,480	991,480	991,480	991,480	991,480	991,480	991,480	991,480
R2	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85
Controls:								
log(Highway market access) × technology	-	-	Yes	Yes	-	-	-	-
log(Airplane market access) × technology	-	-	-	-	Yes	Yes	-	-
log(Telephone share) × technology × time	-	-	-	-	-	-	Yes	Yes

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$

Table 4: Elasticity of new patents to knowledge access, by MSA innovativeness quartile

Column (1) shows the result of Poisson Pseudo Maximum Likelihood (PPML) estimation of $\text{Patents}_{Fiht} = \exp\left[\boldsymbol{\rho}\, log(KA_{iht}) + FE_{Fih} + FE_{it} + FE_{ht}\right] \times \xi_{Fiht}$, for patents filed by establishment of firm F in location i, technology h and time period t. KA_{iht} is knowledge access of establishments in location i technology h and time period t. Column (2) opens the coefficient ρ by the quartile of innovativeness of location i within technology h, computed using patents in 1949-1953. Higher quartile indicates higher initial level of innovativeness. The fourth quartile is used as reference category. Relative to columns (1) and (2), columns (3) and (4) control for technology specific effect of log(highway market access), columns (5) and (6) control for technology specific effect of log(airplane market access), columns (7) and (8) control for technology and time specific effect of log(telephone share). Standard errors clustered at the location-technology ih are presented in parentheses. R2 is computed as the squared correlation between observed and fitted values.

Finally, in Appendix E we include additional robustness checks. We compute different versions of *Knowledge Access*: we use distance-specific β from section 6, we consider the patent stock only of locations j far from i, we do sensitivity analysis using different values of β. Also, we re estimate the effects by quartile of initial innovativeness using patents per capita. Last, we re-do the baseline regression using OLS estimation. Results go in the same direction: an increase in knowledge access leads to an increase in patenting and the effect is stronger in initially less innovative locations.

8. Conclusion

This paper studies how frictions to the mobility of people affect the geography of innovation in the context of the early Jet Age in the United States. With newly digitized data on airlines' flight schedules, we construct a dataset of the flight network in the United States during the 1950s and 1960s. We document the large reduction in air travel time that jet airplanes brought about: around 5 to 6 hours, a 50% reduction, for coast-to-coast travel. Combined with patent data, we find that the reduction in travel time increased knowledge diffusion, especially between research establishments located far apart. The increase in access to knowledge created long-distance spillovers and led to the production of new knowledge.

Our results point to jet airplanes as an important driver behind major changes in the geography of innovation in the United States post World War II: a catching up of the South and the West with the Northeast and the Midwest, and initially less innovative MSAs reducing the gap with more innovative ones.

The results provide policy-relevant insights regarding the impact of passenger transport infrastructure on the emergence of technology clusters. Large R&D policies, like the recent CHIPS and Science Act, frequently include a place-based component to increase technology capacity in regions that lag behind (Gruber and Johnson (2019), Gross and Sampat (2023)). Our results show that connectivity to existing clusters can lead to an increase of local innovation and act as a convergence force between regions.

References

Acemoglu, D., U. Akcigit, and W. R. Kerr (2016). Innovation network. *Proceedings of the National Academy of Sciences 113*(41), 11483–11488.

Agrawal, A., A. Galasso, and A. Oettl (2017). Roads and innovation. *The Review of Economics and Statistics 99*(3), 417–434.

Andersson, D., T. Berger, and E. Prawitz (2017). On the right track: Railroads, mobility and innovation during two centuries'. In *On the Move: Essays on the Economic and Political Development of Sweden*, pp. 203–288. Institute for International Economic Studies, Stockholm University.

Andrews, M. J. and A. Whalley (2021). 150 years of the geography of innovation. *Regional Science and Urban Economics*, 103627.

Arzaghi, M. and J. V. Henderson (2008). Networking off madison avenue. *The Review of Economic Studies 75*(4), 1011–1038.

Audretsch, D. B. and M. P. Feldman (2004). Knowledge spillovers and the geography of innovation. In *Handbook of regional and urban economics*, Volume 4, pp. 2713–2739. Elsevier.

Bai, J. J., W. Jin, and S. Zhou (2021). Proximity and knowledge spillovers: Evidence from the introduction of new airline routes.

Balsmeier, B., M. Assaf, T. Chesebro, G. Fierro, K. Johnson, S. Johnson, G.-C. Li, S. Lück, D. O'Reagan, B. Yeh, G. Zang, and L. Fleming (2018). Machine learning and natural language processing on the patent corpus: Data, tools, and new measures. *Journal of Economics & Management Strategy 27*(3), 535–553.

Baum-Snow, N. (2007). Did highways cause suburbanization? *The quarterly journal of economics 122*(2), 775–805.

Bergé, L. (2018). Efficient estimation of maximum likelihood models with multiple fixed-effects: the R package FENmlm. *CREA Discussion Papers* (13).

Bloom, N., M. Schankerman, and J. Van Reenen (2013). Identifying technology spillovers and product market rivalry. *Econometrica 81*(4), 1347–1393.

Borenstein, S. and N. L. Rose (2014, June). *How Airline Markets Work... or Do They? Regulatory Reform in the Airline Industry*, pp. 63–135. University of Chicago Press.

Borusyak, K. and P. Hull (2023). Nonrandom exposure to exogenous shocks. *Econometrica 91*(6), 2155–2185.

C.A.B. (1951, 1956, 1961, 1966). *Air carrier traffic statistics.* Civil Aeronautics Board.

Caliendo, L. and F. Parro (2015). Estimates of the trade and welfare effects of nafta. *The Review of Economic Studies 82*(1), 1–44.

Campante, F. and D. Yanagizawa-Drott (2017, 12). Long-Range Growth: Economic Development in the Global Network of Air Links*. *The Quarterly Journal of Economics 133*(3), 1395–1458.

Carlino, G. and W. R. Kerr (2015). Agglomeration and innovation. *Handbook of regional and urban economics 5*, 349–404.

Catalini, C., C. Fons-Rosen, and P. Gaulé (2020). How do travel costs shape collaboration? *Management Science 66*(8), 3340–3360.

Caves, R. E. (1962). *Air transport and its regulators: an industry study*. Harvard economic studies v. 120. Cambridge: Harvard University Press.

Chatterji, A., E. Glaeser, and W. Kerr (2014). Clusters of entrepreneurship and innovation. *Innovation policy and the economy 14*(1), 129–166.

Coscia, M., F. Neffke, and R. Hausmann (2020). Knowledge diffusion in the network of international business travel. *Nature Human Behaviour 4*(10).

De Rassenfosse, G. and A. B. Jaffe (2017). Econometric evidence on the r&d depreciation rate. Technical report, National Bureau of Economic Research.

Dhaene, G. and K. Jochmans (2015). Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies 82*(3), 991–1030.

Dijkstra, E. W. et al. (1959). A note on two problems in connexion with graphs. *Numerische mathematik 1*(1), 269–271.

Donaldson, D. and R. Hornbeck (2016). Railroads and american economic growth: A “market access” approach. *The Quarterly Journal of Economics 131*(2), 799–858.

Duranton, G., P. Martin, T. Mayer, and F. Mayneris (2009). The economics of clusters: evidence from france.

Eaton, J. and S. Kortum (2002). Technology, geography, and trade. *Econometrica 70*(5), 1741–1779.

Faber, B. (2014). Trade integration, market size, and industrialization: evidence from china’s national trunk highway system. *Review of Economic Studies 81*(3), 1046–1070.

Feyrer, J. (2019). Trade and income—exploiting time series in geography. *American Economic Journal: Applied Economics 11*(4), 1–35.

Fogel, R. W. (1963). *Railroads and American economic growth: essays in econometric history*. Ph. D. thesis, Johns Hopkins University.

Furman, J. L. and S. Stern (2011, August). Climbing atop the shoulders of giants: The impact of institutions on cumulative research. *American Economic Review 101*(5), 1933–63.

Glaeser, E. (2011). *Triumph of the City*. Macmillan.

Gross, D. P. (2019). The consequences of invention secrecy: Evidence from the uspto patent secrecy program in world war ii. Technical report, National Bureau of Economic Research.

Gross, D. P. and B. N. Sampat (2023). America, jump-started: World war ii r&d and the takeoff of the us innovation system. *American Economic Review 113*(12), 3323–3356.

Gruber, J. and S. Johnson (2019). Jump-starting america: How breakthrough science can revive economic growth and the american dream.

Haines, M. R. et al. (2010). Historical, demographic, economic, and social data: the united states, 1790–2002. *Ann Arbor, MI: Inter-university Consortium for Political and Social Research*.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001, October). The nber patent citation data file: Lessons, insights and methodological tools. Working Paper 8498, National Bureau of Economic Research.

Hansen, B. E. (2021). *Econometrics*.

Hausman, N. (2022). University innovation and local economic growth. *Review of Economics and Statistics 104*(4), 718–735.

Head, K. and T. Mayer (2014). Chapter 3 - gravity equations: Workhorse,toolkit, and cookbook. In G. Gopinath, E. Helpman, and K. Rogoff (Eds.), *Handbook of International Economics*, Volume 4 of *Handbook of International Economics*, pp. 131–195. Elsevier.

Herzog, I. (2021). National transportation networks, market access, and regional economic growth. *Journal of Urban Economics 122*, 103316.

Hovhannisyan, N. and W. Keller (2015, March). International business travel: an engine of innovation? *Journal of Economic Growth 20*(1), 75–104.

Jaffe, A. B., M. Trajtenberg, and M. S. Fogarty (2000, May). Knowledge spillovers and patent citations: Evidence from a survey of inventors. *American Economic Review 90*(2), 215–218.

Jaffe, A. B., M. Trajtenberg, and R. Henderson (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics 108*(3), 577–598.

Jaworski, T. and C. T. Kitchens (2019). National policy for regional development: Historical evidence from appalachian highways. *Review of Economics and Statistics 101*(5), 777–790.

Jayaratne, J. and P. E. Strahan (1996). The finance-growth nexus: Evidence from bank branch deregulation. *The Quarterly Journal of Economics 111*(3), 639–670.

Kantor, S. and A. T. Whalley (2023). Moonshot: Public r&d and growth. Technical report, National Bureau of Economic Research.

Kelly, B., D. Papanikolaou, A. Seru, and M. Taddy (2021). Measuring technological innovation over the long run. *American Economic Review: Insights 3*(3), 303–320.

Kerr, W. R. and F. Robert-Nicoud (2020). Tech clusters. *Journal of Economic Perspectives 34*(3), 50–76.

Kogan, L., D. Papanikolaou, A. Seru, and N. Stoffman (2017, 03). Technological Innovation, Resource Allocation, and Growth*. *The Quarterly Journal of Economics 132*(2), 665–712.

Krugman, P. R. (1991). *Geography and trade*. MIT press.

Magerman, T., B. Van Looy, and X. Song (2006). Data production methods for harmonized patent statistics : patentee name harmonization.

Manson, S., J. Schroeder, D. Van Riper, T. Kugler, and S. Ruggles (2020). Ipums national historical geographic information system: Version 15.0.

Marshall, A. (1890). *Principles of economics*.

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

Moretti, E. (2010). Local multipliers. *American Economic Review 100*(2), 373–377.

Moretti, E. (2021, October). The effect of high-tech clusters on the productivity of top inventors. *American Economic Review 111*(10), 3328–75.

Moretti, E. and P. Thulin (2013). Local multipliers and human capital in the united states and sweden. *Industrial and Corporate Change 22*(1), 339–362.

Pascali, L. (2017). The wind of change: Maritime technology, trade, and economic development. *American Economic Review 107*(9), 2821–54.

Perlman, E. R. (2016). Dense enough to be brilliant: patents, urbanization, and transportation in nineteenth century america. *Work. Pap., Boston Univ.*

Petralia, S. (2019). HistPat International Dataset.

Petralia, S., P.-A. Balland, and D. Rigby (2016). HistPat Dataset.

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

Sanderson, E. and F. Windmeijer (2016). A weak instrument f-test in linear iv models with multiple endogenous variables. *Journal of econometrics 190*(2), 212–221.

Schweiger, H., A. Stepanov, and P. Zacchia (2022). The long-run effects of r&d place-based policies: evidence from russian science cities. *American Economic Journal: Economic Policy 14*(3), 322–351.

Silva, J. M. C. S. and S. Tenreyro (2006). The log of gravity. *The Review of Economics and Statistics 88*(4), 641–658.

Söderlund, B. (2020). The importance of business travel for trade: Evidence from the liberalization of the soviet airspace. Working Paper Series 1355, Research Institute of Industrial Economics.

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*(3), 557–586.

Stock, J. H. and M. Yogo (2005). *Testing for Weak Instruments in Linear IV Regression*, pp. 80–108. Cambridge University Press.

Storper, M. and A. J. Venables (2004, 08). Buzz: face-to-face contact and the urban economy. *Journal of Economic Geography 4*(4), 351–370.

Tsiachtsiras, G. (2021). Transportation networks and the rise of the knowledge economy in 19th century france.

Weidner, M. and T. Zylkin (2021). Bias and consistency in three-way gravity models. *Journal of International Economics 132*, 103513.

Wooldridge, J. M. (2014). Quasi-maximum likelihood estimation and testing for nonlinear models with endogenous explanatory variables. *Journal of Econometrics 182*(1), 226–234.

A. Appendix: Travel Time Data

A.1. Data Construction

We construct a dataset of travel times by plane between US MSAs for the years 1951, 1956, 1961, 1966. We get information of direct flights from airline flight schedules and feed this information into an algorithm to allow for indirect flights. For each MSA pair with airports served by at least one of the airlines in our dataset we compute the fastest travel time in each of the four years.

Using images of flight schedules, we digitized the flight network for six major airlines: American Airlines (AA), Eastern Air Lines (EA), Trans World Airlines (TWA), United Airlines (UA), Braniff International Airways (BN) and Northwest Airlines (NW). Note that the first four in this list were often referred to as the *Big Four*, highlighting their dominant position in the market. They alone accounted for 74% of domestic trunk revenue passenger-miles from February 1955 to January 1956. Together the six airlines accounted for 82% of revenue passenger-miles in that same period, 77% from February 1960 to January 1961 and 78% from February 1965 to January 1966 (C.A.B., 1966). Our sample of airlines thus covers a vast share of the domestic market for air transport. In addition, the airlines were chosen to maximize geographic coverage.

In total we obtain a sample of 5,910 flights. These flights often have multiple stops. If we count each origin-destination pair of these flights separately, our sample contains 17,469 legs.

Table 5 lists the exact dates of when flight schedules we digitized became effective. Due to limited data availability not all flight schedules are drawn from the same part of the year. As seasonality of the network seems limited and given the large market share of the airlines we consider, our data is a good approximation of the network in a given year.

Table 5: Date of Digitized Flight Schedules

Airline	1951	1956	1961	1966
AA	September 30	April 29	April 30	April 24
EA	August 1	October 28	April 1	April 24
TWA	August 1	September 1	April 30	May 23
UA	April 29	July 1	June 1	April 24
BN	August	August 15	April 30	April 24
NW	April 29	April 29	May 28	March 1
PA	June 1	July 1	August 1	August 1

Figure 17 shows two pages of the flight schedule published by American Airlines in 1961. Each column corresponds to one flight. As can be seen, one flight often has multiple stops. Departure and arrival times in most flight schedules are indicated using the 12-hour system. PM times can be distinguished from AM times by their bold print. In the process of digitization we converted the flight schedules to the 24-hour system. Times in most tables are in local time. We thus recorded the time zones that are indicated next to the city name and converted them to Eastern Standard Time.

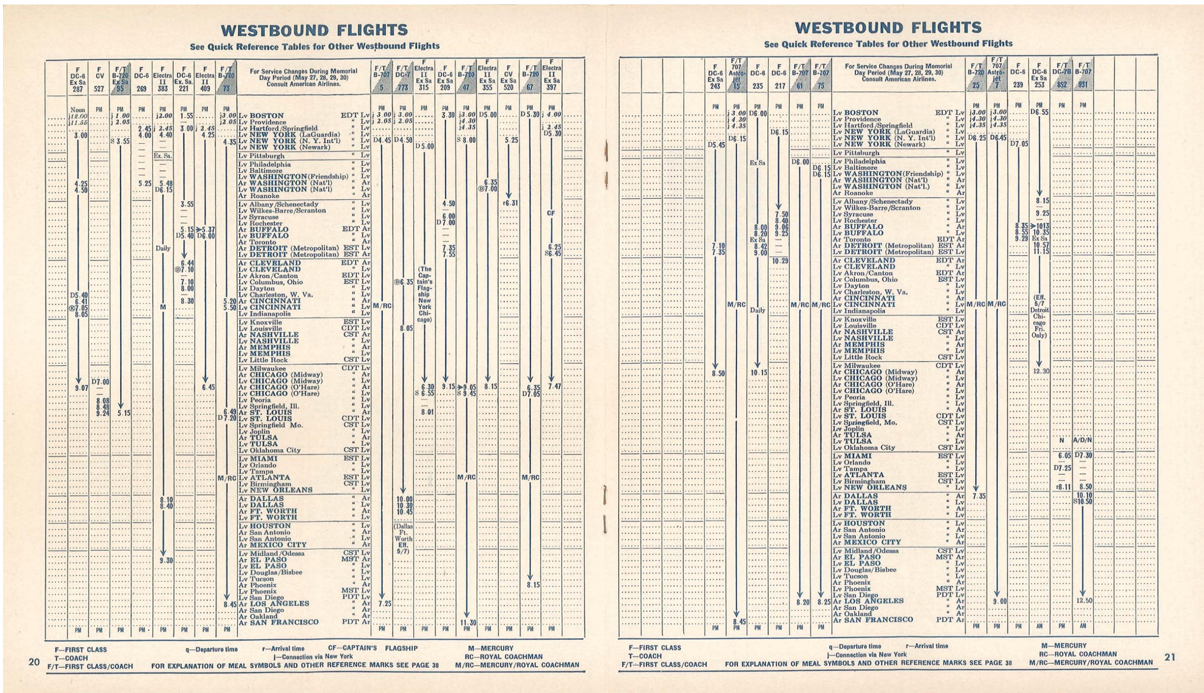


Figure 17: Flight Schedule American Airlines 1961.

To obtain exact geographical information on where airports are located, we match city names to their IATA airport codes. We use the addresses of ticket offices that are indicated on the last pages of the flight schedules. Most of the ticket offices were located directly at the airport, allowing to infer the airport the airline was serving in a given year. For some flight schedules we are missing these last pages and used information from adjacent years in order to identify airports. We also manually verified the airport match using various online sources. We then obtain geographical coordinates from a dataset provided by `https://ourairports.com/` (downloaded July 2020).

From the flight schedule we also collect information on the aircraft model, indicated next to the flight number. Using various online sources, we manually identified aircraft models that are powered by a jet engine. We thus know on which connections airlines were using jet aircraft.

Flight Schedules also contain information on connecting flights. For example, the second column in figure 17 indicates a departure from Boston leaving at 12.00 local time. A footnote is added to the departure time indicating that this departure is a connection via New York. It is thus not operated by flight 287 otherwise described in column 2, but it is just supplementary information for the passenger. As we are interested in the speed of aircraft and the actual travel time on a given link, this information on connecting flights would pollute our data and we thus delete this supplementary information.

As outlined above, the digitization requires human input. It is thus prone error-prone. The travel time calculation relies on each link in the network, and if one important connection has a miscoded flight, it might potentially distort the travel time between many MSA pairs. We thus implement an elaborate method to detect mistakes in the digitization process. In particular, after the initial transcription, we regress the observed duration of the flight on a set of explanatory variables: the full interaction of distance, a set of airline indicators, a set of year indicators and a dummy variable indicating whether the aircraft is powered by a jet engine or not. This linear model yields an

R^2 above 95%. We then compute the predicted duration of each flight and obtain the relative deviation from the observed duration. If the deviation is above 50%, we manually check whether the transcribed information is correct. If we find a mistake, we correct the raw data, rerun the regression and recompute relative deviations, until all the observations with more than 50% deviation have been manually verified.

For 15 connections, the information was correctly transcribed from the flight schedule, but the flight time differed a lot from other flights with similar distances that used the same aircraft. The implied aircraft speed for these cases is either unrealistically high or low, in one case the implied flight time is even negative. These cases seem to be typos introduced when the flight schedule was created (e.g. a ″2″ becomes a ″3″). Instead of inferring what the true flight schedule was which is not always obvious, we drop these cases. Table 6 lists all 15 cases.

Table 6: Dropped Connections

	Airline	Year	Origin	Destination	Departure Time	Arrival Time
0	UA	66	TYS	DCA	1940	2036
1	UA	66	LAX	BWI	2150	1715
2	UA	66	CHA	TYS	1635	1909
3	PA	66	SFO	LAX	2105	1850
4	PA	66	SEA	PDX	705	935
5	PA	56	PAP	SDQ	830	835
6	PA	51	HAV	MIA	800	903
7	PA	51	SJU	SDQ	825	830
8	NW	66	HND	OKA	655	1135
9	EA	66	ORD	MSP	2340	2340
10	EA	56	SDF	MDW	1352	1418
11	EA	56	GSO	RIC	2207	2204
12	AA	56	PHX	TUS	1630	1655
13	PA	51	STR	FRA	1320	1540
14	EA	66	TPA	JFK	1330	1548

As our analysis is at the MSA level, we match airports to 1950 MSA boundaries. Each airport is matched to all MSAs for which it lies inside the MSA boundary or at most

15km away from the MSA boundary. If we focus only on airports contained within MSA boundaries, we would, for example, drop Atlanta's airport. Of 275 US airports, 156 airports are matched to at least one MSA. 18 of these are matched to two MSAs and Harrisburg International Airport is matched to three MSAs: Harrisburg, Lancaster and York. Out of 168 MSAs, 142 are at some point connected to the flight network in our dataset. In table 7 we present the 168 MSAs, the ones that are connected at least once, and the ones that are connected in the four years.

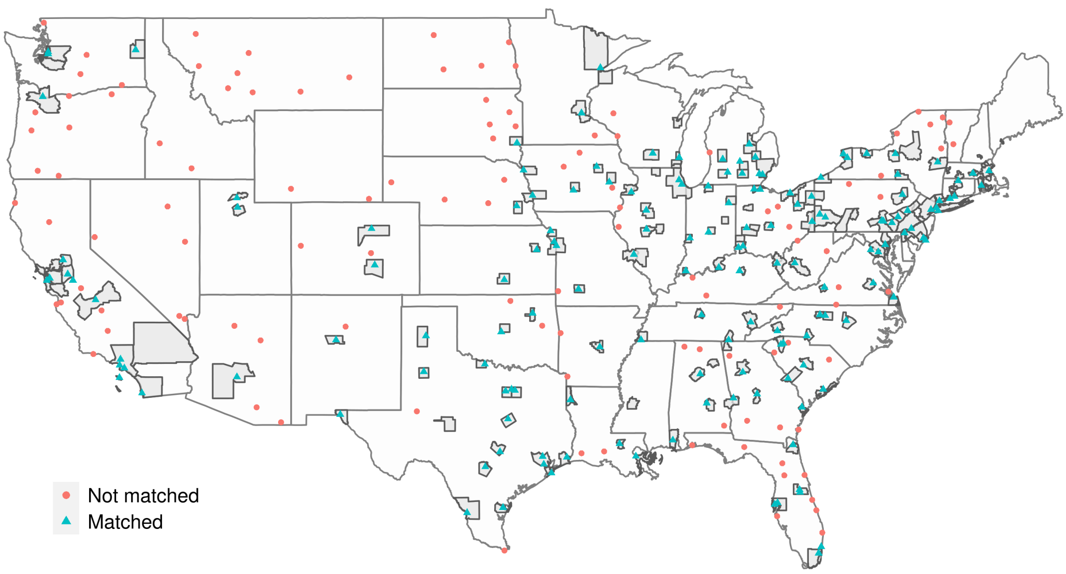


Figure 18: Airports matched to MSAs.

Next, we compute the shortest travel time for every airport pair, and then take the minimum to obtain shortest travel time at the MSA pair level. In particular, we apply Dijkstra's algorithm to compute shortest paths (Dijkstra et al., 1959). We adjust this algorithm to take into account the exact timing of the flight schedules. We consider a possible departure time t from origin city o and then compute the shortest path to destination city d at this time of the day. If getting to d requires switching flights, we account for the required time at the location of the layover. We repeat this procedure for every possible departure time t at origin city o and then take the minimum that gives us the fastest travel time from o to d, τ_{od}.

The flight schedule format requires us to make one assumption. In particular, the flight schedule for a multi-stop flight may either indicate the arrival time or the departure time for a particular stop. If the flight schedule only lists the departure time, we need to infer the arrival time and vice versa. We allow for five minutes between arrival and departure. This is relatively low, but still in the range of observed difference between departure and arrival for cases where we observe both. As correspondences may have been ensured by airlines in reality, i.e. one aircraft waiting with departure until other aircraft arrive, we opted for the lower end of the observed range of stopping times.

Finally, since the shortest travel time measure may not capture the benefits of a highly frequented hub, we also calculate the daily average of the shortest travel time. In particular, we compute the shortest travel time at every full hour of the day and take the average. This measure thus captures the benefits of being located near an airport where flights depart many times per day.

To conclude, we end up with a set of four origin-destination matrices indicating the fastest travel time (and another set with the average daily travel time) between US MSAs in 1951, 1956, 1961 and 1966.

A.2. Descriptive Statistics

Table 8 shows the number of non-stop connections between MSAs by year and airline. It underlines the dominant position of the *Big Four* (AA, EA, TW, UA) which were much bigger than their competitors (BN and NW). The growth of the airline industry is also apparent. All airlines had the lowest number of connections in 1951 and subsequently extended their network. At the same time the average distance of the connections gradually increased over time. Part of this may have been due to jet technology allowing for longer aircraft range. We thus analyze a period where more and longer flights are introduced.

Table 8: Domestic Non-Stop Connections by Airline and Year

Airline	Year	Number of connections	Jet Share (connections)	Jet Share (km)	Mean Distance (in km)
AA	1951	258	0.00	0.00	515.32
AA	1956	367	0.00	0.00	889.66
AA	1961	325	22.15	50.50	768.24
AA	1966	282	73.40	89.52	1020.36
BN	1951	96	0.00	0.00	317.90
BN	1956	210	0.00	0.00	380.60
BN	1961	176	8.52	18.84	460.41
BN	1966	150	72.00	76.64	553.09
EA	1951	345	0.00	0.00	319.87
EA	1956	479	0.00	0.00	412.60
EA	1961	595	3.70	13.28	441.42
EA	1966	492	54.47	75.46	569.01
NW	1951	77	0.00	0.00	521.70
NW	1956	95	0.00	0.00	724.77
NW	1961	127	11.02	32.43	824.59
NW	1966	136	77.94	90.86	945.81
TW	1951	210	0.00	0.00	503.69
TW	1956	253	0.00	0.00	711.78
TW	1961	240	28.75	54.63	807.72
TW	1966	265	86.42	96.05	1143.30
UA	1951	291	0.00	0.00	492.88
UA	1956	361	0.00	0.00	714.39
UA	1961	323	31.89	65.32	803.49
UA	1966	533	49.91	79.54	781.38

While these changes in the network are remarkable, airlines were constrained by the regulator in opening new routes. Accordingly, table 9 shows that the network remains relatively stable over time with more than three quarters of connections remaining intact within a five-year window. Interestingly, during the beginning of the jet age (i.e. 1956 to 1961), the network appears to have been especially stable, with only 11% of connections either disappearing or newly being added. Thus, the rise of jet aircraft did not lead to a vast reshaping of the network. Given the very different technology, this may be surprising, but may partly be due to heavy regulation.

The table also shows that newly introduced routes were over long distances whereas those discontinued were operating on shorter distances. When changes in the network took place, they thus seemed to improve the network for places further apart.

Table 9: Network Changes (weighted by frequency)

Period	Remain connected	Newly connected	Disconnected
Share of Non-stop Connections (%)			
1951 to 1956	78.47	16.79	4.74
1956 to 1961	88.96	6.43	4.6
1961 to 1966	80.64	12.37	6.99
Mean distance (km)			
1951 to 1956	411	1075	337
1956 to 1961	524	914	972
1961 to 1966	568	769	450

Table 10: Network Changes

Period	Remain connected	Newly connected	Disconnected
Connected MSAs			
1951 to 1956	119	7	8
1956 to 1961	122	0	4
1961 to 1966	114	7	8
Non-stop Connections			
1951 to 1956	721	357	124
1956 to 1961	908	231	170
1961 to 1966	912	331	227

Changes in the number of connected MSAs and connections among them. A MSA is connected if in our data it appears as having at least one incoming and one outgoing flight. A non-stop connection refers to a pair of origin MSA-destination MSA between which a non-stop flight operates.

Figure 19 shows all non-stop connections in our data weighted by the (log) frequency. Initially, the network was concentrated in the Eastern states and transcontinental routes were not yet established, due to technological limitations. In contrast, in the 1960s, after the jet is introduced, intercontinental routes quickly emerge and are operated at a high frequency. Similarly, direct connections from the Northeast to Florida intensify. The figure echos the findings from table 10 which illustrates that the overall number of MSA pairs with a direct connection increases over time.

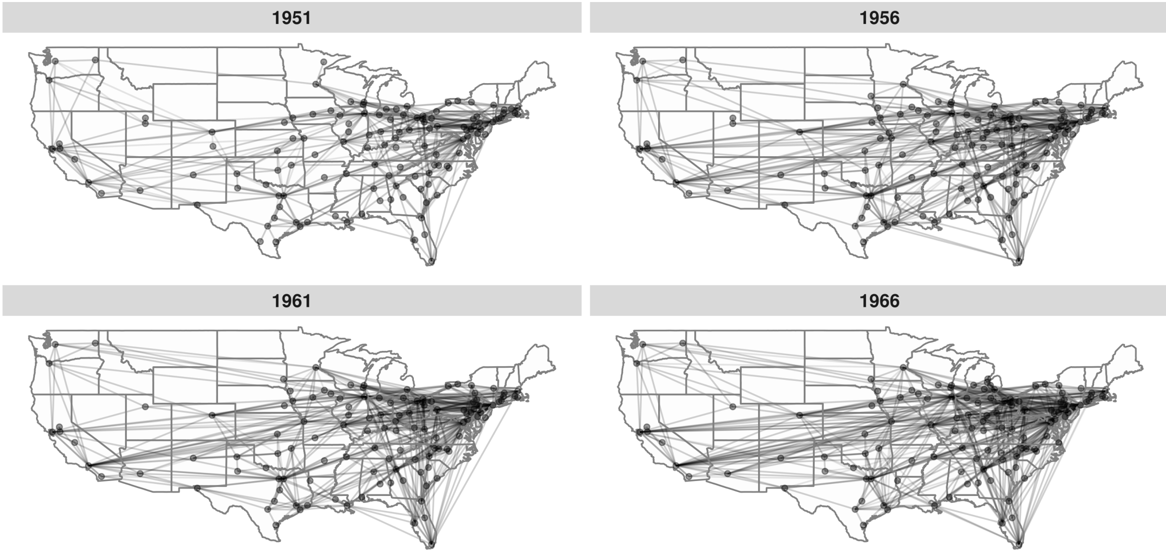


Figure 19: Flight Network by Year. Weighted by log weekly frequency.

Airlines differed in their speed of adoption of the newly arrived jet aircraft. Table 8 shows that, in 1961, 65% of UA's connections between MSAs were flown using a jet aircraft (weighted by distance), whereas this was only true for 13% of EA's connections. While adoption was heterogeneous across airlines, adoption was fast. By 1966, all airlines were operating 75% of their connections with jet aircraft (weighted by distance).

Figure 20 show the average speed of jet and propeller aircraft by distance. Generally, jet aircraft were substantially faster, but especially so on long-distance flights, where they could be up to twice as fast as propeller-driven aircraft. This particularly stark difference in speed for long-haul flights is also reflected by adoption. Figure 21 shows that jet aircraft were first introduced on long-haul flights. Only 50% of MSA pairs at around 1,500 km distance had at least one jet aircraft operating, whereas 100% of pairs above 3,000 km. Then, in the late 1960s, they were also gradually introduced on shorter distances. In fact, for all pairs above 2,000 km there was at least one jet engine-powered flight.

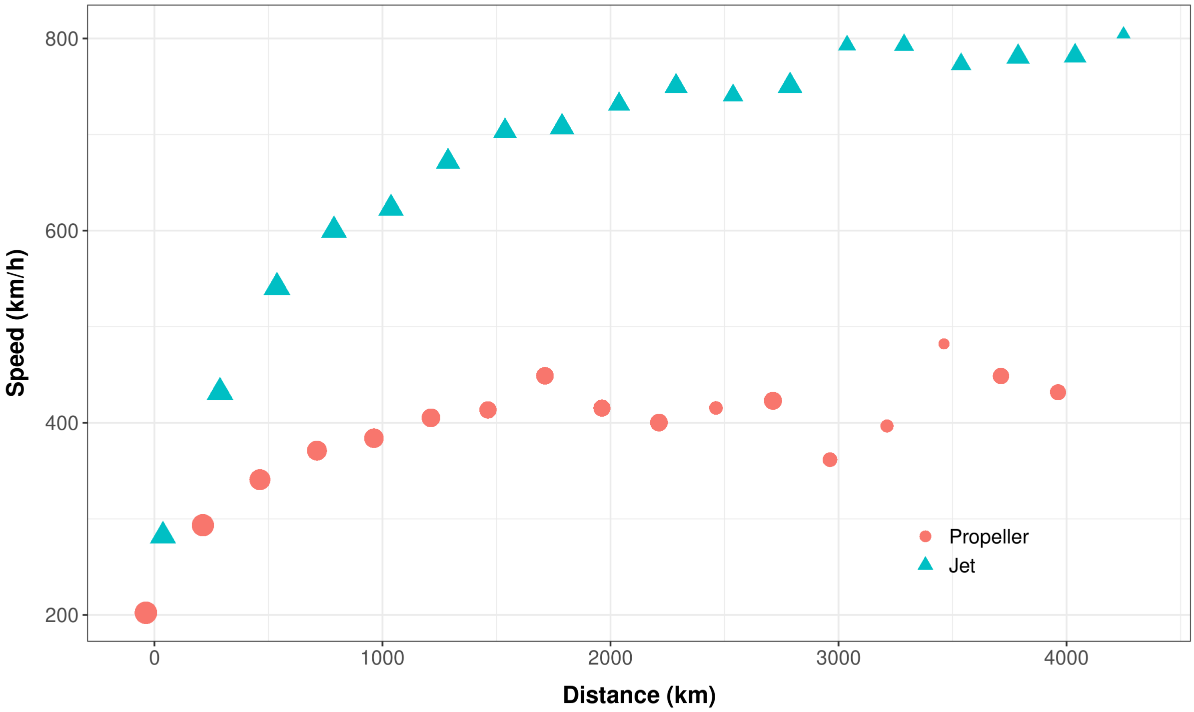


Figure 20: Speed by Aircraft Type. Pooling all Years.

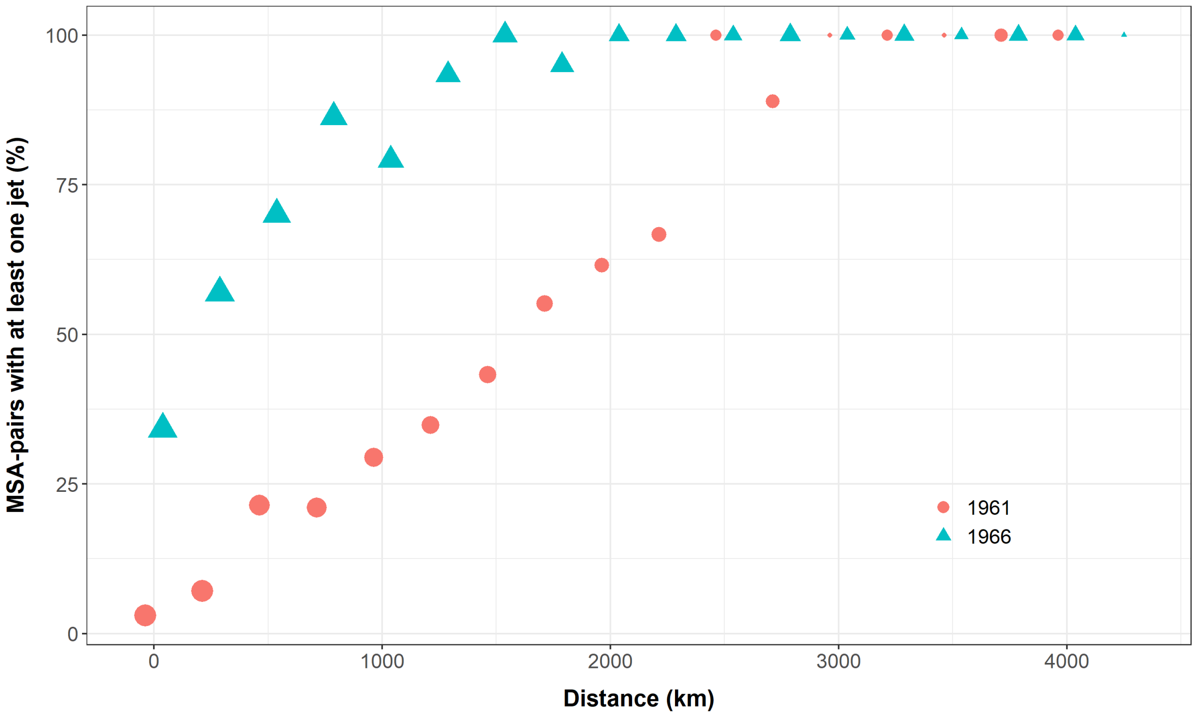


Figure 21: Jet Adoption

Figure 22 shows on which routes jets were operating. In the early days of the jet age it was mainly the transcontinental corridor between New York and California that benefited. In 1966 propeller aircraft were already being phased out and only operating in the dense Eastern part of the US where distances between cities are relatively small.

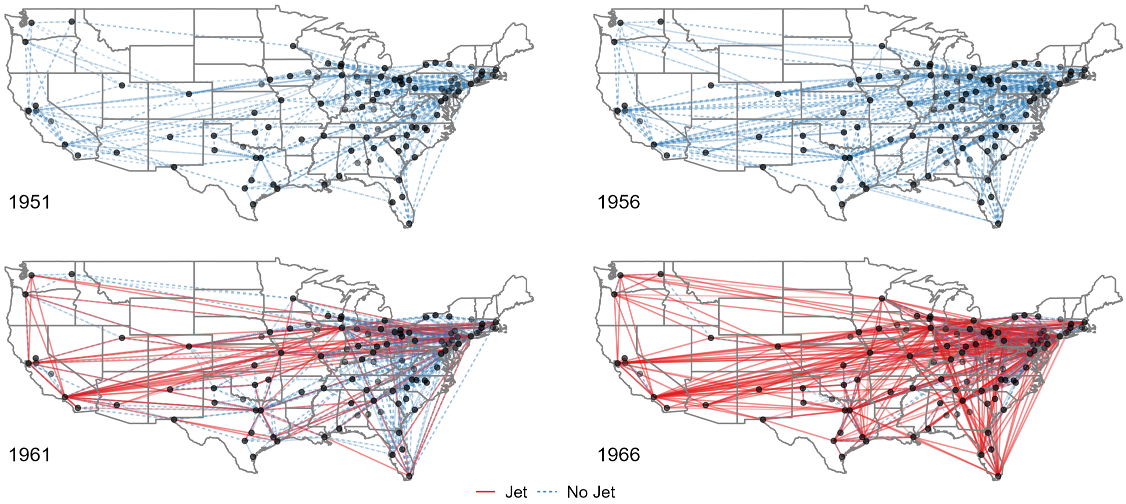


Figure 22: Jet Adoption by Year

The increase in speed due to jet aircraft caused a dramatic reduction in travel times between US cities. When looking at the full origin-destination matrix, i.e. including indirect flights, a network-wide reduction in travel time becomes apparent. Figure 23 shows travel times between US MSAs. While the figure shows a gradual decline in travel time from 1951 to 1966, it also illustrates that conditional on distance and year a large amount of variation in travel time remains, as only a small fraction of all MSA pairs were connected via a direct flight (around 8.5% in 1966).

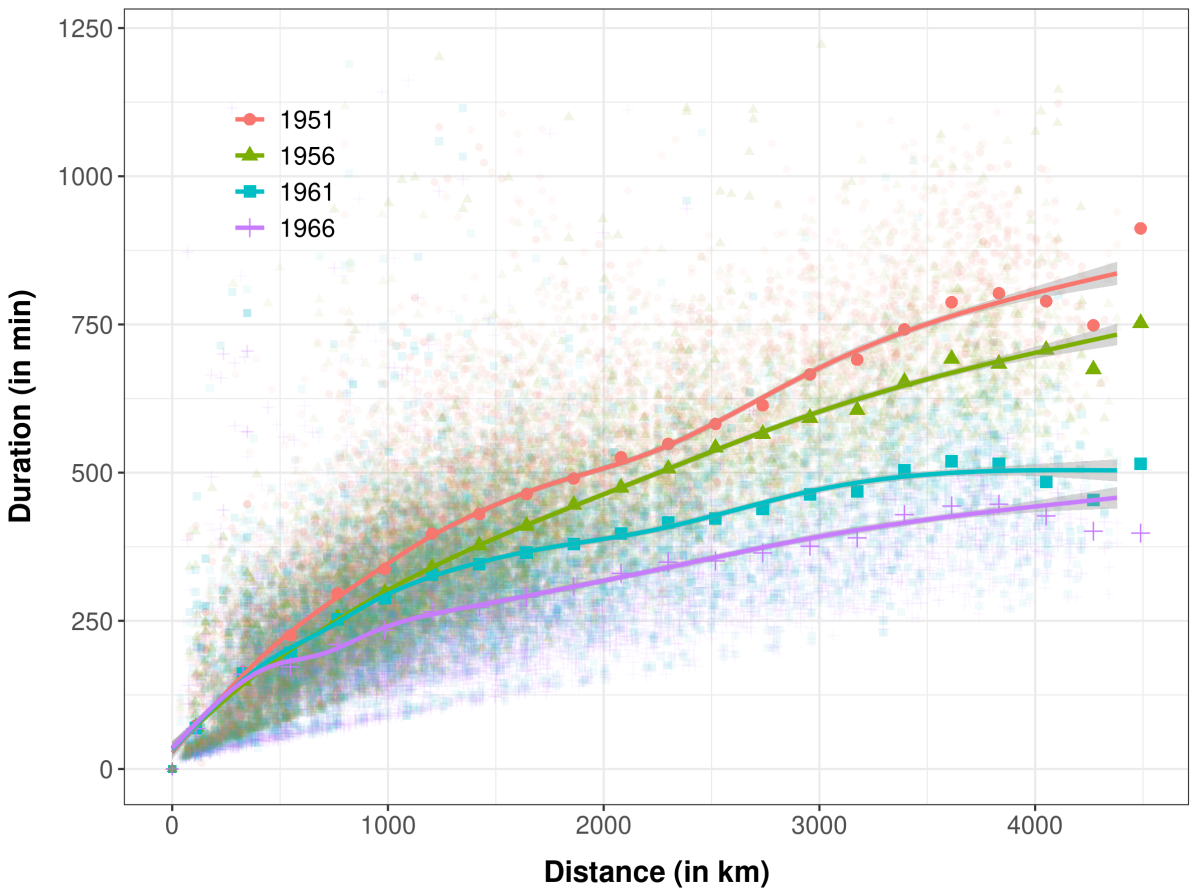


Figure 23: Travel Times between US MSAs.

Figure 24 that the change in travel time is accompanied by a reduction of the amount of legs needed to connect two MSAs at every distance. This reduction is specially marked between 1951 and 1956, and 1961 and 1966. In Figure 25 we open up the change in travel time by the way an MSA pair was connected in 1951 and 1966: either directly (non-stop flight) or indirectly (connecting flight). We observe that much of the increase in travel time for MSA pairs less than 250km apart comes from routes that were operated non-stop and then it needed a connecting flight. Interestingly, for MSA-pairs more than 2,000km apart travel time reduced on average 42% for those pairs that were connected indirectly in both periods, and 51% for those that switched from indirect to direct. This fact shows the relevance of improvements in flight technology even for MSAs not directly connected. It could be the case that a reduction in the amount of legs or an increase in frequency of flights reduces layover time. In Figure 27 we compare the change in travel time from 1951 to 1966 with a fictitious change in travel time in which we eliminate layover time in both time periods. We observe that the average change in travel time is stronger at every distance if we disregard layover time. This implies that the relative importance of layover time over total travel time increases between 1951 and 1966, preventing total travel time to decrease proportionally to the change of in-flight travel time.

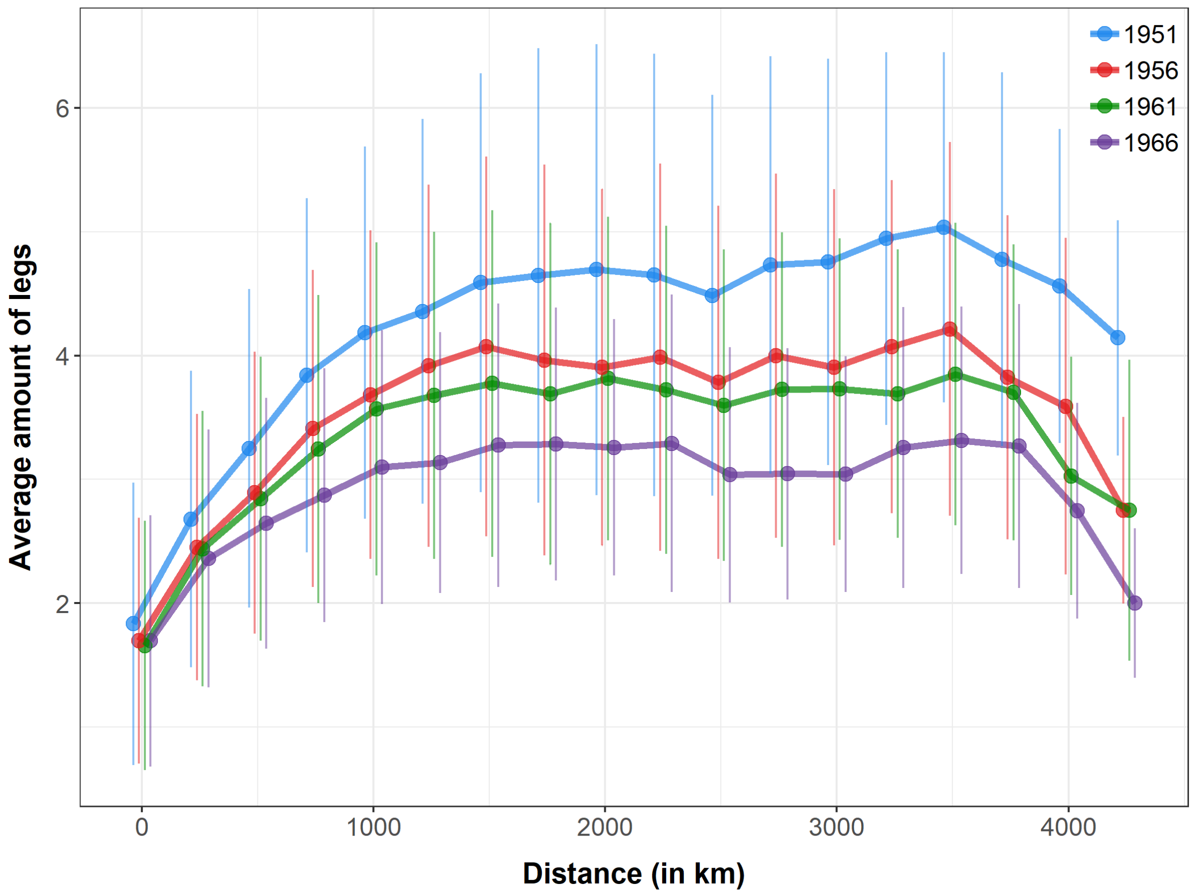


Figure 24: Average amount of legs per route

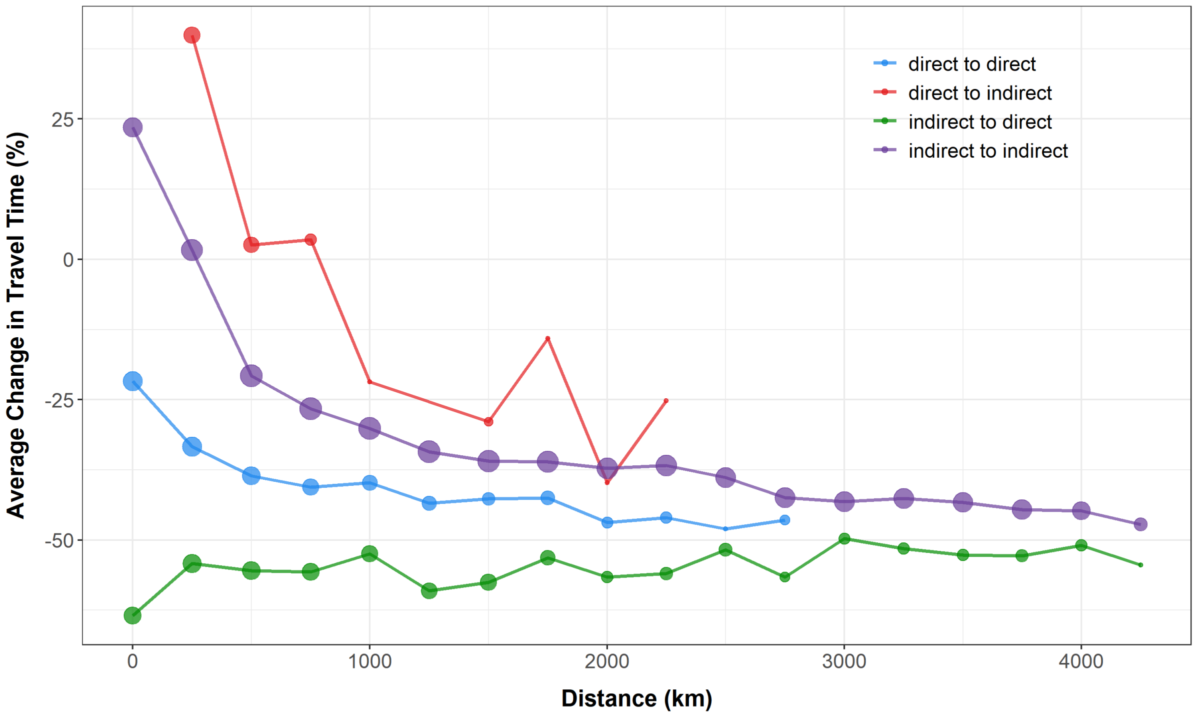


Figure 25: Change in US travel time 1951 to 1966: connections
92

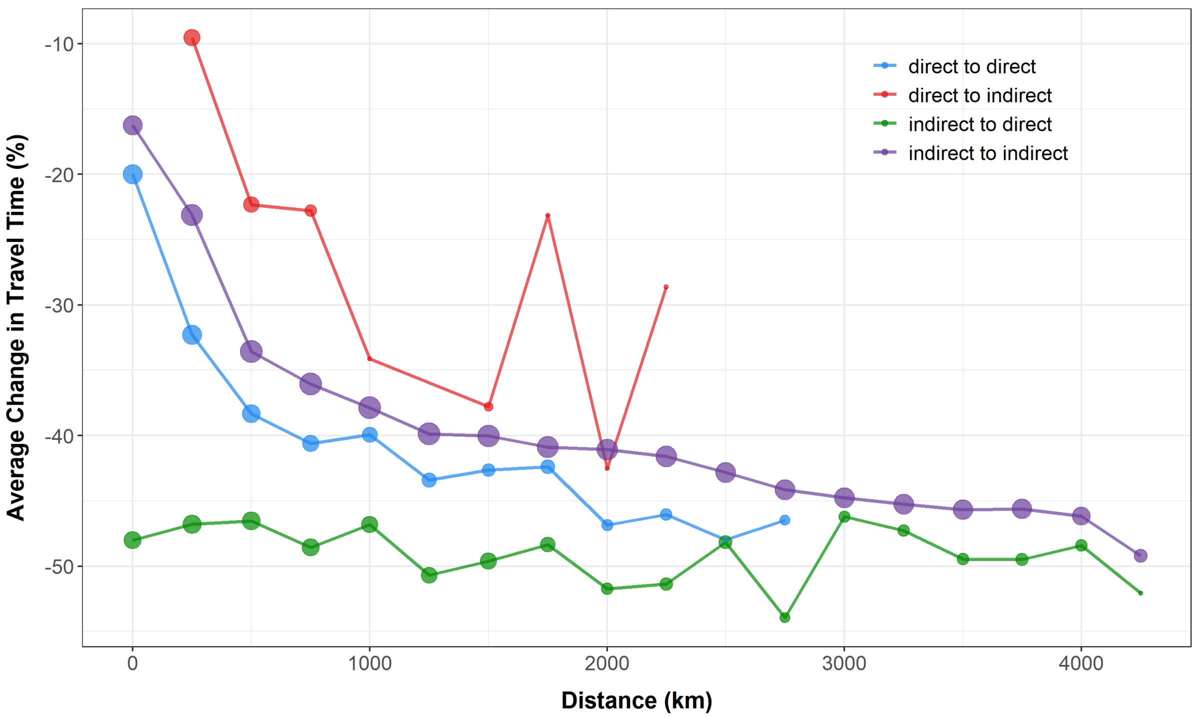


Figure 26: Change in US travel time 1951 to 1966: connections, discarding layover time
93

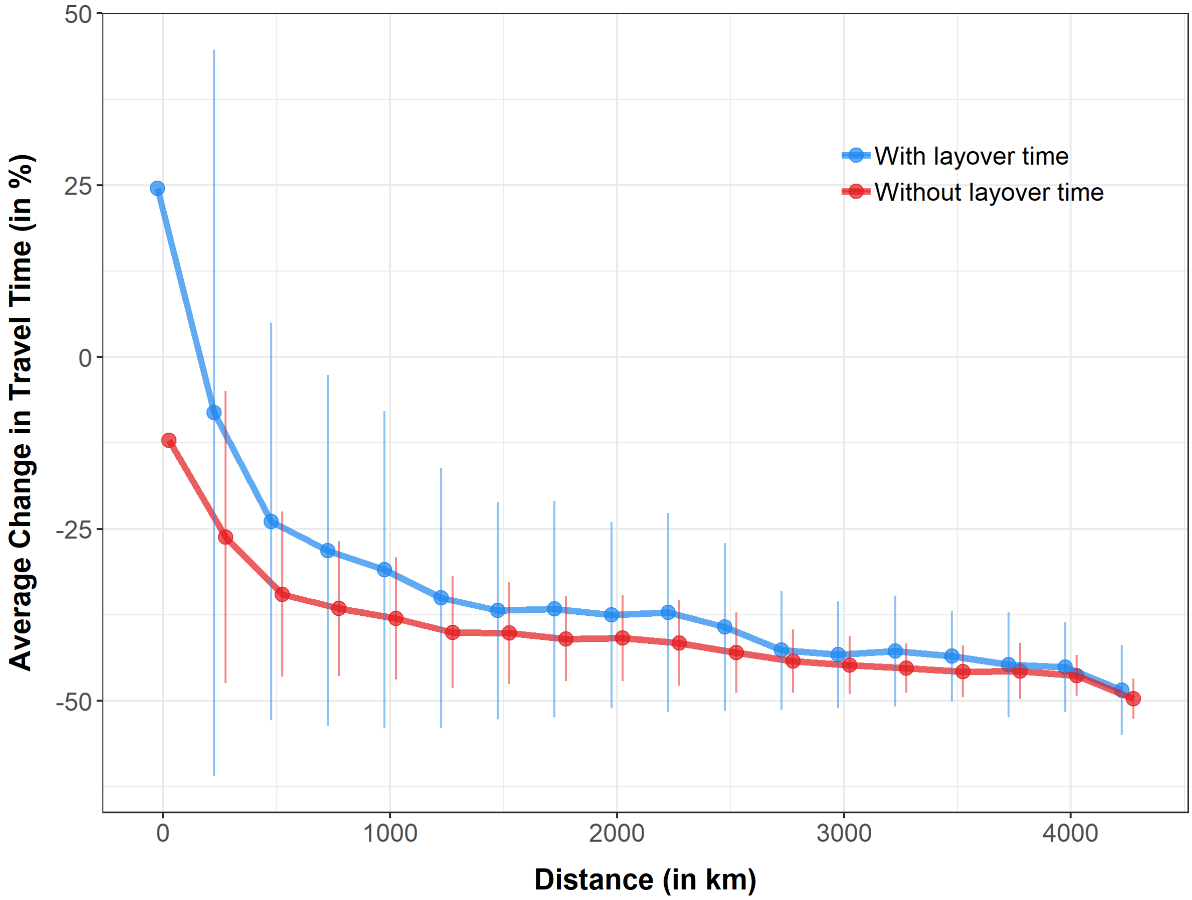


Figure 27: Change in US travel time 1951 to 1966: layover time

In figure 28 we show the average change in travel time in three counterfactual flight networks. The first counterfactual fixes the flight routes[94] and allows aircraft speed to evolve. The second counterfactual fixes aircraft speed and allows flight routes to evolve. The third counterfactual allows both flight routes and aircraft speed to evolve. We obtain that around 90% of the change in travel time is due to the change in speed of aircrafts, while around 10% of the change is due to the change in the flight routes. In the figure 29 in the appendix we show that the proportion is relatively constant for all distances. This confirms that most of the observed changes in the network are due to improvements in the flight technology.

[94]Fixes the origin-destination airports that are connected with a non-stop flight

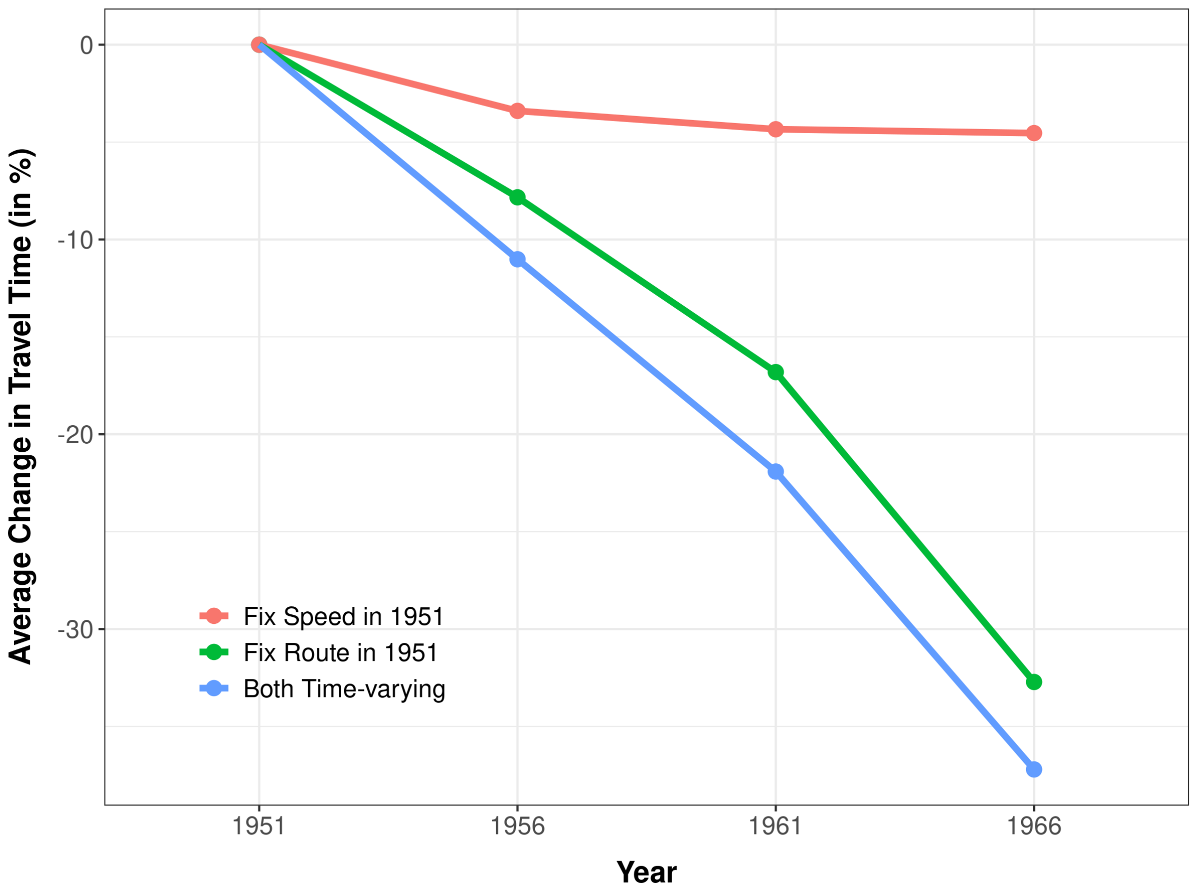


Figure 28: Counterfactual change in travel time

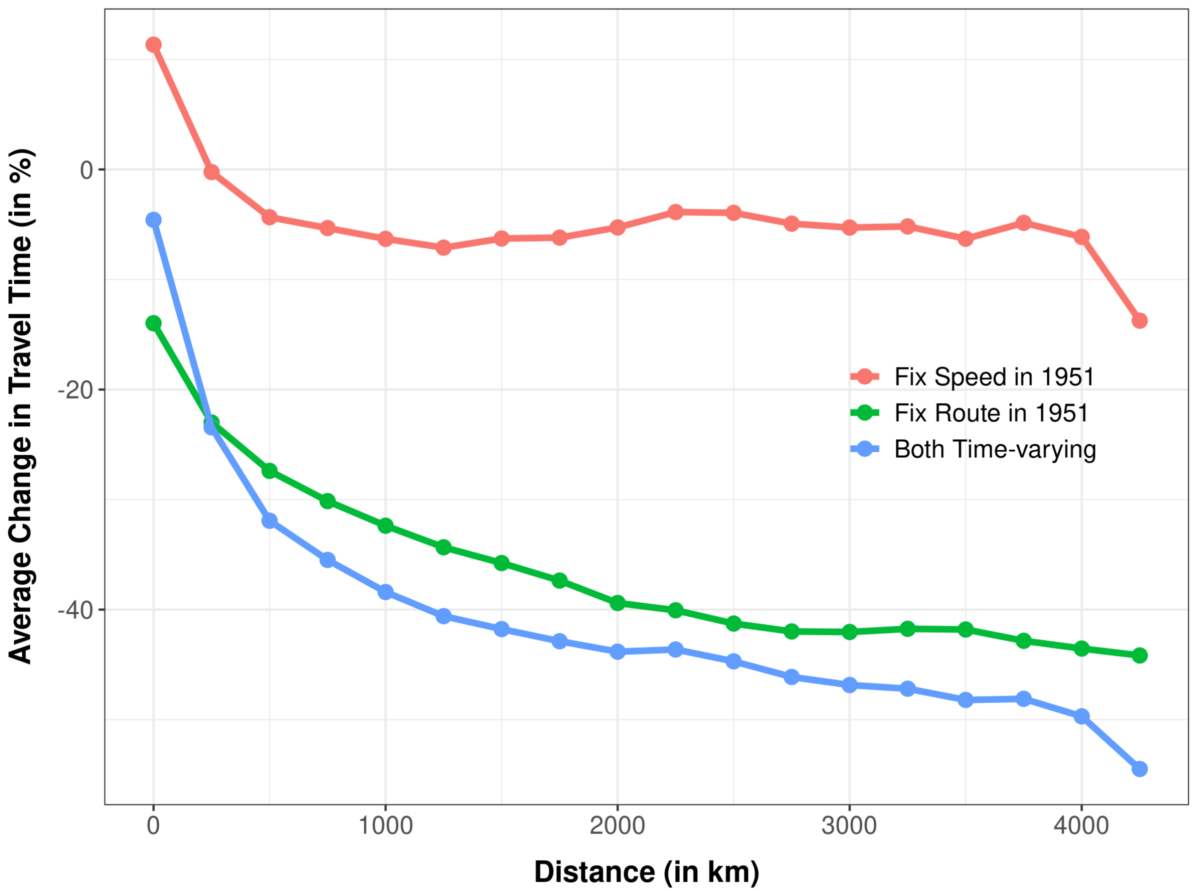


Figure 29: Counterfactual change in travel time 1951-1966

In addition to the changes over time in the network leading to faster travel times, another feature of the US airline industry becomes salient in the data: airlines' regional specialization. As figure 30 shows, while there was competition among the airlines in our dataset on the major routes (Lower West Coast to the Midwest and Upper East Coast to the Midwest), some airlines are very specialized and face no competition from any of the other five airlines on certain routes. In particular, NW controls the routes connecting Seattle to the Midwest and EA controls much of the connections from Florida to New York and surroundings.

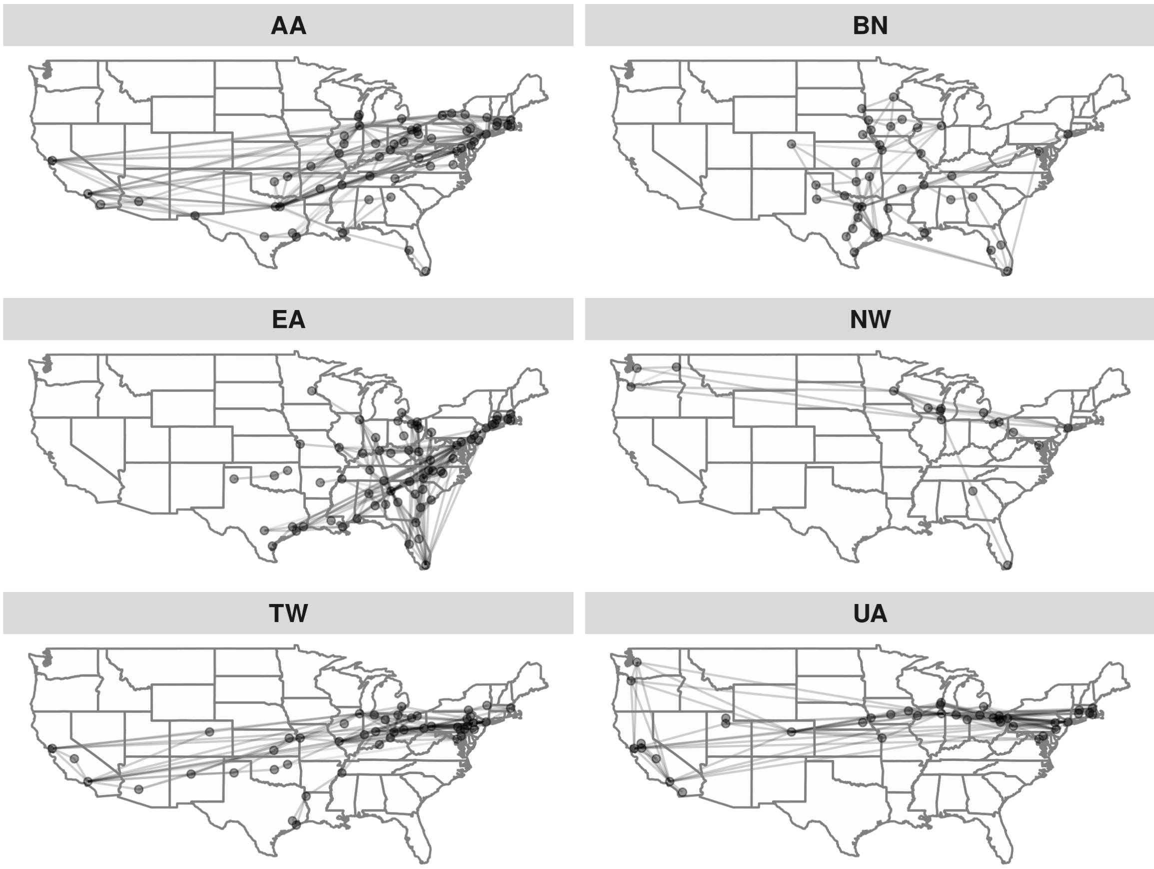


Figure 30: Flight Network in 1956 by Airline (weighted by log frequency).

A.3. Instrumental travel time

In order to construct the instrumental travel time we first estimate, separately for each year, a linear regression of travel time on flight distance using only the fastest non-stop flight in each origin-destination airport pairs. These yearly regressions provide us with the fictitious average airplane of each year: the intercept gives the take-off and landing time of the airplane while the slope provides the (inverse) speed. Results on this estimation are provided in Table 11.

Table 11: Regression of travel time on distance fastest non-stop flights

	Travel time (min)			
Year	1951	1956	1961	1966
	(1)	(2)	(3)	(4)
Constant	25.3***	24.1***	39.5***	29.9***
	(0.809)	(0.656)	(0.921)	(0.678)
Distance (km)	0.146***	0.132***	0.079***	0.068***
	(0.001)	(0.0007)	(0.0010)	(0.0006)
Observations	1,137	1,479	1,438	1,490
R^2	0.93	0.96	0.82	0.90
Implied speed (km/h)	412	453	758	876

The table presents the results of estimating by OLS: travel time$_{ijt}$ $= \alpha_0 + \alpha_1 \times distance_{ij} + \varepsilon_{ijt}$ separately for each year $t \in \{1951, 1956, 1961, 1966\}$. The sample consist of all airport pairs that are connected with a non-stop flight in the respective year. Travel time is the fastest non-stop flight between the airports measured in minutes. The implied speed is calculated as the inverse of the coefficient on distance multiplied by 60.

B. Appendix: Patent data

In this appendix we describe facts that we observe in the US patent data, for patents filed[95] between 1945 and 1975. US patents data containing citations and filing year have been downloaded from Google Patents. Then, it was merged with multiple datasets (see Appendix Patent Data Construction for more details):

- Technology classification: NBER patent database.
- Geographic location of inventors: Histpat and Histpat International for patents

[95]Filing year, also called application year, is the closest date to the date of invention that is present in the data and it represent the date of the first administrative event in order to obtain a patent. In the other hand, publishing or also called granting year, is the later year in which the patent is granted. The difference between filing and granting year depends on diverse non-innovation related factors (as capacity of the patent office to revise applications) and changes over time. Hence filing year is the date in our data that approximates the best to the date of invention.

published until 1975, Fung Institute for patents published after 1975. Both matched to 1950s Metropolitan Statistical Areas (MSAs).

- Ownership: Kogan et al. (2017) for patents owned by firms listed in the US stock market, Patstat for the remaining patents not matched to Kogan et al. (2017).

We highlight two details from the matching process: 1. During filing years 1971-1972 the rate of non-geocoded patents increases, possibly due to Histpat and Fung data not being a perfect continuation one of the other. 2. Kogan et al. (2017) seems to use a matching method based on the patent owner declared in the patent text, as Patstat does. Specially, Kogan et al. (2017) does not explicitly say if it takes into account firm-ownership structure to determine patent ownership, neither does Patstat.

For the analysis presented in this appendix we will use the resulting dataset from the matching procedure, where unless evident or noticed, we will use only patents that have inventors within MSAs. We discard patents that have inventors in multiple MSAs and patents that belong to government organizations or universities. We assign patents to technology categories using fractional count: if a patent is listed in two technology categories, then we assign half a patent to each category. We discard self citations (citations in which the citing patent owner is the same as the cited patent owner) because self-citations may be due to different incentives.

B.1. Matching patents to locations

In figure 31 we observe that the matching rate decreases from around 95% before 1970, to around 80% in 1971 and 1972, and then it stabilizes around 99% after 1975. Hence, geogprahical results during years 1970-1975 will contain an increased amount of measurement error.

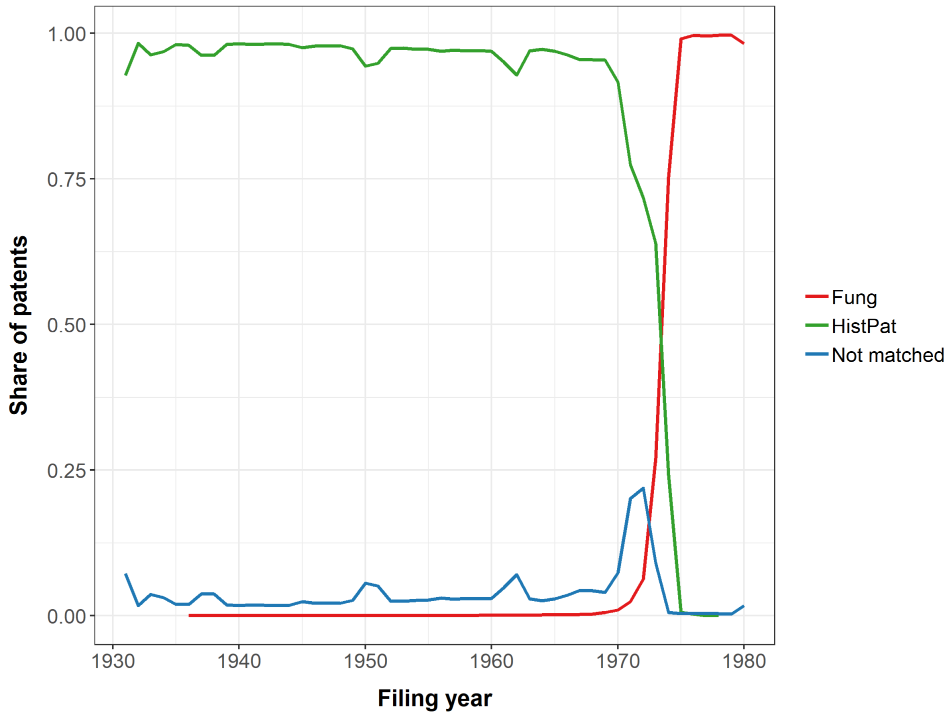


Figure 31: Non-matching rate HistPat, HistPat International and Fung

Figure 32 shows the share of patents that have inventors inside MSAs, and figure 33 displays the same by technology category.[96]

[96]Technologies are aggregated to six big groups, as explained in HJT 2002

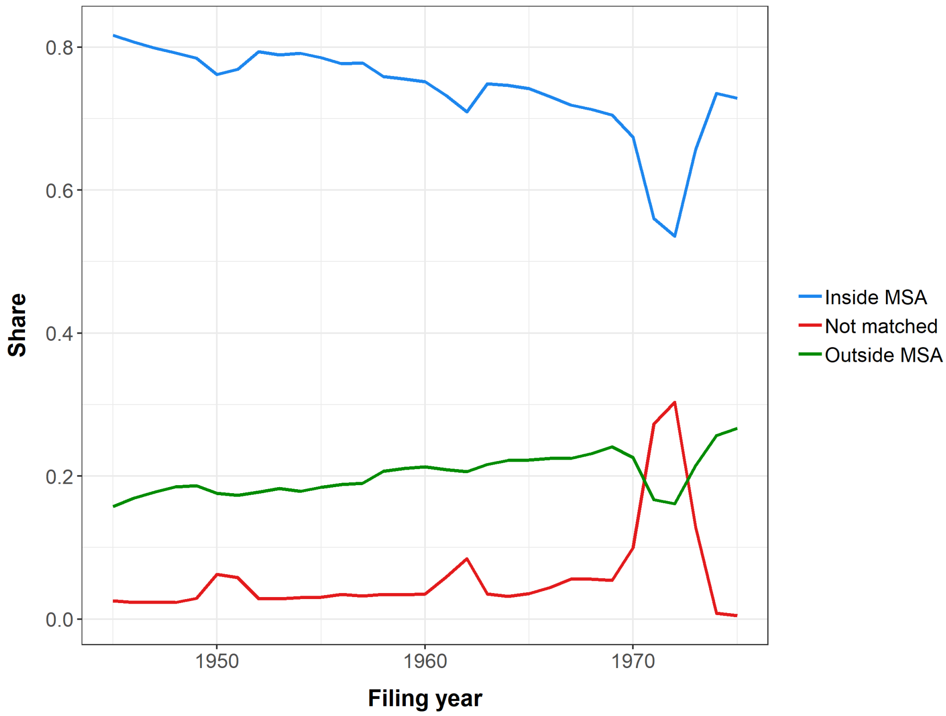


Figure 32: Share patents in Metropolitan Statistical Areas

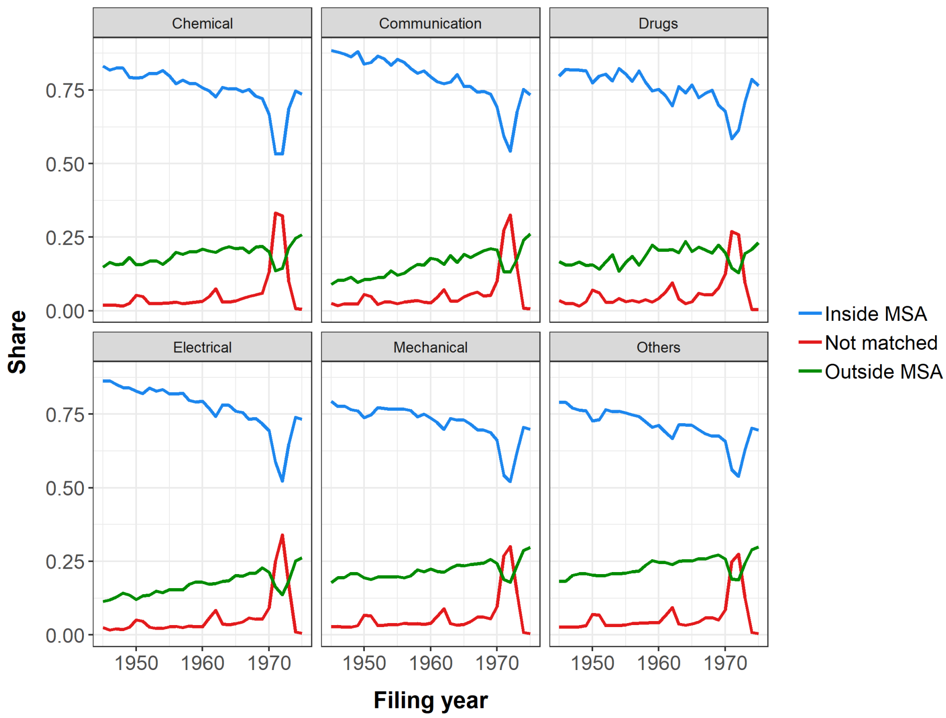


Figure 33: Share patents in Metropolitan Statistical Areas

B.2. Input-Output of patents

In the same spirit as how Input-Output tables of industries are constructed, we can use citations as a reflection of sourced (input) knowledge. In this case, we interpret the cited patent as being a source of knowledge, and the citing patent as being a destination. In Figure 34 we aggregate citations by citing-cited technology category in the years 1949-1953. Rows represent the source technology and columns the destination technology. Columns should sum to 1 (round errors may exist). We highlight in bold those IO coefficients that are higher than 0.1. We observe that the diagonal has coefficients greater than 0.5, implying that technologies rely on themselves to create new knowledge. At the same time, we observe the importance of Electrical to create Communication technologies, and the small relevance of Drugs for every other technology.

Source/Destination	Chemical	Communication	Drugs	Electrical	Mechanical	Others
Chemical	**0.74**	0.01	**0.13**	0.03	0.05	0.05
Communication	0	**0.6**	0	0.07	0.01	0.01
Drugs	0.01	0	**0.6**	0	0	0.01
Electrical	0.03	**0.28**	0.03	**0.7**	0.05	0.04
Mechanical	**0.11**	0.07	0.07	0.1	**0.72**	**0.15**
Others	**0.11**	0.05	**0.16**	0.09	**0.16**	**0.75**
Total	1	1	1	1	1	1

Figure 34: Input-Output of technologies 1949-1953

B.3. Descriptive statistics

Table 12 shows descriptive statistics along each step of the patent data matching and sample selection. The final dataset contains 515,089 patents and 1,639,326 citations.

Table 12: Patent data sample selection

Sample	N patents	N citations	1st quartile cit dist (km)	2nd quartile cit dist (km)	3rd quartile cit dist (km)
Google patents	964,582	4,392,725			
With location	923,150	4,191,886			
US	749,410	3,569,578			
MSA	589,870	2,354,844			
Single location	571,969	2,237,095	213	730	1,682
With owner id	571,824	1,963,644	199	696	1,673
Non gov/univ	565,372	1,932,297	199	696	1,664
With travel time (final sample)	515,089	1,639,326	184	689	1,645

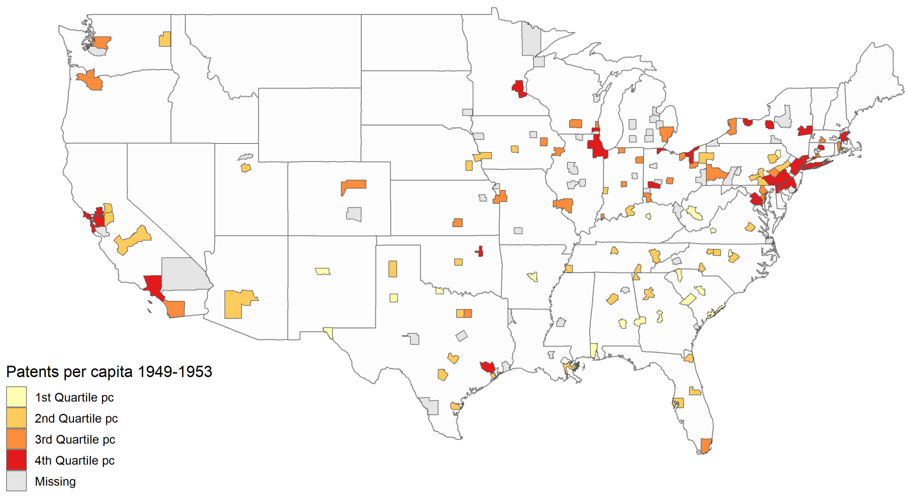


Figure 35: Patents per capita in 1951
Quantiles of patents per capita are computed in each technology and then averaged across technologies. Population is from 1950 Census.

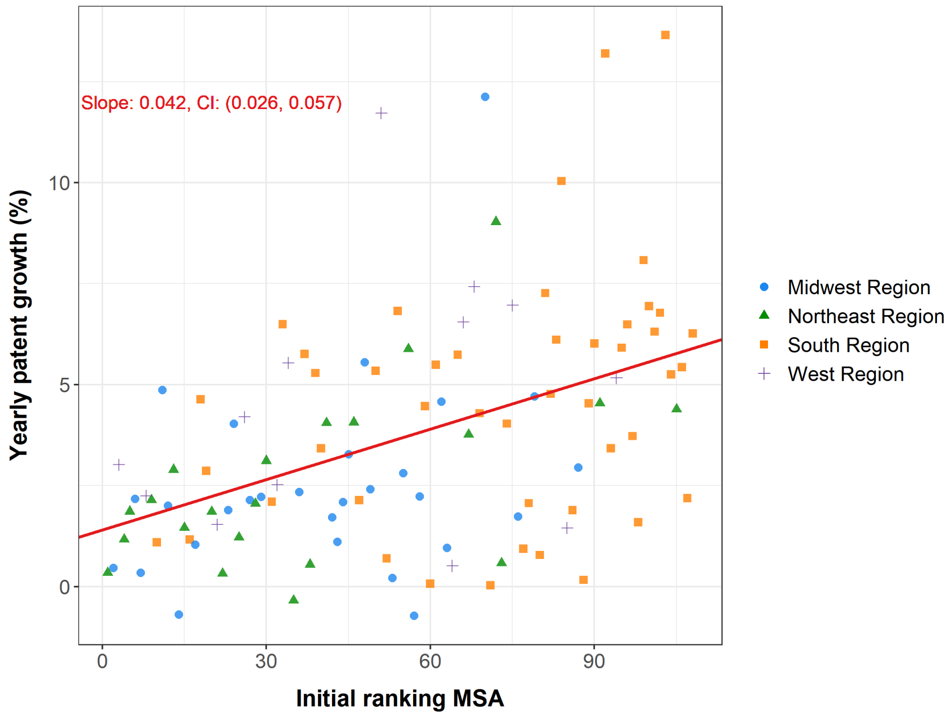


Figure 36: Patent growth by initial innovativeness ranking of MSA

Table 7: Connected MSAs

MSA fips	MSA name	<=3 periods	4 periods	MSA fips	MSA name	<=3 periods	4 periods
80	Akron, OH SMA	X	X	4680	Macon, GA SMA	X	X
160	Albany-Schenectady-Troy, NY SMA	X	X	4720	Madison, WI SMA	X	X
200	Albuquerque, NM SMA	X	X	4760	Manchester, NH SMA		
240	Allentown-Bethlehem-Easton, PA-NJ SMA	X	X	4920	Memphis, TN SMA	X	X
280	Altoona, PA SMA			5000	Miami, FL SMA	X	X
320	Amarillo, TX SMA	X	X	5080	Milwaukee, WI SMA	X	X
480	Asheville, NC SMA	X		5120	Minneapolis-St. Paul, MN SMA	X	X
520	Atlanta, GA SMA	X	X	5160	Mobile, AL SMA	X	X
560	Atlantic City, NJ SMA	X		5240	Montgomery, AL SMA	X	X
600	Augusta, GA-SC SMA	X	X	5280	Muncie, IN SMA		
640	Austin, TX SMA	X	X	5360	Nashville, TN SMA	X	X
720	Baltimore, MD SMA	X	X	5400	New Bedford, MA SMA		
760	Baton Rouge, LA SMA	X		5440	New Britain-Bristol, CT SMA		
800	Bay City, MI SMA	X		5480	New Haven, CT SMA	X	X
840	Beaumont-Port Arthur, TX SMA	X		5560	New Orleans, LA SMA	X	X
960	Binghamton, NY SMA	X		5600	New York-Northeastern NJ, NY-NJ SMA	X	X
1000	Birmingham, AL SMA	X	X	5720	Norfolk-Portsmouth, VA SMA	X	
1120	Boston, MA SMA	X	X	5840	Ogden, UT SMA	X	
1160	Bridgeport, CT SMA	X	X	5880	Oklahoma City, OK SMA	X	X
1200	Brockton, MA SMA			5920	Omaha, NE-IA SMA	X	X
1280	Buffalo, NY SMA	X	X	5960	Orlando, FL SMA	X	X
1320	Canton, OH SMA	X	X	6120	Peoria, IL SMA	X	
1360	Cedar Rapids, IA SMA	X	X	6160	Philadelphia, PA-NJ SMA	X	X
1440	Charleston, SC SMA	X	X	6200	Phoenix, AZ SMA	X	X
1480	Charleston, WV SMA	X	X	6280	Pittsburgh, PA SMA	X	X
1520	Charlotte, NC SMA	X	X	6320	Pittsfield, MA SMA		
1560	Chattanooga, TN-GA SMA	X	X	6400	Portland, ME SMA		
1600	Chicago, IL-IN SMA	X	X	6440	Portland, OR-WA SMA	X	X
1640	Cincinnati, OH-KY SMA	X	X	6480	Providence, RI SMA	X	X
1680	Cleveland, OH SMA	X	X	6560	Pueblo, CO SMA	X	
1760	Columbia, SC SMA	X	X	6600	Racine, WI SMA	X	X
1800	Columbus, GA-AL SMA	X	X	6640	Raleigh, NC SMA	X	X
1840	Columbus, OH SMA	X	X	6680	Reading, PA SMA	X	X
1880	Corpus Christi, TX SMA	X	X	6760	Richmond, VA SMA	X	X
1920	Dallas, TX SMA	X	X	6800	Roanoke, VA SMA	X	X
1960	Davenport-Rock Island-Moline, IA-IL SMA	X	X	6840	Rochester, NY SMA	X	X
2000	Dayton, OH SMA	X	X	6880	Rockford, IL SMA		
2040	Decatur, IL SMA			6920	Sacramento, CA SMA	X	X
2080	Denver, CO SMA	X	X	6960	Saginaw, MI SMA	X	
2120	Des Moines, IA SMA	X	X	7000	St. Joseph, MO SMA	X	
2160	Detroit, MI SMA	X	X	7040	St. Louis, MO-IL SMA	X	X
2240	Duluth-Superior, MN-WI SMA	X		7160	Salt Lake City, UT SMA	X	X
2280	Durham, NC SMA	X	X	7200	San Angelo, TX SMA		
2320	El Paso, TX SMA	X	X	7240	San Antonio, TX SMA	X	X
2360	Erie, PA SMA	X		7280	San Bernardino, CA SMA		
2440	Evansville, IN SMA	X	X	7320	San Diego, CA SMA	X	X
2480	Fall River, MA-RI SMA	X	X	7360	San Francisco-Oakland, CA SMA	X	X
2640	Flint, MI SMA	X		7400	San Jose, CA SMA		
2760	Fort Wayne, IN SMA	X	X	7520	Savannah, GA SMA	X	
2800	Fort Worth, TX SMA	X	X	7560	Scranton, PA SMA	X	X
2840	Fresno, CA SMA	X	X	7600	Seattle, WA SMA	X	X
2880	Gadsden, AL SMA			7680	Shreveport, LA SMA	X	
2920	Galveston, TX SMA	X	X	7720	Sioux City, IA SMA	X	
3000	Grand Rapids, MI SMA	X		7760	Sioux Falls, SD SMA	X	
3080	Green Bay, WI SMA			7800	South Bend, IN SMA	X	X
3120	Greensboro-High Point, NC SMA	X	X	7840	Spokane, WA SMA	X	X
3160	Greenville, SC SMA	X	X	7880	Springfield, IL SMA	X	
3200	Hamilton-Middletown, OH SMA			7920	Springfield, MO SMA	X	
3240	Harrisburg, PA SMA	X	X	7960	Springfield, OH SMA		
3280	Hartford, CT SMA	X	X	8000	Springfield-Holyoke, MA-CT SMA	X	X
3360	Houston, TX SMA	X	X	8040	Stamford-Norwalk, CT SMA	X	
3400	Huntington-Ashland, WV-KY-OH SMA	X		8120	Stockton, CA SMA	X	X
3480	Indianapolis, IN SMA	X	X	8160	Syracuse, NY SMA	X	X
3520	Jackson, MI SMA	X		8200	Tacoma, WA SMA		
3560	Jackson, MS SMA			8280	Tampa-St. Petersburg, FL SMA	X	X
3600	Jacksonville, FL SMA	X	X	8320	Terre Haute, IN SMA	X	X
3680	Johnstown, PA SMA			8400	Toledo, OH-MI SMA	X	X
3720	Kalamazoo, MI SMA	X		8440	Topeka, KS SMA	X	
3760	Kansas City, MO-KS SMA	X	X	8480	Trenton, NJ SMA		
3800	Kenosha, WI SMA			8560	Tulsa, OK SMA	X	X
3840	Knoxville, TN SMA	X	X	8680	Utica-Rome, NY SMA		
4000	Lancaster, PA SMA	X	X	8800	Waco, TX SMA	X	
4040	Lansing, MI SMA	X		8840	Washington, DC-MD-VA SMA	X	X
4080	Laredo, TX SMA	X		8880	Waterbury, CT SMA		
4160	Lawrence, MA SMA			8920	Waterloo, IA SMA	X	
4280	Lexington, KY SMA	X	X	9000	Wheeling-Steubenville, WV-OH SMA	X	
4320	Lima, OH SMA			9040	Wichita, KS SMA	X	X
4360	Lincoln, NE SMA	X	X	9080	Wichita Falls, TX SMA	X	X
4400	Little Rock-North Little Rock, AR SMA	X	X	9120	Wilkes-Barre–Hazleton, PA SMA	X	X
4440	Lorain-Elyria, OH SMA	X	X	9160	Wilmington, DE-NJ SMA	X	X
4480	Los Angeles, CA SMA	X	X	9220	Winston-Salem, NC	X	X
4520	Louisville, KY-IN SMA	X	X	9240	Worcester, MA SMA	X	
4560	Lowell, MA SMA			9280	York, PA SMA	X	X
4600	Lubbock, TX SMA	X	X	9320	Youngstown, OH-PA SMA	X	X

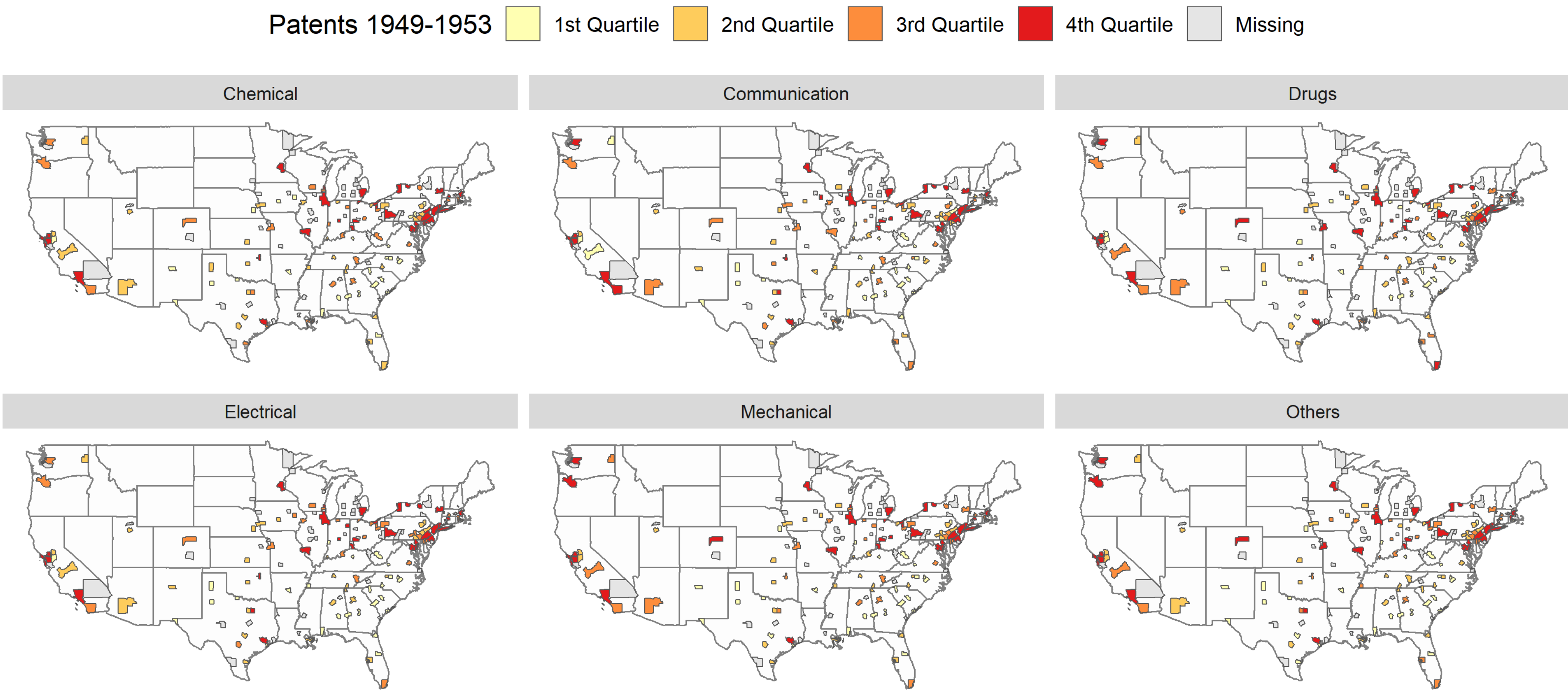


Figure 38: Geography of patenting 1951

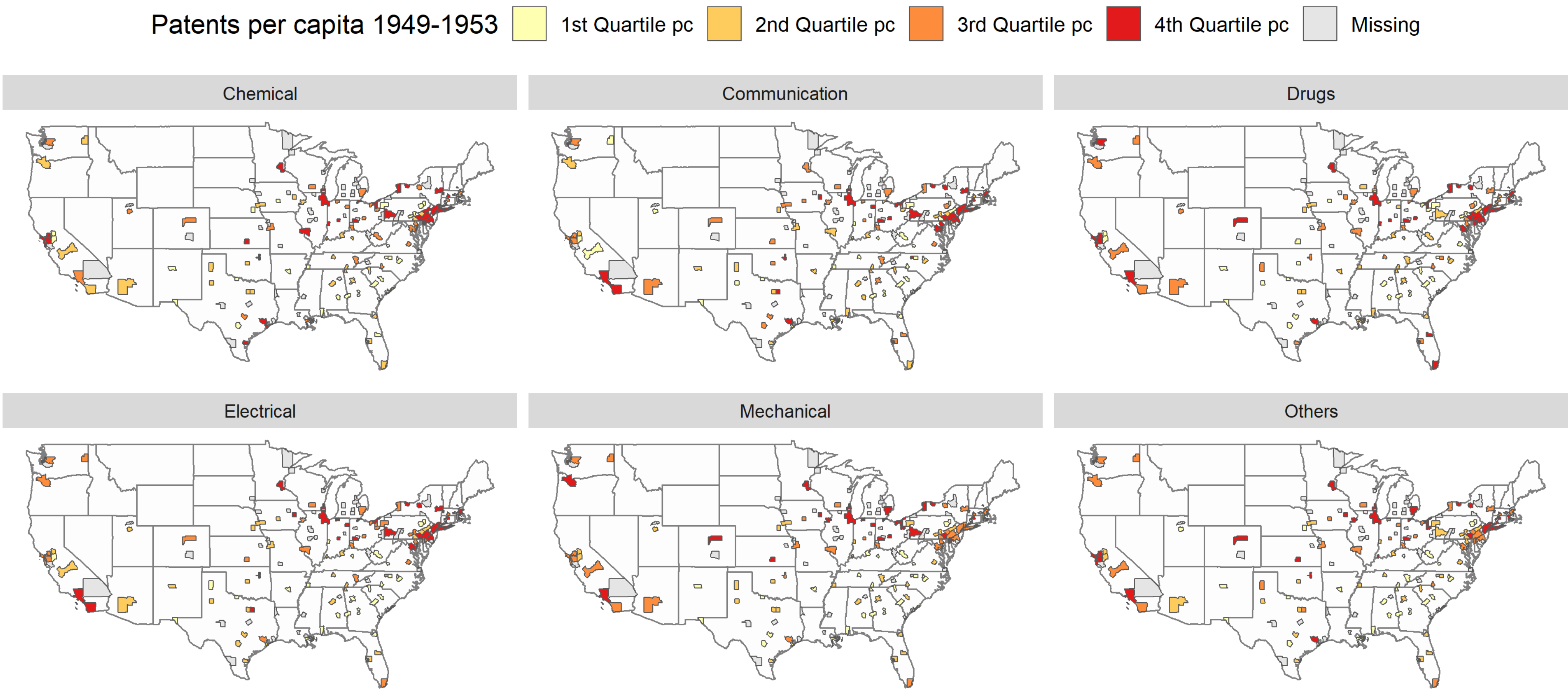


Figure 39: Patents per capita in 1951

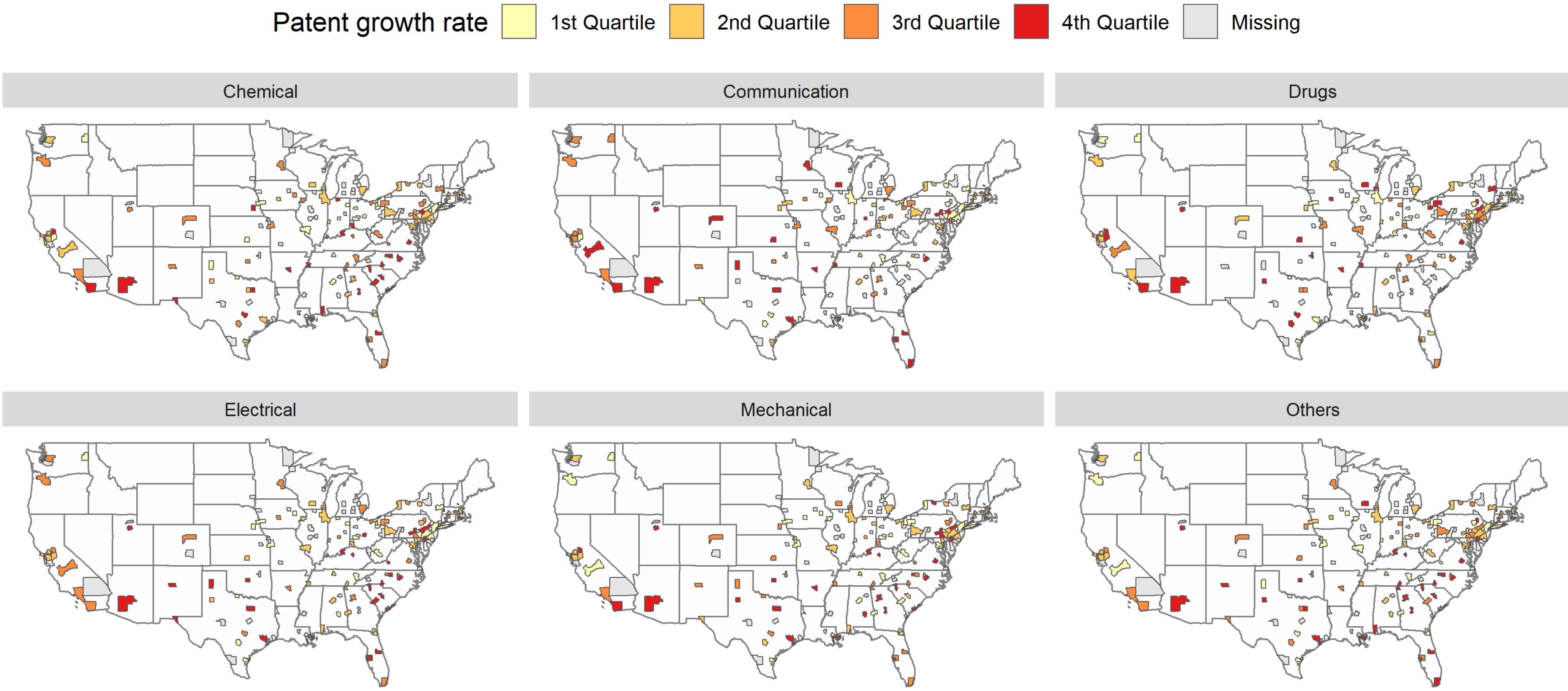


Figure 40: Patent growth rate 1951-1966

B.3.1. Descriptive statistics by technology

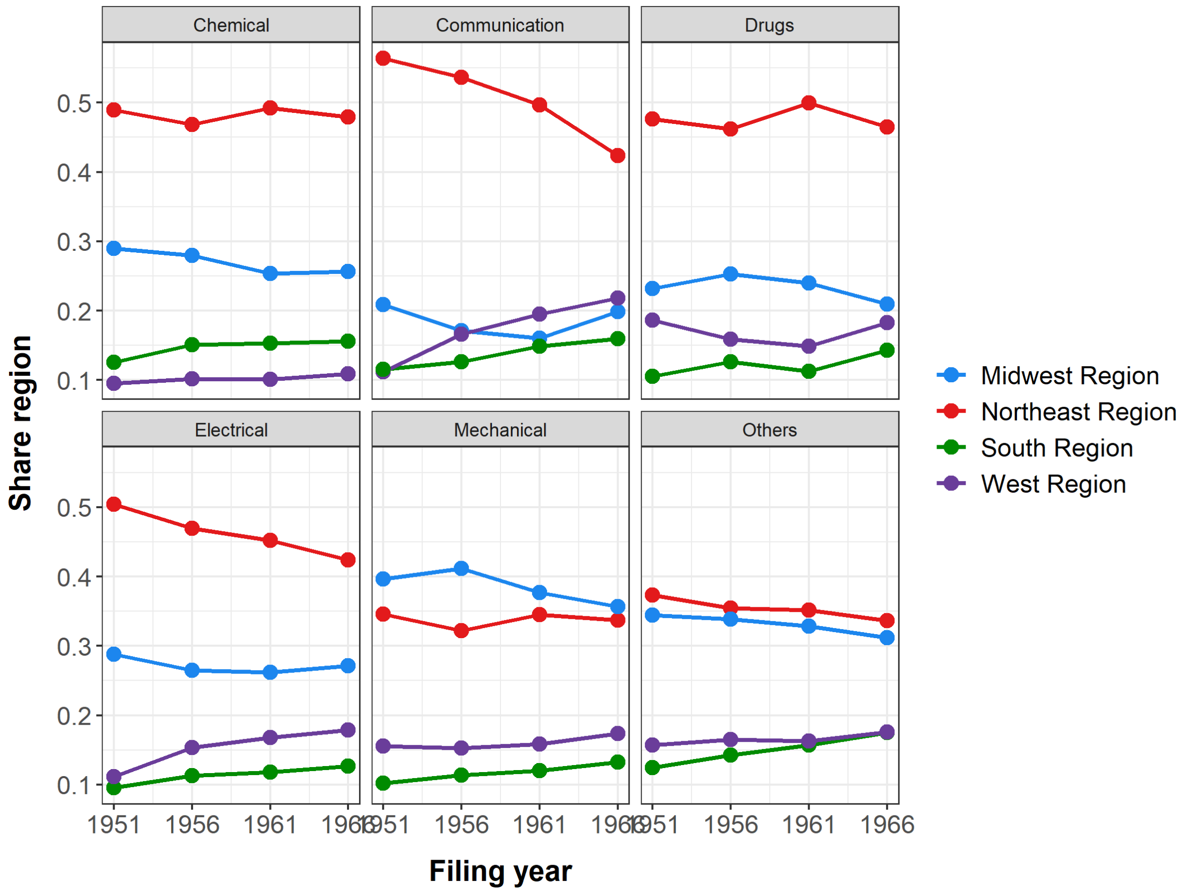


Figure 37: Share of patents by region

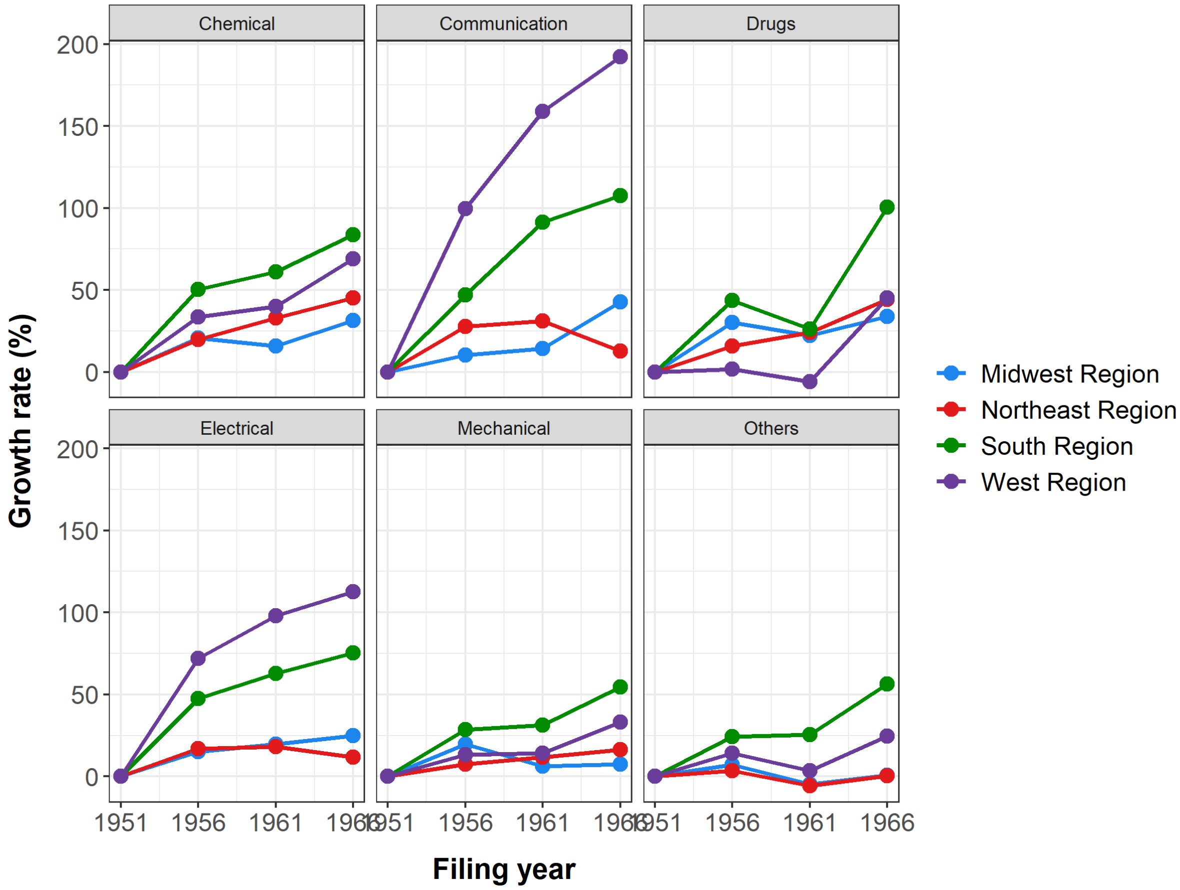


Figure 41: Patent growth rate by region

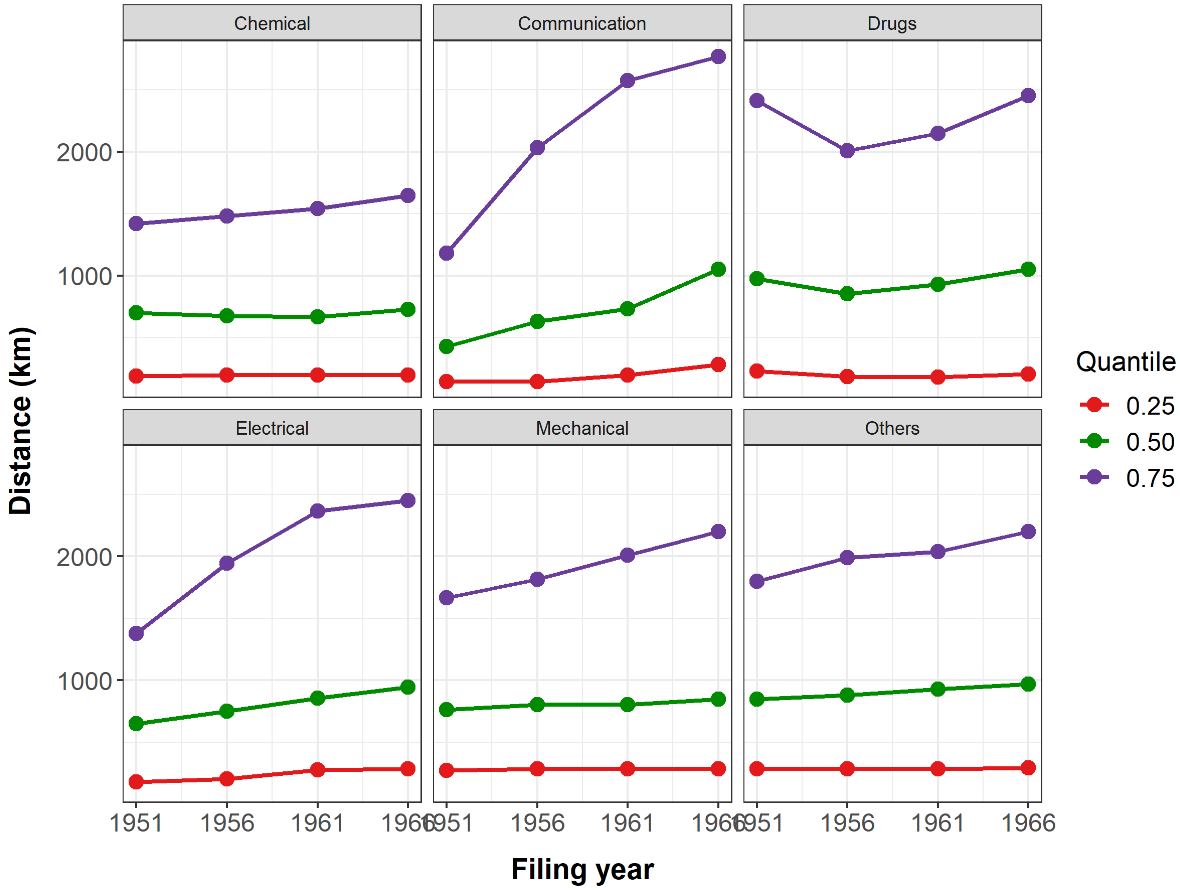


Figure 42: Quantiles of citation distance

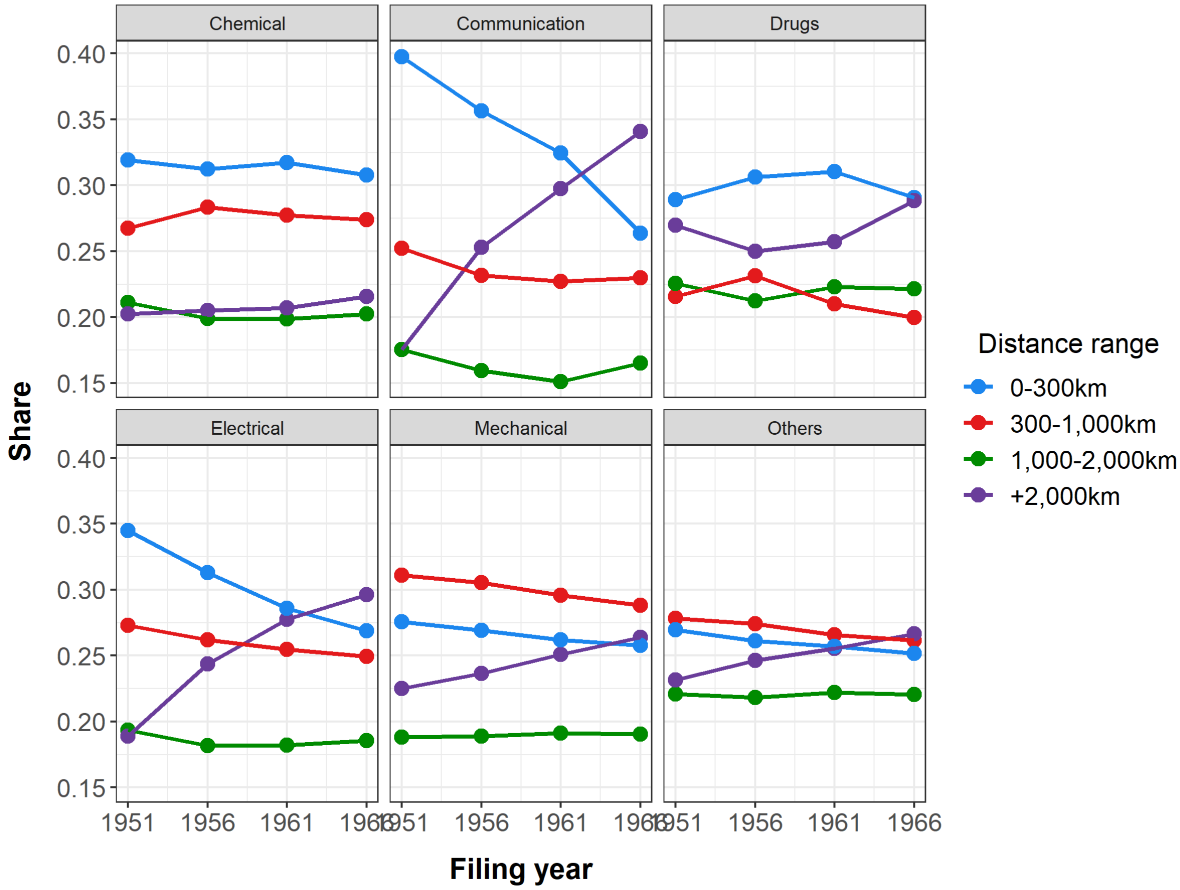


Figure 43: Share of citations by distance

C. Appendix: US Census Regions

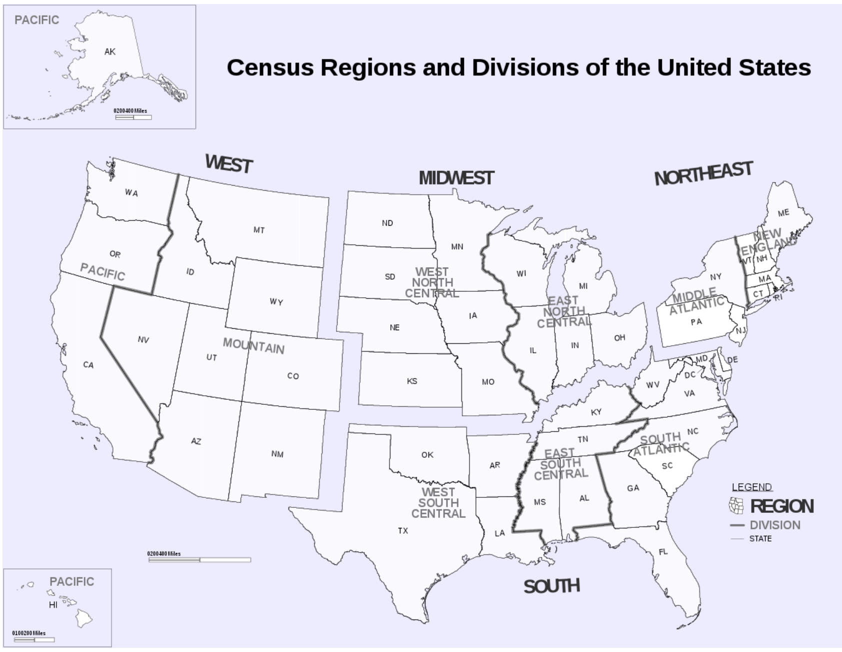


Figure 44: US Census Regions
Source: US Census Bureau

D. Appendix: Bias Correction and IV estimation

D.1. Split-panel jackknife bias correction

Weidner and Zylkin (2021) show that PPML estimation of gravity equations with three-way fixed effects (origin-time, destination-time, origin-destination) is consistent but asymptotically biased. In their words: "*the asymptotic distribution of the estimates is not centered at the truth as* $N \to \infty$" (page 2). The asymptotic bias concerns both point estimates and standard errors. In order to correct the bias we apply their suggested split-panel jackknife bias correction of section 3.4.1 to both point estimates and bootstrap standard errors. The idea of the jackknife bias correction is to estimate the model in many subsamples and then subtract the average coefficients of the subsamples from (twice) the original coefficient.

As suggested in Weidner and Zylkin (2021) when using real world data (as opposite to simulated data), we estimate the bias correction repeatedly. We modify equation (14) in Weidner and Zylkin (2021) to define the bias corrected coefficient as:

$$\tilde{\beta}_N^J := 2 \times \hat{\beta} - \frac{1}{Z} \sum_z \sum_p \frac{\hat{\beta}_{(p,z)}}{4} \tag{11}$$

where p is a random subsample of size 1/4th of the original sample, and Z is the amount of times to subsample.

The *procedure to estimate bias corrected point estimate* $\tilde{\beta}_N^J$ is as follows:

1. Estimate $\hat{\beta}$: the not-bias-corrected estimate of equation (3)
2. Randomly allocate all citing establishment-technology Fih into two equally sized groups (groups are time-invariant). Call them citing groups a and b.
3. Randomly allocate all cited establishment-technology Gjk into two equally sized groups (groups are time-invariant). Call them cited groups a and b.
4. Create four p subsamples of the original data: (a,a), (a,b), (b,a), (b,b). Subsamples keep the same granularity as the original data $FiGjhkt$.
5. Estimate equation (3) (gravity equation of the main text) in each of the subsamples from the previous step to obtain $\hat{\beta}_{(p,z)}$.[97] Store the four estimated coefficients.
6. Repeat Z times steps 2 to 5.
7. Compute equation 11

To compute bias-corrected bootstrap standard errors we need to bias-correct the point estimate $\tilde{\beta}_m^J$ of each bootstrap iteration m. The *procedure to estimate bias corrected standard errors* is as follows:

[97] Given that we require to identify the fixed effects, the *effective subsample* in all four p estimations does not have the same amount of observations. However, in our estimations the *effective subsample* size across p subsamples does not differ by more than 5%.

1. Sample establishment-technology-pairs $FiGjhk$ with replacement such that we obtain a re-sampled data of the same size as the original data (hence, some $FiGjhk$ will be repeated in the re-sampled data). Sampled $FiGjhk$ are kept for all time periods in order to keep the source of identification of β: across time variation within a establishment pair. Label this new dataset $data_m$.

2. Using $data_m$, estimate equation (3) to obtain $\hat{\beta_m}$ (this is a point estimate of the specific $data_m$)

3. Using $data_m$, repeat Z_M times steps 2 to 5 of the *procedure to estimate bias corrected point estimate*. This step provides $Z_M \times 4$ point estimates $\hat{\beta}_{(p,m,z_M)}$

4. Compute the bias corrected point estimate of bootstrap m $\tilde{\beta}^J_m = 2 \times \hat{\beta}_m - \frac{1}{Z_M} \sum_{z_M} \sum_p \frac{\hat{\beta}_{(p,m,z_M)}}{4}$.

5. Store the bias corrected point estimate of bootstrap m

6. Repeat steps 1 to 5 M times to obtain M bias corrected bootstrap point estimates $\tilde{\beta}^J_m$

7. Compute the variance-covariance matrix of bias corrected bootstrap coefficients $\tilde{\beta}^J_m$ and use it to compute standard errors of $\tilde{\beta}^J_N$

The bias correction of point estimates and bias correction of bootstrap standard errors implies estimating $Z \times 4 + Z_M \times M \times 4$ models. This is a computationally demanding task. To estimate columns (1) and (2) of Table 1 we set $Z = 100$, $Z_M = 5$ and $M = 200$, adding up to $1,100$ models to estimate for each column.

As recommended in Hansen (2021), in the Table 13 we repeat Table 1 but reporting 0.025 and 0.975 quantile values of bootstrap estimates (bias corrected for columns (1) and (2)) instead of standard errors:

	PPML		IV PPML	
Dep. variable:		*citations*		
	(1)	(2)	(3)	(4)
log(travel time)	-0.083 $(-0.129; -0.056)$		-0.152 $(-0.210; -0.097)$	
log(travel time) $\times$ 0-300km		0.019 $(-0.054; 0.082)$		-0.076 $(-0.542; 0.384)$
log(travel time) $\times$ 300-1,000km		-0.089 $(-0.141; -0.052)$		-0.134 $(-0.246; -0.066)$
log(travel time) $\times$ 1,000-2,000km		-0.094 $(-0.156; -0.022)$		-0.112 $(-0.192; -0.022)$
log(travel time) $\times$ +2,000km		-0.169 $(-0.277; -0.105)$		-0.203 $(-0.311; -0.136)$
N obs. effective	$4,703,010$	$4,703,010$	$4,703,010$	$4,703,010$
R2	0.88	0.88	0.88	0.88

Table 13: Elasticity of citations to travel time

Column (1) shows the result of Poisson Pseudo Maximum Likelihood (PPML) estimation of $citations_{FiGjhkt} = \exp\left[\boldsymbol{\beta}\, log(\text{travel time}_{ijt}) + FE_{FiGjhk} + FE_{Fiht} + FE_{Gjkt}\right] \times \varepsilon_{FiGjhkt}$, for citations of patents filed by establishment of firm F in location i, technology h and time period t, to patents filed by establishment of firm G in location j and technology k. travel time$_{ijt}$ is the travel time in minutes between location i and j at time period t, and it is set to 1 when $i = j$. When $FiGjhk$ has positive citations in at least one period and no citations in another, we attribute zero citations in the missing period. Column (2) includes the interaction of travel time$_{ijt}$ with a dummy for distance bin between the citing establishment Fi and the cited establishment Gj. Column (3) and (4) show the result of two step instrumental variables estimation, where $log(\text{travel time}_{ijt})$ is instrumented with $log(\text{travel time}_{ijt}^{\text{fix routes}})$, the travel time that would have taken place if routes were fixed to the ones observed in 1951 and in each year routes were operated with the average airplane of the year. 0.025 and 0.975 quantile bootstrap estimates are presented in parentheses. The coefficients and bootstrap estimates in columns (1) and (2) are jackknife bias-corrected. R2 is computed as the squared correlation between observed and fitted values.

D.2. Instrumental variables PPML

To implement the instrumental variables of Poisson estimation we follow the control function approach described in Wooldridge (2014). We explain the procedure using the estimation of the elasticity of citations to travel time. The procedure is similar for the elasticity of (new) patents to knowledge access. We proceed in two steps estimating the following two equations:

$$\begin{aligned} log(\text{travel time})_{FiGjhkt} =& \lambda_2 \, log(\text{instrumental travel time}_{FiGjhkt}) \\ &+ FE_{FiGjhk} + FE_{Fiht} + FE_{Gjkt} + u_{FiGjhkt} \end{aligned} \tag{12}$$

$$\begin{aligned} citations_{FiGjhkt} =& \exp\left[\beta \, log(\text{travel time}_{ijt}) + \lambda \, \hat{u}_{FiGjhkt} \right. \\ &\left. + FE_{FiGjhk} + FE_{Fiht} + FE_{Gjkt}\right] \times v_{FiGjhkt} \end{aligned} \tag{13}$$

In a first step we estimate equation (12) and obtain estimated residuals $\hat{u}_{FiGjhkt}$. In a second step we use the estimated residuals as a regressor in equation (13) which *controls* for the endogenous component of travel time.

To perform inference we bootstrap standard errors in the following way:

1. Sample establishment-technology-pairs $FiGjhk$ with replacement such that we obtain a re-sampled data of the same size as the original data (hence, some $FiGjhk$ will be repeated in the re-sampled data). Sampled $FiGjhk$ are kept for all time periods in order to keep the source of identification of β: across time variation within a establishment pair. Label this new dataset $data_m$

2. Using $data_m$, estimate equations (12) and (13) to obtain the bootstrap estimate $\hat{\beta}_m$. Store $\hat{\beta}_m$.

3. Repeat M times steps 1 and 2.

4. Compute the variance-covariance matrix of $\hat{\beta}_m$ and use it to compute standard errors of $\hat{\beta}$

For columns (3) and (4) of Table 1, and columns (3) and (4) of Table 3 we set $M = 200$.

E. Appendix: Additional results

E.1. Diffusion of knowledge

E.1.1. Heterogeneous effects

First, we investigate if the elasticity varies by the degree of concentration of patents across establishments in the citing technology or cited technology, we find no statistically significant heterogeneous effect. Results are shown in columns (1) and (2) of Table 15.

Second, we check if the elasticity varies by the median forward and backward citation lags of the cited and citing technologies. We find that the elasticity of citations to travel time is *more negative* both for technologies that accumulate citations during a longer time period and for technologies that cite older patents. To be able to precisely show if it is *newer* or *older* technologies that diffuse better as consequence of the jet requires an analysis with the citation level forward and backward lag, and not using the median lag in the technology. Nonetheless, the results seem to suggest that jets improved the diffusion of *older* technologies. Results are shown in columns (3) and (4) of Table 15.

Third, we extend the sample of patents to include patents with a patent owner identified as a government organization or university. Column (5) of Table 15 opens the elasticity of citations to travel time by whether the citing patent belongs to a government organization of university. Column (6) includes a dummy for whether the cited patent belongs to a government organization or university. We do not observe a particular change in the pattern of the elasticity of citations to travel time.

Fourth, we extend the sample to include self citations (citations in which the citing and cited patents belong to the same patent owner F). Column (7) of Table 15 shows that the elasticity is not statistically different for self citations.

Fifth, we check if the elasticity varies with the level of innovativeness of the citing firm. It may be the case that those firms that actually have the -time and monetary- budget to take a plane are only the most innovative ones. We rank firms F in technology h according to the amount of patents filed by F in technology h at the initial time period 1949-1953. We define quantile 0.00 as all those firms that did not file patents in 1949-1953, while quantile 0.01 is assigned to those that filed patents but not as many as to be in the quantile 0.25 or higher. Results are shown in Table 14. We do not find a particular pattern related to the initial innovativeness.

Sixth, we check if the elasticity varies with the citing technology, cited technology and citing-cited technology pair. Results are shown in Table 16 and Table 17. We find that the elasticity is negative and significant mainly when the citing and cited technology are the same. In Appendix B we show that most citations happen within a technology, so most identification power would be when citing and cited technologies are the same.

	Concentration citing	Concentration cited	Cited lag forward	Citing lag backward	Citing govnt & uni	Cited govnt & univ	Self citation
Dep. variable:				*citations*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log(travel time):0-300km	0.103	0.160	-0.045	0.1907	0.021	0.018	0.002
	(0.121)	(0.114)	(0.472)	(0.538)	(0.038)	(0.038)	(0.039)
log(travel time):300-1000km	-0.105	-0.039	-0.546	-0.145	-0.102^{***}	-0.099^{***}	-0.077^{***}
	(0.084)	(0.095)	(0.364)	(0.366)	(0.027)	(0.027)	(0.029)
log(travel time):1000-2000km	-0.138	-0.117	0.086	0.101	-0.094^{**}	-0.093^{**}	-0.094^{**}
	(0.105)	(0.116)	(0.480)	(0.498)	(0.042)	(0.041)	(0.040)
log(travel time):+2000km	-0.287^{***}	-0.268^{***}	0.720^{**}	0.560	-0.185^{***}	-0.188^{***}	-0.153^{***}
	(0.105)	(0.090)	(0.344)	(0.472)	(0.049)	(0.048)	(0.040)
log(travel time):0-300km $\times$ X	-1.180	-2.013	0.028	-0.066	-0.125	0.481	0.038
	(1.843)	(1.712)	(0.185)	(0.211)	(0.367)	(0.543)	(0.252)
log(travel time):300-1000km $\times$ X	0.079	-0.880	0.178	0.018	-0.088	-0.609^{*}	0.077
	(1.188)	(1.366)	(0.144)	(0.145)	(0.265)	(0.330)	(0.127)
log(travel time):1000-2000km $\times$ X	0.634	0.341	-0.073	-0.078	-0.282	-0.370	0.082
	(1.412)	(1.606)	(0.191)	(0.197)	(0.366)	(0.385)	(0.210)
log(travel time):+2000km $\times$ X	1.436	1.157	-0.366^{***}	-0.299	-0.328	0.015	-0.073
	(1.456)	(1.136)	(0.137)	(0.188)	(0.410)	(0.295)	(0.170)
N obs. effective	$4,703,010$	$4,703,010$	$4,703,010$	$4,703,010$	$4,800,144$	$4,800,144$	$4,835,001$
R2	0.88	0.88	0.88	0.88	0.88	0.88	0.94

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$

Table 15: Elasticity of citations to travel time: Heterogeneity (part 1)

Result of Poisson Pseudo Maximum Likelihood (PPML) estimation of $citations_{FiGjhkt} = \exp\left[\sum_d \beta_d \mathbb{1}\{distance_{ij} \in d\}\, log(\text{travel time}_{ijt}) + \sum_d \alpha_d \mathbb{1}\{distance_{ij} \in d\}\, \mathbb{1}\{X_{FiGjhkt}\}\, log(\text{travel time}_{ijt}) + FE_{FiGjhk} + FE_{Fiht} + FE_{Gjkt}\right] \times \varepsilon_{FiGjhkt}$, for citations of patents filed by establishment of firm F in location i, technology h and time period t, to patents filed by establishment of firm G in location j and technology k. travel time$_{ijt}$ is the travel time in minutes between location i and j at time period t, and it is set to 1 when $i = j$. d are distance intervals: $[0 - 300km]$, $(300km - 1000km]$, $(1000km - 2000km]$, $(2000km - max]$. The variable X takes different value depending on the column: in column (1) it is the across-MSA Herfindahl index of the citing technology, in column (2) it is the across-MSA Herfindahl index of the cited technology, in column (3) it is median forward citation lag of the cited technology, in column (4) it is median backward citation lag of the citing technology. In column (5) and (6) the sample includes government and university patents, in column (5) X is a dummy that takes value one if the citing patent belongs to a university or government organisation, in column (6) it is a dummy that takes value one if the cited patent belongs to a university or government organisation. In column (7) the sample includes self citations, the variable X is a dummy that takes value one if the citing firm F cited firm G are the same. When $FiGjhk$ has positive citations in at least one period and no citations in another, we attribute zero citations in the missing period. Standard errors clustered at the non-directional location pair are presented in parenthesis (ij is the same non-directional location pair as ji). R2 is computed as the squared correlation between observed and fitted values.

	Citing quantile	Cited quantile
Dep. variable:	*citations*	
	(1)	(2)
log(travel time) × quantile 0.00	-0.151*** (0.058)	-0.111*** (0.039)
log(travel time) × quantile 0.01	-0.078 (0.114)	-0.084 (0.101)
log(travel time) × quantile 0.25	-0.081 (0.103)	-0.159* (0.093)
log(travel time) × quantile 0.50	-0.139 (0.091)	-0.063 (0.083)
log(travel time) × quantile 0.75	-0.262*** (0.079)	-0.033 (0.068)
log(travel time) × quantile 0.90	-0.029 (0.066)	-0.127** (0.057)
log(travel time) × quantile 0.95	-0.001 (0.037)	-0.123*** (0.038)
log(travel time) × quantile 0.99	-0.130*** (0.035)	-0.066* (0.039)
log(travel time) × quantile 0.999	-0.070 (0.045)	-0.070 (0.045)
N obs. effective	$4,703,010$	$4,703,010$
R2	0.88	0.88

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$

Table 14: Elasticity of citations to travel time: Heterogeneity (part 2)

Column (1) shows the result of Poisson Pseudo Maximum Likelihood (PPML) estimation of $citations_{FiGjhkt} = \exp\left[\sum_q \boldsymbol{\beta}_q \, log(\text{travel time}_{ijt}) \, \mathbb{1}\{quantile_{Fh} \in q\} \, + FE_{FiGjhk} + FE_{Fiht} + FE_{Gjkt}\right] \times \varepsilon_{FiGjhkt}$, for citations of patents filed by establishment of firm F in location i, technology h and time period t, to patents filed by establishment of firm G in location j and technology k. travel time$_{ijt}$ is the travel time in minutes between location i and j at time period t, and it is set to 1 when $i = j$. $quantile_{Fh}$ is the quantile of firm F in the distribution of firms within technology h, using patents applied by F in h in the time period 1949-1953. Column (2) repeats the analysis using the quantile of the cited firm G in technology k. When $FiGjhk$ has positive citations in at least one period and no citations in another, we attribute zero citations in the missing period. When $FiGjhk$ has positive citations in at least one period and no citations in another, we attribute zero citations in the missing period. Standard errors clustered at the non-directional location in parentheses (ij is the same non-directional location pair as ji). R2 is computed as the squared correlation between observed and fitted values.

	PPML	
	Citing technology	Cited technology
Dep. variable:	*citations*	
	(1)	(2)
log(travel time) × Chemical	-0.066	-0.093^{**}
	(0.045)	(0.045)
log(travel time) × Computers & Communications	-0.100	-0.140^{*}
	(0.079)	(0.077)
log(travel time) × Drugs & Medical	-0.053	-0.005
	(0.162)	(0.181)
log(travel time) × Electrical & Electronic	-0.070	-0.054
	(0.048)	(0.046)
log(travel time) × Mechanical	-0.080^{**}	-0.087^{***}
	(0.031)	(0.032)
log(travel time) × Others	-0.147^{***}	-0.113^{**}
	(0.045)	(0.044)
N obs. effective	4,703,010	4,703,010
R2	0.88	0.88

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$

Table 16: Elasticity of citations to travel time by citing and cited technology
Part 1

Column (1) shows the result of Poisson Pseudo Maximum Likelihood (PPML) estimation of $citations_{FiGjhkt} = \exp\left[\sum_{tech} \beta_h \mathbb{1}\{tech = h\} \times log(\text{travel time}_{ijt}) + FE_{FiGjhk} + FE_{Fiht} + FE_{Gjkt}\right] \times \varepsilon_{FiGjhkt}$, for citations of patents filed by establishment of firm F in location i, technology h and time period t, to patents filed by establishment of firm G located in j, in technology k. $\mathbb{1}\{tech = h\}$ is a dummy variable that takes value 1 when the citing technology h is equal to technology $tech$. In column (2) the dummy is modified to $\mathbb{1}\{tech = k\}$ such that it takes value 1 when the cited technology k is equal to technology $tech$. travel time$_{ijt}$ is the travel time in minutes between location i and j at time period t, and it is set to 1 when $i = j$. When $FiGjhk$ has positive citations in at least one period and no citations in another, we attribute zero citations in the missing period. Standard errors clustered at the non-directional location pair are presented in parenthesis (ij is the same non-directional location pair as ji). R2 is computed as the squared correlation between observed and fitted values.

Cited \ Citing	Chemical	Computers & Communications	Drugs & Medical	Electrical & Electronic	Mechanical	Others
Chemical	-0.092^{**} (0.052)	0.219 (0.262)	0.113 (0.199)	-0.299^{***} (0.094)	-0.025 (0.071)	-0.070 (0.068)
Computers & Communications	-0.089 (0.259)	-0.306^{***} (0.095)	-0.657 (0.976)	0.107 (0.090)	0.122 (0.149)	0.095 (0.169)
Drugs & Medical	0.224 (0.239)	0.567 (1.205)	-0.278 (0.268)	-0.230 (0.561)	-0.334 (0.362)	0.358 (0.323)
Electrical & Electronic	0.233^{**} (0.093)	0.171^{*} (0.096)	-0.224 (0.634)	-0.102^{**} (0.056)	0.087 (0.070)	-0.063 (0.079)
Mechanical	-0.060 (0.076)	0.151 (0.145)	-0.152 (0.402)	0.106 (0.082)	-0.129^{***} (0.035)	-0.032 (0.056)
Others	0.042 (0.074)	0.173 (0.169)	0.204 (0.274)	0.052 (0.072)	0.019 (0.053)	-0.209^{***} (0.054)

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$

Table 17: Elasticity of citations to travel time by citing and cited technology
Part 2

Column (1) shows the result of one single Poisson Pseudo Maximum Likelihood (PPML) estimation of $citations_{FiGjhkt} = \exp\left[\sum_{\text{tech pair}} \beta_{hk}\, \mathbb{1}\{\text{tech pair} = hk\} \times log(\text{travel time}_{ijt}) + FE_{FiGjhk} + FE_{Fiht} + FE_{Gjkt}\right] \times \varepsilon_{FiGjhkt}$, for citations of patents filed by establishment of firm F in location i, technology h and time period t, to patents filed by establishment of firm G located in j, in technology k. $\mathbb{1}\{\text{tech pair} = hk\}$ is a dummy variable that takes value 1 when the citing technology h is equal to technology $tech$. In column (2) the dummy is modified to $\mathbb{1}\{tech = k\}$ such that it takes value 1 when the cited technology k is equal to technology $tech$. travel time$_{ijt}$ is the travel time in minutes between location i and j at time period t, and it is set to 1 when $i = j$. When $FiGjhk$ has positive citations in at least one period and no citations in another, we attribute zero citations in the missing period. Standard errors clustered at the non-directional location pair are presented in parenthesis (ij is the same non-directional location pair as ji). R2 is computed as the squared correlation between observed and fitted values. The amount of observation in the effective sample is 4,703,010.

E.1.2. IV PPML: first and second stage estimation

	First stage OLS	Second stage PPML
Dep. variable:	log(travel time)	*citations*
	(1)	(2)
log(travel time fix routes)	0.951^{***}	
	(0.039)	
log(travel time)		-0.152^{***}
		(0.029)
residual		0.094^{***}
		(0.035)
N obs. effective	$10,106,940$	$4,703,010$
R2	0.99	0.88
Within R2	0.38	

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$

Table 18: Elasticity of citations to travel time: first and second stage IV PPML
The table presents the results of 2-step instrumental variables estimation of $citations_{FiGjhkt} = \exp\left[\beta\, log(\text{travel time}_{ijt}) + FE_{FiGjhk} + FE_{Fiht} + FE_{Gjkt}\right] \times \varepsilon_{FiGjhkt}$, where $log(\text{travel time}_{ijt})$ is instrumented with $log(\text{instrumental travel time}_{ijt})$. Column (1) shows the results of the first stage regression estimated by OLS. Column (2) shows the result of the second stage regression estimated by Poisson Pseudo Maximum Likelihood , including the estimated residuals of the first stage as controls. The number of observations in the second stage estimation is smaller due to not being able to identify fixed effects that are required in PPML estimation.

	OLS First stage 0-300km	OLS First stage 300-1,000km	OLS First stage 1,000-2,000km	OLS First stage +2,000km	Second stage PPML
Dep. variable:			log(travel time)		*citations*
	(1)	(2)	(3)	(4)	(5)
log(travel time fix routes) × 0-300km	0.278^{**} (0.122)	0.073 (0.057)	0.024 (0.026)	0.040^{*} (0.022)	
log(travel time fix routes) × 300-1,000km	-0.103^{***} (0.032)	1.113^{***} (0.041)	-0.013 (0.011)	0.010 (0.011)	
log(travel time fix routes) × 1,000-2,000km	-0.064^{***} (0.024)	-0.052^{***} (0.020)	1.059^{***} (0.044)	0.017^{*} (0.009)	
log(travel time fix routes) × +2,000km	-0.058^{***} (0.022)	-0.046^{***} (0.017)	-0.020^{**} (0.010)	1.097^{***} (0.018)	
log(travel time) × 0-300km					-0.076 (0.221)
log(travel time) × 300-1,000km					-0.134^{***} (0.044)
log(travel time) × 1,000-2,000km					-0.112^{**} (0.047)
log(travel time) × +2,000km					-0.203^{***} (0.043)
residual × 0-300km					0.100 (0.196)
residual × 300-1,000km					0.045 (0.053)
residual × 1,000-2,000km					0.026 (0.069)
residual × +2,000km					0.043 (0.078)
N obs. effective	10,106,940	10,106,940	10,106,940	10,106,940	4,703,010
R2	0.99	0.99	0.99	0.99	0.88
Within R2	0.04	0.46	0.80	0.88	

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$

Table 19: Elasticity of citations to travel time: first and second stage IV PPML

The table presents the results of 2-step instrumental variables estimation of Poisson Pseudo Maximum Likelihood of $citations_{FiGjhkt} = \exp\left[\sum_d \beta_d \times \mathbb{1}\{distance_{ij} \in d\} \times log(\text{travel time}_{ijt}) + FE_{FiGjhk} + FE_{Fiht} + FE_{Gjkt}\right] \times \varepsilon_{FiGjhkt}$, where $\mathbb{1}\{distance_{ij} \in d\} \times log(\text{travel time}_{ijt})$ is instrumented with $\mathbb{1}\{distance_{ij} \in d\} \times log(\text{instrumental travel time}_{ijt})$. Given that there are 4 distance segments d there are 4 first stages. Columns (1) to (4) show the results of the first stage regressions which are estimated by OLS. Coefficients of the 4 interactions of the instrument can be identified due to the presence of the fixed effects, e.g. after demeaning by fixed effects there is residual variation that allows to identify the 4 coefficients in each regression of the first stage. Column (5) shows the result of the second stage regression estimated by PPML, including the estimated residuals of the first stage as controls. The number of observations in the second stage estimation is smaller due to not being able to identify fixed effects that are required in PPML estimation.

E.1.3. Robustness

Sample of establishments

During the time period there was entry and exit of research establishments that was not uniform across locations. We may then think that the change in diffusion of knowledge is only consequence of the change in the geographical location of innovation. To test this possibility, in Table **??** we estimate the baseline regression 3 with different samples. In column (1) we include the baseline results.[98] In column (2) we use only citing establishments Fi that filed patents during the initial time period 1949-1953. In column (3) we further restrict the sample to both citing establishments Fi and cited establishments Gj that filed patents in 1949-1953.[99] We find that the coefficient at more than 2,000km remains comparable to the one in the baseline regression, statistically significant at the 1%.

Estimation of log-log gravity equation

We modify equation 3 to have a log-log version:

$$log(citations_{FiGjhkt}) = \kappa\, log(\text{travel time}_{ijt}) + FE_{FiGjhk} + FE_{Fiht} + FE_{Gjkt} + \nu_{FiGjhkt} \quad (14)$$

Results by OLS estimation are presented in Table 20.

[98] Coefficients are not bias corrected.

[99] We require Fi and Gj to have positive amount of patents applied during 1949-1953. However, those establishments need not to have cited each other.

	PPML		OLS	
Dep. variable:	*citations*		$log(citations)$	
	(1)	(2)	(3)	(4)
log(travel time)	-0.083^{***} (0.019)		-0.052 (0.040)	
log(travel time) $\times$ 0-300 km		0.019 (0.036)		0.063 (0.069)
log(travel time) $\times$ 300-1,000 km		-0.089^{***} (0.023)		-0.072 (0.046)
log(travel time) $\times$ 1,000-2,000 km		-0.094^{***} (0.033)		-0.097 (0.072)
log(travel time) $\times$ +2,000 km		-0.169^{***} (0.039)		-0.161* (0.084)
N obs. effective	$4,703,010$	$4,703,010$	2,643,024	2,643,024
R2	0.88	0.88	0.99	0.99

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$

Table 20: Elasticity of citations to travel time: PPML and OLS

Column (1) shows the result of Poisson Pseudo Maximum Likelihood (PPML) estimation of $citations_{FiGjhkt} = \exp\left[\boldsymbol{\beta}\, log(\text{travel time}_{ijt}) + FE_{FiGjhk} + FE_{Fiht} + FE_{Gjkt}\right] \times \varepsilon_{FiGjhkt}$, for citations of patents filed by establishment of firm F in location i, technology h and time period t, to patents filed by establishment of firm G in location j and technology k. travel time$_{ijt}$ is the travel time in minutes between location i and j at time period t, and it is set to 1 when $i = j$. When $FiGjhk$ has positive citations in at least one period and no citations in another, we attribute zero citations in the missing period. Column (2) includes the interaction of travel time$_{ijt}$ with a dummy for distance bin between the citing establishment Fi and the cited establishment Gj. Column (3) and (4) shows the result of OLS estimation of $log(citations_{FiGjhkt}) = \kappa\, log(\text{travel time}_{ijt}) + FE_{FiGjhk} + FE_{Fiht} + FE_{Gjkt} + \nu_{FiGjhkt}$. The coefficients and standard errors in columns (1) and (2) are jackknife bias-corrected. In columns (3) and (4) standard errors are clustered at the MSA-pair. R2 is computed as the squared correlation between observed and fitted values.

Ticket prices

During the period of analysis ticket prices were set by the Civil Aeronautics Board, so airlines could not set prices of their own tickets. Some airlines included a sample of prices in the last page of their booklet of flight schedules a sample of prices, which we digitized. We have digitized American Airlines 1951, 1961, 1966; TWA 1951 and United Airlines 1956 and 1961.[100]. The sample includes prices for 11,590 directional airport pair years. We document multiple facts about prices.

First, prices were set in the form of an intercept plus a variable increment depending on distance between origin and destination (until 1962-1963). A linear regression with an intercept and a slope estimated separately for each year (including 1966), service class (first class or coach service), and aircraft type (propeller or jet) gives a R2 of 0.98 or higher in each regression, with an average R2 of 0.993.

Second, all airlines operating within the same route charged exactly the same price. In 1951, in our digitized price data we have 432 airport pairs in which both American Airlines and TWA were operating and reported the price for first class service. 94% of those airport pairs had exactly the same price in both airlines.

Third, ticket prices of flights operated by jet airplanes had a surcharge of around 6% on top of the one operated by propeller airplanes.

Fourth, the change in prices over time had a similar pattern until 1961: a stronger increase in short distances (probably due to an increase in fixed costs of take-off and landing, although not reflected in the intercept of the linear regressions), and a relatively constant increase for flights between airports more than 1,000 km apart. In the period 1961 to 1966 we observe a drop in prices of around 20% for routes of more than 1,000km distance, breaking the linearity of prices in distance previously observed. We had vi-

[100]The sample of prices digitized was limited due to data availability.

sually inspected price tables and detected that the drop in prices happened in 1962-1963.

Figure 45 shows prices for first class service by year and aircraft type, deflated by the consumer price index to 1951 values. Figure 46 presents the percentage change in deflated prices of first class service. Both figures show the previous facts: prices are generally linear in distance until 1966 in which we observe a break after 1,000 km.

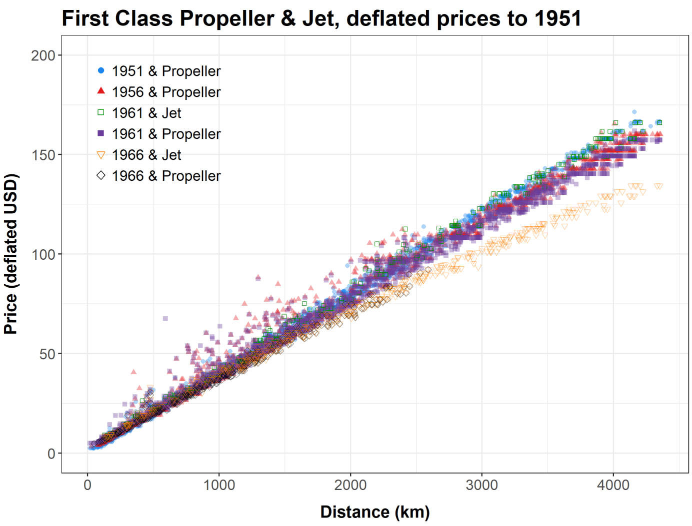


Figure 45: Flight ticket prices, deflated by CPI

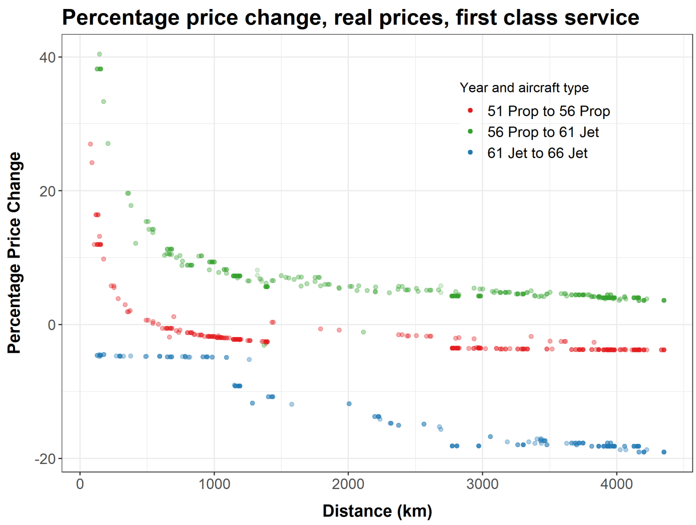


Figure 46: Change flight ticket prices, deflated by CPI

We convert our sample of prices at the airport-pair level to prices of the population of MSA-pairs as follows: first, we obtain a pricing function that can flexibly approximate prices by regressing deflated prices on a cubic polynomial of distance separately for each year. We use prices of first class service for all years, propeller aircraft for 1951 and 1956 and jet aircraft for 1961 and 1966. Second, we predict prices for each MSA-pair and year using the MSA-pair distance and the year's estimated regression.

Highway travel time

Taylor Jaworski and Carl Kitchens have graciously shared with us data on county-to-county highway travel time and nominal travel costs for 1950, 1960 and 1970. Travel time is constructed using maximum speed limit in each highway segment and year. Travel costs uses, for each year, travel time, highway distance, truck driver's wage and petrol costs. See Jaworski and Kitchens (2019) for details. The dataset is constructed using 2010 county boundaries and contains county centroids. We converted it to MSA-to-MSA by matching counties' centroids to 1950 MSAs using the shape file from Manson et al. (2020). We take the minimum travel time and minimum travel costs among all county pairs that belong to the same MSA pair. We convert nominal travel costs to 1950 real travel costs deflating by the consumer price index. We convert 1950, 1960 and 1970 travel times and travel costs to 1951, 1956, 1961 and 1966 by linearly interpolating (e.g. $\text{travel time}_{ij,1951} = \text{travel time}_{ij,1950} \times \frac{1960-1951}{10} + \text{travel time}_{ij,1960} \times \frac{1951-1950}{10}$).

The within MSA-pair correlation of the 1951-1966 change in travel time by highway and airplane is 0.068 for all MSA-pairs, and -0.011 for MSA-pairs more than 2,000 km apart. Figure 47 presents the MSA-pair 1951-1966 change in travel time by highway and airplane, where for exposition we only present MSA-pairs that had a reduction in travel time by both means of transport. Estimating a linear regression of change in air travel time on the change in highway travel time gives a slope of -0.02 not statistically different from zero, with a R2 of 0.00005.[101] Figure 48 repeats the exercise where MSA-pairs are weighted by the amount of establishment-technology pairs used to estimate the elasticity of citations to travel time (equation (3)). In this case the estimated regression has a slope of 0.73 statistically significant at the 1% level and a R2 of 0.09.[102]

In Tables 4 and 21 we present the results of adding highway travel time as control. The low correlation between the change in travel time by highway and airplane implies that the estimated elasticity of citations to air travel time remains almost unchanged,

[101] 8.7% of MSA-pairs had an increase in travel time either by highway or by airplane. The regression with all MSA-pairs has a slope of 0.60 significant at the 1% level. However, the R2 of the regression remains very low: 0.0046.

[102] With all MSA-pairs the slope is 1.01 statistically significant at the 1% level and the R2 is 0.04.

relative to the baseline estimation.[103]

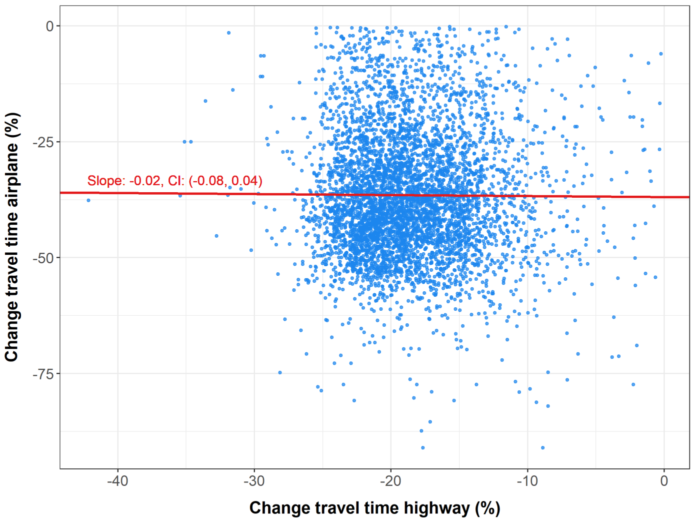


Figure 47: Change travel time by airplane and highway 1951-1966

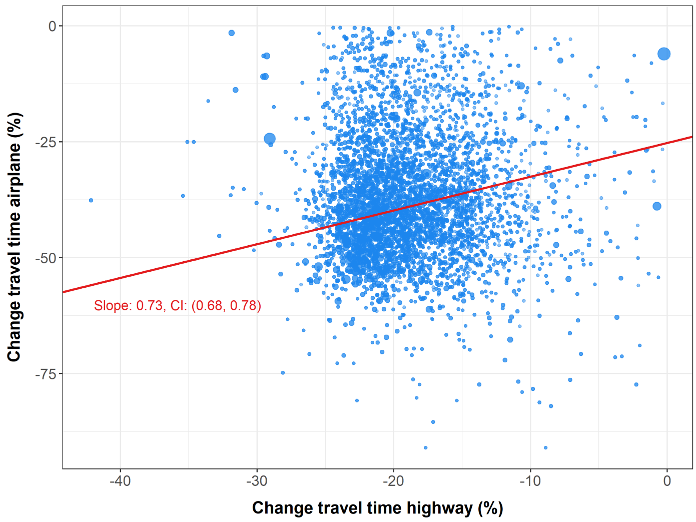


Figure 48: Change travel time by airplane and highway 1951-1966, weighted

[103]In order to perform a test of statistical difference of coefficients we would need to compute the covariance between the two regressions. Assuming the covariance is zero, in columns (1) and (2) 21 the coefficients of air travel time at +2,000km are not significantly different.

	PPML							
Dep. variable:	*citations*							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(travel time) × 0-300km	0.0213 (0.0388)	0.0276 (0.0385)	0.0198 (0.0391)	0.0318 (0.0393)	0.0252 (0.0389)	0.0349 (0.0391)	0.0283 (0.0396)	0.0313 (0.0393)
log(travel time) × 300-1,000km	-0.0990*** (0.0269)	-0.1040*** (0.0292)	-0.0935*** (0.0265)	-0.0745** (0.0303)	-0.1014*** (0.0290)	-0.0857*** (0.0312)	-0.0748** (0.0303)	-0.0861*** (0.0312)
log(travel time) × 1000-2,000km	-0.0928** (0.0418)	-0.1155** (0.0485)	-0.0710* (0.0423)	-0.0395 (0.0523)	-0.0948* (0.0502)	-0.0498 (0.0573)	-0.0318 (0.0520)	-0.0435 (0.0576)
log(travel time) × +2,000km	-0.1848*** (0.0492)	-0.1761*** (0.0531)	-0.1724*** (0.0498)	-0.1238** (0.0587)	-0.1658*** (0.0542)	-0.1052* (0.0607)	-0.1236** (0.0590)	-0.1041* (0.0609)
N obs. effective	4,703,010	4,703,010	4,703,010	4,703,010	4,703,010	4,703,010	4,703,010	4,703,010
R2	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88
Controls:								
log(highway time)	-	Yes	-	-	Yes	Yes	-	Yes
log(telephone share) × time	-	-	Yes	-	Yes	-	Yes	Yes
log(distance) × time	-	-	-	Yes	-	Yes	Yes	Yes

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$

Table 21: Elasticity of citations to travel time: additional controls

Column (1) shows the result of Poisson Pseudo Maximum Likelihood (PPML) estimation of $citations_{FiGjhkt} = \exp\left[\sum_d \beta_d \mathbb{1}\{distance_{ij} \in d\}\, log(\text{travel time}_{ijt}) + \sum_d \alpha_d \mathbb{1}\{distance_{ij} \in d\} \mathbb{1}\{X_{FiGjhkt}\}\, log(\text{travel time}_{ijt}) + FE_{FiGjhk} + FE_{Fiht} + FE_{Gjkt}\right] \times \varepsilon_{FiGjhkt}$, for citations of patents filed by establishment of firm F in location i, technology h and time period t, to patents filed by establishment of firm G in location j and technology k. travel time$_{ijt}$ is the travel time in minutes between location i and j at time period t, and it is set to 1 when $i = j$. d are distance intervals: $[0 - 300km], (300km - 1000km], (1000km - 2000km], (2000km - max]$. Relative to (1), columns (2) to (8) contain additional controls. Log highway time between i and j changes in every time period t. The log mean share of households with telephone line in ij pair interacted in 1960 is interacted with a time dummy. Log distance ij is interacted with a time dummy. When $FiGjhk$ has positive citations in at least one period and no citations in another, we attribute zero citations in the missing period. Standard errors clustered at the non-directional location in parentheses (ij is the same non-directional location pair as ji). R2 is computed as the squared correlation between observed and fitted values.

Frequency adjusted travel time

The frequency of flights may have changed simultaneously with the introduction of jet airplanes. The change in travel time could then be consequence of higher frequency rather than changes in airplanes' speed. Given that some MSA pairs are connected indirectly (with connecting flights), accounting for frequency is not straight forward: the frequency of each leg of the flight route matters (actually, it is not only frequency of each leg but also the synchronization among all potential legs). In order to take into account potential changes in the frequency of flights we computed the daily average travel time. This travel time is the average across all fastest travel times if the passenger was to depart at each full hour (1am, 2am, ..., 1pm, 2pm, etc.). The computation of this travel time includes the waiting time that is affected by frequency: the time until first departure and layover time of each connecting flight. Hence, the daily average travel time is a frequency-adjusted travel time: changes in the daily average travel time that are larger than in the fastest travel time denote that frequency of flights increased and therefore there is less waiting time. If we observe the reverse that means that frequency did not improve as much as the speed of airplanes.

Figure 49 shows the within MSA-pair decrease in the fastest travel time and the daily average travel time.[104] Both measures of travel time follow a similar pattern: slight decrease in 1956, a stronger decrease in 1961 especially for long distance routes, and a further decline in 1966. However, we observe that the decrease of the fastest travel time is on average larger than the one of the daily average travel time: the frequency of flights, if any, attenuated the potential decrease in travel time from the improvements in airplanes' speed. This observation is also in line with a comparison of the fastest travel time with and without layover time (Figure 28 in the Appendix of the paper): layover time attenuated the change in travel time.

In table 22 we estimated the elasticity of citations to travel time using first the fastest travel time (baseline, columns 1 and 2) and the daily average travel time (columns 3 and 4). The estimated elasticity is similar using both measures, which gives confidence that our results are not driven by changes in the frequency of flights.

[104] The within MSA-pair correlation of the (1951-1966) change in fastest travel time and the change daily average travel time is 0.60.

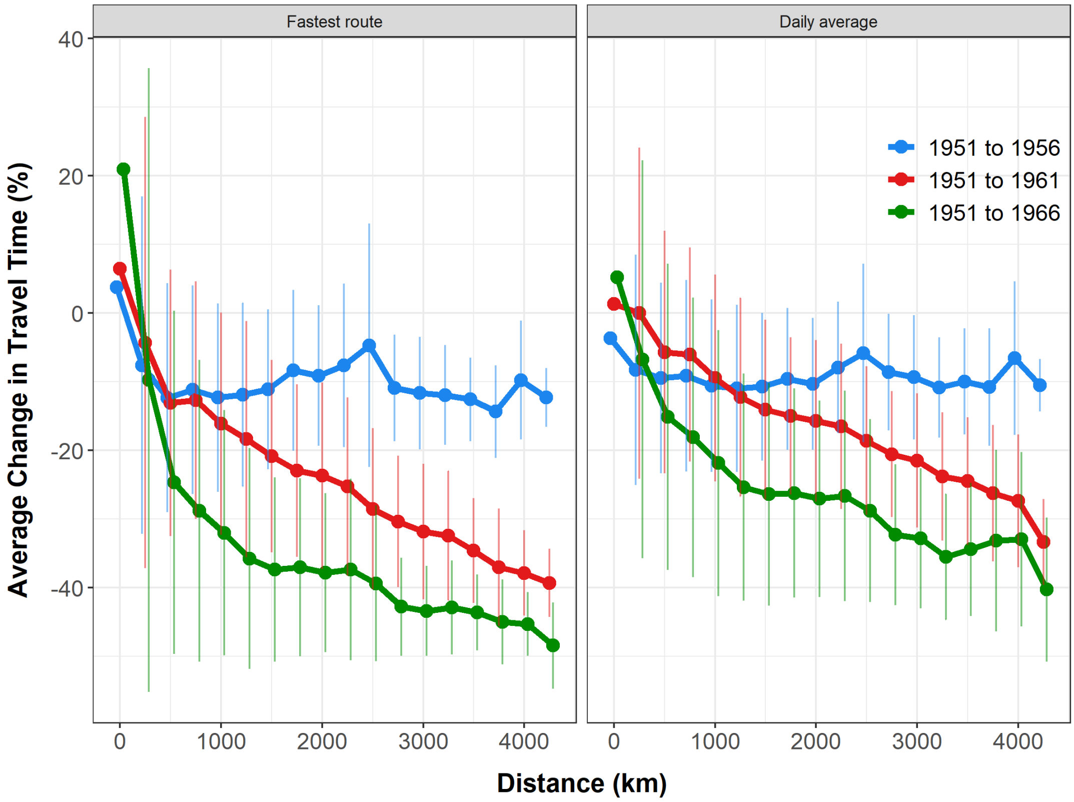


Figure 49: Change in MSAs travel time: fastest travel time and daily average travel time

	PPML not bias-corrected			
Dep. variable:	*citations*			
	(1)	(2)	(3)	(4)
log(travel time)	-0.088^{***} (0.024)			
log(travel time) × 0-300km		0.021 (0.039)		
log(travel time) × 300-1,000km		-0.099^{**} (0.027)		
log(travel time) × 1000-2,000km		-0.093^{**} (0.042)		
log(travel time) × +2,000km		-0.185^{***} (0.049)		
log(travel time daily avg)			-0.100^{***} (0.039)	
log(travel time daily avg) × 0-300km				0.034 (0.037)
log(travel time daily avg) × 300-1,000km				-0.142^{***} (0.047)
log(travel time daily avg) × 1000-2,000km				-0.170^{***} (0.072)
log(travel time daily avg) × +2,000km				-0.236^{***} (0.064)
N obs. effective	4,703,010	4,703,010	4,703,010	4,703,010
R2	0.88	0.88	0.88	0.88

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$

Table 22: Elasticity of citations to travel time: daily average travel time

Column (1) shows the result of Poisson Pseudo Maximum Likelihood (PPML) estimation of $citations_{FiGjhkt} = \exp\left[\boldsymbol{\beta}\, log(\text{travel time}_{ijt}) + FE_{FiGjhk} + FE_{Fiht} + FE_{Gjkt}\right] \times \varepsilon_{FiGjhkt}$, for citations of patents filed by establishment of firm F in location i, technology h and time period t, to patents filed by establishment of firm G in location j and technology k. travel time$_{ijt}$ is the travel time in minutes between location i and j at time period t, and it is set to 1 when $i = j$. When $FiGjhk$ has positive citations in at least one period and no citations in another, we attribute zero citations in the missing period. Column (2) includes the interaction of travel time$_{ijt}$ with a dummy for distance bin between the citing establishment Fi and the cited establishment Gj. Column (3) and (4) use the daily average travel time, which is computed as the average of the fastest travel time departing at every full hour (the average across all 24 potential departing times). Standard errors clustered at the non-directional location are presented between parentheses (ij is the same non-directional location pair as ji). R2 is computed as the squared correlation between observed and fitted values.

E.2. Creation of knowledge

E.2.1. Heterogeneous effects

Dependent Variable:	Patents		Patents quality weighted	
	(1)	(2)	(3)	(4)
log(knowledge access)	10.1*** (3.7)		12.3*** (4.0)	
log(knowledge access quality weighted)		8.0** (3.3)		10.0*** (3.6)
R2	0.85	0.85	0.86	0.86
N obs. effective	991,480	991,480	991,284	991,284

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$

Table 23: Effect of knowledge access on patents, quality weighted

Column (1) shows the result of Poisson Pseudo Maximum Likelihood (PPML) estimation of $\text{Patents}_{Fiht} = \exp[\rho \; log(\text{KA}_{iht}) + FE_{Fih} + FE_{it} + FE_{ht}] \times \xi_{Fiht}$, for patents filed by establishment of firm F in location i, technology h and time period t. KA_*iht* is knowledge access of establishments in location i technology h and time period t. Columns (1) and (2) use number of patents as dependent variable while columns (3) and (4) quality-weighted patents. Columns (1) and (3) use $log(\text{KA}_{iht})$ as explanatory variable while columns (2) and (4) use a quality weighted $log(\text{KA}_{iht})$. Quality weights are the 5-year percentile of quality measure after demeaning by year fixed effects computed in Kelly et al. (2021). Weighting by the 10-year percentile of quality gives similar results. Standard errors clustered at the location-technology level *ih* are presented in parentheses. R2 is computed as the squared correlation between observed and fitted values.

E.2.2. IV PPML: centering instrumental knowledge access

The objective of the recentered instrument is to clean any non-random variation that may be mechanically introduced due to geography. Locations that are geographically far from the initial innovation centers are more likely to have a larger increase in knowledge access with the roll out of jet airplanes, in any realization of the flight network. In order to purge out this potentially non-random variation, we compute the expected value of the instrument considering multiple alternative flight networks and subtract it from the *realized* instrument.

We construct the expected instrument $\mathbb{E}[log(\widetilde{KA}_{iht})]$ as follows:

1. Count the amount of airport-pairs connected by a non-stop flight in 1951, label this the *number of 1951 connections*.

2. Set a new seed number for random draws.

3. For each unique origin airport present in 1951, create a counterfactual connection by drawing a random destination airport (different to the origin) present in 1951.

Repeat as many times until the amount of unique counterfactual connections is equal to the *number of 1951 connections*. As *number of 1951 connections* is larger than the number of origins and destinations, some origins and destinations will be repeated.

4. Check if all 1951 origin and destination airports are present in the randomized connections. Some destinations may not be present due to the random draws. If some destination (origin) is missing, drop a counterfactual connection of a destination (origin) that has at least two origins (destinations). Draw a new random connection for the missing destination (origin). Repeat this step until all origins and destinations are present in the counterfactual network.

5. Check if the counterfactual network is a connected set (i.e. it would be possible to route from any airport to any other airport through intermediary connections). If it is not a connected set, drop this iteration of the counterfactual network and go back to step 2.

6. Predict flight duration of each counterfactual connection in each year using airport-to-airport distance and the estimated intercept and slope of each year

7. Compute the fastest travel time between each airport pair, directly and indirectly connected

8. Match airports to MSA and take minimum travel between MSA-pairs in each year

9. Repeat steps 2 to 9 for 2,000 times

10. With each counterfactual network, compute the counterfactual knowledge access of each MSA-technology-year.

11. Obtain the expected instrument $\mathbb{E}[log(\widetilde{KA}_{iht})]$: within each MSA-technology-year, compute the across-counterfactual network average of the log counterfactual knowledge access

We then recenter the instrument as follows:

$$log(\widetilde{KA}_{iht})_{centered} = log(\widetilde{KA}_{iht}) - \mathbb{E}[log(\widetilde{KA}_{iht})] \tag{15}$$

E.2.3. IV PPML: first and second stage estimation, non-centered instrument

	First stage OLS	Second stage PPML
Dep. variable:	log(knowledge access) (1)	*Patents* (2)
log(knowledge access instrument)	1.01^{***} (0.032)	
log(knowledge access)		11.24^{*} (6.35)
residual		-2.31 (7.20)
N obs. effective	$991,480$	$91,480$
R2	0.99	0.85
Within R2	0.53	

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$

Table 24: Elasticity of patents to knowledge access: first and second stage IV PPML

The table presents the results of 2-step instrumental variables estimation of $\text{Patents}_{Fiht} = \exp\left[\boldsymbol{\rho}\, log(KA_{iht}) + FE_{Fih} + FE_{it} + FE_{ht}\right] \times \xi_{Fiht}$, where $log(KA_{iht})$ is instrumented with $log(\widetilde{KA}_{iht})$. Column (1) shows the results of the first stage regression estimated by OLS. Column (2) shows the result of the second stage regression estimated by Poisson Pseudo Maximum Likelihood , including the estimated residuals of the first stage as controls. The number of observations in the second stage estimation is smaller due to not being able to identify fixed effects that are required in PPML estimation.

	OLS First stage reference quartile	OLS First stage 3rd quartile	OLS First stage 2nd quartile	OLS First stage 1st quartile	Second stage PPML
Dep. variable:		log(knowledge access)			*Patents*
	(1)	(2)	(3)	(4)	(5)
log(knowledge access instrument)	1.00^{***} (0.03)	0.01 (0.06)	0.03 (0.03)	0.00 (0.01)	
log(knowledge access instrument) × 3rd quartile	0.01^{*} (0.004)	1.11^{***} (0.03)	−0.00 (0.01)	−0.00 (0.01)	
log(knowledge access instrument) × 2nd quartile	0.00 (0.01)	−0.01 (0.04)	1.11^{***} (0.03)	−0.00 (0.01)	
log(knowledge access instrument) × 1st quartile	0.01 (0.01)	−0.00 (0.04)	−0.04 (0.04)	1.15^{***} (0.04)	
log(knowledge access)					10.26 (6.38)
log(knowledge access) × 3rd quartile					2.32^{***} (0.66)
log(knowledge access) × 2nd quartile					4.21^{***} (0.84)
log(knowledge access) × 1st quartile					5.77^{***} (1.11)
residual					−2.25 (7.27)
residual × 3rd quartile					−2.55 (1.59)
residual × 2nd quartile					-4.32^{**} (1.97)
residual × 1st quartile					-8.27^{**} (3.28)
N obs. effective	991,480	991,480	991,480	991,480	991,480
R2	1.00	1.00	1.00	1.00	0.85
Within R2	0.53	0.89	0.90	0.90	

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$

Table 25: Elasticity of patents to knowledge access: first and second stage IV PPML

The table presents the results of 2-step instrumental variables estimation of $\text{Patents}_{Fiht} = \exp\left[\sum_q \rho_q \times \mathbb{1}\{quartile_{ih} = q\} \times log(KA_{iht}) + FE_{Fih} + FE_{it} + FE_{ht}\right] \times \xi_{Fiht}$, where $log(KA_{iht})$ is instrumented with $log(\widetilde{KA}_{iht})$. Column (1) to (4) show the results of the first stage regression estimated by OLS. Coefficients of the 4 interactions of the instrument can be identified due to the presence of the fixed effects, e.g. after demeaning by fixed effects there is residual variation that allows to identify the 4 coefficients in each regression of the first stage. Column (5) shows the result of the second stage regression estimated by Poisson Pseudo Maximum Likelihood, including the estimated residuals of the first stage as controls.

E.2.4. IV PPML: first and second stage estimation, centered instrument

	First stage OLS	Second stage PPML
Dep. variable:	log(knowledge access)	*Patents*
	(1)	(2)
centered log(knowledge access instrument)	1.19^{***} (0.05)	
log(knowledge access)		9.86^{*} (5.73)
residual		0.57 (6.41)
N obs. effective	991,480	91,480
R2	0.99	0.85
Within R2	0.51	

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$

Table 26: Elasticity of patents to knowledge access: first and second stage centered IV PPML

The table presents the results of 2-step instrumental variables estimation of $\text{Patents}_{Fiht} = \exp\left[\boldsymbol{\rho}\, log(KA_{iht}) + FE_{Fih} + FE_{it} + FE_{ht}\right] \times \xi_{Fiht}$, where $log(KA_{iht})$ is instrumented with centered $log(\widetilde{KA}_{iht})$. Column (1) shows the results of the first stage regression estimated by OLS. Column (2) shows the result of the second stage regression estimated by Poisson Pseudo Maximum Likelihood , including the estimated residuals of the first stage as controls. The number of observations in the second stage estimation is smaller due to not being able to identify fixed effects that are required in PPML estimation.

	OLS First stage reference quartile	OLS First stage 3rd quartile	OLS First stage 2nd quartile	OLS First stage 1st quartile	Second stage PPML
Dep. variable:		log(knowledge access)			*Patents*
	(1)	(2)	(3)	(4)	(5)
centered log(knowledge access instrument)	1.19*** (0.05)	0.34 (0.29)	0.29** (0.14)	0.01 (0.03)	
centered log(knowledge access instrument) × 3rd quartile	0.00 (0.01)	-1.59*** (0.20)	0.07** (0.03)	0.00 (0.00)	
centered log(knowledge access instrument) × 2nd quartile	0.01 (0.01)	0.10 (0.24)	-1.52*** (0.19)	0.03 (0.02)	
centered log(knowledge access instrument) × 1st quartile	0.00 (0.02)	0.09 (0.23)	0.47** (0.22)	-1.93*** (0.19)	
log(knowledge access)					7.01 (5.83)
log(knowledge access) × 3rd quartile					3.99*** (1.25)
log(knowledge access) × 2nd quartile					7.57*** (2.30)
log(knowledge access) × 1st quartile					9.03*** (2.46)
residual					3.25 (6.53)
residual × 3rd quartile					−2.91 (1.16)
residual × 2nd quartile					−5.51** (2.29)
residual × 1st quartile					−6.01*** (2.30)
N obs. effective	991,480	991,480	991,480	991,480	991,480
R2	1.00	1.00	1.00	1.00	0.85
Within R2	0.51	0.27	0.30	0.43	

***$p < 0.01$; **$p < 0.05$; *$p < 0.10$

Table 27: Elasticity of patents to knowledge access: first and second stage centered IV PPML

The table presents the results of 2-step instrumental variables estimation of $\text{Patents}_{Fiht} = \exp\left[\sum_q \rho_q \times \mathbb{1}\{quartile_{ih} = q\} \times log(KA_{iht}) + FE_{Fih} + FE_{it} + FE_{ht}\right] \times \xi_{Fiht}$, where $log(KA_{iht})$ is instrumented with centered $log(\widetilde{KA}_{iht})$. Column (1) to (4) show the results of the first stage regression estimated by OLS. Coefficients of the 4 interactions of the instrument can be identified due to the presence of the fixed effects, e.g. after demeaning by fixed effects there is residual variation that allows to identify the 4 coefficients in each regression of the first stage. Column (5) shows the result of the second stage regression estimated by Poisson Pseudo Maximum Likelihood, including the estimated residuals of the first stage as controls.

E.2.5. Robustness

	Baseline	Quartile absolute	Quartile per capita
Dependent Variable:		*Patents*	
	(1)	(2)	(3)
log(knowledge access)	10.14***	9.36**	7.77**
	(3.66)	(3.69)	(3.70)
log(knowledge access) × quartile 0.50		2.05***	0.75**
		(0.58)	(0.34)
log(knowledge access) × quartile 0.25		3.80***	1.58***
		(0.90)	(0.50)
log(knowledge access) × quartile 0.00		5.00***	4.03***
		(1.30)	(0.77)
N obs. effective	991,480	991,480	991,480
R2	0.85	0.85	0.85

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$

Table 28: Elasticity of new patents to knowledge access: absolute and per capita MSA innovativeness

Column (1) shows the result of Poisson Pseudo Maximum Likelihood (PPML) estimation of $\text{Patents}_{Fiht} = \exp\left[\boldsymbol{\rho}\, log(KA_{iht}) + FE_{Fih} + FE_{it} + FE_{ht}\right] \times \xi_{Fiht}$, for patents filed by establishment of firm F in location i, technology h and time period t. KA_{iht} is knowledge access of establishments in location i technology h and time period t. Column (2) opens the coefficient ρ by the quartile of innovativeness of location i within technology h, computed within technology using the absolute level of patents in the MSA-technology in 1949-1953. Column (3) computes the quartile of innovativeness using patents per capita in the MSA-technology in 1949-1953 using 1950 population. Higher quartile indicates higher initial level of innovativeness. The fourth quartile is used as reference category. Standard errors clustered at the location-technology ih are presented in parentheses. R2 is computed as the squared correlation between observed and fitted values.

	PPML		β by distance		+300km		+1,000km		+2,000km	
Dependent Variable:	*Patents*									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
log(knowledge access)	10.14*** (3.66)	9.36** (3.69)	18.17*** (4.63)	16.50** (4.76)	10.09** (4.66)	8.70* (4.67)	18.82*** (5.82)	19.08*** (5.74)	12.70 (8.18)	10.26 (7.92)
log(knowledge access) × 3rd quartile		2.05*** (0.58)		2.70*** (0.84)		2.12*** (0.58)		2.08*** (0.53)		1.94*** (0.49)
log(knowledge access) × 2nd quartile		3.80*** (0.90)		5.96*** (1.42)		4.19*** (0.88)		3.97*** (0.81)		3.64*** (0.73)
log(knowledge access) × 1st quartile		5.00*** (1.30)		8.94*** (1.97)		5.49*** (1.25)		5.28*** (1.23)		4.68*** (1.07)
N obs. effective	991,480	991,480	991,480	991,480	991,480	991,480	991,480	991,480	991,480	991,480
R2	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85

***$p < 0.01$; **$p < 0.05$; *$p < 0.10$

Table 29: Elasticity of new patents to knowledge access, varying beta or distance.

Column (1) shows the result of Poisson Pseudo Maximum Likelihood (PPML) estimation of $\text{Patents}_{Fiht} = \exp\left[\boldsymbol{\rho}\, log(KA_{iht}) + FE_{Fih} + FE_{it} + FE_{ht}\right] \times \xi_{Fiht}$, for patents filed by establishment of firm F in location i, technology h and time period t. KA_{iht} is knowledge access of establishments in location i technology h and time period t. Column (2) opens the coefficient ρ by the quartile of innovativeness of location i within technology h, computed using patents in 1949-1953. Higher quartile indicates higher initial level of innovativeness. The fourth quartile is used as reference category. Relative to columns (1) and (2), columns (3) and (4) compute Knowledge Access using four distance-specific β parameter according to distance bins between i and j. The bins are [0km, 300km], (300km, 1000km], (1000km, 2000km], +2,000km. Columns (5) to (10) use the same β as column (1) and (2), but computing Knowledge Access with a truncated sample of j that are further than a certain distance threshold from i. Standard errors clustered at the location-technology ih are presented in parentheses. R2 is computed as the squared correlation between observed and fitted values.

	PPML		OLS	
Dependent Variable:	*Patents*		$log(Patents)$	
	(1)	(2)	(3)	(4)
log(knowledge access)	10.14*** (3.66)	9.36** (3.69)	6.83* (3.19)	6.27* (3.20)
log(knowledge access) × 3rd quartile		2.05*** (0.58)		0.92* (0.51)
log(knowledge access) × 2nd quartile		3.80*** (0.90)		2.64** (1.03)
log(knowledge access) × 1st quartile		5.00*** (1.30)		3.82** (1.79)
N obs. effective	991,480	991,480	300,539	300,539
R2	0.85	0.85	0.87	0.87

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$

Table 30: Elasticity of new patents to knowledge access: PPML and OLS

Column (1) shows the result of Poisson Pseudo Maximum Likelihood (PPML) estimation of $\text{Patents}_{Fiht} = \exp\left[\boldsymbol{\rho}\, log(KA_{iht}) + FE_{Fih} + FE_{it} + FE_{ht}\right] \times \xi_{Fiht}$, for patents filed by establishment of firm F in location i, technology h and time period t. KA_{iht} is knowledge access of establishments in location i technology h and time period t. Column (3) estimates $log(\text{Patents})_{Fiht} = \boldsymbol{\rho}\, log(KA_{iht}) + FE_{Fih} + FE_{it} + FE_{ht} + \xi_{Fiht}$. Columns (2) and (4) open the coefficient ρ by the quartile of innovativeness of location i within technology h, computed within technology using the absolute level of patents in the MSA-technology in 1949-1953. Higher quartile indicates higher initial level of innovativeness. The fourth quartile is used as reference category. Difference in amount of observations is due to dropping zeros in columns (3) and (4). Standard errors clustered at the location-technology ih are presented in parentheses. R2 is computed as the squared correlation between observed and fitted values.

Access to capital

We construct four measures of access to capital using 1949-1953 market capitalization of firms listed in the stock market. The four measures are similar in their essence but differ in the computation of a firm's technology and the firm's location. The measure is computed as follows:

$$\textit{capital access}_{iht} = \sum_{k} \psi_{hk} \sum_{j,j\neq i} \text{Capital stock}_{jk,t=1951} \times \text{travel time}_{ijt}^{\xi} \tag{16}$$

where Capital $\text{stock}_{jk,t=1951}$ is a proxy for the capital which is specific to technology k located in j at the initial time period 1951. ψ_{hk} is an input-output weight of capital flows and ξ is the elasticity of capital flows between to travel time. As a proxy for capital we use market capitalization of firms.

We construct four measures of $\textit{capital access}_{iht}$ which differ on: (i) the way we define the allocation of the firm's capital to each location (either using all inventors' locations or only the assigned headquarters), and (ii) the way we allocate a firm's capital across technologies (using the share of a technology within the firm, or relative to the national share of that technology). We use COMPUSTAT as our source of data for market capitalization.

We proceed as follows:

1. Use share's market price at closure calendar year multiplied by the number shares outstanding. We use the variables *prcc_c* and *csho* to maximize coverage of firms given that other variables have missing value for many firms.

2. Take the yearly average market capitalization to maximize coverage (many firms have missing in a certain year). This step potentially introduces measurement error due to changes in total stock market capitalization but allows us to increase the amount of firms included in the sample.

3. Determine a firm's MSA using patent inventor location. Two ways to determine

the location, 1. only HQ location, 2. all locations where the firm had inventors applying for patents in 1949-1953

4. Determine the share of each technology firm's technology using patent technology. Two ways to determine the share oftechnology: 1. the share of each tech within firm + share within firm relative to national share

5. In the absence of data on a capital input-output weight, assume it is the same as the technology input-output weight, i.e. $\psi_{hk} = \omega_{hk}$

6. In the absence of data on the elasticity of capital flows to travel time assume $\xi = -1$

The four measures of access to capital are as follows:

1. Attribute all capital to headquarters and use the absolute share of each technology in the firm

2. Attribute all capital to headquarters and use the share of each technology in the firm relative to the national share

3. Attribute capital to establishments using their pat share and use the absolute share of each technology in the firm

4. Attribute capital to establishments using their pat share and use the share of each technology in the firm relative to the national share

Table **??** shows the results of estimating the elasticity of new patents to knowledge access while at the same time controlling for capital access.

Sensitivity to β

Indirectly connected MSAs

If the 1951 flight network was constructed in order to connect city pairs that would see future growth in citations, we can alleviate this endogeneity concern by focusing only

β	ρ	$\beta \times \rho$	Predicted yearly growth p.p.	Share yearly growth explained	Predicted yearly growth differential p.p.	Share yearly growth differential explained
-0.186	10.14	-1.89	3.47	0.78	1.1	0.21
-0.1	19.35	-1.94	3.5	0.78	1.07	0.2
-0.2	9.4	-1.88	3.47	0.78	1.1	0.21
-0.3	6.1	-1.83	3.45	0.77	1.14	0.22
-0.4	4.48	-1.79	3.44	0.77	1.16	0.22
-0.5	3.52	-1.76	3.44	0.77	1.19	0.23
-0.6	2.91	-1.74	3.45	0.77	1.2	0.23
-0.7	2.48	-1.73	3.47	0.78	1.22	0.23
-0.8	2.17	-1.73	3.5	0.78	1.22	0.23
-0.9	1.93	-1.73	3.52	0.79	1.24	0.24
-1	1.72	-1.72	3.51	0.79	1.28	0.24
-2	0.58	-1.16	2.8	0.63	1.55	0.3
-5	0.04	-0.19	1.19	0.27	3.65	0.7
-8	0.09	-0.76	8.22	1.84	6.96	1.33
-10	0.11	-1.08	15.16	3.4	8.19	1.56
-20	0.13	-2.63	69.8	15.65	21.66	4.14
-50	0.16	-8.22	531.34	119.16	219.49	41.94
-100	0.12	-12.33	5428.85	1217.49	2971.74	567.91

Table 32: Effect of knowledge access on new patents: varying the value of elasticity of knowledge diffusion

on indirectly connected pairs.

Table 33 presents PPML regressions not bias-corrected. Columns (1) and (2) are the baseline regressions (all MSA-pairs), columns (3) and (4) drop MSA-pairs that are ever connected with one leg (a non-stop flight), and columns (5) and (6) drop MSA-pairs that are ever connected with one flight number. The difference between non-stop and one flight number is that one flight number could serve multiple MSAs by making intermediate stops.[105] The estimated coefficients are in the ballpark of the initial estimates, especially for +2,000km, providing evidence that it is reasonable to use the pre-existing network as the baseline to construct the instrument.

[105] For example, in 1951 NYC-LA was connected with one flight number that included one stop in Chicago, that is two legs but only one flight number: passengers did not have to change airplanes).

	PPML					
	not bias-corrected					
Dep. variable:	*citations*					
	(1)	(2)	(3)	(4)	(5)	(6)
log(travel time)	-0.088^{***} (0.024)		-0.202^{***} (0.051)		-0.241^{***} (0.061)	
log(travel time) × 0-300km		0.021 (0.039)		-0.237^{***} (0.116)		-0.410^{**} (0.165)
log(travel time) × 300-1,000km		-0.099^{**} (0.027)		-0.147^{*} (0.081)		-0.210^{**} (0.095)
log(travel time) × 1000-2,000km		-0.093^{**} (0.042)		-0.157^{*} (0.092)		-0.216^{**} (0.109)
log(travel time) × +2,000km		-0.185^{***} (0.049)		-0.297^{***} (0.085)		-0.242^{***} (0.090)
N obs. effective	4,703,010	4,703,010	1,735,427	1,735,427	1,396,393	1,396,393
R2	0.88	0.88	0.94	0.94	0.94	0.94
Observation selection:						
All	X	X				
Discard one leg			X	X		
Discard one flight number					X	X

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$

Table 33: Elasticity of citations to travel time: dropping directly connected MSA pairs

Column (1) shows the result of Poisson Pseudo Maximum Likelihood (PPML) estimation of $citations_{FiGjhkt} = \exp\left[\boldsymbol{\beta}\, log(\text{travel time}_{ijt}) + FE_{FiGjhk} + FE_{Fiht} + FE_{Gjkt}\right] \times \varepsilon_{FiGjhkt}$, for citations of patents filed by establishment of firm F in location i, technology h and time period t, to patents filed by establishment of firm G in location j and technology k. travel time$_{ijt}$ is the travel time in minutes between location i and j at time period t, and it is set to 1 when $i = j$. When $FiGjhk$ has positive citations in at least one period and no citations in another, we attribute zero citations in the missing period. Column (2) includes the interaction of travel time$_{ijt}$ with a dummy for distance bin between the citing establishment Fi and the cited establishment Gj. Column (3) and (4) discards all ij that are ever connected with one leg (non-stop flight), while columns (5) and (6) discard all ij that are ever connected with one flight number. The difference between non-stop and one flight number is that one flight number could serve multiple MSAs by making intermediate stops. Standard errors clustered at the non-directional location are presented between parentheses (ij is the same non-directional location pair as ji). R2 is computed as the squared correlation between observed and fitted values.