ABSTRACT

Title of Document: Infrastructure Investment and Economic Development

Janet Xiaohui Hao, Ph.D., 2007

Directed By: Professor John Shea, Department of Economics

State and local governments invested 13% of their revenues on infrastructure in 2002, but existing literature provides only mixed evidence that infrastructure contributes to economic activity. To estimate the effect of infrastructure on the economy, Chapter 1 analyzes how the construction of the Interstate Highway System (IHS) contributed to regional development in the United States by expanding intercity trade, using data on the construction of the IHS, intercity trade and regional economic activities. Empirical results provide evidence that the IHS reduced driving times among cities and subsequently increased inter-city trade in the following two ways. First, it increased the volume of trade among existing trading partners. Second, it increased the probability of trade among cities that previously did not trade. Moreover, trade expanded more for nationally traded goods more than for locally traded goods, because the former relied more on the IHS than the latter. By expanding trade, the IHS increased regional output, employment and firm entry.

Existing literature provides mixed evidence on whether infrastructure contributes to economic growth, because of a problem of reverse causality—better infrastructure may not lead to higher growth, but regions with higher growth may invest more in

infrastructure. Chapter 2 uses an instrumental variable to identify the impact of the construction of infrastructure, focusing on the construction of the Interstate Highway System (IHS). A close link exists between the 1956 Interstate Highway (IH) plan and pre-existing economic prosperity. But both historical evidence and regression analysis show that construction priority among highway segments partly depended on how easy it was to build those segments; and ease of construction is plausibly exogenous to economic growth. We instrument the open-to-traffic time for each segment of highway using construction costs from the 1958 Interstate Cost Estimates. We apply the instrument to estimate the effect of new IH on driving times, and find that an OLS regression under-estimates the effect of IH on driving times.

INFRASTRUCTURE INVESTMENT AND ECONOMIC DEVELOPMENT

By

Janet Xiaohui Hao

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2007

Advisory Committee:
Professor John Shea, Chair
Professor John Haltiwanger
Professor Mark Duggan
Professor John McLaren
Professor John Horowitz

Acknowledgements

I thank Dr. John Shea, Dr. John Haltiwanger, Dr. Mark Duggan, Dr. Nuno Limão, Dr. John McLaren, Dr. Allan Drazen, Dr. Michael Pries, Dr. Carlos Vegh, Dr. Peter Murrell, Dr. Oleg Smirnov, Dr. David Luskin, Dr. Hamid Mohtadi, Perry Singleton, Ethan Ilzetzki, Ye Zhang and Drake Jenkins for insightful comments and suggestions. I also thank Richard Weingroff, Eric Gabler and Martin Weiss at the Department of Transportation for providing data on the Interstate Highway System and giving good suggestions.

Table of Contents

List of Tables

List of Figures

1. Infrastructure Investment, Trade Expansion and Regional Development: Evidence from the Interstate Highway System

1.1 Introduction

U.S. state and local governments invested $275 billion in infrastructure in 2002[1]. This investment accounted for 13 percent of the total expenditure of state and local governments. Despite the huge expenditure, a large literature provides only mixed evidence that infrastructure contributes to economic prosperity (Aschauer, 1988, 1989, 1990; Aaron, 1991; Schultze, 1990; Hulten and Schwab, 1991; Jorgenson, 1991). This evidence is mixed because of three major difficulties. First, measuring the contribution of infrastructure is hard. Because infrastructure is not traded in the market place, we usually measure it by government spending, not its economic contribution. Moreover, measurement is often at an aggregated level (state or country) with limited variation. Second, because economists cannot observe the contribution of infrastructure, how infrastructure affects the economy remains a black box. Existing literature does not illustrate the mechanism by which infrastructure affects the economy. Third, the direction of causality is unclear. For example, rich regions tend to invest more in infrastructure than poor regions do.

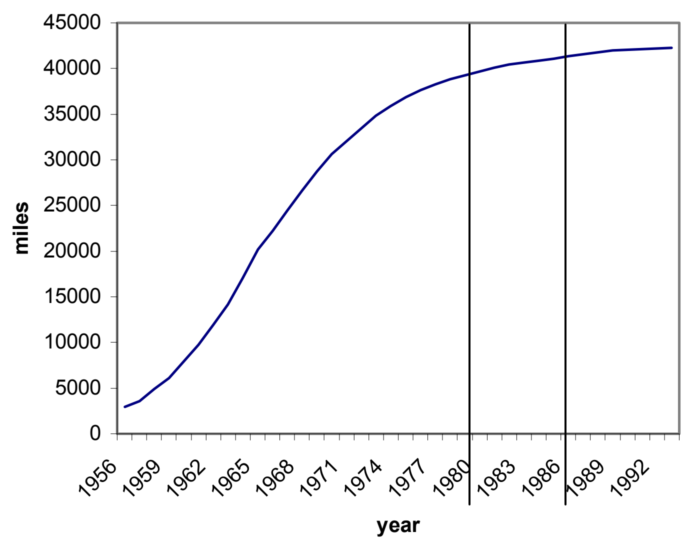
[1] We follow the definition of infrastructure in Duffy-Deno and Eberts (1991). "Public capital includes: (a) sanitary and storm sewers and sewage disposal facilities, (b) roadways, sidewalks, bridges and tunnels, (c) water supply distribution system, (d) public hospitals, and (e) public service enterprises such as airports and ports." Source: US Census Bureau, 2005. State and Local Government Finances by Level of Government and by State: 2001 - 02. Available at http://www.census.gov/govs/estimate/0200ussl_1.html

The Interstate Highway System (IHS) offers a good opportunity for exploring these issues. First, the construction of IHS presents rich cross-sectional variation in infrastructure investment. Because the timing of construction was different among different city pairs, intercity driving times dropped by different magnitudes at different times, allowing us to exploit cross-sectional variation. Second, direction of causality is easier to identify. Historical evidence implies that the timing of construction was partly determined by engineering costs. Third, examining the effect of Interstate Highways on intercity trade opens the black box, since we directly measure the contribution of infrastructure investment (for example, reduction in driving times).

This chapter explores the economic effect of infrastructure, focusing on how the construction of the IHS contributed to regional development in the United States by expanding intercity trade. We ask how new interstate highways (IH) affected trade in goods of two-digit manufacturing industries among major cities between 1967 and 1972, when the IHS was expanding rapidly[2]. The construction of the IHS started in 1956 and was mostly completed by 1972. However, the years 1967-1972 were a period of particularly great expansion in the IHS (Figure 1), with about a quarter of the whole system (11,254 miles) opening to traffic. This, together with an increase in driving speeds, decreased average driving times among the sample cities by 11 percent or 3.28 hours.

[2] We have data on commodity flows in 1963, 1967 and 1972. We focus on trade growth between 1967 and 1972, controlling for the trend of growth from 1963 to 1967.

Figure 1: IHS Mileage Open to Traffic



Source: unpublished Interstate Highway Construction Records (1956—1993), issued by Federal Highway Administration.
Notes: The mileage in 1956 includes toll roads and national high priority corridors existing before 1956 and incorporated into the IHS in the 1956 Federal-Aid Highway Act.

Intercity trade expanded accordingly among cities connected with new IH. Table 1 gives a sense of the extent to which such trade expansion was generated. When approximately 267 miles of new highways opened to traffic between St. Paul and Los Angeles from 1968 to 1973, driving time decreased by 9 hours and 15 minutes (19.3 percent). Shipments of “rubber and miscellaneous plastics products” from St. Paul to Los Angeles rose by 950 percent. In contrast, driving time between St. Paul and Chicago decreased by only 40 minutes (8.3 percent) from 1968 to 1972, and the shipments of similar goods from St. Paul to Chicago rose by only 29 percent. Most of highways I-94 and I-90, connecting St. Paul and Chicago, were already open to traffic by 1968.

Table 1: New Interstate Highways, Reduction in Driving Time and Growth of Trade (1968-1973)

2-digit Industry	Origin	Destination	New miles of Interstate Highways	Reduction in Driving time (hours: minutes & % change)	Growth Rate of Trade[1]
Rubber and misc. plastic products	St. Paul	Los Angeles	I-70 (200 miles) I-80 (67 miles)	9:15 (19.3%)	950%
		Chicago	I-90 (25 miles) I-94 (14 miles)	0:40 (8.3%)	29%
Apparel	Boston	San Francisco	I-80 (327 miles)	10.35 (14.0%)	960%
		Detroit	None	0:15 (1.4%)[2]	39%
Primary metal products	Cleveland	Houston	I-55 (92 miles) I-65 (52 miles) I-71 (43 miles)	3:45 (12.5%)	-0.5%
		St. Paul	I-90 (36 miles) I-94 (14 miles)	0:50 (5.7%)	-82%
Stone, Clay and Glass Products	Pittsburgh	St. Louis	I-70 (190 miles)	2:50 (20.9%)	-18%
		Syracuse	I-79 (70 miles)	0:50 (11.6%)	-45%

Sources: Commodity Transportation Surveys (1963, 1967, 1972), issued by the Bureau of Transportation Statistics, Rand McNally Road Atlas (1968, 1973, 2006), provided by Rand McNally, and unpublished Interstate Highway Construction Records (1956—1993), issued by Federal Highway Administration.
Notes: 1. Trade is outbound shipments (tons) from city of origin to city of destination.
2. Driving time dropped between Boston and Detroit because the average driving speed increased by 3 miles per hour from 1968 to 1972.

There are four potential endogeneity problems in establishing a relationship between driving time and trade growth. First, rich regions build more roads around them. We use disaggregate data to control for endogeneity at the aggregate city level. Industry-level commodity flows and driving times between pairs of cities allow us to control for time-varying factors specific to cities and to industries, such as the output expansion of a region or of an industry. For example, if Chicago expands its output, it expands trade with both Detroit and St. Paul. We examine the difference in the growth rates of trade between those two city pairs and we assume that the difference is exogenous to the output expansion in Chicago. Second, highway authorities may plan routes only between cities with the greatest potential for trade. City-pair fixed effects control for time-invariant characteristics specific to any particular pair of cities, such as the linkage between highway planning and the potential of trade. Third, trade growth may cause congestion and increase driving times. Then we would under-estimate the effect of driving time on trade. Fourth, among cities with planned highways, state governments may build roads earlier for city pairs with the greatest potential for trade. If that is true, we would over-estimate the effect of IHS on trade expansion. Our key assumption is that the timing of construction was exogenous to trade potential. Historical evidence suggests that the timing of construction was partly determined by engineering costs.

We explore the effects of the IHS on the growth of trade on both the intensive and extensive margin. The intensive margin of trade reflects an increase or a decrease in trade volume between existing trade partners, while the extensive margin is the creation or destruction of a trade partnership. To examine the intensive margin, we include only observations with positive trade volumes, regressing (log) change in intercity shipments of 2-digit industries on the (log) change in driving times among 16 major cities from 1967 to 1972. We examine the extensive margin through a linear probability regression where the dependent variable denotes whether trade in a particular 2-digit industry exists between a particular pair of cities.

We find that the 11 percent decline in driving times over the period of study enhanced intercity trade between existing trading partners by 14.0 percent and raised the probability of trade among non-trading areas by 1.32 percentage points. Since new IH are estimated to be responsible for a 3.34 percent decrease in average driving time over our sample period, the IHS increased intercity trade by 4.24 percent and the probability of trade by 0.40 percentage points.

Further, trade in nationally traded goods grew more in response to reduced driving times than that of regionally traded goods, because the former relied more on the IHS than the latter. We define locally traded goods as those dominated by shipments within 200 miles. This result suggests that the benefit of highways to intercity trade in a particular city depends on the tradability of goods produced in that city. Since some cities had already engaged in intra-national trade and thus had produced nationally traded goods, we would expect them to have benefited more from the IHS than other cities.

Finally, we estimate the contribution of improved transportation to the output, employment, investment and firm entry of 2-digit industries in sample cities, and find that industries relying heavily on the IHS responded disproportionately to the IHS, supporting the notion that the IHS caused economic development, not the other way around. For example, if the construction of the IHS was endogeneous---because the regional economy expanded, the government decided to build more roads---we would not observe a relationship between the growth rate of a regional industry and its reliance on the IHS.

The rest of the chapter has the following structure. Section 2 introduces related literature. Section 3 estimates how IH reduced driving times and trade costs. Section 4 derives a gravity equation from a monopolistic competition model. The gravity equation relates bilateral trade to the size of outputs in both regions, bilateral trade costs (in this chapter, driving times) and trade diversion. We focus on the expansion of trade among existing trade partners. Section 5 examines the impact of reduced driving times on the probability of trade among regions that were not previously trading with each other. Section 6 estimates how an improved transportation system contributes to regional growth of output, employment, investment and firm entry. Section 7 concludes.

1.2 Related Literature

The role of infrastructure investment has been studied extensively. However, empirical research on the effect of infrastructure on economic activity has tended to use the value of regional investment as the measure of infrastructure, forcing researchers to regress infrastructure against aggregate regional variables. Since aggregate regional variables, such as regional output, partly determine the amount of infrastructure investment, reverse causality makes evaluating infrastructure projects very difficult. Whether infrastructure contributes to economic prosperity at all remains an open question. Aschauer (1990, 1989, 1988) attributes 60 percent of the slowdown of the growth rate of multifactor productivity in the 1970s to the slowdown of infrastructure investment, especially because the government completed most of the Interstate Highway System by 1970. But Aschauer's striking result may have been driven by the reverse causality noted above (Aaron, 1991; Schultze, 1990; Hulten and Schwab, 1991; Jorgenson, 1991). Hulten and Schwab (1991) compare the "snow belt" and the "sun belt", concluding that differences in private capital and labor, rather than infrastructure, account for the different productivities of those regions. Moreover, the value of investment is a measure of expenditure, not of contribution. Since researchers cannot directly measure the contribution of infrastructure, they cannot identify the mechanisms through which infrastructure affects the economy.

This chapter makes four contributions with respect to previous literature. First, we use a new measure of infrastructure, intercity driving times, which allows us to directly measure the contribution of an important infrastructure. Second, this new measure explores disaggregate city pair-industry commodity flows, controlling for some of the sources of reverse causality. Third, we focus on an important economic activity (intercity trade) that has been largely ignored by previous research on infrastructure. Fourth, we introduce three data sets that have rarely been used by economists.

How intercity trade responds to better infrastructure could mirror the response of international trade. Limão and Venables (2001) measure infrastructure using several variables (density of roads, density of rails, and density of telephone lines) and argue that, especially for landlocked countries, infrastructure largely determines transportation costs. They estimate that if a country's infrastructure were to deteriorate from the median to the 75th percentile, transportation costs would jump by 12 percent and trade volume would fall by 28 percent. This chapter finds that the IHS expands intercity trade, mirroring Limão and Venables' finding that better infrastructure promotes trade on an international scale.

This chapter is also related to a small body of literature on roads. Fernald (1999) applies a growth accounting procedure that separates productivity growth due to road investment from that due to capital or labor and finds that vehicle-intensive industries (transport services) benefit more than other sectors from public investment in roads.

This suggests that roads contribute to productivity. He estimates that road growth was responsible for 1.4 percent of annual TFP growth from 1953 through 1973. After 1973, however, most major commercial routes were so saturated with roads that road growth generated only 0.4 percent of TFP growth. Michaels (2007) finds that the trucking and retail industry expanded in rural counties containing IH, assuming that the construction of rural IH was exogenous to rural economic prosperity. Moreover, he finds that the IHS increased demand for skilled workers in counties with high human capital and decreased it in other counties.

While this chapter examines economic effects of roads, roads also affect other aspects of the society, such as migration and health. Baum-Snow (2007) instruments the number of IH through a city with the 1947 IHS plan and estimates that a new interstate highway through a central city decreases the population of that city by 18 percent. Ashenfelter and Greenstone (2004) estimates that the fatality rate will increase by 10 percent in response to a 1 percent drop in driving time because of higher speed limits.

This chapter focuses on roads. Other types of infrastructure, such as water supply and distribution, also significantly impact the economy. For instance, Duflo and Pande (2007) find that dams in India redistribute income from regions upstream of dams to regions downstream of dams. Moreover, dams in the aggregate increase poverty.

1.3 The Impact of Interstate Highways on Driving Time

Sections 3-5 test whether the IHS promoted trade by decreasing driving times. This section analyzes the empirical relationship between driving times and highway construction. The following two sections then explore the relationship between driving time and intercity trade. To estimate the effects of highway construction on driving times, we need to separate the effects of highway construction and travel velocity on driving times. The theoretical relationship between driving time, velocity, and highway mileage is given by equation (1):

$$drive_{ijt} = \frac{d_{ijt} - I_{ijt}}{v_t} + \frac{I_{ijt}}{1.154 v_t} \tag{1}$$

where d_{ijt} is the mileage of the route between i and j with the shortest driving time, not the straight line distance calculated from the longitude and latitude of cities; I_{ijt} is the total interstate highway mileage on this route; and v_t is the travel velocity on non-interstate highways in year t. The driving time between i and j is the sum of the time on non-interstate highways, $\frac{d_{ijt} - I_{ijt}}{v_t}$, and that on IH, $\frac{I_{ijt}}{1.154 v_t}$. The speed on IH is $1.154 v_t$, because according to Highway Statistics 1973, travel velocity on IH during this period was on average 15.4 percent higher than on non-interstate highways. Therefore, driving time is a function of velocity, distance and the proportion of IH en route.

To separate the effects of highway construction and velocity on driving times, we rewrite equation (1), taking logs of both sides.

$$drive_{ijt} = \frac{(d_{ijt} - I_{ijt})1.154 + I_{ijt}}{1.154 v_t} = \frac{1 + (\frac{1}{1.154} - 1) I_{ijt} / d_{ijt}}{v_t / d_{ijt}} \tag{2}$$

$$\ln drive_{ijt} = -\ln v_t + \ln d_{ijt} + \ln(1 - 0.133 \frac{I_{ijt}}{d_{ijt}}) \tag{3}$$

where $(\frac{1}{1.154} - 1)$ equals -0.133. We then first difference this equation over time for a given city pair ij, for two reasons. First, equation (3) fails to separate the variation of driving times among and within pairs of cities. Since regressions in the rest of the chapter focus on the effect of driving times on trade within pairs of cities, to be consistent, we estimate the effect of IH on driving times within city pairs in this Section. Second, this equation fails to capture heterogeneous driving speeds across the nation. For example, the driving speed in Nevada is on average 62.2 miles per hour during our sample period, while that in Connecticut is 50.2 per hour (Highway Statistics, 1973). First differencing the equation controls for heterogeneous driving speeds on different routes. The first differenced equation is

$$\ln(\frac{drive_{ij,73}}{drive_{ij,68}}) = \alpha + \beta \ln(\frac{d_{ij,73}}{d_{ij,68}}) + \gamma \ln[\frac{1 - 0.133(I_{ij,73} / d_{ij,73})}{1 - 0.133(I_{ij,68} / d_{ij,68})}] + \varepsilon_{ijt} \tag{4}$$

Driving time decreases if the distance drops or if IH open. α captures aggregate variables that may change driving times, such as an increase in speed at the national level. We would expect γ to be positive, because driving time should decrease with the ratio of IH to total mileage (I_{ijt} / d_{ijt}).

The period of time we are interested in is between 1967 and 1972, when the IHS was expanding rapidly (Figure 1) and when the Department of Transportation collected disaggregate commodity flow data. Details on commodity flow data are in Section 4.

This regression uses data on driving times and the mileage of highways open to traffic. Since the data on driving times was not available for 1967 and 1972, we instead use data for 1968 and 1973 from the Rand McNally Road Atlas, which estimates driving times under normal condition during the daytime, considering topography, road speed limits, and congestion (Rand McNally Road Atlas, 1962)[3]. For example, the published driving time between Detroit and Pittsburgh increased by 30 minutes due to congestion between 1968 and 1973, despite constant distances and road networks. The Rand McNally data does not specify time of the day at which driving times are measured. Driving times measured at 1 am are likely to be different from those at 9 am, but as long as they respond to new IH in the same way, the time of measurement should not systematically affect the estimated coefficient.

[3] Rand McNally Road Atlas (1962) says that "Driving time shown is approximate under normal conditions. Consideration has been given to topography, speed laws, and congested areas. Allowances should be made for night driving and unusually fast or slow drivers."

Data on the construction of highways was obtained from the Interstate Highway Construction Records, provided by the Federal Highway Administration. This source lists the construction and operation status of each segment of the IHS. Highway mileage between cities and the fraction of IH on each route were calculated based on the dates on which highway segments were opened to traffic[4].

We focus on the 16 cities and thus the 120 city pairs available in the commodity flow data (Table 2)[5]. Driving times among the 16 areas decreased by 11.0 percent, or 3.28 hours, on average. Driving times decreased between 118 city pairs, but changes in congestion caused driving times to increase between Pittsburgh and Detroit and between Milwaukee and Minneapolis. The percentage reduction in driving time was the largest along routes within the non-northeastern area at 13.3 percent, compared with 11.3 percent within the northeastern area. Routes between the northeastern and non-northeastern areas had the lowest percentage drop at 9.3 percent. In terms of reduction in hours, routes within the non-northeastern area had the largest reduction at 3.97 hours, followed by routes between the northeastern and non-northeastern areas at 3.21 hours. Routes within the northeastern area had the smallest reduction at 0.73 hours.

---

[4] We use the 2006 Rand McNally Road Atlas to find the milepost of each city in order to merge the data on routes with the data on construction status. We cannot calculate the mileposts of cities using the 1968 and 1973 editions, because they have no information on the length of each segment of highway. Since the milepost of a city on the IHS stays almost the same from 1956 to 2006, using the 2006 edition gives an accurate estimate of mileposts. Please see the Appendix for more information on how driving times were constructed.

[5] The 1972 CTS has data on 27 cities, the 1967 CTS has 25 cities, and the 1963 CTS has 18 cities. Because 18 cities are available in all three CTSs, and because the Rand McNally Road Atlases from 1968 and 1973 provide the driving times among 16 cities out of these 18, we have 16 cities in our final data set.

Table 2: 16 Sample Production Areas

Production Areas	Corresponding Standard Metropolitan Statistical Areas (SMSA)
1	Boston, Mass.
	Worcester, Mass.
	Providence-Pawtucket-Warwick, R.I.-Mass.
	Brockton, Mass.
	Lawrence-Haverhill, Mass.-N.H.
	Lowell, Mass.
3	New York, N. Y.
5	Philadelphia, Pa.-N.J.
	Wilmington, Del.-N.J.-Md.
	Trenton, N.J.
9	Syracuse, N.Y.
	Utica-Rome, N.Y.
	Albany-Schenectady-Troy, N.Y.
10	Buffalo, N.Y.
	Rochester, N.Y.
11	Cleveland, Ohio.
	Akron, Ohio.
	Canton, Ohio.
	Lorain-Elyria, Ohio.
	Youngstown-Warren, Ohio.
	Erie, Pa.
12	Pittsburgh, Pa.
	Steubenville-Weirton, Ohio-W. Va.
	Wheeling, W. Va.-Ohio.
13	Detroit, Mich.
	Toledo, Ohio-Mich.
	Ann Arbor, Mich.
14	Cincinnati, Ohio-Ky.-Ind.
	Dayton, Ohio.
	Hamilton-Middletown, Ohio.
	Springfield, Ohio
15	Chicago, Ill.
	Gary-Hammond-East Chicago, Ind.
16	Milwaukee, Wis.
	Kenosha, Wis.
	Racine, Wis.
17	Minneapolis-St. Paul, Minn.
18	St. Louis, Mo.-Ill.
21	Houston, Tex.
	Beaumont-Port Arthur, Tex.
	Galveston-Texas City, Tex.
24	San Francisco-Oakland, Calif.
	Vallejo-Napa, Calif.
	San Jose, Calif.
25	Los Angeles-Long Beach, Calif.
	Anaheim-Santa Ana-Garden Grove, Calif.
	San Bernardino-Riverside-Ontario, Calif.

Sources: Part 2 of Commodity Transportation Survey in Vol. III of Census of Transportation, provided by U.S. Department of Transportation, 1963, 1967 and 1972.

Driving time dropped between cities partly because new IH opened and partly because average speeds increased on both IH and non-interstate highways. The average proportion of IH along routes among the 16 cities increased from 64.8 percent to 79.5 percent, significantly increasing average driving speeds and reducing driving times. Most segments of IH connecting cities within the northeastern region opened to traffic before 1967, but many segments in non-northeastern region opened to traffic between 1968 and 1973 (Rand McNally Road Atlas, 1968 and 1973). Average speed increased on both IH and non-interstate highways. Since the average speed of trucks increased from 54 mph in 1968 to 57 mph in 1973 (Highway Statistics, 1973), we would expect driving times to drop[6]. Changes in distance were probably not important to driving times during this period. Many city pairs had unchanged mileage over this period. IH in some cases provided a shorter route between two locations. For example, the shortest route between Boston and Seattle fell from 3217 to 3102 miles during my sample period. On the other hand, because new IH had to pass through Indianapolis or Louisville to travel between St. Louis and Cincinnati, mileage on this route increased from 338 miles to 358 miles. The average distance among the 16 sample cities increased from 1087 to 1097 miles between 1968 and 1973.

Table 3 reports the effects of distance and highway construction on driving times. The first column lists the estimated coefficients of regression (4). The results suggest that the higher the fraction of IH on a route, the lower the driving time. This result is

[6] Federal regulation of speed limits did not exist before 1974, when all states reduced maximum speed limits to 55 mph.

significant at the 95 percent confidence level. (Log) change in highway mileage between cities does not appear to affect driving times. The second column lists the estimated coefficients using city dummies as additional control variables. We use city dummies to be consistent with regressions in the rest of the chapter, which also use city dummies. Moreover, city dummies capture changes in congestion or road speed specific to regions. Adding city dummies slightly decreases the value of estimated coefficients, but does not affect their significance. The second regression implies that the IHS reduced driving times by 3.34 percent between 1968 and 1973, multiplying the mean of the key explanatory variable in (4) by its estimated coefficient[7].

Table 3: Interstate Highways Reduced Driving Times among Cities

Dependent variable: (log) change in driving times among cities (1968-1973)[1]	(1)	(2)
(Log) change in highway mileage	-0.18	0.05
	(-1.34)	(0.23)
(Log) change in the weighted fraction of interstate highways on a route[2]	1.93	1.56
	(5.51)**	(4.47)**
City of Origin and Destination Dummies	N	Y
# obs	120	120
R-squared	0.20	0.80

Sources: Rand McNally Road Atlas (1968, 1973 and 2006), provided by Rand McNally, and unpublished Interstate Highway Construction Records, provided by the Federal Highway Administration.
Notes: 1. The dependent variable is the driving time between the central cities of the 16 sample production areas.
2. The weighted fraction of interstate highways is defined as (1-0.133*mileage covered by Interstate Highways/total road mileage). Therefore, a positive coefficient indicates that Interstate Highways decreased driving time. Please see text for the derivation of the regression equation.

[7] $1-\exp\{1.56\times\ln[\frac{1-0.133(I_{ij,73}/d_{ij,73})}{1-0.133(I_{ij,68}/d_{ij,68})}]\}=-3.34\%$

1.4 Interstate Highways and Trade Expansion

Trade may have expanded both because the IHS eased trade among existing partners and because it created new trading partners. This section develops a testable model that demonstrates the potential impact of the IHS on trade among existing partners.

1.4.1 Theoretical Framework

We develop a gravity equation from an exact model that allows a relationship between trade costs (specifically driving times) and intercity trade. The gravity equation is a standard framework to predict bilateral trade flows, stating that trade costs, outputs and purchasing powers of two countries determine the volume of trade. Its key assumption is that the specialization of production stays the same over time. A typical gravity equation takes the form of equation (5) below. This equation states that the magnitude of trade flows between country i and country j is a function of the geographic distance between the two, d_{ij}; the sizes of the two economies, measured in the form of output (y_{it} and y_{jt}); and additional variables, X_{ijt} (Deardoff 1984). ε_{ijt} is an error term.

$$\ln(trade_{ijt}) = b_0 + b_1 \ln(d_{ij}) + b_2 \ln(y_{it} y_{jt}) + b'_3 X_{ijt} + \varepsilon_{ijt} \qquad (5)$$

We set up a model of monopolistic competition, following Anderson and van Wincoop (2003). Our model has two major building blocks. First, goods are differentiated by production location. The economy consists of *I* regions. In each region, *K* different goods are produced. Goods of the same index produced in different regions are assumed to be distinct. For example, beer produced in New York and beer produced in Los Angeles are treated as different goods. Second, consumers' preferences are identical and characterized by a CES utility function over the $I \times K$ different goods. We now turn to the decision problems of consumers and producers, and then analyze the equilibrium interregional trade flows resulting from their choices.

1.4.1.1 Consumers' Problem at Location i

Consumer preferences are assumed to be identical both within and across regions. Consumers can therefore be treated as *I* representative consumers. Consumers in region j consume c_{ijk} units of goods k produced in region i to maximize

$$U_j = [\sum_{i=1}^{I}\sum_{k=1}^{K}(c_{ijk})^{\frac{\sigma-1}{\sigma}}]^{\frac{\sigma}{\sigma-1}} \tag{6}$$

subject to the constraint

$$\sum_{i=1}^{I}\sum_{k=1}^{K} p_{ik} t_{ij} t_k c_{ijk} = y_j \tag{7}$$

Here $\sigma > 1$ is the elasticity of substitution among goods, y_j is the nominal income of the consumer in region j, p_{ik} is the producer price of good k in region i, and t_{ij} and t_k are parameters governing trade costs. $p_{ik}t_{ij}t_k$ is the actual price consumers at location j pay to consume one unit of good k produced at location i. Consumers at different locations pay different prices for the same goods because of trade costs.

Trade costs include transportation costs, border-related barriers, information costs, legal costs, and wholesale and retail distribution costs (Anderson and van Wincoop, 2004). Anderson and van Wincoop (2004) estimate that trade costs are equivalent to a 170 percent ad-valorem tax for industrial countries. Transportation costs are equivalent to a 21 percent tax, border-related barriers are equivalent to a 44 percent tax, and distribution costs are equivalent to a 55 percent tax[8]. Since we examine intercity trade, we need not consider border-related costs. Since distribution costs are hard to measure, we use transportation costs as our measure of trade costs.

We apply the standard specification of iceberg transportation costs in trade literature. The transportation costs are equal to $t_{ij}t_k - 1$ units of good k for each unit of good shipped from i to j. t_{ij}>1 is the proportional iceberg transportation cost between regions i and j. t_k>1 is an additional transportation cost factor for industry k goods, allowing for the possibility that some goods incur higher transportation costs

[8] 100%+170%=121%*144%*155% (Anderson and van Wincoop, 2004) .

than others. Heavy goods, for example, may have a higher value of t_k than light goods. We assume that the producer in region i passes the costs on to the consumer[9]. Thus, the revenue that producer k in region i receives from sales to consumers in region j is $x_{ijk} = p_{ik} t_{ij} t_k c_{ijk}$, comprised of the value of goods consumed in region j, $p_{ik} c_{ijk}$, and the transportation costs that the producer passes on to the consumer, $p_{ik}(t_{ij} t_k - 1) c_{ijk}$.

The nominal demand for good k produced in region i by a consumer in region j, derived from the consumer's maximization problem, is

$$c_{ijk} = \frac{(p_{ik} t_{ij} t_k)^{-\sigma} y_j}{P_j^{1-\sigma}} \tag{8}$$

The CES utility function imposes that c_{ijk} is always positive. Here

$P_j = \left(\sum_{g=1}^{I} \sum_{k=1}^{K} (p_{gk} t_{gj} t_k)^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$, a price index of all goods consumed in location j.

This term formalizes trade diversion due to transportation costs (Anderson and van Wincoop, 2003). A decrease in transportation costs between j and a trading partner other than i causes the relative price of region i's goods in region j to increase. This tends to decrease imports from region i. For example, consider the trade from

[9] Depending on who bears the transportation costs, income may or may not include transportation costs. If the firm bears transportation costs and receives a lower price because of the costs, income should not include transportation costs. However, if the consumer bears the costs, income should include the costs.

Detroit to Chicago. If the government builds a new road from Chicago to St. Paul, Chicago may trade more with St. Paul and trade less with Detroit. Equation (8) suggests the demand in region j for good k produced in region i is affected by region j's income as well as by the consumer price of good k relative to the general price index in region j.

Finally, market clearing imposes that income at location i, y_i, is the sum of revenues that producers in region i receives from sales to all regions. Here y_i includes transportation costs.

$$y_i = \sum_{j=1}^{I}\sum_{k=1}^{K} p_{ik}t_{ij}t_k c_{ijk} = \sum_{j=1}^{I}\sum_{k=1}^{K}\left(\frac{p_{ik}t_{ij}t_k}{P_j}\right)^{1-\sigma} y_j \qquad (9)$$

1.4.1.2 Firms' Problem at Location j

We consider here the firms' decision problem, which helps determine goods' prices. Modeling production is necessary, because ignoring the firms' problem could cause us to ignore the simultaneity problem that regions with high levels of output are more likely to trade and also to build more roads. This endogeneity could lead us to over-estimate the effect of roads on trade expansion. We therefore consider the impact of transportation costs on firms' pricing and output decisions.

Firms are monopolists, producing with a production function that is linear in labor. We assume that labor is immobile[10]. No intermediate goods are required for production[11]. Therefore:

$$q_{ik} = \lambda_{ik} l_{ik} \tag{10}$$

where q_{ik} is the quantity of good k produced by a firm in region i, λ_{ik} is a technology parameter, and l_{ik} is labor input. Wages in region i, w_i, are the same across industries. Each firm takes w_i as given. The total quantity demanded of good k produced in location i is $q_{ik} = \sum_{j=1}^{I} t_{ij} t_k c_{ijk} = p_{ik}^{-\sigma} t_k^{1-\sigma} \sum_{j=1}^{I} (\frac{t_{ij}}{P_j})^{1-\sigma} y_j$. A firm chooses its optimal output, q_{ik}, to maximize

$$q_{ik} p_{ik} - \frac{w_i q_{ik}}{\lambda_{ik}} \tag{11}$$

subject to the constraint

$$q_{ik} p_{ik} = p_{ik}^{1-\sigma} t_k^{1-\sigma} \sum_{j=1}^{I} (\frac{t_{ij}}{P_j})^{1-\sigma} y_j \tag{12}$$

Optimal output is as follows:

[10] None of the implications of the model are sensitive to this assumption. At the opposite extreme, if labor were perfectly mobile, the model would yield a similar gravity equation.

[11] We conjecture that the model's results and the results of our empirical analysis would not significantly change with the addition of intermediate goods, because the trade of intermediate goods is determined by the output of downstream industries in a particular city, just as the trade of final goods is determined by the output in a particular city.

$$q_{ik} = \frac{(\frac{\sigma-1}{\sigma})^{\sigma} t_k^{1-\sigma} \lambda_{ik}^{\sigma} \sum_{j=1}^{I} (\frac{t_{ij}}{P_j})^{1-\sigma} y_j}{w_i^{\sigma}} \qquad (13)$$

Firms produce more as the transportation costs of their good decrease ($t_k \downarrow$); as their productivity increases ($\lambda_{ik} \uparrow$); as regional wages decrease ($w_i \downarrow$); and the more geographically proximate they are to wealthy regions ($\sum_{j=1}^{I} (\frac{t_{ij}}{P_j})^{1-\sigma} y_j \uparrow$). We could close the model by adding leisure to the consumer's utility function. The supply curve of labor would be upward sloping and the demand curve would be downward sloping. As the transport costs drop, firms would expand output, hire more workers and drive up wages.

1.4.1.3 Intercity Trade

We now combine the consumers' and firms' problems. Plugging equation (12) into equation (8) and taking logs produces the following gravity equation, which illustrates the determinants of intercity trade of good k between regions i and j.

$$\ln c_{ijk} = \ln q_{ik} + \ln y_j - \sigma \ln t_{ij} - \ln t_k + (\sigma - 1) \ln P_j - \ln \sum_{g=1}^{I} (\frac{t_{ig}}{P_g})^{1-\sigma} y_g \qquad (14)$$

Shipments of industry k from i to j obey the standard form of a gravity equation. Shipments are a function of the output of industry k in region i (q_{ik}), total output in region j (y_j), the trade barriers between i and j (t_{ij} and t_k), and trade diversion (P_j and $\sum_{g=1}^{I}(\frac{t_{ig}}{P_g})^{1-\sigma} y_g$). The first term of trade diversion implies that if consumers in region j face low prices from all the other regions, they buy less from region i. The second term implies that if regions other than j increase their income, they buy more from region i, and therefore region j will buy less from region i. The preferences of consumers imply that c_{ijk} is always positive, implying that cities will always trade the same set of goods.

We take partial derivatives of both sides of the equation to break down the effect of lower transportation costs $\ln t_{ij}$ on trade

$$\frac{\partial \ln c_{ijk}}{\partial \ln t_{ij}} = -\sigma + \frac{\partial \ln q_{ik}}{\partial \ln t_{ij}} + \frac{\partial \ln y_j}{\partial \ln t_{ij}} + (\sigma - 1)\frac{\partial \ln P_j}{\partial \ln t_{ij}} - \frac{\partial \ln \sum_{g=1}^{I}(\frac{t_{ig}}{P_g})^{1-\sigma} y_g}{\partial \ln t_{ij}} \qquad (15)$$

As t_{ij} shrinks, trade *increases* with an elasticity of σ (direct effect). Moreover, since output expands in both trading regions, trade volumes *increase* further (indirect effect). On the other hand, a drop in t_{ij} decreases "multilateral resistance", P_j , and expands market access, $\sum_{g=1}^{I}(\frac{t_{ig}}{P_g})^{1-\sigma} y_g$. The lower the "multilateral resistance", the more likely region j will trade with regions other than region i. The larger the market size, the higher the price, since firms are monopolists. Both these effects *decrease* the volume of trade (indirect effect). Our empirical analysis below controls for changes in q_{ik} as well as for factors specific to cities i and j. Thus we estimate only the direct effect.

1.4.2 Empirical Analysis

1.4.2.1 Estimation Strategy

The theoretical model of the previous section demonstrates that a reduction in transportation costs has direct and indirect effects on the expansion of intercity trade through various channels. This section proposes a regression to estimate the direct effect using the gravity equation (14) derived in the previous section. This equation can be rewritten as:

$$\begin{aligned}\ln c_{ijkt} &= \alpha_0 - \alpha_1 \ln drive_{ijt} + \alpha_2 \ln q_{ikt} + \alpha_3 \ln y_{jt} - \alpha_4 \ln t_{kt} \\ &+ \alpha_5 \ln P_{jt} - \alpha_6 \ln \sum_{g=1}^{I} \left(\frac{t_{igt}}{P_{gt}}\right)^{1-\sigma} y_{gt} + \alpha_7 f_{ij} + \phi_{ijkt}\end{aligned} \quad (16)$$

where:

$$\phi_{ijkt} = e_{it} + e_{jt} + e_{kt} + \varepsilon_{ijkt}$$

We assume that e_{it}, e_{jt}, e_{kt} and ε_{ijkt} are independent from each other and that:

$$\begin{aligned}&cov(\varepsilon_{ijkt}, \varepsilon_{ijht}) = u_{ij}^2 \\ &cov(\varepsilon_{ijkt}, \varepsilon_{mnht}) = 0, \text{ if } ij \neq mn \\ &cov(\varepsilon_{ijk,t}, \varepsilon_{ijk,t-1}) \neq 0\end{aligned}$$

In (16), $drive_{ijt}$ is the driving time between cities i and j in year t, which is our key explanatory variable and proxy for the transportation costs between the two cities. q_{ikt} is the output of industry k in city i. y_{jt} is the total output of all industries in city j. t_{kt}>1 is a time-varying index of the transportation costs of good k, which may depend for instance on weight per unit of value. P_{jt} is a price index of all goods for city j. $\sum_{g=1}^{I}(\frac{t_{igt}}{P_{gt}})^{1-\sigma}y_{gt}$ represents the market access of producers located in city i. Both P_{jt} and $\sum_{g=1}^{I}(\frac{t_{igt}}{P_{gt}})^{1-\sigma}y_{gt}$ are terms capturing trade diversion. Since the preferences of consumers imply that trade is always positive, we use only observations with positive trade volume over time. We consider observations with zero trade in Section 5.

The endogenous variable, c_{ijkt}, is measured in my data as the weight of intercity shipments by all modes of transportation. We use total shipments rather than road shipments alone because data is only available aggregated by all modes of transportation. We therefore measure only trade creation due to the introduction of the IHS, rather than trade diversion to highways from other means of transportation. Trade creation is the expansion of trade via all means of transportation. Trade diversion occurs if, for example, a certain good switches to being shipped via the IHS, rather than via railroad. As demonstrated in Table 4, the share of road shipments in total shipments increased by 8.9 percentage points over the sample period, while the shares of rail and water shipments decreased by 1.3 and 7.7

percentage points, respectively. Thus, there was a net diversion of trade to roads from other forms of transportation.

In our data we can observe $drive_{ijt}$, q_{ikt} and y_{jt}. However, we cannot observe t_{kt}, P_{jt}, or $\sum_{g=1}^{I}(\frac{t_{igt}}{P_{gt}})^{1-\sigma} y_{gt}$. We will use time-varying industry and city dummy variables to capture these unobserved terms in the empirical specification. Using region-specific dummies is a common method to control for unobserved variables in the international trade literature (Hummels, 2001; Roase and van Wincoop, 2001). Anderson and van Wincoop (2003) mention that region-specific dummies lead to consistent estimates of coefficients. In addition to regional-specific dummies, we also use city-pair fixed effects and industry dummies to control for a variety of econometric issues.

Table 4: U.S. Domestic Freight Shipments by Mode of Transportation

Mode	1967	1972
Road	39.6%	48.5%
Rail	32.6%	31.3%
Water	27.5%	19.8%
Other	0.3%	0.4%

Sources: Commodity Transportation Survey (CTS), 1967 and 1972.
Notes: 1. Figures are the percentage of shipments measured by weight, not by value.
2. Shipments by road are the sum of the shipments by motor vehicles and by private trucks.
3. The shipments by air are less than 0.1% for both years.
4. CTS 1972 provides information on 3-digit industries, not on 2-digit industries, so I aggregate 3-digit industries into 2-digit industries. Among the 116 3-digit TCC-code industries, 3 industries are missing in 1967, and 25 industries are missing in 1972. The remaining 88 industries are used to compute the figures of this table.

First, region-specific dummies control for the unobserved variables (P_{jt} , and $\sum_{g=1}^{I}(\frac{t_{igt}}{P_{gt}})^{1-\sigma} y_{gt}$) and other excluded time-varying factors specific to cities. Failure to control for such factors could cause inconsistent estimates of the coefficients on included variables. For instance, regions with high trade volumes may produce more, implying $cov(q_{ikt}, e_{it}) \neq 0$. Alternatively, wealthy regions may have more influence in their state governments; this pressure may cause highways to be built earlier in those regions, implying $cov(drive_{ijt}, e_{it}) \neq 0$ and $cov(drive_{ijt}, e_{jt}) \neq 0$. The city-specific dummies f_{it} and f_{jt} relate to the underlying terms in (16) as follows:

$$f_{it} = \alpha_6 \ln \sum_{g=1}^{I}(\frac{t_{igt}}{P_{gt}})^{1-\sigma} y_{jt} + e_{it}$$
$$f_{jt} = \alpha_2 \ln y_{jt} + \alpha_5 \ln P_{jt} + e_{jt}$$

Second, a city-pair fixed effect dummy, f_{ij}, controls for any time-invariant city-pair specific factors, such as potential correlation between highway planning and pre-existing economic linkages, that may affect highway growth.

Third, a time-varying industry dummy, f_{kt}, solves the problem that our data records the weight rather than the value of shipments, and also eliminates the need to measure the relative transportation costs of a given good, t_{kt}. We use the weight of shipments, since data is only available by weight. Our theory concerns the value of trade, but measuring shipments with weight rather than value will not affect our results, as the dependent variable can be written as:

log(weight of shipments) = log(value of shipments) - log(value per unit weight).

The time-varying industry dummy will soak up the term log(value per unit weight), as long as value per unit weight is constant across regions within a given time period.

Given the proposed empirical strategy, our available data, and proposed dummy variables, the implied regression equation for intercity trade is:

$$\ln c_{ijkt} = \beta_0 + \beta_1 \ln drive_{ijt} + \beta_2 \ln q_{ikt} + \beta_3 f_{kt} + \beta_4 f_{it} + \beta_5 f_{jt} + \beta_6 f_{ij} + \varepsilon_{ijkt} \tag{17}$$

We eliminate the terms involving f_{ij} by estimating (17) in first differences. The benchmark regression becomes the following:

$$\begin{aligned} \ln \frac{c_{ijk,72}}{c_{ijk,67}} &= \gamma_0 + \gamma_1 \ln \frac{q_{ik,72}}{q_{ik,67}} + \gamma_2 \ln \frac{drive_{ij,73}}{drive_{ij,68}} \\ &\quad + \gamma_3 dummy_i + \gamma_4 dummy_j + \gamma_5 dummy_k + \varsigma_{ijk} \end{aligned} \tag{18}$$

where we observe shipments data in 1967 and 1972 and driving time data in 1968 and 1973, and where $dummy_i$, $dummy_j$ and $dummy_k$ are dummies indicating region i, region j and industry k, respectively. Standard errors are clustered by city pairs. In

(18), γ_1 is the elasticity of trade growth with respect to the output of industry k at region i. The key coefficient, γ_2, represents the elasticity of trade growth with respect to the growth of driving time, capturing only the direct effect of driving time on trade and excluding the indirect growth in trade due to output expansion.

A remaining endogeneity problem is that the expectation of rapid trade growth between two cities may cause states to build IH between those two cities earlier. As we describe in Chapter 2 of this dissertation, both historical evidence and empirical analysis show that the priority of construction between 1967 and 1972 was determined in part by engineering cost considerations[12]. If an endogeneity problem exists, we may over-estimate the effect of new IH on trade growth, if states built new IH earlier on routes of large trade potential between 1967 and 1972.

1.4.2.2 Data and Summary Statistics

This analysis uses data on shipments, driving times, the output of origin and destination cities, and industry characteristics. Section 3 describes data on driving times in detail, so we focus here on the rest of the data. Data on shipments, output,

---

[12] The perception of a looming financial crisis in the Interstate Highway System in the 1960s motivated states to complete as many miles of IH as possible, causing them to build those segments that could be completed more rapidly. Historical data prove that construction costs partly determined the timing of opening to traffic. Our data sources are unpublished Interstate Highway Construction Records (1956-1993) and the 1958 Interstate Cost Estimates, both provided by the Federal Highway Administration. We regress opening to traffic time against construction cost estimates, and find that estimated costs have a negative and diminishing impact on the opening to traffic time (Chapter 2). The R-squared of this regression is only 0.01, implying that engineering costs are not the sole determinant of construction priority. Still, engineering costs may be a useful instrument for highway construction, as we explore in Chapter 2.
After the 1980s, political forces started to influence federal highway projects. Using data on 1998 Congressional votes over transportation projects, Knight (2004) estimates that powerful congressmen affected the allocation of federal highway funds to their districts. However, this source of endogeneity does not seem to have been present in my sample period.

and industry characteristics are all from the Department of Transportation's Commodity Transportation Survey (CTS) for 1963, 1967 and 1972.

Data on shipments by all modes of transportation is reported at the 2-digit (TCC code) manufacturing industry level among 25 production areas, each of which consists of one or more Standard Metropolitan Statistical Areas (SMSAs). The shipments data in CTS differs from data on international trade in three ways. First, trade is measured by weight rather than value. Second, regions do not need to have balanced trade. Overall inbound and outbound shipments for each region are highly asymmetric (Table 5), since CTS data records shipments by weight, excludes services and excludes non-sample cities and other regions. Third, CTS records the origin and destination, not the source and final user of shipments. For example, a producer may ship clothes from the factory in Cincinnati to a warehouse in Philadelphia, that then ships those clothes to retail shops. Since the destination may not be the location of final users, output of the destination city may predict the volume of trade poorly. Our regression do not have that problem, because we do not measure the total output in the city of destination. Instead, we use time-varying city dummies to control for output and any other city-specific variables.

Table 5: Outbound and Inbound Shipments by Areas, 1967

Area Code	Area	Outbound shipments (1000 tons)	Inbound shipments (1000 tons)
1	Boston	1,599,381	229,032
3	New York	2,285,385	176,779
5	Philadelphia	2,337,288	772,297
9	Syracuse	371,205	210,100
10	Buffalo	482,557	613,791
11	Cleveland	948,217	1,019,015
12	Pittsburgh	443,226	1,489,419
13	Detroit	1,146,571	1,058,969
14	Cincinnati	530,974	273,132
15	Chicago	1,208,246	1,392,362
16	Milwaukee	398,621	300,253
17	St. Paul	303,200	99,033
18	St. Louis	678,461	433,309
21	Houston	372,496	5,810,924
24	San Francisco	723,841	369,947
25	Los Angeles	1,095,630	676,935

Sources: Commodity Transportation Surveys (1963, 1967 and 1972), provided by the Bureau of Transportation Statistics.
Notes: 1. Each of the 16 areas consists of several Standard Metropolitan Statistical Areas (SMSAs). 7 out of 16 areas are in the northeastern region of US.
2. Outbound shipments are shipments of an area to the other 15 areas. Inbound shipments are similarly defined.

The CTS also records shipments of 2-digit manufacturing industries by distance of shipments and by mode of transportation. Table 6 lists this information for seventeen 2-digit manufacturing industries in 1967. We expect industries trading nationally to benefit more from the IHS than those trading locally. 71.5 percent of the shipments of Stone and Clay travel less than 200 miles, followed by Leather Products at 55.9 percent. In contrast, only 21 percent of the shipments of Electrical and Electronic Machinery travel less than 200 miles, followed by Instruments at 24.1 percent. Moreover, we expect industries relying heavily on roads to respond more to the IHS than other industries. In 1967, Leather Products had the highest share of shipments by road at 92.2 percent, followed by Textile Products at 90.7 percent. Petroleum Refining had the lowest share of shipments by road at 15.6 percent, followed by Paper at 42 percent.

Table 2 lists the production areas included in the sample. The sample includes 16 regions, which reflects the overlap between the availability of CTS data and driving time information in the Rand McNally Road Atlas[13]. There are 1568 city pair-industry combinations with positive trade volumes in 1963, 1967 and 1972; and 2940 combinations with positive trade in at least one of the three years. The cities included in the sample include six in the Northeast, seven in the Midwest, one in the South,

[13] The 1972 CTS has data on 27 cities, the 1967 CTS has 25 cities, and the 1963 CTS has 18 cities. Because 18 cities are available in all three CTSs, and because the Rand McNally Road Atlases from 1968 and 1973 provide the driving times among 16 cities out of these 18, we have 16 cities in our final data set.

Table 6: Outbound Shipments by Industry, 1967 (1000 tons)

TCC code	Industry	Shipments (1000 tons)	Shipments by road	Shipments within 200 miles
20	Food	1,063,248	50.6%	46.2%
22	Textile mill prod.	9,673	90.7%	45.7%
23	Apparel and related prod.	7,327	82.2%	34.1%
25	Furniture and fixtures	16,167	76.3%	29.6%
26	Paper and allied prod.	213,962	42.0%	34.9%
28	Chemicals	947,988	42.3%	46.1%
29	Petroleum refining	3,887,339	15.6%	30.7%
30	Rubber and misc. plastics	60,826	74.7%	34.4%
31	Leather & leather prod.	3,876	92.2%	55.9%
32	Stone, clay, etc.	431,664	63.3%	71.5%
33	Primary metal	1,395,904	45.0%	45.9%
34	Fabricated metal	298,282	71.4%	47.2%
35	Machinery, except electrical	126,310	69.2%	26.1%
36	Electrical and electronic machinery	86,139	65.2%	21.0%
37	Transport equipment	326,128	44.8%	33.0%
38	Instruments	3,451	76.6%	24.1%
39	Misc. manu. prod.	17,034	77.9%	29.0%
	Total	9,171,024		

Sources: Commodity Transportation Surveys (1967), provided by the Bureau of Transportation Statistics.
Notes: 1. Shipments are among 16 areas and of 17 industries in 1963, 1967 and 1972. Each area consists of several Standard Metropolitan Statistical Areas (SMSA). 7 out of 16 areas are in the northeastern region of US. The 17 industries are manufacturing industries at the 2-digit TCC level.
2. The percentage of shipments by road and the percentage of shipments within 200 miles are based on national aggregate data in Commodity Transportation Survey (1967).

and two in the West. Our sample therefore exhibits a broad regional coverage. Industrial composition differed across regions during our sample period, with some industries clustered in specific regions, providing the opportunities for regions to trade with each other. Petroleum refining, for example, was concentrated in Houston. Most of the primary metal industry was located in Detroit, Chicago, and Cleveland.

Table 7 presents the summary statistics of the data. The weight of intercity shipments decreased in our sample from 1967 to 1972. This, however, does not imply a decline in intercity trade, since at the national level the value of shipments per unit of weight increased between these years and since trade was diverted from among the sample regions to other regions in the country. The value of total domestic shipments in the United States of 89 3-digit industries grew by 23.6 percent (in constant 1967 dollars) over the sample period, while their weight increased by only 15.2 percent[14], implying an increase in the value per unit weight of goods. The ratio of shipments between sample regions and non-sample regions to total output in sample regions increased from 55.3 percent to 57.1 percent between 1967 and 1972, while the rate of trade within sample regions to output in sample regions dropped from 31.37 percent to 29.84 percent.

[14] We obtain the weight of shipments from the Commodity Transportation Survey (1967 and 1972), get the value of shipments from the Census of Manufactures (1967 and 1972), and deflate the latter with aggregate GDP deflators. The Census of Manufactures presents the value of output by industry. The Commodity Transportation Survey gives the weight of shipments, including the shipments from a place to itself. The Commodity Transportation Survey 1972 does not provide information on each 2-digit industry. Instead it gives the weight of shipments for some 3-digit industries. Therefore, we merge the data of 3-digit industries between the Census of Manufactures and the Commodity Transportation Survey, and calculate the growth rates of shipments in tons and in dollars, respectively.

Table 7: Summary Statistics

Definition	Obs	Mean	Std. Dev.	Min	Max
Growth rate of intercity driving time between areas j and j from 1967 to 1972	120	-11.0%	6.0%	-25.2%	9.1%
Within the northeastern area	15	-11.3%	4.4%	-20.7%	-4.5%
Within the non-northeastern area	45	-13.3%	5.7%	-23.5%	1.1%
Between the northeastern and non-northeaster area	60	-9.3%	6.0%	-25.2%	9.1%
Change of intercity driving time (hours) between areas j and j from 1967 to 1972.	120	-3.28	3.44	-11.9	0.5
Within the northeastern area	15	-0.73	0.38	-1.34	-0.17
Within the non-northeastern area	45	-3.97	3.53	-11.9	0.08
Between the northeastern and non-northeaster area	60	-3.21	3.54	-11.5	0.5
Growth rate of intercity shipment (tons) from area i to area j of industry k from 1963 to 1967.	2455	-12%	512%	-100%	17772%

Sources: Commodity Transportation Surveys (1963, 1967 and 1972), provided by the Bureau of Transportation Statistics, and Rand McNally Road Atlas (1962, 1968 and 1973), provided by Rand McNally.
Notes: 1. There are 120 observations of driving times among 16 areas ($120 = (15+1)\times 15/2$).
2. Shipments (tons) are among 16 areas and 17 2-digit manufacturing industries in 1963, 1967 and 1972. Each area consists of several Standard Metropolitan Statistic Areas (SMSAs).
3. –100% means intercity shipments (tons) from area i to area j of industry k dropped to zero in 1972.

1.4.2.3 The Impact of Driving Times on Trade

Having shown in Section 3 that the IHS decreased driving times, we now turn to explore how driving time affected intercity trade. Our benchmark regression tests the relationship between the (log) change in the weight of shipments and the (log) change in driving times. The change is measured between the years 1967 and 1972. Our controls include dummies for city of origin and destination, industry dummies, and change in total industry output in the origin city. Results of OLS regressions with standard errors clustered by city pairs are reported in Table 8.

Table 8: Estimated effects of driving time on trade growth

	OLS with standard errors clustered by city pairs				
Dependent Variable: (log) change in intercity trade (1967-1972)	Benchmark (1)	(2)	(3)	(4)	(5)
(Log) change in intercity driving time	-1.27 (-2.45)**	-4.91 (-3.75)**	-1.50 (-0.87)	-5.68 (-2.30)**	-1.33 (-2.60)***
(Log) change in driving time multiplied by the percentage within 200 miles		0.09 (3.15)**		0.10 (3.07)**	
(Log) change in time*roadshare			0.40 (0.13)	1.20 (0.40)	
(Log) change in the output of industry k in the origin area I	0.72 (12.54)**	0.73 (12.63)**	0.72 (12.55)**	0.73 (12.63)**	
Origin dummies	Y	Y	Y	Y	Y
Destination dummies	Y	Y	Y	Y	Y
Industry dummies	Y	Y	Y	Y	Y
Observations	1751	1751	1751	1568	1751
Clusters	120	120	120	120	120
R-squared	0.20	0.20	0.20	0.20	0.11

Sources: Commodity Transportation Surveys (1963, 1967 and 1972), provided by the Bureau of Transportation Statistics; Census of Manufactures (1967), issued by the Department of Commerce; and Rand McNally Road Atlas (1962, 1968 and 1973), provided by Rand McNally.

Notes: 1. The dependent variable is the (log) change in shipments (tons) from area i to area j of industry k from 1967 to 1972. Each area consists of several Standard Metropolitan Statistical Areas (SMSA). 7 of the 16 areas are in the northeastern region of US. The 17 industries are manufacturing industries at the 2-digit TCC level.

2. t-statistics are in parenthesis.

3. Standard errors are clustered by city pairs.

Column (1), the benchmark regression, implies that the effect of the (log) change in driving time on the (log) change in intercity trade is negative and significant at the 95 percent confidence level. The elasticity of trade growth with respect to driving time growth is -1.27. Our estimate implies that the sample average decrease of 11.0 percent in driving time expanded intercity trade by 14 percent. In the previous section, IH were estimated to have reduced driving time by 3.34 percent over our sample period. We can therefore attribute a 4.24 percent increase in intercity trade to interstate highway construction.

The result that a reduction in driving time leads to trade expansion is consistent with existing estimates from the international trade literature. Limão and Venables (2001) measure transportation costs of international trade using the CIF/FOB ratio[15], and estimate that the elasticity of trade with respect to transportation costs is -2.5. Martinez-Zarzoso and Suarez-Burguet (2005) also measure transportation costs with the CIF/FOB ratio, and find an elasticity of -2.3. Our estimated elasticity is smaller in absolute value than previous estimates, possibly because we use a different measure of transportation costs.

Columns (2)-(4) examine heterogeneity in the response of industries to the change in driving times. Fernald (1999) shows that industries with more vehicles experience larger

[15] The CIF price is the cost, insurance and freight price at the port of the importing country, excluding import duties and transportation costs within the importing country. (Source: http://stats.oecd.org/glossary/detail.asp?ID=332) The FOB price is the free on board price at the port of exporting country. The gap between CIF and FOB equals the insurance charges and transportation costs covering the travel between the importing country and the exporting country. (Source: http://forum.europa.eu.int/irc/dsis/coded/info/data/coded/en/gl008349.htm)

productivity growth when roads are expanding. Similarly, according to Shephard's lemma[16], trade in goods that rely heavily on IH should expand more in response to Interstate Highway construction. A priori, we would expect two kinds of industries to benefit most from the IHS: industries that ship goods long distances in a national market and industries that rely heavily on roads rather than air, train, or water transportation. The CTS provides us with the percentage of goods shipped by distance categories and mode of transportation for each industry in 1967.

Column (2) in Table 8 confirms that trade in long-distance industries responds more to the IHS: the interaction between driving time and the share of local shipments (under 200 miles) is positive and significant, and the baseline coefficient on driving time growth (representing the effect of driving time reduction on trade for a completely national industry) is negative and significant. These results remain statistically significant when cutoff distances of 100 miles or 500 are used[17].

Surprisingly, it does not appear that road-intensive industries are significantly more sensitive to the IHS than others. Column (3) shows that the coefficient on the interaction between road-intensity and the (log) change in driving time is insignificant. Column (4)

---

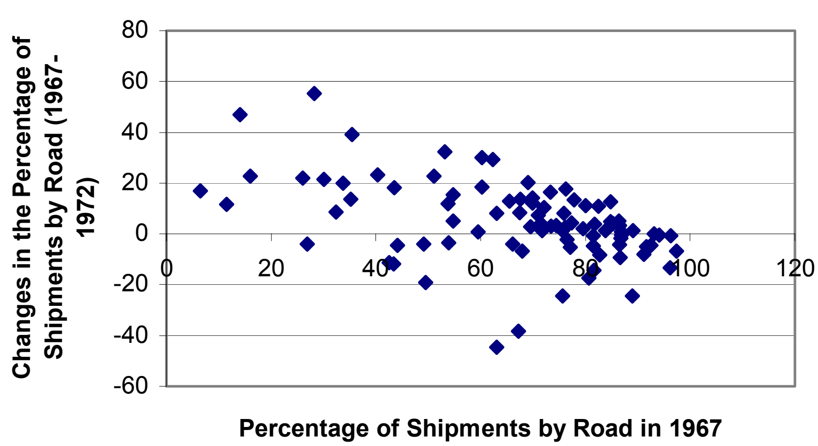
[16] Shephard's lemma states that the partial derivative of the expenditure function of production with respect to the price of a certain input is equal to the amount of that input used in production.
$\frac{\partial e(p^0, u^0)}{\partial p^i} = x_h^i(p^0, u^0), i = 1, \ldots, n.$ Therefore, if an industry uses IH as an important input, the production of that industry will react strongly to new IH. Source: Jehle, Geoffrey A. and Philip J. Reny, Advanced Microeconomic Theory.

[17] If 100 miles is the cutoff, the estimated coefficient on the (log) change in driving time is -3.90, and that on the interaction term is 0.11. Both coefficients are significant at the 99 percent confidence level. If 500 miles is the cutoff, the estimated coefficient on the (log) change in driving time is -8.79, and that on the interaction term is 0.11. Both coefficients are significant at the 99 percent confidence level.

includes both interactions described above. It shows that industries with a high share of shipments by road in 1967 respond less to lower driving times than other industries, although the interaction is not significant. It is possible that the share of shipments by road is a poor measure of likely sensitivity to the IHS, because industries with low road shares in 1967 may have had more opportunities to substitute towards roads in response to the IHS than those with high initial road shares. Indeed, the increase in road share from 1967 to 1972 is negatively correlated with the initial road share (Figure 2).

It is possible that the benchmark regression underestimates the impact of the IHS on intercity trade. First, it does not capture indirect effects of driving time on trade. The

Figure 2: Changes in the Percentage of Shipments by Road for 3-digit Industries



Sources: Commodity Transportation Survey (1963, 1967, 1972), provided by the Bureau of Transportation Statistics., and Census of Manufactures (1963), provided by the US Department of Commerce.
Notes: Shipments are in tons, not in dollars.

theoretical model suggests that when driving time decreases between region i and region j, producers at region i have better access to the domestic market and therefore expand production in response to higher demand. This increases trade with other regions and is not captured in the benchmark regression, since we control for output growth in the origin city-industry. Column (5) in Table 8 shows that if we do not control for output growth in the origin city-industry, the key coefficient increases from 1.27 to 1.33. Second, rate regulation on motor carriers during this time period may have prevented producers and consumers from fully realizing the benefit from interstates. The Department of Transportation regulated shipping rates on a ton-mile basis during this period (Chow, 1991). IH decreased driving time, but not necessarily distance. If motor carriers were not able to pass on their cost savings by decreasing shipping rates, they captured a large part of the benefits of the reduction in driving time in the form of higher profits. Other producers and consumers arguably did not realize the full benefit of interstates until after the deregulation of motor carriers in 1980.

1.4.2.4 Sensitivity Analysis

Tables 9 to 11 report the results of various sensitivity analyses. Columns (1) to (4) of Table 9 check the sensitivity of results to outliers. Coefficients are still significant after excluding 5 percent of observations at each end of the distribution of the dependent variable, or the 5 percent of observations at each end of the distribution of (log) change in driving time. The value of the estimated coefficient changes, however. When excluding possible outliers of the explanatory variable, the coefficient on driving time in the benchmark regression actually increases in absolute value from -1.27 to -1.57. After excluding the possible outliers of the dependent variable, however, the magnitude of the same coefficient drops from -1.27 to -0.66. Removing outliers from regressions including industry heterogeneity yields similar results. Column (5) imposes that the coefficient of $\ln\frac{q_{ik,72}}{q_{ik,67}}$ equals 1. The theoretic model implies that the elasticity of trade with respect to output equals 1. The key coefficient is still negative and significant.

Table 10 demonstrates that the benchmark results are robust to different specifications. Column (1) uses (log) change in output of cities of origin and destination (y_{it} and y_{jt}) instead of regional fixed effects. We conduct this analysis because regional fixed effects soak up changes in driving time that are region specific, leaving less variation in driving time. For example, a truck has to take I-15 to travel from Los Angeles to most of the other 15 production areas. Even though the timing of the construction of I-15 in Nevada and Utah was plausibly exogenous with respect to the economic growth of Los Angeles, the region fixed effects will soak up the effect of I-15. Data on regional output by industry (q_{ikt}) and on regional output (y_{jt}) are aggregated from bilateral shipments. CTS records shipments from a region to the 16 sample regions, to the region itself, and aggregate shipments to all the other regions in the United States, implying that total shipments are the output of industry k in region i in year t, q_{ikt}. The sum of the total output of all industries at region j is the regional output, $y_{jt} = \sum_{k=1}^{K} q_{ikt}$. Controlling for

Table 9: Sensitivity Analysis

Dependent Variable: (log) change in intercity trade (1967-1972)	OLS with standard errors clustered by city-city combinations				
	Excluding 5% of the obs at each end of the distribution of the dependent variable		Excluding 5% of the obs at each end of the distribution of the (log) change in driving times		Imposing the coefficient of (log) change in output to equal 1
	(1)	(2)	(3)	(4)	(5)
(Log) change in intercity driving time	-0.66 (-1.87)*	-3.27 (-3.44)**	-1.57 (-2.27)**	-5.99 (-3.56)**	-5.30 (-3.92)**
(Log) change in driving time multiplied by the percentage of shipments within 200 miles		0.07 (2.76)**		0.12 (3.01)**	0.10 (3.38)**
(Log) change in the output of industry k in the origin area I	0.48 (10.74)**	0.49 (10.82)**	0.72 (11.70)**	0.73 (11.66)**	
Origin dummies	Y	Y	Y	Y	Y
Destination dummies	Y	Y	Y	Y	Y
Industry dummies	Y	Y	Y	Y	Y
Observations	1595	1595	1587	1587	1751
Clusters	120	120	109	109	120
R-squared	0.15	0.15	0.20	0.20	0.05

Sources: Commodity Transportation Surveys (1963, 1967 and 1972), provided by the Bureau of Transportation Statistics; Census of Manufactures (1967), issued by the Department of Commerce; and Rand McNally Road Atlas (1962, 1968 and 1973), provided by Rand McNally.

Notes: 1. The dependent variable is the (log) change in shipments (tons) from area i to area j of industry k from 1967 to 1972. Each area consists of several Standard Metropolitan Statistical Areas (SMSA). 7 of the 16 areas are in the northeastern region of US. The 17 industries are manufacturing industries at the 2-digit TCC level.

2. t-statistics are in parenthesis.

Table 10: Specification Test

Dependent Variable	(1)	(2)	(3)	(4)
	(Log) change in trade (1967-72)	(Log) change in trade (1967-72)	Proxy of change in trade (1967-72)	Proxy of change in trade (1967-72) Obs with positive trade volume only
(Log) change in intercity driving time	-5.19 (-4.51)**			
(Log) change in driving time multiplied by the percentage of shipments within 200 miles	0.10 (3.35)**			
Reduction in intercity driving time (hours)		-0.08 (-2.92)**		
Reduction in driving time (hours) multiplied by the percentage of shipments within 200 miles		0.001 (1.86)*		
Proxy of change in driving time (ratio)			-2.56 (-2.30)**	-3.71 (-3.96)**
Proxy of change in driving time (ratio) multiplied by the percentage of shipments within 200 miles			0.05 (1.94)*	0.07 (3.25)**
(Log) change in the output of industry k in the origin area i	0.82 (14.28)**	0.72 (12.50)**		
Proxy of change in the output of industry in the origin area i			0.39 (6.28)**	0.62 (13.11)**
(Log) change in the output of all industries in origin city	0.02 (0.35)			
(Log) change in the output of all industries in destination city	-0.02 (-0.69)			
Origin dummies	N	Y	Y	Y
Destination dummies	N	Y	Y	Y
Industry dummies	Y	Y	Y	Y
Observations	1751	1751	1959	1751
Clusters	120	120	120	120
R-square	0.18	0.20	0.12	0.20

Sources: Commodity Transportation Surveys (1963, 1967 and 1972), provided by the Bureau of Transportation Statistics
Rand McNally Road Atlas (1962, 1968 and 1973), provided by Rand McNally
Notes: 1. The dependent variable is the (log) change in shipments (tons) from area i to area j of industry k from 1967 to 1972. Each area consists of several Standard Metropolitan Statistical Areas (SMSA). 7 of the 16 areas are in the northeastern region of US. The 17 industries are manufacturing industries at the 2-digit TCC level.
2. Standard errors are clustered by city pairs.
3. t-statistics are in parenthesis.

Table 11

Estimated effects of driving time on trade growth with lagged dependent variable

OLS with standard errors clustered by city-city combinations		
Dependent Variable: log change in intercity trade (1967-1972)	(1)	(2)
(Log) change in intercity driving time	-0.78 (-1.61)	-1.24 (-2.40)**
(Log) change in the output of industry k in the origin area i	0.64 (10.67)**	0.72 (12.54)**
Lagged dependent variable	-0.38 (-15.96)**	
Lagged growth rate of intercity driving time		0.33 (1.43)
Origin dummies	Y	Y
Destination dummies	Y	Y
Industry dummies	Y	Y
Observations	1568	1751
Clusters	120	120
R-squared	0.34	0.20

Sources: Commodity Transportation Surveys (1963, 1967 and 1972), provided by the Bureau of Transportation Statistics; Census of Manufactures (1967), issued by the Department of Commerce; and Rand McNally Road Atlas (1962, 1968 and 1973), provided by Rand McNally.

Notes: 1. The dependent variable is the (log) change in shipments (tons) from area i to area j of industry k from 1967 to 1972. Each area consists of several Standard Metropolitan Statistical Areas (SMSA). 7 of the 16 areas are in the northeastern region of US. The 17 industries are manufacturing industries at the 2-digit TCC level.

2. The (log) change in intercity driving time is between 1968 and 1973, while the lagged (log) change in intercity driving time is between 1962 and 1968.

3. t-statistics are in parenthesis.

(log) output change yields similar results as those in the benchmark regression.

Column (2) replaces our driving time ratio variable, $\ln(drive_{1973} / drive_{1968})$, with the difference in driving times $(drive_{1973} - drive_{1968})$. We do so because we do not know the form of trade costs, since driving time is only a proxy of trade costs. The (log) change of driving time treats a 10-hour decrease of a 50-hour ride equivalently to a 10-minute decrease of a 50-minute ride. Trying out different forms of change in driving times may provide more information on the effect of trade costs on trade expansion. Driving time reduction still has a negative and significant effect on intercity trade with this new specification.

Column (3) approximates the (log) change in variable x with $(x_{1972} - x_{1967})/[(x_{1972} + x_{1967})/2]$, where x can be intercity trade (c_{ijkt}), driving time between i and j ($drive_{ijt}$), or the output of industry k (q_{ikt}). Tornqvist, Vartia and Vartia (1985) suggest that $(x_{1972} - x_{1967})/[(x_{1972} + x_{1967})/2]$ is a valid proxy of (log) change. In this specification, the number of observations increases from 1751 to 1959, because both observations with positive trade in both years and those with zero trade in one year are included. Observations with positive trade in all years give information on trade expansion in response to a reduction in driving time, while observations with zero trade in some years reflect trade creation between new trade partners. The key estimated coefficient γ_2 is still negative and significant. Column (4) carries out the above regression using the same 1751 observations in the benchmark regression, including only trade expansion. The key coefficient remains negative and significant.

Table 11 adds lagged variables to examine the dynamics of trade expansion. Column (1) uses the (log) change in intercity trade between 1963 and 1967 as an additional explanatory variable to control for the possible endogeneity problem that states may build IH earlier between city pairs where trade expanded rapidly before 1967. This endogeneity problem predicts a positive coefficient on lagged trade growth. But the coefficient on lagged trade growth is negative and significant at the 99 percent confidence level, possibly because of measurement error in trade. The estimated coefficient on (log) change in driving time drops to 0.78 and its t-statistic drops to -1.61. These results imply that the dynamics of trade are complicated, and our static benchmark regression is not able to capture the dynamics.

Column (2) uses lagged driving times as an additional explanatory variable to examine the lagged effect of the IHS on trade expansion. We find a positive sign on lagged driving time growth, suggesting regions with good roads before 1967 had a smaller growth of intercity trade. These drops imply that the IHS may have a one-time permanent effect on trade expansion: only areas initially without good roads had rapid trade growth in response to the IHS.

Table 12 lists the results of Two-Stage Least Squares (2SLS) regressions. We instrument the change in driving time (hours) with the change in IH mileage open to traffic to address the endogeneity problem that rich regions may build more highways, but may also generate more traffic congestion, which would bias the OLS estimate of the impact of driving time on intercity trade towards zero. Note that this specification does not address the endogeneity problem that states may build IH earlier along routes with higher trade potential. We drop the (log) change in the output of 2-digit industries in a region, because that change is endogeneous to the dependent variable— (log) change in trade.

Table 12: Estimated effects of driving time on trade growth (2SLS)

Dependent Variable: (log) change in intercity trade (1967-1972)	OLS (1)	OLS (2)	2SLS (3)	2SLS (4)	2SLS (5)
Change in intercity driving time	-0.04 (-3.32)***	-0.09 (-3.20)***	-0.03 (-1.60)	-0.07 (-1.90)*	
Change in driving time multiplied by the percentage within 200 miles		0.001 (-2.10)**		0.001 (-1.38)	
(Log) change in intercity driving time					10.90 (0.50)
Origin dummies	Y	Y	Y	Y	Y
Destination dummies	Y	Y	Y	Y	Y
Industry dummies	Y	Y	Y	Y	Y
Observations	1751	1751	1751	1751	1751
Clusters	120	120	120	120	120
R-squared	0.11	0.11			

Sources: Commodity Transportation Surveys (1963, 1967 and 1972), provided by the Bureau of Transportation Statistics; Census of Manufactures (1967), issued by the Department of Commerce; and Rand McNally Road Atlas (1962, 1968 and 1973), provided by Rand McNally.
Notes: 1. The dependent variable is the (log) change in shipments (tons) from area i to area j of industry k from 1967 to 1972. Each area consists of several Standard Metropolitan Statistical Areas (SMSA). 7 of the 16 areas are in the northeastern region of US. The 17 industries are manufacturing industries at the 2-digit TCC level.
2. t-statistics are in parenthesis.

Surprisingly, Columns (1) and (3) show that the OLS regression over-estimates the effect of driving times on the change in intercity trade. Evaluated at the mean change in driving times (3.28 hours), a 1 hour decrease in driving time is estimated to expand trade by

10.34 percent using 2SLS and by 14.02 percent by OLS[18]. Columns (2) and (4) also imply that OLS over-estimates the effect of driving times on trade growth. Column (5) shows that instrumenting the (log) change in time with new IH mileage leads to an insignificant key coefficient. The reason may be that the length of new IH predicts changes in the level of driving times (hours), but not the percentage change in driving times. The R-squared of the first-stage regression of Column (3) is 0.88, while that of Column (5) is only 0.55.

To sum up, we conclude that results are robust to outliers and to different specifications. A 2SLS regression implies that OLS may over-estimate the effect of reduction in driving time on trade growth, but the effect is still large and significant in the 2SLS regression. A drawback of our analysis is that it fails to capture adequately the dynamics of trade.

1.5 The Extensive Margin of Trade

Two cities can expand their trade in two ways: they can expand the volume of goods already traded (the intensive margin) or they can start trading in new goods (the extensive margin). The extensive margin of trade accounts for a large share of trade growth in response to international trade agreements like NAFTA (Kehoe and Ruhl, 2002). Evenett and Venables (2002) look at the exports of developing countries and also find that the extensive margin constitutes a significant fraction of trade growth. Regressions in the

[18] $e^{0.03\times 3.28}-1=10.34\%$ and $e^{0.04\times 3.28}-1=14.02\%$

previous section focus on the intensive margin, because consumer preferences in the theoretical model impose positive trade in all industries and among all regions.

No model to our knowledge produces a gravity equation treating both the intensive and extensive margin. Our model captures only the intensive margin, while a spatial Dixit-Stiglitz model in Fujita, Krugman and Venables (1999) examines only the extensive margin. Their model uses a CES utility function with preference for variety of goods, predicting that as trade costs drop, a region expands trade by producing and trading new goods.

In this section, we analyze the extensive margin of trade by carrying out two regressions. The first regression is a linear probability regression. The dependent variable is equal to 1 if shipments are greater than zero, and 0 otherwise. Explanatory variables include driving times between pairs of areas in year t, the output of industry k in city k and 2940 city-pair-by-industry dummies, which control for regional and industry heterogeneity as well as any other fixed factors affecting trade between two areas, such as distance. The drawback of this regression is that there are only two possible values of the dependent variable. The second regression complements the first regression by using a dependent variable with a wider range of values. The dependent variable is the number of industries traded between any pair of cities with values ranging from 0 to 17. Since good roads reduce trade costs, we expect new IH both to increase the probability of trade and to increase the number of industries traded.

Trade relationships among U.S. cities changed dramatically along the extensive margin during the sample period. Table 13 shows that among the 2940 sample city pair-by-industry observations, only 1819 observations retained the same status of trade from 1963 to 1972. Among the remaining 1121 observations, 212 observations traded in 1963 and 1967 but stopped trading in 1972; 116 observations traded in 1963 but stopped trading thereafter; 210 observations did not trade before 1972 but started trading in 1972; 183 observations did not trade before 1967 but traded in both 1967 and 1972; 167 observations stopped trading in 1967 and started to trade again in 1972; 233 observations started trading in 1967 and stopped trading in 1972. The volatile trade patterns suggest that examining the extensive margin of trade is necessary.

Table 13: Shifts between Positive and Zero Trade--Number of Observations

Trade in			# observations
1963	1967	1972	
Y	Y	Y	1568
Y	Y	N	212
Y	N	Y	167
N	Y	Y	183
Y	N	N	116
N	Y	N	233
N	N	Y	210
N	N	N	251
	Total		2940

Sources: Commodity Transportation Surveys (1963, 1967 and 1972), provided by the Bureau of Transportation Statistics;
Notes: 1. There are 3 years and 2940 observations for each year. Every observation measures shipments of a 2-digt industry from area i to area j at year t.
2. This analysis defines zero trade from origin to destination as occuring when the origin area has positive shipments of that 2-digit industry.
3. "YYY" means positive trade in all 3 years. "YYN" indicates trade existed in 1963 and 1967, but not in 1972. Other combinations of Y and N are similarly defined.

Table 14 reports the effect of new highways on the extensive margin of trade. Column (1) is a linear probability regression, investigating the link between driving times and the likelihood of trade. It uses panel data of 8820 city-pair-by-industry observations in 1963, 1967 and 1972, among which 918 have zero shipments. Driving time has a negative effect on the probability of trade which is significant at the 99 percent confidence level. A 1 percent increase in driving time decreases the probability of shifting from no trade to trade by 0.12 percentage points. Highways reduced the average driving time by 3.34 percent in my sample. This implies that the probability of trade among areas increased by 0.40 percentage points due to the IHS. Column (2) uses the number of industries traded between any pair of cities as the dependent variable. The results are similar: (log) change in driving time has a negative and significant effect on the number of industries traded. Here a 100 percent increase in driving time decreases the number of 2-digit industries traded between two locations by 0.87. IH reduced the average driving time by 3.44 percent, so that the number of 2-digit industries traded between cities increased by 0.03, on average, due to the introduction of the IHS.

1.6 Interstate Highways and Regional Development

The previous sections estimate that the Interstate Highway System significantly promoted intercity trade. This section examines the effect of improved transportation on regional development. A common difficulty of such research is the endogeneity problem that rich regions invest more in their transportation system.

Table 14: Extensive Margin of Trade

Dependent Variable	$AnyTrade_{ijk,t}$ (1)		$\#IndustryTraded_{ij,t}$ (2)
Log intercity driving time between *i* and *j* in year *t*	-0.12 (-5.43)**	Log intercity driving time between i and j in year t	-0.87 (-2.29)**
Log output of industry *k* in the origin area *i* in year *t*	-0.04 (-4.99)**		
Origin-destination-industry dummy	Y	Origin dummies, destination dummies and origin-destination dummies	Y
#obs	7305	#obs	720
R-squared	0.73	R-squared	0.88

Sources: Commodity Transportation Surveys (1963, 1967 and 1972), provided by the Bureau of Transportation Statistics
Rand McNally Road Atlas (1962, 1968 and 1973), provided by Rand McNally
Notes: 1. The AnyTrade=1 if shipment>0, and =0 if shipment=0.
2. The dependent variable is the (log) change in shipments (tons) from area i to area j of industry k from 1967 to 1972. Each area consists of several Standard Metropolitan Statistical Areas (SMSA). 7 of the 16 areas are in the northeastern region of US. The 17 industries are manufacturing industries at the 2-digit TCC level.
3.Regression (1) is $AnyTrade_{ijk,t} = \alpha + \beta * \ln(drive_{ij,t}) + \gamma * \ln(tons_{ik,t}) + \phi_{ijk} + \varepsilon_{ijk,t}$, t=1963, 1967 and 1972.
4. Regression (2) is $\#IndustryTraded_{ij,t} = \alpha + \beta * \ln(drive_{ij,t}) + \theta_{ij} + f_{it} + f_{jt} + \varepsilon_{ijk,t}$, t=1963, 1967 and 1972.
5. Explanatory variables of primary interest, $drive_{ij,t}$, is the driving time between areas i and j in year t. $tons_{ik,t}$ is the weight of output of industry k in region i.
6. t-statistics are in parenthesis.

We circumvent that endogeneity problem applying the method of Fernald (1999). Fernald (1999) argues that if roads increase productivity, industries with more vehicles will benefit more from road expansion. He applies a growth accounting procedure that separates productivity growth due to road investment from that due to capital or labor and finds that vehicle-intensive industries (transport services) benefit more than other sectors from public investment in roads. We carry out similar regressions to examine in our sample cities how industries with different reliance on the IHS benefited disproportionately from the improved transportation system.

Our basic result is that industries relying on the IHS had higher growth rates of output, employment and firm entry over our sample period. This evidence suggests that industries relying on the IHS benefited disproportionately from the construction of the IHS. This implies that new IH cause economic prosperity, not the other way around. For example, if the construction of the IHS was endogeneous (because when the local economy grew, the government decided to build more roads), or if the relationship between new IH and economic prosperity was totally spurious, we would not expect the growth rates of regional economic variables to increase with a region's concentration of industries that rely on the IHS.

We use data from the Commodity Transportation Survey (1967 and 1972) and the Census of Manufactures (1967 and 1972). The Census of Manufactures (CM) is provided by the Department of Commerce, recording the output, investment and employment of 2-digit industries in major metropolitan areas. All 16 sample cities are in both the CTS and the CM.

We examine the growth rates of value-added, total employment, the number of establishments, and capital expenditure in the 16 sample cities from 1967 and 1972. Table 15 shows that industries differ considerably in the growth rates of these variables at the national level. Value-added ($million) increases the most in industry 30 (Rubber and Misc. Plastic Products) by 49 percent, and decreases the most in industry 31 (Leather and

Table 15: Growth Rates of Economic Activity by Industry (1967-1972)

TCC code	Industry	Shipments within 200 Miles[1] (%)	Vehicle Shares[2] (%)	Shipments by Road (%)	Growth Rates of Economic Activity (%)			
					# Establishments	# Employees	Value Added	Capital Expenditure
20	Food	46.20	1.3	50.60	-14.53	-13.68	24.27	17.47
22	Textile mill prod.	45.70	0.3	90.70	-6.49	0.16	39.34	123.27
23	Apparel and related prod.	34.10	0.3	82.20	-16.76	-12.72	26.61	74.03
25	Furniture and fixtures	29.60	0.6	76.30	-5.48	5.77	36.58	70.66
26	Paper and allied prod.	34.90	0.9	42.00	5.83	1.37	40.86	-4.47
28	Chemicals	46.10	0.7	42.30	-8.34	-2.78	28.85	-25.74
29	Petroleum refining	30.70	1	15.60	-1.62	-6.59	-0.65	-63.74
30	Rubber and misc. plastics	34.40	0.2	74.70	28.61	4.60	48.80	109.07
31	Leather & leather prod.	55.90	0.2	92.20	-33.53	-42.30	-30.58	-8.55
32	Stone, clay, etc.	71.50	2.8	63.30	0.19	0.88	40.36	26.85
33	Primary metal	45.90	0.6	45.00	-6.95	-20.09	1.72	-38.71
34	Fabricated metal	47.20	0.5	71.40	1.58	2.54	39.85	1.29
35	Machinery, except electrical	26.10	0.3	69.20	1.03	-7.16	25.22	-15.63
36	Electrical and electronic machinery	21.00	2.2	65.20	6.23	-22.34	8.95	-53.70
37	Transport equipment	33.00	5.5	44.80	-7.41	-35.92	-3.66	-36.95
38	Instruments	24.10	0.4	76.60	26.56	-32.08	-27.27	-54.91
39	Misc. manu. prod.	29.00	0.4	77.90	-1.77	-1.84	38.67	94.65

Sources: The Commodity Transportation Surveys (1967 and 1972) , The Census of Manufactures (1967 and 1972) and Fernald (1999).
Notes: 1. Defined as shipments within 200 miles by all modes of transportation.
2. Source: Fernald (1999).

Leather Products), by 31 percent. The number of employees (1000 workers) increases the most in industry 25 (Furniture and Fixtures), by 6 percent, and decreases the most in industry 31 (Leather and Leather Products), by 42 percent. The number of establishments increases the most in industry 30 (Rubber and Misc. Products) by 29 percent, while decreasing the most in industry 31 (Leather and Leather Products) by 34 percent. Capital expenditure jumps by 123 percent in industry 22 (Textile Mill Products), while dropping by 64 percent in industry 29 (Petroleum Refining).

To estimate the effect of the transportation system on economic activity, we carry out the following regressions for each measure of activity listed above:

$$y_{ik,t} - y_{ik,t-1} = \alpha_0 + \alpha_1 IHSreliance_k \times (time_{i,t} - time_{i,t-1}) + dummy_i + dummy_k + \psi_{ik} \quad (19)$$
$$cov(\psi_{ik}, \psi_{ih}) \neq 0$$
$$t = 1967, 1972$$

Here y_{ikt} is a measure of activity for industry k in city i. The activity measures we consider are value-added, the number of employees, capital expenditure and the number of establishments. $IHSreliance_k$ measures how heavily a 2-digit industry relies on the IHS. $time_{it}$ is the average driving time between cities i and all the other 15 sample cities in year t. $time_{it} = \frac{1}{15}\sum_{j \neq i} time_{ij,t}$. $dummy_k$ controls for different industry trends that may or may not have something to do with the IHS. $dummy_i$ absorbs any city-specific variables, such as regional economic growth or regional improvement in the transportation system, that may affect $(y_{ik,t} - y_{ik,t-1})$. ψ_{ik} is an error term. We carry out separate regressions for three measures of the reliance on the IHS---percentage of shipments over 200 miles (CTS, 1967), percentage of shipments by road (CTS, 1967) and vehicle shares (Fernald, 1999). We cluster standard errors by cities.

The key explanatory variable is the interaction, $IHSreliance_k \times (time_{i,t} - time_{i,t-1})$. Our theory predicts that if an industry relied more on the IHS than other industries and if the city it located in improved more on the transportation system (as proxied by declines in driving times) than other cities, that industry-region should have had more growth over our sample period than other industry-regions. Thus we predict that the coefficient α_1 should be negative and significant.

Tables 16-18 lists the results of the regressions. Regressions in Table 16 measure the reliance on the IHS with the percentage of shipments over 200 miles (national goods). Column (1) implies that value-added increases more in industries that produce more national goods in response to a drop in driving time. The estimated effect of driving time on value-added is large. Evaluated at the median of the share of national goods (0.344), a one-hour decline in driving time increases the value-added of a 2-digit industry in a city by $29.35 million. The mean of value-added of 2-digit industries in the 16 sample regions is $282 million. An increase of $29.35 million is a 10.41 percent increase.

Table 16: The Effect of IHS on Economic Development (National Goods)

Dependent Variable: Growth Rates of Economic Variables	Value-added	# Employees	Hours worked of production Workers	# Production Workers
	(1)	(2)	(3)	(4)
Percentage of Shipments more than 200 miles multiplied by Change in Average Driving Time	-85.31 (-2.92)**	-3.54 (-2.14)**	-4.05 (-1.51)	-2.32 (-1.87)*
City dummies	Y	Y	Y	Y
Industry dummies	Y	Y	Y	Y
#obs	250	250	250	250
R^2	0.12	0.22	0.21	0.22
Dependent Variable: Growth Rates of Economic Variables	**Capital Expenditure**	**# Total Establishments**	**# Establishments ≤ 20 Employees**	**# Establishments > 20 Employees**
	(5)	(6)	(7)	(8)
Percentage of Shipments more than 200 miles multiplied by Change in Average Driving Time	-2.04 (-0.65)	-21.79 (-1.85)*	2.60 (0.10)	8.87 (0.36)
City dummies	Y	Y	Y	Y
Industry dummies	Y	Y	Y	Y
#obs	233	252	252	252
R^2	0.19	0.28	0.33	0.24

Sources: The Commodity Transportation Surveys (1967 and 1972) and The Census of Manufactures (1967 and 1972).
Notes: Regressions are Ordinary Least Squares with standard errors clustered by cities.

Table 17: The Effect of IHS on Economic Development (Vehicle Shares)

Dependent Variable: Growth Rates of Economic Variables	Value-added	# Employees	Hours worked of production Workers	# Production Workers
	(1)	(2)	(3)	(4)
Vehicle Shares Multiplied by Change in Average Driving Time	-36.16 (-2.48)**	-0.82 (-1.40)	-0.75 (-0.70)	-0.43 (-0.88)
City dummies	Y	Y	Y	Y
Industry dummies	Y	Y	Y	Y
#obs	250	250	250	250
R^2	0.23	0.24	0.22	0.23
Dependent Variable: Growth Rates of Economic Variables	**Capital Expenditure**	**# Total Establishments**	**# Establishments ≤ 20 Employees**	**# Establishments > 20 Employees**
	(5)	(6)	(7)	(8)
Vehicle Shares Multiplied by Change in Average Driving Time	-0.49 (-0.44)	0.98 (0.45)	4.54 (1.33)	1.97 (1.05)
City dummies	Y	Y	Y	Y
Industry dummies	Y	Y	Y	Y
#obs	233	252	252	252
R^2	0.19	0.28	0.33	0.24

Sources: Vehicle shares of 2-digit industries are from Fernald (1999). All the other variables are from The Commodity Transportation Surveys (1967 and 1972) and The Census of Manufactures (1967 and 1972).
Notes: Regressions are Ordinary Least Squares with standard errors clustered by cities.

Table 18: The Effect of IHS on Economic Development (Share of Shipments by Road)

Dependent Variable: Growth Rates of Economic Variables	Value-added	# Employees	Hours worked of production Workers	# Production Workers
	(1)	(2)	(3)	(4)
Percentage of Shipments by Road multiplied by Change in Average Driving Time	66.19 (2.62)**	0.23 (0.18)	-0.40 (-0.17)	-0.17 (-0.15)
City dummies	Y	Y	Y	Y
Industry dummies	Y	Y	Y	Y
#obs	250	250	250	250
R^2	0.12	0.22	0.21	0.22
Dependent Variable: Growth Rates of Economic Variables	**Capital Expenditure**	**# Total Establishments**	**# Establishments ≤ 20 Employees**	**# Establishments > 20 Employees**
	(5)	(6)	(7)	(8)
Percentage of Shipments by Road multiplied by Change in Average Driving Time	-0.77 (-0.10)	-24.61 (-1.27)	-35.82 (-0.84)	-19.08 (-1.46)
City dummies	Y	Y	Y	Y
Industry dummies	Y	Y	Y	Y
#obs	233	252	252	252
R^2	0.19	0.28	0.33	0.25

Sources: The Commodity Transportation Surveys (1967 and 1972) and The Census of Manufactures (1967 and 1972).
Notes: Regressions are Ordinary Least Squares with standard errors clustered by cities

Column (2) shows that total employment increases more in industries that produced more national goods in response to a drop in driving time. The estimated effect of driving time on total employment is large. Evaluated at the median of the share of national goods (0.344), a one-hour drop in driving time increases the employment of a 2-digit industry in a city by 1,218 workers. A 2-digit industry in a sample region has 19,600 employees on average. 1,218 workers are 6.2 percent of the total employment. Column (3) shows that hours worked of production workers increase more in industries producing national goods than in those producing local goods in response to a drop in driving time, but the effect is not significant. Column (4) indicates that the number of production workers also increases more in industries producing national goods than in those producing local goods in response to a drop in driving time. Column (5) finds that the effect of distance shipped on capital expenditure is insignificant.

Column (6) shows that firm entry is higher in industries that produced more national goods in response to a drop in driving time. The estimated effect of driving time on firm entry is large. Evaluated at the median of the share of national goods (0.344), a one-hour decline in driving time increases the firm entry of a 2-digit industry in a city by 8 firms. On average, a 2-digit industry in a sample city has 324 firms. 8 firms amount to 2 percent of the number of existing firms. Columns (7) and (8) show that the key coefficient becomes insignificant if we examine large firms (more than 20 employees) and small firms separately.

Table 17 uses the vehicle shares in Fernald (1999) to measure how heavily industries rely on roads. Fernald (1999) defines the vehicle shares as the input share of vehicles in total revenue. Industries providing transport services have larger vehicle shares than other industries. Column (1) shows that value-added increases more in industries that have larger vehicle shares than other industries in response to a drop in driving time. Column (2) shows that total employment also expands more in those industries, but the effect is insignificant. Results in Columns (5)-(8) show that the investment and firm entry of industries with large vehicle shares do not expand more than other industries in response to a drop in driving time.

Table 18 uses the percentage of shipments by road to measure how heavily industries rely on IHS. Column (1) shows that value-added drops in industries that rely heavily on roads in response to a drop in driving time. It is possible that industries with low road shares in 1967 may have had more opportunities to substitute other modes of transportation towards roads in response to the IHS than those with high initial road shares (Figure 2). Columns (2)-(8) suggest that the employment, investment and firm entry of industries with large road shares do not expand more than other industries.

Overall, the effects of new IH on value-added, employment, and firm entry are both large and significant.

1.7 Conclusions

Investment in infrastructure accounts for 13 percent of the total expenditure of state and local governments[19]. Despite the large amount of spending, existing literature provides only mixed evidence that infrastructure contributes to the economy.

We analyze the effect of infrastructure on economic prosperity, focusing on how the construction of the Interstate Highway System (IHS) contributed to regional development in the United States by expanding intercity trade. We use three unique data sets and a new measure of transportation infrastructure, namely, intercity driving times. This new measure allows us to exploit disaggregate data on bilateral trade flows and overcome some of the typical problems in research on infrastructure.

We find that the IHS reduced driving times among my sample cities by 3.34 percent between 1967 and 1972, which subsequently increased trade volume among existing trading partners by 4.24 percent, and raised the probability of trade among non-trading areas by 0.40 percentage points. Moreover, trade in national goods grew more than trade in regional goods, because the former relied more on the IHS than the latter.

We estimate that industries using the IHS intensively benefited disproportionately from the IHS, implying that the IHS caused regional development. In particular, the growth rates of regional output, employment and firm entry were higher in industries that sold

[19] Source: US Census Bureau, 2005. State and Local Government Finances by Level of Government and by State: 2001 - 02. Available at http://www.census.gov/govs/estimate/0200ussl_1.html

goods on a national market. This suggests that regions with different industry compositions will benefit differently from the new infrastructure.

This research starts from the basic insight that cross-section variation among countries or within a country provides important information. Existing research focuses on the variation of infrastructure at high levels of aggregation without much cross-section variation. By contrast, the rich variation of infrastructure investment in this chapter suggests a large effect of infrastructure on economic activities. Future evaluations of the contribution of infrastructure to economic growth should focus on welfare analysis and should consider all aspects of the economy, such as the productivity gains from division of labor.

Another direction of future research is to estimate and predict the contribution of infrastructure investment in developing countries. Governments in developing countries invested an average of 1.4 percent of their GDP in infrastructure in 2000 (IMF statistics). We can use the method of this chapter to analyze the impact of investment in infrastructure in developing countries.

2. Engineering Costs and the Construction of the Interstate Highway System

U.S. state and local governments invested $275 billion in infrastructure in 2002[20]. This investment accounted for 13% of the total expenditure of state and local governments. Despite the size of this investment, the large literature on the contribution of infrastructure to economic activities provides only mixed evidence on whether infrastructure contributes to economic growth at all. A major difficulty is identifying the direction of causality. To see this difficulty, consider a regression of regional income on regional investment in infrastructure. Such a regression usually presents a positive relationship. This relationship, however, may be a result of reverse causality: better infrastructure may not lead to higher growth, but regions with higher growth may invest more in infrastructure. Therefore, an Ordinary Least Squares regression may suffer from reverse causality and overestimates the effect of infrastructure on economic growth.

Because of reverse causality, economists disagree on whether infrastructure contributes to economic growth at all. Aschauer (1990, 1989, 1988) estimates that if the stock of infrastructure increases by 1%, multifactor productivity will increase by 0.4%, and that productivity slowed after 1970 because local governments neglected infrastructure. Moreover, he finds that infrastructure is 4 times more productive than private capital. Aschauer's work has taken heavy criticism. Many economists (Aaron, 1991; Schultze, 1990; Hulten and Schwab, 1991; Jorgenson, 1991) argue that the reverse causality may

[20] We follow the definition of infrastructure in Duffy-Deno and Eberts (1991). "Public capital includes: (a) sanitary and storm sewers and sewage disposal facilities, (b) roadways, sidewalks, bridges and tunnels, (c) water supply distribution system, (d) public hospitals, and (e) public service enterprises such as airports and ports." Source: US Census Bureau, 2005. State and Local Government Finances by Level of Government and by State: 2001 - 02. Available at http://www.census.gov/govs/estimate/0200ussl_1.html

have driven Aschauer's results. Hulten and Schwab (1991) compare the "snow belt" and the "sun belt", concluding that it is the difference in private capital and labor, not infrastructure, that accounts for the different productivities in those regions. Also, they stress that the problem of reverse causality aggravates any research on infrastructure.

Moreover, results in this literature are sensitive to econometric methods (Hulten and Schwab, 1991). For example, Holtz-Eakin (1988) and Aschauer (1990) use similar data sets, but Holtz-Eakin (1988) allows for non-stationary data while Aschauer (1990) does not. Holtz-Eakin (1988) thus finds that the effect of infrastructure on economic growth is zero, while Auschauer (1990) estimates a large contribution of infrastructure to economic growth.

Economists have used three ways to control for reverse causality. The first way is to use disaggregate data and to control for regional or industry heterogeneity. For example, Chapter 1 of this dissertation examines city pair-industry commodity flow data, using city dummies to control for the fact that rich regions built more roads. But Chapter 1 does not control for a potential endogeneity problem at the city-pair level—that the first roads governments built may have been those between cities with the highest trade potential.

The second way is to explore the growth of industries with different reliance on roads. Fernald (1999) examined the economic growth of 2-digit industries during the period of expansion of the IHS. He argued that because industries with higher vehicle intensity grew more during the expansion, the IHS contributed to that growth. Chapter 1 uses the same strategy to examine the growth of 2-digit industries in major cities, finding that industries that relied heavily relied on roads benefited more from the construction of the IHS. The strength of this method is that the explanatory variable—vehicle intensity or another measure of road reliance—does not directly cause industry growth. The weakness is that while we can identify that an industry grows because of improvements in the transportation system, we cannot tell which part of the transportation system—a particular infrastructure project, or improvements to vehicles—caused the growth.

The third method, instrumental variables, is under-developed because macroeconomic variables such as infrastructure investment are usually strongly affected by other macro variables. Baum-Snow (2007) uses a simple instrumental variable to estimate the effect of highways on urban development. He instruments the number of highways through a city with the original 1947 IHS plan, assuming that the purpose of the system was for national defense, not for economic growth. He finds that each additional Interstate Highway going through a city causes 18% of the city's population to move from the city to its suburbs. But his assumption that the original IHS plan was designed solely for national defense may not hold. Also, the number of highways through a city measures infrastructure in an overly simple way. Research papers such as Aschauer (1988, 1989 and 1990) and Hulten and Schwab (1991) use the stock of or investment in infrastructure as the explanatory variable. The number of highways may not reflect precisely the stock of or investment in infrastructure.

This paper has two goals. The first goal is to develop an instrument for a new measure of infrastructure, focusing on how driving times among cities dropped when the government built the IHS. OLS regressions of intercity driving times on IH construction may produce biased results. Rich regions may build more highways but may also generate more traffic congestion, which would bias OLS estimates of the impact of road construction on driving time downward. The second goal is to provide additional evidence for a key assumption in Chapter 1, that ease of construction heavily influenced which road segments were built first. Chapter 1 controls for the possibility that rich cities may build more roads, and estimates that the IHS expanded intercity trade. However, we assume that state governments decided the timing of construction of each highway for reasons other than trade potential along the route. This paper provides some historical and empirical evidence that ease of construction was an important determinant of the timing of construction.

The basic idea is that a looming Federal Highway Trust Fund crisis in the 1960s forced states to prioritize which road segments they would build based on how much the segments cost, and the crisis drove states into a race to complete as much mileage as soon as possible. States built cheap segments earlier than expensive segments, and sometimes built them in a random, disconnected fashion. This identification strategy is similar to that of Duflo and Pande (2007). They estimate that dams redistribute income from upstream regions to downstream regions of dams. They argue that some areas are not conducive to dams because of geography, and instrument the location of dams with river gradients.

We assume that engineering costs are exogeneous to regional economic performance. If this assumption fails to hold, we would most likely under-estimate the effect of engineering costs on the timing of construction. The reason is that rich regions, which usually have high construction price levels, may build IH faster than poor regions, while our theory argues that high construction prices lead to a delay of construction. Therefore, engineering costs can be used as an instrument for differences in the timing of construction and declines in driving times among cities.

This instrument can be used to measure the contribution of IH construction to a novel measure of infrastructure, namely driving times, which can help open the black box of the mechanisms through which infrastructure affects the economy. Driving times measure infrastructure more precisely than the measure used in Baum-Snow (2007), because change in driving times is a continuous variable that responds both to the number and length of new segments of roads, while Baum-Snow (2007) counts only the number of new roads.

In the rest of the paper, we first estimate how engineering costs determined the timing of construction for each segment of highway and then predict the estimated mileage open at particular dates of particular IH connecting cities. We find that engineering costs are important determinants of the timing of construction. Then we estimate how much IH reduce driving times among cities, using engineering costs as an instrument for the length of interstate highways on a given route.

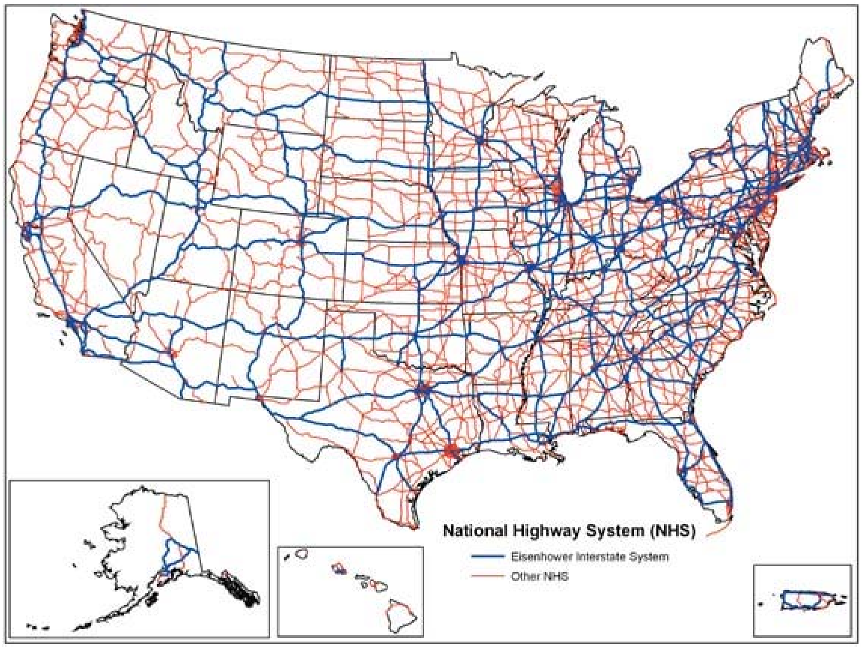
2.1 Historical evidence on the construction of the Interstate Highway System

We are interested in the change in Interstate Highway connections among cities from 1962 to 1973, a period for which data on driving times is available. States built the IHS at similar speeds because the federal government allocated the construction funds assuming that all states would finish in 1972 (Federal Highway Administration, 1998). Within each state, however, historical evidence and empirical analysis suggest that state governments decided to build the cheapest roads first. Engineering costs of different road segments can thus instrument for the timing of road construction.

2.1.1 A brief history of the Interstate Highway System.

The System is currently 46,876 miles long (Federal Highway Administration, 2007), consisting of 10 major highways from East to West and 7 highways from North to South (Figure 3). The US government built the IHS to connect industrial centers and major cities. The 1944 Federal Highway Act said that IH should be "… so located as to connect by routes, as direct as practicable, the principal metropolitan areas, cities, and industrial centers…"(Federal Highway Administration, 1996). For example, I-90 connects Seattle with Boston, going through Chicago, Buffalo and Syracuse. I-95 connects Boston with Miami, going through New York, Philadelphia and Baltimore. Since 53% of goods (by value) were shipped by road in 1967, the IHS is critical to enabling cross-country shipping, integrating what would otherwise be a group of isolated regional economies into a national market[21].

Figure 3: The Interstate Highway System



Source: Federal Highway Administration. Available at
http://www.fhwa.dot.gov/hep10/images/nhsjpg.jpg

States governments constructed most of the IHS between 1956 and 1973 (Figure 4). State governments incorporated 1,930 miles of existing highways into the IHS, then built

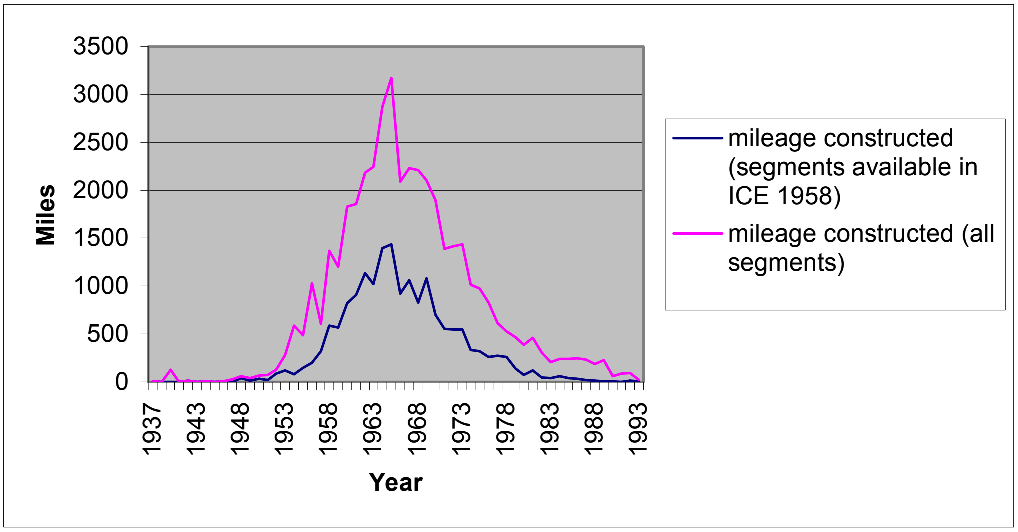
[21] We get the value (1967 dollars) of outputs of 2-digit manufacturing industries from the Census of Manufactures (1967), and the weight of outputs and the percentage of shipments by road from the Commodity Flow Survey (1967). Then we get the value of shipments by road using the formula:

$$(\sum_i weight\_ship_i \times value\_per\_ton_i \times percent\_road_i)/(\sum_i weight\_ship_i \times value\_per\_ton_i)$$

i denotes each 2-digit industry.

7,899 miles of highways between 1956 and 1961, 25,260 miles from 1962 to 1973, and 7,427 miles from 1974 to 1993[22].

Figure 4: Mileage Constructed in Each Year (1956-1993)



Sources: *Interstate Highway Construction Records* (1937-1993) and *Interstate Cost Estimate 1958*, provided by the Federal Highway Administration.

2.1.2 The difference in the timing of construction across states.

Political power may have affected the location of IH during the planning period of the IHS. Using data on 1998 Congressional votes over transportation projects, Knight (2004) estimates that powerful congressmen affect the allocation of federal highway funds to their districts. However, in theory political power should not affect the speed of construction in states. All states should have constructed the System at a similar speed.

[22] Source: Interstate Highway Construction Records (1937-1993), provided by the Federal Highway Administration.

The US government set the agenda and financial rules of construction in the 1956 Federal Highway Act. The 1956 Federal Highway Act stated that the state governments should finish building the IHS by 1972. The federal government would finance 90% of the costs, while the state governments would finance 10%. The federal government set up a Highway Trust Fund to finance the construction, allocating money to state governments to assure that all states could finish construction simultaneously (Federal Highway Administration, 1998).

In practice, differences in state sizes and the existing mileage of toll roads made construction speed different across states. A small state such as Rhode Island had only 69 miles to build, so they built their portion of the IHS faster than other states. Many northeastern states had several major toll roads built before 1956. The federal government incorporated those toll roads into the IHS. We control for different speeds of construction at the state level in our regressions.

2.1.3 The difference in the timing of construction within states.

At the state level, governments typically built the cheapest (in terms of cost per mile) segments first, because of a looming financial crisis in the IHS[23]. All states but Utah built rural highways earlier than urban highways, because reallocating urban residents is

[23] Source: Interviews with historians and engineers at the Federal Highway Administration conducted in 2004.

costly and time-consuming[24]. They built segments first on plain terrain and not through the mountains. They sometimes built segments of highways randomly and in a disconnected fashion. For example, the map of the IHS in 1964 shows disconnected segments between Houston and New Orleans (Highway Statistics, 1965). Cheap segments were built earlier than expensive segments in response to a perceived financial crisis of the Highway Trust Fund.

The financial crisis began in the late 1950s. The estimated costs of the whole System increased from $25 billion to $36 billion[25] in 1959 for the following 2 reasons (Metz and Ritter, 2006). First, the IHS was a large national project, increasing construction wages and the prices of specialized raw materials[26]. Second, Congress added an extra 1000 miles into the System in 1956 without providing extra financial resources[27]. As a result of the looming financial crisis, federal highway officers toured the country, stating that if state governments failed to finish the construction by 1972, they might not get enough funds for construction after 1972. In particular, the federal authorities emphasized that state governments would have to finish half of the designed mileage by 1967 in order to

---

[24] Source: Interviews with historians and engineers at the Federal Highway Administration conducted in 2004.

[25] The estimated costs of the IHS increased to $129 billion (1998 dollars) in 1991 (Federal Highway Administration, 1998).

[26] Source: Interviews with historians and engineers at the Federal Highway Administration conducted in 2004.

[27] Source: http: Answers.com. //www.answers.com/topic/interstate-highway-system

finish the whole system by 1972 (Metz and Ritter, 2006). Indeed, state governments finished half by 1967[28].

2.2 Empirical Strategy

Our goal is to estimate the change in intercity driving times due to new IH, instrumenting the length of new IH using engineering costs. OLS regression may under-estimate the effect of highways on driving time, because rich regions may build more highways, but may also generate more traffic congestion and thus longer driving times. The basic idea can be described by a 2-stage regression. The 1st stage estimates the effect of engineering costs on the timing of construction. Then we construct a fitted value to show how many IH miles should have been open on a particular route by each year. The 2nd stage examines how new IH changed driving times among cities, using the fitted value as an instrumental variable for actual IH mileage.

The 1st regression tests the assumption that state governments built cheap segments earlier than expensive segments. In particular, we test the hypothesis that segments with large estimated costs per mile opened to traffic later than those with small estimated costs. The dependent variable is the open-to-traffic time (1958-1993), and the key explanatory variable is the engineering costs per mile estimated in 1958. We add the square of costs-per-mile as an additional explanatory variable, because the effect of costs

[28] After Congress passed legislation to provide more financial resources for the IHS in the mid 1970s, federal and state governments were able to continue with construction and expand the system to over 46,000 miles. Source: Interviews with historians and engineers in the Federal Highway Administration.

on the open-to-traffic time is likely to diminish as costs increase. For example, the average cost per mile in the data is $0.74 million (1958 dollars), while the cost of a tunnel or bridge can climb to $83 million. We expect a $1 million increase from $0.74 million to $1.74 million to have more of an effect on the open-to-traffic time than a $1 million increase from $83 million to $84 million. Other explanatory variables include state dummies. States with higher prices and higher wages would incur larger construction costs. Also, states of different sizes built highways at different speeds. We estimate the following equation.

$$\begin{aligned} &time\_open_i = \alpha_0 + \alpha_1 cost_i + \alpha_2 cost_i^2 + dum\_state_i + \mu_i \\ &\mathrm{var}(\mu_i) = \sigma_i^2 \end{aligned} \tag{20}$$

The subscript i denotes each segment of IH. $costs_i$ is the construction costs per mile estimated in 1958. dum_state_i represents a series of state dummies, controlling for state-specific variables, such as state price level and state size. States with high price levels are more likely to incur large engineering costs, while our hypothesis is that large costs delay open-to-traffic time. Therefore, without state dummies, an OLS regression is likely to underestimate the effect of costs on the open-to-traffic time. We expect α_1 to be positive and α_2 to be negative, since states will delay the construction of expensive segments and since the effect of costs on the open-to-traffic time is diminishing.

This equation omits many other factors that may affect open-to-traffic time, such as local wages. We expect state dummies to absorb differences in local price levels across states.

This regression will at worst under-estimate the true effect of engineering costs. After controlling for the state price levels, engineering costs may still be correlated with regional economic performance. Rich regions yet may have higher wages and prices and may build roads faster than poor regions. Therefore, higher costs would predict earlier construction, while our hypothesis is that higher costs predict a delay of construction.

The 2nd regression tests for a relationship between engineering costs and the change in driving times among cities over the period 1962-1973. The theoretical relationship between driving time, velocity, and highway mileage is given by equation (1) in Chapter 1:

$$drive_{ijt} = \frac{d_{ij} - I_{ijt}}{v_t} + \frac{I_{ijt}}{1.154 v_t} = \frac{d_{ij}}{v_t} - 0.133 \frac{I_{ijt}}{v_t},$$
$$t = 1962, 1968, 1973.$$

d_{ij} is the mileage of the route between city *i* and city *j* with the shortest driving time, which in general is not equal to the straight line distance calculated from the longitude and latitude of cities. Distance remained mostly unchanged from 1962 to 1973. I_{ijt} is the total interstate highway mileage on this route; and v_t is the travel velocity on non-interstate highways in year *t*. The driving time between *i* and *j* is the sum of the time on non-interstate highways, $\frac{d_{ij} - I_{ijt}}{v_t}$, and that on IH, $\frac{I_{ijt}}{1.154 v_t}$. The speed on IH is $1.154 v_t$, because according to Highway Statistics 1973, travel velocity on IH during this period was on average 15.4 percent higher than on non-interstate highways. Therefore, driving time is a function of distance and the portion of IH en route.

We add year dummies and city-pair fixed effects to the above equation. The year dummies capture factors, such as an increase in speed at the national level, that may impact overall driving times. City-pair fixed effects absorb any pre-existing factors, like terrains and distance, that may affect driving times. The equation becomes:

$$drive_{ijt} = \beta_0 + \beta_1 I_{ijt} + \beta_2 dum\_year + \beta_3 f_{ij} + u_{ijt}$$
$$t = 1962, 1968, 1973$$

f_{ij} is the fixed effect of city pairs.

We first-difference the above equation, and estimate the following equation.

$$drive_{ij,t} - drive_{ij,t-1} = \alpha_0 + \alpha_1 (I_{ij,t} - I_{ij,t-1}) + \alpha_2 time\_dummy_t + \varphi_{ij,t} \qquad (21)$$
$$\mathrm{var}(\varphi) = \xi^2$$
$$t = 1962, 1968, 1973$$

We use the fitted value from (20) to construct a measure $\hat{I}_{ij,t}$, which shows how many IH miles should have been open on the route between city *i* and city *j* by year *t*. We expect α_1 to be negative, since IH increase driving speed and reduce driving times. If α_1 is significant, and since the construction costs were reported in 1958, we can conclude that construction costs estimated in 1958 predict the change in driving times from 1962 to 1973.

The regression equation is different from that in Chapter 1, because the dependent variable is the change in time (hours), while in Chapter 1, the dependent variable is (log) change in time. Chapter 1 uses (log) change in time to be consistent with the form of variables in the benchmark regression for intercity trade.

2.3 Data and Summary Statistics

We use three major data sources: Interstate Highway Construction Records (1937-1993), Interstate Cost Estimate (1958), and the Rand McNally Road Atlas (1962, 1968 and 1973). Interstate Highway Construction Records (IHCR) provide the value of the dependent variable of equation (20). This data records the month and year that each segment of IH opened to traffic. For example, the segment between 0 mileage and 10.4 mileage of Interstate 10 in Alabama was open to traffic in December 1965.

The Interstate Cost Estimate (1958) provides the key explanatory variables of equation (20). Starting in 1956, state governments submitted an Interstate Cost Estimate (ICE) to the federal government every two years. The federal government allocated construction funds based on the ICEs to ensure that every state would complete construction by 1972

(Federal Highway Administration, 1998). The 1st ICE in 1956 was inaccurate, so we use the 1958 ICE[29].

The 1958 ICE provides values of 14 items of construction costs, including (1) Preliminary Engineering, (2) Right of Way, (3) Clear and Grub, (4) Utility Adjustments, (5) Grade and Drain, (6) Subbase, Base, Surfacing, and Shoulders, (7) Railroad Grade Separations, (8) Highway Grade Separations Without Ramps, (9) Interchanges, (10) Other Bridges and Tunnels, (11) Walls, (12) Guardrail, Fencing, Lighting, and Traffic Control Devices, (13) Roadside Improvement and (14) all other items. Table 19 shows that the average estimated cost to build a mile of IH was $0.74 million (1958 dollars). Bridges and tunnels were the most expensive items on a per-mile basis. The most expensive bridge or tunnel costs $83 million per mile.

Table 19: Summary Statistics

Variable	#obs	Mean	Std. Dev.	Min.	Max.
Length of Segment (miles)	5397	3.21	3.46	0.02	36.1
Year opening to traffic	5397	1966	7.10	1937	1993
Cost per mile ($million)[1]	5397	0.74[2]	3.46	0.01	83.33

Source: *Interstate Cost Estimate 1958*, provided by the Federal Highway Administration.
Notes: 1. Cost per mile is the sum of 14 items of construction costs divided by the length of the segment. Costs are in millions of 1958 dollars.
2. It equals average cost per mile weighted by the length of each segment.

[29] Richard Weingroff, a historian at the Federal Highway Administration, suggested that we use the ICE 1958 instead of ICE 1956 because state governments made many mistakes preparing the first ICE. State governments improved their ICE preparation after 1956.

We have data on estimated costs for over 20,000 out of the 42,000 miles of the IHS. About 20,000 miles of data is missing, because the ICEs of 10 states are missing, and because some routes are missing for states with ICE. We use data on 17,871 miles out of the 20,000 miles of highways for the following reasons. First, state governments changed parts of the highway planning after 1958. For example, 276 miles of I-70 in Colorado in the current system were not in the original plan. The exact location of I-91 in Vermont changed. Since the average length of segment is only 3.21 miles (Table 19), a change in the location of a route makes it difficult to identify the open-to-traffic time and its corresponding engineering costs. Second, the maps we use for ICE are not accurate enough, so in some cases we are not able to identify the beginning of a route. Around some cities, we find it difficult to distinguish the beginning of a highway from the city bypass. If the error is more than 10 miles, we drop that segment from the sample.

Our data set is representative of the whole IHS. It roughly matches the general trend of construction of the IHS (Figure 4). The bottom line is the mileage constructed in each year in our sample, and the top line is that of the whole IHS. Both lines indicate that the peak of construction occurred in 1967, and that state governments built most of the system in the 1960s and early 1970s.

Our sample has 5397 segments of highways. The earliest opening-to-traffic time is 1937[30], and the latest opening-to-traffic time is 1993. The most expensive segment is $83 million per mile, and cheapest a mere $10,000. Therefore, our sample gives us enough variation to explore the relationship between costs and opening-to-traffic time.

[30] The federal government incorporated some high-grade roads and toll roads into the IHS in 1956.

The third major data source is the Rand McNally Road Atlas (1962, 1968, and 1973), which provides estimates of driving times between major U.S. cities. Rand McNally's estimated driving times depend on distance, road speed limits, and congestion[31]. For example, the published driving time between Detroit and Pittsburgh increased by 30 minutes due to congestion between 1968 and 1973, despite constant distances and road networks. We construct driving times among cities using the map of driving times among contiguous cities at the end of each Atlas. We consider different potential routes for traveling from one city to another and find the route with the shortest driving time (see Appendix).

Merging those data sets required an enormous amount of time. No researcher, to our knowledge, ever tried to construct a similar data set. The IHCR and ICE identify each segment by the mileage post on each route. The mileage is labeled 0 on the southern border of a state for highways going from south to north and labeled 0 on the western border for highways going from west to east. The map of driving times in the Rand McNally Road Atlas, however, provides no information on which route is used to construct driving times. We use Yahoo! Maps to find the exact route, assuming routes have remained the same from the 1960s to the present. We need the mileage posts of each city and each intersection in order to merge driving times with IHCR and ICE. The only way, to our knowledge, to find each mileage post is to count the segments mile by

[31] Rand McNally Road Atlas (1962) says: "Driving time shown is approximate under normal condition. Consideration has been given to topography, speed laws, and congested areas. Allowances should be made for night driving and unusually fast or slow drivers."

mile using detailed highway maps of each state. Merging the data sets for the 21 cities[32] used in Chapter 1 took us about 300 hours. Because of time constraints, our sample consists of just these 21 cities.

2.4 Results

This analysis instruments the timing of construction with engineering costs and estimates how much the construction of IH affects driving times among cities. The 1st stage explains the timing of construction with engineering costs. The 2nd stage relates driving times among cities with the timing of construction predicted by engineering costs.

2.4.1 First-stage Regressions

The benchmark regression is Equation (20), using only observations that opened to traffic after 1958. The number of observations is 4785. Column (1) of Table 20 implies that state governments delayed the construction of expensive segments. Moreover, the effect of engineering costs on the open-to-traffic time is diminishing. Both coefficients are significant at the 90% confidence level. The estimated relationship predicts a positive slope between open-to-traffic time and engineering costs for most observations. Only 78 out of 4785 segments fall into the downward sloping part of the quadratic relationship. The two coefficients are jointly significant at the 80% confidence level. The R squared

[32] Chapter 1 uses only 16 pairs of cities. We, however, calculated the driving times among 21 pairs of cities available in both Rand McNally Road Atlas and CTS (1967) and then dropped 5 cities that are not available in CTS (1963). So for this chapter we do have data on 21 cities.

is 0.01. Thus, engineering costs are only one of the determinants of open-to-traffic time. Our model ignores many other factors that determine open-to-traffic time.

That regression suggests that, at the mean engineering cost, a $1 million increase in engineering costs will delay the open-to-traffic time by 3.88 months[33]. Thus a one standard deviation of the costs ($ 3.46 million) is expected to delay the open-to-traffic time by 13.42 months—a large effect of engineering costs on open-to-traffic time. This regression underestimates the effect of engineering costs on open-to-traffic time, because of a simultaneity problem. Rich regions have larger potential for growth and political power, so state governments may choose to build roads around rich regions earlier than around poor regions. At the same time, rich regions incur large engineering costs because of high prices and wages. We would associate large engineering costs with earlier open-to-traffic time, contrary to our key hypothesis.

Table 20: The Effect of Engineering Costs on Open-to-traffic Time

Explanatory Variables (Costs per mile)	(1) Benchmark	(2) w/ state dummies	(3) 1937-1993
Total Costs	4.10	3.98	3.39
	(1.66)*	(1.56)	(2.39)**
Square of total	-0.15	-0.13	-0.07
costs	(-1.81)*	(-1.57)	(-3.31)***
State dummies		Y	
Constant	816.72	853.61	800.60
	(185.20)***	(289.05)***	(142.74)***
# obs	4785	4785	5397
R-squared	0.01	0.10	0.01

Sources: *Interstate Highway Construction Records* (1937-1993) and *Interstate Cost Estimate 1958*, provided by the Federal Highway Administration.
Notes: 1. The regression is OLS with standard errors clustered by states.
2. Regressions (1)-(3) uses only observations that opened to traffic after 1958.
3. The Dependent variable is open-to-traffic time (month+12*year).
4. All Costs are the sum of 14 items of construction costs.

[33] 4.10-2*0.15*0.74=3.88

Column (2) uses state dummies to control for difference across states. States with high price levels are more likely to incur large engineering costs, while we hypothesize that large costs delay open-to-traffic time. The key coefficients become insignificant, but the t-statistics are still large. The key coefficients are jointly significant at the 70% confidence level. Evaluated at the mean, a $1 million increase in engineering costs will delay the open-to-traffic time by 3.71 months, smaller than the effect estimated in Column (1). This result is opposite to our prediction that we would under-estimate the key coefficients without state dummies.

Column (3) uses all the segments that opened to traffic from 1937 to 1993. If states built highways in order of expense even before 1958, we would expect the key coefficients to remain close to those in Column (1). The number of observations increases from 4785 to 5397. The estimated coefficients imply a larger effect of costs on the timing than that in column (1). The key coefficients are joint significant at the 99% confidence level. This result may just be an artifact of how costs for already-built roads were estimated. If these costs are just set equal to actual costs, these may have risen over time, so that more recent roads would cost more.

Those regressions suggest that engineering costs estimated in 1958 predict the open-to-traffic time from 1958 to 1993. Engineering costs are only one of the determinants of open-to-traffic time. Our model ignores many other factors that determine open-to-traffic time. Because of the potential correlation between engineering costs and those omitted variables, such as local price levels, our regressions are likely to under-estimate the effect of costs on open-to-traffic time.

2.4.2 Second-stage regression.

The 1st-stage regressions predict the open-to-traffic time of each segment. Based on the prediction of these regressions, we calculate the fitted length of highways open to traffic on any given intercity route in 1962, 1968, and 1973. For 2nd stage regressions, we regress the driving times against the actual length of highways open to traffic among cities, instrumenting the actual length with the fitted length of highways open to traffic.

A difficulty is that we have no data on construction costs for some segments of highways. Chapter 1 constructed the driving times and routes connecting cities for 21 cities[34]. There are altogether 210 pairs of cities, but we only have information on highway segments for 69 pairs of cities. Further, for those 69 city pairs, we know the costs for only some of the segments. For example, the distance between New York and Harrisburg is 162 miles. We have data on costs for 64 miles. The percentage available is 39.5%. Table 21 gives a sense about the information available for those 69 pairs of cities. 2 pairs of cities have information available for less than 20% of the mileage, 31 pairs have information available for 40%--60% of the mileage, and only 2 pairs have information available for more than 80% of the mileage.

[34] Chapter 1 uses only 16 pairs of cities. We, however, calculated the driving times among 21 pairs of cities available in both Rand McNally Road Atlas and CTS (1967) and then dropped 5 cities are not available in either CTS (1963) or CTS (1972). So we do have data on 21 cities.

We assume that the predicted speed of construction on available segments is an unbiased estimate of the predicted speed on other segments of the same route. For example, we only know the costs of 64 miles out of the 162 miles between New York and Harrisburg. Our fitted values from the first regression imply that 0 of these 64 miles should have been open to traffic in 1962, 7 miles in 1968 and 63 miles in 1973. We assume that the fitted speed of construction is the same for the other 98 miles. As long as factors causing missing information are uncorrelated with changes in driving times, the estimated impact of IH on driving times is unbiased.

Table 21: Percentage of Mileage with Data on Costs

% Available	# City Pairs
<20%	4
20%-40%	22
40%-60%	31
60%-80%	10
80%-100%	2
Total	69

Source: *Interstate Cost Estimate 1958*, provided by the Federal Highway Administration.
Notes: For example, the distance between area 3 and area 8 is 162 miles. We have data on costs for 64 miles. Therefore the % available is 64/162=39.5%.

Table 22 and Table 23 report the results of 2SLS regressions. All regressions except for regression (4) of Table 23 use data on 69 pairs of cities for two periods (1962-1968 and 1968-1973). The number of observations is 138. Table 22 reports the results of the 1st stage regressions and Table 23 reports those of the 2nd stage. For Column (1) of Table 22 and Table 23, we first estimate the regression of open-to-traffic time on engineering costs, controlling for state dummies. We form the fitted open-to-traffic time by holding state dummies fixed at their sample mean and using variation only in the costs. Then we get the fitted mileage of IH open to traffic by 1962, 1968 and 1973, respectively. In Column (1) of Table 22, we regress the actual length of IH on each route on this fitted mileage. The R-squared is equal to 0.79. The key coefficient is 0.50, which is significant at the 99% confidence level. When building highways between two cities is expensive, state governments delay the construction. However, we expect the first stage coefficient in Table 22 to be close to 1, while it is only 0.50, which implies that many other factors in addition to engineering costs affect the open-to-traffic time. Column (1) of Table 23 finds that each mile of IH decreases driving time between a pair of cities by 0.009 hours or 0.54 minutes, significant at the 99% confidence level. The coefficient is reasonable, because it implies that, for example, building IH between cities of a distance of 200 miles apart would decrease driving time by 1 hour and 48 minutes. In Column (2) of Table 22 and Table 23, we estimate the regression of open-to-traffic time on engineering costs, controlling for state dummies. We form the fitted open-to-traffic time using variation both in the costs and the state dummies. Then we get the fitted mileage of IH open to traffic by 1962, 1968 and 1973, respectively. The R-squared and key coefficient are the same as those in Column (1) of Table 22. Column (2) of Table 23 estimates a similar effect of IH on driving times as Column (1).

Column (3) in Table 23 is the OLS regression. The estimated key coefficient is 0.008, less than the estimated 2SLS coefficients. The OLS regression under-estimates the effect of highways on driving time, possibly because of an endogeneity problem that rich regions may build more highways, but may also generate more traffic congestion.

Table 22: Engineering Costs and Open-to-traffic Time

Dependent Variable: Length of new Interstate Highways	(1) Engineering costs	(2) Engineering costs and state dummies
Length of new Interstate Highways	0.50	0.50
(estimated from engineering costs)	(16.85)***	(16.88)***
Time dummy	-223.07	-223.52
=1 if the change is from 1962 to 1968	(-4.73)***	(-4.74)***
Constant	193.74	194.05
	(9.73)**	(9.76)***
#obs	138	138
R-square	0.79	0.79

Sources: *Interstate Highway Construction Records* (1937-1993) and *Interstate Cost Estimate 1958*, provided by the Federal Highway Administration.

Table 23: The Effect of New Interstate Highways on Driving Times among Cities

Dependent variable: Change in driving times (hours)	Instrument Engineering Costs only (1)	Instrument Engineering Costs and State dummies (2)	OLS Segments available in ICE (3)	Instrument >40% of mileage available in ICE Engineering Costs only (4)	OLS All segments of the IHS (5)
Length of new Interstate Highways	-0.009 (-6.25)***	-0.009 (-6.24)***	-0.008 (-7.19)***	-0.009 (-4.27)***	-0.011 (-8.90)***
Year dummy =1 if the change is from 1962 to 1968	3.39 (3.91)***	3.38 (3.91)***	3.20 (3.99)***	3.58 (2.88)**	3.02 (4.28)***
Constant	-4.47 (-8.31)***	-4.46 (-8.32)***	-4.56 (-8.87)***	-4.50 (-5.98)***	-3.55 (-6.80)***
#obs	138	138	138	88	138
R-sq (within)	0.28	0.28	0.28	0.25	0.37

Sources: *Interstate Highway Construction Records* (1937-1993) and *Interstate Cost Estimate 1958*, provided by the Federal Highway Administration. *Rand McNally Road Atlas* (1962, 1968, 1973), provided by Rand McNally.
Notes: 1. The dependent variable is the change in driving times (hours) from 1963 to 1968 and from 1968 to 1973.
2. Column (1) uses engineering costs as the only instrument. Column (2) uses both state dummies and engineering costs as instrument variables.
3. Many segments of IHS are unavailable in ICE. ICE provides information on only 17,871 miles out of the 42,000 miles of the whole IHS.

Column (4) and Column (5) in Table 23 examine the problem that missing information on costs for many segments may affect the estimated coefficients. These two regressions provide mixed information on the problem of missing information. Column (4) repeats the specification of column (1), but using only observations on routes with more than 40% of the mileage available in ICE. The key estimated coefficient in Column (4) is the same as in Column (1), implying that the speed of construction on available segments may be representative of that on all segments. In contrast, Column (5) repeats the OLS specification of column (3), using mileage for all segments rather than just mileage on ICE segments, and finds that missing information makes a difference to the estimated coefficient.

In sum, high engineering costs predict a delay in IH construction, because state governments built cheap segments earlier than expensive segments. Therefore, we can instrument the change in driving times among major cities using engineering cost estimates from 1958. Regressions suggest that OLS regressions under-estimate the effect of new highways on driving times.

2.5 Conclusion.

This paper proposes an instrumental variable for investment in infrastructure, focusing on the construction of the Interstate Highway System. We explored the natural experiment that a looming financial crisis drove states to build Interstate Highways as fast as possible. Both historical and regression evidence suggest that state governments built cheap segments earlier than expensive segments. We instrument the open-to-traffic time of highway segments using engineering costs and find that an OLS regression under-estimates the contribution of new highways to the reduction in driving time.

Moreover, this paper provides a new instrument for research on infrastructure. If construction cost estimates for highway segments are available, then finding an instrumental variable for the timing of construction is possible. This new instrumental variable can be related to driving times, which are a direct measure of the contribution of infrastructure, giving researchers the opportunity to address many problems in the literature estimating the contribution of public investment to regional development. A limitation of this instrumental variable is that it applies to only the period of the financial crisis of the Highway Trust Fund, but not other periods.

Appendix: Data Sets

I constructed the data set of driving times using maps of driving times from the Rand McNally Road Atlases of 1962, 1968, and 1973 (Rand McNally). Then I found the starting and ending mileage posts of each segment of highway, and merged the data sets of driving times with the Interstate Highway Construction Records (1938-1993) (IHCR) and Interstate Cost Estimate from 1958 (ICE).

1. Constructing the data set of driving times and driving distances.

Rand McNally provides a map of driving times among contiguous major cities in the United States. Rand McNally calculates the driving times using road mileage, topography, speed limits, and congestion. Driving a truck during the night will require less time than what is listed in the Atlases. The Rand McNally from 1962 says

> *"Driving time shown is approximate under normal condition. Consideration has been given to topography, speed laws, and congested areas. Allowances should be made for night driving and unusually fast or slow drivers."*

I calculated the shortest driving times among the 21 cities in the Commodity Transportation Survey (1967)[35]. I first found all the possible routes connecting two cities and their corresponding driving times, and then found the route with the shortest time. The routes of shortest driving times may change from year to year because new roads open to traffic.

Since the Rand McNally data provide driving times only between contiguous cities, I must assume that routes between distant cities go through major cities in between. For example, I

[35] Commodity Transportation Survey (1967) has 25 cities, but only 21 cities are available on the map.

assume that if you drive from Seattle to Cleveland in 1962, you will pass through Chicago. This constraint will not affect the estimated driving times much because routes from one city to another usually do go through the major cities between them, as shown in Yahoo! Map. For example, Yahoo! Map also finds that the best route for driving from Seattle to Cleveland goes through Chicago.

I find the driving distances on the chart of driving distances in Rand McNally. The driving distances remained almost the same from 1962 to 1973.

2. Merging the Rand McNally driving times with Interstate Highway Construction Records and Interstate Cost Estimate.

The following example illustrates. Driving from St. Louis to Detroit took 9 hours and 55 minutes in 1973. I am interested in the lengths of IH that opened to traffic in 1973 and the engineering costs of each IH segment. Thus, I need to merge the data of driving times with IHCR and ICE. I had several problems in merging these data sets. First, I knew nothing about the routes between cities other than the cities that might be between them on that route. So I used Yahoo! Map to find more precise routes in 2006. For example, Yahoo! Map gives the following driving directions from the center of St. Louis to the center of Detroit.

1. Starting in St. Louis on Tucker Blvd go toward Walnut Str-go 0.4 mi
2. Turn R on Washington Ave-go 0.5 mi
3. Turn L on N 4th St.-go 0.1 mi
4. Turn R on Dr. Martin Luther King Jr. Mem. Brg-go 1.5 mi
5. Take L ramp onto I-70 East-go 100.4 mi
6. Take L fork onto I-57 North toward Chicago-go 181.3 mi
7. Take exit #345A onto I-80 East toward I-294/Indiana-go 27.1 mi
8. I-80 East becomes I-94 East-go 242.9 mi
9. Take exit #213B onto I-96 East-go 1.3 mi

10. Take the M-10 L exit toward Civic Center-go 2.2 mi
11. Take the Larned St. L exit toward Cobo Center-go 0.5 mi
12. Turn L on Woodward Ave. [M-1]-go 0.2 mi
13. Arrive at the center of Detroit, MI.

Using the state maps of the Rand McNally Road Atlas 2006, I found that the major IH from St. Louis to Detroit are I-70 and I-57 in Illinois, I-94 in Indiana, and I-94 in Michigan. If any of those four IH had zero mileage open before 1973, I dropped that IH and assumed that drivers took local roads of similar distance along that segment.

Then I found the starting and ending mileage post of those four IH segments, since IHCR identifies highways segments with starting and ending mileage posts. The mileage is labeled 0 on the southern border of a state for highways going south to north, and labeled 0 on the western border for highways going west to east. I counted the mileage of each point of changing routes using state maps in Rand McNally Road Atlas. For example, a driver can change IH from I-70 to I-57 at Effingham, Illinois, and can change from I-57 to I-80 at Tinley Park, Illinois. To find the starting mileage posts of I-57, I added up the length of each segment labeled in the state map from the southern border of Illinois to Effingham, Illinois. To find the ending mileage posts of I-57, I added up the length of each segment from the southern border to Tinley Park, Illinois. After finding the starting and ending mileage posts of each IH segments, I recorded the following in an Excel file.

From St. Louis to Detroit			
State	IH	Starting Mileage Post	Ending Mileage Post
Illinois	I-70	0	76
Illinois	I-57	167	348
Indiana	I-94	0	46
Michigan	I-94	0	195

Using the open-to-traffic time of IH segments in IHCR, I got the lengths of highways that opened to traffic in 1962, 1968, and 1973.

The next step was to merge IHCR with ICE, which also presented difficulties. First, ICE has a different number system for IH from the standard numbering of Interstate Highways. For example, each state may name I-95 differently, so I had to find the corresponding Interstate route number for each highway in each state. Second, the plan of some IH changed after 1958 and thus the length of each IH changed significantly, making it difficult to identify those segments in ICE, because the average length of segments in ICE is only 3.21 miles. If the mileage of IH in ICE is over plus-or-minus 10 miles the actual mileage, I considered it inaccurate to identify ICE segments on that IH and dropped that IH. Fourth, for routes starting in big cities, distinguishing their starting points from city bypasses is difficult because the state maps of ICE were ambiguous. For example, you can take two different exits from I-495 to I-270 in Maryland, so I could not identify which exit is the starting point of I-270. I dropped routes that may lead to errors over 10 miles.

Using the starting and ending mileage posts of segments in ICE and in routes connecting cities, I merge the data set of driving times and ICE 1958 with IHCR (1938-1993).

Bibliography

Aaron, Henry J., “Comments on ‘Why is Infrastructure Important?’” in Alicia H. Munnell (ed.) *Is There a Shortfall in Public Capital Investment?* Federal Reserve Bank of Boston, Boston, 1991.

Anderson, James E., and van Wincoop, Eric, “Gravity with Gravitas: A Solution to the Border Puzzle,” *American Economic Review*. vol. 93, 2003. pp. 170-192.

Anderson, James E., and van Wincoop, Eric, “Trade Costs,” *Journal of Economic Literature*. vol. 42, 2004. pp. 170-192.

Aschauer, David A., “Government Spending and the ‘Falling Rate of Profit’”, *Economic Perspectives*. Federal Reserve Bank of Chicago. vol. 13, 1988. pp. 17-25.

Aschauer, David A., “Is Public Expenditure Productive?” *Journal of Monetary Economics*. vol. 23, 1989. pp. 177-200.

Aschauer, David A., “Public Spending for Private Profit”, *Wall Street Journal*. March 14, 1990.

Ashenfelter, Orley and Michael Greenstone, “Estimating the Value of a Statistical Life: The Importance of Omitted Variables and Publication Bias,” *American Economic Review*. May 2004, v. 94, iss. 2. pp. 454-60.

Baum-Snow, Nathanial, “Do Highways Cause Suburbanization?” *Quarterly Journal of Economics*, forthcoming.

Chow, Garland, “US and Canadian Trucking Policty,” in Kenneth Button and David Pitfield eds., *Transport Deregulation: an international movement*. New York: St. Martin's Press, 1991.

Duffy-Deno, Kevin T. and Randall Eberts, “Public Infrastructure and Regional Economic Development: A Simultaneous Equations Approach,” *Journal of Urban Economics*. Vol. 30, 1991. pp. 329-343.

Duflo, Esther and Rohini Ponde, “Dams,” *Quarterly Journal of Economics*, forthcoming.

Federal Highway Administration, “Public Roads”, Special Edition 1996. p. 11.

Federal Highway Administration, “The Dwight D. Eisenhower System of Interstate and Defense Highways”, 1998.

Federal Highway Administration, “50th Anniversary of the Interstate Highway System,” available at http://www.fhwa.dot.gov/interstate.

Fernald, John G., “Roads to Prosperity? Assessing the Link Between Public Capital and Productivity,” *American Economic Review*. Vol. 89(3), 1999. pp. 619-638.

Highway Statistics, “The National System of Interstate and Defense Highways: Status of Improvement as of September, 1964”, Public Roads Administration, Federal Works Agency. 1965.

Highway Statistics, Public Roads Administration, Federal Works Agency. 1973.

Hulten, Charles R. and Robert M. Schwab, “Public Capital Formation and the Growth of Regional Manufacturing Industries,” *National Tax Journal.* Vol. 44(4), 1991. pp. 121-134.

Jorgenson**,** Dale W., “Investing in Productivity Growth”, Technology and economics: Papers commemorating Ralph Landau's service to the National Academy of Engineering. Washington, D.C.: National Academy Press. 1999. pp. 57-63.

Knight, Brian, “Parochial Interests and the Centralized Provision of Local Public Goods: Evidence from Congressional Voting on Transportation Projects,” *Journal of Public Economics*. vol. 88(3-4), 2004. pp. 845-866.

Limão, Nuno, Anthony J. Venables, “Infrastructure, Geographical Disadvantage, Transport Costs and Trade,” *The World Bank Economic Review*. Vol. 15, 2001. pp. 451-479.

Martinez-Zarzoso, Inmaculada and Celestino Suarez-Burguet, “Transport Costs and Trade: Empirical Evidence for Latin American Imports from the European Union,” *Journal of International Trade and Economic Development.* Vol. 14(3), 2005. pp. 353-371.

Metz, W. L., and Joyce Ritter, “Building the Interstate. Federal Highway Administration,” 2006. Available at http://www.fhwa.dot.gov/infrastructure/build.htm

Michaels, Guy, “The Effect of Trade on the Demand for Skill---Evidence from the Interstate Highway System,” CEPR DP 6056, 2007.

Rand McNally Road Atlas. Rand McNally. 1962.

Tornqvist, Leo, Pentti Vartia and Yrjo O. Vartia, “How Should Relative Changes be Measured?” *The American Statistician.* Vol. 39 (1), 1985. pp. 43-46.

Schultze, Charles E. “The Federal Budget and the Nation's Economic Health,” in Henry J. Aaron (ed), *Setting National Priorities: Policy for the Nineties*. The Brookings Institute, Washington. 1990.

US Census Bureau, 2005. State and Local Government Finances by Level of Government and by State: 2001 - 02. Available at http://www.census.gov/govs/estimate/0200ussl_1.html

Weingroff, Richard, “Creating the Interstate System: The Federal-Aid Highway Act of 1956,” *Public Roads*. Special Edition, June 1996. pp.9-16.