DISCUSSION PAPER SERIES

DP16734

Transportation Costs in the Age of Highways: Evidence from United States 1955-2010.

Sylvain Barde and Alexander Klein

ECONOMIC HISTORY

ISSN 0265-8003

Transportation Costs in the Age of Highways: Evidence from United States 1955-2010.

Sylvain Barde and Alexander Klein

Discussion Paper DP16734
Published 18 November 2021
Submitted 17 November 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre’s research programmes:

- Economic History

Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Sylvain Barde and Alexander Klein

Transportation Costs in the Age of Highways: Evidence from United States 1955-2010.

Abstract

This paper constructs general road transport costs in the United States between 1955 and 2010 combining stock measures of transportation network with fuel consumption, driving speed, fuel prices, and labour costs. This results in a novel data set of 3105×3105 county-pairs for seven benchmark years. Using a county-level counterfactual analysis, we precisely quantify the reduction of the transport cost generated by Interstate Highway System. We document an inverted U-shape pattern for road transport costs, peaking in 1980, explained by initially increasing labor costs, followed by cost reductions due to trucking industry deregulation and the completion of the IHS.

JEL Classification: N72, N92, O18, R41

Keywords: transport costs, Interstate Highway System, Road Network, Dijkstra's algorithm

Sylvain Barde - s.barde@kent.ac.uk
University of Kent, School of Economics

Alexander Klein - a.klein-474@kent.ac.uk
University of Kent, School of Economics and CEPR

Acknowledgements
We would like to thank the Division of Human and Social Sciences, University of Kent, for their financial support, Peter Matthews for excellent research assistance, and the participants of LSE Economic History Seminar Series, and European Urban Economic Association Meeting in Amsterdam 2019 for helpful comments.

Transportation Costs in the Age of Highways: Evidence from United States 1955-2010

Sylvain Barde*[1] and Alexander Klein[1]

[1]University of Kent, School of Economics

Abstract

This paper constructs general road transport costs in the United States between 1955 and 2010 combining stock measures of transportation network with fuel consumption, driving speed, fuel prices, and labour costs. This results in a novel data set of 3105×3105 county-pairs for seven benchmark years. Using a county-level counterfactual analysis, we precisely quantify the reduction of the transport cost generated by Interstate Highway System. We document an inverted U-shape pattern for road transport costs, peaking in 1980, explained by initially increasing labor costs, followed by cost reductions due to trucking industry deregulation and the completion of the IHS.

JEL classification: N72, N92, O18, R41

Keywords: Transport costs, Interstate Highway System, Road Network, Dijkstra's algorithm

*We would like to thank the Division of Human and Social Sciences, University of Kent, for their financial support, Peter Matthews for excellent research assistance, and the participants of LSE Economic History Seminar Series, and European Urban Economic Association Meeting in Amsterdam 2019 for helpful comments.

1 Introduction

The question of the effect that the transportation network has on economic activity is at the heart of our understanding of spatial distribution of economic activities. Recently, it has sparked a renewed interest with studies analysing effects of transport infrastructure on GDP, suburbanization, trade, employment, earnings, housing costs, amenities, road congestion or urban growth.[1] In all studies, the quantification of transportation infrastructure was crucial. Measures related to the physical layout of the transport network are used most often, such as a categorical variable indicating whether transport infrastructure passes through a geographical unit (e.g. city, county, region), or a stock measure such as the length of infrastructure (e.g. the length of highways, number of highway rays). High correlation of these measures with transportation costs then allow spatial and temporal changes in the transport network to be interpreted as changes in the transportation costs.

Stock measures or a categorical variable offer important insights on how the transportation network impacts the economy when assessing the impact of a newly built transport network, or the expansion of an existing one. They capture an *extensive margin* of transport costs: a county without an access to a highway or a railway line would see its transportation costs decline once it gained access to them. Transportation costs, however, also include fuel consumption, driving speed, fuel prices, and labour costs which stock measures or categorical variables might not fully capture. Furthermore, stock measures and categorical variables can not capture the time variation of transportation costs when the physical layout of transportation network is near its completion or fully completed, hence changes little over time.[2]

The purpose of this paper is threefold. First, it goes beyond the stock or categorical variable measures of transportation costs and constructs general road transport costs that combine stock measures of transportation network evolution with fuel consumption, driving speed, fuel prices, and labour costs for post-World War II United States. These transport costs measure the marginal costs per single journey between county centroids. This results in a new data set comprising $3105\times3104\times1/2$ pairs of transport costs in 2010 US$ between the population centroids of 3105 US counties in 1955, 1960, 1970, 1980, 1990, 2000, and 2010. Second, the level of detail achieved allows us to document novel stylised facts about spatial and temporal changes of road transport costs over the course of second half of the twentieth century. Third, it quantifies the impact of the Interstate Highway System in the United States on county-pair road transport costs between 1955 and 2010.

Our paper is closely related to the literature on generalized transport costs, which combine all components of transport costs into one measurable index (e.g. Condeço-Melhorado et al., 2011; Koopmans et al., 2013). In that respect, we contribute to it by applying the methodology of Combes and Lafourcade (2005) and calculate generalized road transport costs for the post-World War II US road network. This

[1] See for instance Baum-Snow (2007); Duranton and Turner (2012); Brinkman and Lin (2019); Jaworski et al. (2020).
[2] This is recognised for example in footnote 17 of Herzog (2021).

paper also relates to the literature on US road infrastructure, especially the IHS (Baum-Snow, 2007, 2020; Chandra and Thompson, 2000; Duranton and Turner, 2012; Duranton et al., 2014; Herzog, 2021; Jaworski and Kitchens, 2019; Jaworski et al., 2020; Michaels, 2008; Brinkman and Lin, 2019; Allen and Arkolakis, 2019; Frye, 2021). Most of this literature uses stock measures or a categorical variable, except for Jaworski and Kitchens (2019), Jaworski et al. (2020), and Allen and Arkolakis (2019) where transport costs focus on speed and minimisation of travel time. There are other studies which examine the role of transport infrastructure in developing countries (Baum-Snow et al., 2017; Datta, 2012; Faber, 2014; Ghani et al., 2016; Qin, 2017), or historical context (Donaldson and Hornbeck, 2016; Donaldson, 2018; Alvarez-Palau et al., 2020). These are indirectly related to our paper, either methodologically through the use of Dijkstra's algorithm or because they focus on the same historical time period.

Methodologically, Jaworski and Kitchens (2019) and Herzog (2021) are most closely related to our work. Jaworski and Kitchens (2019) also apply the Combes and Lafourcade (2005) method and Dijkstra's algorithm to obtain over 4 million pairwise US county transport costs for 1960, 1995 and 2010. Our work improves on this in two ways. First, we refine the analysis both in the time dimension, by considering more years, and in the spatial dimension, by using state-level transport cost determinants such as driving speeds, fuel prices and fuel consumption, rather than the US-wide aggregate data used by Jaworski and Kitchens (2019). Second, in their analysis Dijkstra's algorithm is used to first minimise the travel time between counties, with the resulting travel time and distance combined with the cost determinants to obtain the cost for the route. Our approach instead follows Combes and Lafourcade (2005) more closely, by first calculating the time-related and distance-related cost of travelling on each road, and then using Dijkstra's algorithm to minimise the overall cost of travelling. As will be discussed further below, this allows route distance vs. route time trade-offs to be integrated into the decision problem of finding the cheapest route. A similar set of arguments apply to Herzog (2021), which follows a similar strategy of using Dijkstra's algorithm to calculate minimum travel times, again on the basis of aggregate US driving speeds.

Our paper is also related to the literature that calculates the contribution of the IHS to welfare or GDP. Allen and Arkolakis (2014, 2019), Jaworski et al. (2020) offer various estimates.Allen and Arkolakis (2014) estimate that, depending on the parameters of their model, the removal of IHS in 2000 would lead to 1.1%-1.4% decline in welfare. Allen and Arkolakis (2019) offer various welfare estimates of improving IHS in terms of lane-miles added to each highway and Jaworski et al. (2020) estimate that removing IHS in 2010 would lower GDP by 3.9%. Our estimates show that removing IHS would lead to about 2% reduction in county-pairwise marginal transport cost per journey and offer a different perspective on the contribution of IHS to economic activities.

The remainder of the paper is organized as follows. Section 2 present a historical context of building a road network in the United States. Section 3 discusses the methodology, data sources and main stylised facts about the determinants of transport costs. The evolution of the minimal pairwise transport costs are discussed in Section 5, and the specific impact of the IHS is presented in Section 6. Section 7 concludes.

2 A History of US Road Network

Designing and building a system of interstate highways has a long history, with the IHS marking a culmination of the decades-long efforts to create standardized, high-speed, access-controlled highways connecting US states north to south and east and west. This section will discuss these efforts and will focus on the following areas: (i) highway legislation and financing of the highway construction (ii), construction and design.[3]

The construction of a long-distance road network goes back to the early nineteenth century with early turnpike and toll roads such as turnpike roads in Connecticut, New York-Philadelphia road, or the road built by North-western Turnpike from Winchester to Parkersburg on the Ohio River. Building and maintenance of roads dropped in priority during the railway era, until the rise of motorized vehicles at the turn of the twentieth began to incentivize road building (U.S. Federal Highway Administration, 1977). Indeed, railroad competition rendered toll roads unprofitable, causing them to cease their operation and surrender their charters. These were taken over by local authorities and used as common roads. Rapid urbanization in the late nineteenth century brought about widening disparities between the urban areas, with quality local urban transportation system on one side, and rural areas with poorly maintained and often seasonally inaccessible roads on the other. The Good Roads Movement of the 1890s organized campaigns to raise awareness and nationwide support towards the passing of 'good roads laws', as did North Carolina and Iowa (U.S. Federal Highway Administration, 1977, pp 41-42). The movement helped to bring the federal government and the Congress to view a national road system as a Federal interest and initiate legislation in its support.[4] Legislative efforts since then included the issues of financing the road construction (especially the share of federal funds being appropriated for that purpose), the constitutional delimitation of federal vs. state powers, and the implementation of the road-building program itself. The important milestones include the creation of the permanent Office of Public Roads in 1905, the 1916 Federal Aid Road Act, the 1921 Federal Aid Highway Act, the formation of the American Association of State Highway Officials (AASHO) in 1914, the interstate highway system proposal by the Federal Works Administrator in 1947, and the 1956 Federal-Aid Highway Act.[5]

The 1916 Act was the first federal highway legislation providing federal funds to states to build state-wide roads, on a fifty-fifty matching basis. It facilitated the establishment of state highway agencies to administer these funds. The later act of 1921 reflected the demand for a national highway system which

[3]This section draws closely from *America's highways, 1776-1976* (U.S. Federal Highway Administration, 1977).

[4]In early years, the interest of the federal government in good quality roads was linked to the delivery of the post.

[5]There were many other important milestones related to the financing of the interstate highway system, organization of state and federal road agencies, technological developments such as the design of roads, road-surface materials, or road safety measures. For details see e.g. U.S. Federal Highway Administration (1977).

emerged during the World War I, the rise of trucking industry as well as substantial legislative efforts during and after the war. It established, among other others, the first set of quality standard for interstate roads, opened offices of the Bureau of Public Roads in eleven western states, and approved the selection of "7% system" in which not more than 7% of all state roads would be designated to the federal interstate road system. This rule laid the basis for the future expansion of the IHS, and also established that state highway departments in cooperation with county officials would select which state roads would form part of the interstate system co-financed with the federal government (Federal-Aid Highway Act 1921, section 'Selection of System of Road'). The Bureau of Public Roads led the expansion of systematic planning work for the interstate highway network with a series of transportation surveys, and generated the first map of roads planned in the event of war, called the "Pershing Map".

The pressure from the soaring number of motor vehicles after the First World War led states to build roads and highways even without the support from the federal government. As a result, the interwar decades saw a substantial increase in highways financed and maintained by states only, often crossing state boundaries and creating a de-facto system of highways connecting all contiguous US states. A Joint Board on the Interstate Highways requested by the AASHO and appointed by the Secretary of Agriculture submitted a recommendation 1925 for a comprehensive and uniform interstate route designation scheme amounting to 75,884 miles (Report of Joint Board on Interstate Highways, 1925, Table III). This created the U.S. Route or U.S. Highway system, an integrated network of highways numbered within a nationwide grid. The AASHO was also instrumental in creating roads standards and active in adopting standard practices in road building to ensure a uniformity of roads from state to state (U.S. Federal Highway Administration, 1977, pp. 390-395).

The Second World War paused the efforts to expand the existing interstate highway network, but the post-war boom, with a rapid associated recovery of motor traffic, brought the issue of an efficient and suitable highway system to the center of government attention. The 1944 Act appropriated $150 million in each of the first three postwar years, let states select the interstate routes and limited the total mileage of the interstate highways to 40,000 miles. While the states chose the routes it was the Public Road Administration, a successor to the Bureau of Public Roads, which set the standards in collaboration with the AASHO. Furthermore, a map of the entire interstate highway system was created and approved by the Federal Works Administrator on August 2, 1947 which was subsequently amended with additional routes through urban areas and approved as the official National System of Interstate Highways on September 15, 1955.

Post-war federally funded interstate road construction started slowly, mostly due to the high price of highway structures, resulting from increased demand for construction materials during the post-war housing boom, causing a shortage of labour, materials, equipment and contractors. At the same time, the high demand for high-speed highways made toll roads and turnpikes profitable again. With the

Pennsylvania Turnpike turning profit, Maine opened a turnpike connecting Portland, Maine with the borders of New Hampshire in 1947, and the New Jersey Turnpike Authority built a road connecting the George Washington Bridge and the Delaware River. All of these revealed a high latent demand for modern highways free of congestion which resulted in a toll-road boom and by 1955 over 1,200 miles of toll roads over seventeen states were built costing \$1.55 billion (U.S. Federal Highway Administration, 1977, pp. 167-168). Furthermore, the toll roads were high-speed with wide right-of-way, geometric standards equal or better than the ones recommend by the AASHO, and amenities for their users. This had become a thorny issue for the planners of the future IHS since the toll roads could have pre-empted federally funded highway location. It was resolved by new laws which provided no federal support for the toll roads, and where the roads followed planned Interstate system they would be incorporated in it provided they satisfy the standards. In addition, no new toll roads were permitted on the planned IHS.

Continuing increase of motor vehicles and the increase in freight resulting from the Korean War put strain on the existing highways. New plans to speed up the process of building the interstate highway system involved both Congress and President Eisenhower. After much legislative wrangling, the Federal-Aid Highway Act of 1956 was signed into law on June 29, 1956. It established a Highway Trust Fund and appropriated federal taxes on motor fuel, tires and tread rubber and a portion of excise taxes on automobiles, trucks, and buses. The funds were allocated to the states with a matching rate of 90% federal aid to 10% state funds. The Act called for the adoption of uniform interstate highway system design standards to facilitate high-speed, access-controlled roads. Specifically, the standard included, 12-foot travel lane width, 10-foot minimum shoulder width, elimination of railroad grade crossings, highway at-grade intersections, design speeds of 50, 60, and 90 mph for mountainous, rolling, and flat terrain conditions respectively, and geometric standards for curvature, gradient, width and number of lanes to allow such speeds. The standards of the IHS was the fundamental attribute that distinguished IHS highways from other U.S. highways, especially controlled-access which requires that the traffic across highways is carried by overpasses and underpasses, can be access only by ramps and the traffic is unhindered by traffic signals, intersections, and at-grade crossings with other roads or railways. The construction of the IHS started in a context where the U.S. Highway system had been in place since mid-1920s, with states raising their own funds and already constructing a network of roads. This was reflected in the choice of routing and the new highways were most often built as an upgrade or replacement of the exiting U.S. highways, especially in high road density areas on the East coast. Since the standards dictated controlled-access and unhindered traffic flow, mere upgrading was not possible in urban areas and the new highways were built either bypassing them or running through with multilevel crossing. Bypassing the urban areas often happened along the existing route, using so called parallel routing. For example, several sections of Interstate 5 in Southern California were constructed prior to the 1956 Federal Highway Act, including the Aliso Street Viaduct (built in 1948), portions of former U.S. 101 Santa Ana Freeway located the south of Los Angeles, and portions of U.S. 101 Montgomery Freeway south of San Diego. These sections were added to Interstate 5 and U.S. 101 was decommissioned in 1964. The rest of the route northward, I-5 parallels and replaces the former U.S. 99, which was decommissioned in stages between 1964 and 1972.[6]

The originally planned IHS was declared finished in 1992, but further expansions followed. The pace and geographical variation of the construction will be presented further in sections 3.1 and 4.1. Overall, the US interstates created a uniform road network of access-controlled roads which facilitated fast speeds and cross the entire US territory. Their impact on the transportation costs, along with other non-IHS related transport costs will be discussed in further below.

3 Data Sources and Methodology

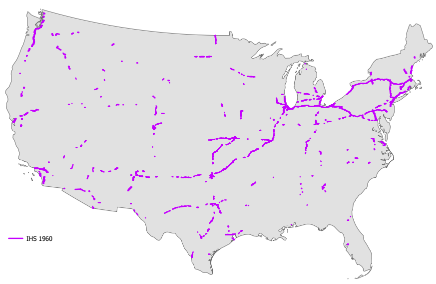
We broadly follow the methodology developed by Combes and Lafourcade (2005) in order to calculate the cheapest route on the US road network connecting pairs of US county population centroids for seven benchmark years: 1955, 1960, 1970, 1980, 1990, 2000, and 2010. This requires using a network representing the US road system, calculating the cost of travelling on each segment of the network, then using Dijkstra's algorithm to calculate the cheapest route between pairs of centroids on the network. The size of the US road network and the large number of pairwise county routes involved in the analysis mean that a preliminary data reduction step is required in order to make the calculations tractable, detailed in appendix B. In order to be able to accurately calculate transport costs given the large geographical and time dimensions involved, we also collected rich, spatially disaggregated data which allow us to model closely both distance and time-related travel costs and thus capture the spatial and time variation of transport cost determinants.

3.1 The US Road Network

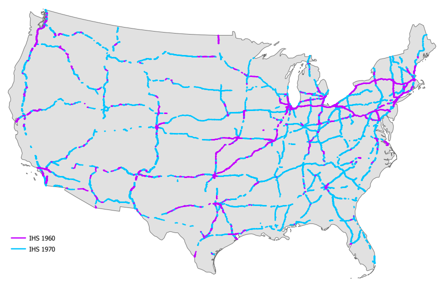
The first task involves creating a digitized map of the US road system that reflects the state of the road network for each of our benchmark years of 1955, 1960, 1970, 1980, 1990, 2000, and 2010, The starting point of our analysis is the digitized National Highways Planning Network 14.05 GIS shape file (henceforth NHPN), which contains the full IHS and US highway system.[7] Every signed road is split into very fine-grained segments, often on the scale of a hundred meters, leading to a large network containing 625,610 road segments. Each segment is assigned a county and state identifier, in addition to information relating to signage and distance, allowing us to assign state or county level attributes, such as the rural or urban nature of the county. This rural/urban classification, which is detailed further in appendix A, is based on the US Department of Agriculture Rural-Urban Continuum Codes, which allows us to track

[6] https://www.interstate-guide.com/i-005/#history, accessed April 17, 2020.
[7] US Federal Highway Administration, 2014

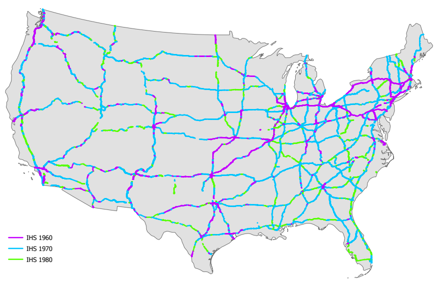
Figure 1: Construction of the Interstate Highway System 1960-2010



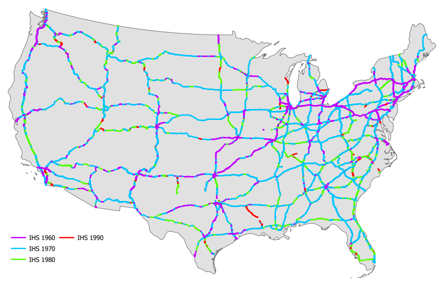
(a) 1960



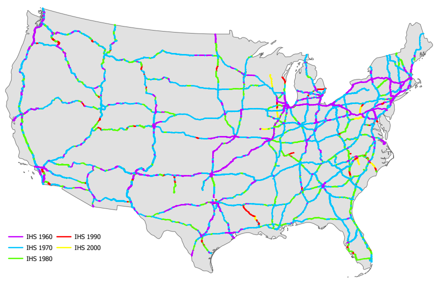
(b) 1970



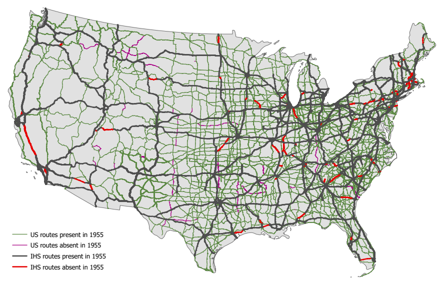
(c) 1980



(d) 1990



(e) 2000



(f) 1955 routing diagnostics

the progress of urbanisation in the post-war period in the 48 contiguous US states. Combined with a distinction between interstate vs. non interstate roads, we therefore consider 4 categories of roads: urban interstates, rural interstates, urban non-interstate and rural non-interstate. The rationale for this specific classification is that this allows us to make use of the historical, state-level, average driving speed data discussed further below.

Three data sources were used to determine the time-evolution of the IHS in each benchmark year. First, the 'Interstate Density Maps' which were published by the Federal Highway Administration and show the evolution of the interstate highways every ten years between 1950 and 2000.[8] Second, we use the Federal Highway Administration PR-511 records, which contain information relating to the time when each segment of the IHS was open to public traffic. This data has been used before including the studies by Chandra and Thompson (2000), Baum-Snow (2007), Nall (2015) or Frye (2021). Finally, this was supplemented with historical sources including Interstate-Guide, an online comprehensive guide to the IHS and its history.[9] Given the very fine resolution of the NHPN shape file, we were able to determine very accurately which portion of which highway in which county was open to the public by the end of each decade. The staggered construction of the IHS over the decades of our analysis are shown in Figure 1.

As discussed in Section 2, the construction of the IHS followed the practices of so called parallel routing – building interstate highways alongside existing roads – and upgrading the existing state highways. This included upgrades of the road surface, increases in the number of lanes, and building of over- and underpass crossings (e.g. Interstate-Guide (2021), Georgia Department of Transportation 2007, Colorado Department of Transportation 2002, Oregon Department of Transportation 2004, California Department of Transportation,[10].). Indeed, many parallel routed roads still exist, often as frontage roads, though some roads were decommissioned such as the well-known US Route 66 which existed from 1926 until 1985. As a result, the physical layout and direction of the IHS followed the existing layout of state and US routes, which is indeed confirmed when we overlay the interstate highway system map with state and US roads map from 1955. This fact helps us to deal with the segments of the IHS in the NPHN shape files which did not exist in a particular benchmark year. Because of the practices of parallel routing and upgrading of existing state highways, these segments are recoded as non-interstate roads prior to their construction date. In order to ensure this approach was appropriate, we manually checked each signed interstate segment against existing maps of US routes in 1950 and 1955 as well as on Interstate-Guide to establish whether a pre-existing route was present in that location in 1955. As shown in the first column of Table 1 and in Figure 1(f), over 92% of IHS mileage already existed in some form in 1955.

Table 1 reveals that US routes and state routes make up most of the remaining roads in the NHPN

[8] `https://www.fhwa.dot.gov/interstate/densitymap.cfm`
[9] `https://www.interstate-guide.com/`
[10] `https://www.cahighways.org/itypes-history.html`, accessed on February 15, 2021.

Table 1: Composition of the US road network

	Interstates	US routes	State routes	Other	Total
2014 Mileage	45,637.31	132,715.53	209,166.82	49,371.34	436,891.48
% of total	10.45	30.38	47.88	11.3	100
% pre 1955	92.44	97.11	94.52	-	-

Source: National Highway Planning Network version 14.05.
'Other' category contains county, township, municipal and unsigned roads.

shapefile, making up 78% of the network mileage. As is the case for the IHS, the overwhelming majority of the physical layout of non-interstate highway was in place by 1955, as state and US routes were being built since the 1920s. Indeed, the manual check carried out for US routes against the 1950 map, shown in Figure 1(f), confirmed that 97% percent of the network of US routes present in the 2014 NPHN map already existed in 1955. The network of state routes was also checked using state-level transportation sources to establish the opening date of each signed state route, and again confirmed that around 95% of state routes were open by 1955. [11] The remaining 11% of the network is made up of county, municipal and township routes, which cannot be checked due to lack of signage. However, these routes are extremely local to settled areas within a county.

It is important to emphasise that the fact that we use the 2014 NPHN network as the basis of our work does not mean that we ignore the lengthy construction of the IHS itself, nor that construction activity occurred on non-interstate highways between 1955 and 2010. Most of these efforts, however, focused on improving existing roads, by building additional lanes, resurfacing, and building bypasses and tunnels to avoid dangerous segments of the roads. Clearly all this infrastructure investment will impact transport costs and it is therefore crucial to account for this improvement of the road system over time. However, while this affects the cost of travelling on a road segment in the network, it does not affect the existence or location of the segment itself. Instead, these qualitative improvements will be captured by the reference cost data used to calculate the cost of traversing the segment.

3.2 Reference Costs and Minimal Transport Costs

In addition to providing the topology of the network itself, the NHPN data provides two types of information for each road segment, which will be referred to as an *edge*: the distance of each the road segments d_e and the road type of the edge in a given year, $r_{e,t} \in \{1, 2, 3, 4\}$. This serves as the basis of the calculation of the dollar cost associated with each edge. Following Combes and Lafourcade (2005), we consider two types of reference costs: distance-related costs per mile travelled and time-related costs per hour spent travelling. It is important to mention two general comments. First, as previously mentioned, given the much larger geographical scope of the analysis compared to the French case in Combes and Lafourcade (2005), both types costs vary not only across years and the type of edges but also across US

[11] Because state routes represent nearly half the network, these are not displayed in Figure 1(f) due to space constraints

states, providing a high-level of spatial disaggregation. Second, due to the federal nature of the USA and to the fact that our analysis goes back to 1955, it has not been possible to gather data relating to the additional maintenance, depreciation, insurance and accommodation costs that are included in Combes and Lafourcade (2005) for the French case. In this respect, we are in a similar situation to Jaworski and Kitchens (2019), who similarly report lack of data availability for these ancillary costs in the USA.

The time-related reference cost for an edge are the costs incurred by the truck drivers and are calculated as:

$$T_{a,t} = w_t \frac{d_e}{s_{r_{e,t},t}} \tag{1}$$

where w_t is the nominal average hourly earnings in trucking in year t, the time spent traversing edge e is given by the edge's distance d_e, in miles, divided by the driving speed on the edge in year t, $s_{r_{e,t},t}$, measured in miles per hour. The hourly wage were calculating using the weekly earnings in trucking, taken from the publications of the Bureau of Labor Statistics: US Department of Labor Bulletin, Employment, Hours, and Earning, United States 1904-1994, and Occupational Employment Statistics Survey.

As mentioned previously, the average driving speed data encodes improvements in the quality of the road network, especially the specific contribution of the IHS, as well as in the capabilities of motor vehicles. We collected average driving speeds for the four types of roads we consider in each of seven benchmark years, and in each of forty-eight contiguous states. This results in 1,344 different speeds ($4 \times 7 \times 48$), providing detailed variation over time, geography, and type of roads. The main source was the Federal Highway Statistics supplemented with the U.S. Historical Statistics, U.S. Bureau of Transportation Statistics, and Department of Commerce, with further details discussed in Appendix A.

The distance related reference costs for each edge and year are calculated as:

$$D_{e,t} = p_{e,t} \times gpm_{e,t} \times d_e \tag{2}$$

This is simply the product of fuel price on the edge in a given year, $p_{e,t}$, the fuel consumption on the edge in gallons per mile $gpm_{e,t}$ and the edge's distance d_e. State-level fuel prices were taken from the U.S. Department of Agriculture's 'Agricultural Prices: Annual Summary', the U.S. Energy Information Administration's average price of motor gasoline series and from the Bulletin of U.S. Department of Labour. Fuel consumption in each year and each of the forty eight contiguous states was calculated by dividing vehicle-miles travelled by total fuel consumption, using data from the Federal Highway Statistics 1950-2010. As is the case for driving speeds, the fuel consumption data capture qualitative improvements over time, specifically the overall improvement in fuel efficiency seen in the motor industry over the 1955-2010 period.

Once each edge in the road network has been allocated a cost for each year of the analysis, the cost in a given year of travelling on a route $R_{i,j}$, denoted $\tau_{i,j,t}$, between counties i and j is simply the sum of the costs associated with each edge in $R_{i,j}$.

$$\tau_{i,j,t} = \sum_{e \in R_{i,j,t}} (T_{e,t} + D_{e,t}) \tag{3}$$

The minimal cost of travelling from county i to county j in year t, $\tau^*_{i,j,t}$, can be found using Dijkstra's algorithm. Specifically, Dijkstra's algorithm searches over the space of all possible routes $R_{i,j}$ and returns the route associated with the lowest cost.

$$\tau^*_{i,j,t} = \min_{R_{i,j}} \tau_{i,j,t} \tag{4}$$

It is important to point out that given the large size of the NHPN dataset, the road network is simplified prior to running Dijkstra's algorithm, in order to keep the problem tractable. Details of this procedure, which does not affect the optimality of the solution found by Dijkstra's algorithm, are provided in appendix B.

4 Stylised facts of transport cost determinants

Before presenting the county-pair transportation costs obtained using Dijkstra's algorithm, it is important to present and discuss the road network and reference costs that enter the analysis. Because these new and large datasets underpin the temporal and geographical variation of the minimal county-pair transport costs obtained with Dijkstra's algorithm, it is important that we check that on aggregate they are consistent with known stylised facts.

4.1 The time-evolution of the IHS

The five digitized interstate highway maps for the years 1960, 1970, 1980, 1990 and 2000, are presented in Figure 1. The digitized maps shows that the interstate highways were often built as disjointed highway segments that were gradually linked together, which is consistent with the fact the creation of the IHS was mainly an improvement to selected roads from the existing road network. In order to ensure the accuracy of this time-evolution of the network, we specifically checked the sequence of maps against the data digitized by Baum-Snow (2007) on the total number of miles of interstate highways open to public in each year and each county.

The mileage of the IHS, presented in Figure 2(a) with a breakdown by regions in Table 2, shows a substantial temporal and regional variation. The largest increase in the mileage was in the 1960s, followed by a steep decline and eventually petering out of the construction activities in the 1990s and 2000s. Indeed, fifty percent of the total IHS mileage was built by 1970, in the first fourteen years of the

Table 2: Mileage of the Interstate Highway System by Decades

	1956-1960	1961-1970	1971-1980	1981-1990	1991-2000	2001-2010	1956-2010
			IHS mileage built				*Total*
New England	464	933	348	103	3	12	1,863
Middle Atlantic	1,220	1,418	541	218	209	129	3,735
East North Central	1,225	3,440	1,477	314	297	17	6,770
West North Central	810	2,955	1,347	311	3	2	5,428
South Atlantic	320	3,091	2,133	774	336	212	6,866
East South Central	93	2,200	827	244	22	45	3,431
West South Central	952	2,773	1,393	389	66	0	5,573
Mountain	706	4,009	1,993	527	25	10	7,270
Pacific	755	2,045	835	225	31	1	3,892
US	*6,545*	*22,864*	*10,894*	*3,105*	*992*	*428*	*44,828*
			Share in IHS mileage built				*Final*
New England	7.1	4.1	3.2	3.3	0.3	2.9	4.2
Middle Atlantic	18.6	6.2	5.0	7.0	21.1	30.1	8.3
East North Central	18.7	15.0	13.6	10.1	29.9	4.0	15.1
West North Central	12.4	12.9	12.4	10.0	0.3	0.4	12.1
South Atlantic	4.9	13.5	19.6	24.9	33.9	49.5	15.3
East South Central	1.4	9.6	7.6	7.9	2.2	10.6	7.7
West South Central	14.5	12.1	12.8	12.5	6.7	0.1	12.4
Mountain	10.8	17.5	18.3	17.0	2.5	2.4	16.2
Pacific	11.5	8.9	7.7	7.2	3.1	0.2	8.7
US	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*

Sources for 1956-2000: PR-511, Baum-Snow (2007).
2001: `https://www.fhwa.dot.gov/ohim/hs01/hm41.htm`
2010: `https://www.fhwa.dot.gov/policyinformation/statistics/2010/hm20.cfm`

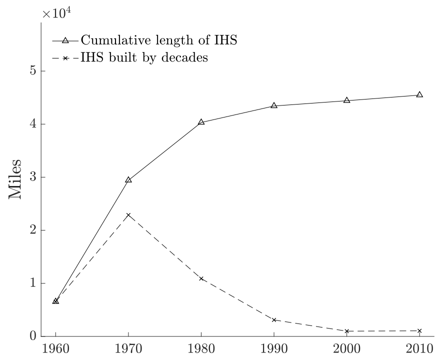
system. Geographically, the construction was concentrated first in the West, Middle Atlantic region and Midwest. This has changed over time as the share of the interstate highways built in the South Atlantic region was increasing by each decade. Upon its completion, the distribution of IHS across regions was roughly bi-modal: five regions have roughly similar share of interstate highways ranging from twelve to sixteen percent while the other four share between four and eight percent.

4.2 Fuel efficiency, fuel price and distance-related costs

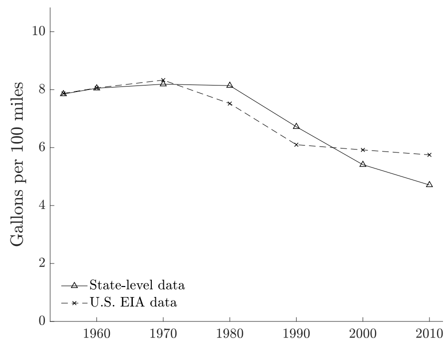
Table 3 presents the regional variation in distance related costs and its determinants over the period of analysis. This reveals that US-wide distance-related costs declined from \$15.94 to \$10.97 per 100 miles in 2010 \$US between 1955 and 2010, a 31% drop in real terms. This pattern holds for all US regions and was driven largely by the decline in fuel consumption.

Both determinants of distance costs were checked against the stylised facts seen in aggregate data. Figure 2(b) compares the average of the state-level fuel consumption calculated from the U.S. Highway Statistics and used in our analysis against the overall U.S. fuel consumption as reported by the U.S. Energy Information Administration. We see a good agreement between the series, and a similar pattern of a stable fuel consumption until the 1970s, followed by a decline until 2010. Figure 2(c) similarly compares average of the state-level data used in our analysis to the annual time series of retail price and tank-trunk prices, all expressed in 2010 \$US, and again we see that the decadal state-level averages

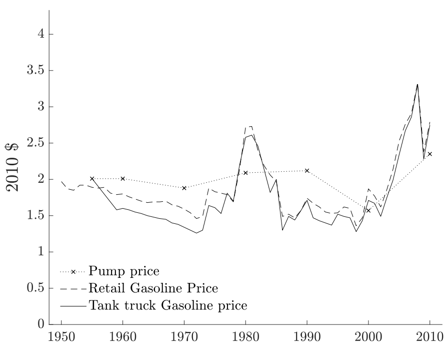
Figure 2: Determinants of distance costs



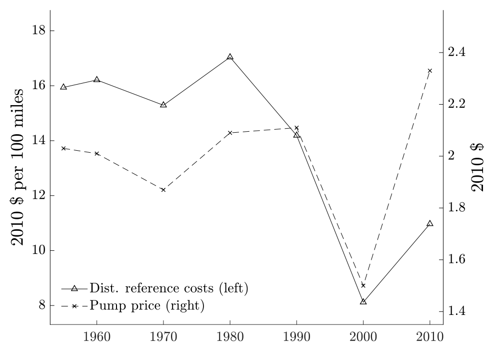
(a) Length of the IHS by decade



(b) Fuel Consumption



(c) Gasoline Prices



(d) Distance Reference Cost and Gasoline Price

Table 3: Distance Reference Costs

	1955	1960	1970	1980	1990	2000	2010	1955-2010
			Fuel efficiency (Gallons per 100 miles)					
New England	7.84	7.89	7.94	7.63	6.35	5.41	5.00	-36%
Middle Atlantic	7.70	7.34	7.99	7.94	6.64	5.50	4.88	-37%
East North Central	7.75	8.00	8.09	8.17	6.77	5.42	4.67	-40%
West North Central	7.84	8.52	8.43	8.61	7.16	5.65	4.75	-39%
South Atlantic	7.86	7.95	8.04	7.77	6.64	5.37	4.58	-42%
East South Central	8.24	8.55	8.50	8.04	7.09	5.30	4.40	-47%
West South Central	8.31	8.66	8.90	8.94	6.98	5.54	4.93	-41%
Mountain	8.05	8.33	8.37	8.32	6.84	5.43	4.52	-44%
Pacific	7.58	8.11	7.88	8.00	6.27	5.24	4.76	-37%
US	*7.85*	*8.05*	*8.19*	*8.14*	*6.72*	*5.41*	*4.71*	*-40%*
			Pump price of gasoline (2010 US$ per gallon)					
New England	1.89	1.89	1.84	2.13	2.20	1.69	2.42	28%
Middle Atlantic	1.87	1.85	1.83	2.13	2.08	1.58	2.35	25%
East North Central	2.07	2.03	1.91	2.10	2.12	1.53	2.35	13%
West North Central	2.05	2.03	1.86	2.10	2.13	1.53	2.33	14%
South Atlantic	2.00	1.96	1.85	2.07	2.14	1.51	2.30	15%
East South Central	2.05	2.03	1.91	2.09	2.10	1.48	2.26	11%
West South Central	1.92	1.92	1.78	2.01	2.06	1.50	2.25	17%
Mountain	2.22	2.18	1.97	2.09	2.13	1.61	2.40	8%
Pacific	2.06	2.19	1.97	2.12	2.09	1.70	2.52	22%
US	*2.03*	*2.01*	*1.87*	*2.09*	*2.11*	*1.50*	*2.33*	*15%*
			Distance reference costs (2010 US$ per 100 miles)					
New England	14.84	14.89	14.61	16.22	13.97	9.12	12.08	-19%
Middle Atlantic	14.43	13.62	14.64	16.96	13.80	8.68	11.45	-21%
East North Central	16.02	16.21	15.47	17.19	14.32	8.28	10.96	-32%
West North Central	16.09	17.32	15.68	18.06	15.26	8.65	11.07	-31%
South Atlantic	15.71	15.58	14.89	16.10	14.20	8.13	10.52	-33%
East South Central	16.87	17.36	16.20	16.77	14.89	7.84	9.96	-41%
West South Central	15.95	16.63	15.89	17.96	14.39	8.33	11.08	-31%
Mountain	17.84	18.20	16.51	17.40	14.57	8.76	10.84	-39%
Pacific	15.62	17.79	15.56	16.98	13.12	8.88	11.96	-23%
US	*15.94*	*16.21*	*15.29*	*17.04*	*14.19*	*8.12*	*10.97*	*-31%*

Source: Detailed description of sources is in Appendix A.2.
Note: Distance reference costs are calculated by multiplying fuel consumption by gasoline price.

follow the same pattern as the aggregate data. Figure 2(d) finally compares the average fuel price and average distance reference costs from our dataset, to show that while the two series co-move, there is a clear downward trend in the distance costs compared to the fuel prices over decades, which stems from the increase in fuel efficiency.

In addition to the time-variation, Table 3 reveals considerable regional variation for both components of distance costs. Regarding fuel consumption, most regions experienced increases until 1970. However there is a considerable regional variation in the 1970s, where some regions continued to see increases in fuel consumption, while many, especially in the South and West, saw it declines. By the 1980s, fuel consumption was declining in all US regions. Regarding fuel prices, Table 3 shows that there is considerable spatial variation in addition to the well-known time trends. Indeed, even though real gasoline prices increased in all regions between 1955 and 2010, the Mountain region experienced the smallest whilst New England the largest increase. Interestingly, in the 1970s – the decade of the oil shocks – the Western

Table 4: Average Speed by Type of Roads, 1955-2010

	1955	1960	1970	1980	1990	2000	2010
				Urban			
(1) Interstate	-	50.04	56.94	56.54	58.64	57.82	56.66
(2) Other	46.42	48.22	54.87	54.48	56.41	55.00	53.90
				Rural			
(3) Interstate	-	51.74	58.88	58.53	61.16	61.72	61.95
(4) Other	48.45	49.98	56.87	56.53	56.74	57.20	57.42
				All roads			
	47.43	49.99	56.89	56.52	58.24	57.94	57.48

Source: Detailed description of sources is in Appendix A.4.
Notes: Speeds in miles per hour. 1960: average of 1958 and 1959; 1970: average of 1967 and 1969; 1980: average of 1977 and 1979; 1990: average of 1987 and 1989.

and Southern oil-producing regions saw the smallest increases in gasoline prices, whilst the North-East saw the largest.

Because the overall distance-reference costs are determined by both components, the end result is that they also exhibit significant time and regional variation. Although they declined substantially between 1955 and 2010 across all regions, a noteworthy regional variation emerges: the Mountain and East-South Central regions experienced the largest decrease in distance reference cost (about forty percent) whilst New England and Middle Atlantic saw the smallest (about 18% and 20% respectively).

4.3 Driving speeds and time-related costs

The driving speed data we have collected forms a central part of the analysis, as it allows us to quantify the qualitative improvement brought on by the IHS relative to other types of roads. As discussed in the data section, we have collected data on average driving speeds on four types of roads, for all forty-eight contiguous border US states, over the benchmark years of 1955, 1960, 1970, 1980, 1990, 2000, and 2010, resulting in 1,344 data points. Table 4 and 5 summarise this driving speed data by type of road, and by region respectively. They confirm that driving speeds are, on average, higher on interstate highways than on non-interstate roads, and higher in rural than urban areas. Average speeds increased by about 21% between 1955 and 2010 and, as was the case for the time-reference costs, exhibit a considerable variation over time and geographical regions.

Regarding the time variation, average speeds increased between 1960 and 1970, and then again between 1980 and 1990, but remained stable between 1970 and 1980, and declined slightly after 1990. The stability of average speeds in the 1970s conceals the fact that they dropped quite dramatically in the early 1970s, following the oil shocks, and it took the whole decade to return to the levels seen in 1970 (Highway Statistics 1979). The decline after 1990 was driven by the lower speeds in urban areas, both on the interstates and other roads, reflecting increasing congestion.

Two facts stand out from the spatial breakdown of average driving speeds: first, the increase in speeds

Table 5: Time Reference Costs

	1955	1960	1970	1980	1990	2000	2010	1955-2010
				Average driving speed (miles per hour)				
New England	43.65	50.31	57.25	56.76	58.35	56.96	56.43	29%
Middle Atlantic	45.22	48.68	55.40	55.71	58.35	57.98	57.20	26%
East North Central	47.73	51.28	58.36	57.51	59.07	58.88	58.86	23%
West North Central	47.50	50.62	57.60	56.58	58.41	57.63	57.52	21%
South Atlantic	46.98	50.04	56.94	56.56	57.75	58.06	58.11	24%
East South Central	49.16	49.04	55.81	55.92	58.05	58.16	57.86	18%
West South Central	47.73	50.44	57.40	57.26	58.55	59.52	59.17	24%
Mountain	50.34	49.75	56.61	56.05	57.30	57.98	58.18	16%
Pacific	48.58	49.77	56.63	56.32	58.31	56.24	54.00	11%
US	*47.43*	*49.99*	*56.89*	*56.52*	*58.24*	*57.94*	*57.48*	*21%*
				Hourly earnings of truck drivers (2010 US$)				
US	16.6	18.3	21.2	25.5	21.8	18.8	18.2	10%
				Time reference costs (2010 US$ per 100 miles)				
New England	38.02	36.38	37.02	44.91	37.46	32.97	32.29	-15%
Middle Atlantic	36.70	37.63	38.29	45.79	37.48	32.42	31.91	-13%
East North Central	34.77	35.69	36.32	44.32	37.02	31.88	30.95	-11%
West North Central	34.94	36.17	36.80	45.07	37.46	32.61	31.73	-9%
South Atlantic	35.33	36.61	37.25	45.10	37.88	32.37	31.40	-11%
East South Central	33.75	37.43	38.09	45.72	37.73	32.36	31.57	-6%
West South Central	34.77	36.30	36.94	44.52	37.36	31.58	30.82	-11%
Mountain	32.97	36.83	37.48	45.53	38.14	32.38	31.33	-5%
Pacific	34.17	36.79	37.43	45.27	37.53	33.41	33.75	-1%
US	*34.99*	*36.63*	*37.27*	*45.12*	*37.55*	*32.43*	*31.72*	*-9%*

Source: Detailed description of sources is in Appendix A.2 and A.3.
Notes: Speed: 1960: average of 1958 and 1959; 1970: average of 1967 and 1969; 1980: average of 1977 and 1979; 1990: average of 1987 and 1989, Wages: except for 1955 and 1960, hourly earning are decadal averages, e.g. 1970 earnings are the average of 1961-1970.
Earnings in 1960 are the average of 1958-1960.
Time reference costs = (earning of truckers/speed)×100

between 1955 and 2010 was unequal across regions, with the Pacific and Mountain regions experiencing the lowest increases relative to other regions. Second, the decrease in average speeds between 2000 and 2010 was confined mostly to the states in Middle Atlantic and Pacific regions. Given that the states in Middle Atlantic and Pacific regions contain New York City, Los Angeles, and San Francisco, i.e. some of the most urbanized metropolitan areas of the country, this is consistent with the fact, presented in Table 4, that the decrease in US-wide average speeds over the period was due to lower speeds in urban areas.

As previously explained, the time-reference costs for a given distance are obtained by dividing the hourly earnings of truck drivers, shown in the middle of Table 5, by average driving speed. Table 5 also presents the regional breakdown of these time reference by region and decade, and reveals an inverse U-shape pattern. Even though the costs declined by 9% on average between 1955 and 2010, they increased significantly between 1960 and 1980, and then declined strongly by 2010. It is important to note that the biggest increase was in the 1970s, due to a 20% increase in the real earnings of truckers and relatively stable driving speed in that decade. The decline of the time-reference costs after 1980 is closely related to the decline of real earnings of truckers, a trend well noticed in the literature and attributed, among other things, to the deregulation beginning in the late 1970s and the subsequent rise of non-union

Table 6: Total Transport Reference Costs per 100 miles, 2010 US$

	1955	1960	1970	1980	1990	2000	2010
				Urban			
(1) Interstate	-	52.79	52.51	62.11	51.45	40.56	43.08
(2) Other	51.68	54.16	53.91	63.81	52.92	42.22	44.72
				Rural			
(3) Interstate	-	51.58	51.28	60.58	49.92	38.51	40.34
(4) Other	50.18	52.83	52.55	62.12	52.70	40.91	42.65
				All roads			
	50.93	52.84	52.56	62.16	51.75	40.55	42.70

Source: Detailed description of sources is in Appendix A.2 to A.4.

trucking providers (Hirsch, 1988; Rose, 1987; Belman and Monaco, 2001). Though still decreasing, the first decade of the millennium witnessed a slower decline of time-reference costs, a development driven largely by declining speed in urban areas.

4.4 Total transport reference costs

The total reference transport costs were calculated by summing distance-reference and time-reference costs for forty-eight contiguous states, four types of roads, and all benchmark years. Tables 6, and 7 present the breakdown by type of road and by US region respectively. Table 6 shows a 16% decrease between 1955 and 2010, from $50.9 to $42.7 per 100 miles in 2010 $US, while exhibiting an inverse U-shape pattern peaking in 1980 driven by the real wages of truckers and higher fuel prices. The breakdown by the type of roads confirms that it was cheaper to drive on rural roads than in urban areas, and on the IHS compared to non-interstates, which is explained by the distribution of driving speeds. Transport costs increased considerably in the 1970s, reflecting the increase in gasoline prices, stagnant driving speeds and high earning labour costs. Costs dropped following this, caused by the sharp drop in the fuel consumption and the fall in labour costs. The first decade of the millennium witnessed a slight increase in the transport costs driven entire by the increasing cost in the urban areas, and especially on the interstate roads.

The regional breakdown in Table 7 confirms that all regions saw the highest costs in 1980. The subsequent decline varied by regions, but the steepest decline was confined to Mountain region and southern parts of the US. Not surprisingly, the costs declined least in New England, Middle Atlantic and Pacific region, which is consistent with the decline in driving speed as seen in Table 5.

Table 7: Total Transport Reference Costs per 100 miles by Regions, 2010 US$

	1955	1960	1970	1980	1990	2000	2010	1955-2010
New England	52.86	51.27	51.63	61.13	51.42	42.10	44.37	-16%
Middle Atlantic	51.13	51.25	52.92	62.75	51.28	41.09	43.36	-15%
East North Central	50.79	51.91	51.79	61.51	51.34	40.16	41.91	-17%
West North Central	51.03	53.49	52.49	63.12	52.72	41.26	42.80	-16%
South Atlantic	51.04	52.19	52.14	61.20	52.07	40.50	41.92	-18%
East South Central	50.63	54.79	54.28	62.49	52.62	40.21	41.53	-18%
West South Central	50.72	52.93	52.82	62.47	51.74	39.91	41.90	-17%
Mountain	50.81	55.04	53.98	62.94	52.71	41.14	42.18	-17%
Pacific	49.78	54.58	52.99	62.25	50.65	42.29	45.71	-8%
US	*50.93*	*52.84*	*52.56*	*62.16*	*51.75*	*40.55*	*42.70*	*-16%*

Source: Detailed description of sources is in Appendix A.2 to A.4.

5 The evolution of pairwise county transports costs in the USA, 1955-2010

With 3015 counties taken from the 48 contiguous US states, the output of the Dijkstra analysis is a set of 4,818,960 (3105 × 3104 × 1/2) distinct county-pair optimal routes $R^*_{i,j}$.[12] Because the output for each origin/destination pair records the entire route, we can also obtain the mileage travelled by type of roads in addition to the value of the minimal transport costs. The composition of the route is especially important since it allows us to calculate the share of distance travelled on the IHS and thus investigate a relationship between the county-pair transportation costs and the IHS share in section 6.1. It is important to point out that the transport costs calculated using this method correspond to the minimal marginal cost of a single trip between two county centroids. As a result, average costs for a given origin county discussed below are unweighted arithmetic means taken over all 3014 destination counties, and do not integrate the distribution of trips between counties. This would require data on the volume of trips carried out between pairs of counties in a given period, which is not available.

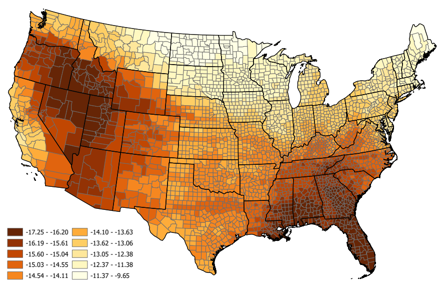
5.1 The spatial evolution of pairwise minimal transport costs

In order to illustrate the spatial evolution of the pairwise minimal transport costs $\tau^*_{i,j,t}$, we start by averaging them over all destination counties j for all the benchmark years, $\bar{\tau}^*_{i,y} = \sum_j \tau^*_{i,j,t}/(n-1)$ and then compute the percentage change between the benchmark years. Figure 3(a) presents the long-run change between 1955 and 2010, with the other panels presenting the changes decade by decade.

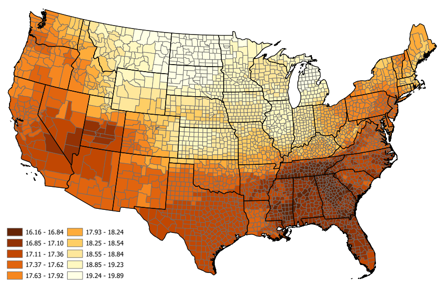
Two facts stand out from the long-run county-level changes presented in Figure 3(a). First, there is a widespread decline of the transportation costs across all US counties of between 9% and 17%. Second, there is a noticeable Northwest – Southeast gradient in which the largest decline happened in the Pacific northwest and Atlantic southeast, and the smallest in the North and Northeast of the country. We can thus distinguish three broad parts of the US: (i) a part with the lowest decline in the transportation

[12] Because a route is simply a set of connected edges in the network between two vertices, we have $R^*_{i,j,t} = R^*_{j,i,t}$, and $\tau^*_{i,j,t} = \tau^*_{j,i,t}$. Because counties are represented by a centroid, we also have $\tau_{i,i,t} = 0$.

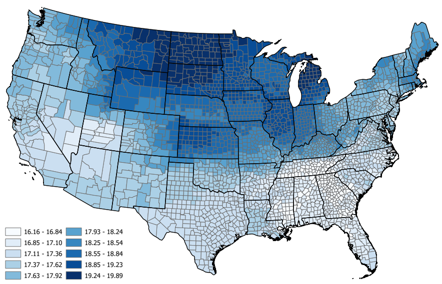
Figure 3: Percentage Change in County Average Real Transportation Costs



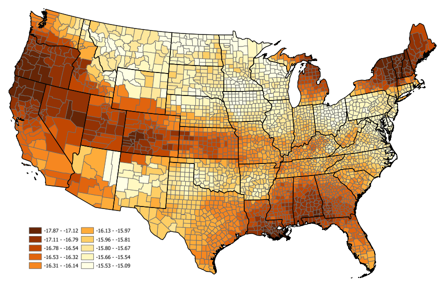
(a) 1955-2010



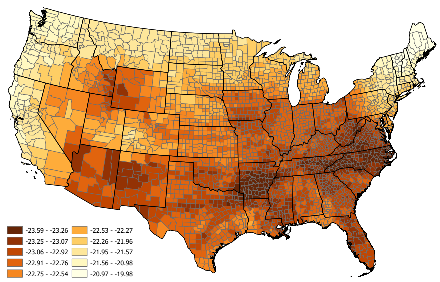
(b) 1960-1970



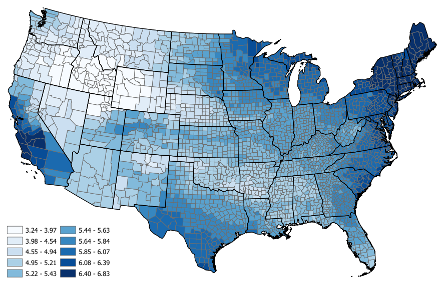
(c) 1970-1980



(d) 1980-1990



(e) 1990-2000



(f) 2000-2010

costs that includes the northern parts of the US and stretches from eastern Montana to Maine, (ii) a part with medium decline of transportation costs which includes middle regions of the US from Nebraska to Texas, most of the state of Washington, California, Kentucky and both Virginias, (iii) and a part with the highest decline of the transportation costs which includes the south east, most of the Rocky Mountains, and western part of Arizona. It is interesting to observe that the two areas that see the largest reductions are not spatially joined.

The long-term spatial changes captured in Figure 3(a) conceal variation across decades. As shown in the section 4, the reference costs used in the analysis, i.e. average driving speeds, fuel prices and fuel consumption, vary both across time and across US regions. Furthermore, the IHS was being built over the period of our analysis, and the pace construction differed substantially across US states, as seen in Figure 1. Therefore, it is also instructive to examine the percentage change in the county-level real transportation costs for each decade.

For the 1960-1970 period, Figure 3(b) shows that the largest decline occurred in the Western parts of the US, forming a horseshoe pattern starting in Arizona, going through California and ending in the western parts of Montana. The smallest decline is observed in New England and Middle Atlantic. The observed pattern is consistent with the transportation costs in Table 7: The Pacific coast and Mountain regions saw the largest decadal decline of transport reference costs while New England and Middle Atlantic were among the lowest. It is important to note that the magnitude of the percentage decline in transportation costs is small, about 3% percent at most. This is due to the fact that despite falls in fuel consumption and petrol prices, and the rapid progression of construction on the IHS progressed, the wages of truck drivers actually increased, as shown in Table 5, offsetting the decline of the other costs.

Figure 3(c) reveals that the 1970s saw an unequivocal and substantial increase in costs of between 16% to 20%. The largest increase was concentrated mostly in the northern states of Montana, the Dakotas, Minnesota, and Michigan, and to a lesser extent Illinois and Kansas. The smallest increase was in the Southern states, especially in the Southeast. The increase in fuel prices due to the two oil shocks was certainly one of the main factors. However, we need to be careful to attribute the visible increase in costs only to this factor. Table 5 shows that the time reference costs also increased by about twenty percent between 1970 and 1980, driven by an almost twenty percent increase in the real earnings of truck drivers with essentially unchanged driving speeds.

The 1980s was the first decade to see a substantial decline in average county transportation costs, with Figure 3(d) showing a 15-20% reduction across the US. Regionally, New England, Middle Atlantic and Michigan experienced among the largest decline in that decade, along with northern California, Oregon, Nevada, Utah, Colorado, and the southern states of Louisiana, Mississippi, and Alabama. This reduction was driven by a roll-back of the high fuel costs seen during the 1970s, the decline of the real earning of truckers, and finally by the near completion of the IHS, offering higher driving speeds across the continent.

In particular, this is visible on the map through the clear pattern of declining transportation costs in the corridor along the routes of I-5 from California to Washington, I-80 from California to Nebraska, and I-82 from Oregon to Idaho. This reduction in transport costs continued in the 1990s, as shown in Figure 3(e), with a further reduction of 20-23%. The regional pattern of this is consistent with that in Table 7 which shows the regional distribution of the total transport reference costs, in which reductions in underlying costs are indeed concentrated in the Southern states.

Finally, Figure 3(f) shows the percentage change in county-level transportation costs in the first decade of the new millennium. Here we see a modest increase, ranging from about 3-7%, with the Northeastern regions, coastal California and Southern Texas, seeing the highest increase. Again, this is consistent with the underlying increases in the total transport reference costs in Table 6, with distance-reference costs increasing slightly due to fuel prices and time-reference costs increasing following the fall in average driving speeds in urban areas. This is visible in the highly urbanized metropolitan in Figure 3(f): The eastern seaboard from New England to the Carolinas, and a strip of counties from San Francisco to San Diego.

5.2 Shift - Share analysis: the contribution of transport cost determinants

Whilst the construction of the IHS explicitly aimed at increasing driving speeds, and was one of the largest post-WWII infrastructure projects in the United States, it was not the only factor contributing to the decrease in the transportation costs over the past half a century. As discussed in section 4, changes in fuel consumption, fuel prices, and wages of truck drivers were also important factors in the evolution of transport costs. Therefore, in this section, we provide a first pass at decomposing the contribution of the distance-related and time-related costs to the overall changes the transportation costs, as well as the effect of improvements in the road network itself. As our minimal cost methodology follows Combes and Lafourcade (2005), we also use their shift-share approach to analyse these impacts. The results of the analysis are provided in Table 8.

Panel A shows the share of the two components in the total US transportation costs. These are calculated from the cheapest pairwise county-level routes obtained from the Dijkstra analysis, by averaging the distance travelled and time spent across routes and multiplying the resulting averages by the corresponding reference costs: distance costs from Table 3 and hourly wages from Table 5.[13] We can see that the general trend up to 2000 is a fall in the share of the distance costs, and a corresponding increase in the share of time costs. This is consistent with the pattern seen in Tables 3 and 5 where both reference costs fell over time, though distance costs saw a larger reduction. The 2000-2010 period saw a reversion of the trend, with the share of distance costs increasing, a result of increasing fuel prices over

[13]Note that it is possible to first calculate the distance cost and time cost share for each individual optimal route, and then average across routes. The reason we calculate these shares by first calculating average distance/time, and then multiplying by reference costs is so that we can calculate the impact of the road network in Panel D. Both calculations provide very similar shares.

Table 8: Shift-Share Analysis of the Sources of Transportation Costs 1955-2010.

	Panel A: Percentage in total transportation costs						
	1955	*1960*	*1970*	*1980*	*1990*	*2000*	*2010*
Distance cost (1)	31.95	30.27	28.71	27.02	26.85	19.49	24.96
Time cost (2)	68.05	69.73	71.29	72.98	73.15	80.51	75.04
	Panel B: Percentage change in average costs						
	1955-1960	*1960-1970*	*1970-1980*	*1980-1990*	*1990-2000*	*2000-2010*	*1955-2010*
Distance cost (3)	1.70	-5.65	11.40	-16.63	-42.70	35.15	-30.99
Time cost (4)	10.02	1.70	21.16	-15.89	-13.14	-1.66	-2.60
Total cost (5)	7.36	-0.52	18.36	-16.09	-21.08	5.51	-11.67
	Panel C: Contribution to the changes in transportation costs (%)						
	1955-1960	*1960-1970*	*1970-1980*	*1980-1990*	*1990-2000*	*2000-2010*	*1955-2010*
Distance cost (6)	0.54	-1.71	3.27	-4.49	-11.46	6.85	-9.90
Time cost (7)	6.82	1.19	15.09	-11.60	-9.61	-1.34	-1.77
Total cost (8)	7.36	-0.52	18.36	-16.09	-21.08	5.51	-11.67
	Panel D: Contribution to the changes in transportation costs less IHS (%)						
	1955-1960	*1960-1970*	*1970-1980*	*1980-1990*	*1990-2000*	*2000-2010*	*1955-2010*
Distance cost (9)	0.54	-1.71	3.28	-4.51	-11.48	6.85	-9.95
Time cost (10)	7.03	11.00	14.45	-10.42	-10.35	-2.43	6.57
Total cost (11)	7.57	9.29	17.72	-14.93	-21.83	4.42	-3.39
	Panel E: Contribution of IHS (%)						
	1955-1960	*1960-1970*	*1970-1980*	*1980-1990*	*1990-2000*	*2000-2010*	*1955-2010*
Difference (8)-(11)	-0.21	-9.81	0.64	-1.16	0.76	1.10	-8.28

Source: Detailed description of sources is in Appendix A.2 to A.4.

that decade.

Panel B shows the percentage change in the two cost components over the benchmark years. These are multiplied to the shares in Panel A to obtain the contribution of the two cost components to the changes in the total transportation costs, shown in Panel C. This confirms that over the 1955 to 2010 period, falling distance-related costs were the driving force behind the overall decline in the transportation costs, with the fall time-related costs providing a modest boost, mainly in the later years. There is a considerable decade-by-decade variation, however, again consistent with the discussion in Section 4. The 1955-1980 periods were driven by high wage costs, although rising fuel consumption costs also played a significant role in 1970-1980 and then again in 2000-2010. The decrease in total transport costs was largest post 1980: in 1980-1990 mostly due to the decline in wages, 1990-2000 mainly because of the declining fuel consumption.

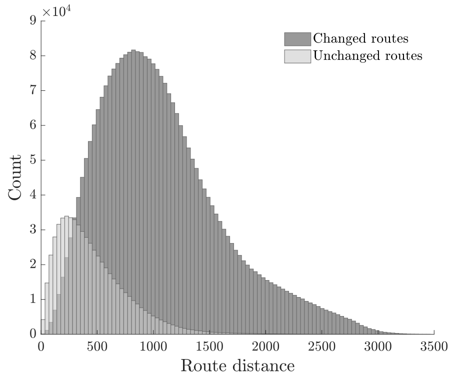
The shift-share analysis can also be used to get a first measurement of the aggregate effect of road network improvements on transportation costs, including the IHS. This is done in Panel D, which carries out the same calculations as in Panels A-C, except that it uses the average distances and times of the year 1955, as is done in Combes and Lafourcade (2005). In doing so we freeze the road network in its 1955, pre-IHS state, thus 'turning-off' the contribution of the road network, especially the higher driving speeds. The specific contribution of the road network in subsequent years is then obtained by subtracting these frozen network changes in transport costs from the ones that include the full time-evolution of the road network. This is shown in Panel E, which confirms that the aggregate contribution of the improved road network, including the IHS, to the decrease the transportation costs between 1955 and 2010 is about 8%. The 1960s and 1980s saw the largest decreases in transport costs, while 1970s, 1990s, and 2000s saw the contribution of the road network actually increase of the transportation costs, although these effects are small. In the 1970s, this is explained by the fact that driving speeds actually dropped following the oil shock in 1973. The situation in the 1990s and 2000s is similar in the sense that it was the declining speed on urban interstates, as we saw in Table 4, which caused the transportation costs on the IHS to increase.

The facts that improvements in the road network seem to sometimes to increase the overall transportation costs, albeit by a small amount, might be counterintuitive and is a result of the assumption used in the shift-share analysis to obtain the counterfactual. Specifically, the impact of time and distance costs net of the road network in a given year years were calculated by assuming that the average route time and distance is given by the 1955 values, which raises two problems. First, it means that the variation in the distance and time reference costs did not affect the distance travelled or time spent in the counterfactual, as these are exogenously kept constant, which means that the routes are no longer optimal in the Dijkstra sense and explains why the contribution of the road network improvements to transport costs is occasionally positive. Secondly, the reliance on 1955 travel times/distances for the counterfactual analysis means that it is not possible to separate the specific contribution of the IHS from the more general improvement in average driving speeds over the entire network. Therefore, in the next section, we focus specifically on the impact of the IHS by finding the cost-minimizing routes with and without the interstates but, unlike the shift-share analysis, taking into account changes in the distance and time reference costs.

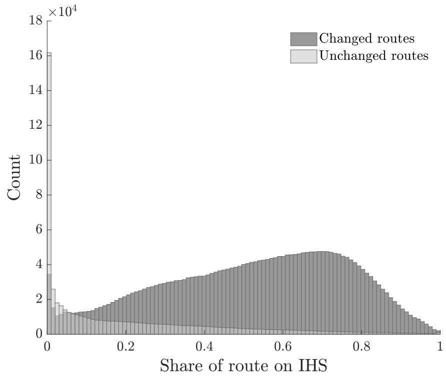
6 The impact of the Interstate Highway System

Having examined the time and geographical evolution of the overall pattern of county-level transport costs, we now move to examining the role of the IHS. Specifically, we investigate a relationship between the distance travelled from county i to j and the share of the trip carried out on the IHS, and the effect of the IHS on the county-pairwise transport costs. Here, we take advantage of the fact that the Dijkstra algorithm returns the entire cost-minimizing route $R^*_{i,j}$ between two counties i and j, enabling us to calculate the percentage of that route taken on the US interstates. Figure 4(a) and (b) present the distribution of route distances and IHS shares. One interesting aspect is that a significant minority of routes do not change over time, in the sense that the cost-minimising route travelled in 1955 remains optimal in 2000. The first two panels of Figure 4 reveal, however, that these routes tend to be much shorter than average, and the great majority of them have a zero or very low interstate share. This is

Figure 4: County-Pair Distance and IHS Share

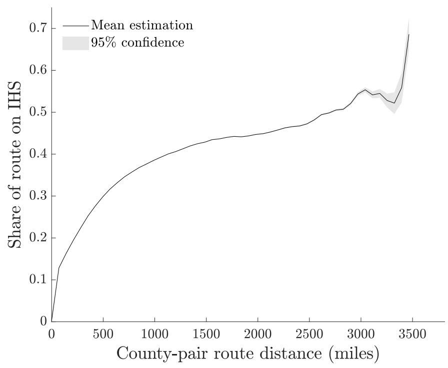


(a) Distribution of route distances, 2000



(b) Distribution of IHS share, 2000

Note: 'Unchanged routes' refers to pairwise Dijkstra optimal routes that are the same in 1955 and 2000.



(c) Polynomial regression, 1960-2010

consistent with local trips to neighbouring counties or states, which would not necessarily following the IHS, particularly in the western part of the country where the IHS network is sparser.

6.1 Minimal costs, route distance and interstate share

In order to assess if the minimal pairwise county transport costs are decreasing in the share of the optimal route travelled on the IHS, we estimate the following regression equation:

$$\log \tau^*_{i,j,t} = \beta_1 \log d_{i,j,t} + \beta_2 IHS_{i,j,t} + \omega T_{i,j,t} + X'_{i,j,t}\delta + \gamma_t + \epsilon_{i,j,t} \qquad (5)$$

where t indexes the year, i and j index the origin and destination counties, $\log \tau^*_{i,j,t}$ is the log of the county pair-wise minimal transport costs, $\log d_{i,j,t}$ is the log of minimal cost route's distance between counties i and j, $IHS_{i,j,t}$ the is percentage of the route between county i and j taken on the US interstates, $T_{i,j,t}$ is a linear time trend, γ_t are time fixed effects, $X'_{i,j,t}$ are additional controls and $\epsilon_{i,j,t}$ is the error term. Since the main coefficient of interest is β_2, we estimate this equation for the period 1960-2010 excluding 1955 when IHS was not built yet. We include time fixed effects and county-pair time-trends to control for time-related factors which may affect the transportation costs. Indeed, as we discussed in

Table 9: County-Pairwise Transport Costs and Share of Route on IHS, 1960-2010.

	Pooled OLS I	Panel Fixed-Effect II
log (Distance)	0.998***	0.5282***
	(1.06e-05)	(0.001)
% Route on IHS	-0.0003***	-0.0004***
	(2.34e-07)	(4.26e-07)
Time trend	-2.54e-07***	-1.52e-07***
	(1.68e-06)	(1.66e-06)
Year FE	YES	YES
Origin State FE	YES	YES
Destination State FE	YES	YES
Observations	28,884,696	28,913,760

Note: The dependent variable is log of county-pairwise transport costs per 100 miles in 2010 US$ in 1960, 1970, 1980, 1990, 2000, and 2010. The percentage of interstates is calculated as the percentage of IHS miles taken on the optimal route between counties i and j. Time trend is a linear trend for each county-pair. Robust standard errors are reported in parenthesis.

the Section V, there has been a steady decline of distance-related costs as well as driving speed which is likely correlated with the spread of the interstate highways as well as oils shocks in the 1970s.

We initially estimate equation (5) with pooled OLS, including state-origin and state-destination dummies in the vector $X'_{i,j,t}$ to control for any state-level effects that might be correlated with transportation costs, such as driving speed limits. We then estimate equation (5) with a fixed-effect panel data estimator to also control for the unobserved county-pairwise effects. Table 9 presents the results for both estimations. All variables are highly significant, due to the large number of observations, and the controls have expected sign in both columns: the negative time trend is consistent with the fall in the transport costs previously discussed and distance is positively related to transport costs. In the pooled OLS case, distance is unit-elastic, reflecting the fact that both cost components (1) and (2) are linearly related to distance at the level of each edge. In the fixed effect model, however, this elasticity falls significantly below one. In other words, deviations from the average distance between two counties in a given year are associated to proportionally smaller changes in minimal costs. This effect is the result of optimisation by the agents, who can trade off time-related and distance-related costs when picking optimal route, thus loosening the one-to-one link between distance and cost that exists at the level of an edge.[14]

Crucially, the parameter on the IHS share is negative, confirming that all other things equal the larger the share of the route between two counties taken on the US interstates, the lower the transport costs. The log-linear specification with respect to the IHS share means that we need to multiply the estimated value of β_2 by 100 to interpret it in terms of percentages. The panel fixed effect estimate suggests an economic effect of 0.04, i.e. increasing the IHS share on a route by one percentage point decreases transport costs

[14]This is consistent with the every day experience that a longer route might be cheaper overall, i.e. we might choose a less direct route between origin and destination because it allows us to join a high speed highway, leading to a quicker (and cheaper) trip, despite the route itself being longer. This is also why it is important to optimise the choice of route over both distance and time costs, especially in a counterfactual analysis.

by 0.04%. While this effect might seem low, calculating the economic impact for the average IHS share, which is 36% percent between 1960 and 2010, yields a 1.44% decrease of transport cost. For 2010 the IHS share is 46% which leads to an average decrease in costs of 1.84%. A further discussion of the magnitude of the cost savings brought by the IHS will be discussed after the counterfactual analysis in section 6.2.

The IHS was built not only to facilitate high speed travel but also to facilitate travel on long-distance, cross-continental journeys. Therefore, our second hypothesis of interest is that we expect a positive relationship between the county-pair distance and the share of the route US taken on the US interstates. This is investigated by estimating the following non-parametric regression:

$$IHS_{i,j,t} = f(d_{i,j,t}) + \epsilon_{i,j,t} \tag{6}$$

where $d_{i,j,t}$ is the distance in miles of the cost-minimising route between county i and j in each year t, and $IHS_{i,j,t}$ is the share of that distance travelled on the US interstates. We use a local polynomial regression of degree one with an Epanechnikov kernel, using cross-validation to estimate the optimal bandwidth. Figure 4(c) presents the result, and shows that the share of routes taken on the US interstates indeed increases monotonically with the distance, but the rate is not uniform. The share of a trip taken on the IHS initially increases quite rapidly as journey distance increases to about 1000 miles, then grows more slowly from that point onwards, and again increases rapidly for distances above 5000 miles. At the lower end of the trip distance distribution, this confirms that despite the IHS representing only 10% of the road network, using it for a significant portion of a journey very rapidly becomes a valid cost-minimising strategy. At the higher end of the distance spectrum, as trips start spanning the width of the continent, this suggests that the IHS is indeed fulfilling its design requirements of facilitating long distance travel.

6.2 Counterfactual analysis: a world without the IHS

The counterfactual analysis is based on the comparison of the optimal county-pair transport costs presented above with a second set of optimal county-pair costs $\tau^c_{i,j,t}$, also obtained using the Dijkstra algorithm on the same time and distance reference costs, but using a road network which does not contain the interstate highways. This counterfactual network is obtained by treating interstate roads as if they were non-interstate roads, thus removing their speed advantage, shown in Table 4. This counterfactual network reflects a world where the planned US interstates were never built, and the road network entering the Dijkstra algorithm remained a mix of state and US routes, but where improvements in road surfaces and motor vehicles occurred as normal. Unlike the shift-share analysis in Section 5.2, all the routes in the full counterfactual analysis are cost-minimizing routes found with Dijkstra's algorithm, and thus optimally take into account all changes in distance and time-reference costs over the period. This is particularly important in view of the discussion of the fixed effects estimate above, which show that

Table 10: Actual and Counterfactual County-Pair Transport Costs per Journey

	Average county-pair transport costs in 2010 $US							
	All Routes			Affected Routes Only			Average	
	with IHS	No IHS	$\overline{\%\Delta}$	with IHS	No IHS	$\overline{\%\Delta}$	IHS share	$P_{01}(\%\Delta)$
1955	492.7	492.7	-	0	0	-	-	-
1960	529.5	530.2	-0.15	569.8	570.7	-0.15	7.34	-1.31
1970	524.1	527.9	-0.72	564.0	568.2	-0.75	30.06	-3.00
1980	617.8	624.3	-1.04	664.7	671.9	-1.08	39.95	-4.03
1990	518.2	525.5	-1.41	557.3	565.5	-1.46	43.90	-3.90
2000	401.0	408.8	-1.95	431.3	440.0	-2.02	47.56	-5.81
2010	422.9	430.7	-1.84	454.9	463.6	-1.90	46.22	-5.46

Note: This table compares actual county-pair transport costs 2010 US$ with counterfactual calculated for the counterfactual scenario of no IHS.
'Affected routes only' restricts the analysis only to those routes that changed with the introduction of the IHS.
$P_{01}(\%\Delta)$ is the 1^{st} percentile of the distribution of percentage differences between the actual and counterfactual cases.

cost-minimising behaviour can loosen the otherwise unit-elastic relationship between distance and cost.

The aggregate effect of this analysis is presented in Table 10, and confirm that optimal counterfactual costs are higher than the actual ones, hence the presence of the IHS does reduce transport costs. Another important result is that the cost savings clearly increase between 1960 and 2000. In other words the IHS grows more effective as its mileage and the share of trips on it increases. The small drop for 2010 reflects the decrease in average driving speeds over that decade. Importantly, the magnitude of the cost reduction, discussed further below, is notably lower than the 8% of the shift-share analysis, which overestimated the effect by ignoring the significant improvement in driving speeds for non inter-state roads from 1955 to 2010.

In addition to better identifying the specific contribution of the IHS to the evolution of transport costs, this disaggregated counterfactual analysis allows us to look at the geographical variation in the impact of the IHS, shown in Figure 5. As was there case for Maps 2 and 3, we obtain both the actual and counterfactual average transport cost in an origin county i by averaging the pairwise transport cost $\tau^*_{i,j,t}$ and $\tau^c_{i,j,t}$ over all destinations j.

$$\overline{\Delta\tau_{i,t}} = \frac{\sum_j \left(\tau^*_{i,j,t} - \tau^c_{i,j,t}\right)}{3104} \tag{7}$$

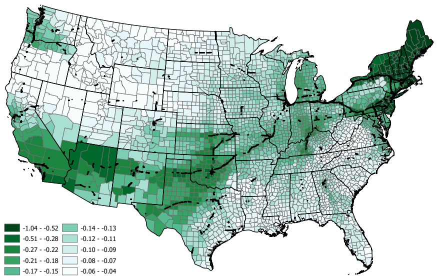
We see that, similarly to Table 10, the presence of the IHS generates a fall in transport costs, with the magnitude of the effect increasing from 1960 to 2000. Importantly, the geographical variation of the fall in transport costs closely reflects the gradual construction of the interstates, which is overlaid onto the maps. Indeed, the largest savings on the transport costs are either in regions with dense IHS infrastructure, such as the East coast and Appalachians, or in regions with initially relatively poor connections to the rest of the country. These poor connections can be due to the region itself being geographically peripheral, with relatively longer land routes to the rest of the country, as is the case for Florida or New England, or because the transport infrastructure connecting the region to the rest of the country was lacking. This is the case for the Pacific coast, which is connected to the rest of the country via a relatively sparse network spanning the Rocky mountains. In both cases, the presence of even small segments of the IHS immediately help disenclave the region, bringing that largest reductions in cost. Two important conclusions can be drawn from the geographical distribution of the gains from the IHS. First, while the gains are clearly unevenly distributed, it is the case that all counties benefited to some extent from the presence of the IHS. Second, by the time the IHS is completed in 2000, it is clear that the largest gains are distributed on the two main continental seaboards, which is strongly evidence that the IHS achieved its stated aim of facilitating "speedy intercontinental travel".

One final issue that needs further discussion is the absolute magnitude of the impact of the IHS on the cost of transport, which may appear low given the magnitude of the investment represented by the IHS. Indeed, the average contribution of the IHS, from Table 10, is at just under 2%, with the geographical variation in Figure 5 suggesting that the largest contribution at county level is 4% of transport costs. Several factors need to be considered, both to explain the size of the effect, and also to put the size effect into context.

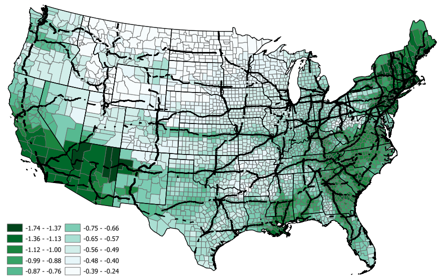
First of all, there is the effect of averaging the pairwise route costs over destinations. Even for the more detailed county level results shown in Figure 5, the cost reduction is a straight average across all possible 3104 destinations, therefore individual pairwise costs might see bigger reductions individually. As an illustration, the final column of Table 10 shows the 1$^{\text{st}}$ percentile of the distribution of impacts, i.e. the 99$^{\text{th}}$ percentile of reductions, showing that the impact of the IHS on individual routes can be large. In addition, as already highlighted, these impacts are obtained by a straight average across all possible 3104 destination counties and do not reflect the true, but unobserved, distribution of trips across destinations. As a result the effective historical transport cost reductions experienced in a given county will not necessarily match these simple averages.

The fact that the IHS seems to have generated a relatively small 2% additional average reduction compared to the background decline in transport costs does not mean, however, that the effect is not economically significant. First of all, this is in line with the effect sizes found by Allen and Arkolakis (2014), Allen and Arkolakis (2019), Jaworski et al. (2020) and mentioned in the introduction, albeit on transport costs themselves rather than on GDP or welfare. A second element to consider is the fact that profit margins in the US trucking industry are typically very low compared to other economic sectors. Using historical sectoral profit margin data, one can see that over the 1998-2019 period, the average operating margin after tax for the trucking sector was 6.05%, with a standard deviation of 3.22. The net margin was even tighter at 2.27%, with a standard deviation of 3.14.[15] Given the low margins present in this sector, even seemingly marginal reductions in transports costs will nevertheless be very valuable,

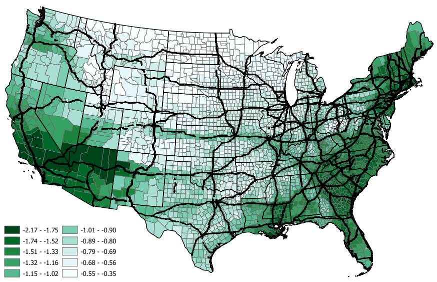
[15]This is calculated using Prof. Aswath Damodaran's historical corporate finance dataset, which is publicly available at `http://pages.stern.nyu.edu/~adamodar/`.



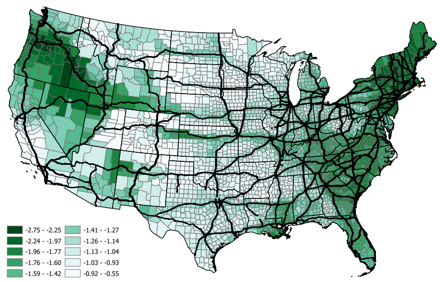
(a) 1960



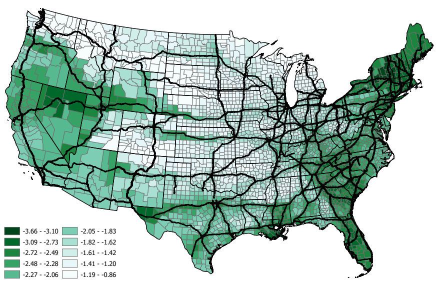
(b) 1970



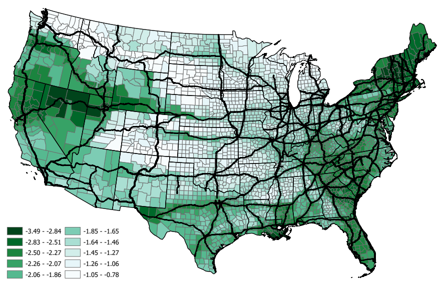
(c) 1980



(d) 1990



(e) 2000



(f) 2010

Figure 5: Percentage Difference between Actual and Counterfactual Real County Transport Costs

and will affect decision-making with regards to optimal routing and industry location.

A final consideration when assessing the magnitude of the IHS cost reductions comes from the literature on the estimation of agglomeration economies, which aims amongst other things to estimate the increasing returns to density which drive the agglomeration of economic activity into urban areas. Typically, this is estimated as an elasticity of productivity with respect to the size or population density of a city. Rosenthal and Strange (2004) provide both a review of the early literature as well as a summary of available estimates and conclude that the value of this elasticity is between 0.03 and 0.08, in other words doubling the density of a city increases productivity by 3 to 8%. Because these studies often rely on wages as a measure of productivity, estimation is affected by endogeneity problems as well as sorting of workers and firms. These issues were refined in a later set of contributions (Combes et al., 2008, 2011, 2012) which refine the estimates of the elasticity down to the 2-4% range, again a range that is similar to the IHS impact found in the counterfactual analysis. This regional science literature has established that spatial dynamics, agglomeration of activity in this case, are driven over the long run by mechanisms, such as increasing returns to density, that appear to be very modest in magnitude but continuously apply over long periods. The point of the comparison with the IHS is that because this is a long-run policy intervention, even a small instantaneous effect can be significant enough to affect the spatial dynamics of the US economy over the long run.

7 Conclusion

The purpose of the paper is to provide precise measurements of point-to-point marginal costs of transport on the post-World War II US road network. In doing so we take into account not only the evolution of the road network itself, most notably the construction of the IHS, but also fuel prices, fuel efficiency, truck driver wages and driving speeds for urban/rural and interstate/non-interstate roads. The resulting minimal pairwise county transport costs allow us to shed light on key factors behind the evolution of transport costs in the post- World War II age of highways. Overall, the transport costs, driven by a combination of higher driving speeds, improved fuel efficiencies and lower wages of truck drivers, fell significantly between 1955 and 2010. This fall, however, was uneven and the transport costs exhibit an inverted U-shape pattern peaking in 1980. This was due to the oil shocks of the 1970s which, combined with driver unionisation, pushed up fuel prices and wages, thus offsetting the beneficial impact of the IHS and better fuel efficiencies. After that, transport costs generally declined as the transport sector de-unionized, driving speeds continued to increase, and the benefits of the IHS occurred. Only in the first decade of the twenty first century did we witness a slight increase in the transport costs due to the decline in urban speed caused by congestion.

A complete set of pairwise minimal cost routes between counties makes it possible to run counterfactual analyses to isolate the county-level impact of individual components. We illustrate this by isolating the contribution of the IHS to transport cost reductions. The aggregate effect is in line with the existing literature on the impact of the IHS as well as the literature on agglomeration economies, however, the added value of the exercise comes from examining the spatial distribution of the impact of the IHS and noting for instance that the greatest gains are seen on both seaboards of the contiguous USA. Finally, it is our ambition to provide a detailed transport cost dataset that can facilitate future research on the long-impact of transport cost changes on the US economy. As highlighted throughout, when calculating average transport costs in one location we purposefully assumed away the distribution of traffic over destinations. Similarly, the impact of the IHS is evaluated in terms of reductions to transport costs, not GDP or welfare. This is not because we believe that such factors are unimportant, but instead because the scope of such research exceeds that of this paper. Given the central role that transport costs play in regional science for building measures of market access or explaining the spatial distribution of activity, the level of detail provided by this dataset will help improve the understanding of the spatial dynamics of the US economy in the latter half of the 20th century.

References

Allen, Treb and Costas Arkolakis (2014) "Trade and the Topography of the Spatial Economy," *The Quarterly Journal of Economics*, Vol. 129, pp. 1085–1140.

——— (2019) "The welfare effects of transportation infrastructure improvements," *NBER Working Paper Series*.

Alvarez-Palau, Eduard J, Dan Bogart, Oliver Dunn, Max Satchell, and Leigh Shaw Taylor (2020) "Transport and urban growth in the first industrial revolution," *working paper*.

Baum-Snow, Nathaniel (2007) "Did highways cause suburbanization?" *The quarterly journal of economics*, Vol. 122, pp. 775–805.

——— (2020) "Urban transport expansions and changes in the spatial structure of us cities: Implications for productivity and welfare," *Review of Economics and Statistics*, Vol. 102, pp. 929–945.

Baum-Snow, Nathaniel, Loren Brandt, J Vernon Henderson, Matthew A Turner, and Qinghua Zhang (2017) "Roads, railroads, and decentralization of Chinese cities," *Review of Economics and Statistics*, Vol. 99, pp. 435–448.

Belman, Dale L and Kristen A Monaco (2001) "The effects of deregulation, de-unionization, technology, and human capital on the work and work lives of truck drivers," *ILR Review*, Vol. 54, pp. 502–524.

Brinkman, Jeffrey and Jeffrey Lin (2019) “Freeway Revolts!,” *Federal Reserve Bank of Philadelphia Working Papers.*

Chandra, Amitabh and Eric Thompson (2000) “Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system,” *Regional Science and Urban Economics*, Vol. 30, pp. 457–490.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon (2008) “Spatial wage disparities: Sorting matters!,” *Journal of urban economics*, Vol. 63, pp. 723–742.

——— (2011) “The identification of agglomeration economies,” *Journal of economic geography*, Vol. 11, pp. 253–266.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sébastien Roux (2012) “The productivity advantages of large cities: Distinguishing agglomeration from firm selection,” *Econometrica*, Vol. 80, pp. 2543–2594.

Combes, Pierre-Philippe and Miren Lafourcade (2005) “Transport costs: measures, determinants, and regional policy implications for France,” *Journal of economic geography*, Vol. 5, pp. 319–349.

Condeço-Melhorado, Ana, Javier Gutiérrez, and Juan Carlos García-Palomares (2011) “Spatial impacts of road pricing: Accessibility, regional spillovers and territorial cohesion,” *Transportation Research Part A: Policy and Practice*, Vol. 45, pp. 185–203.

Datta, Saugato (2012) “The impact of improved highways on Indian firms,” *Journal of Development Economics*, Vol. 99, pp. 46–57.

Donaldson, Dave (2018) “Railroads of the Raj: Estimating the impact of transportation infrastructure,” *American Economic Review*, Vol. 108, pp. 899–934.

Donaldson, Dave and Richard Hornbeck (2016) “Railroads and American economic growth: A “market access” approach,” *The Quarterly Journal of Economics*, Vol. 131, pp. 799–858.

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

Duranton, Gilles and Matthew A Turner (2012) “Urban growth and transportation,” *Review of Economic Studies*, Vol. 79, pp. 1407–1440.

Faber, Benjamin (2014) “Trade integration, market size, and industrialization: evidence from China’s National Trunk Highway System,” *Review of Economic Studies*, Vol. 81, pp. 1046–1070.

Frye, Dustin (2021) “Transportation networks and the geographic concentration of employment,” *Working Paper.*

Ghani, Ejaz, Arti Grover Goswami, and William R Kerr (2016) "Highway to success: The impact of the Golden Quadrilateral project for the location and performance of Indian manufacturing," *The Economic Journal*, Vol. 126, pp. 317–357.

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

Hirsch, Barry T (1988) "Trucking regulation, unionization, and labor earnings: 1973-85," *Journal of Human Resources*, pp. 296–319.

Jaworski, Taylor, Carl Kitchens, and Sergey Nigai (2020) "Highways and Globalization," Technical report, National Bureau of Economic Research.

Jaworski, Taylor and Carl T Kitchens (2019) "National policy for regional development: Historical evidence from Appalachian highways," *Review of Economics and Statistics*, Vol. 101, pp. 777–790.

Koopmans, Carl, Wim Groot, Pim Warffemius, Jan Anne Annema, and Sascha Hoogendoorn-Lanser (2013) "Measuring generalised transport costs as an indicator of accessibility changes over time," *Transport Policy*, Vol. 29, pp. 154–159.

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

Nall, Clayton (2015) "The political consequences of spatial policies: How interstate highways facilitated geographic polarization," *The Journal of Politics*, Vol. 77, pp. 394–406.

Qin, Yu (2017) "'No county left behind?'The distributional impact of high-speed rail upgrades in China," *Journal of Economic Geography*, Vol. 17, pp. 489–520.

Rose, Nancy L (1987) "Labor rent sharing and regulation: Evidence from the trucking industry," *Journal of Political Economy*, Vol. 95, pp. 1146–1178.

Rosenthal, Stuart S and William C Strange (2004) "Evidence on the nature and sources of agglomeration economies," in *Handbook of regional and urban economics*, Vol. 4: Elsevier, pp. 2119–2171.

U.S. Federal Highway Administration (1977) *America's highways, 1776-1976: A History of the Federal Aid Program*: U.S. Government Print Office.

A Data: Sources and preparation

A.1 NHPN IHS build data and urban/rural classification

The process of generating the map of the extant IHS in 1960, 1970, 1980, 1990, 2000, and 2010 respectively was as follows. For each signed interstate (I-2 to I-99) we used the 'Interstate Density Maps' to identify which interstate highways sections were built by county and decade respectively and created the corresponding indicators in the NHPN shape file. Since the NHPN shape file splits each numbered highway into very fine-grained segments, this enabled us to determine not only whether an interstate highway was built in a county by the end of each decade, but also the length completed. This enabled us to verify that the total length was correct against the Baum-Snow (2007) digitized PR-511 data, and correct any discrepancies to closely match that data. Most counties, especially rural ones, are traversed at most by a single interstate highway, making it straightforward to determine the length built by the end of each decade. For counties with more than one interstate highway, it is also straightforward to determine the extant network if all interstate highways were open to public by the end of a given decade. For the remaining counties, in which different interstate highway segments was finished in different decades, we referred to Interstate-Guide to determine which portion of which interstate highway was opened to public by the end of each decade. Again, since the NHPN shape file splits each highway into very detailed segments on the scale of hundred meters, we were able to determine very accurately which portion of which highway and in which county was open to the public by the end of each decade. The resulting six digitized interstate highway maps for the years 1960, 1970, 1980, 1990, 2000, and 2010 are presented in Figure 1.

The NHPN data contains a rural/urban classification code for the road segments, however this only applies to the vintage of the NHPN data (2014), and does not provide historical information. In order to track the evolution of urbanisation over time, we use the rural-urban continuum codes provided by the US Department of Agriculture. This encodes the level of urbanisation for each county with a range that goes from 1 (Counties in metro areas of 1 million population or more) to 9 (Completely rural or less than 2,500 urban population, not adjacent to a metro area). Because our distinction is a binary one, we treat codes 1-7 as urban and 8-9 as rural.

A.2 Fuel: prices and consumption

The sources used and how the raw data were cleaned up are detailed below. The crucial aspect that needs to be clarified is the fact that we have used data on gasoline prices and consumption in the analysis, despite the fact that most US road transport vehicles have run on diesel since at least the mid 1950s. The main obstacle to using diesel prices and consumption is the difficulty in acquiring reliable data for the pre 1975 period. The U.S. Energy Information Administration does not provide price data prior to 1975, and the

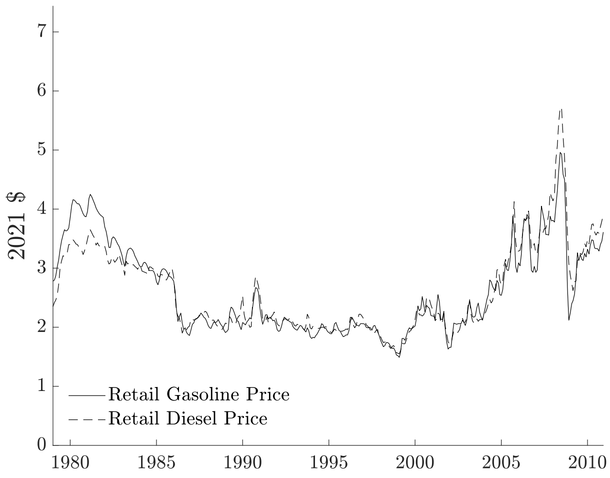


Figure 6: US monthly fuel prices 1979-2010, U.S. Energy Information Administration

only source for state-level historical prices are the USDA agricultural price statistics. Examination of the diesel prices in those publications revealed abnormally low values relative to gasoline prices, including after 1975 where the USDA can be directly compared to the USEIA data. Because the USDA data collects prices paid by farmers, it is very likely that these diesel prices include tax-exempt off-road diesel (also known as 'red diesel'), which is heavily used in the agricultural sector for running farm machinery and other off-road equipment. Because using these tax-exempt prices as determinants for on-road costs would introduce distortions, it was decided to rely on gasoline data instead. This decision was further justified by the fact that the US aggregate gasoline and diesel monthly price series reported by the USEIA track very closely, as visible in figure 6.

The raw data sources for gasoline prices were as follows:

- **1955**: U.S. Department of Labor, Bulletin No. 1197, Bureau of Labor Statistics 1956, Table 3
- **1959-1974**: U.S. Department of Agriculture, Crop Reporting Board, Economics and Statistics Service, Agricultural Prices. Annual Summary, edition 1959-1974
- **1975-2010**: U.S. Energy Information Administration, State Energy Data System

State-level gasoline prices for the period 1975-2000 were taken from U.S. Energy Information Administration, series MGTCD (average price of motor gasoline), which was straightforward. Similarly, prices from 1959-1974 were obtained from scanned copies of paper publications produced by USDA. While this required a large digitisation effort to convert the scanned documents to tabular format, the data collection itself was also straightforward.

Obtaining state-level gasoline prices for 1955, before the inception of the Interstate Highway System, required more work. We used retail prices of gasoline in fifteen cities in fifteen U.S states for 1955 from the Bulletin of U.S. Department of Labour. To calculate the prices in the remaining states, we assumed that the ratio of state to U.S average gasoline price is stable between 1955 and 1959 and used U.S average retail gasoline price prices in 1955 and state gasoline prices in 1959. Since the prices for the years 1959-2010 are tank-trunk prices, we scaled down 1955 retails prices using the state ratios of retail to tank trunk prices in 1959.

For fuel consumption, the sources were as follows:

- **State-level gasoline consumption data**: Federal Highway Statistics 1950-2010
- **U.S. average gasoline consumption**: U.S. Energy Information Administration, `https://www.eia.gov/totalenergy/data/browser/?tbl=T01.08#`

The state-level fuel consumption data is what was used as an input to the analysis, with the US average data serving as a validation check.

A.3 Earnings of truck drivers

Several sources were used for the earnings of truck drivers:

- **1955**: Union Wages and Hours, United States Department of Labor, Bulletin No. 1195, 1955, Table 7
- **1958-1993**: 'Employment, Hours, and Earning Unites States 1909-1994, Vol II'. We used nominal wages of non-supervisory workers in SIC421 'Trucking and courier services except for air' for the period 1964 to 1994. For the period 1958-1964, no data for trucking are available and so we used nominal average weekly earnings only for SIC 42 'Motor freight transportation and warehousing'.
- **1997-2010**: Occupational Employment Statistics Survey, Bureau of Labor Statistics, Department of Labor, average wages of two occupation categories: (i) truck drivers, heavy and tractor-trailer, (ii) truck drivers-light, include delivery/route workers.[16]

A.4 Driving Speeds

The state-level driving speeds used in the analysis were obtained as follows:

- **1955**: Speed data for urban, and rural roads are from 1957 Federal Highway Statistics (page 22), and 1964 U.S. Department of Commerce report 'Accidents on Main Rural Highways Related to Speed, Driver, and Vehicle' (page 6, Table 3) respectively. The Department of Commerce report provides information on speed in eleven states; speed in the remaining states was imputed by assuming that the speed in the states located in the same U.S Census region is the same. Urban speed is calculated using U.S. average speed from the Federal Highway Statistics and assuming

[16] `https://www.bls.gov/oes/tables.htm`, accessed August 3, 2020

that the ratio of state to U.S. average speed is the same in rural and urban areas. Since there was no Interstate Highway System, there are only two types of roads: urban non-interstate, and rural non-interstate.

- **1958-1970**: U.S.-wide average speeds are from U.S Historical Statistics, Series Q188. To calculate state-level speed data, we used a ratio of state to U.S. average speed. Specifically, we assume that the ratio of state speed to US average speed is the same as in the second half of the 1970s. Using second half of the 1970s is justifiable as the average speed at that time was similar to the levels of 1960s after a decline in the early 1970s. The stability of state speed to US average speed ratio over time was examined using data from 1955, and 1976-1991 for which Federal Highway Statistics provides detailed state-level breakdown. The calculated ratios are remarkably stable over time and regression analysis showed no statistically significant time trend. This confirmed that whilst average US speed was generally increasing over time, as shown in Table 4, the state-level average speed relative to US average speed showed no discernible time trend.

- **1971-1991**: : Speed data are from the Federal Highway Statistics 1981-1991, tables VS-1. These provide near-complete data for all types of road, with only a few missing observations imputed as the average speed of the U.S Census region in which the state with the missing data is located. Unfortunately, the reporting of speed data in the Federal Highway Statistics was discontinued after 1991.

- **2000-2010**: Data in 2010 are proxied using 2012 data on average speed in selected metropolitan areas published in 'Freight Facts and Figures 2017, U.S. Department of Transportation' (Table 4-1), and speed data US interstates published in 'Top 25 Commodity Corridor Performance Measures' provided by the Federal Highway Administration. This data enables us to cover thirty states. The remaining data is imputed by again assuming that the ratio of state-level speed to U.S. average speed is the same as in 1991. We have tested this assumption by comparing the imputed speed to the observed data for the 30 states for which we have speed data. This comparison revealed a very close match, indicating again that the imputation works well. An exception is Connecticut, Oregon, and Washington where the speed would be about ten percent higher, hence the transport costs for the routes through those states are upper bounds. As there is no speed data available for 2000, we take the average of 1991 and 2010.

B Network simplification and analysis

The US road network is modelled as a graph $\Gamma = (V, E)$, with vertices V representing the intersections in the road network, and edges E representing the road segments themselves. Given two vertices v and

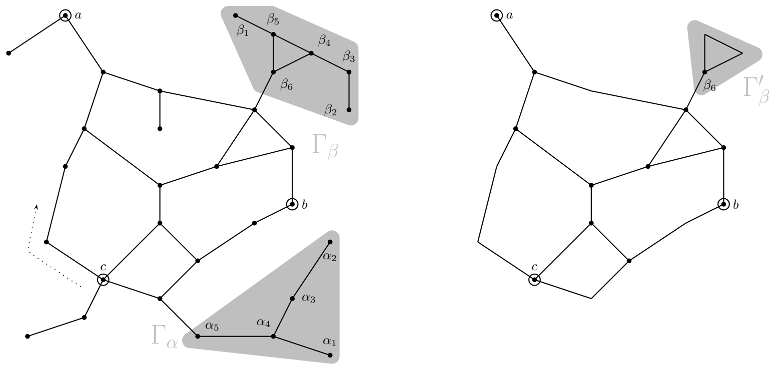
u, (v, u) represents an edge $e \in E$ connecting v and u. n_v is the degree of vertex v, i.e. the number of edges that connect to v. Indices i and j will be used to index vertices that are also county centroids, and these will form a set $C \subset V$. Finally $R_{i,j}$, which will be referred to as a route, is a path on Γ connecting centroids i and j. $R^*_{i,j}$ will be used specifically to denote the optimal route between i and j.

B.1 GIS stage: Network generation

The raw NHPN GIS data only contains road segments, with related information such as the county/state FIPS code, the length of the segment and the signage of the road. This gives us the edges E of the network, but not the vertices. In order to generate a proper graph Γ, we first generate the two endpoints of each road segment. Whenever two road segments are connected, this will result in two vertices being added at the location of this connection, one for each of the two corresponding edges. As a second step, therefore, all duplicate vertices are deleted, resulting in a set of unique vertices V. Crucially, the number of duplicates in the same location directly provides the degree of each vertex, n_v, which will play a central role in the simplification strategy presented below.

Before carrying out the simplification, two additional tasks are carried out. The first of these is the calculation of the transport costs on each edge in E in a given year, τ_e. This is done with (1) and (2), using the spatially disaggregated reference cost data for driving speeds, wages, fuel prices and fuel consumption described in Appendix A, combined with the distance and state FIPS code of each edge provided by the NHPN data. The NHPN data also provides the signage of the road segment corresponding to each edge, allowing the identification of interstates, although this is modified as explained above to reflect the construction of the IHS. The resulting transport costs for each year of analysis are stored in a set of symmetric $|V| \times |V|$ matrices τ_t.

The second task is the identification of the county centroids, which will form the start and end points of the routes in Dijkstra's algorithm. The strategy used for determining the coordinates for the centroid of each county is to average the coordinates of all the vertices in that county. Because the density of the road network is directly correlated to the density of the population, this allows us to obtain centroids that very closely track the largest population centre in each county without requiring very disaggregated population data. Rather than adding a new set of vertices to Γ, which would then also requires a new set of edges to connect them to the rest of the graph, we simply locate the closest existing vertex to the average and flag it as the centroid. The average distance between the average of vertex locations and the vertex selected as a centroid is small, 1.77 miles, with a standard deviation of 2.03 miles. The largest errors are located in the low population density Western states, with Nye County, Nevada seeing a maximal error of 29.94 miles. These relatively small deviations relative to the scale of the contiguous USA validates this choice of approach. Those vertices of Γ that are flagged as centroids form a subset $C \in V$.



(a) Before simplification (b) After simplification

Figure 7: Network simplification

At the end of this process, we have a full graph $\Gamma = (V, E)$ representing the US road network: a set of edges E representing the roads, a set of vertices V representing the intersections, with information about costs associated with edges, the degree of the vertices, and a set of identifiers C indicating which vertices represent a county centroid. Because of the high fidelity of the NHPN data, this graph is very large, with $|V| = 592,828$ vertices and $|E| = 625,610$ edges.

B.2 Simplification stage 1: Removal of degree 1 vertices

Before running Dijkstra's algorithm for pairs of US county centroids, we first simplify the graph Γ obtained in the GIS step. This is because the algorithm needs to be run $3105 \times 7 = 21,735$ times, once for every origin county and every benchmark year, and the time-complexity of Dijkstra's algorithm increases quadratically with $|V|$. Given the size of the raw graph, in order to keep the problem tractable, we need to first simplify the graph to reduce the number of edges and vertices. We do this by removing those road intersections that do not enter the decision problem of finding the cheapest route between two vertices in C, in order to ensure that the optimal solution found by Dijkstra's algorithm is not affected by the simplification process. Figure 7 is provided to help illustrate the simplification process, via a toy example where a, b and c represent centroids that will be the origin or destinations of our routes.

The first step is to iteratively eliminate all vertices of degree 1, as well as the edges that lead to them. These vertices represent the termination point of a dead-end road, or a cul-de-sac, and such locations can only be visited by a cheapest route if they happen to be either the destination or the origin of that route. The proof of this is trivial, and intuitively experienced by any driver who accidentally turns onto a dead-end road! Suppose that a route $R_{i,j}$ visits a degree 1 vertex v that is not its origin or destination: because $n_v = 1$, it is connected to a single edge (v, u), and the route will traverse (v, u) twice: once on the way to v, and once on the way back. Therefore, there also exist at least one vertex that is visited twice, u. A cheaper route can be obtained by removing (u, v) and (v, u) from $R_{i,j}$, and visiting u only once.

Algorithm 1 Removal of degree 1 vertices

Require: V: Vertices of Γ
Require: n: vector, degree of each vertex
Require: E: Edges of Γ
Require: C: Set of vertices in V that are centroids

while $\exists\, v \in V : n_v = 1 \;\wedge\; v \notin C$ **do**
  $v \leftarrow \text{find}(v \in V : n_v = 1 \;\wedge\; v \notin C)$ ▷ find degree 1 vertex
  $u \leftarrow \text{find}(u \in V : (v, u) \in E)$ ▷ find find connected vertex
  $V \leftarrow V \setminus v$ ▷ remove v from vertices
  $n \leftarrow n \setminus n_v$
  $E \leftarrow E \setminus (v, u)$ ▷ remove (v, u) from edges
  $n_u \leftarrow n_u - 1$ ▷ Reduce degree of u by 1
end while

return V, E, n

In fact, a much more general statement can be made: suppose that Γ contains a subgraph that is connected to the rest Γ by a single edge, and additionally suppose that this subgraph does not contain any centroids. Then, using the same logic as above, the cheapest route between any pair of centroids will never visit this subgraph. Figure 7(a) shows two such subgraphs, labelled Γ_α and Γ_β, which will never be visited by the optimal routes between a, b and c. [17] In theory, such subgraphs can therefore be removed from the original graph without affecting the choice of optimal route.

The removal of degree 1 vertices is carried out iteratively, using algorithm 1. This very simple and fast implementation will remove single branches, such as the road to the lower left of centroid c, as well as single-edge connected subgraphs, as long as the subgraph is a tree. This is the case for subgraph Γ_α, which is entirely removed by the procedure: in the first round, the algorithm will remove α_1 and α_2, and reduce the degree of α_3 to 1 and α_4 to 2. Subsequent rounds will sequentially remove α_3, then α_4 and α_5. Algorithm 1, however, cannot cope with subgraphs that contain a cycle, such as Γ_β, where the cycle formed by vertices β_4, β_5 and β_6 will not be removed. This is because after the elimination of the single branches containing β_1, β_2 and β_3, the vertices in the cycle will all have a degree of at least 2, and will therefore be ignored by the algorithm. This leaves a remainder subgraph, labelled Γ'_β in Figure 7(b). Unfortunately the problem of identifying, and then removing single-edge connected subgraphs that do not contain centroids, such as Γ_α and Γ_β is computationally complex and time-demanding. Because the purpose of the network simplification is to reduce the overall computation time required, and because the remainder subgraphs left by Algorithm 1 will never be visited by an optimal route in any case, it was decided that this simpler and faster simplification, albeit slightly sub-optimal, was sufficient. This process removes 16,984 vertices and 16,976 edges from Γ.

[17] Such topologies are common on real-life road networks, as many residential neighbourhoods are designed in this manner. There is a single collector road connecting the area to the main road network, precisely to avoid excess traffic being routed through the neighbourhood.

Algorithm 2 Removal of degree 2 vertices

Require: V: Vertices of Γ
Require: E: Edges of Γ
Require: W: Edge weights (cost, distance, time, etc.)
Require: C: Set of vertices in V that are centroids

$V' \leftarrow \{v \in V : n_v > 2 \lor v \in C\}$ ▷ get critical vertices
$E', W', M \leftarrow \varnothing$ ▷ initialise empty return variables
$L_{(v,u)} \leftarrow 0 \quad \forall (v,u) \in E$ ▷ label all edges as unvisited

for all $s \in V'$ **do** ▷ for all critical vertices
  $V_s \leftarrow \{u \in V : (s,u) \in E\}$ ▷ find vertices connected to s
  for all $u \in V_s$ **do**
    $v \leftarrow s$
    if $L_{(v,u)} = 0$ **then** ▷ only run on unvisited paths
      $P \leftarrow \varnothing$
      $W_P, stop \leftarrow 0$
      while $stop = 0$ **do**
        $P \leftarrow P \cup \{(v,u)\}$ ▷ add edge to path
        $L_{(v,u)} \leftarrow 1$ ▷ flag edge as visited
        $W_P \leftarrow W_P + W_{v,u}$ ▷ increment path weight
        if $u \in V'$ **then** ▷ if u is a critical vertex, we are done
          $E' \leftarrow (s,u)$ ▷ add new single edge (s,u)
          $M \leftarrow ((s,u), P)$
          $W'_{s,u} \leftarrow W_P$
          $stop \leftarrow 1$
        else ▷ else, move to next edge on the path
          $v' \leftarrow v' \in V : (u,v') \in E \land v' \neq v$
          $v \leftarrow u$
          $u \leftarrow v'$
        end if
      end while
    end if
  end for
end for

return V', E', W', M

B.3 Simplification stage 1: Removal of degree 2 vertices

Once all non-centroid degree 1 vertices are removed the next step is to remove degree 2 vertices. Unlike the degree 1 vertices, which are cannot be part of an optimal route unless they are a centroid, these can be visited by an optimal route. Despite this, they are irrelevant to the decision problem of Dijkstra's algorithm, because they do not represent a decision point where any real choice is available to a driver. Such a vertex has one edge leading to it, one edge leading away from it, and is therefore not a true intersection.[18]. Using Figure 7(a) as an illustration, suppose a driver travelling from c to a decides to follow the route indicated by the dotted arrow: when the driver arrives at the next vertex, which has degree 2, there is no choice but to continue along the route. The only alternative is to return to c, which

[18] The overwhelming majority of degree 2 vertices in the NHPN data represent administrative boundaries in the data, where the same signed road is broken up according to county boundaries, minor changes in road signage, or maintenance identifiers, etc.

as established in the previous section, means the route cannot be the cheapest. Thus, for the purposes of finding optimal routes in Γ, any two edges connected by a vertex of degree 2 can essentially be treated as a single edge. This insight forms the basis of the second simplification algorithm.

Algorithm 2 constructs a new graph $\Gamma' = (V', E')$ and an edge mapping M, allowing to reconstruct in the original graph Γ any route $R_{i,j}$ found in Γ'. The new set of vertices V' consists of all the vertices in V that either have a degree greater than 2, or that represent a centroid location.[19] In other words, V' is the subset of the critical vertices of Γ, representing origins and destinations of routes, as well as those vertices in the graph that represent true decision points for a route. Let v, u be two such critical vertices, appearing both in V and V', and assume that they are connected in Γ by a path $v - u$ made up entirely of degree 2 vertices. Let P be the subset of E containing the edges in the $v - u$ path. Then in Γ' v and u are directly connected by a single edge $(v, u) \in E'$, where the weight of the edge (be it a travel cost, a distance, or a time) associated with (v, u) are simply the sum over the edges in the disaggregated path P, i.e. $W_{(v,u)} = \sum_e W_e \quad \forall e \in P$.

Figure 7(b) illustrates the result of this process. The simplified graph Γ' preserves the core topology of the full graph Γ, as well as the edge cost or distance information required to run Dijkstra's algorithm, while greatly reducing the number of vertices and edges. In our case, the reduction in the graph size is dramatic, with only $|V'| = 51,602$ vertices and $|E'| = 84,403$ remaining in the simplified graph Γ'. This represents an order-of-magnitude reduction in the number of vertices, which given the $\mathcal{O}(|V|^2)$ time-complexity of Dijkstra's algorithm represents roughly a two order-of magnitude improvement in the runtime of our analysis.

B.4 Dijkstra's algorithm and route reconstruction.

As previously stated, the optimal routes $R^*_{i,j}$ on the simplified graph Γ' between centroids $i, j \in C$ are found using Dijkstra's algorithm. In our implementation, each individual run of the algorithm involves picking a starting point $i \in C$ and a set of target destinations $\{j, k, \dots\} \in C$, and results in a corresponding set of optimal routes to the targets $R^*_{i,j}, R^*_{i,k}, \dots$ as well as a vector of minimal transport costs from i to the set of targets $\tau^*_{i,j}$. Building a full matrix of transport costs in a given year thus requires $|C| = 3105$ runs, one for each origin county. This needs to be repeated for each of the 7 years in the analysis.

The network simplification carried out above reduces the computation time of a single run, from 1 origin county to 3104 destinations (producing 3105 optimal routes), to an average of 241 seconds, which is an extremely significant gain. Nevertheless, given the number of runs required, further optimisation of our implementation was required. First, the implementation takes into account the symmetry of the transport cost matrix, i.e. $R^*_{i,j} = R^*_{j,i}$ and $\tau^*_{i,j} = \tau^*_{j,i}$, to shrink the target set as runs proceed and reduce the overall computation required by a factor of 2. For the very first run, using centroid i as the origin,

[19] The vertices a and b in Figure 7(b) respectively have degree 1 and 2 yet because they represent centroids, they are not removed as part of the overall simplification.

the target set is $C \setminus \{i\}$. For the second run, starting in j, the target set is smaller, $C \setminus \{i, j\}$, given that $R^*_{i,j}$ was already produced as part of the first run. The target set for the next iteration, starting in k will be $C \setminus \{i, j, k\}$, and so on and so forth. As a second step we took advantage of the trivially parallel nature of the problem, as all runs are independent from each other, differing only in their starting point and target sets. The analysis was therefore run in parallel, using 72 threads on a 36-core cluster, which enabled the 3105 runs required to obtain optimal costs for a given year to be processed in just under 3 hours, thus making the overall analysis tractable.

Finally, while Dijkstra's algorithm is run on a simplified network Γ' compared to the full network Γ, because the simplification only removes vertices that are not relevant to the decision process of finding a cost-minimising route, the minimal cost of reaching j from i will be the same in both graphs. Similarly, an optimal route $R^*_{i,j}$ on Γ' is also optimal on Γ, once the single edges between critical vertices in $R^*_{i,j}$ are replaced by the corresponding sequence of edges containing degree 2 vertices. This is done using the edge mapping M produced by Algorithm 2, which gives us the ability to 'undo' the simplification if required, and recover information that is lost in the simplification process: the signage of a road segment, it's rural/urban nature etc. In particular, this is what enables us, for example, to calculate the share of an optimal route's distance travelled on the IHS.