Roads and Trade: Evidence from the US[§]

Gilles Duranton[*]
University of Pennsylvania

Peter M. Morrow[‡]
University of Toronto

Matthew A. Turner[‡]
University of Toronto

Revised: 30 August 2013

ABSTRACT: We estimate the effect of interstate highways on the level and composition of trade for US cities. Highways within cities have a large effect on the weight of city exports with an elasticity of approximately 0.5. We find little effect of highways on the total value of exports. Consistent with this, we find that cities with more highways specialize in sectors producing heavy goods.

Keywords: interstate highways, transportation costs, trade and specialization.

JEL classification: F14, R41, R49

[§]Thanks to Kristian Behrens, Ig Horstmann, Steve Redding, Rosa Sanchis, Aloysius Siow, Dan Trefler, the anonymous referees, the editor Imran Rasul, and seminar and conference participants for comments and suggestions. Financial support from the Canadian Social Science and Humanities Research Council is gratefully acknowledged. We are also grateful to Magda Biesiada and Christian Dippel for their assistance in preparing the data used in this project and to Nathaniel Baum-Snow for sharing his data with us.

[*]Wharton School, University of Pennsylvania, 3620 Locust Walk, Philadelphia, PA 19104, USA (e-mail: duranton@wharton.upenn.edu; website: https://real-estate.wharton.upenn.edu/profile/21470/). Also affiliated with the Centre for Economic Policy Research, the Rimini Centre for Economic Analysis and the Spatial Economic Centre at the London School of Economics.

[‡]Department of Economics, University of Toronto, 150 Saint George Street, Toronto, Ontario M5S 3G7, Canada (e-mail: peter.morrow@utoronto.ca; website: http://www.economics.utoronto.ca/morrow/).

[‡]Department of Economics, University of Toronto, 150 Saint George Street, Toronto, Ontario M5S 3G7, Canada (e-mail: mturner@chass.utoronto.ca; website: http://www.economics.utoronto.ca/mturner/index.htm).

Support from the Einaudi Institute for Economics and Finance during part of the time that this research was conducted is gratefully acknowledged.

1. Introduction

We investigate the effect of highways between and within cities on the weight and value of bilateral trade between large US cities. We base our investigation on data describing trade flows, the interstate highway network, city employment by sector since 1956 and other city characteristics. Identifying the causal effect of highways on trade is difficult. Regressing a measure of trade at the city level on a measure of city roads may not isolate the causal effect of roads on trade since highway construction may be more prevalent in cities that trade more (if highways are built to support trade) or, alternatively, in cities that trade less (if highways are built to foster trade). We resolve this inference problem by exploiting exogenous variation in exploration routes between 1528 and 1850, in railroad routes circa 1898 and in a 1947 plan of the interstate highway network. We find that a 10% increase in highways within a city causes about a 5% increase in the weight of its exports. We find weak evidence of a small effect on the value of these exports. It follows that highways within cities cause them to specialize in sectors that have high weight to value ratios. We corroborate this conclusion with city level employment data. Finally, we also find that variations in highways between cities cause large variations in the weight and value of trade as they reduce travel distances between cities.

Understanding the effect of highways on trade is important for a number of reasons. First, trade and specialization play a central role in the process of economic growth. Consequently, understanding the forces that promote trade is among the most fundamental problems that economists face. Second, the construction and maintenance of transportation infrastructure is among our most costly policy endeavors and our understanding of the effects of such investments remains rudimentary.[1]

Our analysis of how highways change trade flows consists of three components: a theoretical framework which allows us to understand the effects of highways on trade, motivates our econometric specification and highlights some identification issues; high quality data describing highways, trade flows and relevant covariates; and a strategy for resolving the possible endogeneity of highways to trade flows. The theoretical model which forms the basis for our analysis is an extension of Anderson and van Wincoop's (2003) framework to an economy with multiple productive sectors. This model leads to a gravity equation describing the effects of distance on equilibrium trade flows. The model also implies that cities with a relative abundance of highways specialize in sectors that are relatively intensive in their use of city highways. Taken together, we have a logically consistent framework to examine the effect of within- and between-city highways on the weight and value of inter-city trade and to assess the effect of highways on the composition of production and trade for a given city. Our theoretical model leads to a two-step estimation strategy resembling Redding and Venables (2004). The first step estimates a gravity model for the value and weight of trade flows, each as a function of distance and exporter- and importer-specific fixed effects. These fixed effects measure a city's propensity to export or import value or weight

[1] According to Couture, Duranton, and Turner (2012), the US spends nearly 200 billion dollars a year on road construction and maintenance; the value of capital stock associated with road transportation in the US tops 5 trillion dollars; the average driver spent about 72 minutes a day driving; and the median household devotes 18% of its budget to road travel.

conditional on distance and trading partner characteristics. In the second step, we explain these fixed effects as functions of within-city highways and other city characteristics. In a separate empirical exercise, we examine the sectoral composition of exporter employment as a function of within-city highways.

Our three primary data sources are the US Commodity Flow Survey (CFS), the Highway Performance Monitoring System (HPMS) and the County Business Patterns (CBP). From the CFS we calculate bilateral domestic trade flows for a cross-section of major US cities. From the HPMS we calculate the interstate highway distance between cities. This distance reflects the cost of traveling between cities. From the HPMS we also calculate the stock of interstate highway kilometers within each city's boundaries. This stock of city highways measures the city's capacity to move trucks through the congested portion of the city onto the inter-city portion of the highway network. We use the CBP to construct control variables for employment in our main empirical exercise and use historical employment data going back to 1956 to analyse the effect of highways on city specialisation. Our data also describe a rich set of city level control variables.

It is possible that the desire to trade will cause cities to build highways, or that some unobserved city characteristics cause both road building and trade. Such endogeneity and missing variable problems may confound estimates of the relationship between highways and trade. Resolving these inference problems is the third part of our analysis. We rely on instrumental variables estimation using instruments based on maps of the routes of major expeditions of exploration between 1528 and 1850, a map of major railroad routes in 1898 and a preliminary plan of the interstate highway system from 1947.

Our theoretical model requires that the cost of trading heavy goods be more sensitive to transportation costs than trade in light goods. Thus, changes to a city's highways should affect comparative advantage in heavy goods. Our data confirm this qualitative conclusion and suggest that the effect of roads on trade and specialization is economically important. Our main results are that a 10% increase in highways at the city of departure causes a near 5% increase in tons exported and that a change in highways causes at most a small change in the value of a city's exports. In a separate analysis we investigate the relationship between city sectoral employment and the stock of within-city highways. This investigation shows that cities with more highways employ more people in sectors producing heavy goods.

While the literature on trade investigates the effect of trade barriers and the role of transportation costs, infrastructure receives less attention. Limão and Venables (2001) find a strong positive association between an index of infrastructure quality and trade costs. In a similar vein, Clark, Dollar, and Micco (2004) also uncover a strong positive correlation between port efficiency and exports. This type of finding has been confirmed by more recent research (see Behar and Venables, 2011, for a review).[2] However, there are very few papers which can claim to identify the causal

[2]There is also a small literature that examines direct measures of transportation costs (Glaeser and Kohlhase, 2003, Combes and Lafourcade, 2005, Hummels, 2007) and one that focuses on the time costs associated with trade (Hummels, 2001, Evans and Harrigan, 2005, Djankov, Freund, and Pham, 2010). Harrigan (2010) considers the role of goods characteristics in determining transportation costs and whether shipments arrive by air, sea, or ground. Finally the literature in urban economics that looks at patterns of economic activity in cities only pays scant attention to trade (Duranton and Puga, 2000).

effects of infrastructure on trade.[3] As a rare exception, Michaels (2008) examines the effect of access to the interstate highway network on rural counties in the US.

In a recent paper closely related to our work, Donaldson (2014) examines the historical effects of railroads in India on price differentials across regions, trade flows and incomes. He finds strong evidence that railroads decreased trade costs and the price gaps between regions, and increased the value of trade between these regions. Interestingly he also finds that railroads led to higher incomes in real terms and that most of these gains can be accounted for by the gains from trade. Our findings differ since we find a small and often insignificant effect of roads on trade in value whereas we find strong effects on the specialization of economic activity. We think these contrasting results reflect contextual differences between railroads in colonial India and the US interstate network in 2007. Colonial India relied mostly on agriculture and agricultural goods were nearly impossible to trade before the railroads were built. While we do not deny the importance of the US interstate network, we note that there are some feasible alternatives to road trade in the US such as rail, airplane, or water transportation. We also note that workers are more freely mobile in modern America than they were in colonial India, and so respond more rapidly to regional wage differences.

Our work and Donaldson's (2014) also differ in their methodologies. Donaldson's empirical results rely strongly on his model. Given data limitations, he can only compare the rail network that was built to placebo networks that were never built. Our approach is less reliant on a particular model. To some extent, we are less confident that the greater complexity of today's transportation networks and economic activity can be fully captured by a tractable model. On the other hand, we face fewer data limitations and can assess the empirical robustness of our findings more thoroughly.

Our work is also related to a small emerging literature that assesses the effects of investments in transportation infrastructure on the evolution of cities and countries. Fernald (1999) examines the relationship between infrastructure spending and productivity for US states, Baum-Snow (2007) examines the role of the interstate highway system in the suburbanization of US cities, and Duranton and Turner (2012) examines how city populations depend on the interstate highway system. While the particular questions addressed in these papers differ from ours, like them we work towards a better understanding of the effects that transportation infrastructure has on economic development.

2. Data

We conduct two main empirical exercises. In one, we examine the relationship between roads and pairwise trade flows in a cross-section of cities. In the other, we consider a panel of cities and examine the effect of roads on city level employment patterns in traded sectors. We here describe the data on which the two exercises are based. Appendix A provides more detail.

[3] Like us, Feyrer (2009) also uses an IV approach. His focus nonetheless differs radically from ours. He uses the closure of the Suez Canal as a shock on trade costs to assess the effect of trade on growth. His estimation strategy relies on the exclusion restriction that, for countries, nothing correlated with the increase in trade distances during the eight year closure of the Suez Canal also affected economic growth.

Table 1: Weight and value of all 2007 CFS trade for 66 CFS cities

Sector	NAICS 2007	Value	Weight	$\frac{\text{Weight}}{\text{Value}}$	$\log(\frac{\text{Weight}}{\text{Value}})$	Count $N_i^k > 0$
Apparel	315	24,381	1,463	0.06	-2.81	66
Beverages and Tobacco	312	129,804	143,530	1.11	0.10	66
Chemicals	325	713,674	594,262	0.83	-0.18	66
Computers and Electronics	334	389,399	5,416	0.01	-4.28	66
Electrical Equipment, Appliances	335	128,868	18,771	0.15	-1.93	65
Fabricated Metal	332	338,290	118,350	0.35	-1.05	66
Food	311	585,676	568,950	0.97	-0.03	66
Furniture and Related	337	83,494	18,700	0.22	-1.50	66
Leather and Allied Products	316	5,441	636	0.12	-2.15	63
Machinery	333	343,262	40,523	0.12	-2.14	66
Mining (except Oil and Gas)	212	85,730	3,638,118	42.44	3.75	66
Miscellaneous	339	144,279	10,941	0.08	-2.58	66
Nonmetallic Mineral Products	327	124,713	1,060,926	8.51	2.14	66
Paper	322	174,780	166,472	0.95	-0.05	66
Petroleum and Coal Products	324	608,090	1,415,099	2.33	0.84	66
Plastics and Rubber Products	326	209,268	66,753	0.32	-1.14	66
Primary Metals	331	251,412	201,339	0.80	-0.22	65
Printing and Related Activities	323	98,848	33,661	0.34	-1.08	66
Textile Mills	313	35,936	8,989	0.25	-1.39	66
Textile Product Mills	314	28,448	7,022	0.25	-1.40	66
Transportation Equipment	336	715,294	94,023	0.13	-2.03	66
Wood Products	321	100,923	218,834	2.17	0.77	66
Mean		241,818	383,308	2.84	-0.83	65.8

Notes: Totals from 2007 CFS for all pairwise trade between 66 CFS regions for all variables. Unit values are for aggregate U.S. shipments. Import and export values are given in millions of nominal dollars. Weights are in thousands of tons. The last column reports count of cities with positive sectoral employment in 2007 county business patterns.

2.1 Trade data for 2007

We rely on the 2007 Commodity Flow Survey (CFS) to measure trade flows and we base our sample on the geography of this survey. The CFS divides the continental US into 121 CFS regions, each an aggregation of adjacent counties. Our sample consists of the 66 such regions organized around the core county of a US metropolitan area. We discard CFS regions like "Rest of Texas", which mix rural and small urban areas, and CFS regions that comprise an entire state (e.g., Idaho).[4] We refer to our 66 regions as CFS cities. Figures 1 to 3 below illustrate our sample of CFS cities as shaded polygons with dots representing employment centroids.

[4] CFS cities are often larger than the corresponding (consolidated) metropolitan statistical areas. For instance, Miami-Fort Lauderdale and West Palm Beach-Boca Raton in Florida are two separate metropolitan areas according to the 1999 US Census Bureau definitions but they are part of the same CFS region. On the other hand, a small metropolitan area like Gainesville (FL) is not part of a well identified CFS region. We exclude Washington DC, which the CFS defines as only the District of Columbia without the rest of its metropolitan region.

The CFS reports bilateral sectoral trade flows in both value (US dollar) and weight (tons) between CFS regions aggregated across sectors for 2007. Bilateral trade flows are also disaggregated by mode of transport; road (trucks), railroad and all modes. Bilateral trade flows are not available by sector.[5] The CFS nonetheless reports shipments from and to each region disaggregated by sector. Table 1 describes the weight and value of trade between our 66 cities by sector as well as the weight per value, here tons per million dollars, of an average shipment by sector. This table highlights considerable heterogeneity across sectors.

There is some censoring of trade flows in the CFS, more often of weight data than value. Such censoring reflects disclosure rules and mostly affects small values. In section 7.2, as a robustness check we generalize our econometric model to explicitly model censoring. This exercise suggests that censoring is not important econometrically. Given this, we generally ignore censored trade flows.

2.2 *Roads, 1956-2007*

We rely on several sources to construct data describing roads between and within CFS cities. To calculate interstate highway distances between cities, we first find the employment weighted centroid of each CFS region. Next, we calculate the pairwise distance between CFS cities as the interstate highway network distance between centroids using the 2005 National Highway Performance Network (NHPN) map of the interstate highway system.[6] Remarkably, with only a handful of exceptions, the centroids of all CFS cities are within a few kilometers of an interstate highway. We also calculate pairwise Euclidean distance between the same centroids. To measure the 'distance of a city to itself', we approximate the distance of within-city trade by $0.66(\text{area}/\pi)^{0.5}$, as in Redding and Venables (2004). Since interstate highways in rural areas are generally uncongested, we do not attempt to measure the capacity of the inter-city road connection. Absent congestion, the availability of extra lanes or alternate routes has little effect on transportation costs. Moreover, the preponderance of rural interstate highways have exactly two lanes in each direction so that there is little econometrically useful variation in capacity.

To describe contemporaneous roads within the boundaries of CFS regions, we mainly rely on the 2007 Highway Performance Monitoring System (HPMS). The US federal government administers the HPMS through the Federal Highway Administration in the Department of Transportation. This annual survey, which is used for planning purposes and to apportion federal highway money, collects data about the entire interstate highway system. The 2007 HPMS allows us to calculate the kilometers of interstate highways within the boundaries of each CFS city in 2007. It also reports the number of lanes on each segment of interstate highway, the lane kilometers of major urban roads and the lane kilometers of *urban* interstate highway.[7] This allows us to calculate the corresponding CFS city totals. As a robustness check, and to learn about the role of network configuration, we

[5]In fact, the 2007 CFS does report bilateral trade flows by sector, but these data are too heavily censored to be useful for our analysis.

[6]This map is only available for 2005 and 2012.

[7]The HPMS gives us two pieces of information about the location of each segment; the county in which it lies and whether or not it is in an urbanized area within that county.

construct an index of radial road capacity for each city. We construct this radial road index using the algorithm developed in Baum-Snow, Brandt, Henderson, Turner, and Zhang (2012) to count the number of highway rays on the 2005 NHPN map.[8]

We also construct a panel of road data with a less detailed description of the transportation network than is available in our 2005/2007 cross-section. This panel covers the period from about 1956 until about 2007, loosely, from the first construction of the US interstate highway system to the present.

For the years 1957, 1967, 1977 and 1987, we base our roads data on the PR-511 data used in Baum-Snow (2007). These are administrative data describing the construction of the interstate highway system. In particular, they assign each segment of the interstate highway system to a county and report its length and date of opening. For the years 1987, 1997 and 2007, our roads data are based on the Highway Performance Monitoring System (HPMS) from which the same information about segment length is available. Our HPMS and PR-511 data overlap in 1987 and in 1993 (not used) and agree almost exactly in both years.

2.3 *Employment, 1956-2007*

Since CFS cities are aggregations of counties, we can construct employment data from the County Business Patterns (CBP) data. These data are constructed from social security payroll tax records and enumerate establishments and employment by county and sector. Our panel of employment data is based on CBP data from 1956, 1970-1, 1977-8, 1987-8, 1996-7 and 2006-7. Assembling these data into a panel requires that we solve four problems.

First, some county/sector level employment values are censored. Since establishment counts are not censored, we impute censored employment values from establishment counts.[9] To further reduce measurement error, we average over pairs of adjacent CBP years for all years but 1956. The resulting panel describes employment in nominal years 1956, 1970, 1977, 1987, 1996, 2006. Second, the industrial classification system changes several times over the course of this 50 year period. We construct correspondence tables that map contemporaneous industrial classifications into 2007 NAICS, the classification used in the 2007 CFS. Third, the 1956, 1970 and 1971 CBP identify counties using a social security administration code rather than the now pervasive FIPS codes. This requires the construction of an additional correspondence table. Finally, the CBP data are available in electronic form only as far back as 1970. To extend our panel of employment data back to the earliest days of the interstate highway system, we convert the 1956 CBP from microprint images of the original paper books to electronic form. This process is sufficiently difficult that

[8]For a given city, this algorithm operates as follows. First, draw two circles, one of radius 20km and one of radius 40km, both centered on the city centroid. Second count the number of intersections of each circle with the road network. Third, calculate the smaller of these two numbers. This is our index of radial road capacity. More detail and illustrations are available in Baum-Snow *et al.* (2012).

[9]While the details of this imputation vary slightly by year as the details of the establishment count data vary, the basic idea is to regress county sectoral employment on establishment counts, which are sometimes available by size of establishment, and to use the resulting regression to impute employment on the basis establishment counts. The R^2 of the regressions used for these imputations is often as high as 0.99.

we base this 'year' of data on a single year of the CBP rather than the average of adjacent years.[10] The resulting employment data describe employment in each CFS city at approximately 10 year increments. Deviations from this uniform time increment either reflect electronic data availability or allow us to use pairs of adjacent years based on the same industrial classification system.

2.4 *Historical and planned networks*

In both of our econometric exercises we are concerned that roads are assigned to cities on the basis of their propensity to trade. To address this possibility, we construct instruments from a 1947 plan of the interstate highway network, from a map of railroad routes circa 1898, and from a map of routes of expeditions of exploration. We now describe these data, but postpone a discussion of instrument validity.

Our 1898 railroad data is based on a digital image of a map of major railroad lines (Gray, c. 1898). We convert this image to a digital map with the same format and projection as the map of 2005 interstate highways, as in Duranton and Turner (2011). We then calculate 1898 railroad distance between cities, kilometers of 1898 railroad in each CFS city and an index of radial capacity for 1898 railroads in each CFS city. Figure 1 shows 1898 railroad routes and CFS cities.

Our map of the 1947 plan of the interstate highway system is based on a digital image of the paper record of this plan. We create a digital map of the 1947 highway plan from this paper record (United States House of Representatives, 1947). We then calculate kilometers of 1947 planned interstate highway in each CFS city and compute planned distance between cities. Figure 2 shows the planned highway network.

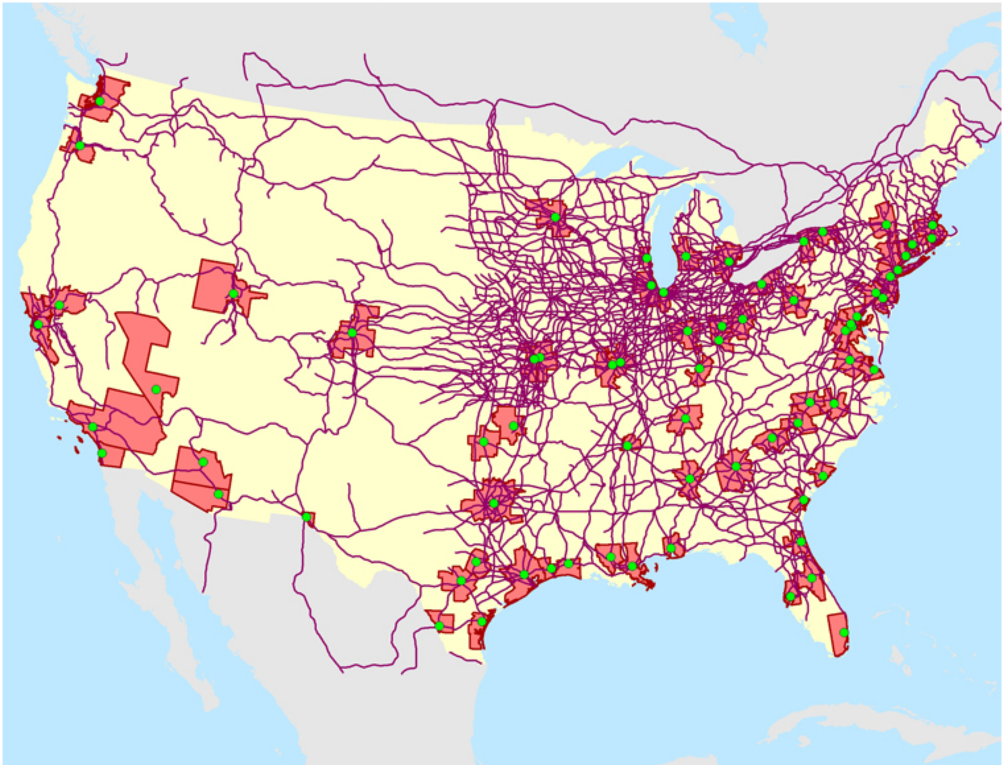
Our data on exploration routes is derived from maps in United States Geological Survey (1970). The National Atlas of the United States of America (1970) describes the routes of major expeditions of exploration that occurred during each of five available time periods; 1528–1675, 1675–1800, 1800–1820, 1820–1835 and 1835–1850. We digitize each map and count 1 km by 1 km pixels crossed by an exploration route in each CFS city. We then compute our index by summing these counts across all maps. Following this procedure, routes used throughout the 1528-1850 period receive more weight than those used for a shorter period of time. Note that the electronic representation of the exploration routes data does not allow us to calculate network distances along historical exploration routes. Figure 3 shows exploration routes between 1528 and 1850 and CFS cities.

2.5 *Other data*

We obtain past and contemporaneous populations for each CFS city from county population data in the 1920, 1950 and 2000 censuses for reasons we discuss below. We also use the 2000 census to compute a variety of other controls such as the share of adult population with at least a college

[10] As far as we know, this is the first use of the complete 1956 CBP data in the economics literature. We also note that 1956 is the first year for which the geographical coverage of the CBP is comprehensive and that the next year of available data is 1959.

Figure 1: 1898 railroads



Source: Map based on Gray (c. 1898).
Notes: The lines are 1898 railroads. The shaded polygons are the CFS regions in our sample. The small dots indicate CFS region centroids.

degree and a measure of income per capita. We also use a several geographical characteristics collected from various sources.

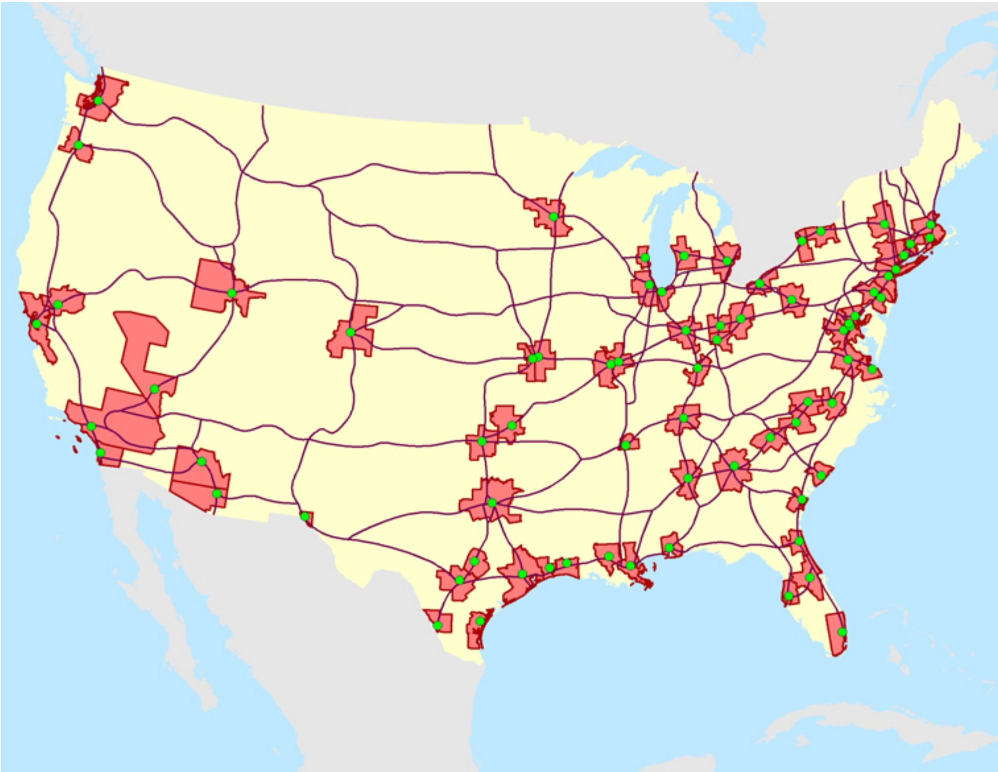
In an extension of our cross-sectional investigation, we also consider the effect of railroads on trade flows. We rely on GIS railroad maps from the North American Atlas and from the National Atlas. We compute kilometers of railroad track within each CFS city in 2004 and an index of radial rail capacity that mirrors the radial road index. Finally, we use these railroad maps to compute rail network distance between the centroids of CFS cities.

Table 2 and figure 4 provide descriptive statistics for our trade, road and rail data. Table 2 shows that road is the dominant mode of transportation, while railroads play a minor role. Table 2 also shows that exports to other cities represent only 62% of all shipments in value and 31% in weight. Figure 4 refines this second point and shows that the weight and value of trade flows decay rapidly with distance.

3. Model

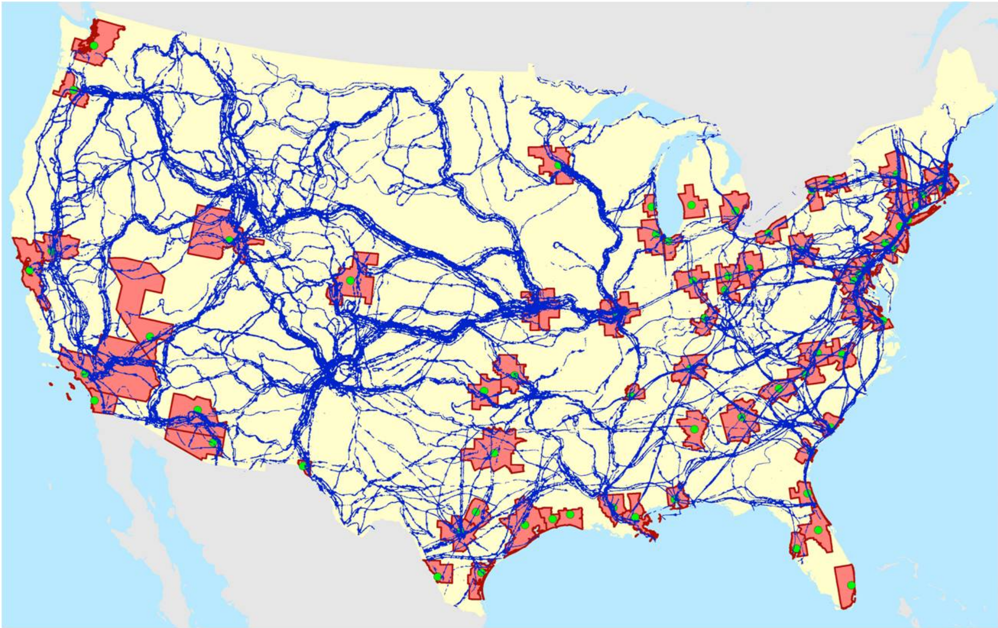
We require an econometric strategy that allows us to make inferences about the effect on trade of roads within and between cities. This strategy must respond to three problems. First, patterns of

Figure 2: 1947 planned highways



Source: Map based on United States House of Representatives (1947).
Notes: The lines are planned 1947 interstate highways. The shaded polygons are the CFS regions in our sample.The small dots indicate CFS region centroids.

Figure 3: 1528-1850 exploration routes



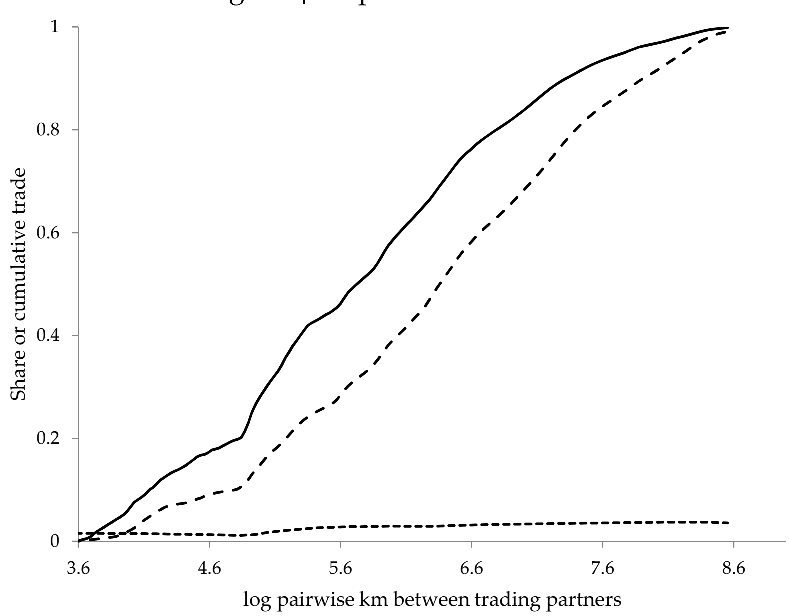
Source: Map based on United States Geological Survey (1970).
Notes: The lines are exploration routes between 1528 and 1850. The shaded polygons are the CFS regions in our sample. The small dots indicate CFS region centroids.

Table 2: City infrastructure endowments and trade flows by weight, value and mode

Variable	Mean	Std. Dev.	Minimum	Maximum
Value of exported shipments, road	50,552	62,810	1,295	406,851
(log) Value of exported shipments, road	10.35	0.99	7.17	12.92
Value of imported shipments, road	50,552	57,213	6,182	360,004
(log) Value of imported shipments, road	10.45	0.83	8.73	12.79
Weight of exported shipments, road	52,615	49,855	1,512	297,702
(log) Weight of exported shipments, road	10.52	0.89	7.32	12.60
% Road in exported value	70.7%	14.4%	17.1%	90.1%
% Rail in exported value	1.1%	2.2%	0%	10.0%
% Road in exported weight	86.2%	16.2%	23.5%	100%
% Rail in exported weight	1.5%	2.4%	0%	13.8%
% export in all shipments, value	62.1%	11.3%	40.6%	91.8%
% export in all shipment, weight	31.3%	15.3%	7.3%	78.2%
Employment, 2007	1,129,117	1,180,287	66,006	6,759,481
Section km of interstate highway, 2007	381	247	61	1,661
Railroad km, 2004	335	228	65	1304
Planned highway km, 1947	252	162	56	1,016
Railway km, 1898	619	405	91	2,104
Exploration routes index, 1518-1850	6,329	5,386	225	36,049

Notes: means across 66 CFS regions for all variables. All trade figures are for 2007. Import and export values are given in millions of nominal dollars. Weights are in thousands of tons.

Figure 4: Exports and distance



Source: Computed from the 2007 Commodity Flow Survey.
Notes: The plain line represents the cumulative distribution function of exports in weight. The (long) dashed line represents the cumulative distribution function of exports in value. The (short) dashed line represents the railroad share of total tons exported by railroad or road.

trade between any pair of cities may depend on the availability of alternative trading partners for each member of the pair. Second, roads may affect the composition as well as the level of trade, as cities specialize (or not) in transportation intensive production. Third, roads may be endogenously assigned to cities (or city pairs) on the basis of trade patterns.

To address these problems, we develop a simple theory of roads and trade. Since our goal is a tractable econometric specification, we consistently make the strongest simplifying assumptions consistent with our data: making the model more detailed than our data adds complexity without informing our estimation. Specifically, we suppose that each city produces a unique set of differentiated goods as in Anderson and van Wincoop (2003). We note, however, that the results of Arkolakis, Costinot, and Rodríguez-Clare (2012) suggest that our analysis is robust to alternative microfoundations, such as the Ricardian model of Eaton and Kortum (2002).

We analyze trade between a large but finite set of cities, indexed by $i \in \{1, \ldots, I\}$. City i houses $N_i \in \mathcal{R}_{++}$ identical consumers, each of whom purchases goods and supplies a single unit of labor in their home city. Consumers are mobile across sectors but, to begin, immobile across cities. As a stylized description of the fact that city employment in most sectors is small relative to total city employment, we consider a continuum of sectors k, for $k \in [0,1]$.

Firms in each city and sector produce the same city-specific variety with a constant returns to scale technology,

$$Q_i^k = A_i \, N_i^k, \tag{1}$$

where A_i is the productivity of labor in city i and N_i^k is the endogenously determined employment in city i's sector k. Consumers in all cities have identical preferences. A consumer in city j chooses consumption q_{ij}^k from all cities i in all sectors k to maximize the utility function,

$$U_j = \left[\sum_{i=1}^{I} \int_0^1 (q_{ij}^k)^{\frac{\sigma-1}{\sigma}} \, \mathrm{d}k \right]^{\frac{\sigma}{\sigma-1}} , \tag{2}$$

where $\sigma > 1$. The consumer's budget constraint is

$$W_j \geq \sum_{i=1}^{I} \int_0^1 P_{ij}^k \, q_{ij}^k \mathrm{d}k \, , \tag{3}$$

where W_j is the unit wage in city j and P_{ij}^k is the price of the variety from city i in sector k when consumed in city j. In equilibrium, consumers maximize utility, price-taking firms maximize profit, there is free entry in all sectors and cities, the labor market clears in all cities and output markets clear for all varieties.

The fact that each city produces its own variety in each sector k, together with our assumption that the elasticity of substitution is finite and constant, implies positive employment in all sectors in all cities. As we see in table 1, this a good approximation to our aggregated data and eliminates corner solutions. It is possible, but cumbersome and uninformative, to extend our model to more realistic substitution patterns. Equation (1) allows for cities to have different levels of aggregate productivity but abstracts from Ricardian comparative advantage in the production technology. We investigate other sources of comparative advantage, e.g., a city's endowment of human capital, in our empirical work.

Let X_{ij}^k denote expenditure in city j for the variety produced by city i in sector k. That is, X_{ij}^k describes the *value* of pairwise trade in a particular good and is one the main objects of study. We postpone a consideration of the *weight* of pairwise trade, which derives from our analysis of value.

Maximizing utility (2) subject to the budget constraint (3) and aggregating across consumers implies that

$$X_{ij}^k = P_{ij}^k \, Q_{ij}^k = \left(\frac{\mathbb{P}_j}{P_{ij}^k} \right)^{\sigma - 1} N_j \, W_j \,, \tag{4}$$

where

$$\mathbb{P}_j \equiv \left[\sum_{i=1}^{I} \int_0^1 (P_{ij}^k)^{1-\sigma} \, \mathrm{d}k \right]^{\frac{1}{1-\sigma}} \tag{5}$$

is the price index for city j. Equation (4) gives city j's demand for imports from city i in sector k. To arrive at the corresponding equilibrium quantity we must specify trade costs and derive the relevant supply relationship.

To describe transportation costs we assume that for one unit of output from city i's sector k to reach importer city j, the exporter must ship $\tau_{ij}^k \geq 1$ units. Thus, the cost to produce the variety from sector k and city i is multiplied by τ_{ij}^k at destination j. Competitive production of each variety implies marginal cost pricing so that

$$P_{ij}^k = \tau_{ij}^k \, \frac{W_i}{A_i} \,, \tag{6}$$

Inserting equation (6) into (4) and aggregating the resulting pairwise sector flows across sectors yields the aggregate value of equilibrium shipments from i to j,

$$X_{ij} = \left(\frac{A_i}{W_i} \right)^{\sigma - 1} \left[\int_0^1 \left(\tau_{ij}^k \right)^{1-\sigma} \mathrm{d}k \right] \mathbb{P}_j^{\sigma-1} \, N_j \, W_j \,. \tag{7}$$

Finally, the equilibrium for city i is such that its labor market clears. Using the production function (1), the demand for labor in city i and sector k is $N_i^k = Q_i^k / A_i$. Recalling that $\tau_{ij}^k \, Q_{ij}^k$ units must be shipped from city i for Q_{ij}^k units to be consumed in city j, inserting equation (6) into (4), aggregating across importers, and using the resulting expression to substitute for the quantities imported by the destination cities yields employment in each city and sector,

$$N_i^k = \sum_{j=1}^{I} \frac{\tau_{ij}^k \, Q_{ij}^k}{A_i} = \sum_{j=1}^{I} \frac{\tau_{ij}^k \, X_{ij}^k}{A_i \, P_{ij}^k} = \frac{A_i^{\sigma-1}}{W_i^{\sigma}} \sum_{j=1}^{I} \frac{\mathbb{P}_j^{\sigma-1}}{(\tau_{ij}^k)^{\sigma-1}} \, N_j \, W_j \,. \tag{8}$$

Aggregate labor supply in city i, N_i, equals the sum of labor demand for all sectors in city i. Thus, using equation (8)

$$N_i = \frac{A_i^{\sigma-1}}{W_i^{\sigma}} \left(\sum_{j=1}^{I} N_j \, W_j \int_0^1 \frac{\mathbb{P}_j^{\sigma-1}}{(\tau_{ij}^k)^{\sigma-1}} \mathrm{d}k \right) . \tag{9}$$

Given an equilibrium in all cities but i, we can solve for the equilibrium in city i. Knowing population in this city, labor productivity and transportation costs in all sectors, equation (9) yields the wage W_i. We can use this wage in equations (7) and (8) to compute the value of exports from i to j and city i employment in any sector k.[11]

[11] Inserting equation (6) into (5) and substituting into (9) shows that the wage in city i can be written as increasing and strictly concave function of wages in other cities. This can be used to show the existence and uniqueness of the general equilibrium of the model.

We so far rely on a description of transportation costs that abstracts from the role of roads and from intrinsic differences in the difficulty of moving different commodities, both of which are central to our inquiry. To proceed, let R_i be a measure of the road capacity in city i, most often 'kilometers of interstate highways within city boundaries,' and let R_{ij} describe the roads connecting city i to city j, usually 'distance between i and j along the interstate highway network.' To allow transportation costs to differ across commodities, let V_k denote the weight of one unit of sector k output. Without loss of generality, rank sectors by V^k where sector 0 produces the lightest goods and sector 1 the heaviest: $V^0 \leq ... \leq V^k \leq ... \leq V^1$.

With this notation in place, we can allow transportation costs to vary with; sector, between-city roads and within-city roads, according to

$$\tau_{ij}^k \equiv \tau^k(R_i, R_{ij}, R_j) = \tau_x^k(R_i) \times \tau_{xm}(R_{ij}) \times \tau_m(R_j)\,. \tag{10}$$

Thus, we decompose the cost of shipping a unit of output from city i and sector k to city j into three components; the cost of leaving exporter city i in sector k, the cost of going from exporter city i to importer city j and the cost of entering importer city j.[12,13] To lighten notation, where possible we use $\tau_i^k \equiv \tau_x^k(R_i)$, $\tau_{ij} \equiv \tau_{xm}(R_{ij})$ and $\tau_j \equiv \tau_m(R_j)$.

The cost of transportation between cities, $\tau_{xm}(R_{ij})$, increases with the network distance between cities, so that $\partial\tau_{xm}(R_{ij})/\partial \log R_{ij} > 0$. The cost of entering city j, $\tau_m(R_j)$, decreases in with its stock of roads, so that $\partial\tau_m(R_j)/\partial \log R_j < 0$. Note that this restricts the cost of between-city travel and the cost of entering a city to be the same across sectors.

The cost of shipping goods out of city i, $\tau_x^k(R_i)$, also declines with the stock city highway kilometers, i.e., $\partial\tau_x^k(R_i)/\partial \log R_i < 0$. However, we allow the effect of roads on transportation costs to vary by sector and we assume that sectors producing heavier goods are more sensitive to roads. More specifically, assume that the elasticity of the cost to exit city i with respect to city roads, i.e., $\rho_R^{\tau,k} \equiv \partial \log \tau_x^k(R_i)/\partial \log R_i$, is larger in absolute value for sectors producing heavier goods. It follows that the ranking of sectors by weight induces a corresponding ranking of these elasticities, $|\rho_R^{\tau,0}| \leq ... \leq |\rho_R^{\tau,k}| \leq ... \leq |\rho_R^{\tau,1}|$.

The description of transportation costs given in equation (10) is central to our analysis. As in much of the literature, transportation costs affect destinations costs multiplicatively. In fact, we can imagine that some transportation costs may be additive. To preserve clarity we maintain the multiplicative assumption in our theory, but investigate alternatives in our empirical work. The assumption that the cost of shipping goods is more sensitive to city roads in sectors producing heavier goods is a natural consequence of our prior that bulky goods are more 'road intensive' and benefit more from the presence of highways. While the opposite assumption is defensible, as we will see, our assumption is consistent with our empirical results. While we allow the cost of leaving a city to vary by sector, we restrict the cost of travelling between cities to be the same for all sectors.

[12]The multiplicative formulation of transportation costs between $\tau_{xm}(R_{ij})$ and $\tau_m(R_j)$ is not needed to derive the theoretical predictions presented below. We impose it only because it leads to a simple empirical specification.

[13]Although we introduce more structure on shipping costs than the previous literature, we nonetheless refrain from a full modeling of a shipping sector. This would be beyond the scope of this paper and require us to make modeling decisions about the location of this sector, the wages it pays, the fact that for any city trade will balance in value but not in weight making some trucks run empty, etc. See Behrens and Picard (2011) for a model exploring the implications of some of these issues.

Since bilateral trade flows in the CFS are not available at the sector level, allowing $\tau_{xm}(R_{ij})$ to vary by sector would add complexity without informing our estimations. We also assume that the cost of entering a city is constant across sectors. This implies that cities choose patterns of sectoral specialization according to how many roads they have and not how many roads their neighbors have. Allowing the cost of entering a city to differ across sectors would cause specialization to be determined in part by demand from other cities, and would complicate our analysis enormously. Again, we investigate this issue in our empirical work and find no evidence that city roads affect the weight of imports.

While this relationship is not explicit in our model, we do not preclude highways from having an effect on the quality of goods in the spirit of Alchian and Allen (1964). It would be straightforward to relabel 'weight' as 'quality' and to associate heavier goods with lower quality. It is also possible that highways affect productivity directly. While our model considers only allocative gains from trade, it could be extended to consider efficiency gains from trade by allowing productivity A_i to depend on highways R_i. In this case, highways would affect trade directly by determining transportation costs and indirectly through their effect on city productivity. We do not pursue this line of inquiry because we do not have good measures of city productivity. Our estimates generally reflect the sum of these two effects. Our specialization results nonetheless show that the specialization channel between sectors emphasized by our model is empirically important.

4. Roads and equilibrium trade

We now turn to a characterization of the equilibrium relationship between roads and trade. Characterizing the equilibrium relationship between roads and trade allows us to evaluate several comparative statics. Empirical testing of these comparative statics is the principal object of the paper. Our analysis also leads naturally to the econometric model which forms the basis of our main empirical exercise.

To begin, insert (10) into (7) and take logs to get

$$\log X_{ij} = \delta_i^X + (1-\sigma)\log \tau_{ij} + \delta_j^M . \tag{11}$$

In this equation δ_i^X and δ_j^M are expressions, defined below, involving the economic fundamentals of importer and exporter cities. We can describe δ_i^X as the 'propensity to export', or when necessary 'propensity to export value'. Similarly, δ_j^M is the 'propensity to import' or the 'propensity to import value'. In the context of the empirical model developed below, we often call δ_i^X an 'exporter fixed effect' or an 'exporter value fixed effect' and δ_j^M an 'importer fixed effect'. We use analogous language to describe cities' propensities to trade weight.

We can now state the first of our comparative statics about equilibrium trade flows.

Comparative Static 1 A reduction in road distance between two cities increases the value of trade between these two cities but does not affect its composition.

A formal proof of this result (and those that follow) is given in Appendix B. A reduction in the distance between two cities reduces trade costs which increases trade in all sectors. Since the cost of intercity distance is the same for all sectors by equation (10), a reduction in the distance between two cities also leaves the composition of trade unchanged.

From equations (7), (10) and (11), and then making use of equation (9) to eliminate the endogenous wage, W_i, we can write the propensity to export as

$$\begin{aligned} \delta_i^X &= \log(N_i W_i) - MA_i^X \\ &= S(R_i) + \frac{\sigma-1}{\sigma}\log A_i + \frac{\sigma-1}{\sigma}\log N_i - \frac{\sigma-1}{\sigma} MA_i^X . \end{aligned} \tag{12}$$

Equation (12) involves both a function of within-city roads, $S(R_i)$, and an export market access term, MA_i^X. The first of these depends on the cost of exporting from city i and is defined as

$$S(R_i) \equiv \frac{1}{\sigma}\log\int_0^1 \left(\tau_i^k\right)^{1-\sigma} \mathrm{d}k . \tag{13}$$

For city i, MA_i^X is an export market access term, a form of market potential, and it is given by

$$\begin{aligned} MA_i^X &\equiv \log\sum_{j=1}^{I} \frac{\mathbb{P}_j^{\sigma-1}}{(\tau_{ij}\,\tau_j)^{\sigma-1}} N_j W_j \\ &= \log\sum_{j=1}^{I} e^{(1-\sigma)\,\log\tau_{ij}+\delta_j^M} , \end{aligned} \tag{14}$$

where the second line results from substituting equations (7), (10) and (11) into the first. Because pairwise transportation costs $(1-\sigma)\log\tau_{ij}$ and propensity to import δ_j^M appear in equation (11), we can use an estimate of equation (11) to calculate MA_i^X for cities in our sample.

Our next comparative static follows.

Comparative Static 2 Export market potential, MA_i^X, negatively affects city i's propensity to export, δ_i^X.

This apparently counter-intuitive prediction follows from the fact that equation (11), which describes δ_i^X, already accounts for distance to importers and their propensity to import. Equation (12) then only captures a negative indirect wage effect. To see this, note that by equation (7) we have $\delta_i^X = (A_i/W_i)^{\sigma-1}$, that is higher wages reduce the propensity to export. To obtain our final expression in equation (12) we substitute market potential for wages. Since wages and market access are positively related as per equation (9), we must thus obtain a negative effect of market potential on propensity to export.

Our third comparative static also follows from equation (12).

Comparative Static 3 The effect of within-city roads on the propensity to export value should be positive but small.

Expression (12) shows that within-city roads affect the value exports only to the extent that they affect wages. The literature suggests that the elasticity of the value of exports with respect to within-city roads should be small, perhaps about 0.05.[14] Given the size of our sample, such a small effect is probably undetectable.

[14] Duranton and Turner (2012) find that the elasticity of city population with respect to within-city interstate highways is about 0.10 after 10 years. Beaudry, Green, and Sand (2014) find that the elasticity of city population with respect to city wages is about 2 after 10 years. This suggests an elasticity of city wages with respect to within-city interstate highways of 0.10/2=0.05. Assuming a more mobile labour force would lead to even smaller numbers.

Turning to the propensity to import, Appendix B shows that

$$\begin{aligned} \delta_j^M &= \log(N_j W_j) - MA_j^M \\ &= S(R_j) + \frac{\sigma-1}{\sigma}\log(A_j) + \frac{\sigma-1}{\sigma}\log(N_j) + \frac{\sigma}{(\sigma-1)^2} MA_j^X - MA_j^M\,, \end{aligned} \tag{15}$$

where MA_j^M is an import market access term defined in the same appendix. In this expression, the city's propensity to import depends on city roads, population, unobserved productivity, export market access and an import market access term.

Equation (15) gives us our next comparative static.

Comparative Static 4 The effect of roads on the propensity to import value should be positive but small.

The logic behind this comparative static is the same as the logic for comparative static 3. Roads affect imports only through wages and we expect the effect of roads on wages to be small.

We now derive expressions for the weight of equilibrium trade. For Q_{ij}^k units of the variety from sector k to reach city j, $\tau_{ij}^k Q_{ij}^k$ units need to be shipped from city i. The weight of those goods is $\tau_{ij}^k Q_{ij}^k V^k$. Using equations (4) and (6) and aggregating across all sectors implies that the weight of the exports from city i to city j is

$$T_{ij} = \frac{A_i}{W_i}\int_0^1 X_{ij}^k V^k \,\mathrm{d}k\,. \tag{16}$$

Using equations (4), (6), (9) and (10), the weight of trade can be expressed as

$$\log T_{ij} = \delta_i^T + (1-\sigma)\log\tau_{ij} + \delta_j^M, \tag{17}$$

where city i's propensity to export weight is

$$\delta_i^T = S^T(R_i) + \log A_i + \log N_i - MA_i^X\,, \tag{18}$$

and

$$S^T(R_i) \equiv \log\int_0^1 \left(\tau_i^k\right)^{1-\sigma} V^k \,\mathrm{d}k - \log\int_0^1 \left(\tau_i^k\right)^{1-\sigma} \mathrm{d}k\,. \tag{19}$$

A similar derivation leads to a corresponding expression for the propensity to import weight.

Equation (19) leads to a comparative static about roads and the weight of trade.

Comparative Static 5 The difference between the effect of within-city roads on the propensity to export value and weight is increasing in the heterogeneity of sector weights.

With one important difference, the determinants of the propensity to export weight in equation (18) are the same as those of the propensity to export value in equation (12); roads, productivity, population and market access. Equation (19) shows that the effect of roads on exported weight also depends on the difference between a term that aggregates the effect of city roads on transportation costs across sectors weighting them by the weight of goods in this sector and a similar term with unweighted transportation costs. If weight differences across sectors are small then this term will be small. That is, the effect of roads on exported weight increases in the heterogeneity of sector weights. From comparative static 3, we know that the effect of roads on exported value is small.

Comparative static 5 follows immediately. From table 2, we see that the weight to value ratio varies across sectors by about a factor of 42,000.

Our final and most important comparative static is intuitively obvious, but follows less immediately from the analysis.

Comparative Static 6 An increase in roads within a city causes a decrease in employment by sectors producing light goods and an increase in employment by sectors producing heavy goods.

An increase in city roads makes exporting less costly. Lower export costs lead to an increase in the demand for the varieties produced by this city, which in turn leads to a greater value of exported output. Thus, an increase in city roads is actually an increase in productivity that affects sectors producing heavy goods more than sectors producing light goods. This means that an increase in city roads alters the patterns of comparative advantage and causes greater specialization in the production and export of heavier goods.

Remarkably, we find that the data support all of the predictions of this admittedly stylized model except comparative static 1 (which requires that intercity road distance affect the weight and value of trade equally). Unsurprisingly, we find that trade in value is less sensitive to the road distance between cities than trade in weight. It appears to be straightforward to generate this prediction in our model by allowing intercity roads to affect transportation costs more in sectors producing heavier goods. However, this extension would come at a high price in complexity and our lack of bilateral trade data disaggregated by sector limits our ability to explore it empirically.

We close this section by noting that the analysis above treats city population as exogenous. We consider the implications of labor mobility across cities in Appendix B. Endogenizing population results in competing expressions for the propensities to export or import value or weight that do not depend on the productivity shifter A_i and that do depend on an import market access term. Given that import and export market access have a correlation of 0.97, this does not affect the estimates of our variable of interest.

5. Econometric model

Although we focus on the determinants of trade flows rather than wages, our econometric approach resembles the "new economic geography" estimations pioneered by Redding and Venables (2004). To begin, note that our theory provides little intuition about the form of the transportation cost term in equation (11), $(1-\sigma)\log\tau(R_{ij})$. Given this, we let $O_l(z)$ denote an order l polynomial in z and define

$$(1-\sigma)\log\tau(R_{ij}) \equiv O_l(\log R_{ij}) + \epsilon_{ij}. \tag{20}$$

That is, we write the transportation cost term of equation (11) as an arbitrary polynomial in pairwise road distance and a residual. Substituting (20) into (11) now leads naturally to our 'first-step' estimating equation,

$$\log X_{ij} = \delta_i^X + O_l(\log R_{ij}) + \delta_j^M + \epsilon_{ij}\,. \tag{21}$$

This is a "gravity" equation for trade flows with fixed effects. The error term ϵ includes approximation error that results from replacing the arbitrary between-city transportation cost function,

$\tau_{xm}(R_{ij})$, with a polynomial in highway distance and any component of $\tau_{xm}(R_{ij})$ unrelated to distance.

A necessary condition for an unbiased OLS estimate of the effect of road distance on trade flows is that our approximation error be uncorrelated with R_{ij}. Such a correlation might arise if more direct highway links connect pairs of cities that trade more. We pursue three strategies to address this problem. First, we consider different polynomial approximations of R, of order 1,2 and 4. Second, we consider estimates based on Euclidean rather than highway distance. Third, we conduct an instrumental variables estimation in which we use pairwise distance along the 1898 railroad and 1947 highway plan to predict 2005 interstate distance. Our estimates are robust to all of these variations. We discuss instrument validity below.

Within-city roads do not appear explicitly in estimating equation (21) or its theoretical precursor, equation (11). This follows immediately from the structure of transportation costs in equation (10). Roads in city i affect transportation costs to all of its trading partners equally. Thus, the effect of within-city roads helps to determine a city's propensity to trade, or, in the language of equation (21), its importer and exporter fixed effects. To determine the effect of within-city roads on trade, we conduct second-step regressions which predict importer and exporter fixed effects as a function of within-city roads.

Equation (12) suggests regressing the exporter fixed effect δ_i^X on within-city highways, productivity, population and market access. However, our model does not suggest a functional form for the relationship between roads and propensities to trade. Specifically, we see in equation (13) that the relationship between roads and the propensity to trade is determined by, $S(R_i)$, which depends on the unknown functions $\tau^k(R_i)$. While we experiment with other possibilities in our robustness checks, we usually use the logarithm of kilometers of interstate highway within a city as our measure of capacity of a city's road network. Hence, from equation (12) our second-step regression is

$$\hat{\delta}_i^X = \beta_0 + \rho_R^X \log R_i + \beta^{'} C_i + \mu_i\,, \tag{22}$$

where C_i is a vector of city i characteristics. Consistent with equation (12), in our baseline specification C_i contains log 2007 employment and a measure of market access computed according to equation (14) using the results of the first-step estimation of equation (21).

Inspection of equations (22) and (12) suggests a structural interpretation of the error term in this equation. Specifically, this error term consists of two components. The first reflects the difference between $S(R_i)$ and our approximation of this term, here $\rho_R^X \log R_i$. The second reflects unmeasured city productivity, A_i, which occurs explicitly in (12) but not in (22).

Our estimates of equation (22) must respond to four inference problems. First, the propensity to export may cause highway development in cities: more highways may be built in cities that export intensively (or more highways may be built to help cities that do not export much). Second, missing variables which affect both the propensity of cities to export value and the provision of city roads may confound our estimates. Third, the error arising from our approximation of $S(R_i)$ by $\rho_R^X \log R_i$ may be correlated with the propensity to export and R_i, and hence confound our estimates. Finally, our market access controls may themselves be endogenous.

We pursue three strategies to deal with these problems. To investigate the role of approximation error, we consider a variety of measures of the road network. To consider the role of omitted variables we experiment widely with control variables that may be correlated with city productivity. Finally, to address both reverse causation and omitted variable we instrument for within-city roads with instruments derived from our data on exploration routes, 1898 railroads and the 1947 highway plan. We consider a number of possible solutions to the issues raised by market access below.

We close this discussion with three final comments on our model. First, we expect that sectors with high transportation costs locate close to their markets. Thus, patterns of trade reflect both the direct effect of high transport costs and their indirect effect on the location of production. Separating these two effects is an open question in the literature and would require a source of exogenous variation in firm location, as well as a source of exogenous variation in transportation infrastructure. The fact that firm locations may reflect transportation costs affects the interpretation of our estimates but does not affect the validity of our estimates, provided that our instruments are not correlated with firm location choices. In our robustness checks, we investigate this possibility and fail to find evidence to support it.

Second, in equation (11), all else equal, a one unit increase in the propensity to export, δ_i^X, causes a one unit increase in the log value exported to each of i's trading partners. In this sense, changes in δ_i^X directly reflect changes in the log of exports. Thus, a change in δ_i^X resulting from a log change in another variable is an elasticity of export value. An analogous intuition applies to the propensity to import, δ_i^M. Also, note that the dependent variable in the second step is an estimate of the exporter fixed effect, $\hat{\delta}_i^X$, which is equal to its true value plus an error term. This affects the estimation of the standard errors for the coefficients estimated in equation (22). We follow a simple FGLS procedure to recover corrected standard errors.

Finally, equation (22) describes the relationship between within-city roads and a city's propensity to export value. We can derive a corresponding equation to describe a city's propensity to import value from equation (15). The main difference between the resulting expression and our expression for the exporter effect, equation (22), is that the importer effect equation also includes an expression for import market access. To derive the corresponding two-step estimation procedure for the effect of roads on the weight of trade, we use equations (17) and (19).

5.1 *Instrument validity*

Our three instruments; 1528-1850 exploration routes, 1898 railroads and 1947 planned highways, all predict the modern network of interstate highways. As a result, they also predict the number of kilometers of interstate highways within a city. Exploration routes result from a search for an easy route to travel on foot, horseback, or wagon. Since such a route will likely be a good route for a car, contemporary highways often follow exploration routes. A similar argument applies to 1898 railroads. In addition, building both railroad tracks and automobile roads requires leveling and grading a roadbed. Hence, an old railroad track is likely to become a modern road because old railroads may be converted to automobile highways without the expense of leveling and grading. Finally, the 1947 highway plan results from a prolonged effort by President Franklin D. Roosevelt, who began planning for a national highway system in 1937. Many interstate highways described by this plan were subsequently built. Building started after the 1956 Federal Aid Highway and Highway Revenue Acts and was completed by the early 1990s. More formally, table 12 in Appendix C reports a number of first-stage specifications predicting log 2005 highway kilometers where our instruments are highly significant. In addition, we report weak instrument test statistics developed by Stock and Yogo (2005) with our TSLS results. Unless otherwise mentioned, our instruments are not weak.

A valid instrument must also be orthogonal to the structural equation error term. We first consider the 1898 railroad network. A first argument for the validity of 1898 railroad kilometers rests on the fact that the US rail network was developed during and immediately after the civil war, and during the industrial revolution. At that time, the US economy was much smaller and more agricultural than today.[15] A second argument rests on the circumstances surrounding the development of the US rail network. Early US railroads were developed mainly to transport grain, livestock and lumber as well as passengers over long distances (Fogel, 1964, Fishlow, 1965, Cronon, 1991). These flows of people and agricultural commodities little resemble 2007 trade flows of (mostly) manufactured goods described by our data. Moreover, the rail network was constructed by private companies expecting to make a profit from railroad operations in a not too distant future.

While the 1898 railroad network plausibly determines contemporary highways for some reason unrelated to trade flows in 2007, 1898 rail might affect trade in 2007 through a number of other channels. First, larger cities in the late 19th century were more likely to receive railroads. Because population is persistent, large cities in 1898 tend to be large today and large cities trade more than small cities. To avoid this problem, we control for contemporaneous population as well as population in 1920 (the closest we can get to 1898 with existing data) in most of our regressions.

Second, more productive cities might have received more kilometers of railroad tracks in 1898. These cities might still be more productive today and, as required by our model and by common sense, export more as a result. Controlling for population in 1920, 1950 and 2000 helps to resolve this problem since we expect population to move to more productive cities: there is little evidence that city productivity is persistent conditional on population size (Glaeser and Gottlieb, 2009). In some regressions we also use income per capita and the share of adult population with a college degree to capture unobserved city productivity.

Third, 1898 railroads may cause cities to specialize in manufacturing and this specialization may be persistent through 1947 into the present. If so, the component of the modern road network predicted by the 1898 rail network will predict modern manufacturing simply because 1898 railroads predict modern manufacturing specialization. If so, the 1898 railroad network is not a valid instrument for the modern highway network in our regressions. We pursue two strategies to investigate this possibility. First, in separate appendix E we investigate the relationship between 1898 railroads and 1956 sectoral specialization. To the extent that our data can inform us about this

[15] At the peak of railroad construction, around 1890 (United States Bureau of Statistics, 1899, pp. 151 and 362), the US population was 55 million, with 9 million employed in agriculture or nearly 40% of the workforce (United States Bureau of Statistics, 1899, pp. 10 and 23). By 2007, the population of the US was 302 million with 2.2 million employed in agriculture, about 1.5% of the workforce (United States Bureau of the Census, 2011, pp. 18 and 399).

relationship, 1898 railroads do not appear to affect manufacturing specialization in 1956. Second, we can control for contemporary and 1956 manufacturing specialization and for the weight of a city's exports in 1956. We will see that, although these are important explanatory variables, they do not affect our estimates of the effect of modern highways on trade patterns. That is, neither investigation suggests that 1898 railroads affect modern patterns of trade and specialization through some mechanism that is correlated with, but different from modern roads.

Finally, there may be geographical features which are correlated with both 1898 railroads and 2007 trade flows. To condition them out we consider a variety of geographical controls such as distance to the nearest body of water, slope and census region fixed effects.

Our maps of exploration routes describe major expeditions of exploration ranging over three centuries. The motivations for these expeditions were as varied as the explorers and times in which they lived; from the search for the fountain of youth or gold, to the establishment of fur trading territories, to finding emigration routes to Oregon, or to the expansion of the US territory towards the Pacific Ocean.

Using different instruments, for which threats to validity differ, allows for informative over-identification tests. Old explorations routes are an attractive instrument in this respect. We are concerned that 1898 railroad routes are not a valid instrument because of the possible correlation between 1898 railroads and 2007 trade flows through persistent population patterns or persistent manufacturing specialization. However, early explorations of the US took place in areas that were sparsely populated by indigenous peoples where manufacturing was nonexistent. We are also concerned that a city's stock of 1898 railroads reflects persistent productivity and 2007 trade flows. This story seems implausible for early exploration routes. While some explorers were looking for gold and other minerals, these factors play little role in 2007 manufacturing production.

Our third instrument is 1947 planned highway miles. The 1947 plan was first drawn to *"connect by routes as direct as practicable the principal metropolitan areas, cities and industrial centers, to serve the national defense and to connect suitable border points with routes of continental importance in the Dominion of Canada and the Republic of Mexico"* (United States Federal Works Agency, Public Roads Administration, 1947, cited in Michaels, 2008). Historical evidence confirms that the 1947 highway plan was, in fact, drawn to this mandate (see Mertz, undated, and Mertz and Ritter, undated, as well as other sources cited in Chandra and Thompson, 2000, Baum-Snow, 2007, and Michaels, 2008).

Planned 1947 highways, like 1898 railroads and 1528-1850 exploration routes also face some threats to validity. First, planned highways could be correlated with 2007 trade flows and with some persistent determinant of population. Duranton and Turner (2012) find that 1947 planned highways are uncorrelated with population growth in the 1940s and 1950s. Thus, in accordance with their mandate, planners in 1947 tried to connect population centers, not to anticipate future population levels and trade patterns. From this, it follows that controlling for 1950 and 2007 population levels should condition out persistence in population as a threat to the validity of this instrument.

6. The effect of highways on trade

6.1 First-step results

Table 3 reports estimates of cities' pairwise trade in weight and value using the gravity specification (21) for the value of trade or the corresponding equation for weight. In panel A the dependent variable is the weight of bilateral trade by road. Panel B reports corresponding results for value. Samples differ across the two sets of regressions because of the higher incidence of censoring in the CFS weight data.

Column 1 reports an OLS regression of trade flows on exporter and importer fixed effects and log 2005 highway distance. Column 2 reports the corresponding TSLS regression using network distances from 1947 planned highways and 1898 railroads as instruments for 2005 highway distance. Columns 3 and 4 add a quadratic term in log 2005 highway distance to the regressions of column 1 and 2, while columns 5 and 6 include cubic and quartic terms. Column 7 returns to the simple specification of column 1 but uses Euclidian distances. Column 8 mixes Euclidian and highway distances. Although we report only trade elasticities with respect to distance the regressions of table 3 also estimate importer and exporter fixed effects. These are our propensities to import and export and are the dependent variables in our second-step regressions. We first discuss the elasticities reported in table 3 before turning to the fixed effects.

In column 1 the elasticity of trade with respect to distance is -1.90 for weight and -1.41 for value. Column 2 shows that instrumenting 2005 highway distances with 1898 railroad and planned 1947 highway distances does not change these estimates. Columns 3-6 confirm the similarity of OLS and IV results. Although column 3 and 4 column indicate the effect of highway distance is U-shaped, trade always decreases with distance for pairwise distances in our sample. Comparing column 7 with column 1 we see that trade elasticities are about the same for Euclidean and highway distances. That each of these specifications yields similar elasticity estimates reflects the high correlation between Euclidean distances and network distances along 1898 railroads, planned 1947 highways and 2005 highways.

Our estimates of the mean elasticity of trade with respect to distance range between -1.63 and -1.91 for weight and between -1.17 and -1.41 for value. From equation (11), we see that a larger elasticity of substitution between goods implies a greater sensitivity of trade to distance. Since the US is a highly integrated economy trading highly substitutable goods, this suggests that our estimates should be at least marginally larger than what is found for cross-country trade. In fact, our estimates are at the high end of those collected by Disdier and Head (2008).

In a separate web appendix (Appendix D) we show that road trade is more sensitive to distance than all trade. Consistent with this, we also show that elasticities of trade with respect to distance are smaller for rail trade than road trade. Finally, we show that the elasticity of trade with respect to distance is moderately sensitive to how we calculate the distance for a trade between a city and itself.

Estimates of the exporter and importer fixed effects are stable across specifications. The pairwise correlations between the sets of exporter effects implied by panel A are all 0.99 or above. The same

Table 3: First-step results

	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS	(5) OLS	(6) TSLS	(7) OLS	(8) OLS
Panel A. Dependent variable: (log) Weight of bilateral trade flows, road trade.								
log(hwy. dist.)	-1.90***	-1.90***	-2.86***	-2.87***	-3.72	-1.49		-1.84***
	(0.023)	(0.022)	(0.17)	(0.16)	(6.95)	(6.57)		(0.20)
log(hwy. dist.)^2			0.077***	0.078***	0.76	0.22		
			(0.013)	(0.013)	(1.77)	(1.67)		
log(hwy. dist.)^3					-0.13	-0.070		
					(0.19)	(0.18)		
log(hwy. dist.)^4					0.0071	0.0050		
					(0.0078)	(0.0075)		
log(Euclid dist.)							-1.91***	-0.062
							(0.024)	(0.20)
Mean effect	-1.90	-1.90	-1.74	-1.74	-1.63	-1.63	-1.91	-1.90
Median effect	-1.90	-1.90	-1.72	-1.72	-1.73	-1.71	-1.91	-1.90
R^2	0.86	-	0.87	-	0.87	-	0.86	0.86
First stage F		153,426		20,514		2,206		
Panel B. Dependent variable: (log) Value of bilateral trade flows, road trade.								
log(hwy. dist.)	-1.41***	-1.41***	-2.18***	-2.19***	-1.82	-0.53		-1.32***
	(0.019)	(0.019)	(0.14)	(0.14)	(5.96)	(5.77)		(0.16)
log(hwy. dist.)^2			0.062***	0.062***	0.44	0.13		
			(0.011)	(0.011)	(1.52)	(1.47)		
log(hwy. dist.)^3					-0.092	-0.061		
					(0.17)	(0.16)		
log(hwy. dist.)^4					0.0057	0.0046		
					(0.0068)	(0.0066)		
log(Euclid dist.)							-1.41***	-0.088
							(0.020)	(0.16)
Mean effect	-1.41	-1.41	-1.28	-1.28	-1.18	-1.17	-1.41	-1.41
Median effect	-1.41	-1.41	-1.27	-1.27	-1.26	-1.25	-1.41	-1.41
R^2	0.83	-	0.83	-	0.84	-	0.83	0.83
First stage F		161,034		20,163		2,187		

Notes: All regressions include importer and exporter fixed effects for all cities. The both panels report the same regressions but use different dependent variables: weight of trade flows for panel A and value of trade flows for panel B. Regressions in panel A are based on 2,476 observations and 2,705 observations in panel B. In column 2, 4 and 6, highway distance terms are instrumented by their corresponding 1947 planned highway and 1898 railroad distance terms. Robust standard errors in parentheses. ***, **, *: significant at 1%, 5%, 10%.

holds for importer fixed effects. The corresponding correlations for the fixed effects estimated in panel B are at least equally strong.[16] Our second-step regressions, which we report below, generally rely on importer and exporter fixed effects estimated in column 6, although in our robustness checks we experiment with importer and exporter fixed effects based on other first step specifications.

It is also of interest to investigate the relative importance of physical remoteness and highway distance for trade flows. To address this issue, column 7 uses log Euclidean pairwise distance rather than log network distance as a control, while column 8 includes both Euclidean and network distance measures. In column 7 we see that the effect on trade flows of a change in Euclidean distance is almost exactly the same as the effect of a change in network distance. In column 8 we see that the network distance coefficient is significant and of about the same magnitude as in column 1, while the coefficient of Euclidean distance is tiny and not statistically different from zero. Thus column 8 strongly supports the hypothesis that it is travel distance and not physical proximity which determines trade flows.

6.2 Second-step results

We now turn to estimates of equation (22) in order to examine the effect of within-city roads on the propensity to trade. Table 4 reports results for OLS regressions using the exporter fixed effects estimated in the first step as the dependent variable. The second through the fourth rows report; robust standard errors, OLS standard errors and FGLS standard errors. Our preferred specification uses fixed effects estimates based on column 6 of table 3. Column 1 is a rudimentary OLS specification using only log 2005 lane kilometers of within-city highways as an explanatory variable. Column 2 is the simplest specification consistent with our model (equation 12) and includes within-city highways, employment and market access. We compute market access as suggested by equation (14), using the results from the same first-step specification that we use to estimate exporter fixed effects.[17] Column 3 controls for log population from 1920, 1950 and 2000.

Column 4 also controls for the log share of manufacturing employment in 2003. Our model assumes that all goods are tradable when the output of many service industries actually is not. This means that column 3 measures the total effect of within-city highways on the extensive margin (the production of tradable vs. non tradable goods) and the intensive margin (the production of heavier goods within manufacturing), whereas the specification in column 4 measures only the intensive margin.

Columns 5 to 8 perform the same regressions as columns 1 to 4, but use exporter value fixed effects rather than exporter weight fixed effects as the dependent variable.

[16]In addition, the correlations between the weight and value fixed effects estimated from column 6 are 0.92 for exporters and 0.97 for importers. The correlations between importer and exporter fixed effects are 0.73 for weight and 0.75 for values.

[17]We do not include a city's own importer effect when doing this computation: to do so could lead to simultaneity biases. To see this, consider a situation where some cities import intermediate goods and export final goods. We return to this issue in our robustness checks when we consider alternative measures of market access.

Table 4: Second-step results, OLS for exporter fixed effects

Exporter fixed effect	(1) weight	(2) weight	(3) weight	(4) weight	(5) value	(6) value	(7) value	(8) value
log highway km	1.17***	0.53***	0.38**	0.27**	1.26***	0.24	0.094	-0.037
Robust s.e.	(0.14)	(0.18)	(0.14)	(0.13)	(0.16)	(0.19)	(0.15)	(0.15)
Non-robust s.e.	(0.12)	(0.16)	(0.16)	(0.14)	(0.14)	(0.17)	(0.16)	(0.13)
Corrected s.e.	(0.12)	(0.18)	(0.16)	(0.16)	(0.14)	(0.17)	(0.16)	(0.16)
log employment		0.55***	0.73*	0.49		0.88***	1.19*	0.90*
		(0.12)	(0.41)	(0.36)		(0.11)	(0.64)	(0.46)
Market access (export)		-0.45***	-0.66***	-0.65***		-0.18	-0.38**	-0.36***
		(0.14)	(0.15)	(0.12)		(0.12)	(0.15)	(0.11)
log 1920 population			-0.38	-0.29			-0.35	-0.23
			(0.27)	(0.25)			(0.32)	(0.33)
log 1950 population			1.02**	0.65			0.95*	0.49
			(0.43)	(0.42)			(0.52)	(0.55)
log 2000 population			-0.74	-0.17			-0.85	-0.13
			(0.51)	(0.49)			(0.79)	(0.64)
log % manuf. emp.				0.66***				0.83***
				(0.13)				(0.17)
R^2	0.59	0.73	0.79	0.84	0.56	0.77	0.81	0.88

Notes: 66 observations per column. All regressions include a constant. Robust standard errors in parentheses unless otherwise mentioned. ***, **, *: significant at 1%, 5%, 10%.

The coefficient on city highway kilometers is large for both trade in weight and value in columns 1 and 5. When we control for employment and export market access in columns 2 and 6 the highways coefficients decrease and remain significant for trade in weight but not value. Adding controls for 1920, 1950 and 2000 population in columns 3 and 7 further reduces the highways coefficient. Finally, controlling for manufacturing employment reduces the coefficient on within-city highways still further, to 0.27, for trade in weight in column 4. In column 8, this coefficient is small and insignificant for trade in value.

Other results also merit discussion. First, the exporter value fixed effect is roughly proportional to employment in columns 6 to 8, while the exporter weight fixed effect is less than proportional to employment. Bigger cities export lighter goods. This is consistent with well known facts about patterns of specialization in US cities (e.g., Duranton and Puga, 2000). Second, the coefficients on market access are negative and generally significant. This confirms our third comparative static. Third, correcting for the sampling error associated with our use of an estimated dependent variable affects the standard errors only slightly. Given this, we suppress corrected standard errors in subsequent estimations.

Panel A of table 13 in Appendix C replicates table 4 using importer rather than exporter fixed

Table 5: Second-step results, TSLS for exporter fixed effects

Exporter fixed effect	(1) weight	(2) weight	(3) weight	(4) weight	(5) value	(6) value	(7) value	(8) value
log highway km	1.13***	0.57***	0.47***	0.39***	1.10***	0.17	0.070	-0.028
	(0.14)	(0.16)	(0.14)	(0.12)	(0.17)	(0.16)	(0.14)	(0.12)
log employment		0.52***	0.69*	0.46		0.91***	1.20**	0.89**
		(0.11)	(0.39)	(0.34)		(0.091)	(0.59)	(0.43)
Market access (export)		-0.45***	-0.65***	-0.63***		-0.19	-0.38***	-0.36***
		(0.14)	(0.14)	(0.11)		(0.12)	(0.14)	(0.11)
log 1920 population			-0.38	-0.29			-0.35	-0.23
			(0.25)	(0.23)			(0.31)	(0.30)
log 1950 population			1.00**	0.64*			0.95*	0.49
			(0.40)	(0.38)			(0.49)	(0.52)
log 2000 population			-0.74	-0.18			-0.85	-0.13
				(0.49)				(0.74)
log % manuf. emp.				0.64***				0.83***
				(0.12)				(0.16)
Overid. p-value	0.100	0.043	0.15	0.30	0.081	0.071	0.28	0.55
First-stage Stat.	97.5	90.3	80.4	85.2	97.5	90.3	80.4	85.2

Notes: 66 observations per column. All regressions include a constant and use log 1947 planned highway km, log 1898 railroad km and log 1528-1850 exploration routes index as instruments for log kilometers of interstate highways. The Stock-Yogo critical values for the first-stage statistics are 13.91 for the relative IV bias (at 5%) and 22.30 for the IV size bias (at 10%). Robust standard errors in parentheses. ***, **, *: significant at 1%, 5%, 10%.

effects as the dependent variable.[18] The results are weaker than with exporter fixed effects in table 4. Our model predicts that the weight of imports should not be affected directly by within-city roads because the cost of entering a city is the same for all sectors, and this prediction is broadly confirmed in the data. We impose this assumption primarily to facilitate the analytic solution of our model, but we might expect that roads cause sectors producing heavier goods to face relatively lower costs to enter a city with more highways. Given this, we should expect to see that cities import relatively heavier bundles of goods as their within city road network improves. That we are unable to measure such an effect in our data suggests that changes in relative prices are small relative to their elasticities of substitution. A similar statement applies to the import of intermediate products. Our data does not appear to allow us to pursue this hypothesis further.

Table 5 mirrors table 4 but uses TSLS estimation and instruments 2005 kilometers of interstate highways with 1528-1850 exploration routes, 1898 railroads and 1947 planned highways. The TSLS and OLS coefficient on instrumented lane kilometers of highways are less than one standard deviation apart in all specifications. The coefficients of other explanatory variables are also stable.

[18]Following the specification derived in Appendix B, regressions in table 13 control for both an importer and an exporter market access variable, although the 0.97 correlation between importer and exporter market access makes separate identification of their coefficients impractical.

These small differences between OLS and TSLS estimates suggest that highways allocated to cities at random have about the same effect on trade as highways allocated through the prevailing allocation process.

The first-stage statistics show that the instruments are strong. We pass overidentification tests when we control for population in 1920 (shortly after the construction of railroads), in 1950 (at the time of the design of the highway plan) and 2000 (to control implicitly for long run growth since 1950). That is, the decadal population variables control for persistent productivity that determines the assignment of highways to cities and trade flows. For this reason, column 3 in table 5 is our preferred specification.

In column 2, the elasticity of weight of exports with respect to lane kilometers of highways is 0.57 and highly significant. For the value of trade, the corresponding coefficient in column 6 is smaller at 0.17 and insignificant. In our preferred specification, column 3, the elasticity of the weight of exports with respect to lane kilometers of highways is 0.47 and significant. For the value of trade in column 7, the coefficient is again smaller and insignificant.

Our comparative static 5 suggests that the weight of trade should be sensitive to within-city highways if differences in weight per unit of value are large across sectors. In fact, table 1 shows that these differences are large for contemporary US manufacturing industries. On the other hand, comparative static 3 suggests that the elasticity of the value of trade with respect to highways should be low. In table 5 we see that the coefficient on within-city highways is larger for trade in weight than for trade in value, and that this difference is statistically significant. We also see that effect of roads on the value of trade, is statistically indistinguishable from zero with a small point estimate. This is consistent with comparative static 3 which predicts that the effect of roads on the value of exports will be small.

In column 4 we add the share of manufacturing employment as a control. At 0.39, the coefficient on within-city highways is slightly smaller than in column 3 and remains significant at 1%. The corresponding coefficient for trade in value in column 8 remains insignificant. The small difference in the within-city highway coefficient between columns 3 and 4 suggests that highways increase the weight of trade by affecting the composition of exports rather than by promoting more manufacturing employment. This is also consistent with the small and insignificant effect of within-city highways on exported values.

The effect of within-city highways on the weight of exports is economically large. Using equations (21) and (22), we calculate that multiplying the stock of within-city highways by a factor α changes the weight exported by a factor of $\alpha^{0.47}$. Thus, doubling the stock of within-city highways causes a 39% increase in expected export weight. To place this magnitude in the context of our sample, we consider Milwaukee (WI) and Indianapolis (IN). These two cities have roughly the same CFS area population, 1.7 and 1.8 million. Milwaukee is in the second decile in the distribution of interstate lane kilometers, Indianapolis is in the ninth, and Indianapolis has 151% more highways. Our estimates predict that Indianapolis should export $2.51^{0.47} = 1.54$ or 54% more tons than Milwaukee. This prediction slightly understates the 56% difference in exported tons between these two cities in the CFS data, although the value of exports from the two cities is approximately equal.

Panel B of table 13 in Appendix C performs the same TSLS regressions as in table 5, but uses

Table 6: Second-step results, robustness to choice of instruments and econometric technique

	(1) TSLS	(2) TSLS	(3) TSLS	(4) TSLS	(5) TSLS	(6) TSLS	(7) LIML	(8) GMM
Panel A. Employment, market access and past populations as controls.								
log highway km	0.50***	0.45***	0.55***	0.49***	0.83***	0.11	0.47***	0.41***
	(0.15)	(0.14)	(0.17)	(0.15)	(0.25)	(0.31)	(0.14)	(0.13)
Instruments								
log 1528-1850 exploration	N	Y	Y	N	N	Y	Y	Y
log 1898 railroad km	Y	N	Y	N	Y	N	Y	Y
log 1947 highway km	Y	Y	N	Y	N	N	Y	Y
Overid. p-value	0.061	0.24	0.089	.	.	.	0.043	0.15
First-stage Stat.	70.1	105	46.1	139	45.4	14.9	80.4	80.4
Panel B. Employment, market access, past populations and manufacturing share of employment as controls.								
log highway km	0.39***	0.38***	0.51***	0.38***	0.64***	0.34*	0.39***	0.36***
	(0.13)	(0.12)	(0.16)	(0.13)	(0.22)	(0.19)	(0.13)	(0.12)
Overid. p-value	0.12	0.83	0.25	.	.	.	0.30	0.30
First-stage Stat.	64.1	113	46.7	128	40.9	23.9	85.2	85.2

Notes: 66 observations per column. The dependent variable is exporter fixed effect for weight in all columns of both panels. Regressions in panel A use log 2007 highway kilometers, log 2007 employment, export market access, and log population for 1920, 1950 and 2000 as explanatory variables and include a constant. Panel B also controls for the log share of 2003 manufacturing employment. Each column of panel B replicates the instrumentation strategy of the corresponding column of panel A. Robust standard errors in parentheses. ***, **, *: significant at 1%, 5%, 10%.

importer fixed effects as the dependent variable, instead of exporter fixed effects. The coefficient on highways for the weight of imports is positive but small and insignificant. The coefficient on highways for the value of imports is also insignificant with small point estimates. These TSLS results confirm the OLS results of panel A of the same table. As predicted by comparative static 4, highways do not appear to affect the propensity to import value (nor weight).[19]

7. Robustness of main results

We now verify the robustness of our estimates. We consider the importance of: instrument choice; the possible persistence of manufacturing specialisation in sectors producing heavy goods; alternate control variables; the possible endogeneity of our market access control variables; alternate specifications of our first-step regression; and an alternate one-step estimate strategy.

[19] That highways foster exports in weight but not value is an important result against the alternative hypothesis that the effect of highways on exports is only due to an increase in wholesale and warehousing activity (which should arguably affect values as well as weights. We return to this issue below.)

7.1 *Instrument choice*

Table 6 checks the robustness of our TSLS estimations to choice of instruments and econometric technique. Columns 1 to 3 of panel A replicate column 3 of table 5, our preferred specification for exporter fixed effects in weight, but considers pairs of our three instruments. The different pairwise combinations of instruments yield the same results as our three instruments taken together. The coefficient on within-city highways varies between 0.45 and 0.55, close to our estimate of 0.47 in table 5. Columns 4 to 6 also replicate this specification, but use the instruments one at a time. While planned 1947 highways yield a coefficient close to our preferred estimate of 0.47 in table 5, the coefficient on highway kilometers obtained from 1898 railroads is larger at 0.83 whereas that obtained from old exploration routes is smaller and insignificant at 0.12.

Panel B reports the same regressions as panel A but adds the log share of manufacturing employment as a control variable, as in column 4 of table 5. In columns 5 and 6, the difference between regressions using old exploration routes and 1898 railroads as instruments decreases, as do the p-values for overidentification tests in columns 1 to 3. Thus, the difference in estimates caused by instrument choice in panel A reflects the fact that old exploration routes are negatively correlated with the share of employment in manufacturing while 1898 railroads are positively correlated with this variable. This is may reflect the fact that more explorations took place in areas with difficult geography and less subsequent manufacturing.

Finally, in columns 7 and 8 of both panels, we replicate the estimation of columns 3 and 4 of table 5 using LIML and IV-GMM instead of TSLS.[20] LIML does not change our results, while IV-GMM leads to a slightly smaller coefficient on within-city highways.

7.2 *Other robustness checks*

We now present a summary of a number of other robustness checks. The details of these results are reported in separate web appendices. First, we may worry that 1898 railroads caused cities to specialize in heavy manufacturing and that such early specialisation was persistent. This is a natural worry given that rail transportation played an important role in American industrialization (Atack, Haines, and Margo, 2011) and that patterns of economic activity are persistent (Bleakley and Lin, 2012). In separate Appendix E, we show that, despite some persistence in city specialisation in sectors producing heavy goods, there is no evidence that our IV estimates of the effect of interstate highways on the weight of exports reflects early manufacturing specialization caused by the 1898 railroad network. We show this by augmenting the specifications of table 5 with two further controls, the share of manufacturing employment in 1956 and a weight index for 1956 manufacturing employment. We construct the 1956 weight index by interacting 1956 manufacturing sector employment shares with contemporary weight per unit value for these sectors and summing over sectors. The resulting estimates are virtually indistinguishable from those of column 4 of table 5 where the contemporaneous share of manufacturing employment

[20]Limited information maximum likelihood (LIML) is a one-stage IV estimator. Compared to TSLS, it provides more reliable point estimates and test statistics with weak instruments. As for estimating instrumental variable regressions with the generalized method of moments, it has a clear advantage in the presence of heteroskedasticity of unknown form.

is already included. Second, as noted earlier, we fail to find evidence that 1898 railroads affect specialization patterns in 1956.

More generally, our instruments may still fail the exclusion restriction associated if they are correlated with a variable that is not among our controls and affects a city's propensity to export. In Appendix F, we consider several other variables that might be correlated with both our instruments and exports. More specifically, we augment our OLS and TSLS specifications in columns 3 and 4 of table 4 and 5 with variables describing: average distance to the nearest body of water; land gradient; dummy variables for census regions; log share of the fraction of adult population with a college degree or more; log average income per capita; log share of employment in wholesale trade; and log average daily traffic on the interstate highways in 2005. None of our main results is affected by these controls, even when we use each of our instruments individually. It follows that if our instrumental variables estimates are confounded by a missing variable, then this missing variable would need to be correlated in the same way with all three instruments and not be related to any of the variables we listed above.

In our main tables of results, tables 4 and 5, our measure of within-city roads is the log of within-city interstate highway kilometers, so that its coefficient is an elasticity. However, nothing in our theory suggests either this particular measure or our choice of functional form. In Appendix G, we plot the residuals of the regressions in tables 4 and 5 and log city kilometers of highway to show that the relationship between exports and kilometers of within-city highways is well approximated by a log-log specification. We also consider an alternative, network-based measure of within-city highways, the log of the number of radial rays of interstate highways. This yields similar results for both OLS and TSLS specifications. We also experiment with other alternatives such as weighting interstates highway kilometers by their lanes or considering only 'urban' interstates. These other measures of interstate highways generally yield similar results.

Among our control variables, market access appears to play an important role. Regardless of the specification, its coefficient is always large in magnitude and highly significant. Ignoring market access in our preferred OLS specification raises the coefficient on interstate highways by one standard deviation. In Appendix H, we consider a variety of specifications in which we use alternative measures of market access computed directly from population or aggregate income in other cities and the distance of these cities. We also experiment with the possible simultaneity of market access by instrumenting contemporaneous market access with a measure of market access using 1920 populations. Relative to the results of our preferred specification, the coefficient of within-city highways is about the same, while the coefficient for market access remains between -0.5 and -0.8. Overall, we find that our treatment of the market access variable has little effect on the estimate of the effect of within-city highways.

Our dependent variable, the propensity of cities to export weight, is estimated from a particular first-stage regression. In Appendix I, we verify that our results are not sensitive to the exact manner with which we estimate the propensity of cities to export weight. We obtain similar results when weighting observations in the first step, estimating it with a type 2 TOBIT to explicitly account for censoring of pairwise trade flows, using the output of each first-step variant reported in table 2, assigning a log pairwise distance of zero to all within-city trades, and ignoring within-city trade altogether. None of these alternatives affects our estimates of the effect of within-city highways on exports.

Finally, although our two-step approach is a natural outcome of our model but one could also estimate the effect of within-city highways on exports in a single step. In Appendix J, we report results showing that our conclusions are not affected by our choice of a two-step vs. a single-step estimation.

8. Extensions

8.1 Short- and long-distance trade

Our description of trade costs, equation (10), requires that the cost of exiting a city is proportional to a fixed fraction of a shipment regardless of the final destination. Alternatively, common sense suggests that within-city roads may matter less for shipments over longer distances.

To investigate this conjecture, we construct subsamples of shipments travelling less than 1,000 kilometers and more than 1,000 kilometers, and repeat our two-step estimation for each subsample. Shipments travelling less than 1,000 kilometers account for more than 90% of all trade in weight in our data. Panels A and B of table 7 report results. Columns 1 and 2 reproduce the two OLS specifications of columns 3 and 4 of table 4 for trade in weight. Columns 3 and 4 reproduce the two corresponding TSLS specifications of columns 3 and 4 of table 5. Finally, columns 5 to 8 repeat the same exercise for trade in value.

Panel A reports coefficient estimates based on shipments travelling less than 1,000 km. Interestingly, we see that the elasticity of exports with respect to roads is larger for trading distances below 1,000 km than for all trade. For our preferred specification, the coefficient on within-city highways is 0.81 for trading distances less than 1,000 kilometers, instead of 0.47 for all trade. On the other hand, for shipments travelling more than 1,000 kilometers, in panel B, the coefficient on within-city highways is 0.23 and insignificant. Hence, within-city highways have a larger effect on shipments travelling shorter distances. We note that less than 6% of tons trade over distances father than 1,000 kilometers.[21] In panel C of table 7 we also consider trade over long distances, but here define long distance trade to be shipments over 700 kilometers. The TSLS coefficient on within-city highways for trade is now significant and, at 0.27, closer to our preferred estimate of 0.47 for all trade. This suggests that within-city highways are irrelevant only to the small proportion of shipments travelling more than 1000 km. We also note in columns 5, 7 and 8 of panel A that the estimates for the effects of within-city highways on trade in value are now significant and larger than their corresponding estimates of tables 4 and 5.

We replicate table 7 for imports. In a separate web appendix (Appendix K). Much as for our other results, we find that the coefficients of within-city highways are generally insignificant except in the case of imported weights at distance below 1,000 kilometers where they are marginally significant, positive and small.

[21]From the figure, 1000 km is at the 82nd percentile of exports to other cities and from table 1 that only 31% of tons are exported to other cities. Thus, we have $0.31 \times 0.18 = 0.06$.

Table 7: Second-step results for short and long distance trade

Exporter fixed effect	(1) weight OLS	(2) weight OLS	(3) weight TSLS	(4) weight TSLS	(5) value OLS	(6) value OLS	(7) value TSLS	(8) value TSLS
Panel A. Short distance trade (less than 1,000 km).								
log highway km	0.65***	0.56***	0.81***	0.75***	0.31**	0.20	0.36**	0.29**
	(0.15)	(0.15)	(0.16)	(0.15)	(0.15)	(0.15)	(0.15)	(0.13)
% manuf. emp.	N	Y	N	Y	N	Y	N	Y
R^2	0.75	0.79			0.81	0.87		
Overid. p-value			0.18	0.39			0.29	0.25
First-stage Stat.			80.8	85.5			80.8	85.5
Panel B. Long distance trade (more than 1,000 km)								
log highway km	0.16	0.084	0.23	0.17	-0.098	-0.22	-0.15	-0.24
	(0.16)	(0.16)	(0.16)	(0.15)	(0.19)	(0.19)	(0.17)	(0.16)
R^2	0.73	0.77			0.78	0.84		
Overid. p-value			0.025	0.058			0.12	0.22
First-stage Stat.			83.0	87.9			83.0	87.9
Panel C. Long distance trade (more than 700 km)								
log highway km	0.23	0.16	0.31*	0.27*	-0.058	-0.16	-0.088	-0.16
	(0.17)	(0.17)	(0.16)	(0.15)	(0.17)	(0.17)	(0.16)	(0.15)
R^2	0.77	0.79	0.77	0.79	0.80	0.85	0.80	0.85
Overid. p-value			0.0068	0.022			0.096	0.17
First-stage Stat.			85.0	89.3			85.0	89.3

Notes: 66 observations per column. All regressions use log 2007 employment, export market access, and log population for 1920, 1950 and 2000 as explanatory variables and include a constant. All TSLS regressions use log 1947 planned highway km, log 1898 railroad km and log 1528-1850 exploration routes index as instruments for log kilometers of interstate highways. Robust standard errors in parentheses. ***, **, *: significant at 1%, 5%, 10%.

Overall these results suggest that within-city highways matter more for shipments travelling less than 1000 km than for longer shipments. Our multiplicative model of trade costs implies instead that log within-city highways should have the same coefficient on long- and short-distance shipments. It is natural to suspect that an additive specification would imply that city roads matter more for short distance trade. With additive transportation costs, the cost of exiting a city represents a larger fraction of the total cost of trade.

The case for such a model is ambiguous. An additive specification of trade costs would make our model intractable and represents a long standing challenge in the international trade literature. More importantly, additive specifications for trade costs imply that aggregate trade flows no longer follow the standard gravity specification we use in the first step of our analysis. The results of table 3 show that a simple gravity specification is an excellent first approximation for bilateral trade flows.[22] While we leave a definitive resolution of this puzzle to future research, in a separate web appendix (Appendix L) we explore possible biases in our estimates if the true effect of within-city roads is purely additive. In a simplified partial equilibrium version of our model, we find that, unless the cost of exiting a city represents an implausibly large fraction of the total cost of trade, we are likely to underestimate the true elasticity of transportation costs with respect to city roads.

8.2 Trade internal to cities

Trade within cities is an important component of trade. From table 2 we see that internal trade accounts for about 69% of the weight and 38% of the value of all trade in our sample. In this section we investigate the effect of within-city highways on internal trade.

We note that our model delivers some predictions for the log share of internal trade. We develop these in a separate web appendix (Appendix M). A caveat is that the exact specification depends on the details of how we model transportation cost for trade within cities. To avoid having results that rely on arbitrary choices made about this issue in the theoretical model, we experiment with a variety of specifications in table 8.

Panel A of table 8 reproduces the same OLS regressions as table 4 but uses the log weight and the log value of shipments internal to the city. That is, it conducts estimates of equation (22) with the log of internal trade, by weight or value, as the dependent variable. Panel B of table 8 replicates the same exercise for the TSLS regressions of table 5.

Comparing the results of panels A and B to those of tables 4 and 5, we find that the coefficients on within-city highways are slightly larger than their corresponding estimates in table 4 or 5 but that the difference is less than one standard deviation. Although underlying regressions are quite different, these coefficients are comparable: in table 8, within-city highway coefficients are elasticities of internal trade, while in tables 4 and 5 coefficients are also elasticities of inter-city trade.

Interestingly, if we compare the estimates of panels A and B of table 8 to those obtained in panel A of table 7, which focuses on short distance trade, the effects are smaller. Hence, although within-city highways have a slightly larger effect on internal trade than on all trade, this effect is smaller than that on trade taking place at distances below 1,000 km.

Panel C of table 8 reproduces the regressions of panels A and B, but uses the log share of internal shipments as the dependent variable. Panel D adds log internal distance to the specifications used in panel C. These regressions allow us to check whether within-city highways cause cities to change their trading patterns in favor of internal trade. These two sets of regressions do not produce robust estimates. To the extent that within-city highways do shift a city's trade patterns in favor of internal trade, this effect is too small to measure in our sample. This is consistent with panels A and B, which suggest that the effect of within-city highways on trade is about the same as the effect of within-city highways on external trade.

[22]To be precise,with a constant elasticity for the cost of distance, an additive specification for the cost of exiting a city would no longer imply a gravity specification since log trade would no longer be proportional to log distance except in some extremely specific cases.

Table 8: Trade internal to cities

Exporter fixed effect	(1) weight	(2) weight	(3) weight	(4) weight	(5) value	(6) value	(7) value	(8) value
Panel A. Dependent variable: log internal trade, OLS.								
log highway km	1.26***	0.38**	0.41**	0.38**	1.45***	0.33***	0.26**	0.20*
	(0.14)	(0.17)	(0.16)	(0.17)	(0.12)	(0.12)	(0.12)	(0.11)
Controls.	0	1	2	3	0	1	2	3
R^2	0.63	0.82	0.82	0.83	0.66	0.90	0.91	0.92
Panel B. Dependent variable: log internal trade, TSLS.								
log highway km	1.22***	0.54***	0.56***	0.54***	1.31***	0.31***	0.27**	0.22*
	(0.13)	(0.15)	(0.16)	(0.16)	(0.13)	(0.12)	(0.12)	(0.11)
Overid. p-value	0.72	0.77	0.67	0.84	0.26	0.21	0.25	0.65
First-stage Stat.	97.5	90.3	80.4	85.2	97.5	90.3	80.4	85.2
Panel C. Dependent variable: log share of internal trade, TSLS								
log highway km	0.043	-0.035	-0.030	-0.024	0.13*	0.064	0.084	0.10
	(0.030)	(0.035)	(0.034)	(0.034)	(0.068)	(0.081)	(0.078)	(0.074)
Overid. p-value	0.13	0.20	0.098	0.14	0.073	0.098	0.28	0.42
First-stage Stat.	97.5	90.3	80.4	85.2	97.5	90.3	80.4	85.2
Panel D. Dependent variable: log internal trade, TSLS controlling for internal distances.								
log highway km	0.95***	0.33	0.33	0.32	1.09***	0.36**	0.16	0.13
	(0.24)	(0.22)	(0.24)	(0.22)	(0.24)	(0.17)	(0.20)	(0.20)
log internal distance	0.61*	0.40	0.40	0.39	0.49	-0.081	0.20	0.16
	(0.35)	(0.26)	(0.30)	(0.28)	(0.40)	(0.22)	(0.29)	(0.26)
Overid. p-value	0.37	0.57	0.54	0.75	0.093	0.21	0.20	0.58
First-stage Stat.	34.9	42.8	31.7	34.9	34.9	42.8	31.7	34.9

Notes: 66 observations per column. The set controls 0 is a constant; 1 adds log 2007 employment and export market access; 2 further adds log population for 1920, 1950 and 2000; 3 also consider log share of manufacturing employment. All TSLS regressions use log 1947 planned highway km, log 1898 railroad km and log 1528-1850 exploration routes index as instruments for log kilometers of interstate highways. Robust standard errors in parentheses. ***, **, *: significant at 1%, 5%, 10%.

8.3 Rail

Although railroads carry only a small fraction of all trade, it is interesting to ask if railroads in cities foster rail trade.[23] To assess the effect of railroad tracks on trade, we replicate our main analysis using rail variables. In the first step, we regress bilateral rail trade between cities on a quartic of track distance between them, and importer and exporter fixed effects. We report results in a separate web appendix (Appendix D). Unsurprisingly, rail trade is far less sensitive to distance than road trade. The coefficient on distance for rail trade in weight is -0.57 instead of -1.90 for road

[23] We also investigated possible patterns of complementarities and substitution between road and rail for overall trade. The strong correlation between railroad kilometers and interstate highway kilometers leads unfortunately to unstable results. In addition, we could not find strong enough instruments for the rail-road interaction terms.

Table 9: Second-step results for rail trade

Exporter fixed effect	(1) weight	(2) weight	(3) weight	(4) weight	(5) value	(6) value	(7) value	(8) value
Panel A. OLS estimations.								
log(rail km)	0.46	0.82***	0.87***	0.94***	0.13	0.078	-0.061	-0.016
	(0.31)	(0.27)	(0.28)	(0.27)	(0.27)	(0.31)	(0.31)	(0.34)
Controls.	0	1	2	3	0	1	2	3
Observations	40	34	34	34	39	39	39	39
R^2	0.09	0.25	0.28	0.38	0.00	0.02	0.21	0.22
Panel B. TSLS estimations.								
log(rail km)	0.066	0.72**	0.65**	1.03***	-0.044	-0.39	-0.36	-0.24
	(0.35)	(0.32)	(0.30)	(0.30)	(0.39)	(0.48)	(0.41)	(0.41)
Controls.	0	1	2	3	0	1	2	3
Observations	40	34	34	34	39	39	39	39
First-stage F	43.7	22.9	17.4	17.6	62.9	38.4	35.9	32.5

Notes: The set controls 0 is a constant; 1 adds log 2007 employment and export market access; 2 further adds log population for 1920, 1950 and 2000; 3 also consider log share of manufacturing employment. All TSLS regressions use log 1898 railroad km as instrument for log kilometers of rail tracks. Robust standard errors in parentheses. ***, **, *: significant at 1%, 5%, 10%.

trade.

Panel A of table 9 duplicates our main set of second step OLS results, i.e., table 4, for railroads. Panel B repeats our main TSLS results of table 5 for railroads using 1898 railroads as an instrument for 2004 kilometers of within-city railroad track. The main result from table 9 is the strong effect of railroads on rail trade in weight. In column 3 of panel B, the coefficient on railroad kilometers is 0.65 whereas the corresponding estimate for the effect of interstate highways on road trade is 0.47. Controlling for the manufacturing share of employment in column 4 yields an even larger estimate. Otherwise, the results for rail trade are similar to those for road trade. We see no effect of city rail tracks on the value of rail trade and the difference between OLS and TSLS coefficients is small. We note that these rail results rely on much smaller trade volumes and a smaller number of cities than do the results for road trade, and so should be regarded as somewhat speculative.

9. Roads and employment specialization

Our results so far show that within-city highways cause an increase in the weight of a city's trade, but not its value. A natural corollary of this finding is that highways should have a positive effect on the weight per unit value of exports (tons per dollar), and indeed, regression results verify this relationship (not reported). That within-city highways cause an increase in the weight but not the value of a city's trade also implies that an increase in within-city highways causes cities to become more specialized in the production of heavy sectors according to comparative static 6. We now conduct our second main empirical exercise and examine city level sectoral employment data to check this implication directly.

We use our panel of road and employment data to search for a systematic relationship between the extent of a city's highway network and employment in sectors producing heavy goods. As previously, let N_i^k denote employment in sector k and city i, R_i the kilometers of interstate highway within a city, and C_i a vector of control variables such as log employment, log population in 1920, 1950 and 2000, and market access.

Our estimations proceed in two steps. First, we predict city level employment in each sector as a function of the city's highways. For each sector, we estimate

$$\ln N_i^k = \beta_0^k + \rho_R^{N,k} \ln R_i + \beta^{k'} C_i + \epsilon_i^k \,, \tag{23}$$

where ϵ_i^k is a mean zero error. The coefficient of interest is $\rho_R^{N,k}$, the elasticity of sector k employment with respect to within-city highways. As in our model, V^k is the average tons per dollar of output for sector k. Table 1 enumerates the sectors in the CFS and reports weight per dollar for trade in these sectors. We next use $\hat{\rho}_R^{N,k}$ as the dependent variable in a regression where V^k is the explanatory variable of interest. That is, we estimate

$$\hat{\rho}_R^{N,k} = \gamma_0 + \gamma_1 \ln V^k + v^k. \tag{24}$$

If $\gamma_1 > 0$ then cities specialize in sectors producing heavier goods when they have more within-city highways, and we confirm comparative static 6.

Table 10 reports estimates of equation (24). In column 1 of panel A, the dependent variable is the highway elasticity of sector employment from OLS estimates of equation (23). We find that cities with more roads specialise in the production of heavier goods. The remaining columns of table 10 check the robustness of this finding by varying the specification used to estimate $\rho_R^{N,k}$ in equation (23). In column 2, our dependent variable is a TSLS estimate of $\rho_R^{N,k}$ where we use all three of our instruments. In column 3, we include log distance to the nearest body of water and log average land gradient as control variables in estimates of $\rho_R^{N,k}$. In column 4, we include census region indicators. Column 5 includes the log share of adult population with a college degree. To control for natural resource abundance, column 6 includes the proportion of employment in mining. In columns 7 and 8, we include log income per capita and wholesale employment as controls, respectively.

As we move down the panels of table 10, we vary the control variables included in our estimates of the first-step equation (23). Panel B repeats the estimates of panel A while controlling for manufacturing employment in the second step. In all columns of panels A and B, we find that $\gamma_1 > 0$. This suggests that cities specialize employment in heavy sectors when they have more highways. Long-lived, but omitted variables may be correlated with sectoral specialization and the development of the interstate highway system and precede their construction. If so then the results reported in table 10 may reflect patterns of sectoral specialization *prior* to the construction of the highway system. Panels C and D address this concern. Panel C replicates panel A with a slightly different dependent variable. In particular, it uses estimates of $\rho_R^{N,k}$ based on a variant

Table 10: Main specialization results

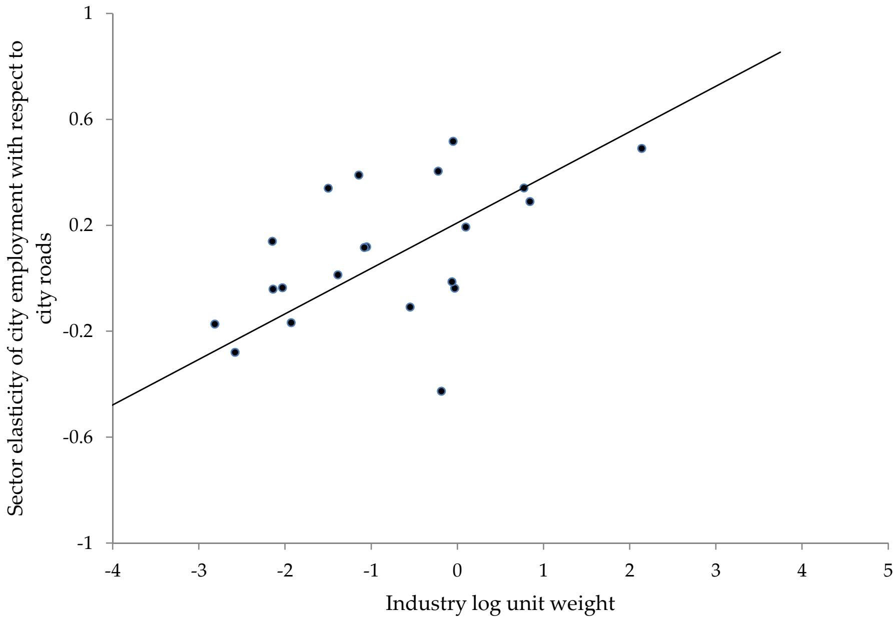
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable: sector specific coefficient on interstate highways								
estimated with:	OLS	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS
using additional controls:	-	-	Water & Slope	Census Div.	% College	Mining	Income pc.	Wholesale
Panel A: Baseline								
log weight per unit value	0.13^{**}	0.16^{***}	0.22^{**}	0.12^{**}	0.16^{***}	0.14^{**}	0.17^{***}	0.16^{***}
	(0.047)	(0.051)	(0.082)	(0.055)	(0.048)	(0.049)	(0.053)	(0.050)
R^2	0.28	0.33	0.26	0.18	0.34	0.29	0.34	0.32
Panel B: With 2007 manufacturing employment								
log weight per unit value	0.16^{***}	0.18^{***}	0.22^{**}	0.13^{**}	0.16^{***}	0.16^{***}	0.18^{***}	0.17^{***}
	(0.038)	(0.044)	(0.079)	(0.050)	(0.045)	(0.043)	(0.048)	(0.043)
R^2	0.46	0.44	0.28	0.24	0.39	0.40	0.41	0.45
Panel C: With 1956 sectoral employment								
log weight per unit value	0.12^{**}	0.16^{***}	0.11^{**}	0.10^{**}	0.18^{***}	0.17^{***}	0.19^{***}	0.16^{***}
	(0.045)	(0.042)	(0.043)	(0.044)	(0.037)	(0.048)	(0.039)	(0.043)
R^2	0.26	0.42	0.26	0.22	0.54	0.39	0.54	0.42
Panel D: With 1956 sectoral and 2007 manufacturing employment								
log weight per unit value	0.14^{***}	0.17^{***}	0.13^{***}	0.12^{**}	0.13^{***}	0.18^{***}	0.17^{***}	0.17^{***}
	(0.041)	(0.041)	(0.039)	(0.047)	(0.033)	(0.043)	(0.040)	(0.041)
R^2	0.37	0.47	0.35	0.23	0.45	0.47	0.49	0.47

Notes: 22 observations per column for each panel. Standard errors in parentheses. The dependent variable is the sector specific coefficient on interstate highways estimated from regression (23) using log kilometers of interstate highways, log 2007 employment, log 1920, 1950 and 2000 population and log market access. Column 1 is estimated with OLS whereas columns 2-7 are estimated with TSLS using log 1947 planned highway km, log 1898 railroad km and log 1528-1850 exploration routes index as instruments for log kilometers of interstate highways. The regression estimating the highways elasticity of employment also includes log distance to water and log slope (column 3), census region dummies (column 4), log share college graduates (column 5), log share mining employment (column 6), log share manufacturing employment (column 7) and log share employment in wholesale (column 8). Robust standard errors in parentheses. ***, **, *: significant at 1%, 5%, 10%.

of equation (23) which also controls for 1956 employment in sector k in city i. Since within-city highways were zero for all cities in 1956, this year marks the beginning of interstate construction, this first step regression is equivalent to a regression in first differences. Panel D also includes the share of manufacturing employment in city i as a control variable in equation (23).[24] Figure 5 plots the data underlying column 2 of Panel D. This figure shows that in sectors producing heavy goods, employment increases with city highways whereas in sectors producing light goods,

[24] We include the initial level of employment in sector k in city i on the right-hand side as opposed to constraining its coefficient to one and the placing it on the left hand side due to evidence that mean reversion in city-sector employment is a major feature of the data as found by Simon (2004) and Duranton (2007). Duranton and Turner (2012) also find that aggregate city employment only adjusts gradually with city highways. This requires controlling for initial levels.

Figure 5: Specialization



Notes: The horizontal axis plots log weight unit per dollar by sector. The vertical axis plots the elasticity of sector employment with respect to city kilometers of interstate highways estimated in column 2 of table 10.

it decreases with city highways.[25] More generally, the results reported in all the panels of table 10 show a complementarity between within-city highways and employment in heavy sectors exactly as comparative static 6 predicts.

The results reported in table 10 are based on the two-step approach described by equations (23) and (24). We can also estimate the complementarity between roads and employment in heavy sectors in a single step. To conduct this estimation we regress log employment in each sector on log highways and log highways interacted with log unit weight. We report results corresponding to panel A of table 10 based on a single step econometric procedure in a separate web appendix (Appendix N). These results are similar to those of table 10.

Our data also allow us to track the way that the development of the US interstate highway system gradually led US cities to specialize. A simple way to do this is to estimate equation (23) in first differences between 1956 and 1970, 1977, 1987, 1997 and 2007. In this regression, the dependent variable is a change in sectoral employment and the explanatory variable of interest is the corresponding change in within-city highway kilometers. We can estimate this equation with OLS. We can also use TSLS to correct for the probable endogeneity of within-city highways.

Table 11 reports results for increasingly long first differences, all of which take 1956 as their initial year. In panel A, we estimate the sensitivity of changes in sectoral employment to changes in within-city highways using OLS in the first step. In panel B we estimate the first step with TSLS using planned 1947 highways, 1898 railroads and old exploration routes as instruments. The results of

[25] Results are robust to the exclusion of the point in the lower left hand corner which is NAICS industry 334 which is "Computer and Electronic Product Manufacturing."

Table 11: Specialization results, dynamics

	(1) 1956-1970	(2) 1956-1977	(3) 1956-1987	(4) 1956-1997	(5) 1956-2007
Panel A. Dependent variable: sector specific coefficient on interstate highways in first-difference, OLS					
log weight per unit value	-0.0075	-0.0082	0.093**	0.15***	0.15***
	(0.038)	(0.026)	(0.042)	(0.036)	(0.045)
R^2	0.00	0.01	0.19	0.46	0.36
Panel B. Dependent variable: sector specific coefficient on interstate highways in first-difference, TSLS					
log weight per unit value	0.074	0.037	0.15***	0.21***	0.20***
	(0.050)	(0.031)	(0.044)	(0.040)	(0.045)
R^2	0.16	0.07	0.37	0.57	0.50

Notes: 22 observations per column for each panel. Standard errors in parentheses. The dependent variable is the sector specific coefficient on interstate highways estimated from a first-differenced version of regression (23) using log kilometers of interstate highways. The regressions in panel A are estimated with OLS whereas those of panel B are estimated with TSLS. Robust standard errors in parentheses. ***, **, *: significant at 1%, 5%, 10%.

table 11 show that the effects of changes in interstate highways on sectoral employment is not correlated (or only weakly correlated) with weight per unit value in 2007 for the earlier periods, 1956-1970 and 1956-1977. Increasing specialisation in heavier sectors in cities with more roads becomes more apparent over 1956-1987 and 1956-1997. The results for 1956-2007 are the same as those for 1956-1997 and suggest no further effects after 1997, many years after the substantive completion of the US interstate highway system. To confirm these results, we further explore the timing of sectoral specialization in relation to roads by replicating panel C of table 10 while controlling for employment in increasingly recent years. In a web appendix (Appendix N, appendix table 15), we report results showing again that most of the changes in employment specialization caused by the within-city highways took place between 1977 and 1997. Overall, these results are suggestive that changes in sectoral employment followed the development of the interstate highway system which mostly took place during the 1960s and 1970s with a lag of 10 to 20 years.

10. Conclusion

This paper examines the causal effect of highway infrastructure on the level and composition of trade in U.S. cities. We find that the weight and value of bilateral trade decrease rapidly with the highway distance between cities, with the weight of trade decreasing more rapidly than its value. A 1% reduction in the travel distance between trading partners increases the value of trade between them by 1.4% and the total weight by 1.9%. These large effects imply that reducing these pairwise distances through expanding the highway system is likely to have large effects on trade and welfare. While a high correlation between Euclidean and highway network distance makes it difficult to untangle econometrically their separate roles, we note that the median highway network distance between a pair of cities is 20% further than the median Euclidean distance, so further improvements may be possible.

We also find that within-city highways affect trade. A 10% increase in a city's stock of highways causes approximately a 5% increase in the weight of exports, but does not cause a measurable change in the value of exports. While we do find an effect of highways on the value of trade at short distances including within the city, the effect on trade in weight is also larger than at longer distances. Consequently, an increase in within-city highways causes a city to specialize in the export of heavy goods.

Using independent data on county-level employment, we find evidence of specialization in response to differences in within-city roads as suggested by our results using trade data. Both time series and cross-section variation in these data suggest that improvements to the within-city road network cause cities to specialize in employing workers in sectors with heavy output. These data further suggest that this specialization takes about 10 or 20 years to complete: the effect of the interstate highway system on employment specialization seems to have been completely realized by the late 1990s, about 20 years after the bulk of construction was completed.

In sum, our results provide an important new insight into the way transportation infrastructure affects trade and the organization of economic activity. Better transportation encourages trade in general, and in particular, encourages the production and trade of things that are hard to move. We also estimate the magnitude of these effects. In addition to their academic interest, these estimates help to inform policy makers as they make infrastructure policy.

To assess the relevant cost-benefit tradeoff involved in the construction of new roads, recall that an average city in our sample has 381 kilometers of interstate highway within its boundaries. Increasing the extent of this network by 1% requires the construction of about 4 kilometers of new highways. This results in about a 0.29% increase in the value of short distance trade and no discernable increase in the total value of trade. Given that trade below 1000 kilometers accounts for about 43 billion dollars per year for an average city, these new 4 km of within-city highways cause about 125 million dollars per year in new trade at short distances and cost approximately 12 million dollars per year (Duranton and Turner, 2012).

These findings suggest at least three directions for further research. First, that the distribution of production appears sensitive to marginal changes in transportation costs while the value of local production is not, suggests that while transportation infrastructure may be an important determinant of where production occurs, it does not have a large effect on the total value of output. If true, this calls into question the wisdom of centralized subsidies for local roads as such subsidies may simply reallocate production and not increase it. However, it suggests that building roads can potentially divert economic activity to depressed areas making regional road construction a rare example of an effective place-based policy.[26] With this said, we do not currently have a basis for assessing the welfare implications of such a policy, and building roads within a city purely to reallocate economic activity seems unlikely to be good policy in general.

[26]Indeed, Duranton and Turner (2012) find some evidence that this is how highway construction is used in the US.

Second, our results indicate that roads are complementary to the production of heavy goods. This suggests that infrastructure policy should allocate roads to cities with a comparative advantage in the production of such goods. Such a policy presumes at least a rudimentary understanding of this sort of comparative advantage and developing this understanding is an important area for further research. Third, if there is a systematic relationship between skills and the weight of production, then we should expect roads to lead to a reallocation of skills across the landscape. An investigation of the magnitude of this effect is an important question for future research.

References

Alchian, Armen A. and William R. Allen. 1964. *University Economics*. Belmont, CA: Wadsworth.

Anderson, James E. and Eric van Wincoop. 2003. Gravity with gravitas: A solution to the border puzzle. *American Economic Review* 93(1):170–192.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare. 2012. New trade models, same old gains? *American Economic Review* 102(1):94–130.

Atack, Jeremy, Michael Haines, and Robert A. Margo. 2011. Railroads and the rise of the factory: Evidence for the United States, 1850-1870. In Paul W. Rhode, Joshua L. Rosenbloom, and David F. Weiman (eds.) *Economic Evolution and Revolutions in Historical Time*. Palo Alto CA: Stanford University Press, 162–179.

Baum-Snow, Nathaniel. 2007. Did highways cause suburbanization? *Quarterly Journal of Economics* 122(2):775–805.

Baum-Snow, Nathaniel, Loren Brandt, J. Vernon Henderson, Matthew A. Turner, and Qinghua Zhang. 2012. Roads, railroads and decentralization of chinese cities. Processed, University of Toronto.

Beaudry, Paul, David A. Green, and Ben Sand. 2014. Spatial equilibrium with unemployment and wage bargaining: Theory and estimation. *Journal of Urban Economics* (forthcoming).

Behar, Alberto and Anthony J. Venables. 2011. Transport costs and international trade. In André de Palma, Robin Lindsey, Emile Quinet, and Roger Vickerman (eds.) *A Handbook Of Transport Economics*. Cheltenham: Edward Elgar, 97–115.

Behrens, Kristian and Pierre M. Picard. 2011. Transportation, freight rates, and economic geography. *Journal of International Economics* 85(2):280–291.

Bleakley, Hoyt and Jeffrey Lin. 2012. Portage and path dependence. *Quarterly Journal of Economics* 127(2):587–644.

Chandra, Amitabh and Eric Thompson. 2000. Does public infractsructure affect economic activity? Evidence from the rural interstate highway system. *Regional Science and Urban Economics* 30(4):457–490.

Clark, Ximena, David Dollar, and Alejandro Micco. 2004. Port efficiency, maritime transport costs, and bilateral trade. *Journal of Development Economics* 75(2):417–450.

Combes, Pierre-Philippe and Miren Lafourcade. 2005. Transport costs: Measures, determinants, and regional policy implications for France. *Journal of Economic Geography* 5(3):319–349.

Couture, Victor, Gilles Duranton, and Matthew A. Turner. 2012. Speed. Processed, University of Toronto.

Cronon, William. 1991. *Nature's Metropolis: Chicago and the Great West*. New York: W. W. Norton & Co.

Disdier, Anne-Célia and Keith Head. 2008. The puzzling persistence of the distance effect on bilateral trade. *Review of Economics and Statistics* 90(1):37–48.

Djankov, Simeon, Caroline Freund, and Cong S. Pham. 2010. Trading on time. *Review of Economics and Statistics* 92(1):193–204.

Donaldson, David. 2014. Railroads of the Raj: Estimating the impact of transportation infrastructure. *American Economic Review* (forthcoming).

Duranton, Gilles. 2007. Urban evolutions: The fast, the slow, and the still. *American Economic Review* 97(1):197–221.

Duranton, Gilles and Diego Puga. 2000. Diversity and specialisation in cities: Why, where and when does it matter? *Urban Studies* 37(3):533–555.

Duranton, Gilles and Matthew A. Turner. 2011. The fundamental law of road congestion: Evidence from US cities. *American Economic Review* 101(6):2616–2652.

Duranton, Gilles and Matthew A. Turner. 2012. Urban growth and transportation. *Review of Economic Studies* 79(4):1407–1440.

Eaton, Jonathan and Samuel Kortum. 2002. Technology, geography, and trade. *Econometrica* 70(5):1741–1780.

Evans, Carolyn and James Harrigan. 2005. Distance, time, and specialization: Lean retailing in general equilibrium. *American Economic Review* 95(1):292–313.

Fernald, John G. 1999. Roads to prosperity? Assessing the link between public capital and productivity. *American Economic Review* 89(3):619–638.

Feyrer, James. 2009. Distance, trade, and income — The 1967 to 1975 closing of the suez canal as a natural experiment. Processed, Darthmouth College.

Fishlow, Albert. 1965. *American Railroads and the Transformation of the Ante-Bellum Economy*. Cambridge (MA): Harvard University Press.

Fogel, Robert. 1964. *Railroads and American Economic Growth: Essays in Econometric History*. Baltimore (MD): Johns Hopkins University Press.

Glaeser, Edward L. and Joshua D. Gottlieb. 2009. The wealth of cities: Agglomeration economies and spatial equilibrium in the United States. *Journal of Economic Literature* 47(4):983–1028.

Glaeser, Edward L. and Janet E. Kohlhase. 2003. Cities, regions and the decline of transport costs. *Papers in Regional Science* 83(1):197–228.

Gray, Charles P. c. 1898. *Gray's New Trunk Railway Map of the United States, Dom. of Canada and Portion of Mexico. Compiled from latest official sources, showing carefully selected list of cities & towns in readable type for quick reference.* New York, NY. Library of congress catalogue # G3701.P3 1898 G7 RR 67.

Harrigan, James. 2010. Airplanes and comparative advantage. *Journal of International Economics* 82(2):181–194.

Hummels, David. 2001. Time as a trade barrier. Processed, Purdue University.

Hummels, David. 2007. Transportation costs and international trade in the second era of globalization. *Journal of Economic Perspectives* 21(3):131–154.

Limão, Nuno and Anthony J. Venables. 2001. Infrastructure, geographical disadvantage, transport costs and trade. *World Bank Economic Review* 15(3):451–479.

Mertz, W. Lee. Undated. Origins of the Interstate. Processed, Federal Highway Administration (United States Department of Transportation).

Mertz, W. Lee and Joyce Ritter. Undated. Building the Interstate. Processed, Federal Highway Administration (United States Department of Transportation).

Michaels, Guy. 2008. The effect of trade on the demand for skill - Evidence from the Interstate Highway System. *Review of Economics and Statistics* 90(4):683–701.

Redding, Stephen and Anthony J. Venables. 2004. Economic geography and international inequality. *Journal of International Economics* 62(1):63–82.

Simon, Curtis J. 2004. Industrial reallocation across US cities, 1977-1997. *Journal of Urban Economics* 56(1):119–143.

Stock, James H. and Motohiro Yogo. 2005. Testing for weak instruments in linear IV regression. In Donald W.K. Andrews and James H. Stock (eds.) *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Cambridge: Cambridge University Press, 80–108.

United States Bureau of Statistics. 1899. *Statistical abstract of the United States: Twenty-First Number 1898*. Washington, DC: US Government printing office.

United States Bureau of the Census. 2011. *Statistical abstract of the United States 2011*. Washington, DC: US Government printing office.

United States Bureau of Transportation Statistics. 2010. *Transportation. 2007 Commodity Flow Survey*. Washington, DC: US Government printing office.

United States Department of Transportation, Federal Highway Administration. 2005. HPMS *Field Manual*. Washington, DC.

United States Federal Works Agency, Public Roads Administration. 1947. *Press release*.

United States Geological Survey. 1970. *The National Atlas of the United States of America*. Washington, DC.

United States House of Representatives. 1947. *A Report on the National Interregional Highway Committee, Outlining and Recommending a National System of Interregional Highways*. Washington, DC: US Government printing office.

Appendix A. Data

Definition of the CFS *regions:* In the CFS, there are 174 areas within the United States. Dropping 50 aggregate states, the District of Columbia, 49 "remainder" regions and Honolulu, leaves 73 areas. Restricting attention to CFS regions with non-zero realizations of each of the instrumental variables leaves us with a sample of 66 CFS areas. See United States Bureau of Transportation Statistics (2010) for details about the CFS.

Road infrastructure: We rely on three distinct data sources to describe the road network. To calculate the number of kilometers of interstate highway in each CFS region we use the US Highway Performance and Monitoring System (HPMS) "Universe" data for 1987, 1997 and 2007. These data are described in United States Department of Transportation, Federal Highway Administration (2005) and Duranton and Turner (2011). HPMS data are not available prior to 1983. To calculate the number of kilometers of interstate highway in each CFS region in earlier years, we rely on 'PR511' data obtained from Nathaniel Baum-Snow. These are administrative data collected in the course of constructing the interstate highway system. They record the length of interstate highway sections in each county annually from 1957 to 1993. We make use of the years 1957, 1967, 1977 and 1987. Details about these data are available in Baum-Snow (2007).
The HPMS and PR511 are tabular data sets and so do not provide information about network configuration. To calculate network distance between CFS centers and our radial road index, we rely on the 2005 National Highway Performance Network (NHPN) map of the interstate highway system. This map contains the necessary network information to allow the calculation of pairwise network distances and also allows us to calculate our radial road index. We note that this map is not available for 2007. This map is available for download from the Federal Highway Administration at `http://www.fhwa.dot.gov/planning/processes/tools/nhpn/`.

To calculate interstate highway distances between cities we first identify the centroid of each CFS region. To do this, we find the area weighted centroid of the urbanized portions of each county comprising the CFS. We then calculate the centroid of the CFS city as the employment weighted average of the centroids of the component counties. In this calculation we rely on census shapefiles describing county and urbanized area boundaries in 2000. Next, we calculate the pairwise distance between CFS cities as the interstate highway network distance between centroids, relying on the map of the interstate highway system contained in the NHPN map of the US road network for 2005. We note that the algorithm we use to find pairwise distances includes the minimum distance from a centroid to the interstate highway network as part of the bilateral travel distance. With very few exceptions the centroids of CFS cities are within a few kilometers of an interstate highway.

An analogous procedure allows us to find distances between cities along the 1947 highway plan, along 1898 railroad routes and along both of our maps of modern railroad networks.

Contemporary Rail infrastructure: We rely on two maps to describe the modern railroad network. The first is the 2004 railroad map from the North American Atlas. This is a joint publication of the US, Canadian and Mexican Geological services, and describes rail links between major cities or rail lines servicing resource extraction. This map is available for download at: `http://nationalatlas.gov/atlasftp.html`. Our second map of railroad networks is from the

US National Atlas and provides a much more detailed description of US railroads. In particular, it does not restrict attention to lines linking major cities but includes many smaller lines and even commuter rail lines. This second map results from a compilation of several railroads maps by the USGS, all of them circa 2004. These USGS railroad maps are also available for download at: http://nationalatlas.gov/atlasftp.html. These two maps form the basis for our calculation of the kilometers of railroad track within each CFS region, for our calculation of the rail network distance between CFS centroids, and for our index of radial rail capacity.

Employment data: We use the County Business Patterns (CBP) to construct a panel of employment data by county and industry for the years, 1956, 1970, 1977, 1987, 1997 and 2007. To reduce measurement error, our '1970' CBP data is based on the county level mean of 1970 and 1971 CBP, our '1977' CBP data is based on the county level mean of 1976 and 1977, our '1987' data is based on the mean of 1987 and 1988, our '1997' data on the mean of 1996 and 1997, and our '2007' data on the mean of 2006 and 2007. Our 1956 data is based only on 1956 CBP

Organizing these employment data into a panel consistent with our data for roads presents several challenges. First, county business patterns are organized by contemporaneous counties. Thus, aggregating these data to constant boundary 2007 CFS regions requires that we track all of changes in county boundaries.

Second, our CFS data is organized on the basis of 22 CFS industrial classification codes. County business patterns, however, is organized by different versions of SIC and NAICS classifications in different years. We reconcile these various classification systems into classifications consistent with the 2007 CFS. Our choice of pairs of CBP years, e.g., '1970' is 1970 and 1971 but '1977' is 1976 and 1977, is driven by efforts to keep industrial classification systems constant within pairs.

The various years of CBP data area available in different formats and from different sources. Data are available in electronic form from the census for the years 1986-2010 from the US census at http://www.census.gov/econ/cbp/download/index.htm. Data are available in electronic form from the University of Virginia library for the years 1977-2001 at http://fisher.lib.virginia.edu/collections/stats/cbp/, although only at the level of two digit classification. Data are available in electronic form from the National Archives as far back as 1970 and can be accessed through this link http://www.census.gov/econ/cbp/historical.htm. CBP data does not appear to be available in electronic form prior to 1970. Our 1956 CBP data is digitized from 'microprint' images of the original hardcopy, available from the University of Toronto library. Because of the difficulty of this process, our 1956 cross-section of CBP data is based on just a single year's survey.

Since CFS cities are defined as aggregations of counties, these data allow us to construct measures of aggregate employment for 1956, 1970, 1977, 1987, 1997 and 2007. To our knowledge, this is the first time that such a long panel of county-sector level employment data has been constructed.

Population and demographics: We use decennial censuses from 1920, 1950 and 2000. 1920 is earliest year for which we can construct population numbers for the whole country on the basis of the 1999 county boundaries that form the basis for CFS regions. To measure CFS socio-demographics we use aggregated county level data from the 2000 US census. These data record a wide variety of characteristics such as educational attainment, income, etc.

Appendix B. Theory

Derivation of the importer effect

We can use equations (7), (10) and (11) to write

$$\delta_j^M = (\sigma - 1)(\log \mathbb{P}_j - \log \tau_j) + \log(N_j\, W_j)\,. \tag{B1}$$

From equation (B1), a city's propensity to import depends on its price index $\mathbb{P}_j$, the cost of entry, τ_j and its total income $N_j\, W_j$. From equations (7), (10) and (11), we can also write

$$e^{\delta_i^X + (1-\sigma)\log \tau_{ij}} = \int_0^1 \left(\frac{\tau_i^k\, \tau_{ij}\, W_i}{A_i} \right)^{1-\sigma} \mathrm{d}k\,. \tag{B2}$$

Summing both sides of (B2) across all cities i; using equations (5), (6) and (10) again; and taking logs yields

$$MA_j^M \equiv \log \sum_{i=1}^{I} e^{\delta_i^X + (1-\sigma)\log \tau_{ij}}\, \mathrm{d}i \;= (1-\sigma)(\log \mathbb{P}_j - \log \tau_j). \tag{B3}$$

Following Redding and Venables (2004), the term on the left-hand-side can be naturally interpreted as an import market access. Next, we can insert (B3) into (B1) to obtain

$$\delta_j^M = \log(N_j\, W_j) - MA_j^M\,. \tag{B4}$$

This expression is interesting because it shows that city j's propensity to import, δ_j^M, depends only on its income, $N_j\, W_j$, and its import market access, MA_j^M, which in turn depends on the distance to other cities and the propensity to export of the latter. Roads in city j do not appear directly in this expression. As a result, regardless of how much of a shipment disappears when entering city j, the value of this shipment when it leaves exporting city i is not directly affected. There is nonetheless an indirect effect. Roads in j will affect imports through their effect on the income of that city and through its propensity to export. To see this, we can use equation (B5) into (B4) to eliminate city j's wage and obtain equation (15) in the main text.

Proof of comparative statics results

CS1. We can insert equation (6) into (4) and use the fact that the cost of distance τ_{ij} between any two cities i and j is the same for all sectors by equation (10) to show that the ratio of exports between any two sectors $X_{ij}^{k'}/X_{ij}^{k}$ does not depend on τ_{ij}. Hence the composition of trade between i and j is not affected by τ_{ij}.

CS2. This follows directly from equation (12).

CS3. The fact that exports in value should increase with within-city roads conditionally on market potential is again obvious from equation (12). This results also holds unconditionally. To see this, we can set the numéraire to be W_i and focus on equation (7). By equation (10) an increase in τ_{ij} has direct negative effect on X_{ij}. It can only lead to an increase in X_{ij} if the general equilibrium term $\mathbb{P}_j^{\sigma-1} N_j W_j$ increases so much as to more than offset the direct effect of the increase in prices through τ_{ij}. This would then imply an increase in the imports of city j from all other cities. As shown by the equation (B5) below, this increase in imports in j should increase the wages of all other cities. In turn, this increase in wages should increase the demand for the goods produced by city i. Such an increase would violate the labour market clearing condition in this city. Hence, the direct negative effect of an increase in τ_{ij} on X_{ij} must dominate the indirect general equilibrium effects.

CS4. Given that trade in value is balanced, this result follows directly from CS3.

CS5. This result follows directly from equation (19).

CS6. First, let $\rho_R^X = \partial \log X_i / \partial \log R_i$ denote the elasticity of the value of exports with respect to city roads, $\rho_R^W = \partial \log W_i / \partial \log R_i$ the elasticity of wages with respect to city roads, and $\rho_R^{Q,k} = \partial \log Q_i^k / \partial \log R_i$ the elasticity of output in sector k with respect to city roads.

The aggregate value of goods from city i is $X_i = W_i N_i$. Constant population then implies $\rho_R^X = \rho_R^W$. Inserting equation (10) into (9) and rearranging yields

$$W_i = \left[\frac{A_i^{\sigma-1}}{N_i} \left[\int_0^1 \left(\tau_i^k \right)^{1-\sigma} \, \mathrm{d}k \right] \left[\sum_{j=1}^{I} \frac{\mathbb{P}_j^{\sigma-1}}{(\tau_{ij}\, \tau_j)^{\sigma-1}} W_j N_j \right] \right]^{1/\sigma} . \tag{B5}$$

Using this equation to compute ρ_R^W we have

$$\rho_R^X = \rho_R^W = \frac{1-\sigma}{\sigma} \frac{\int_0^1 \rho_R^{\tau,k} \, (\tau_i^k)^{1-\sigma} \mathrm{d}k}{\int_0^1 (\tau_i^k)^{1-\sigma} \mathrm{d}k} . \tag{B6}$$

Together with our assumptions that $\sigma > 1$ and $\rho_R^{\tau,k} < 0$ for all k, equation (B6) implies that the aggregate value of a city's exports increase with its roads. Thus, $\rho_R^X > 0$.

Next, we insert equation (1) into (8) to obtain an expression for output in sector k and city i, Q_i^k. Differentiating this expression with respect to R_i, using (10), and rearranging leads to

$$\rho_R^{Q,k} = (1-\sigma)\, \rho_R^{\tau,k} - \sigma\, \rho_R^W . \tag{B7}$$

Because $\rho_R^W > 0$ and $\rho_R^{\tau,k} < 0$, the sign of $\rho_R^{Q,k}$ is ambiguous. However, total output in city i is given by $Q_i = \int_0^1 Q_i^k \mathrm{d}k = A_i N_i$ and is therefore constant. This implies that not all sectors can contract or expand with city roads. Since k ranks sectors by both weight and by $\rho_R^{\tau,k}$, it follows that output contracts in sectors producing light goods and expands in sectors producing heavy goods.

Finally, because sectors producing heavy goods expand and sectors producing light goods contract while total quantities produced remains constant, it follows that the weight of goods from a city increases as roads in that city increase. □

Worker mobility

We now explore the effect of worker mobility on our estimation strategy. First, from (4) individual consumption in city i of the variety from sector k and city i is

$$q_{ji}^k = \frac{\mathbb{P}_i^{\sigma-1}}{(P_{ji}^k)^\sigma} W_i . \tag{B8}$$

Inserting this expression into the utility function (2) and making use of equation (5) implies the following indirect utility: $U_i = W_i/\mathbb{P}_i$. Then, worker mobility implies that in equilibrium utility in city i must be equal to the common utility achieved in the other cities. Taking logs, this implies $\log U_i = \log W_i - \log \mathbb{P}_i = \underline{u}$. Combining this expression with (B3) leads to

$$\log(W_i) = \underline{u} + \log \tau_i - \frac{MA_i^M}{\sigma - 1}. \tag{B9}$$

After using equation (B9) into the first part of equation (12) we write the propensity to export value under perfect labor mobility as

$$\delta_i^X = \underline{u} + \log \tau_i + \log N_i - MA_i^X - \frac{MA_i^M}{\sigma - 1}, \tag{B10}$$

where MA_i^M is an import market access term defined in Appendix B and $\underline{u}$ is the log of the reservation level of utility for mobile workers. A comparison with equation (12) shows immediately that the exporter effect under full labor mobility does not contain any productivity term but adds an import market access term. The coefficient on population and export market access also changes slightly. In addition, the transportation term directly measures the effect of city roads on the cost of exiting a city. In the same spirit, we can also express cities' propensity to export weight under perfect labor mobility as

$$\delta_i^T = \frac{\sigma}{\sigma-1}\underline{u} + S_2^T(R_i) + \frac{\sigma}{\sigma-1}\log N_i - \frac{\sigma}{\sigma-1}MA_i^X - \frac{\sigma}{(\sigma-1)^2}MA_i^M, \tag{B11}$$

where $S_2^T(R_i) = S^T(R_i) + \frac{\sigma}{\sigma-1}(\log \tau_i - S(R_i))$. It is easy to see that labor mobility has very similar effects on the propensity to import value and weight.

Appendix C. Other first stage and second-step results

Table 12: First-stage results, OLS for log highway km

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log 1947 highway km	0.67***	0.55***	0.63***	0.58***		0.64***		
	(0.084)	(0.074)	(0.070)	(0.059)		(0.057)		
log 1898 railroad km	0.17***	0.050	0.019		0.35***		0.39***	
	(0.062)	(0.045)	(0.049)		(0.053)		(0.061)	
log 1528 explorations	0.12***	0.100***		0.095***	0.20***			0.23***
	(0.034)	(0.019)		(0.019)	(0.033)			(0.047)
log employment		0.19	0.23	0.21	0.078	0.23	0.14	0.24
		(0.18)	(0.19)	(0.19)	(0.25)	(0.19)	(0.29)	(0.37)
Controls	N	Y	Y	Y	Y	Y	Y	Y
R^2	0.86	0.94	0.92	0.94	0.84	0.92	0.77	0.76

Notes: 66 observations per column. The dependent variable is log 2005 highway kilometers. All regressions include a constant. The controls are log 2007 employment, export market access, log population for 1920, 1950, 2000 and log share of 2003 manufacturing employment. Robust standard errors in parentheses. ***, **, *: significant at 1%, 5%, 10%.

Table 13: Second-step results for importer fixed effects

Importer fixed effect	(1) weight	(2) weight	(3) weight	(4) weight	(5) value	(6) value	(7) value	(8) value
Panel A. OLS estimations.								
log highway km	0.92***	0.12	0.13	0.16	0.90***	0.041	0.032	0.067
Robust s.e.	(0.25)	(0.23)	(0.16)	(0.15)	(0.23)	(0.21)	(0.15)	(0.14)
Non-robust s.e.	(0.15)	(0.17)	(0.17)	(0.17)	(0.14)	(0.16)	(0.16)	(0.17)
Corrected s.e.	(0.15)	(0.22)	(0.17)	(0.17)	(0.14)	(0.20)	(0.16)	(0.16)
log employment		0.70***	-0.50	-0.42		0.74***	-0.23	-0.16
		(0.11)	(0.56)	(0.49)		(0.099)	(0.54)	(0.48)
Market access (export)		-0.30	-0.49	-0.58		-0.52	-0.66*	-0.74*
		(0.37)	(0.39)	(0.45)		(0.36)	(0.38)	(0.44)
Market access (import)		-0.52	-0.32	-0.24		-0.18	-0.055	0.025
		(0.34)	(0.35)	(0.39)		(0.34)	(0.37)	(0.41)
log 1920 population			0.061	0.027			-0.020	-0.053
			(0.30)	(0.30)			(0.29)	(0.30)
log 1950 population			-0.22	-0.10			-0.046	0.070
			(0.47)	(0.48)			(0.40)	(0.42)
log 2000 population			1.43**	1.25**			1.10*	0.92*
			(0.67)	(0.56)			(0.61)	(0.53)
log % manuf. emp.				-0.23				-0.22
				(0.41)				(0.38)
R^2	0.37	0.72	0.75	0.75	0.39	0.72	0.75	0.75
Panel B. TSLS estimations.								
log highway km	0.87***	0.16	0.16	0.18	0.82***	0.035	0.029	0.051
	(0.23)	(0.15)	(0.15)	(0.15)	(0.21)	(0.15)	(0.15)	(0.14)
log employment		0.68***	-0.51	-0.43		0.74***	-0.23	-0.15
		(0.11)	(0.53)	(0.46)		(0.10)	(0.51)	(0.45)
Market access (export)		-0.29	-0.49	-0.58		-0.52	-0.66*	-0.75*
		(0.36)	(0.36)	(0.42)		(0.35)	(0.36)	(0.41)
Market access (import)		-0.53	-0.32	-0.24		-0.18	-0.055	0.024
		(0.32)	(0.33)	(0.37)		(0.33)	(0.34)	(0.38)
log 1920 population			0.061	0.027			-0.020	-0.053
			(0.27)	(0.27)			(0.27)	(0.28)
log 1950 population			-0.23	-0.10			-0.046	0.072
			(0.44)	(0.44)			(0.38)	(0.39)
log 2000 population			1.43**	1.24**			1.10*	0.93*
			(0.63)	(0.53)			(0.57)	(0.49)
log % manuf. emp.				-0.23				-0.22
				(0.38)				(0.35)
Overid. p-value	0.86	0.23	0.14	0.22	0.90	0.48	0.27	0.31
First-stage Stat.	97.5	89.0	80.1	81.6	97.5	89.0	80.1	81.6

Notes: 66 observations per column. All regressions include a constant. All TSLS regressions use log 1947 planned highway km, log 1898 railroad km and log 1528-1850 exploration routes index as instruments for log kilometers of interstate highways. Robust standard errors in parentheses. ***, **, *: significant at 1%, 5%, 10%.