Roads and Trade: Evidence from the US[§]

Gilles Duranton[*]
University of Toronto
Peter M. Morrow[‡]
University of Toronto
Matthew A. Turner[‡]
University of Toronto

This draft: 17 July 2011

ABSTRACT: We estimate the effects 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. There is no discernible 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, Aloysius Siow, Dan Trefler, and seminar and conference participants for comments. 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.

[*]Department of Economics, University of Toronto, 150 Saint George Street, Toronto, Ontario M5S 3G7, Canada (e-mail: gilles.duranton@utoronto.ca; website: http://individual.utoronto.ca/gilles/default.html). Also affiliated with the Center for Economic Policy Research, and the Center for Economic Performance 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).

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, and city characteristics. Identifying the causal effect of highways on trade is difficult 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 and has no measurable 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 changes to highways between cities cause large increases in the weight and value of trade as they reduce travel distances between cities.

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. While the literature on trade investigates the effect of trade barriers, the role of transportation costs and infrastructure receives less attention. Limão and Venables (2001) find a strong positive association between an index of infrastructure quality and trade. 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).[1] However, there are very few papers which can claim to identify the causal effects of infrastructure on trade.[2] Michaels (2008) examines the effect of access to the interstate highway network on rural counties in the US. Donaldson (2010) examines the historical effects of railroads in India on price differentials across regions, trade flows, and incomes. To our knowledge, our paper is the first to examine the effect of transportation infrastructure on urban administrative units in a developed country, as well as the first to examine the role of infrastructure in determining patterns of specialization.

The construction and maintenance of transportation infrastructure is among our most costly policy endeavors. In spite of this, our understanding of the effects of these investments on the evolution of cities and countries is based only on a small emerging literature. 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

[1]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).

[2]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.

Duranton and Turner (2010) examines how city populations depend on the interstate highway system. Donaldson (2010) examines the effect of railroads on India's development. 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.

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 and motivates our econometric specification; high quality data describing highways, trade flows, and relevant covariates; a strategy for resolving the probable 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 (1) examine the effect of within and between city highways on the weight and value of inter-city trade, and (2) assess the effect of highways on the composition of production and trade for a given city. Our theoretical model also leads to a two-step estimation strategy resembling Redding and Venables (2004). In the first step we estimate gravity models 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 and weight conditional on distance and trading partner characteristics. In the second step, we explain these fixed effects as functions of city highways and other city characteristics. In a distinct empirical exercise, we examine the sectoral composition of exporter employment as a function of city highways.

Our two primary data sources are the US Commodity Flow Survey (CFS) and the Highway Performance Monitoring System (HPMS). 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 lane 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. 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. Hence, changes to a city's highways should affect comparative advantage in heavy goods. Our data confirm this qualitative conclusion and suggests 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 5% increase in tons exported, and that a change in highways does not cause a measurable change in the value of a city's exports. These results require that an increase in within city highways leads a city to become more specialized in the export of goods with low values per unit weight. 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.

2. Model

We estimate the effect of transportation costs and infrastructure on trade flows using data that describe bilateral trade flows between cities. The first problem we face is that trade flows between any two cities i and j will reflect the cost of transportation that i and j face when they trade with other cities, as well as the cost of transportation between i and j. This is precisely the problem that Anderson and van Wincoop (2003) (and much of the related literature) address, and we generalize their model to suit our purpose. Specifically, we first elaborate the description of transportation to allow an investigation of the effects on trade flows of roads both within and between cities. Second, we consider a multi-sector framework to allow an investigation of comparative advantage based on city level road abundance and across sector variation in dependence on roads. Third, we explore the implications of labor mobility, an important issue for data describing a system of cities rather than a system of countries.

Preferences, technology, and geography

Most cities are small relative to the aggregate economy. To represent this fact, we consider a continuum of cities on the interval [0,1]. Conveniently, this implies that each city can be treated as a small open economy. City i hosts N_i identical workers/residents and labor is mobile across sectors. City employment in most sectors is also small relative to total city employment. Again, we consider a continuum of sectors on the interval [0,1]. For now, we take labor to be immobile across cities, although we relax this assumption below.

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

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

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. Equation (1) allows for cities to have different levels of aggregate productivity but abstracts from Ricardian comparative advantage in the production technology. Although we do not examine other forces for comparative advantage in our theoretical section, the empirical section explicitly examines other forces for comparative advantage such as the city's endowment of human capital. Consistent with constant returns to scale and free entry, we assume that producers in all cities and sectors are price takers.

Consumers in all cities have identical preferences. A consumer in city j maximizes the utility function,

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

by choosing consumption q_{ij}^k of the varieties from all cities i in all sectors k with $\sigma > 1$. The consumer's budget constraint is

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

where W_j is the wage of a worker in city j and P_{ij}^k is the price of the variety from city i in sector k when consumed in city j.

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. Not only is this a good approximation to our aggregated data, but it also eliminates the complications associated with corner solutions. It is possible, but cumbersome and uninformative, to extend our model to more realistic substitution patterns.

In equilibrium all 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.

Maximizing utility (2) subject to the budget constraint (3) and aggregating across consumers implies that expenditure in city j for the variety produced by city i in sector k is

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

where N_j is population and

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

is the price index for city j. X_{ij}^k must also equal the value of shipments from city i to j in sector k.

Trade between cities is costly. To describe these 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. This standard iceberg assumption implies that the cost of 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 across sectors yields the value of aggregate shipments from i to j,

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

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

$$N_i^k = \int_0^1 \frac{\tau_{ij}^k\, Q_{ij}^k}{A_i}\,\mathrm{d}j = \int_0^1 \frac{\tau_{ij}^k\, X_{ij}^k}{A_i\, P_{ij}^k}\,\mathrm{d}j = \frac{A_i^{\sigma-1}}{W_i^{\sigma}} \int_0^1 \frac{\mathbb{P}_j^{\sigma-1}}{(\tau_{ij}^k)^{\sigma-1}}\, W_j\, N_j\,\mathrm{d}j\,. \tag{8}$$

Finally, the equilibrium for city i is such that its labor market clears. Aggregate city labor supply, N_i equals the sum of labor demand in all sectors. Using equation (8), we obtain

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

Given an equilibrium in all cities but i, it would now be easy to 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. This wage can then be used in equations (7) and (8) to compute the value of exports from i to j and city i employment in any sector k.[3]

Transportation

To ease exposition, we have so far relied on a description of transportation costs which is too simple to allow an investigation of the effect of roads on trade. We now allow transportation costs to vary with sector, roads between exporter and importer cities, and roads at both locations,

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

In this equation, R_i measures roads in exporter city i. R_{ij} measures roads between exporter city i and importer city j. R_j measures roads in importer city j. 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; the cost of going from exporter city i to importer city j; and the cost of entering importer city j.[4,5] 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 leaving city i, $\tau_x^k(R_i)$, depends on the roads within this city: $\partial\tau_x^k(R_i)/\partial \log R_i < 0$. This cost also differs across sectors in the following fashion. One unit of labor in city i produces A_i units of good k with weight V^k per unit. Without loss of generality, we can rank sectors by V^k where sector 0 produces the lightest goods and sector 1 the heaviest: $V^0 \leq ... \leq V^k \leq ... \leq V^1$. We assume that sectors producing heavier goods are more sensitive to roads. Hence, the elasticity of the transportation 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. This implies that for any level

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

[4]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.

[5]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.

of within city roads 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 cost of transportation between cities, $\tau_{xm}(R_{ij})$, depends on the roads between cities with $\partial\tau_{xm}(R_{ij})/\partial \log R_{ij} < 0$. Unlike the cost of leaving a city, we restrict the cost of traveling between cities to be the same for all sectors. While relaxing this condition is clearly of interest, our data does not provide pairwise trade flows by sector. Thus, such a generalization would add complexity without informing our empirical work.

The cost of entering city j, $\tau_m(R_j)$, depends on its roads with $\partial\tau_m(R_j)/\partial \log R_j < 0$. Again, we assume no heterogeneity across sectors in entry costs. 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. With this said, this is a restrictive assumption and we examine its validity in our empirical work.

Implications of the model for trade flows

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. Our model now allows us to make a number of predictions, summarized in the following proposition.

Proposition 1 *An increase in roads in a city causes an increase in the weight and value of its exports. It also causes a decrease in employment by sectors producing light goods and an increase in employment by sectors producing heavy goods.*

Proof To derive the first part of the result, note that because no city is of positive mass, the aggregate value of exported 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,

$$W_i = \left[\frac{A_i^{\sigma-1}}{N_i} \left[\int_0^1 \left(\tau_i^k\right)^{1-\sigma} \, \mathrm{d}k \right] \left[\int_0^1 \frac{\mathbb{P}_j^{\sigma-1}}{(\tau_{ij}\,\tau_j)^{\sigma-1}} \, W_j \, N_j \, \mathrm{d}j \right] \right]^{1/\sigma} . \tag{11}$$

Using this equation to compute ρ_R^W we have,

$$\rho_R^X = \rho_R^W = \frac{1-\sigma}{\sigma} \frac{\int_0^1 \rho_R^{\tau,k} \, (\tau_i^k)^{1-\sigma} \mathrm{d}k}{\int_0^1 (\tau_i^k)^{1-\sigma} \mathrm{d}k} . \tag{12}$$

Together with our assumptions that $\sigma > 1$ and $\rho_R^{\tau,k} < 0$ for all k, equation (12) implies that the aggregate value of a city's exports increase with its roads. Thus, $\rho_R^X > 0$.

Next, we insert (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,

$$\rho_R^{Q,k} = (1-\sigma)\, \rho_R^{\tau,k} - \sigma\, \rho_R^W . \tag{13}$$

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 output, and hence exports to other cities, also become heavier as city roads increase. □

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 higher wages and 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.

The results derived in proposition 1 are interesting beyond these qualitative predictions since they also suggest magnitudes for these effects. According to expression (12), the elasticity of the value of exports with respect to city roads should be the same as the corresponding elasticity for wages. The recent literature suggests only modest effects of roads on wages (e.g., Fernald, 1999, Duranton and Turner, 2010). Thus, we expect roads to have only small effects on the value of exports.

On the other hand, the effect of city roads on the composition of trade might be much larger. As made clear by equation (13), the elasticity of sector output with respect to roads depends on the difference between how sensitive to city roads are transportation costs for this sector and the roads elasticity of wages. This difference is magnified by the elasticity of substitution between goods, σ, for which most estimates in the literature are above 5 (e.g., Anderson and van Wincoop, 2004, for a review). This suggests that city roads can affect the specialization of cities in a dramatic way and, in turn, possibly have a large effect on the weight of trade flows.

Implications of the model for estimation

To inform the specification of regressions predicting the value of inter-city trade flows, we insert (10) into (7) and take logs to get,

$$\ln X_{ij} = \delta_i^X + (1-\sigma)\ln \tau_{ij} + \delta_j^M \,. \tag{14}$$

In this equation δ_i^X and δ_j^M are expressions – precisely defined below – involving the economic fundamentals of importer and exporter cities. These parameters play a central role in our analysis. 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 will ultimately investigate the effect of roads on the weight of trade, and in this investigation use analogous language to describe cities' propensities to trade weight.

Note that in equation (14), all else equal, a one unit increase in the propensity to import 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 reflect changes in the log of exports, and a change in δ_i^X against a log change in another variable is an elasticity of export value. We note this intuition here, since we eventually exploit it to interpret our regression results. Analogous intuition applies to importer and exporter fixed effects.

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

$$\delta_i^X = S(R_i) + \frac{\sigma - 1}{\sigma} \ln A_i + \frac{\sigma - 1}{\sigma} \ln N_i - \frac{\sigma - 1}{\sigma} MA_i^X \,, \tag{15}$$

where

$$S(R_i) \equiv \frac{1}{\sigma} \ln \int_0^1 \left(\tau_i^k\right)^{1-\sigma} \, \mathrm{d}k \tag{16}$$

depends on the cost of exporting from city i and, following proposition 1, increases with roads in this city. MA_i^X is an export market access term, a form of market potential. For city i it is given by,

$$\begin{aligned} MA_i^X \;&\equiv\; \ln \int_0^1 \frac{\mathbb{P}_j^{\sigma-1}}{(\tau_{ij}\,\tau_j)^{\sigma-1}} W_j\, N_j \, \mathrm{d}j \\ &=\; \ln \int_0^1 e^{(1-\sigma)\,\ln \tau_{ij} + \delta_j^M} \, \mathrm{d}j \,, \end{aligned} \tag{17}$$

where the second line results from substituting (7), (10), and (14) into the first. Because pairwise transportation costs $(1 - \sigma) \ln \tau_{ij}$ and propensity to import δ_j^M appear in equation (14), we can use the output of this equation to calculate MA_i^X.

It is interesting to note that we expect the export market access of a city, MA_i^X, which weights the economic size of other cities by a measure of their price index and an inverse measure of distance, to affect that city's propensity to export, δ_i^X, *negatively*. The reason for this apparently counter-intuitive prediction is that equation (14), which delivers δ_i^X, already accounts for distance to importers and their propensity to import. Equation (15) then only captures a negative indirect wage effect. To see this, recall from equation (14) that the propensity to export is negatively affected by wages. At the same time, wages and market access are positively related, as per equations (7) and (17). We thus obtain a negative effect of market potential on propensity to export when substituting market potential for wages, as we do in equation (15).

Turning to the propensity to import, we can use equations (10) and (7) to write,

$$\delta_j^M = (\sigma - 1)(\ln \mathbb{P}_j - \ln \tau_j) + \ln(W_j N_j) \,. \tag{18}$$

From equation (18), a city's propensity to import depends on its price index $\mathbb{P}_j$, the cost of entry, τ_j, and its total income $W_j N_j$. Appendix A provides an alternative expression using only variables exogenous to the importing city. 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. On the basis of the derivation in Appendix A, we conclude that the effect of city roads on imports should be small, since cities with more roads import more only to the extent that more roads makes them richer through exports. As mentioned before, the literature tends to find small effects of roads on wages for the US.

Turning to the weight of trade, we note first that 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

$$T_{ij} = \frac{A_i}{W_i} \int_0^1 X_{ij}^k V^k \, \mathrm{d}k \, . \tag{19}$$

Using (4), (6), (10), and (11), it is easy to show that the weight of trade can be expressed as $\ln T_{ij} = \delta_i^T + (1-\sigma) \ln \tau_{ij} + \delta_j^M$, where city i's propensity to export weight is,

$$\delta_i^T = S^T(R_i) + \ln A_i + \ln N_i - MA_i^X \, , \tag{20}$$

where

$$S^T(R_i) \equiv \ln \int_0^1 \left(\tau_i^k\right)^{1-\sigma} V^k \, \mathrm{d}k - \ln \int_0^1 \left(\tau_i^k\right)^{1-\sigma} \, \mathrm{d}k \, . \tag{21}$$

Following proposition 1, $S^T(R_i)$ increases with roads in city i.[6]

So far our model treats city population as exogenous. This reflects the fact that the city population levels experience small changes in response to changes in their stock of roads (Duranton and Turner, 2010). We nevertheless consider the implications of labor mobility across cities for our approach in Appendix B. Under perfect labor mobility the propensity to export value becomes,

$$\delta_i^X = \underline{u} + S_2(R_i) + \ln N_i - MA_i^X - \frac{MA_i^M}{\sigma - 1} \, , \tag{22}$$

where $S_2(R_i) = S(R_i) + \ln \tau_i$, MA_i^M is an import market access term defined in Appendix A, and $\underline{u}$ is the log of the reservation level of utility for mobile workers. A comparison with (15) shows immediately that the exporter effect under full labor mobility does not contain productivity but adds an import market access term. The coefficient on population and export market access also changes slightly. In addition, the transportation term now sums the cost of exit for exports and the cost of entry for imports. In the same spirit, we can also express cities' propensity to export weight under perfect labor mobility as

$$\delta_i^T = \frac{\sigma}{\sigma - 1} \underline{u} + S_2^T(R_i) + \frac{\sigma}{\sigma - 1} \ln N_i - \frac{\sigma}{\sigma - 1} MA_i^X - \frac{\sigma}{(\sigma - 1)^2} MA_i^M \, , \tag{23}$$

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

To sum-up, equations (14) and (15) (or the corresponding equations for the weight of trade) suggest a two-step approach to estimating the effect of roads on inter-city trade. In the first step, trade flows between any pair of cities i and j are explained by δ_i^X, city i's propensity to export, the cost of distance between these two cities, τ_{ij}, and city j's propensity to import, δ_j^M. In the second step, we explain the propensity to import and export as functions of city level variables. From equation (15), a city's propensity to export depends on city roads, productivity, population, and

[6] Differentiating expression (21) with respect to R_i and using the fact that the road elasticity of transportation costs is highest for industries producing heavier goods also gives this result.

Table 1: Summary statistics for our main variables

Variable	Mean	Std. Dev.	Minimum	Maximum
Value of exported shipments	75,652	95,944	1,757	605,126
Value of imported shipments	75,652	90,146	6,884	550,016
Weight of exported shipments	66,507	72,539	1,488	393,540
Weight of imported shipments	66,507	72,112	6,361	392,469
Employment, 2007	1,142,440	1,188,091	67,628	6,734,743
Lane km of interstate highway, 2005	3,131	2,325	434	14,440
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: 65 observations 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.

market access as defined by (17).[7] As we make clear below, this two step analysis maps directly into a two-step econometric approach.

3. Data

We now turn attention to our data. Our source of trade data is the 2007 Commodity Flow Survey (CFS). We combine the CFS with road data from the 2003-2005 Highway Performance Monitoring System (HPMS), employment data from the 2007 County Business Patterns (CBP), population counts from various US decennial censuses, and historical transportation data that we describe in further detail below. Summary statistics for our main variables are reported in table 1. Appendix C provides more detail.

The CFS reports bilateral commodity trade flows in both value (US dollar) and weight (tons) between CFS regions aggregated across sectors for 2007. Data are for all modes of transportation but are predominantly by road as we discuss in section 4. The lower 48 contiguous states are covered by 121 non-overlapping CFS regions, each constructed as an aggregate of adjacent counties. Our sample of "cities" consists of the 65 CFS 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). Our 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 areas like Gainsville (FL) is not part of a well identified CFS region. We

[7]Our approach shares some common features with the "new economic geography" estimations pioneered by Redding and Venables (2004). Like them, we use a two-step estimation with a gravity regression in the first step and the output of the first step is needed to construct a key second-step regressor. However, instead of focussing on a wage equation (our equation 11) our focus is on the determinants of trade flows.

exclude Washington DC, which the CFS defines as only the District of Columbia without the rest of its metropolitan region. Figures 1 to 3 represent our sample of CFS regions as shaded polygons.[8]

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. To measure city road capacity we use the 2005 HPMS and calculate lane kilometers of interstate highway in each CFS city. We also exploit the 2005 HPMS to compute pairwise highway distances between the employment centroids of all CFS cities. In robustness checks, we use traffic data from the 2003 HPMS.

We use CBP data from 2007 to calculate employment levels in each CFS city. Specifically, we use the CBP to calculate total city employment, the share of employment in manufacturing sectors, and city by sector level employment. 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 degree and a measure of income per capita.

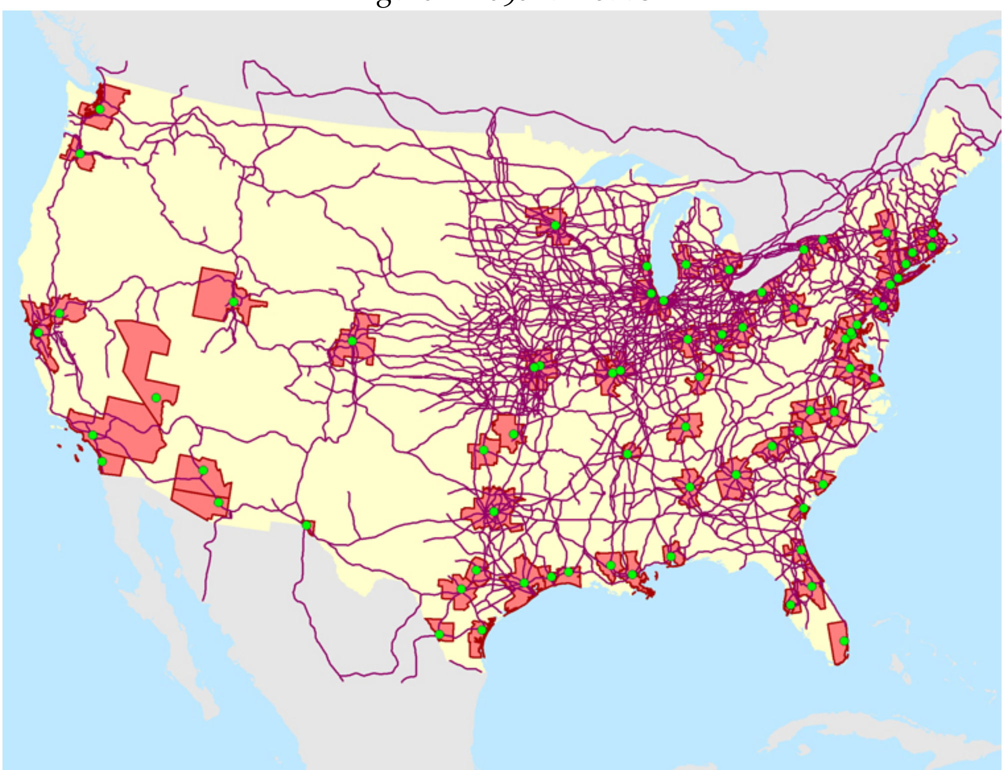
We also use three historical measures of the US transportation network in cities. The first is based on railroad routes around 1898. To construct this variable, we obtain a digital image of a map of major railroad lines (Gray, c. 1898) and convert this image to a digital map with the same format and projection as the map of 2005 interstate highways, as in Duranton and Turner (2011). Figure 1 is a map representing 1898 railroads and CFS cities. We then calculate kilometers of 1898 railroad in each CFS city. We use the same map of major railroad lines in 1898 to compute railroad distance between cities.

Our second historical transportation measure derives from the 1947 plan of the interstate highway system. To construct this variable we create a digital map of the 1947 highway plan from its paper record (United States House of Representatives, 1947). We then calculate kilometers of 1947 planned interstate highway in each MSA. Figure 2 shows a map of the planned highway network. We use the same map of the 1947 highway network to compute planned distance between cities.

Our third historical transportation measure is based on exploration routes in early US history. 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. A map of exploration routes between 1528 and 1850 is displayed in figure 3. Unlike 1898 railroads and 1947 planned highways, the nature of exploration routes data does not permit the calculation of route distances between cities.

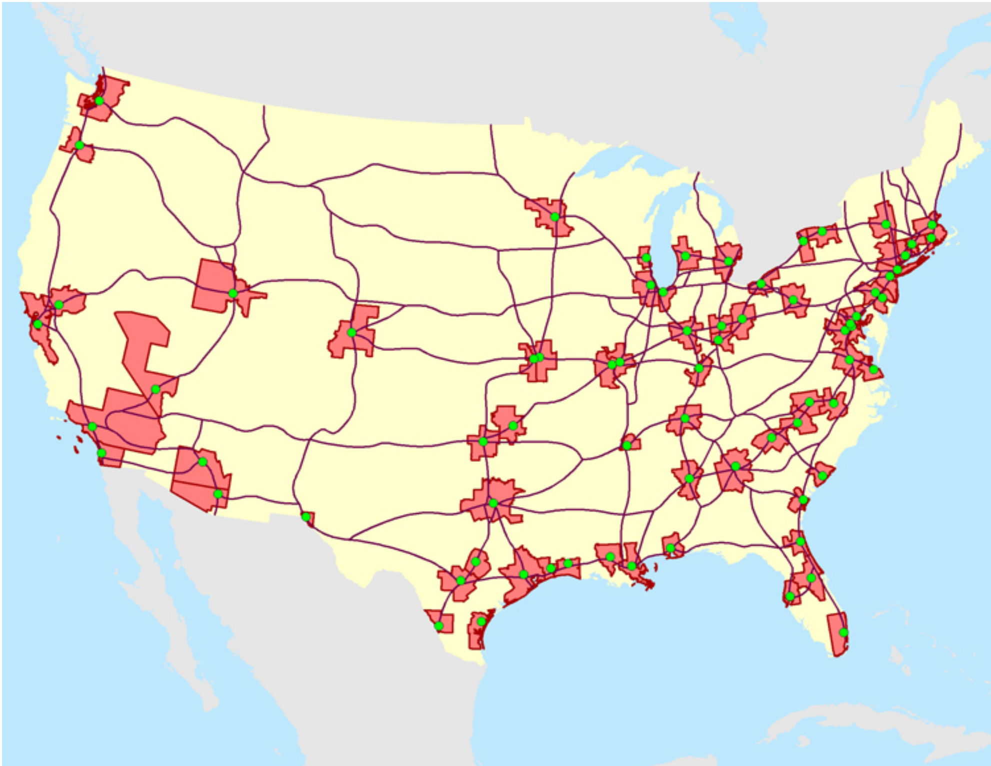
[8] Unfortunately the public use data does not break the data into exporter CFS region-importer CFS region-mode or exporter CFS region-importer CFS region-commodity observations because of Census disclosure standards.

Figure 1: 1898 railroads



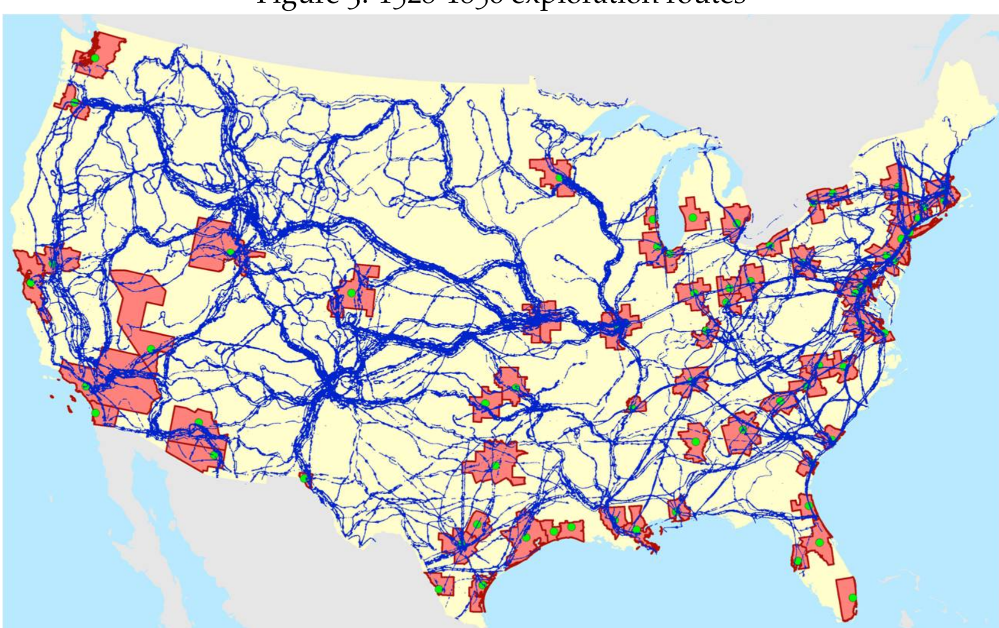
Source: Map based on Gray (c. 1898).
Notes: The lines are 1898 railroads. The shaded polygons are the CFS regions in our sample. The small dots indicate CFS region centroids.

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.

4. Estimation strategy

Econometric specification

Following directly from equation (14), our first-step regression equation is,

$$\ln X_{ij} = \delta_i^X + D_{ij}\alpha + \delta_j^M + \epsilon_{ij}\,. \tag{24}$$

In equation (24), D_{ij} is a polynomial of the instrumented log interstate highway distance between the employment centroids of cities i and j.[9] Equation (24) is of course a "gravity" equation for trade flows with fixed effects in the spirit of Redding and Venables (2004).[10] We apply it to the weight and value of trade flows between US cities.

Following directly from equation (15) our second-step regression is,

$$\hat{\delta}_i^X = \beta_0 + \rho_R^X \ln R_i + \beta' C_i + \mu_i\,, \tag{25}$$

where R_i is lane kilometers of roads in city i and C_i is a vector of city i characteristics. To measure city roads, we use lane kilometers of interstate highway within city boundaries. This measure reflects the capacity of a city's road network to move trucks into or out of the city quickly. Consistent with equation (15), in our baseline specification C_i contains log 2007 employment and a measure of market access computed according to equation (17) using the results of the first-step estimation of equation (24). There is a corresponding estimation explaining the weight of trade which uses δ_i^T as the dependent variable.

In our estimations, we focus more on the specification with fixed labor (15) than with mobile labor (23) for two reasons. First, population adjusts slowly to changes in roads (Duranton and Turner, 2010), so that the fixed labor model more nearly approximates what we observe than does the model based on free-mobility. Second, the specification with fixed labor faces a missing variable problem that does not arise under labor mobility. Our focus on the fixed labor model requires that we address this problem. With this said, the two specifications are similar. The main difference is the presence of an import market access term with mobile labor. In addition, import and export market access have a correlation of 0.97 so that their separate identification is not possible.

We face a number of inference problems in this second step. First, we do not have a direct measure of city productivity, A_i, which appears in equation (20). Thus missing variables which affect both the propensity of cities to export value or weight and the provision of city roads may confound our estimates. Second, the propensity to export may cause highway development in cities: more highways and more lanes may be built in cities that export intensively, or more highways may be built to help cities that do not export much. In either case, regressing cities' propensity to export on their stock of within city highways does not yield valid estimates of the

[9] 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.

[10] There are two differences between our specification and standard gravity estimation (e.g., Disdier and Head, 2007, Redding and Venables, 2004). Our use of a fourth-order polynomial for log distance yields imprecise estimates for the elasticity of trade with respect to distance, the main coefficient of interest in much of this literature. Because we use data for trade within a unified country, we do not include controls regarding language, existence of a common border, etc.

effect of city highways on exports. These inference problems can be resolved by instrumental variables estimation.

Instrumentation

Valid instruments need to predict current city highways but be otherwise uncorrelated with cities' propensity to export. To supplement our IV approach, we also consider a number of additional control variables. These control variables help to circumvent the problem of missing variables directly and help to preclude a correlation between our instruments and exports unrelated to city highways.

Estimating equation (25) with instrumental variables estimators can yield an unbiased estimate of the coefficients on lane kilometers of highway, ρ_R^X and ρ_R^T, if our instruments satisfy two conditions, relevance and exogeneity. Formally, these conditions are

$$\text{Cov}(R_i, Z_i|.) \neq 0 \tag{26}$$

and

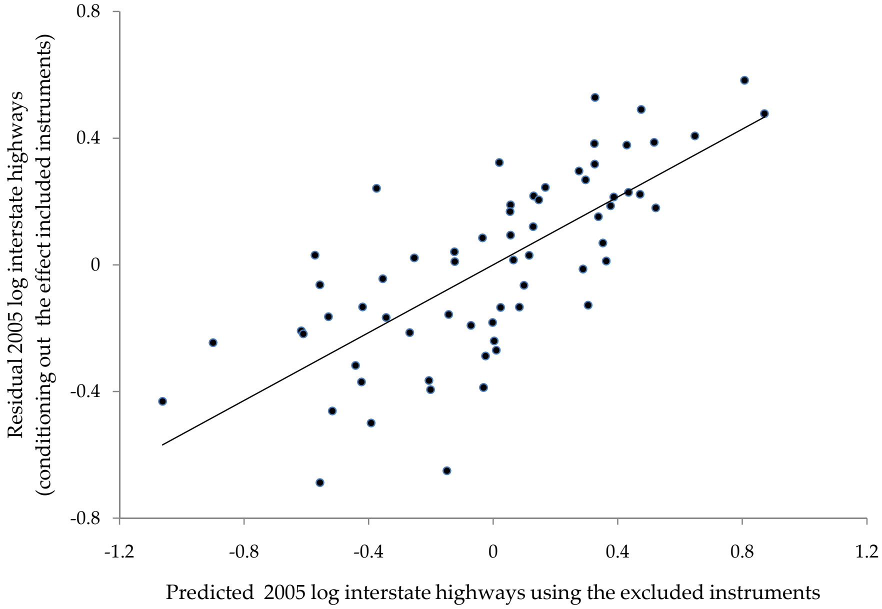
$$\text{Cov}(\mu_i, Z_i|.) = 0\,. \tag{27}$$

Similar conditions apply to the instrumental variable estimation of equation (24), the first step gravity equation.

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 have some ability to predict the stock of highways within a city. Starting with exploration routes, we note that they result from a search for an easy way to get from one place to another on foot, horseback, or wagon. Since a good route for a man, horse or wagon will likely be a good route for a car, exploration routes will often be good routes for contemporary highways. 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. Even without conversion, roads are much cheaper to build alongside railroad tracks. 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.

The raw data confirm that the 1947 highway plan, the 1898 railroad network, and 1528-1850 exploration routes predict the 2005 highway network. The correlation between log lane kilometers of highways and log planned 1947 kilometers of highways is 0.82, the correlation between log lane kilometers of highways and log 1898 kilometers of railroads is 0.71, and the correlation between log lane kilometers of highways and the log of our 1528-1850 exploration routes index is 0.46. While these correlations are interesting, equation (26) makes clear that the validity of an instrument depends on the partial correlation of the instrumental variables and the endogenous regressor.

Figure 4: First stage: Predicted highways and residual 2005 interstate highways



Notes: The horizontal axis plots "predicted highways" using 1528-1850 exploration routes, 1898 railroads, and 1947 planned interstate highways. The vertical axis plots residual 2005 interstate highways after conditioning out the effect of the included instruments. Both axes use the first stage of columns 4 and 8 of table 4.

To show the ability of our three instruments to predict log 2005 lanes kilometers of highway cities, figure 4 plots this variable after conditioning out the effect of employment, population, and market potential against the prediction coming from our three instruments. This graph shows that our three instruments provide a good prediction of this highway residual. 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.

Equation (27) requires that a valid instrument be orthogonal to the error term. In other words, the difficulty of inferring the effect of highways on trade in the estimation of equation (25) arises either because highways were allocated on the basis of future trade or because of a missing variable that drives both the allocation of highways and trade. A valid instrument can affect trade patterns only through its effects on contemporaneous highways, conditional on the controls. We now develop *a priori* arguments that our instruments satisfy this condition when the appropriate controls are included.

We first consider the 1898 railroad network. One argument for the validity of 1898 railroad kilometers as an instrument rests on the length of time since these railroads were built and the fundamental changes in the nature of the economy since 1898. The US rail network was developed during and immediately after the civil war, and during the industrial revolution (United States Bureau of Statistics, 1899, pp. 151 and 362). At the peak of railroad construction, around 1890, 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). Thus, as the country was building the 1898 railroad network its economy was much smaller and more agricultural than today.

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. Indeed, many manufactured goods traded in 2007 did not even exist in 1898. Moreover, the rail network was constructed by private companies expecting to make a profit from railroad operations in a not too distant future. It is difficult to imagine that these railroad entrepreneurs anticipated contemporary trade flows or city productivity in their planning.

Condition (27) requires that our instruments be uncorrelated with second stage errors, conditional on controls. Practically, this condition has two parts. First, that our instrument plausibly determines contemporary highways for some reason unrelated to trade flows in 2007, and second, that it not affect trade and roads through any channel orthogonal to our control variables. We now turn attention to the second part of this task and consider channels through which 1898 rail might directly affect trade in 2007.

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). This drastically reduces the likelihood that some unobserved factor which determines 1898 railroads also determines 2007 population and hence trade.

Second, more productive cities might have received more kilometers of railroad tracks in 1898. These cities might still be more productive today and, as suggested by our model, 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. In addition, there is little evidence that city productivity is persistent conditional on population size.[11] Thus, including three decadal population measures as controls should control for any persistent city productivity that explains both 1898 railroads and 2007 trade flows. 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 might cause an initial and persistent specialization into manufacturing. Such a manufacturing specialization could plausibly cause more trade independently of modern roads. This would invalidate 1898 railroads as an instrument. To address this possibility, we include the employment share of manufacturing as a control in some regressions. A positive, significant coefficient on 2005 lane kilometers of highways when this control is included indicates that cities with more highways export more even if we condition out their tendency to specialize into manufacturing activities.

Finally, there may be geographical features which are correlated with both 1898 railroads and 2007 trade flows. For example, the great lakes region might have fewer roads and railroad due to

[11] Glaeser and Gottlieb (2009) report strong mean reversal in income per capita for US cities for each decade between 1960 and 2000. The decline of erstwhile highly productive cities like Detroit or Buffalo illustrates this tendency.

exogenous limits imposed by the presence of water. 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.

We now turn to our maps of exploration routes. These maps describe major expeditions of exploration ranging over three centuries. These expeditions include those of the early Spanish explorers Hernando de Soto and Álvar Núñez Cabeza de Vaca, who explored the American South and Southwest in the mid-16th century, the great French explorer, Robert de LaSalle, who explored the Mississippi and the Great Lakes region in the late 17th century, the famous Lewis and Clark expedition of 1804 and Frémont's explorations of the West in the mid-1800's. 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.

While additional control variables help to make the case for instrument validity, using different instruments, for which threats to validity differ, allows for informative over-identification tests. For this reason, old explorations routes are an attractive instrument. We are concerned that 1898 railroad routes are not a valid instrument because of the possible correlation between 1898 railroads, persistent population patterns, and 2007 trade flows. However, early explorations of the US took place in areas that were sparsely populated by indigenous peoples. 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 face the same threats to validity as 1898 railroads and 1528-1850 exploration routes. First, planned highways could be correlated with 2007 trade flows and with some persistent determinant of population. Duranton and Turner (2010) find that coefficient on planned highways in a regression of log 1947 kilometers of planned highway on log 1950 population is almost exactly one. 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. In addition, we note that the claim that persistent city productivity led to more planned 1947 highways and greater trade today finds no support in the mandate for the 1947 highway plan. This mandate makes no mention of productivity and the planned 1947 allocation is well explained by population in 1950 but not by population growth around 1950.

Finally, planned 1947 highways are also attractive because geography appears to have played little or no role in their determination.

We are concerned, however, that the 1947 highway plan is correlated with city level patterns of sectoral specialization. That the planned network was required to serve the national defense was indeed a response to problems that emerged in coordinating the war effort in the early 1940s. This raises the possibility that cities specializing in sectors serving the national defense may be better served in the plan, and such initial differences in specialization may persist and affect 2007 trade flows. While this story is complicated and may be limited to a small number of defense-related sectors, we attach less weight to this instrument for this reason.

To conclude, for our three instruments to be invalid in our estimation of the effect of highways on exports there would need to be a mysterious variable that is correlated with all three of our instruments and with 2007 exports. This variable should not be associated with city populations past and present, persistent city productivity (to the extent that this can be captured by population growth and city education variables), specialization in manufacturing, and geography. To the degree that remaining endogeneity problems affect instruments in different ways, our overidentification tests are informative. In addition, the robustness section shows that all of our results are robust to pairwise inclusion of two of the three instruments being included at a time.

Other estimation and interpretation issues

In addition to these fundamental inference problems, other problems of inference and interpretation merit discussion. First, the stocks of other forms of infrastructure in a city, e.g., air or rail, might be correlated with city highways and affect trade. This is a special case of the missing variable problem which our IV estimation strategy addresses. Moreover, according to the Bureau of Transportation Statistics, single-mode truck shipments represent 88% of the value of all single mode shipments, as opposed to 5% for pipelines, 4% for rail, 2% for air, and 1% for water.[12] Thus, the fact that we do not measure other stocks of transportation infrastructure does not appear to be an important problem economically or econometrically.[13]

Second, we expect that higher quality products are more likely to be exported (Alchian and Allen, 1964, Hummels and Skiba, 2004). It would be straightforward to change labels in our model and think of lighter goods as being higher quality within a given sector. Our finding in section 5 that city highways have a strong positive effect on weight traded and an insignificant effect on value traded is consistent with both interpretations. In section 6, we provide direct evidence that more city highways lead to specialization in sectors that produce heavier goods. This does not preclude highways from having an effect on the quality of goods as well.

[12]Single-mode shipments represent more than 80% of all shipments. A large fraction of "multiple-modes" shipments are by courier companies for which no detailed breakdown is available. For multiple-modes shipments with a known combination of shipments, a large majority involve trucks.

[13]With this said, our findings pertain to the effects of the road infrastructure for internal trade within the US. The bulk of US international trade is by sea and by air. According to the Bureau of Transportation Statistics, single-mode truck shipments represent only 20% of the value of single-mode US exports (instead of 88% for internal trade) whereas water represents nearly 40% and air 36% of exported value (but less than 1% of tons shipped) in the 2007 CFS. While capacity is also clearly a fundamental issue for ports and airports, our results do not apply directly to these different contexts.

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. Because we do not have good measures of city productivity, our estimate of the coefficient for city highways in equation (25) reflects the sum of these two effects.

It may also be that city highways foster trade but not changes in production. This could happen if, for example, cities with more highways tend to serve as logistical centers for the trucking and warehousing sectors. We address this issue in two different ways. First, we control for employment in wholesale when estimating (25). Second, in section 6 we verify directly that city highways change the patterns of industrial specialization of cities as predicted by our model.

Finally, 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 (25). We follow the same FGLS procedure as Combes, Duranton, and Gobillon (2011) to recover corrected standard errors.

5. Estimating the effect of highways on trade

First-step results

We first estimate cities' import and export propensities for weight and value using the gravity specification (24). Results are reported in table 2. The dependent variable is the weight of bilateral trade flows in panel A and value of bilateral trade flows in panel B.[14] Column 1 regresses trade flows on exporter and importer fixed effects and log Euclidian distance using OLS. Columns 2 and 3 duplicate this specification using log 1947 highway distance and log 2005 highway distance, instead. Column 4 instruments log 2005 highway distance with log 1898 railroad distance. Columns 5 and 6 also include the square of log 2005 highway distance using OLS and TSLS, respectively. Columns 7 and 8 add cubic and quartic log distance terms.

Our estimates of the elasticity of the weight of trade with respect to distance are -1.16 to -1.18 and -0.81 to -0.82 for value for columns 1-4 of each panel. These estimates accord with those collected by Disdier and Head (2007). Given our use of importer and exporter fixed effects, R-squareds are appropriately high. Using Euclidian or highway distance yields similar results. Instrumenting 2005 highway distances with 1898 railroad distances does not change our results.[15] While the choice of distance variable and the estimation method do not matter, the specification does. The mean and median distance effects vary depending on whether they are estimated with a linear, quadratic, or quartic specification.

[14] The samples are not the same for both types of trade flows because of greater censoring for trade flows when measured in weight. In general, missing observations are due to disclosure requirements and, while containing small values, are not generally true zeroes. Using a consistent sample for both sets of regressions makes no change to the results. Because of this, we do not explore the issue of missing trade flows any further here.

[15] Instrumenting 2005 road distance by Euclidian distances gives very similar results. Because the 1947 highway plan was meant in part to reduce trade costs between cities, 1947 highway distances would be less suitable as instrument (but still yields similar results). Our data do not allow us to calculate network distances along exploration routes.

Table 2: First-step results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	OLS	OLS	TSLS	OLS	TSLS	OLS	TSLS
Distance	Eucl.	1947 hwy	2005 hwy	2005 hwy	2005 hwy	2005 hwy	2005 hwy	2005 hwy
Panel A. Dependent variable: Weight of bilateral trade flows.								
log distance	-1.18[a]	-1.16[a]	-1.16[a]	-1.16[a]	-0.082	-0.077	0.68	0.68
	(0.027)	(0.027)	(0.027)	(0.026)	(0.052)	(0.051)	(0.73)	(0.71)
log distance^2					-0.12[a]	-0.12[a]	-0.10	-0.071
					(0.0055)	(0.0053)	(0.33)	(0.32)
log distance^3							-0.051	-0.060
							(0.050)	(0.049)
log distance^4							0.0047[c]	0.0053[b]
							(0.0025)	(0.0024)
Mean effect	-1.18	-1.16	-1.16	-1.16	-1.82	-1.83	-1.52	-1.51
Median effect	-1.18	-1.16	-1.16	-1.16	-1.86	-1.87	-1.65	-1.64
R^2	0.80	0.79	0.79	-	0.83	-	0.85	-
First-stage Stat.				560,640		77,689		199
Panel B. Dependent variable: Value of bilateral trade flows.								
log distance	-0.82[a]	-0.81[a]	-0.81[a]	-0.81[a]	-0.21[a]	-0.20[a]	-0.41	-0.55
	(0.019)	(0.020)	(0.020)	(0.020)	(0.046)	(0.045)	(0.60)	(0.60)
log distance^2					-0.066[a]	-0.066[a]	0.39	0.47[c]
					(0.0044)	(0.0044)	(0.27)	(0.27)
log distance^3							-0.12[a]	-0.13[a]
							(0.040)	(0.040)
log distance^4							0.0078[a]	0.0086[a]
							(0.0020)	(0.0020)
Mean effect	-0.82	-0.81	-0.81	-0.81	-1.16	-1.16	-0.88	-0.87
Median effect	-0.82	-0.81	-0.81	-0.81	-1.18	-1.18	-1.01	-1.01
R^2	0.82	0.82	0.82	-	0.83	-	0.85	-
First-stage Stat.				529,916		79,482		222

Notes: All regressions include importer and exporter fixed effects for all cities. The same regressions are run in both panels with 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 2,639 observations and 3,299 observations in panel B. In column 1, distance is Euclidian distance. Column 2 uses highway distance from the 1947 planned highways map. Columns 3-8 use 2005 highway distance. In column 4, 6, and 8, highway distance terms are instrumented by their corresponding 1898 railroad distance terms. Robust standard errors in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%.

For most of our second-step regressions below we use the fixed effects estimated in column 8. Despite differences in the estimated effect of distance, the estimates for the exporter and importer fixed effects are very stable. The pairwise correlations between the exporter effects estimated in panel A are all 0.98 or above. The same holds for importer fixed effects. The corresponding correlations for the fixed effects estimated in panel B are equally strong or stronger. Equally strong results are obtained when estimating Spearman rank correlations.[16]

We draw three conclusions from these estimations. First, the choices of distance variable, specification, and estimation method do not affect our estimates of importer and exporter fixed effects. This is important because these fixed effects are used as dependent variables in what follows. Second, the similarity between OLS and TSLS results is strongly suggestive of no endogeneity bias for highway distance between cities. Third, highway distance has a large effect on trade flows but the high correlation between highway distance and Euclidian distance suggests few opportunities for improvement.

Main second-step results

We now turn to the estimation of equation (25). Table 3 reports results for OLS regressions using the exporter fixed effects estimated in the first step as dependent variable. The second through the fourth rows contain standard errors that are robust, standard OLS, and calculated according to the method proposed by Combes *et al.* (2011), respectively. Column 1 is a rudimentary OLS specification using only log 2005 lane kilometers of city highways as an explanatory variable for exporter weight fixed effects. Column 2 adds log employment in 2007. Column 3 is the baseline specification coming from theory (equation 15) which includes city highways, employment, and market access. Market access is computed as suggested by equation (17) using the results from the same first-step specification that we use to estimate the exporter fixed effects.[17] Column 4 is our preferred specification as it includes includes log population from 1920, 1950, and 2000 to control for the assignment of highways and rails based on historical and not just contemporary population patterns. 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.

The coefficient on city highways is high in the two rudimentary specifications of columns 1 and 5. However, as soon as employment is added, this coefficient shrinks dramatically and becomes insignificant. We nonetheless note that the point estimates for the coefficient on highway lane kilometers are higher when using the exporter weight fixed effect as dependent variable in columns 2-4 than for value in columns 6-8 within each specification. Other results also merit discussion. First, the exporter value fixed effect is roughly proportional to employment in column

[16]We also note that the correlations between the weight and value fixed effects estimated from column 8 are 0.88 for exporters and 0.95 for importers. The correlations between importer and exporter fixed effects are 0.69 for weight and 0.78 for values.

[17]We do not include a city's own importer effect when doing this computation. There are two reasons for that. First, our gravity regression is estimated from trade flows external to cities. Second, adding a city's own effect can potentially create serious simultaneity biases. Think for instance of situations where some cities imports intermediate goods and exports final goods.

Table 3: 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 lane km	0.97[a]	0.28	0.30	0.19	1.17[a]	0.10	0.11	0.043
Robust s.e.	(0.18)	(0.24)	(0.22)	(0.19)	(0.16)	(0.15)	(0.14)	(0.13)
Non-robust s.e.	(0.14)	(0.21)	(0.20)	(0.18)	(0.14)	(0.16)	(0.16)	(0.14)
Corrected s.e.	(0.14)	(0.21)	(0.20)	(0.18)	(0.14)	(0.16)	(0.16)	(0.14)
log employment		0.68[a]	0.68[a]	0.70		1.06[a]	1.05[a]	1.74[a]
		(0.15)	(0.15)	(0.44)		(0.099)	(0.089)	(0.57)
Market access (export)			-0.58[a]	-0.97[a]			-0.30[c]	-0.57[a]
			(0.21)	(0.19)			(0.15)	(0.19)
log 1920 population				-0.51[c]				-0.51
				(0.29)				(0.35)
log 1950 population				1.49[a]				1.32[b]
				(0.46)				(0.53)
log 2000 population				-0.91				-1.49[b]
				(0.55)				(0.74)
R^2	0.43	0.55	0.60	0.72	0.53	0.77	0.79	0.84

Notes: 65 observations per column. All regressions include a constant. Robust standard errors in parentheses unless otherwise mentioned. *a*, *b*, *c*: significant at 1%, 5%, 10%.

6 and 7 whereas the exporter weight fixed effect is less than proportional to employment: bigger cities export lighter goods. Second, the coefficients on market access are negative and significant as predicted by the theory. Third, correcting for the sampling error associated with our use of an estimated dependent variable effects only the third decimal of the standard errors. Hence, we suppress corrected standard errors for our subsequent TSLS estimation. Finally, R-squareds are high and are higher for values than weights. Table 8 in Appendix D replicates table 3 using importer rather than exporter fixed effects as the dependent variable.[18] The results are generally the same as with the exporter fixed effects.

Table 4 mirrors table 3 but, instead of OLS, uses TSLS estimation and instruments 2005 lane kilometers of interstate highways with 1528-1850 exploration routes, 1898 railroads, and 1947 planned highways. In all TSLS specifications, the coefficient on instrumented lane kilometers of highways increases relative to the corresponding OLS coefficient, while coefficients of other explanatory variables remain unchanged. In column 3, the elasticity of weight of exports with respect to lane kilometers of highways is 0.72 and highly significant. For the value of trade (column 7), the coefficient is lower at 0.25 and insignificant. In the preferred specification of column 4 for weight, the elasticity of the weight of exports with respect to lane kilometers of highways is 0.51 and significant. For the value of trade, the coefficient is again lower and insignificant. The

[18]Following the specification derived in Appendix A, regressions in table 8 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.

Table 4: 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 lane km	1.14[a]	0.84[a]	0.72[a]	0.51[b]	1.19[a]	0.32	0.25	0.11
	(0.17)	(0.29)	(0.25)	(0.22)	(0.17)	(0.22)	(0.20)	(0.16)
log employment		0.32	0.41[b]	0.56		0.92[a]	0.96[a]	1.71[a]
		(0.19)	(0.17)	(0.46)		(0.15)	(0.13)	(0.54)
Market access (export)			-0.60[a]	-0.96[a]			-0.30[b]	-0.57[a]
			(0.21)	(0.19)			(0.15)	(0.18)
log 1920 population				-0.55[b]				-0.52
				(0.25)				(0.32)
log 1950 population				1.51[a]				1.32[a]
				(0.40)				(0.49)
log 2000 population				-0.98[c]				-1.50[b]
				(0.55)				(0.71)
Overid. p-value	0.32	0.28	0.10	0.41	0.21	0.06	0.05	0.20
First-stage Stat.	49.3	21.9	23.3	21.0	49.3	21.9	23.3	21.0

Notes: 65 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 lane kilometers of interstate highways. Robust standard errors in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%.

first-stage statistics indicate that the instruments are not weak. Consistent with the discussion in section 3, we pass overidentification tests easily 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.

Table 9 in Appendix D performs the same TSLS regressions as in table 4 but uses importer instead of exporter fixed effects as the dependent variable. The results are essentially the same as those reported in the corresponding OLS table (table 8). The coefficient on highways for the weight of imports is positive but insignificant. The coefficient on highways for the value of imports is also insignificant, but with smaller point estimates.

For weight exported, and to a lesser extent for value exported, the TSLS road coefficients are larger than their OLS counterparts. It is possible that this difference reflects classical measurement error in our "highway lane kilometers" variable. This seems implausible: given the role of the HPMS in the allocation of highway funding, misreported data would be noticed. Alternatively, it may be that old exploration routes, 1898 railroads, and 1947 planned highways are better at predicting highways that are important for trade. To eliminate this possibility, we confirmed that the difference between OLS and TSLS estimates persists when we use different measures of highways. Having eliminated alternative explanations, we conclude that highways allocated at random have a larger effect on trade than highways allocated through the prevailing allocation process.

Duranton and Turner (2010), Baum-Snow (2007), and Duranton and Turner (2011) all arrive at a similar conclusion in somewhat different circumstances. Duranton and Turner (2010) argue that the difference between TSLS and OLS estimates of the effect of roads on city level population growth reflect an association between negative growth shocks in cities and greater employment in road construction. That is, roads really are allocated to cites that are less productive. This finding is consistent with the evidence presented by Glaeser and Gottlieb (2008) that highway funding in practice is disproportionately allocated to poorer low density locations. In turn, this suggests two non-exclusive explanations for the smaller impact on trade of highways assigned by the equilibrium process. First, highway construction could be used as "make-work" in poorly performing cities. Second, land costs might be higher in cities that export more. If so, then constant per capita highway funding rates will lead high exporting cities to be underprovided with highways.[19]

The effect of highways on the weight of exports is economically large. Using equations (24) and (25), we calculate that multiplying the stock of city highways by a factor α changes the weight exported by a factor of $\alpha^{0.51}$. Thus, doubling the stock of city highways causes a 42% expected increase in export weight. To place this magnitude in the context of our sample, we consider Buffalo (NY) and Nashville (TN). These two cities have roughly the same CFS area population, 1.2 and 1.3 million. Buffalo is at the 25th percentile in the distribution of interstate lane kilometers whereas Nashville is at the 75th percentile, with Nashville having 144% more highways. Our estimates predict that Nashville should export $2.44^{0.51} = 1.58$ or 58% more tons than Buffalo. This prediction slightly understates the 64% difference in exported tons between these two cities in the CFS data. On the other hand, the values exported by the two cities are approximately equal.

Alternatively, using the -1.51 value of elasticity of trade in tons with respect to distance estimated in column 8 of table 2, the increase in trade associated with a movement from the 25th to the 75th percentile of lane kilometers of highways could also be achieved by a reduction of highway distance to other cities of 26%. This is equivalent to moving New York to Chicago with respect to shipments to San Francisco.[20]

Robustness checks

We now check that our estimates of the effects of city highways on exports are robust to; our econometric technique, our choice of instruments, our use of a two-step rather than one-step econometric specification, the precise econometric specification used to generate the fixed effects in the first step, and additional control variables introduced to evaluate alternate hypotheses.

Table 10 in Appendix D assesses the robustness of our baseline and preferred TSLS estimations to choice of econometric technique and instruments. Using LIML instead of TSLS does not change

---

[19]The literature provides further support for these explanations. From the descriptions of Knight (2002, 2004) and Duranton and Turner (2010), funding for interstate highways in the US is determined by a formula, but with the possibility for legislators in the relevant committees to alter the distribution. Since the formula aims to equalize the distribution of funds per capita, it is biased against road provision where land is expensive and difficult to assemble.

[20]In our model with a continuum of cities, it is easy to compute the effect of shorter distances on trade. With discrete cities, roads between cities affect the price indices everywhere leading to a more complicated calculation (Anderson and van Wincoop, 2003).

our results.[21] This is expected given the strength of our instruments. This table also shows that different pairwise combinations of instruments yield the same results as our three instruments taken together. Consequently, no single instrument appears to be driving our results. For our preferred specification the coefficient on highways varies between 0.48 and 0.56, and is thus very close to our estimate of 0.51 in table 4. This is of course consistent with the failure to reject the overidentification test associated with this estimate. For our baseline estimation, which does not control for population in 1920, 1950, and 2000, there is more variation for the highway coefficient (between 0.57 and 0.90). This is also consistent with the marginal overidentification test statistic in column 3 of table 4.

While our theoretical model maps naturally into a two-step estimation procedure, the following regression allows a similar estimate of the effects of highways on trade in one step,

$$\ln X_{ij} = \lambda_0 + \rho_R^X \ln R_i + \lambda^{X'} C_i + \alpha' D_{ij} + \rho_R^M \ln R_j + \lambda^{M'} C_j + v_{ij} \,. \tag{28}$$

As previously, C_i and C_j are city level controls, D_{ij} is a vector of polynomials of highway distances between i and j, and R_i and R_j are lane kilometers of highways in exporting city i and importing city j, respectively.

This one-step approach suffers from two drawbacks relative to our two-step estimation. First, our theoretically derived measure of market access is estimated from a first step gravity equation. By construction we cannot use the market access variable proposed by theory in the one-step estimation of equation (17). Hence, for this one-step estimation we use a more ad-hoc measure of market access where, for each city, we sum aggregate income in other cities weighted by inverse distance. The second drawback of one-step estimations is that changing the set of city controls can affect the estimates of the effect of distance between cities, arguably an undesirable property. When estimating (28), we need to account for the fact that the error term v_{ij} has three components: one associated with city i, another associated with city j, and one associated with the pair ij. This suggests a two-way clustering and we follow the procedure developed by Cameron, Gelbach, and Miller (2010). In addition, we try to keep our instrumentation strategy simple. In our TSLS estimations of (28) we only instrument lane for kilometers of city highways but not highway distances between cities.

Table 11 in Appendix D presents results for one-step estimations. Panel A corresponds to the OLS results of table 3 and 8 and panel B to the TSLS results of table 4 and 9. In all cases the one-step estimates are within one standard error of their two-step counterparts and often differ only in the second decimal place. The coefficient on exporter highways using the one-step version of our preferred specification is 0.47 instead of 0.51 in the two-step estimation.

In tables 3 and 4, our dependent variable, the propensity to export, is an exporter fixed effect estimated from a particular gravity regression. In addition, an important explanatory variable, market access, results from a theoretical construct (equation 17) which also makes uses of coefficients estimated in the first step. Table 12 in Appendix D reports the results of a number of specifications using fixed effects from alternative gravity estimations.[22] From that table, we can see that using an

[21] 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.

[22] We also adjust the computation of market access to the firsts-step specification being used.

Table 5: Second-step results, TSLS for weight exporter fixed effects with extra controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A.	Water+Slope		Census div.		% College		Income p.c.	
log highway lane km	0.68^a	0.39^b	0.57^a	0.45^b	0.54^b	0.39^c	0.70^a	0.49^b
	(0.25)	(0.20)	(0.21)	(0.19)	(0.24)	(0.20)	(0.25)	(0.22)
log populations 20, 50, 00	N	Y	N	Y	N	Y	N	Y
Overid. p-value	0.23	0.72	0.40	0.16	0.03	0.60	0.09	0.43
First-stage Stat.	29.4	27.1	25.1	23.1	23.2	20.5	23.0	20.9
Panel B.	Share manuf.		Share wholesale		Traffic		Urban highways	
log highway lane km	0.49^b	0.42^b	0.70^a	0.53^b	0.74^a	0.70^a	1.28^b	0.93^c
	(0.20)	(0.20)	(0.23)	(0.21)	(0.25)	(0.24)	(0.51)	(0.50)
log populations 20, 50, 00	N	Y	N	Y	N	Y	N	Y
Overid. p-value	0.46	0.69	0.28	0.54	0.09	0.43	0.33	0.52
First-stage Stat.	21.6	20.1	21.8	19.2	24.8	24.5	6.13	3.65

Notes: 65 observations per column. All regressions include a constant, log 2007 employment, and export market access, and use log 1947 planned highway km, log 1898 railroad km, and log 1528-1850 exploration routes index as instruments for log lane kilometers of interstate highways. Robust standard errors in parentheses. In panel A, columns 1 and 2 include average land gradient and distance to the nearest body of water. Columns 3 and 4 include dummy variables for census regions. Columns 5 and 6 include the log share of the fraction of adult population with a college degree or more. Columns 7 and 8 include the log of average income per capita. In panel B, columns 1 and 2 include the log of the share of manufacturing employment. Columns 3 and 4 include the log of the share of employment in wholesale trade. Columns 5 and 6 include the log of average daily traffic on the interstate highways in 2005. Columns 7 and 8 use log interstate highway lane kilometers in the urbanized part of CSA regions instead of interstate highways in the entire region. All columns are estimated with TSLS and except for columns 7 and 8 in panel B which are estimated with LIML. *a*, *b*, *c*: significant at 1%, 5%, 10%.

OLS linear gravity specification in the first step leads to marginally higher points estimates for the elasticity of exports with respect to highways. Using an OLS polynomial specification for gravity instead of its corresponding TSLS makes no difference. Using an ad-hoc measure of market access based on income leads to slightly lower estimates. None of these estimates is statistically different from the corresponding estimates in table 4.

Finally, it is possible that the exclusion restriction (27) might not be satisfied even in our preferred specification. In table 5, we experiment with including a variety of additional control variables to our preferred estimates of the effect of city highways on propensity to export. In columns 1 and 2 of panel A of table 5, we duplicate our baseline and preferred specifications adding two key geographical controls: the log distance to the closest body of water (Ocean, Gulf or Great Lakes) and an index of average land gradient. In columns 3 and 4, we conduct the same exercise, adding dummy variables for census regions. In columns 5 and 6, we include the log of the share of adult population with a college degree. This specification also evaluates the Heckscher-Ohlin hypothesis that specialization comes from differences in the abundance of skilled labor. In columns 7 and 8, we include log income per capita. In columns 1 and 2 of table 5 panel B, we include the log share of manufacturing employment. In columns 3 and 4, we include the log share of employment in wholesale. In columns 5 and 6, we include highway average annual daily traffic. Finally, in columns 7 and 8, we use lane kilometers of interstate highways in the "urbanized" parts of the CFS cities as an alternative measure of city highways.

None of the additional control variables we use in table 5 changes our results. For our preferred specification, these additional controls sometimes increase our preferred estimate. In other cases, they reduce it. All alternative estimates lie within one standard error of our preferred estimate, often much closer. The largest change in the coefficient on highways occurs when using a more restrictive measure of urban interstate highways in columns 7 and 8 of panel B. However, the standard errors are also larger and our instruments are weak in these regressions.[23] Nonetheless it is striking that the estimated effect of highways on the propensity to export is larger when we consider the more central and congested highways where we expect most bottlenecks to occur (interstate highways in urbanized areas) or when we control for traffic.

6. Results about specialization

So far we have found that highways have a large and significant effect on exported weights and a small and insignificant effect on exported values. A natural corollary is that highways have a positive effect on weights per unit value, i.e., tons per dollar of exports as measured by total weight divided by total value. In table 6 we verify that this is indeed the case by using exported weight per unit value (i.e., tons per dollar) as our dependent variable. Columns 1 and 2 report OLS export results for two second-step specifications that correspond to our baseline and preferred explanatory variables. Columns 3 and 4 report TSLS results for the same two specifications. Columns 5 to 8 repeat the exercise for imported weight per unit value.

The results in table 6 are fully consistent with our findings so far. The coefficient on log weight per value is typically the difference between the corresponding coefficients in regressions to predict weight and value, respectively. More highways cause cities to export heavier goods.

We now conduct a direct test of the claim in Proposition 1 that more city highways increase employment in sectors producing heavier goods. To begin, for each sector, we predict city level employment as a function of city highways. That is, we estimate,

$$\ln N_i^k = \beta_0^k + \rho_R^{N,k} \ln R_i + \beta^{k'} C_i + \epsilon_i^k \,, \tag{29}$$

where N_i^K is employment in sector k and city i. Aside from within city highways, R_i, the regression also includes the vector of control variables C_i from our preferred regressions, i.e., log employment, population in 1920, 1950, and 2000, and market access. The coefficient of interest is $\rho^{N,k}$, the elasticity of sector employment with respect to lane kilometers of highway.

We next use $\hat{\rho}_R^{N,k}$ as the dependent variable in a regression where the explanatory variable of interest is average tons per dollar of output for sector k, denoted UV^k as measured by US total

[23]Our instruments predict overall CFS highways well but are less good at predicting highways in the smaller urbanized portions of CFS regions.

Table 6: Second-step results, weight per unit value (tons per dollar)

log unit weight	(1) OLS export	(2) OLS export	(3) TSLS export	(4) TSLS export	(5) OLS import	(6) OLS import	(7) TSLS import	(8) TSLS import
log highway lane km	0.26[c]	0.21	0.53[a]	0.44[b]	0.10	0.088	0.18[c]	0.16
	(0.15)	(0.15)	(0.18)	(0.19)	(0.078)	(0.080)	(0.099)	(0.10)
log employment	-0.38[a]	-0.90[c]	-0.55[a]	-1.00[b]	-0.066	-0.56[b]	-0.12[c]	-0.59[b]
	(0.11)	(0.53)	(0.12)	(0.48)	(0.059)	(0.23)	(0.069)	(0.24)
Market access (export)	-0.32[b]	-0.44[a]	-0.32[b]	-0.43[a]	-0.32[a]	-0.36[a]	-0.33[a]	-0.36[a]
	(0.15)	(0.16)	(0.15)	(0.15)	(0.079)	(0.083)	(0.078)	(0.080)
log 1920 population		0.11		0.077		0.038		0.028
		(0.29)		(0.26)		(0.18)		(0.17)
log 1950 population		0.064		0.082		-0.030		-0.024
		(0.44)		(0.38)		(0.30)		(0.27)
log 2000 population		0.45		0.40		0.54[c]		0.52[c]
		(0.68)		(0.60)		(0.31)		(0.31)
R^2	0.26	0.31	-	-	0.32	0.38	-	-
Overid. p-value	-	-	0.81	0.77	-	-	0.22	0.25
First-stage Stat.	-	-	23.3	21.0	-	-	23.3	21.0

Notes: 65 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 lane kilometers of interstate highways. Robust standard errors in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%.

Table 7: Specialization, OLS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable: industry specific coefficient on interstate highways								
estimated with:	OLS	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS	TSLS
using additional controls:	-	-	Water + slope	Census regions	College	Mining emp.	Share manuf.	Share wholesale
log unit weight	0.053	0.15^a	0.15^a	0.13^a	0.16^a	0.15^a	0.18^a	0.15^a
	(0.031)	(0.024)	(0.024)	(0.024)	(0.027)	(0.031)	(0.022)	(0.023)
R^2	0.17	0.39	0.39	0.44	0.38	0.30	0.49	0.40

Notes: 22 observations per column except for column 6 with 21 observations. The dependent variable is the industry specific coefficient on interstate highways estimated from regression (29) using log lane kilometers of interstate highways, log 2007employment, log 1920, 1950, and 2000 population and log market access. 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 graduate (column 5), log share manufacturing employment (column 7), and log share employment in wholesale (column 8). Robust standard errors in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%.

weight divided by total value shipped,

$$\hat{\rho}_R^{N,k} = \gamma_0 + \gamma_1 \ln UV^k + v^k. \tag{30}$$

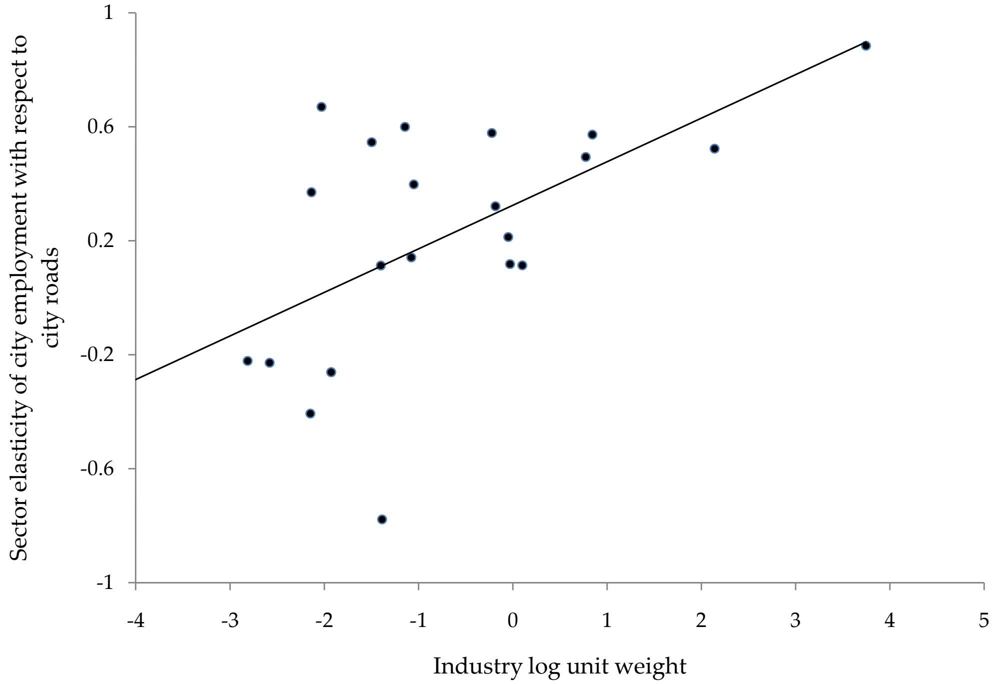
If $\gamma_1 > 0$ then cities specialize in sectors producing heavier goods when they have more highways. Table 7 reports estimates of equation (30). In column 1, the dependent variable is the highway elasticity of employment when (29) is estimated with OLS. To check the robustness of this finding, the remaining columns in table 7 replicate the regression of column 1, but experiment with different ways of estimating the highways elasticity of employment. 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 or water and log average land gradient as control variables when estimating $\rho_R^{N,k}$. In column 4, we use census region indicators as extra controls. In column 5, we use the log share of adult population with a college degree. In column 6, we estimate include the proportion of employment in mining to control for natural resource abundance. In columns 7 and 8 we use the log share of manufacturing and wholesale employment as controls.

The table shows a stable positive correlation between weights per unit value by sector and the elasticity of sector employment with respect to highways. Figure 5 plots the data underlying column 2 of table 7. This figures shows that in sectors producing heavy goods, employment increases with city highways whereas in sectors producing light goods, it decreases with city highways. These results correspond exactly to the prediction of Proposition 1.

7. Conclusion

This paper examines the effect of highways infrastructure on the level and composition of trade in US cities. We find that the weight and value of bilateral trade decreases rapidly with the highway

Figure 5: Specialization



Notes: The horizontal axis plots log weight unit per dollar by sector. The vertical axis plots the elasticity of industry employment with respect to lane kilometers of interstate highways estimated as in column 2 of table 7.

distance between cities, with the weight of trade decreasing more rapidly than its value.

We also find that within city highways affect trade. A 10% increase in a city's stock of highways causes about a 5% increase in the weight of exports, but does not cause a measurable change in the value of exports. This means that an increase in within city highways causes a city to specialize in the export of heavy goods. In fact, our estimates require a 10% increase in within city highways to cause about a 5% decrease in the unit value of the city's exports. Within city highways give cities an economically important comparative advantage in the production of heavy goods.

The large effect on trade of reductions in pairwise distance means that adding segments to the interstate highway system which reduce these pairwise distances is likely to have large effects on trade and welfare. However, the high correlation between Euclidean and highway distance suggests that there are few opportunities to build such segments. That city highways do not increase the value of exports suggests that the changes in trade caused by city highways probably do not have large welfare effects. Together with the observation in Duranton and Turner (2011) that trucks account for only about 12% of within city traffic, this suggests planners should not give much consideration to trade effects when planning a city's highway network. One possible exception to this conclusion might occur if a city has a comparative advantage in the production of heavy goods that cannot be exploited because the city's highway network is inadequate. In this case, our results suggest that building highways could have an important impact on the city's ability to specialize in the production of heavy goods.

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.

Anderson, James E. and Eric van Wincoop. 2004. Trade costs. *Journal of Economic Literature* 42(3):691–751.

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

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, (forthcoming).

Behrens, Kristian and Pierre M. Picard. 2011. Transportation, freight rates, and economic geography. Processed, UQAM.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2010. Robust inference with multiway clustering. *Journal of Business and Economic Statistics* 29(2):238–249.

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, Gilles Duranton, and Laurent Gobillon. 2011. The costs of agglomeration: Land prices in French cities. Processed, University of Toronto.

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.

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

Disdier, Anne-Célia and Keith Head. 2007. 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. 2010. Railroads of the Raj: Estimating the impact of transportation infrastructure. Processed, MIT.

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. 2010. Urban growth and transportation. Processed, University of Toronto.

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

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. 2008. The economics of place-making policies. *Brookings Papers on Economic Activity* (1):155–253.

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.

Hummels, David and Alexandre Skiba. 2004. Shipping the good apples out? An empirical confirmation of the Alchian-Allen conjecture. *Journal of Political Economy* 112(6):1384–1402.

Knight, Brian. 2002. Endogenous federal grants and crowd-out of state government spending: Theory and evidence from the federal highway aid program. *American Economic Review* 92(1):71–92.

Knight, Brian. 2004. Parochial interests and the centralized provision of local public goods: Evidence from congressional voting on transportation projects. *Journal of Public Economics* 88(3-4):845–866.

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.

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. Full derivation of the importer effect

From (7), (10), and (14) it is easy to show that

$$e^{\delta_i^X + (1-\sigma)\ln\tau_{ij}} = \int_0^1 \left(\frac{\tau_i^k\,\tau_{ij}\,W_i}{A_i}\right)^{1-\sigma} \mathrm{d}k\,. \tag{A1}$$

Integrating both sides of (A1) across all cities i, using (5), (6), and (10) again, and taking logs yields

$$MA_j^M \equiv \ln \int_0^1 e^{\delta_i^X + (1-\sigma)\ln\tau_{ij}}\,\mathrm{d}i \;=(1-\sigma)(\ln\mathbb{P}_j - \ln\tau_j). \tag{A2}$$

Following Redding and Venables (2004), the term on the left-hand-side can be naturally interpreted as an import market access. We can now insert (A2) into (18) to obtain

$$\delta_j^M = \ln(N_j\,W_j) - MA_j^M\,. \tag{A3}$$

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 in this expression. This is because cities are of measure zero. 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 affected. There is nonetheless an indirect effect. Roads in j will affect imports through their effect on the income of that city. To see this, we can use equation (11) into (A3) to eliminate city j's wage and obtain

$$\delta_j^M = S(R_j) + \frac{\sigma-1}{\sigma}\ln(A_j) + \frac{\sigma-1}{\sigma}\ln(N_j) + \frac{\sigma}{(\sigma-1)^2}MA_j^X - MA_j^M\,. \tag{A4}$$

Like its corresponding expression on the export side (equation 15), equation (A4) contains a productivity term that is not observed in the data.

Appendix B. Worker mobility

In this appendix, we 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

$$q_{ji}^k = \frac{\mathbb{P}_i^{\sigma-1}}{(P_{ji}^k)^\sigma}W_i\,. \tag{B1}$$

Inserting this expression into the utility function (2) and making use of (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 $\ln U_i = \ln W_i - \ln\mathbb{P}_i = \underline{u}$. Combining this expression with (A2) leads to

$$W_i = \underline{u} + \ln\tau_i - \frac{MA_i^M}{\sigma-1}\,. \tag{B2}$$

From (10) and (7), it is also easy to get $\delta_i^X = S_2(R_i) + (\sigma - 1)\ln A_i - (\sigma - 1)\ln W_i$ where the exporter effect δ_i^X is as in (14) and the effect of the roads $S(R_i)$ is defined by (16). Using (15) and (B2) into this expression, we can eliminate A_i and W_i and write expression (22) in the main text. Applying a similar reasoning to the weight of exports in equation (20) is straightforward and leads to expression (23) in the main text.

Appendix C. 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 at least one non-zero realization of the instrumental variables leaves us with a sample of 65 CFS areas. See United States Bureau of Transportation Statistics (2010) for details about the CFS.

Road infrastructure: To measure each CFS city's stock of highways we use the US Highway Performance and Monitoring System (HPMS) "Universe" data for 2003. These data are described extensively in United States Department of Transportation, Federal Highway Administration (2005) and Duranton and Turner (2011).

To calculate Interstate Highway distances between cities we first identify the centroid of each CFS city. To do this, we find the area weighted centroid of the urbanized portions of each county comprising the 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 National Highway Planning Network (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 and along 1898 railroad routes.

Employment data: To measure employment we use the County Business Patterns data from the US Census Bureau. These data are available annually from 1977 to 2003. Since CFS cities are defined as aggregations of counties, these data allow us to construct measures of aggregate employment for 1983, 1993, and 2003. We also construct disaggregated employment data at the three digit-level.

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 D. Other second-step results

Table 8: Second-step results, OLS for importer fixed effects

Importer fixed effect	(1) weight	(2) weight	(3) weight	(4) weight	(5) value	(6) value	(7) value	(8) value
log highway lane km	0.81[a]	0.11	0.23[c]	0.18	0.86[a]	-0.019	0.038	0.015
Robust s.e.	(0.21)	(0.17)	(0.12)	(0.11)	(0.17)	(0.13)	(0.10)	(0.10)
Non-robust s.e.	(0.14)	(0.21)	(0.16)	(0.15)	(0.13)	(0.16)	(0.14)	(0.14)
Corrected s.e.	(0.14)	(0.21)	(0.16)	(0.16)	(0.13)	(0.16)	(0.14)	(0.14)
log employment		0.69[a]	0.61[a]	-0.71		0.87[a]	0.84[a]	0.20
		(0.17)	(0.16)	(0.49)		(0.14)	(0.14)	(0.48)
Market access (export)			-0.26	-0.44			-0.43	-0.64[b]
			(0.40)	(0.35)			(0.33)	(0.31)
Market access (import)			-1.35[c]	-1.20[c]			-0.53	-0.17
			(0.71)	(0.67)			(0.62)	(0.57)
log 1920 population				0.079				-0.013
				(0.32)				(0.27)
log 1950 population				-0.058				-0.050
				(0.50)				(0.40)
log 2000 population				1.44[b]				0.77
				(0.59)				(0.53)
R^2	0.36	0.50	0.73	0.76	0.43	0.67	0.79	0.80

Notes: 65 observations per column. All regressions include a constant. Robust standard errors in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%.

Table 9: Second-step results, TSLS for importer fixed effects

Importer fixed effect	(1) weight	(2) weight	(3) weight	(4) weight	(5) value	(6) value	(7) value	(8) value
log highway lane km	0.86[a]	0.36	0.14	0.17	0.82[a]	0.020	-0.13	-0.10
	(0.23)	(0.25)	(0.18)	(0.16)	(0.18)	(0.21)	(0.16)	(0.15)
log employment		0.53[a]	0.67[a]	-0.83[c]		0.85[a]	0.95[a]	0.22
		(0.18)	(0.13)	(0.49)		(0.16)	(0.10)	(0.44)
Market access (export)			-1.56[a]	-1.69[a]			-1.05[a]	-1.24[a]
			(0.53)	(0.60)			(0.38)	(0.48)
Market access (import)			0.50	0.60			0.30	0.49
			(0.45)	(0.52)			(0.33)	(0.44)
log 1920 population				0.023				0.015
				(0.30)				(0.23)
log 1950 population				-0.18				-0.16
				(0.49)				(0.38)
log 2000 population				1.75[a]				0.89[c]
				(0.62)				(0.54)
Overid. p-value	0.75	0.65	0.39	0.51	0.84	0.67	0.64	0.73
First-stage Stat.	49.3	21.9	23.5	20.9	49.3	21.9	23.5	20.9

Notes: 65 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 lane kilometers of interstate highways. Robust standard errors in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%.

Table 10: Second-step results, robustness to instruments and instrumenting method

	(1) LIML	(2) LIML	(3) TSLS	(4) TSLS	(5) TSLS	(6) TSLS	(7) TSLS	(8) TSLS
log highway lane km	0.77[a]	0.53[b]	0.90[a]	0.52[b]	0.76[a]	0.56[b]	0.57[b]	0.48[b]
	(0.27)	(0.22)	(0.27)	(0.24)	(0.27)	(0.23)	(0.25)	(0.21)
log populations 1920, 1950, 2000	N	Y	N	Y	N	Y	N	Y
Instruments								
log 1528-1850 exploration routes index	Y	Y	Y	Y	N	N	Y	Y
log 1898 railroad km	Y	Y	Y	Y	Y	Y	N	N
log 1947 planned highway km	Y	Y	N	N	Y	Y	Y	Y
Overid. p-value	0.10	0.41	0.11	0.18	0.04	0.39	0.35	0.22
First-stage Stat.	23.3	21.0	16.1	22.2	29.5	27.2	29.1	27.8

Notes: 65 observations per column. The dependent variable is exporter fixed effect for weight in all columns. All regressions include a constant, log 2007 employment, and export market access as controls. Robust standard errors in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%.

Table 11: Results for one-step estimations

Bilateral trade flows	(1) weight	(2) weight	(3) weight	(4) weight	(5) value	(6) value	(7) value	(8) value
Panel A. OLS estimations.								
log highway lane km	0.92[a]	0.29	0.27[c]	0.21	1.09[a]	0.086	0.072	0.028
for exporter	(0.15)	(0.19)	(0.16)	(0.14)	(0.13)	(0.13)	(0.12)	(0.12)
log highway lane km	0.79[a]	0.21	0.17	0.13	0.84[a]	0.019	0.018	0.0072
for importer	(0.21)	(0.16)	(0.11)	(0.11)	(0.14)	(0.11)	(0.086)	(0.088)
log employment		0.64[a]	0.69[a]	0.47		1.04[a]	1.06[a]	1.46[a]
for exporter		(0.13)	(0.11)	(0.36)		(0.093)	(0.084)	(0.44)
log employment		0.57[a]	0.64[a]	-1.03[c]		0.82[a]	0.85[a]	0.082
for importer		(0.19)	(0.15)	(0.61)		(0.13)	(0.11)	(0.51)
Market access			-0.63[a]	-1.13[a]			-0.33[b]	-0.68[a]
for exporter			(0.20)	(0.22)			(0.16)	(0.22)
Market access			-1.24[a]	-1.33[a]			-0.78[a]	-0.81[a]
for importer			(0.22)	(0.22)			(0.15)	(0.15)
log populations 20, 50, 00	N	N	N	Y	N	N	N	Y
R^2	0.63	0.67	0.72	0.74	0.55	0.70	0.72	0.74
Panel B. TSLS estimations.								
log highway lane km	1.10[a]	0.79[a]	0.60[a]	0.47[a]	1.10[a]	0.28	0.18	0.069
for exporter	(0.15)	(0.24)	(0.19)	(0.17)	(0.15)	(0.19)	(0.17)	(0.14)
log highway lane km	0.87[a]	0.53[b]	0.23	0.17	0.81[a]	0.058	-0.087	-0.10
for importer	(0.23)	(0.25)	(0.17)	(0.15)	(0.16)	(0.18)	(0.14)	(0.14)
log employment		0.33[b]	0.49[a]	0.43		0.92[a]	1.00[a]	1.44[a]
for exporter		(0.16)	(0.13)	(0.40)		(0.13)	(0.11)	(0.44)
log employment		0.35[c]	0.60[a]	-1.04[c]		0.80[a]	0.92[a]	0.14
for importer		(0.20)	(0.15)	(0.60)		(0.14)	(0.11)	(0.49)
Market access			-0.62[a]	-1.10[a]			-0.33[b]	-0.68[a]
for exporter			(0.20)	(0.22)			(0.16)	(0.22)
Market access			-1.23[a]	-1.33[a]			-0.78[a]	-0.82[a]
for importer			(0.22)	(0.22)			(0.15)	(0.15)
log populations 20, 50, 00	N	N	N	Y	N	N	N	Y
Overid. p-value	0.61	0.49	0.28	0.66	.	.	0.14	.
First-stage Stat.	36.0	18.3	19.6	21.0	27.3	19.2	22.0	20.2

Notes: 2,639 observation for columns 1 to 4 and 3,299 for columns 5 to 8. All regressions include a constant and quartic function of 2005 highway distance between i and j. Columns 4 and 8 include log population in 1920, 1950, and 2000 for both exporters and importers. In panel B, we use (same city) log 1947 planned highway km, log 1898 railroad km, and log 1528-1850 exploration routes index as instruments for log lane kilometers of interstate highways for both the importer and exporter cities. Robust standard errors in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%. In some cases, the overidentification statistics is not produced because the estimated covariance matrix of moment conditions is not of full rank due to an insufficient number of clusters using the procedure of Cameron *et al.* (2010).

Table 12: Second-step results, robustness to definition of exporter effect and market access with TSLS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log highway lane km	0.83^a	0.64^a	0.83^a	0.62^a	0.72^a	0.51^b	0.69^a	0.40^b
	(0.26)	(0.23)	(0.26)	(0.22)	(0.25)	(0.21)	(0.24)	(0.20)
log populations 20, 50, 00	N	Y	N	Y	N	Y	N	Y
Overid. p-value	0.13	0.46	0.14	0.47	0.10	0.42	0.10	0.56
First-stage Stat.	22.8	21.0	23.5	21.5	23.3	21.0	23.4	21.1

Notes: The dependent variable is exporter fixed effect for tons in all columns. 65 observations per column. All regressions include a constant, log 2007 employment, and export market access as controls. All regressions use log 1947 planned highway km, log 1898 railroad km, and log 1528-1850 exploration routes index as instruments for log lane kilometers of interstate highways. Columns 1 and 2 use, as dependent variable, exporter fixed effects estimated with an OLS specification linear in log highway distance in the first step. Columns 3 and 4 use exporter fixed effects estimated with an OLS specification linear in log Euclidian distance in the first step. Columns 5 and 6 use exporter fixed effects estimated with an OLS specification using a quartic function of log highway distance in the first step. Columns 7 and 8 use a "ad-hoc" market access variable computed using city level incomes. Robust standard errors in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%.