The US Interstate Highway's Effect on Agglomeration

Thor Dodson

March 15, 2021

Abstract

The US Interstate Highway System had a significant impact on market accessibility and transportation costs between regions. Whether this should lead to increased agglomeration of economic activity due to increased 'economic centripetal forces' or a dispersal from 'centrifugal forces' depends on factors that differ by industry. This study suggests the impact depends on truck transportation utilization and backward linkages. Travel time estimates constructed by representing the US highway system as a network over time and data on the spatial inequality of earnings are used for a panel estimation with interactions, individual and time effects, and regional variation for identification.

Keywords: transportation, agglomeration, spatial inequality, new economic geography, United States

JEL Codes: N72, R12, R41

A thousand thanks to my advisors Robert Halvorsen and Levis Kochin for their patience and pushes, Yoram Barzel for hosting seminars and listening to too many ideas, my colleagues Matt Daniel, Kwong-Yu Wong, Andrew Boomer, Kris Overbo, Shuo Jiang, Junwei Mao, and Qing Zhang for help in every dimension, Richard Weingroff from the FHWA for his help clarifying details on the PR-511 data, Interstate financing, and providing completion progress maps, Mauricio Ortiz and Levi from the BEA for their help calculating the spatial inequality indices, Mike Babb for his assistance in learning how to use GIS and network packages, undergraduates Yean Kim, Nick Hall, Sin Myung Han, and Sainyam Goel for help with data work, and to my family and friends for their support throughout this process.

1 Introduction

By altering the landscape of transportation costs road systems facilitate the agglomeration as well as dispersal of industries. A long literature exists examining the impact of roads on the spatial distribution of economic activity (Chandra and Thompson, 2000; Michaels, 2008; Rothenberg, 2011; Redding and Turner, 2014; Frye, 2016), but uncertainty remains about how specific industries respond and what are the characteristics influencing their response. Understanding the impact of road infrastructure is important for regional policy makers as the spatial distribution of the amount and type of earnings has lasting implications for structural inequality and regional divergence (Redding, 2005; Paredes et al., 2016; Niehbuhr et al., 2012).

A strand of literature on agglomeration describes economic centripetal and centrifugal forces that influence the relative locations of firms (Marshall, 1890; Fujita et al., 1999; Cook et al., 2007; Pelegrín and Bolancé, 2008). These forces are difficult to directly measure, but industry characteristics creating sensitivity to the forces can be used as proxies to predict the response. As the transportation costs change, the ways in which industries are sensitive to the affected forces will influence the changes in spatial distributions.

In this paper I examine how the US Interstate Highway System impacted the spatial distribution of different industries and characteristics that can explain the varying responses. Based on location theory and the benefits of agglomeration and dispersal, I suggest that industries with a higher truck transportation share of inputs and backwards linkage measure are more likely to disperse in response to the reduction in travel time.

To measure the effect of the Interstate Highway System I construct a novel data set of travel times between metropolitan regions in the US for each year between 1950 and 1993 using the completion dates of road segments to build edge weighted networks. The travel time is an important component of the transportation cost between regions affecting the price paid to drivers, supply timing, and inventory holding requirements. I add to the literature examining detailed road data (Rothenberg, 2011; Faber, 2014; Donaldson and Hornbeck, 2016; Alder, 2016; Jaworski et al., 2018; Morten and Oliveira, 2018) with my travel time estimates and the methodology. By looking at the road system as a network with weighted edges the marginal benefit of specific roads on travel times throughout the system can be observed and used to examine many questions, although this paper focuses on the impact of the aggregated changes on agglomeration.

Using data on county level earnings by industry in the US I construct a spatial GINI index measuring how unequal the distribution of economic activity is across all counties for each year. This index reveals how clustered or agglomerated different industries are and is commonly used in research on spatial distribution (Rey and Smith, 2012; Sutton, 2012; Panzera and Postiglione, 2019). This index does not tell us about the exact distribution of activity, as multiple distributions can lead to the same spatial GINI, but changes in the spatial GINI do tell us whether industries are becoming concentrated into fewer counties or spreading out. This measure of agglomeration does not speak to location within counties (Börjesson et al., 2019), nor does it speak to specialization within industries which is another common indicator of agglomeration (O'Donoghue and Gleave, 2004).

I use a panel data set with interaction effects to detect the industry varying effect the change in the travel time index has on the spatial GINI index. I perform robustness checks including adjusting for county area, alternate measures of spatial inequality, additional controls, and alternate regression specifications. I conduct simulations with artificial data verifying the appropriateness of the preferred specification given the likelihood of lags and leads in response.

Additionally I exploit regional variation in the timing and magnitude of road completion to estimate the causal effect conditional on region, industry, and time effects controlling for unobserved variables. The eight regions are as defined by the U.S. BEA for economic comparison[1]. Due to regional factors orthogonal to the change in location of industries such as varying state institutions, weather, terrain, and construction delays, different regions

[1] `https://www.icip.iastate.edu/maps/refmaps/bea`

completed their roads at different times. If regions that built their roads earlier also observed a change in spatial GINI earlier, than it is likely the change is caused by the roads.

I find that industries with a higher trucking share of inputs and a higher backwards linkage disperse more when travel times are reduced. The average highway travel time between metropolitan regions decreased by about 18%, with varying declines across regions. The spatial GINI for total personal income declined slightly between 1969 and 1985, but rose to its previous level by 2000 with little change afterward, while the spatial GINI for population declined slightly until 1980 and has been slightly increasing ever since. This combined with the significant movements in industry specific spatial GINI suggest there is not a large change in the overall spatial distribution of economic activity, but there is significant relocation of where specific types of industry occur.

The paper proceeds as follows: Section 2 discusses the theory of why different industries will respond differently to an improvement in the road system, Section 3 describes the data and methodology, Section 4 reports and discusses the estimation results, and Section 5 concludes.

2 Theoretical Background

Roads alter the time it takes to traverse an area, effectively warping space and bringing regions closer together by facilitating the faster movement of cars and trucks. This reduction in travel time lowers the cost of moving goods by lowering the wage paid to the drivers, reducing the uncertainty associated with waiting, facilitating smoother production flows, and reducing required inventories as stocks or parts can more quickly be replenished. The last three effects are particularly important, as observed in the global rise of "just-in-time" manufacturing and inventory management during the 1970s and 80s (Sayer, 1986; Brox and Fader, 1997), as well as the premium placed on overnight shipping (Stecke and Zhao, 2007). Although rail and water can typically transport materials at a lower cost per unit, the speed offered by roads is crucial for supply coordination, and the access provided by roads to regions not adjacent to rail or water necessitate their use for the 'first and last mile' for intermodal shipping.By providing access for vehicles and lowering the cost of transportation between regions, roads play a crucial role in shaping the location decisions of firms.

Agglomeration is the clustering of economic activity in space. This applies to multiple scales, including countries, cities, and districts. The benefits of agglomeration are aptly summarized by Marshall (1890) who points to three sources: 1) knowledge spillovers—the idea that information is "in the air" and technical processes and innovation are propagated through proximity by increased interactions, 2) pooled labor—the increased matching of needs to skills for employers and employees from both having access to a larger pool, 3) forward and backward linkages—the reduced costs from proximity to markets and sources of inputs, as transport is costly. The third type is the most explored by the new economic geography and 'market access' literature (Fujita et al., 1999; Duranton et al., 2014; Donaldson and Hornbeck, 2016). We can think of these benefits as 'centripetal forces' that pull activities towards each other, resulting in clustering. However, being near other firms has a trade-off—wagesand the price of land are pushed up due to competition, acting as 'centrifugal forces' pushing firms to locate away from clusters. Furthermore, proximity to multiple sources of demand and inputs may be a relevant consideration pushing a plant away from any particular market center and towards a point of centralized distribution, as elaborated by Weber's (1909) point of minimum transport.

Different industries have different sensitivities to each of these forces based on what they do and how they do it. Thünen (1826) captured this idea with his model of agricultural land use and this was extended by Alonso's (1960) bid-rent theory; an example of which is shown in Figure 1. The key idea is how much 'land rent'[2] an industry is able to generate at a particular location, based on the difference between the value of their product at the market and the costs of inputs and transportation incurred from operating at that position. Industries' that

[2]Thünen defines land rent as value generated in excess of all input costs, although there are some competing definitions of this concept

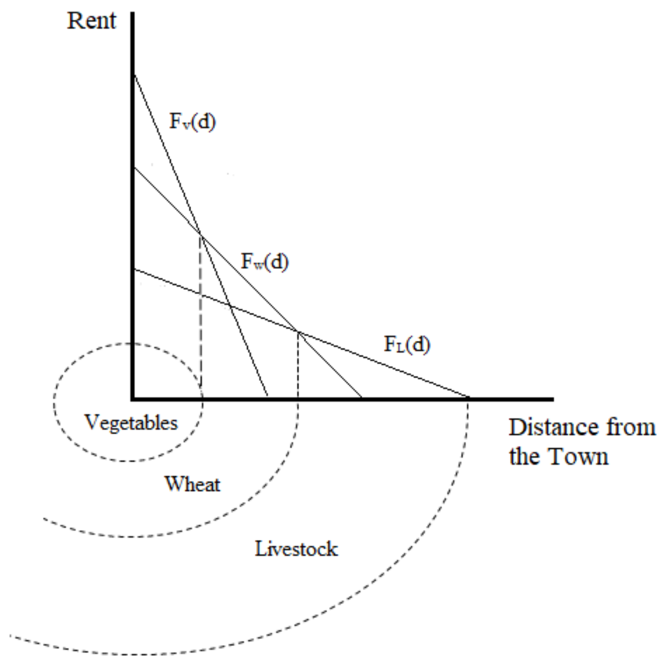

Figure 1: A Bid-Rent Curve

generate a higher rent for any given location are more likely to locate there since they can outbid other types of industry. In its simplest form we conceive a single market existing at a point in a uniform plane where economic activity can take place, but it can be extended to incorporate multiple market centers and surfaces with varying transportation costs such as a river or road system. For the single market framework, the vertical intercept represents the rent an industry can offer for being at the center of the market—the point where the benefits of agglomeration are the highest, and the slope represents how the rent an industry can offer changes with distance from the market—a combination of the transportation cost for that industries' product and how the total cost of inputs changes with distance. Industries that benefit from agglomeration tend towards the market, and industries with goods that can be moved cheaply tend to be pushed away from the market. In a multiple market framework this is more complicated as firms within industries may choose to deal with just one market or multiple markets, but still we would observe that industries benefiting more from agglomeration would tend towards market centers and industries with costs that decline more rapidly with distance would locate away from market centers. In reality markets do not operate at single points in space, but the same logic applies for distributed markets as long as there is some varying concentration of market activity across space.

From this lens, an improvement in the road system does two things. By lowering the cost of transporting materials, the slope of the bid-rent curve is flattened as it is less costly to be located away from the market center. This effect pushes industries outward from market centers and makes more distant locations viable points of operation. However, an improvement in the road system also facilitates increased access to a market center as customers and employees from a wider radius can commute in. This increases the agglomeration benefits of an area by creating a larger labor pool firms can pull from, increasing the suite of interactions that lead to knowledge spillovers, and increasing market accessibility. Effectively, the market center becomes larger and has increased capacity for agglomeration. By lowering transportation costs and facilitating access, improved roads push some industries out and pull other industries in.

Industries that have a larger truck transportation share of inputs benefit more from the decline in transportation costs due to improvements to the road system. While the reduction in transportation costs reduces the slope of the bid-rent curve for all industries, the slope becomes more flat for industries that utilize trucking more. This makes it comparatively less costly for these industries to be farther away and hence pushes them outward, away from the market centers/central business districts. Based on this, we suggest the main hypothesis of the paper—that the coefficient on the interaction term between travel time and truck transport share of inputs will be positive.

The stage in the product life-cycle is an additional influence on the sensitivity to the benefits of agglomeration and dispersal (Eriksson et al, 2020). The conception of a product life-cycle distinguishes four stages in a product's life: introduction, growth, maturity, and saturation. The first two and last two can be grouped together as early and late respectively. Early stage products involve design, the supply chain is not well formed, demand must be created, and there is low competition; thus they benefit more from the knowledge spillovers and access to pooled high skilled labor of agglomeration. Late stage products face high competition and low prices, deal with complex supply chains and mass production, and profitability/survival is more based on production/distribution efficiency— thus they benefit more from the lower wages, cost of land, and centralized distribution offered by dispersal. When the road is improved both agglomeration and dispersal are further facilitated, exasperating the location preferences for both early and late stage products. A direct measure of life-cycle stage is not available but backwards linkage, the total increase in production stemming from an increase in the final demand for a particular industry because of the additional inputs required to produce it, the additional inputs required to produce those, and so on, is a reasonable proxy. If an industry is in late stage production with a complex supply chain involving many industries as inputs, this will appear as a higher number in this measure, as late stage industries tend to have lower profit margins from the high competition. Because the inputs and outputs are measured in dollars, as the price of the output decreases from increased competition the ratio of inputs to outputs will be higher, therefore for a given increase in output there will be a larger increase in inputs, and hence a higher measure of backward linkage. This effect could be mitigated if the industry inputs are moving through the life-cycle at the same time and undergoing a similar process, or if the reduction in industry input use from increased efficiency is greater than the reduction in price from increased competition.

In summary, because of the differing effects of centripetal and centrifugal economic forces on industries, when the road is improved we suspect that industries that utilize trucking more will disperse, industries categorized as information services and material transformations will agglomerate and disperse respectively, and industries dealing in early and late stage products will agglomerate and disperse respectively.

3 Data and Method

The Interstate Highway System began construction in 1956, although the call for an updated national highway system had been building since the 1930's (Weingroff, 2017). While there was already a sizeable road system in place and most places could be accessed, the road conditions were often poor[3], many of them unpaved. The Interstate standards enabled high speed travel due to the quality of the surface, the curvature, sight distance, grade and superelevation design restrictions, the minimum of two lanes in each direction separated by a median, and the limited access restriction with no stop lights or driveways. In 1955 the US had around 3,418,214 miles of public roads (US DOT, 1985), and although only 48,440 miles were eventually constructed as part of the Interstate System it carries about 20% of the nation's traffic (Weingroff, 2006).

The Interstate Highway System can be viewed as accomplishing two things: 1) connecting and providing or improving access to regions, 2) lowering the cost of moving goods and people through reductions in travel time and facilitating larger trucks. The key statistic I utilize is the average transportation time between metropolitan regions for each year of its construction.

I build an edge-weighted network representation of the US road system for each year between 1950 and 1993 as the Interstate Highway System was developed and use this to estimate the travel times between metropolitan statistical areas with a shortest path algorithm. I do this by combining two geographic information system (GIS) road files and converting them to an edge-weighted network that the Dijkstra algorithm can be performed on.

The first GIS file is formed by isolating the interstate highways from the PA_NHS 2012 shapefile[4] detailing all US roads at that time. The second GIS shape file I form by manually tracing a 1954 map image[5] produced by the US government detailing the principle highways

[3] see Figure 6 in the appendix

[4] Accessed from the FHWA website,
https://www.fhwa.dot.gov/policyinformation/hpms/shapefiles_2017.cfm

[5] https://www.raremaps.com/gallery/detail/38608/a-pictorial-map-of-the-united-states-of-america-show

and arterials in existence at that time, what we refer to now as the US numbered highways. I approach the road system in this way because in addition to entirely new roads the Interstate Highway System replaced many segments of the previous highway system, so many portions of Interstate Highways were still beneficial although the entire road was not yet completed. This method does not include additional non-interstate highways that were constructed during this period, which biases the travel time reduction estimates downward.

Next, using the "PR-511" dataset, a construction log[6] detailing the completion date of each Interstate segment, the active segments of Interstate Highway are overlaid with the pre-existing highway system to construct a representation of the total highway system for each year between 1953-1994.[7]

With the highway system in place and converted to a network, the Dijkstra algorithm[8] finds the shortest weighted path between any two points in the network to estimate the travel time for each year. The weights on each road segment are the travel time based on the distance and speed. 65 mph is assumed for Interstate Highways; 50 mph is assumed for the non-interstate highways, differing slightly from the assumptions made in Jaworski et al (2018)[9]. This is done for every metropolitan-statistical-area (MSA) pair to generate a travel time matrix for each year. Figure 2 shows the average of this travel time[10]index matrix for each year. On average the Interstate Highway System reduced travel times between MSA's by about 18%, although the actual reduction in travel time (unobserved) is partially due to vehicle improvements[11].

[6]This dataset was digitized and made available by Baum-Snow (2007), available here `https://www.dropbox.com/s/wq5cp6gm4ocxjo4/CD-ROM.rar?dl=0`

[7]The PR-511 has a range of statuses 1-6. Status 1 is fully complete and up to standards. Status 2 is mostly complete and open to traffic, and this is the measure of completion used.

[8]I use the python modules 'networkx' to shape the network, and 'igraph' to implement the Dijkstra.

[9]These speed assumptions are a simplification based on travel time estimates provided by AAA maps from 1955, 1996, and 2018, to isolate the speed changes from the road and vehicle improvements. Routes without an interstate segment experienced a rise in speed of about 5mph, likely from improvements in car technology, while routes receiving interstate segments experienced rises in speed between 10-20mph, with variance likely due to congestion. Thanks to John King for providing his personal copy of the 1955 AAA map.

[10]The units are coordinate distance per mph

[11]There were notable policy changes during this period—the National Maximum Speed Law established in 1974 and the Motor Carrier Act of 1980. I dismiss the National Maximum Speed Law as it was reportedly

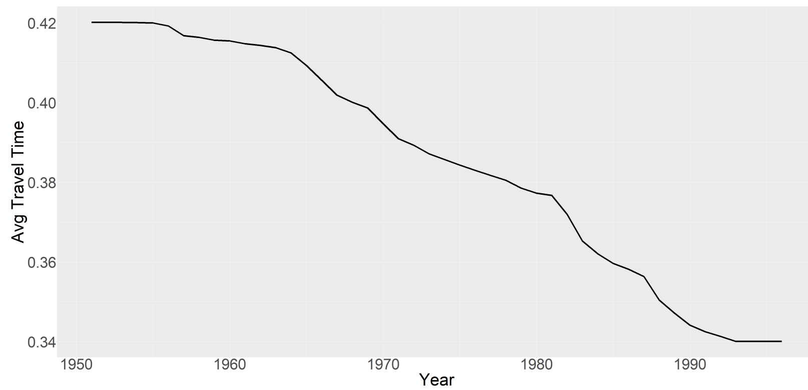

Figure 2: Average Travel Time between MSA's

For the regional travel time estimates, I take the average travel time from each county within that region to every other county within 425 miles, chosen based on the average distance for over-night shipping availability. This number reflects the travel time within the region and to counties near the edge of the region. The regional travel times are normalized to the national travel times to facilitate comparison of regression coefficients.

My data adds to the literature explicitly representing road systems as a network of transportation costs, such as Rothenberg (2011) who utilizes a mapping between road quality and speeds to estimate the travel time changes in Indonesia, Faber (2014) who constructs least cost path spanning tree networks examining China's National Trunk Highway System, Donaldson and Hornbeck (2016) who calculate lowest-cost county-to-county freight routes in the US, Alder (2016) who constructs a grid of cells with different speeds to use a shortest path algorithm examining bilateral travel times in India, and Jaworski et al (2018) who utilize decennial maps with surface information, mileage, and travel time estimates to construct internal trade costs for the US. The benefit of my method is the level of detail at the annual level, allowing a wide range of travel times to be estimated and compared with other variables changing during this time frame. Furthermore, the regional variation in the timing

not followed or enforced. The Motor Carrier Act of 1980 deregulating the trucking industry had many impacts potentially lowering transportation costs, which would bias the estimate of the effect of the change in travel time on agglomeration upwards. Similarly, congestion from traffic is unaccounted for, which would bias the estimated travel times upward and therefore the effect on agglomeration downwards

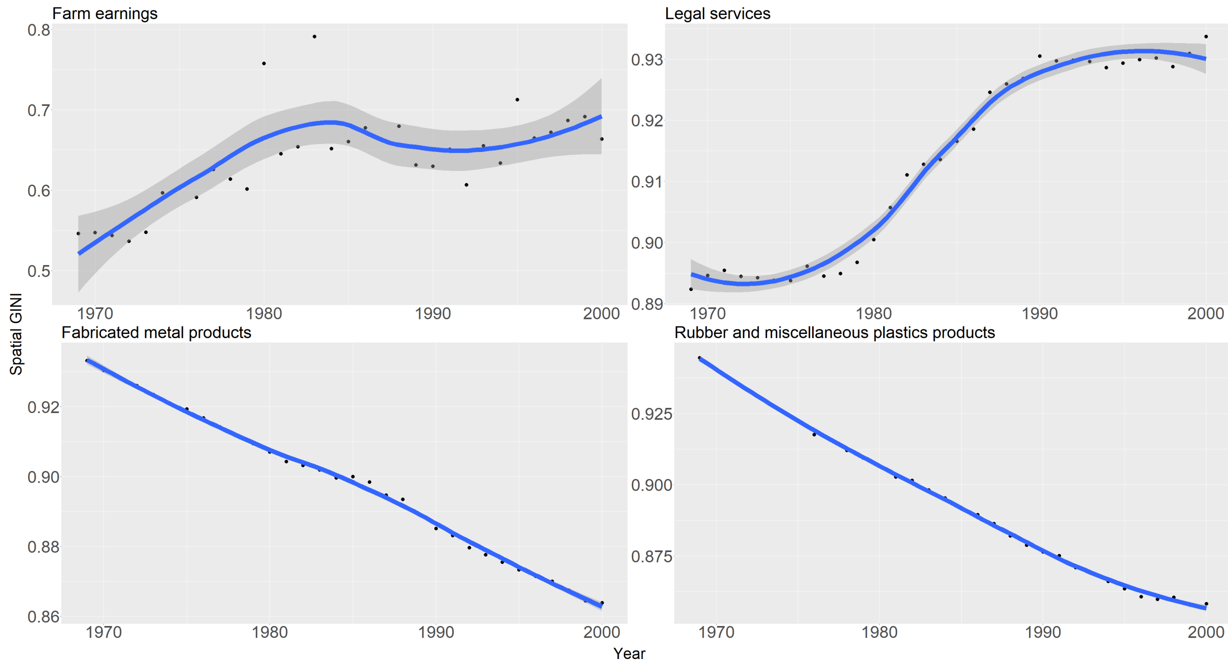

Figure 3: Spatial GINI for Select Industries

and magnitude of road completion can be exploited to parse out the causal effect based on the timing of response in the dependent variable, conditional on unobserved variables being controlled for through the use of effects.

Using detailed BEA data on county earnings by industry I construct a measure of spatial inequality over time using the same principle as the GINI coefficient of income inequality. For 50 SIC defined industries I calculate the cumulative share of the nation's income from each of the 3081 US counties, and arrange these in order of lowest to highest to form the distribution from which the GINI coefficient can be calculated. Figure 4 shows the distribution for farming and non-farming income in 1969. The non-farming income is more bowed in, revealing the income is concentrated in fewer counties, indicating a higher degree of spatial inequality. The spatial GINI is calculated for each industry for each year between 1969-2000[12] , some sample industries are shown in Figure 5. The earnings data are reported based on where the earner lives, so any commuting across counties will bias the estimate downward.

[12]For some industries negative earnings appear at times, distorting the range of the GINI. These observations are set to zero.

Figure 6 highlights the change in spatial GINI from 1969 to 2000 for the industries. A striking feature is the dispersal of industries like rubber and miscellaneous plastics, fabricated metal, lumber and wood products, stone, clay, and glass products, industry machinery and equipment, and miscellaneous manufacturing—all industries dealing with physical goods. On the other side we see the agglomeration of industries such as legal services, depository and nondepository institutions, and communications—industries that deal with information.

Some industries dealing with physical goods that go against this trend include oil and gas extraction, forestry and fishing, mining, textile mills, and coal production. But these are industries directly dealing with the extraction or cultivation of natural resources and may be tied to specific locations, and thus not as susceptible to the changing dispersal forces as much as the changing availability of sites from resources running out or being discovered. Another oddity is farming, which saw the largest increase in agglomeration of all the industries, but I suspect this is more from the farming specific technology changes from the Green Revolution than changes in the road system. Retail trade agglomerated while wholesale trade dispersed, aligning with the prediction of response from the road improvement based on their varying use of land and preference for centralized distribution. The dispersal of business services and insurance carriers highlights another tension–the benefits of proximity to information hubs and the benefit of moving to where the customers are located. As Hoover and Vernon (1959) discussed in their analysis of the distribution of people and jobs in the New York Metropolitan region certain operations, such as banking and life insurance, become standardized they find less of a need to be near the information sharing hubs, and more of a need to locate near their increasingly suburbanized customer base, especially as technology like the telephone and internet facilitate the exchange of information across distances. Information industries under this sort of influence may still disperse despite the increased ability to cluster.

Figure 4: Change in Spatial GINI

These spatial GINI estimates fit into the literature examining measures of spatial distribution including Rey and Smith (2012) who introduce a spatial decomposition of the GINI coefficient that exploits the contiguity matrix, Sutton (2012) who constructs spatial GINI from nighttime satellite imagery and population density, and Panzera and Postiglione (2019) who propose an index based on the GINI that introduces regional importance weighting.

The truck transportation share of inputs is calculated from the BEA input-output 'use' table, detailing each industries use of other industries in dollars. Ideally, we would like a measure that reflects how much an industry relies on truck transportation for both inputs and outputs, and it is not clear how this is attributed in the input-output table.[13]

The theory suggests that the stage in the product life-cycle will impact agglomeration and dispersal. As a proxy for these I utilize an industry measure of 'Rasmussen backward linkages'—the column sum of the 'Leontief inverse' or 'total requirements matrix' calculated from the input-output table of industry interactions. This measure reflects the total increase in production stemming from an increase in the final demand for a particular industry because of the additional inputs required to produce it, the additional inputs required to produce those, and so on.

3.1 Method

In this section I layout the methodology used, potential issues, and how I address them.

I utilize a fixed effect regression with interaction terms to test if changes in travel time change the spatial GINI and if differences between industries explain the differences in the change of the spatial GINI across industries. Furthermore, I construct 'meaningful' marginal effects and standard errors as in Brambor, Clark and Golder (2014), I verify the results are robust to alternate specifications and measures of agglomeration, and I estimate a regional

[13]The input-output use table uses the NAICS industry codes, while the BEA county earnings uses the SIC industry code. Industries were matched based on the US BLS concordance guide and unmatched industries were dropped.
`https://www.bls.gov/bls/exit_BLS.htm?a=true&url=https://www.census.gov/eos/www/naics/concordances/2002_NAICS_to_1987_SIC.xls`

form of the model to control for time effects and utilize variation in road construction.

I estimate a model of the following form:

$$\text{spatialGINI}_{it} = \alpha + \alpha_i + \beta_0 tt_t + \beta_1 ts_{it} + \beta_2 bl_{it} + \beta_3 tt_t ts_{it} + \beta_4 tt_t bl_{it} + \epsilon_{it}$$

where tt_t is the index of average travel time between MSA's, ts_{it} is the truck transportation share of inputs, and bl_{it} is the measure of backward linkage for industry i at year t.

If tt and ts are not endogenous to travel time, the effect of reducing travel time on the spatialGINI is

$$\frac{\partial \text{spatialGINI}_{it}}{\partial tt_t} = \underset{(-)}{\beta_0} + \underset{(+)}{\beta_3 ts_{it}} + \underset{(+)}{\beta_4 bl_{it}}$$

where the hypothesized signs for the coefficients are noted. Conditional on a trucking input share of zero and a backwards linkage of zero, we expect the reduction in travel time to lead to an increase in the spatial GINI, that is, agglomeration. For industries with a high trucking input share and high backwards linkage, this effect will be mitigated to the point of being reversed so that a reduction in travel time leads to a decrease in the spatial GINI, that is, dispersion. If trucking input share and backwards linkage are changing in response to the changes in travel time, the marginal effect is more complicated, but this is unlikely as the change in these variables across time is negligible (I explore this more in the appendix).

As detailed in Brambor, Clark and Golder (2014), when including interaction terms for testing conditional hypotheses, care must be taken in the implementation and interpretation of the results. Specifically, the constitutive effects must be included and must not be interpreted as unconditional marginal effects, and 'meaningful' marginal effects and standard errors should be reported. That is, for the specification above, the appropriate standard error formulation for the marginal effect of travel time is shown below.

$$\hat{\sigma}_{sg_{it}} = \sqrt{var(\hat{\beta}_0) + ts_{it}^2 var(\hat{\beta}_3) + bl_{it}^2 var(\hat{\beta}_4) + 2ts_{it}cov(\hat{\beta}_0\hat{\beta}_3) + 2bl_{it}cov(\hat{\beta}_0\hat{\beta}_4) + 2ts_{it}bl_{it}cov(\hat{\beta}_3\hat{\beta}_4)}$$

When regressing non-stationary trends spurious correlation is a major concern, however in this case I find it appropriate and necessary to address another problem. Because firms are forward looking, the road construction was generally known in advance, the plant lifetimes can potentially be very long, and there are potential benefits to being a first mover, it is highly likely that some firms would relocate or expand operations in anticipation of the road completion. On the other hand, relocating is expensive, and firms may prefer to postpone relocation or expansion as the desirability of locations depends on the changing travel times as well as the locations of other firms. That is, the effect of the changing travel time index could lead or lag behind the effect on spatial GINI and the timing could vary by industry. This is supported by cross-correlation results between the industry specific spatial GINI's and the lagged travel time index (see Figure 9 in the appendix). Because of this, transforming the series with first difference requires the regression to precisely specify the leads and lags structure, a well-known problem in the literature (Hannan and Robinson, 1973; Andrews and Fair, 1992; Vaisey and Miles, 2014). By regressing the levels and not specifying leads or lags however, the long run effect is captured. I perform simulations with artificial data to verify the efficacy of this specification, finding that the levels regression with only contemporaneous variables accurately estimates the true long run effect regardless of the leads and lags distribution, while the first difference regression parameter estimates are extremely sensitive to the lead and lag specification. See the Appendix C for more information on this issue and the simulation results.

This still leaves the possibility of an unobserved change across time, such as technology change leading to industry restructuring, being responsible for the change in spatial distribution of industries. I address this in two ways. First, the Interstate Highway was completed in 1993 and the spatial GINI is not changing by as much after the year 2000[14]. Even after accounting for a potential lagged response, unless the unobserved change also finished around the time of the Interstate completion, this suggests the roads did have an effect. Second, by performing the same analysis at the regional level any unobserved time effects that affect all regions can be controlled for while facilitating estimation due to the variation in travel time across regions. The combination of these would require that in order for the change in travel time to be spuriously correlated there would have to be an unobserved simultaneous change across time that concludes around the same time as highway construction and also varies across regions in the same way the road completion dates do.

The regional specification adds a dimension to the dependent variable, the spatial GINI, as well as the travel time as shown below.

$$\text{spatialGINI}_{itk} = \alpha + \alpha_i + +\alpha_k + \beta_0 tt_{tk} + \beta_1 ts_{it} + \beta_2 bl_{it} + \beta_3 tt_{tk} ts_{it} + \beta_4 tt_{tk} bl_{it} + \epsilon_{itk}$$

In addition to facilitating the time effect to control for unobserved variables affecting all regions, this specification captures the regional variation in magnitude and timing of completion in the coefficients on travel time. Intuitively, if regions that complete their highway portion earlier also agglomerate/disperse earlier, than this suggests that the change is due to the road completion, and this will be picked up by the coefficients. This identification strategy will be valid unless the unobserved parallel trend also varies at the state level in the same way as completion timing, or if there are unobserved region specific variables changing that happen to cause a change in spatial distribution at the same time the roads are being completed.[15]

[14]The variance of the change in spatial GINI from 1969 to 2000 across industries is .0014, while it is .00063 from 2001 to 2017. Furthermore, the average of the absolute value of the change in spatial GINI is .028 and .013 for 1969-2000 and 2001-2017 respectively

[15]in order to protect business confidentiality, many county earnings are suppressed for certain industries. The suppression rate in a given year varies from less than 5% to 50% depending on the industry. This should not interfere with the overall patterns of spatial distribution but when looking at the state or regional level these suppressions become a significant issue generating movements in the data that are more a product of

To verify the robustness of the results to variable definitions I perform additional tests. The spatial GINI discussed is based on county level earnings, but some counties have significantly different land areas, which could obscure the change in clustering when economic activity moves between counties of different sizes. To account for this I compute another set of spatial GINI's based on county earnings per land area. Additionally, I construct alternate measures of spatial inequality: the Theil index and the 80:40 ratio, to verify the robustness of the results. I also add controls for the boating, rail, and air transport shares of input. These results support the central finding and can be seen in Appendix B.

4 Result

The regression results for the national regression with various specifications are shown below in Table 1. From the coefficients we can see that a reduction in travel time is correlated with increased clustering, but for industries with a high truck transportation share of inputs and a high measure of backward linkage this is smaller and can even be negative, implying a correlation with dispersal rather than agglomeration. This is similar to the results in Rothenberg (2011) who finds that road surface quality improvements in Indonesia lead to a dispersal of durable goods manufacturers relative to nondurable goods manufacturers using the Ellison and Glaeser index.

The results are similar for both random and fixed effects, suggesting heterogeneity bias is not a problem, and this is further validated by the estimator from Bell and Jones (2015). As discussed, the first difference estimation is not reliable without knowledge of the structure of the response leads and lags.

The standard errors and marginal effects are shown by industry in Figure 7. The average z-score of the marginal effect of tt across time and across industries is 4.08 with a standard deviation of 1.99, indicating that the estimate is statistically significant for most industries

suppression policy change than actual industry relocation. The regional spatial GINI's were obtained with the cooperation of the BEA running my algorithm on the unsuppressed data, but because of this the source data is not available for replication.

Table 1: Estimation Results

Coef.	RE	FE1	REWB	FE2	FE3	FD	RS
tt	-.81***	-.82***	-.82***	-	-.74***	.24	-1.29***
	(.12)	(.12)	(.12)		(.13)	(.41)	(.081)
ts	-1.31***	-1.27***	-1.27***	-2.85***	-1.23***	-1.24*	-.50***
	(.23)	(.23)	(.23)	(.90)	(.23)	(.73)	(.122)
bl	-.15***	-.16***	-.15***	-.18*	-.15***	.14*	-.26***
	(.02)	(.02)	(.02)	(.1)	(.02)	(.07)	(.014)
*tt*ts*	3.58***	3.50***	3.50***	6.85***	3.45***	3.69*	1.63***
	(.56)	(.56)	(.56)	(2.44)	(.56)	(2.01)	(.30)
*tt*bl*	.42***	.42***	.42***	.52*	.43***	-.30	.72***
	(.06)	(.06)	(.06)	(.28)	(.06)	(.19)	(.037)

Signif. codes: .01 '***' .05 '**' .1 '*'
tt-travel time, *ts*-trucking share, *bl*-backward linkage
FE1 is individual 'within' fixed effects
REWB is random effects with industry averages to capture the between group effects while controlling for heterogeneity bias as discussed in Bell and Jones (2015)
FE2 is time 'within' fixed effects
FE3 is two ways 'within' fixed effects
FD is first difference
RS is the regional variation specification

most of the time. While it is beneficial to incorporate the noise from each of the interacting terms into the standard errors, in this context because the travel time is constant across industries the variations in marginal effect within industry are driven by changes in *ts* and *bl*.

The estimated marginal effects of travel time echo support for the theories discussed due to the signs of the estimated coefficients. Most industries have a positive predicted marginal effect, suggesting they are dispersing in response to the reduction in travel time. This includes almost every industry involved in producing physical goods as they generally have a higher trucking share. Industries that have a low measure of backward linkage (they do not pull on as many industries for inputs) are more likely to have a negative predicted marginal effect, consistent with the benefits of centralized distribution from dispersal being larger for industries with high backward linkage. These marginal effects are overall consistent with Redding and Turner (2014) who survey the existing literature finding that highways

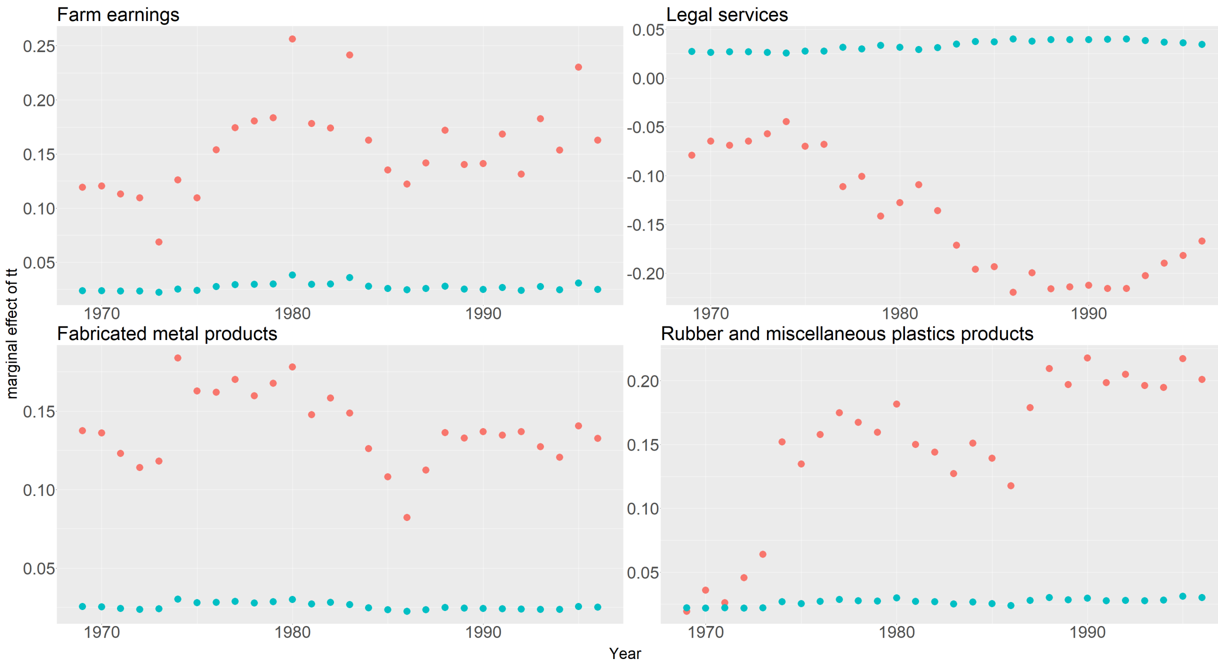

Figure 5: Marginal Effect of Travel Time on Spatial GINI and Standard Errors by Industry

tend to decentralize urban populations and manufacturing activity while different sectors appear to respond differently.

The regressions for spatial GINI with land area control, alternate measures, and additional controls support the central findings and can be found in Appendix B.

The coefficients from this regression support the hypothesis as the signs are unchanged and the standard error diminishes. By adding the regional variation in travel time and spatial GINI the coefficients reflect the differences in timing and magnitude of the change, and the time effects control for unobserved variables affecting all industries and regions. The magnitude of travel time and backwards linkage is increased, while the magnitude of trucking share of inputs diminishes, suggesting that within regions these variables have slightly different importance which may stem from the prominence of different types of industries in each region.

5 Conclusion

Industries are subject to economic centripetal and centrifugal forces influencing the patterns of their relative positions in space. Differences between industries will result in differing sensitivities to these forces. As the road system is improved, both agglomeration and dispersal are facilitated, leading to some industries clustering more densely in fewer counties and some industries spreading throughout more counties. These differing responses can partially be explained by truck transportation utilization and backward linkages—industries with higher measures in both tend to disperse in response to a reduction in travel times.

This finding is relevant for countries building limited access highway systems as well as regions building roads, as they should consider the impact on the spatial distribution of earnings and structural inequality. As certain firms increasingly cluster in population centers while other firms disperse to capture lower wages the inequality between regions is exacerbated. On average across industries earnings per person in city counties was 59% higher than in rural counties in 1969, and in 2000 this ratio rose to 76% while the population distribution did not change substantially. While roads connect regions they also drive them apart. Like other trade cost reductions contributing to globalization, roads bring benefits that may need to be tempered with other policies.

This paper expands the understanding of how the clustering of economic activity responds to changes in the road system and contributes new data on the changes in travel time in the US from the construction of the Interstate Highway System. The spatial GINI is not a novel concept, but the application in the context of road improvements is original and may be useful to other researchers.

These findings are robust to multiple specifications, but there are limits to the interpretation. This does not tell us about where economic growth will occur, only about the response in clustering behavior. There are still challenges to understanding the patterns of spatial distribution such as the importance of history, the tendency for positive feedback, and the influence of new technologies such as phones, computers, and the internet.

The high detail of the travel time data set leaves opportunities for future research, including examining metrics of spatial distribution other than the spatial GINI, examining the market access of different regions and how changes influenced economic growth, as well as the effect of the travel time on other data such as traffic congestion, patterns of trade, and the impact on the changing economic make-up of regions.

Appendix A: Additional Figures

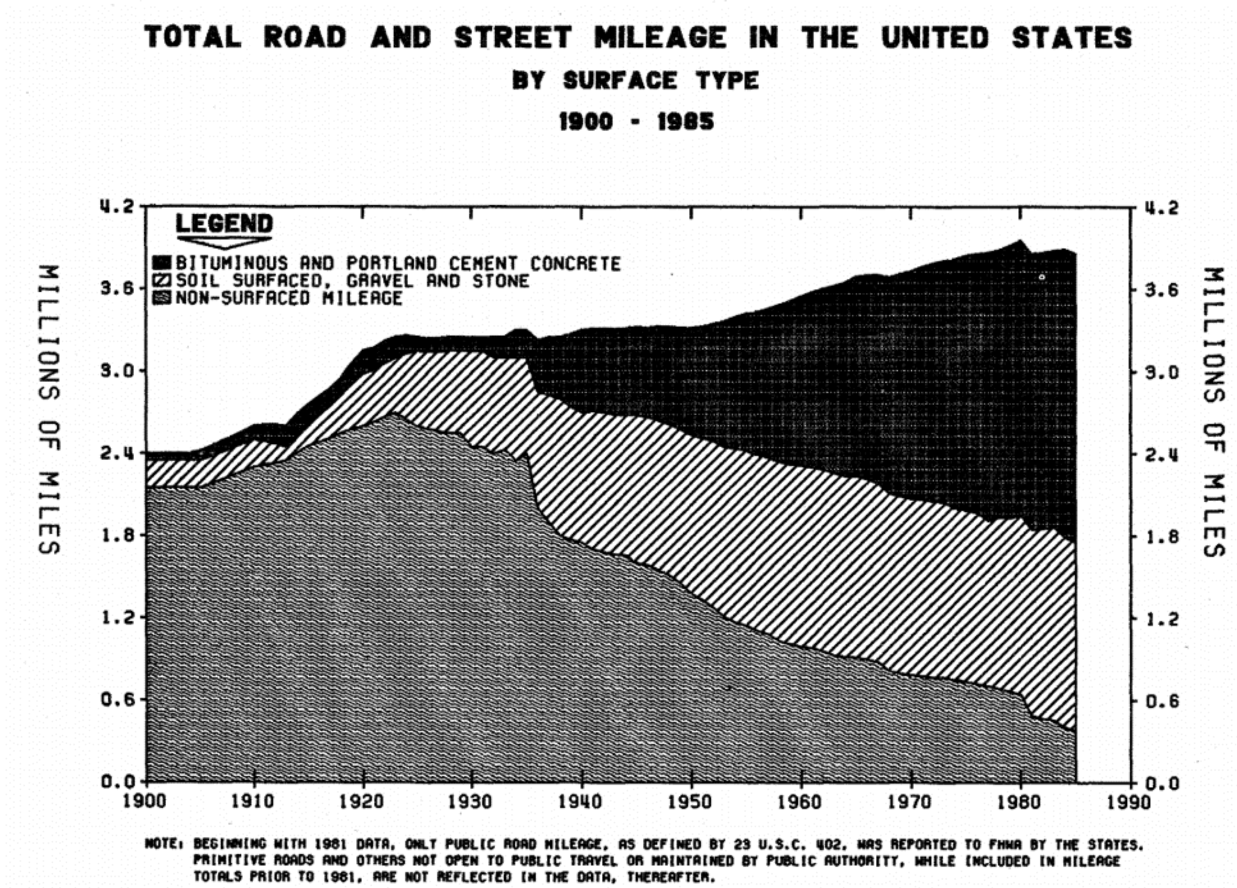

Figure 6: Surface Quality of US Roads, Source: US DOT 1985

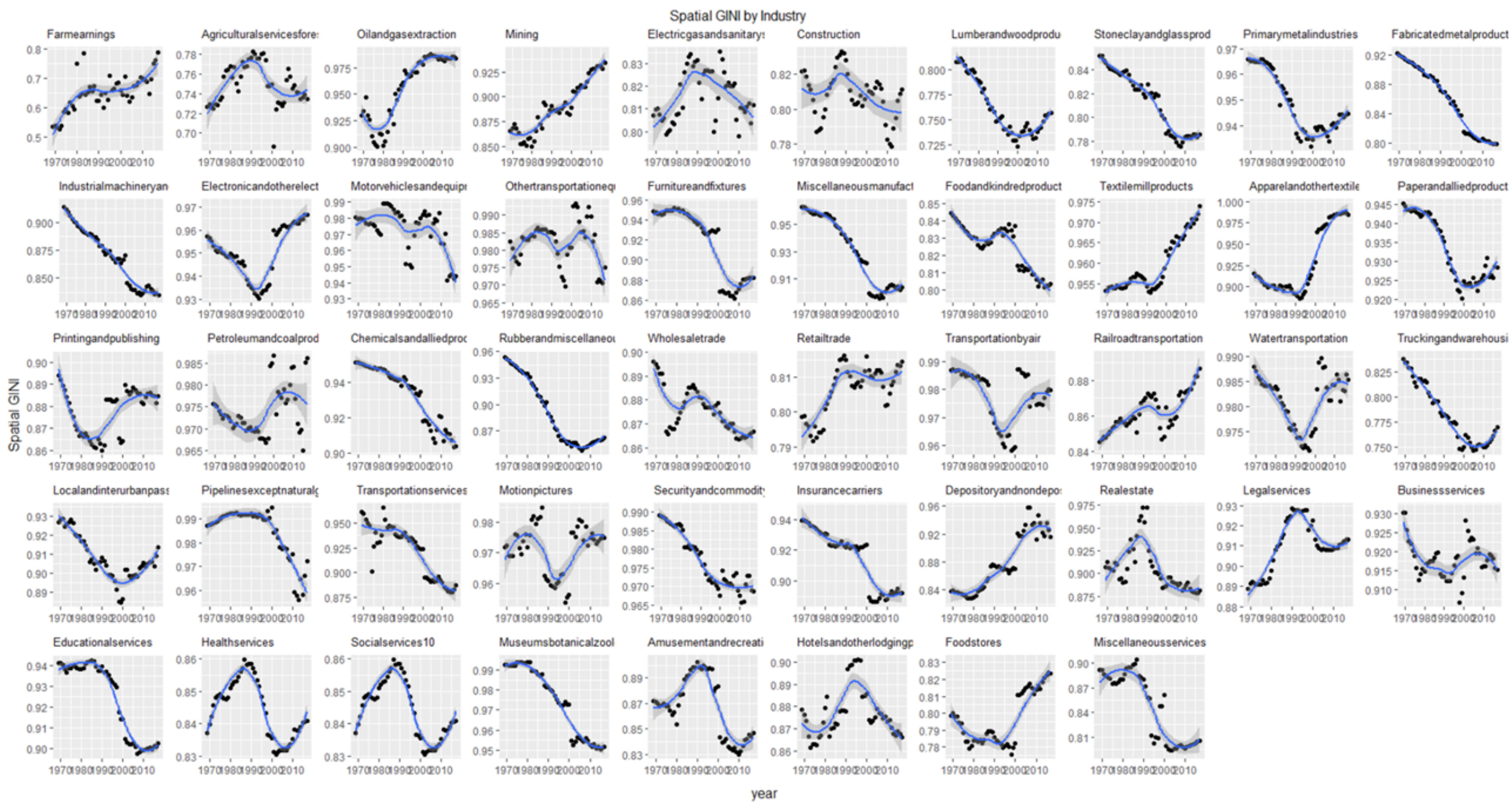

Figure 7: Spatial GINI by Industry before and after 2000

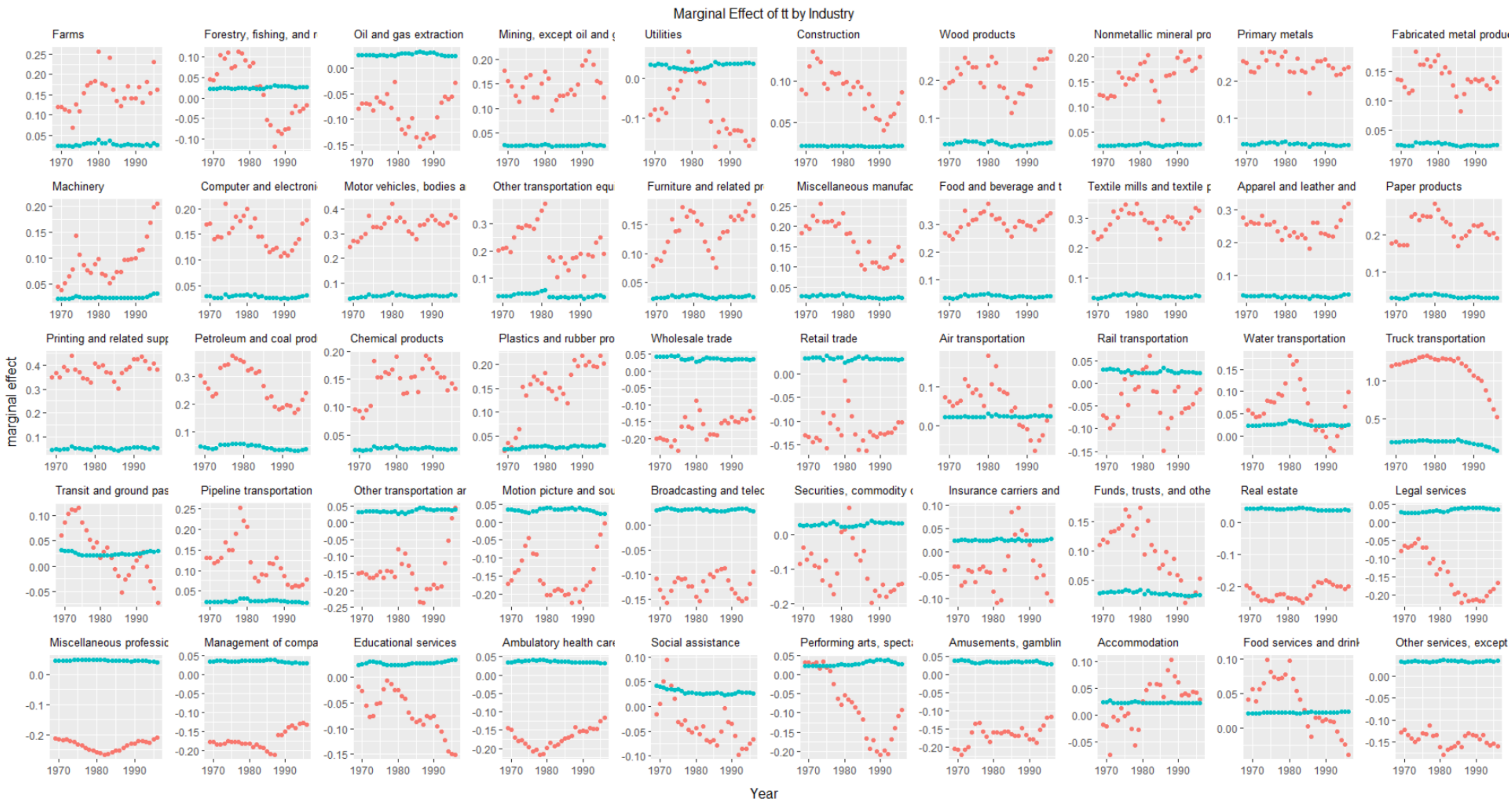

Figure 8: Marginal Effect of tt on Spatial GINI and Standard Errors by Industry

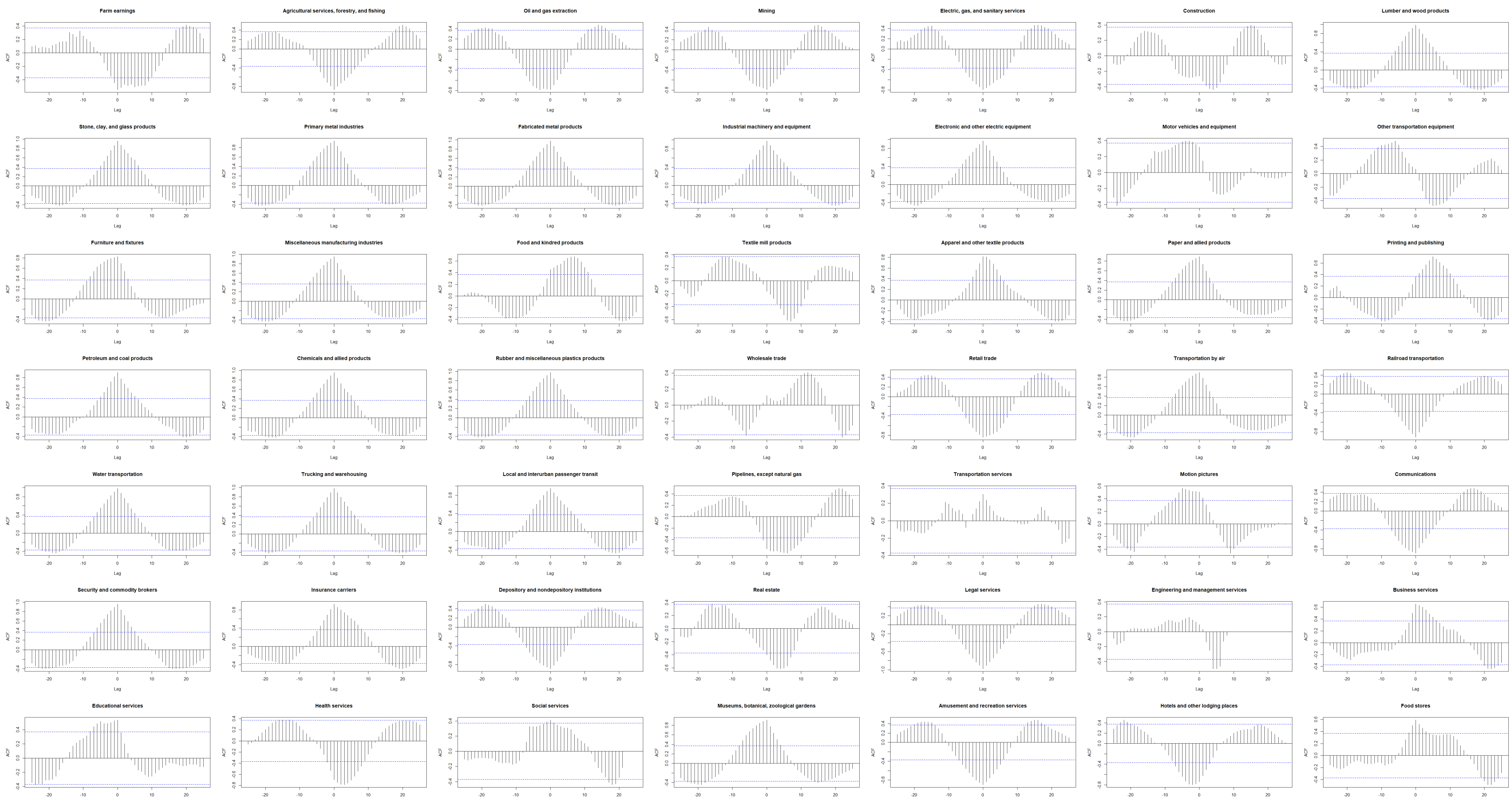

Figure 9: Cross Correlations for Lagged Values of Travel Time and Industry Spatial GINI

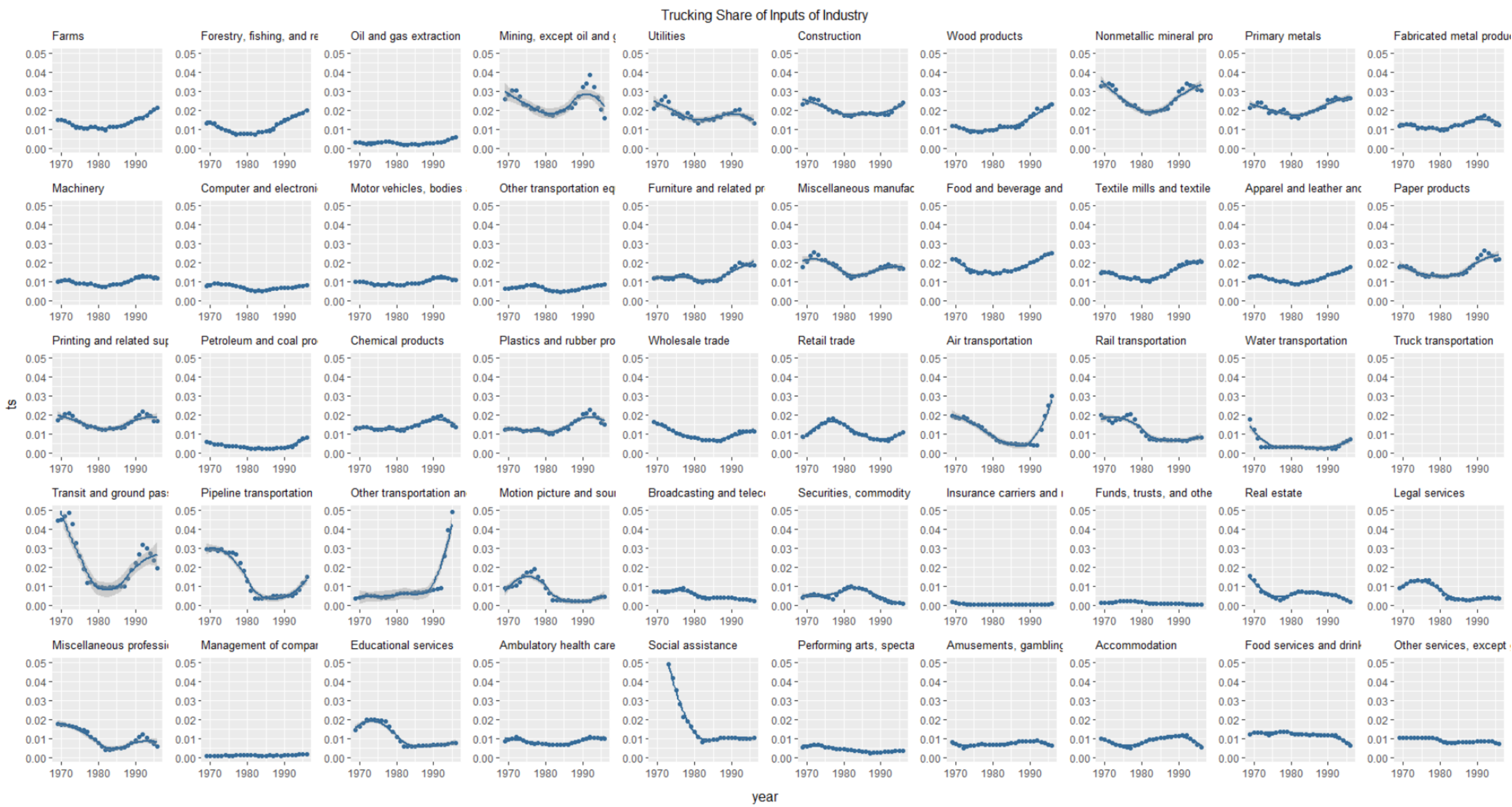

Figure 10: Trucking Share of Inputs by Industry

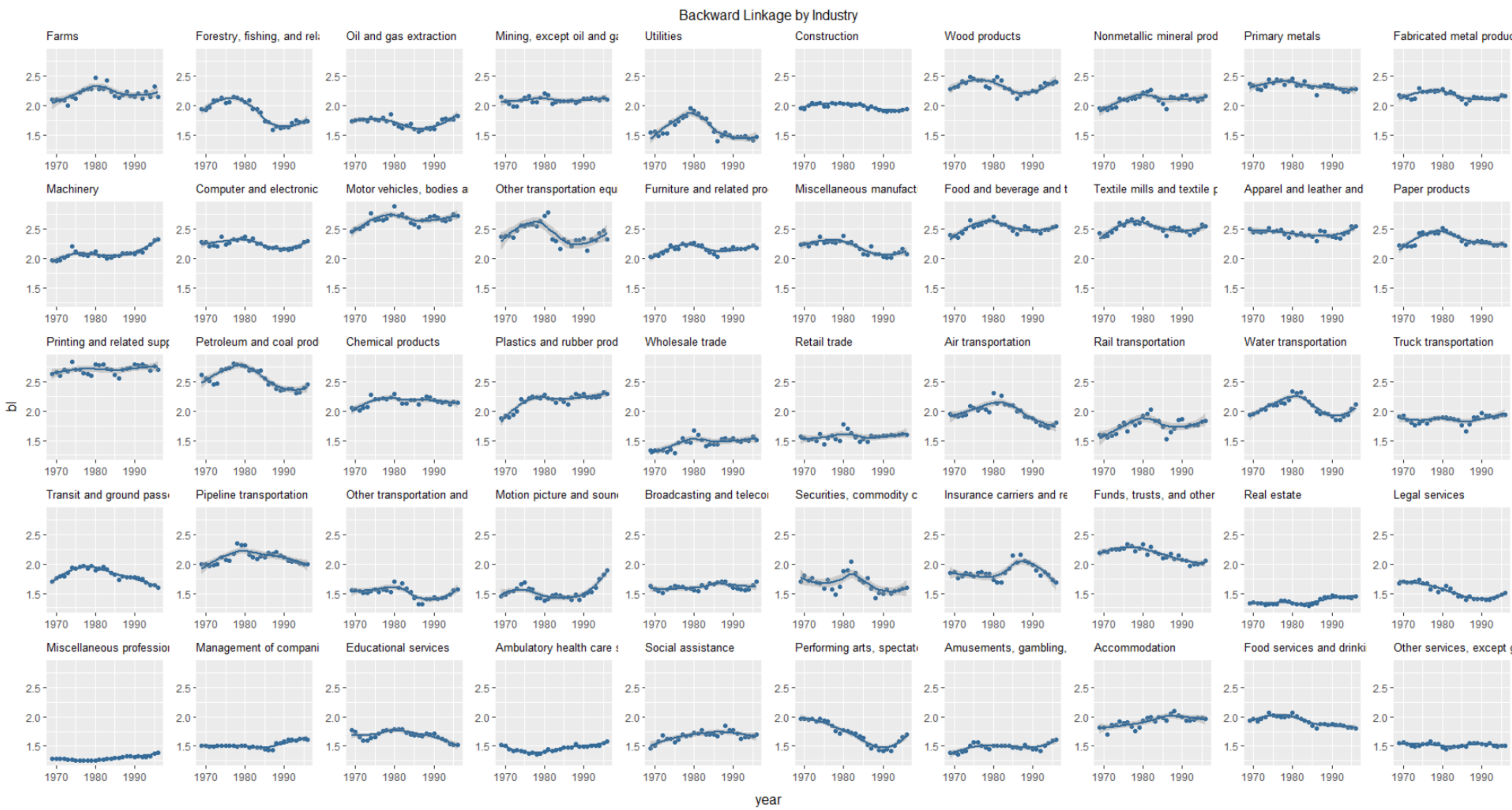

Figure 11: Rasmussen Measure of Backwards Linkages by Industry

Name(SIC)	mean(sGINI)	sGINI 69-96	mean(ts)	ts 69-96	mean(bl)	bl 69-96
Farmearnings	0.625	0.123	0.013	0.007	2.197	0.048
Agriculturalservicesforestryandfishing	0.757	0.054	0.012	0.007	1.88	-0.21
Oilandgasextraction	0.936	0.041	0.003	0.003	1.706	0.102
Mining	0.872	0.025	0.024	-0.01	2.089	-0.046
Electricgasandsanitaryservices	0.816	0.02	0.018	-0.008	1.606	-0.079
Construction	0.811	-0.018	0.02	0.001	1.978	-0.015
Lumberandwoodproducts	0.776	-0.064	0.013	0.012	2.332	0.129
Stoneclayandglassproducts	0.833	-0.032	0.026	-0.002	2.095	0.201
Primarymetalindustries	0.955	-0.028	0.021	0.005	2.334	-0.084
Fabricatedmetalproducts	0.892	-0.066	0.012	0	2.166	-0.014
Industrialmachineryandequipment	0.888	-0.046	0.01	0.002	2.085	0.364
Electronicandotherelectricequipment	0.945	-0.023	0.007	0	2.242	0.016
Motorvehiclesandequipment	0.977	-0.031	0.01	0.001	2.655	0.275
Othertransportationequipment	0.982	-0.011	0.007	0.002	2.406	-0.046
Furnitureandfixtures	0.943	-0.02	0.013	0.007	2.157	0.145
Miscellaneousmanufacturingindustries	0.95	-0.036	0.018	-0.001	2.185	-0.154
Foodandkindredproducts	0.833	-0.01	0.018	0.003	2.518	0.143
Textilemillproducts	0.955	0.001	0.015	0.006	2.514	0.118
Apparelandothertextileproducts	0.9	-0.013	0.012	0.006	2.428	0.052
Paperandalliedproducts	0.939	-0.02	0.017	0.004	2.317	-0.002
Printingandpublishing	0.874	-0.012	0.016	-0.001	2.71	0.083
Petroleumandcoalproducts	0.971	-0.009	0.004	0.002	2.554	-0.167
Chemicalsandalliedproducts	0.945	-0.014	0.014	0.001	2.164	0.082
Rubberandmiscellaneousplasticsproducts	0.916	-0.094	0.014	0.002	2.175	0.413
Wholesaletrade	0.88	-0.019	0.01	-0.005	1.458	0.191
Retailtrade	0.804	0.01	0.011	0.002	1.564	0.043
Transportationbyair	0.979	-0.024	0.012	0.01	1.97	-0.14
Railroadtransportation	0.858	0.024	0.012	-0.012	1.753	0.237
Watertransportation	0.98	-0.016	0.004	-0.01	2.057	0.182
Truckingandwarehousing	0.803	-0.065	0.333	-0.195	1.861	0.034
Localandinterurbanpassengertransit	0.913	-0.031	0.022	-0.025	1.815	-0.104
Pipelinesexceptnaturalgas	0.991	0.003	0.014	-0.015	2.106	-0.004
Transportationservices	0.941	-0.03	0.011	0.053	1.506	0.007
Motionpictures	0.971	-0.009	0.007	-0.005	1.527	0.444
Communications	0.877	0.028	0.005	-0.005	1.606	0.074
Securityandcommoditybrokers	0.982	-0.018	0.005	-0.003	1.659	-0.111
Insurancecarriers	0.928	-0.019	0	-0.001	1.862	-0.167
Depositoryandnondepositoryinstitutions	0.848	0.03	0.001	-0.001	2.165	-0.129
Realestate	0.919	-0.008	0.006	-0.014	1.365	0.115
Legalservices	0.908	0.039	0.007	-0.005	1.538	-0.166
Engineeringandmanagementservices11	0.923	X	0.01	-0.012	1.296	0.107
Businessservices	0.918	-0.012	0.001	0.001	1.524	0.104
Educationalservices	0.939	-0.01	0.011	-0.007	1.68	-0.256
Healthservices	0.851	0.009	0.008	0.001	1.461	0.054
Socialservices10	0.833	X	0.021	-0.041	1.672	0.229
Museumsbotanicalzoologicalgardens	0.987	-0.02	0.004	-0.002	1.709	-0.275
Amusementandrecreationservices	0.881	0.026	0.007	-0.002	1.484	0.225
Hotelsandotherlodgingplaces	0.878	0.022	0.009	-0.005	1.922	0.15
Foodstores	0.787	-0.017	0.012	-0.006	1.925	-0.141
Miscellaneousservices	0.875	-0.082	0.009	-0.003	1.519	-0.046

Name(SIC)	Name(NCIS)
Farmearnings	Farms
Agriculturalservicesforestryandfishing	Forestryfishingandrelatedactivities
Oilandgasextraction	Oilandgasextraction
Mining	Miningexceptoilandgas
Electricgasandsanitaryservices	Utilities
Construction	Construction
Lumberandwoodproducts	Woodproducts
Stoneclayandglassproducts	Nonmetallicmineralproducts
Primarymetalindustries	Primarymetals
Fabricatedmetalproducts	Fabricatedmetalproducts
Industrialmachineryandequipment	Machinery
Electronicandotherelectricequipment	Computerandelectronicproducts
Motorvehiclesandequipment	Motorvehiclesbodiesandtrailersandparts
Othertransportationequipment	Othertransportationequipment
Furnitureandfixtures	Furnitureandrelatedproducts
Miscellaneousmanufacturingindustries	Miscellaneousmanufacturing
Foodandkindredproducts	Foodandbeverageandtobaccoproducts
Textilemillproducts	Textilemillsandtextileproductmills
Apparelandothertextileproducts	Apparelandleatherandalliedproducts
Paperandalliedproducts	Paperproducts
Printingandpublishing	Printingandrelatedsupportactivities
Petroleumandcoalproducts	Petroleumandcoalproducts
Chemicalsandalliedproducts	Chemicalproducts
Rubberandmiscellaneousplasticsproducts	Plasticsandrubberproducts
Wholesaletrade	Wholesaletrade
Retailtrade	Retailtrade
Transportationbyair	Airtransportation
Railroadtransportation	Railtransportation
Watertransportation	Watertransportation
Truckingandwarehousing	Trucktransportation
Localandinterurbanpassengertransit	Transitandgroundpassengertransportation
Pipelinesexceptnaturalgas	Pipelinetransportation
Transportationservices	Othertransportationandsupportactivities
Motionpictures	Motionpictureandsoundrecordingindustries
Communications	Broadcastingandtelecommunications
Securityandcommoditybrokers	Securitiescommoditycontractsandinvestments
Insurancecarriers	Insurancecarriersandrelatedactivities
Depositoryandnondepositoryinstitutions	Fundstrustsandotherfinancialvehicles
Realestate	Realestate
Legalservices	Legalservices
Engineeringandmanagementservices11	Miscellaneousprofessionalscientificandtechnicalservices
Businessservices	Managementofcompaniesandenterprises
Educationalservices	Educationalservices
Healthservices	Ambulatoryhealthcareservices
Socialservices10	Socialassistance
Museumsbotanicalzoologicalgardens	Performingartsspectatorsportsmuseumsandrelatedactivities
Amusementandrecreationservices	Amusementsgamblingandrecreationindustries
Hotelsandotherlodgingplaces	Accommodation
Foodstores	Foodservicesanddrinkingplaces
Miscellaneousservices	Otherservicesexceptgovernment

Appendix B. Robustness

	Dependent variable:			
	g			
	(1)	(2)	(3)	(4)
tt	-0.817***	-1.189***	-2.533***	
	(0.119)	(0.185)	(0.731)	
ts	-1.268***	-1.294***	-0.853	0.135***
	(0.230)	(0.230)	(1.237)	(0.041)
bl	-0.155***	-0.151***	-0.432***	0.014***
	(0.023)	(0.023)	(0.147)	(0.005)
Year		-0.001***	-0.004**	0.002***
		(0.0003)	(0.002)	(0.0003)
tt:ts	3.502***	3.554***	2.462	
	(0.560)	(0.559)	(3.074)	
tt:bl	0.421***	0.420***	1.116***	
	(0.060)	(0.060)	(0.365)	
ts:Year			-0.003	-0.009***
			(0.008)	(0.001)
bl:Year			0.002*	-0.001***
			(0.001)	(0.0001)
Constant	0.926***	1.068***	1.611***	0.597***
	(0.045)	(0.070)	(0.294)	(0.010)
Observations	1,375	1,375	1,375	1,375
R^2	0.960	0.960	0.960	0.960
Adjusted R^2	0.958	0.959	0.959	0.958
Residual Std. Error	0.015 (df = 1320)	0.015 (df = 1319)	0.015 (df = 1317)	0.015 (df = 1320)
F Statistic	587.684*** (df = 54; 1320)	579.677*** (df = 55; 1319)	560.182*** (df = 57; 1317)	585.927*** (df = 54; 1320)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 3: Regressions with a Time Trend

Table 4: Alternate Measures and Controls

	Controls		Without bl		Theil		80:40	
Coef.	RE	FE1	RE	FE1	RE	FE1	RE	FE1
α	1.09	-	.89	-	6.55	-	226	-
	(.05)		(.04)		(.85)		(13.6)	
tt	-.57	-.56	-.04***	-.05***	-13.9	-13.59	-621	-617
	(.13)	(.13)	(.04)	(.04)	(2.25)	(2.24)	(36.3)	(36.3)
ts	-.75	-.63	-.70	-.58*	-10.0*	-7.77**	-108**	-81.0***
	(.24)	(.24)	(.24)	(.24)	(4.22)	(4.35)	(64.7)	(69.4)
bs	7.89*	7.96*	8.58*	8.65*	84.0***	80.4***	-4060	-3940
	(3.56)	(3.56)	(3.58)	(3.58)	(63.9)	(63.9)	(1000)	(1010)
rs	3.18	3.61	2.64*	3.10	31.4***	42.7*	-850	-750*
	(1.12)	(1.12)	(1.12)	(1.12)	(19.85	(20.1)	(308)	(319)
as	1.07***	1.19***	1.96***	2.05**	104	105	-21.3***	-31.9***
	(1.22)	(1.22)	(1.21)	(1.21)	(20.9)	(20.8)	(347)	(347)
bl	-.10	-.10	-	-	-3.09	-2.93	-72.6	-68.9
	(.02)	(.02)			(.43)	(.43)	(6.84)	(6.92)
*tt*ts*	2.27	2.02	2.10	1.86	24.5*	19.6**	295**	239***
	(.59)	(.59)	(.59)	(.59)	(10.4)	(10.6)	(161)	(168)
*tt*bs*	-21.0*	-21.2*	-23.0*	-23.1*	-248***	-239***	10600	10400
	(9.03)	(9.02)	(9.1)	(9.06)	(162)	(162)	(2540)	(2560)
*tt*rs*	-4.35***	-5.06**	-2.30***	-3.10***	-93.2**	-111*	2150	2040
	(2.75)	(2.75)	(2.72)	(2.72)	(49.1)	(49.2)	(771)	(781)
*tt*as*	-2.75***	-3.09***	-5.1***	-5.36***	-291	-296	37.4***	49.5***
	(3.29)	(3.29)	(3.26)	(3.25)	(56.3)	(56.1)	(933)	(935)
*tt*bl*	.28	.27	-	-	9.30	9.00	211	206
	(.06)	(.06)			(1.13)	(1.13)	(18.2)	(18.3)
R^2_{adj}	.13	.09	.12	.08	.13	.11	.44	.44

FE1 is individual 'within' fixed effects
Removing the trucking industry makes ts large significant in 80:40 and Theil (ts outlier)
Signif. codes: .01 ' ' .05 '*' .1 '**' 1 '***' (the stars are reversed)

	Dependent variable:	
	g	
	(1)	(2)
tt	-0.817***	-0.280***
	(0.119)	(0.073)
ts	-1.268***	-0.316**
	(0.230)	(0.142)
bl	-0.155***	-0.068***
	(0.023)	(0.014)
tt:ts	3.502***	0.829**
	(0.560)	(0.345)
tt:bl	0.421***	0.188***
	(0.060)	(0.037)
Constant	0.926***	0.756***
	(0.045)	(0.027)
Observations	1,375	1,375
R^2	0.960	0.967
Adjusted R^2	0.958	0.965
Residual Std. Error (df = 1320)	0.015	0.009
F Statistic (df = 54; 1320)	587.684***	707.411***
Note:	*p<0.1; **p<0.05; ***p<0.01	
	(1)-National suppressed	
	(2)-Earnings Divided by Land Area	

Table 5: Regression with Control for County Land Area

Appendix C. Empirical Specification

On Lags and Leads

In considering the impact of improving the road system on the spatial distribution of industry it is reasonable to believe the effect could lead or lag. Because firms are forward looking, the road construction was generally known in advance, the plant lifetimes can potentially be very long, and there are potential benefits to being a first mover, it seems probable that some firms would relocate or expand operations in anticipation of the road completion. On the other hand, relocating is expensive, and firms may prefer to postpone relocation or expansion as the desirability of locations depends on the changing travel times as well as the locations of other firms.

The states were required to submit the completion status for the various segments of the Interstate Highway System as it was constructed. The status categories are:
1—fully completed and open to traffic,
2—mostly complete and open to traffic,
3—under construction and not open to traffic,
4—planning, specification, estimates, contracting, right-of-way acquisitions underway,
5—mileage designation underway (public hearings, route location studies).
Based on changes between these statuses (for which only parts of the sample are represented), the average time from construction to opening was 5 years (3→2 or 1, 14% of observations)
the average time from planning to opening was 18 years (4→2 or 1, 14% of observations)
the average time from designation to opening was 4 years (5→2 or 1, 52% of observations).

This information could be used to inform the leads structure, as seemingly firms should have knowledge of where the road will be about 4 or 5 years ahead of time. A lag structure in this case is not immediately apparent but is nevertheless important as misspecification can bias coefficients and even flip the sign of the coefficient as shown in Vaisey and Miles (2014). A common practice is to try multiple lag structures and see which one performs best under a criteria such as the Akaike information criterion (AIC) or Schwartz information criterion (BIC), however this does not solve the problems presented by misspecification. Furthermore, this approach underreports the standard errors, as recognized in Schmidt (1973) and Frost (1975), typically being computed as though the lag length is fixed. Some demo results are presented here to see the implications of this issue.

Using explanatory lags is common in the reduced form roads literature, such as Li and Whitaker (2018) and Jiwattanakulpaisarn et al (2011), while using explanatory leads is less common, Leduc and Wilson (2012) being the only example I know of. In the market access literature, lags are not commonly utilized as the economic structural model is not dynamic.

Andrews and Fair (1992) present a method for adjusting the standard errors of coefficient estimates from the polynomial distributed lag technique when the lag length is uncertain. By allowing the lags to be continuous (with a mapping to discrete) and specifying the lag length as a parameter, the regression function is differentiable with respect to the lag length and the effect from changing the lag length can be included in the standard errors.

Below are the results of estimations with varying lag lengths, using generated data, where X is a trend with noise and $\epsilon \sim$ N(0, 200)

$$Y_t = +ß_0 X_t + ß_1 X_{t-1} + ß_2 X_{t-2} + ß_3 X_{t-3} + ß_4 X_{t-4} + \epsilon_t$$

The true parameter is listed in the far left column. Notice that: 1) the coefficients are inaccurate when the model is underspecified (too few lags) 2) the coefficients are still accurate when the model is overspecified (too many lags) 3) the standard errors are not affected by overspecification

	Dependent variable:										
	y										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
x: 2.068	1.975***	4.028***	3.685***	3.152***	2.065***	2.063***	2.049***	2.051***	2.054***	2.057***	2.066***
	(0.065)	(0.127)	(0.144)	(0.141)	(0.061)	(0.063)	(0.063)	(0.064)	(0.064)	(0.065)	(0.065)
x1: -4.215		-2.294***	-2.618***	-3.208***	-4.166***	-4.167***	-4.189***	-4.188***	-4.184***	-4.181***	-4.172***
		(0.127)	(0.142)	(0.141)	(0.061)	(0.062)	(0.063)	(0.063)	(0.064)	(0.065)	(0.065)
x2: -0.865			0.706***	0.175	-0.863***	-0.865***	-0.885***	-0.884***	-0.882***	-0.877***	-0.869***
			(0.143)	(0.141)	(0.061)	(0.062)	(0.063)	(0.064)	(0.064)	(0.065)	(0.065)
x3: 0.772				1.725***	0.784***	0.782***	0.765***	0.767***	0.770***	0.772***	0.786***
				(0.141)	(0.061)	(0.062)	(0.063)	(0.063)	(0.064)	(0.064)	(0.065)
x4: 4.209					4.145***	4.144***	4.124***	4.125***	4.128***	4.132***	4.138***
					(0.061)	(0.062)	(0.063)	(0.063)	(0.064)	(0.065)	(0.065)
x5: NA						0.008	-0.007	-0.006	-0.003	-0.0001	0.011
						(0.063)	(0.063)	(0.064)	(0.064)	(0.065)	(0.065)
x6: NA							0.110*	0.111*	0.115*	0.117*	0.128**
							(0.063)	(0.064)	(0.064)	(0.064)	(0.065)
x7: NA								-0.010	-0.007	-0.004	0.003
								(0.064)	(0.064)	(0.065)	(0.065)
x8: NA									-0.023	-0.021	-0.011
									(0.065)	(0.065)	(0.065)
x9: NA										-0.028	-0.021
										(0.065)	(0.065)
x10: NA											-0.093
											(0.065)
Constant	-20.355	99.529***	80.644**	49.175	-0.600	-0.668	-1.359	-1.319	-1.248	-1.193	-1.146
	(38.877)	(34.366)	(34.184)	(31.957)	(13.413)	(13.430)	(13.421)	(13.431)	(13.438)	(13.444)	(13.437)
Observations	990	990	990	990	990	990	990	990	990	990	990
R^2	0.481	0.610	0.619	0.670	0.942	0.942	0.942	0.942	0.942	0.942	0.942
Adjusted R^2	0.481	0.609	0.618	0.669	0.942	0.942	0.942	0.942	0.942	0.942	0.942
Residual Std. Error	630.934 (df = 988)	547.229 (df = 987)	540.895 (df = 986)	504.030 (df = 985)	211.233 (df = 984)	211.338 (df = 983)	211.117 (df = 982)	211.222 (df = 981)	211.316 (df = 980)	211.404 (df = 979)	211.294 (df = 978)
F Statistic	916.373*** (df = 1; 988)	772.259*** (df = 2; 987)	535.052*** (df = 3; 986)	499.763*** (df = 4; 985)	3,201.225*** (df = 5; 984)	2,665.025*** (df = 6; 983)	2,289.537*** (df = 7; 982)	2,001.353*** (df = 8; 981)	1,777.416*** (df = 9; 980)	1,598.368*** (df = 10; 979)	1,454.749*** (df = 11; 978)
Note:											*p<0.1; **p<0.05; ***p<0.01

Figure 12: Simulation Results: Simple Trend

These results are a bit surprising, as the lagged independent variables are highly autocorrelated, and I expected multicollinearity to be a problem, which it does not seem to be here. The results 1)-3) above are robust to:

-X being a purely random variable (no trend)

-the true model only having lagged coefficients (no Xt)

-the true model skipping certain lags (for instance Xt-2 and Xt-4, but not Xt-3)

However, these results are not robust to

-drastically reducing the sample size

-drastically increasing the error variance

-drastically reducing the size of the coefficients

When the true coefficients are distributed according to a polynomial the unrestricted model is able to accurately estimate them, but if the degrees of freedom are a concern then the polynomial distributed lag technique may be desirable. The table below shows the results for varying lag lengths when the true coefficients are distributed according to a 2nd order polynomial and the last lag is restricted to be zero.

The true coefficients are accurately picked out when the correct lag length is specified, but when further lags are included the model is not able to reject the null hypothesis that they are zero, although it still performs fairly well. After accounting for the uncertainty of the lag length as in Andrew and Fairs (1992) the standard errors increase significantly when the model is misspecified. This suggests that without applying the Andrew and Fairs (1992) method, one could easily accept coefficient estimates that are in reality far from the true value.

Lag	true_b	β_2	std.err_2	std.err_AF_2	β_3	std.err_3	std.err_AF_3	β_4	std.err_4	std.err_AF_4
0	-17.37	-9.31	10.55	11.13	-14.34	2.88	3.10	-17.27	0.45	0.49
1	-16.13	-21.22	1.79	2.23	-17.47	0.86	0.95	-16.01	0.18	0.18
2	-13.77	-18.12	1.81	3.10	-16.13	0.33	1.15	-13.66	0.02	0.18
3	-10.29				-10.30	0.60	1.13	-10.20	0.10	0.24
4	-5.70							-5.65	0.09	0.18
5										
6										
7										
8										
9										
10										

Lag	β_5	std.err_5	std.err_AF_5	β_6	std.err_6	std.err_AF_6	β_7	std.err_7	std.err_AF_7	β_8	std.err_8	std.err_AF_8
0	-21.58	1.14	1.18	-23.29	1.10	1.14	-23.83	0.96	1.03	-23.48	0.81	0.87
1	-16.81	0.54	0.56	-17.36	0.58	0.59	-17.69	0.55	0.55	-17.64	0.49	0.49
2	-12.51	0.11	0.47	-12.30	0.19	0.45	-12.45	0.23	0.38	-12.61	0.24	0.36
3	-8.68	0.19	0.61	-8.10	0.11	0.59	-8.12	0.05	0.51	-8.39	0.05	0.45
4	-5.31	0.29	0.62	-4.77	0.26	0.68	-4.69	0.19	0.63	-4.97	0.12	0.57
5	-2.42	0.23	0.42	-2.31	0.29	0.62	-2.16	0.26	0.65	-2.36	0.20	0.63
6				-0.72	0.21	0.39	-0.54	0.26	0.56	-0.56	0.24	0.61
7							0.18	0.17	0.34	0.43	0.21	0.49
8										0.62	0.13	0.29
9												
10												

Figure 13: Standard Errors Adjusted for Lag Length Uncertainty from a Polynomial Distributed Lag Regression

To see if these techniques are appropriate for my situation, I generate data that is distributed similar to mine but where the true relationships are known. I generate a panel dataset consisting of: a monotonically changing trend t (representing the travel time) which only decreases but by different amounts for 42 periods, a variable ts that varies across 5 'industries' with some noise across time, fixed effects for each industry, and the dependent variable which depends on lags and interaction terms

$$Y_{it} = \alpha_i + \zeta_i ts_{it} + \sum_{j=0}^{3} \beta_j t_{t-j} +_j t_{t-j} * ts_{it} + \epsilon_{it},$$

where the coefficients are either randomly generated or set manually and $\epsilon \sim \mathrm{N}(0, \sigma^2)$. This is parallel to the actual data and desired specification, where the level of travel time affects the spatial GINI of each industry differently based on it's truck transport share of inputs. The primary coefficients of interest are the β_j and $_j$ on travel time and the interaction term with truck share. The results from varying lag specifications for both are shown below.

Regardless if the coefficients are generated randomly, linearly, or distributed according to a polynomial, the results are the same—as more lags are added the regression is unable to differentiate which lags the true effects are coming from, but the sum of the coefficients is very close to the sum of the true coefficients, even when the standard errors on the coefficients are too high to be statistically different from zero. In the previous data generation process the autocorrelation was fairly high but not enough to cause multicollinearity, however in this case when ts_{it} and tt are interacted the autocorrelation is much higher, which is likely causing the inability to distribute the coefficients correctly.

This approach is able to pick out the sum of the coefficients for both the travel time and the interaction terms, suggesting that the long run effect of the change in roads on spatial distribution can accurately be inferred, but the timing of the effect may be unknown. This is true even when leads are included in the true model as shown in the figure below. To pick out specifically which lag the effect is coming from, a first difference regression with lags seems tempting, but the same issue of multicollinearity appears, and furthermore the sum of the coefficients is not equal to the true sum, so it is not able to pick up the total effect as with the levels. A table for these results are shown below.

Based on these simulation results, while it is likely there are lagged and lead effects from the road construction, the levels regression is able to pick out the long run effect on spatial distribution for different industries, even with the interaction term, so this is the preferred specification.

	Dependent variable:										
	X										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
l2: -0.121	-0.129***	-0.118***	-0.117***	-0.121***	-0.117***	-0.117***	-0.117***	-0.117***	-0.118***	-0.116***	-0.115***
	(0.022)	(0.020)	(0.020)	(0.019)	(0.019)	(0.019)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
l3: -0.781	-0.798***	-0.778***	-0.775***	-0.784***	-0.777***	-0.776***	-0.777***	-0.775***	-0.777***	-0.773***	-0.771***
	(0.036)	(0.034)	(0.034)	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.034)	(0.035)	(0.035)
l4: -0.051	-0.091*	-0.059	-0.055	-0.068	-0.058	-0.057	-0.058	-0.055	-0.059	-0.052	-0.049
	(0.053)	(0.050)	(0.050)	(0.048)	(0.048)	(0.048)	(0.049)	(0.049)	(0.050)	(0.051)	(0.052)
l5: 0.385	0.322***	0.363***	0.368***	0.351***	0.364***	0.365***	0.365***	0.368***	0.363***	0.372***	0.376***
	(0.069)	(0.064)	(0.064)	(0.062)	(0.062)	(0.062)	(0.063)	(0.063)	(0.064)	(0.066)	(0.067)
ts: -1	-0.909***	-1.013***	-1.044***	-1.072***	-1.085***	-1.069***	-1.066***	-1.052***	-1.044***	-1.049***	-1.063***
	(0.130)	(0.120)	(0.120)	(0.116)	(0.116)	(0.117)	(0.117)	(0.117)	(0.119)	(0.120)	(0.122)
t0: 8	19.978***	9.740***	8.300***	7.581***	7.692***	7.988***	8.047***	8.473***	8.651***	8.876***	8.852***
	(0.181)	(1.824)	(1.911)	(1.842)	(1.925)	(1.932)	(1.960)	(1.975)	(2.057)	(2.101)	(2.112)
t1: 6		10.282***	7.482***	5.717***	5.528**	4.638*	4.564*	3.600	3.576	3.690	3.465
		(1.823)	(2.184)	(2.151)	(2.281)	(2.354)	(2.393)	(2.478)	(2.494)	(2.516)	(2.617)
t2: 4			4.279**	0.201	0.021	0.857	0.942	1.465	1.301	1.310	1.369
			(1.879)	(2.122)	(2.183)	(2.249)	(2.299)	(2.318)	(2.386)	(2.395)	(2.419)
t3: 2				6.611***	6.273***	7.067***	6.955***	6.069**	6.124**	5.725**	5.706**
				(1.835)	(2.189)	(2.254)	(2.329)	(2.403)	(2.426)	(2.543)	(2.556)
t4					0.605	2.439	2.354	3.670	3.411	3.600	3.920
					(1.936)	(2.285)	(2.337)	(2.512)	(2.636)	(2.668)	(2.794)
t5						-2.904	-3.132	-2.620	-2.430	-2.926	-3.082
						(1.945)	(2.327)	(2.346)	(2.435)	(2.618)	(2.676)
t6							0.358	2.309	2.507	2.801	2.987
							(1.939)	(2.387)	(2.461)	(2.544)	(2.624)
t7								-2.901	-2.421	-2.120	-2.448
								(2.072)	(2.583)	(2.651)	(2.840)
t8									-0.672	0.289	0.323
									(2.102)	(2.766)	(2.780)
t9										-1.220	-1.643
										(2.302)	(2.601)
t10											0.590
											(1.792)
ts:t0: -8	-19.717***	-7.221**	-5.260*	-4.512	-4.229	-4.693	-4.618	-5.420*	-5.856*	-6.400**	-6.313*
	(0.273)	(2.769)	(2.883)	(2.780)	(2.913)	(2.924)	(2.960)	(2.987)	(3.108)	(3.188)	(3.206)
ts:t1: -6		-12.551***	-8.649***	-6.807**	-7.113**	-5.722	-5.841	-4.055	-4.017	-4.267	-3.621
		(2.764)	(3.288)	(3.245)	(3.442)	(3.554)	(3.605)	(3.741)	(3.764)	(3.789)	(3.946)
ts:t2: -4			-5.920**	-1.474	-1.775	-3.081	-2.922	-3.942	-3.517	-3.533	-3.731
			(2.751)	(3.101)	(3.197)	(3.304)	(3.378)	(3.418)	(3.525)	(3.540)	(3.573)
ts:t3: -2				-7.076**	-7.723**	-8.892***	-9.085**	-7.475**	-7.590**	-6.721*	-6.692*
				(2.746)	(3.314)	(3.392)	(3.515)	(3.627)	(3.658)	(3.820)	(3.839)
ts:t4					0.981	-1.874	-1.986	-4.319	-3.730	-4.090	-4.928
					(2.949)	(3.480)	(3.553)	(3.799)	(3.983)	(4.026)	(4.238)
ts:t5						4.456	4.090	3.155	2.657	3.809	4.237
						(2.882)	(3.456)	(3.487)	(3.628)	(3.912)	(3.996)
ts:t6							0.565	-2.968	-3.389	-4.148	-4.717
							(2.909)	(3.567)	(3.678)	(3.801)	(3.921)
ts:t7								5.269*	4.068	3.402	4.366
								(3.093)	(3.848)	(3.945)	(4.239)
ts:t8									1.659	-0.527	-0.503
									(3.135)	(4.152)	(4.174)
ts:t9										2.815	3.903
										(3.446)	(3.888)
ts:t10											-1.705
											(2.684)
Constant	-0.552***	-0.498***	-0.478***	-0.440***	-0.435***	-0.447***	-0.446***	-0.455***	-0.457***	-0.457***	-0.453***
	(0.074)	(0.068)	(0.068)	(0.066)	(0.066)	(0.067)	(0.067)	(0.067)	(0.068)	(0.068)	(0.069)
Observations	160	160	160	160	160	160	160	160	160	160	160
R^2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Adjusted R^2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Residual Std. Error	0.057 (df = 152)	0.052 (df = 150)	0.052 (df = 148)	0.049 (df = 146)	0.049 (df = 144)	0.049 (df = 142)	0.050 (df = 140)	0.049 (df = 138)	0.050 (df = 136)	0.050 (df = 134)	0.050 (df = 132)
F Statistic	47,606.760*** (df = 7; 152)	44,556.930*** (df = 9; 150)	37,245.940*** (df = 11; 148)	34,324.200*** (df = 13; 146)	29,738.730*** (df = 15; 144)	26,322.110*** (df = 17; 142)	23,331.960*** (df = 19; 140)	21,259.610*** (df = 21; 138)	19,186.150*** (df = 23; 136)	17,515.180*** (df = 25; 134)	16,057.470*** (df = 27; 132)
Note:											*p<0.1; **p<0.05; ***p<0.01

Figure 14: Regression Results with Multiple Lags for Simulated Data

	Dependent variable:													
	x													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
l2: 0.73	0.691***	0.695***	0.702***	0.705***	0.698***	0.699***	0.699***	0.700***	0.693***	0.693***	0.693***	0.693***	0.694***	0.694***
	(0.033)	(0.029)	(0.024)	(0.020)	(0.019)	(0.019)	(0.019)	(0.020)	(0.019)	(0.019)	(0.020)	(0.020)	(0.020)	(0.020)
l3: 1.348	1.289***	1.296***	1.310***	1.318***	1.305***	1.307***	1.308***	1.309***	1.295***	1.295***	1.296***	1.294***	1.297***	1.297***
	(0.054)	(0.047)	(0.040)	(0.033)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.033)	(0.033)	(0.034)
l4: -0.001	-0.098	-0.086	-0.066	-0.055	-0.076*	-0.074	-0.073	-0.071	-0.092**	-0.091**	-0.091**	-0.093**	-0.089*	-0.089*
	(0.076)	(0.066)	(0.056)	(0.046)	(0.045)	(0.045)	(0.045)	(0.045)	(0.045)	(0.045)	(0.046)	(0.046)	(0.047)	(0.047)
l5: 0.245	0.137	0.152*	0.180**	0.196***	0.168***	0.171***	0.173***	0.175***	0.147**	0.147**	0.148**	0.145**	0.150**	0.150**
	(0.102)	(0.088)	(0.075)	(0.062)	(0.060)	(0.060)	(0.060)	(0.061)	(0.061)	(0.061)	(0.061)	(0.061)	(0.063)	(0.064)
ts: -1	-1.353***	-1.329***	-1.333***	-1.164***	-1.075***	-1.041***	-1.037***	-1.041***	-1.013***	-1.008***	-1.010***	-0.991***	-1.015***	-1.017***
	(0.215)	(0.188)	(0.160)	(0.136)	(0.134)	(0.135)	(0.134)	(0.135)	(0.133)	(0.134)	(0.135)	(0.137)	(0.143)	(0.145)
t-3: -1.476	17.254***	0.488	-1.574	1.911	0.579	-0.182	0.630	0.782	0.899	0.546	0.455	0.237	-0.141	-0.188
	(0.312)	(3.266)	(2.807)	(2.359)	(2.302)	(2.344)	(2.372)	(2.427)	(2.380)	(2.524)	(2.530)	(2.608)	(2.647)	(2.739)
t-2: -0.551		17.257***	4.344	0.035	2.590	2.723	0.681	0.563	0.778	1.033	0.231	0.323	0.964	0.982
		(3.350)	(3.706)	(3.107)	(3.083)	(3.075)	(3.256)	(3.298)	(3.332)	(3.410)	(3.529)	(3.564)	(3.639)	(3.674)
t-1: 1.768			15.400***	-2.172	-2.897	-1.681	-1.195	-1.531	-1.996	-1.527	-0.868	-0.360	-0.352	-0.286
			(2.944)	(3.382)	(3.258)	(3.333)	(3.323)	(3.483)	(3.465)	(3.654)	(3.720)	(4.030)	(4.049)	(4.227)
t0: 8				19.063***	9.969***	9.461**	10.483***	10.502***	10.410***	10.131***	10.859***	10.606**	8.724*	8.710*
				(2.564)	(3.648)	(3.652)	(3.672)	(3.702)	(3.686)	(3.775)	(3.909)	(4.056)	(4.499)	(4.544)
t1: 6					8.834***	4.663	4.516	4.873	5.258	4.728	4.491	4.116	5.326	5.199
					(2.604)	(3.660)	(3.638)	(3.793)	(3.728)	(3.933)	(3.953)	(4.109)	(4.313)	(4.885)
t2: 4						4.238	-0.831	-1.108	-1.278	-1.024	-2.557	-2.413	-1.889	-1.787
						(2.646)	(3.808)	(3.914)	(4.058)	(4.115)	(4.471)	(4.501)	(4.551)	(4.855)
t3: 2							5.039*	4.062	3.960	4.407	4.570	5.088	4.808	4.822
							(2.730)	(3.847)	(3.841)	(4.041)	(4.052)	(4.354)	(4.384)	(4.464)
t4								1.197	1.173	0.847	3.044	3.011	2.618	2.597
								(3.317)	(4.492)	(4.614)	(5.256)	(5.277)	(5.316)	(5.387)
t5									0.128	-1.275	-2.706	-3.408	-4.386	-4.410
									(3.391)	(4.664)	(5.022)	(5.556)	(5.670)	(5.740)
t6										1.502	-1.339	-1.060	0.990	0.911
										(3.487)	(4.616)	(4.721)	(5.181)	(5.339)
t7											3.261	4.222	4.104	4.257
											(3.652)	(5.053)	(5.078)	(5.625)
t8												-0.929	-4.943	-4.964
												(3.421)	(5.355)	(5.408)
t9													3.673	3.421
													(3.775)	(5.676)
t10														0.234
														(3.906)
ts:t-3: -1.436	-20.154***	-6.356	-4.717	-8.747**	-7.147**	-5.937*	-6.990*	-7.103*	-7.103*	-6.402*	-6.314	-5.561	-5.120	-4.958
	(0.462)	(4.918)	(4.243)	(3.569)	(3.485)	(3.553)	(3.600)	(3.690)	(3.616)	(3.832)	(3.841)	(3.973)	(4.032)	(4.187)
ts:t-2: -1.027		-14.150***	-4.909	0.024	-3.151	-3.407	-0.868	-0.780	-2.482	-2.987	-2.493	-2.881	-3.667	-3.718
		(5.051)	(5.527)	(4.644)	(4.615)	(4.604)	(4.898)	(4.972)	(5.046)	(5.156)	(5.341)	(5.389)	(5.503)	(5.558)
ts:t-1: -0.262			-11.170**	8.906*	9.836**	7.937	7.363	7.666	9.434*	8.441	8.019	6.223	6.280	6.043
			(4.447)	(5.098)	(4.909)	(5.028)	(5.017)	(5.270)	(5.282)	(5.572)	(5.662)	(6.161)	(6.190)	(6.454)
ts:t0: -8				-21.772***	-10.833*	-10.040*	-11.306**	-11.340**	-10.058*	-9.465	-9.915	-8.921	-6.683	-6.644
				(3.895)	(5.559)	(5.561)	(5.590)	(5.639)	(5.614)	(5.741)	(5.986)	(6.186)	(6.855)	(6.923)
ts:t1: -6					-10.584***	-4.110	-3.954	-4.260	-4.502	-3.430	-3.277	-1.963	-3.442	-2.937
					(3.942)	(5.550)	(5.516)	(5.740)	(5.633)	(5.952)	(5.989)	(6.248)	(6.581)	(7.439)
ts:t2: -4						-6.545	-0.423	-0.190	-2.143	-2.611	-1.598	-2.045	-2.622	-3.007
						(3.950)	(5.678)	(5.855)	(6.049)	(6.142)	(6.694)	(6.742)	(6.807)	(7.297)
ts:t3: -2							-6.041	-5.152	-3.569	-4.571	-4.650	-6.468	-6.139	-6.239
							(4.043)	(5.748)	(5.758)	(6.037)	(6.058)	(6.546)	(6.591)	(6.697)
ts:t4								-1.074	4.243	4.996	3.495	3.446	3.854	3.941
								(5.089)	(6.861)	(7.023)	(7.998)	(8.027)	(8.078)	(8.181)
ts:t5									-6.127	-3.242	-2.319	0.375	1.567	1.697
									(5.234)	(7.165)	(7.790)	(8.602)	(8.780)	(8.891)
ts:t6										-3.111	-1.279	-2.376	-4.785	-4.524
										(5.245)	(6.861)	(7.032)	(7.714)	(7.968)
ts:t7											-2.104	-5.924	-5.781	-6.354
											(5.540)	(7.645)	(7.683)	(8.565)
ts:t8												3.684	8.460	8.539
												(5.128)	(8.049)	(8.133)
ts:t9													-4.409	-3.452
													(5.770)	(8.644)
ts:t10														-0.896
														(5.944)
Constant	-0.072	-0.146	-0.192**	-0.342***	-0.392***	-0.417***	-0.423***	-0.420***	-0.424***	-0.427***	-0.424***	-0.429***	-0.413***	-0.412***
	(0.125)	(0.109)	(0.093)	(0.080)	(0.078)	(0.079)	(0.079)	(0.080)	(0.078)	(0.079)	(0.079)	(0.081)	(0.083)	(0.085)
Observations	140	140	140	140	140	140	140	140	140	140	140	140	140	140
R^2	0.999	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Adjusted R^2	0.999	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Residual Std. Error	0.079 (df = 132)	0.069 (df = 130)	0.058 (df = 128)	0.048 (df = 126)	0.046 (df = 124)	0.046 (df = 122)	0.046 (df = 120)	0.046 (df = 118)	0.045 (df = 116)	0.045 (df = 114)	0.045 (df = 112)	0.046 (df = 110)	0.046 (df = 108)	0.046 (df = 106)
F Statistic	25,071.180*** (df = 7; 132)	25,784.780*** (df = 9; 130)	29,439.030*** (df = 11; 128)	36,776.250*** (df = 13; 126)	34,490.350*** (df = 15; 124)	30,629.940*** (df = 17; 122)	27,753.360*** (df = 19; 120)	24,733.570*** (df = 21; 118)	23,544.020*** (df = 23; 116)	21,364.190*** (df = 25; 114)	19,743.080*** (df = 27; 112)	18,260.580*** (df = 29; 110)	16,929.250*** (df = 31; 108)	15,616.730*** (df = 33; 106)
Note:														*p<0.1; **p<0.05; ***p<0.01

Figure 15: Regression Results with Multiple Lags and Leads for Simulated Data

	Dependent variable:										
	X										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
I2: 0.708	0.696^{***}	0.695^{***}	0.697^{***}	0.698^{***}	0.698^{***}	0.697^{***}	0.696^{***}	0.696^{***}	0.696^{***}	0.692^{***}	0.692^{***}
	(0.019)	(0.018)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
I3: 0.541	0.486^{***}	0.484^{***}	0.489^{***}	0.490^{***}	0.489^{***}	0.488^{***}	0.485^{***}	0.487^{***}	0.485^{***}	0.479^{***}	0.477^{***}
	(0.034)	(0.032)	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.034)	(0.034)	(0.034)
I4: 1.173	1.077^{***}	1.074^{***}	1.082^{***}	1.084^{***}	1.083^{***}	1.081^{***}	1.077^{***}	1.079^{***}	1.076^{***}	1.068^{***}	1.065^{***}
	(0.049)	(0.046)	(0.047)	(0.047)	(0.047)	(0.048)	(0.047)	(0.047)	(0.048)	(0.049)	(0.049)
I5: 1.733	1.620^{***}	1.615^{***}	1.626^{***}	1.629^{***}	1.627^{***}	1.625^{***}	1.619^{***}	1.622^{***}	1.619^{***}	1.608^{***}	1.604^{***}
	(0.064)	(0.061)	(0.061)	(0.062)	(0.062)	(0.063)	(0.062)	(0.062)	(0.063)	(0.064)	(0.064)
ts: -1	-0.894^{***}	-0.926^{***}	-0.910^{***}	-0.926^{***}	-0.958^{***}	-0.976^{***}	-0.947^{***}	-0.930^{***}	-0.922^{***}	-0.940^{***}	-0.945^{***}
	(0.080)	(0.079)	(0.079)	(0.084)	(0.090)	(0.092)	(0.095)	(0.097)	(0.098)	(0.100)	(0.103)
t0: 8	8.905^{***}	8.802^{***}	8.747^{***}	8.571^{***}	8.272^{***}	8.498^{***}	8.519^{***}	8.296^{***}	8.285^{***}	8.235^{***}	8.169^{***}
	(2.151)	(2.049)	(2.045)	(2.082)	(2.102)	(2.138)	(2.126)	(2.152)	(2.168)	(2.164)	(2.203)
t1: 6		-0.089	0.117	0.157	-0.117	-0.239	-0.172	-0.320	-0.296	-0.345	-0.324
		(2.032)	(2.031)	(2.044)	(2.066)	(2.081)	(2.085)	(2.098)	(2.112)	(2.108)	(2.126)
t2: 4			3.249	3.232	3.274	3.106	3.100	2.818	2.798	2.944	2.944
			(2.015)	(2.027)	(2.021)	(2.043)	(2.035)	(2.082)	(2.119)	(2.120)	(2.135)
t3: 2				-1.158	-1.252	-1.236	-1.223	-1.058	-1.032	-0.779	-0.739
				(2.046)	(2.046)	(2.056)	(2.038)	(2.056)	(2.095)	(2.100)	(2.127)
t4					-1.539	-1.712	-1.683	-1.634	-1.662	-1.273	-1.212
					(2.001)	(2.027)	(2.009)	(2.016)	(2.049)	(2.064)	(2.085)
t5						-1.521	-1.573	-1.800	-1.783	-2.346	-2.382
						(2.192)	(2.277)	(2.304)	(2.324)	(2.362)	(2.386)
t6							-0.151	0.366	0.381	0.181	0.128
							(2.210)	(2.308)	(2.335)	(2.335)	(2.386)
t7								1.736	1.667	1.991	2.017
								(2.274)	(2.368)	(2.380)	(2.396)
t8									-0.170	-0.794	-0.787
									(2.197)	(2.243)	(2.258)
t9										-2.827	-2.905
										(2.175)	(2.268)
t10											-0.319
											(2.232)
ts:t0: -8	-9.633^{***}	-9.165^{***}	-9.084^{***}	-8.815^{***}	-8.099^{**}	-8.517^{***}	-8.187^{**}	-7.771^{**}	-7.787^{**}	-7.764^{**}	-7.566^{**}
	(3.250)	(3.096)	(3.090)	(3.141)	(3.172)	(3.224)	(3.205)	(3.240)	(3.264)	(3.258)	(3.319)
ts:t1: -6		5.707^{*}	5.393^{*}	5.313^{*}	5.942^{*}	6.204^{*}	6.699^{**}	7.027^{**}	7.074^{**}	7.211^{**}	7.164^{**}
		(3.073)	(3.074)	(3.096)	(3.125)	(3.152)	(3.156)	(3.176)	(3.197)	(3.192)	(3.218)
ts:t2: -4			-4.730	-4.708	-4.865	-4.542	-4.960	-4.357	-4.186	-4.442	-4.393
			(3.102)	(3.120)	(3.111)	(3.146)	(3.135)	(3.195)	(3.258)	(3.257)	(3.280)
ts:t3: -2				1.802	2.020	2.004	1.764	1.380	1.498	1.084	0.945
				(3.034)	(3.032)	(3.046)	(3.020)	(3.047)	(3.108)	(3.114)	(3.155)
ts:t4					4.098	4.460	4.613	4.514	4.363	3.707	3.562
					(3.002)	(3.043)	(3.016)	(3.024)	(3.080)	(3.103)	(3.141)
ts:t5						2.860	2.022	2.503	2.485	3.498	3.630
						(3.296)	(3.416)	(3.454)	(3.481)	(3.536)	(3.574)
ts:t6							-3.000	-4.116	-4.054	-3.721	-3.527
							(3.305)	(3.464)	(3.504)	(3.505)	(3.577)
ts:t7								-3.719	-3.869	-4.437	-4.509
								(3.438)	(3.568)	(3.581)	(3.607)
ts:t8									-0.761	0.308	0.260
									(3.315)	(3.380)	(3.405)
ts:t9										5.176	5.508
										(3.325)	(3.463)
ts:t10											1.208
											(3.339)
Constant	-0.832^{***}	-0.832^{***}	-0.849^{***}	-0.841^{***}	-0.827^{***}	-0.818^{***}	-0.819^{***}	-0.827^{***}	-0.827^{***}	-0.812^{***}	-0.811^{***}
	(0.022)	(0.025)	(0.027)	(0.031)	(0.037)	(0.040)	(0.042)	(0.044)	(0.045)	(0.047)	(0.051)
Observations	155	155	155	155	155	155	155	155	155	155	155
R^2	0.977	0.980	0.980	0.980	0.980	0.980	0.981	0.981	0.981	0.982	0.982
Adjusted R^2	0.976	0.978	0.978	0.978	0.978	0.978	0.979	0.978	0.978	0.978	0.978
Residual Std. Error	0.050 (df = 147)	0.048 (df = 145)	0.048 (df = 143)	0.048 (df = 141)	0.048 (df = 139)	0.048 (df = 137)	0.048 (df = 135)	0.048 (df = 133)	0.048 (df = 131)	0.048 (df = 129)	0.048 (df = 127)
F Statistic	891.326^{***} (df = 7; 147)	770.324^{***} (df = 9; 145)	633.228^{***} (df = 11; 143)	529.681^{***} (df = 13; 141)	462.447^{***} (df = 15; 139)	404.549^{***} (df = 17; 137)	370.867^{***} (df = 19; 135)	334.223^{***} (df = 21; 133)	301.783^{***} (df = 23; 131)	278.788^{***} (df = 25; 129)	254.779^{***} (df = 27; 127)
Note:											*p<0.1; **p<0.05; ***p<0.01

Figure 16: First Difference Regression Results with Multiple Lags and Leads for Simulated Data

Endogeneity of Regressors

The travel time is likely exogenous to the spatial distribution of industries for two reasons. First, the Interstate Highway was planned as a national defense network independent of the economic growth of different regions, a fact utilized by the literature for identification (Baum-Snow, 2007; Michaels, 2008; Duranton and Turner, 2014). Second, the travel time and spatial GINI's are aggregates for the entire nation. Any changes in spatial distribution are unlikely to change the travel time because the roads were already going to be built, the route is influenced by external factors such as the cost of construction based on the grade and strata, and the timing of construction is influenced by random factors as well such as weather, local politics, and construction delays. Second, even if the placement of roads is systematically adjusted by industry stakeholders lobbying, it is unclear the effect this would have on the aggregate travel time as it could raise or lower it depending on the position of the industry in relation to paths in-between MSA's.

If ts and bl are affected by the change in travel time, the change in spatial GINI from a change in travel time would be:

$$\frac{\partial \text{spatialGINI}_{it}}{\partial tt_t} = \beta_0 + \beta_3 ts_{it} + \beta_4 bl_{it} + \beta_1 \frac{\partial ts_{it}}{\partial tt_t} + \beta_2 \frac{\partial bl_{it}}{\partial tt_t} + \beta_3 tt_{it} \frac{\partial ts_{it}}{\partial tt_t} + \beta_4 tt_{it} \frac{\partial bl_{it}}{\partial tt_t}$$

To refute this, we observe that the change in trucking input share and backward linkages is very low across time. The largest mean normalized variance (index of dispersion) across time among industries for ts is .017, while the mean is .0018, both of which are considered to be not very dispersed. For bl across industries the largest mean normalized variance across time is .022 and the mean is .0055, which again is not very dispersed. Because these two terms are changing very little across time we can consider the last four terms in the differential equation to be zero. See Figures 10 and 11 in the appendix.

Regarding the possibility of an unobserved variable driving the change in spatial distribution– although including a time effect makes it impossible to identify the coefficient on tt, it still allows for the interaction of unobserved time effects, yet including a dummy variable for year does not substantially change the outcome (seen as FE2 in the regression table), nor does including a time trend. Additionally, leaving out tt entirely and replacing it with a time trend does not yield the same results, suggesting the movements in tt are meaningful beyond it's trend component. See Table 3 in the appendix for these regression results.

References

Andrews, Donald W.K. and Ray C. Fair. 1992. Estimation of polynomial distributed lags and leads with end point constraints. *Journal of Econometrics*, 53:1-3, 123-139.

Alder, Simon. 2016. Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development. University of North Carolina at Chapel Hill.

Alonso, William. 1960. *Location and Land Use*. Harvard University Press, Cambridge MA.

Baum-Snow, Nathaniel. 2007. Did Highways Cause Suburbanization?. *The Quarterly Journal of Economics* 122 (April): 775-805.

Börjesson, Maria, Gunnar Isacsson, Matts Andersson, and Christer Anderstig. 2019. Agglomeration, productivity and the role of transport system improvements. *Economics of Transportation*, 18:27-39.

Brambor, Thomas, William Roberts Clark, and Matt Golder. 2006. *Political Analysis* 14: 63-82.

Brox, James A. and Christina Fader, 1997. Assessing the impact of JIT using economic theory. *Journal of Operations Management* 15: 371-388.

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

Cook, Gary A. S., Naresh R Pandit, Jonathan V Beaverstock, Peter J Taylor, and Kathy Pain. 2007. The role of location in knowledge creation and diffusion: evidence of centripetal and centrifugal forces in the City of London financial services agglomeration. *Environment and Planning A* 39: 1325-1345. DOI: 10.1068/a37380.

Donaldson, Dave and Richard Hornbeck. 2016. Railroads and American Economic Growth: A "Market Access" Approach. *The Quarterly Journal of Economics*, 131:2, 799–858.

Duranton, Gilles, Peter Morrow, and Matthew Turner. 2014. Roads and Trade: Evidence from the US. *The Review of Economic Studies* 70:5, 1741-1779.

Eriksson, Katherine, Katheryn Russ, Jay C. Shambaugh and Minfei Xu. Trade Shocks and the Shifting Landscape of U.S. Manufacturing. *National Bureau of Economic Research Working Papers*. Cambridge, MA.

Faber, Benjamin. 2014. Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System. *Review of Economic Studies* 81: 1046-1070.

Frye, Dustin. 2016. Transportation Networks and the Geographic Concentration of Industry. Vassar College, Poughkeepsie, New York.

Fujita, Masahisa, Paul Krugman, Anthony J. Venables. 1999. *The Spatial Economy. Cities, regions and international trade*. MIT Press.

Hannan, E. J., and P. M. Robinson.1973. Lagged Regression with Unknown Lags. *Journal of the Royal Statistical Society. Series B (Methodological)*, 35:2, 252-67.

Hoover, Edgar M. and Raymond Vernon. 1959. *Anatomy of a Metropolis*. Harvard University Press. Cambridge, Massachusetts.

Jaworski, Taylor, Carl T. Kitchens and Sergey Nigai. 2018. The Interstate Highway System and the Development of the American Economy. University of Colorado, Florida State University, University of Colorado.

Leduc, Sylvain and Daniel Wilson. 2012. Roads to Prosperity or Bridges to Nowhere? Theory and Evidence on the Impact of Public Infrastructure Investment. *National Bureau of Economic Research Working Papers*. Cambridge, MA.

Li, Jianling and Elizabeth Whitaker. 2018. The impact of governmental highway investments on local economic outcome in the post-highway era. *Transportation Research Part A: Policy and Practice*, 113: 410-420.

Marshall, Alfred. 1890. *Principles of Economics*. Macmillan, London.

Michaels, Guy. 2008. The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System. *Review of Economics and Statistics*, 90:4, 683-701.

Morten, Melanie and Jaqueline Oliveira. 2018. The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City. Stanford, CA.

Niebuhr, Annekatrin, Nadia Granato, Anette Haas, and Silke Hamann. 2012. Does Labour Mobility Reduce Disparities between Regional Labour Markets in Germany? *Regional Studies*, 46:7 841-858.

O'Donoghue, Dan and Bill Gleave. 2004. A Note on Methods for Measuring Industrial Agglomeration. *Regional Studies*, 38:4, 419-427.

Panzera, Domenica and Paolo Postiglione. 2019. Measuring the Spatial Dimension of Regional Inequality: An Approach Based on the Gini Correlation Measure. *Social Indicators Research*, 148: 379-394.

Paredes, Dusan, Victor Iturra and Marcelo Lufin. 2016. A Spatial Decomposition of Income Inequality in Chile, *Regional Studies*, 50:5, 771-789.

Pelegrín, Angels and Catalina Bolancé. 2008. Regional Foreign Direct Investment in Manufacturing. Do Agglomeration Economies Matter? *Regional Studies*, 42:4, 505-522.

Jiwattanakulpaisarn, Piyapong, Robert B. Noland and Daniel J. Graham. 2012. Marginal Productivity of Expanding Highway Capacity. *Journal of Transport Economics and Policy*, 46:3, 333-347.

Redding, Stephen. 2005. Spatial income inequality. *Swedish Economic Policy Review*, 12:29-55.

Redding, Stephen and Matthew Turner. *Transportation Costs and the Spatial Organization of Economic Activity*. NBER Working Paper, 2014.

Rey, Sergio and and Richard John-Smith. 2012. A spatial decomposition of the Gini coefficient. *Letters in Spatial and Resource Sciences* 6: 1-16.

Rothenberg, Alexander. 2013. Transport Infrastructure and Firm Location Choice in Equilibrium: Evidence from Indonesia's Highways. University of California, Berkeley.

Sayer, Andrew. 1986. New developments in manufacturing: the just-in-time system. *Capital Class*, 10:3, 43-72.

Stecke, Kathryn, and Xuying Zhao, 2007. Production and Transportation Integration for a Make-to-Order Manufacturing Company with a Commit-to-Delivery Business Mode. *Manufacturing and Service Operations Management* 9(2): 206-224.

Sutton, Paul. 2012. The 'Spatial GINI' Coefficient: An empirical satellite imagery derived metric characterizing the co-distribution of light and people. University of Denver.

U.S. Department of Transportation. Federal Highway Administration. 1985. *Highway Statistics: Summary to 1985*. Washington: Government Printing Office. [p.179-185]

Vaisey, Stephen and Andrew Miles. 2014. What You Can—and Can't—Do With Three-Wave Panel Data. *Sociological Methods Research* 46 (January): 44-67.

von Thunen, Johann Heinrich. 1826. *The Isolated State*. Hamburg: Perthes. English translation. Oxford: Pergamon, 1966.

Weber, Alfred. 1909. *Theory of the Location of Industries*. [translated by Carl J. Friedrich] Chicago: The University of Chicago Press, 1929.

Weingroff, Richard. 2006. The Size of the Job. FHWA. `https://www.fhwa.dot.gov/infrastructure/50size.cfm`.

Weingroff, Richard. 2017. Before the Federal-Aid Highway Act of 1956: Francis V. Du Pont In Context . FHWA. `https://www.fhwa.dot.gov/highwayhistory/dupont.pdf`.