The Impact of Roads on Agglomeration

Thor M Dodson

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Robert Halvorsen, Chair

Levis Kochin

Alan Griffith

Program Authorized to Offer Degree:
Economics

University of Washington

Abstract

The Impact of Roads on Agglomeration

Thor M Dodson

Chair of the Supervisory Committee:
Professor Robert Halvorsen
Department of Economics

This dissertation contains three essays on the impact of roads on agglomeration. In the first essay, I show the US Interstate Highway System had a significant impact on the agglomeration and dispersal of different industries due to differing sensitivities to increased economic centripetal and centrifugal forces generated by a reduction in transportation costs. This study suggests the impact depends on truck-transportation utilization and backward linkages. I construct travel time estimates by representing the US highway system as a network each year during its construction utilizing a dataset of completion dates for each segment. I combine this with county level earnings data by industry to construct a measure of spatial inequality as a proxy for agglomeration. I conduct a panel regression with multiple industries across time including interaction efects, individual and time effects, utilize regional variation in the timing of highway completion to support the finding.

The second essay examines the set of literatures regarding roads and the economy. Following the timeline of the development of thought, I examine and discuss some of the key works from early location theory, central place theory, urban economics, cost-benefit analysis, the new economic geography, market access, and graph theory.

The third essay is an exploratory spatial monopolistic competition model in the spirit of the new economic geography. In a monopolistic competition framework firms produce different types of substitutable goods using labor and other goods as intermediate inputs, competing in price and wage over a two-dimensional space where transport is costly. Households commute to work for firms and use their income to consume each type of good with a preference for variety within each type. Firms enter and exit based on profitability, establishing clustering patterns over time based on the fixed distribution of households and the relative positions of other firms. This paper fits into the literature on production networks and spatial equilibrium in that it utilizes an exogenous input-output matrix with endogenous market power to explore the locations of activities in relation to each other and agglomeration, but the spatial forces stem only from trade and does not feature any ad hoc benefits of agglomeration or household migration. I find that by altering the space of transport costs roads influence the patterns of regional activity by facilitating competition and that industries with no linkages exhibit more dispersion.

Table of Contents

List of Figures

Acknowledgments

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.

Chapter 1

The US Interstate Highway's Effect on Agglomeration

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.

1.1 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 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 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.

1.2 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 uncertainty, facilitating smoother production flows, and reducing required inventories for stocks and parts. The last three effects are particularly important, as observed in the global rise of "just-in-time" manufacturing and inventory management during the 1970s

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

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 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—wages and 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 pushes a plant away from any particular market center and towards a point of centralized distribution, as elaborated by Weber's (1909) conception of point of minimum transport.

Different industries have different sensitivities to each of these forces based on what they do. Thünen (1826) captured this idea with his model of agricultural land use and this was extended by Alonso's (1960) bid-rent theory (see Appendix for visualization). The key idea is how much 'land rent' 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 generate a higher rent for any given location are more likely to locate there as 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 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. 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. Therefore, we suspect 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 influences the sensitivity to the benefits of agglomeration and dispersal (Eriksson et al, 2020). The product life-cycle distinguishes four stages: 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, and industries dealing in early and late stage products will agglomerate and disperse respectively.

1.3 Data and Method

The Interstate Highway System began construction in 1956 enabling 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. 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 during 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. The first GIS file is formed by isolating the interstate highways from the PA_NHS 2012 shapefile[2] detailing all US roads at that time. The second GIS shape file I form by manually digitizing a 1954 map image[3] produced by the US government detailing the principle highways and arterials in existence at that time, what are now referred to as the US numbered highways. I approach the road system in this way because in addition to different routes the Interstate Highway System replaced portions of earlier roads, relying on the rest for connection. 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 [4] 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.[5]

By converting the highway system to a network, the Dijkstra algorithm[6] 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 length 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)[7]. This is done for every

---

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

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

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

[5]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.

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

[7]These speed assumptions are a simplification based on travel time estimates from AAA maps from 1955, 1996, and 2018, after isolating the speed changes from the road and vehicle improvements. Routes without

metropolitan-statistical-area (MSA) pair to generate a travel time matrix for each year.[8] On average the Interstate Highway System reduced travel times between MSA's by about 18%, although the true change in travel time reduced by faster cars and reduced by congestion. [9].

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 likelihood for over-night shipping availability. This number reflects the travel time most significant for trade within the region and to counties near the edge of the region. The regional travel times are mean-normalized to the national travel times to facilitate comparison of regression coefficients.

My data adds to the literature explicitly representing road systems, 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 and magnitude of road

---

an interstate segment experienced a rise in speed of about 5mph, likely from improvements in cars, while routes receiving interstate segments experienced rises in speed between 10-20mph, with variance likely due to congestion.

[8]The units are coordinate distance per mph

[9]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 not followed or enforced. The Motor Carrier Act of 1980 deregulating the trucking industry had many impacts potentially lowering transportation costs, which could 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

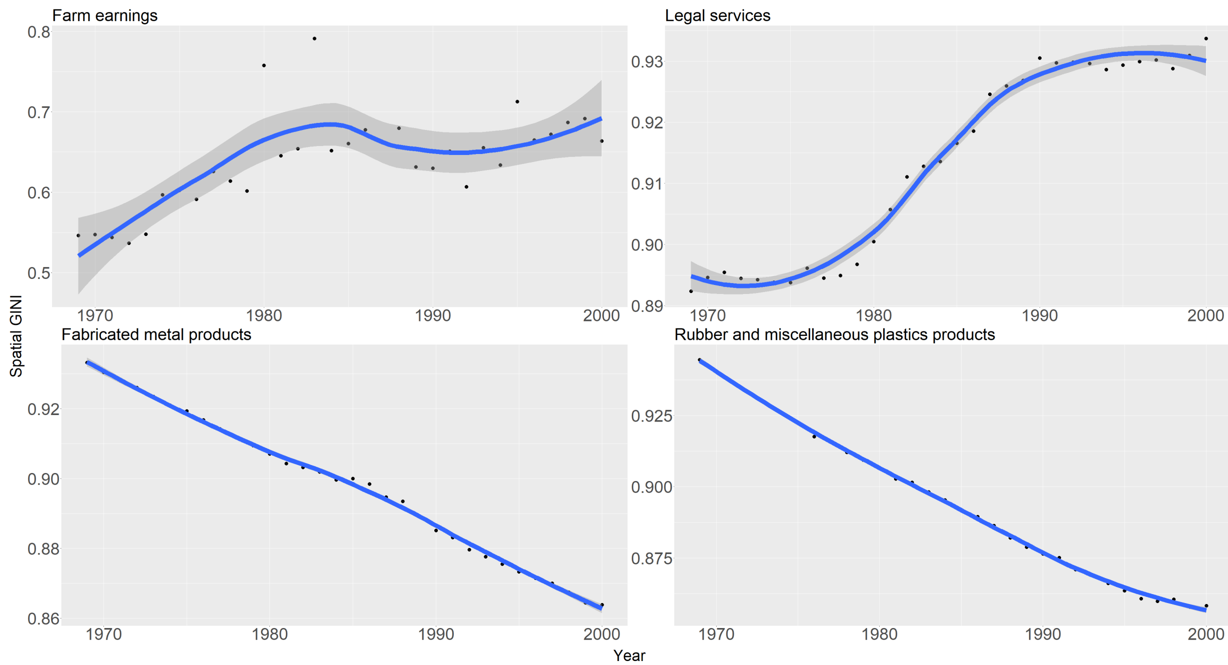


Figure 1.1: Spatial GINI for Select Industries

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 individual, regional, and time effects.

Using BEA data on county earnings by industry I construct a spatial GINI over time.[10] As shown in Dixon et al (1988), a consistent estimator for the Gini is given by

$$spatialGINI_{it} = 1 - \frac{2}{n-1}(n - \frac{\sum_{j=1}^{n} j y_{jit}}{\sum_{j=1}^{n} y_{jit}})$$

for industry i at year t. This is done for each industry for each year between 1969-2000[11] , some sample industries are shown in Figure 1. The earnings data are reported based on where the earner lives, so any commuting across counties will bias the estimate downward.

Figure 2 highlights the change in spatial GINI from 1969 to 2000 for all industries. A

[10]See this measure in detail at `https://spatial-gini-dash.herokuapp.com/`

[11]Negative earnings are set to zero.

Figure 1.2: Change in Spatial GINI

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 (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) discuss in their analysis of the distribution of people and jobs in the New York Metropolitan region, as 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 sub-urbanized customer base, especially as technology like the telephone and internet facilitate the exchange of information across distance. Information industries under this sort of influence may still disperse despite the increased ability to cluster.

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 and details each industries use of other industries in dollars. Separate measures reflecting how much an industry relies on truck transportation inputs and outputs, rather than a mix, would be preferred but are not available. With a higher truck-transportation-share of inputs an industry is more sensitive to changes in shipping costs.[12]

The theory suggests that the stage in the product life-cycle will impact agglomeration and dispersal. As a proxy for these, I utilize a 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. As detailed in Kula (2008), the backwards linkage measures the impact on supplier industries from a unit increase in final demand, and is given by

$$bl_j = \sum l_{ij}$$

$$L = (I - A)^{-1}$$

where l_{ij} is the ij^{th} element of the Leontief matrix, L, where I is the identify matrix and A is the input coefficients matrix obtained by dividing the amount of inputs by the total produced for each industry.

[12]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`

1.3.1 Method

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 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 $\underline{i}$ at year $\underline{t}$.

If $\underline{\text{ts}}$ and $\underline{\text{bl}}$ are not affected by 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 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[13]. 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 controlling 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

[13]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

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.[14]

To verify the robustness of the results to alternate 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.

1.4 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.

[14]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 significant, generating movements in the data that are more a product of 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 unsuppressed data source is not available for replication.

Table 1.1: 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

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 3. The average z-score of the marginal effect of $\underline{tt}$ 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 most of the time.

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 and backwards linkage. 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 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 the regional 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

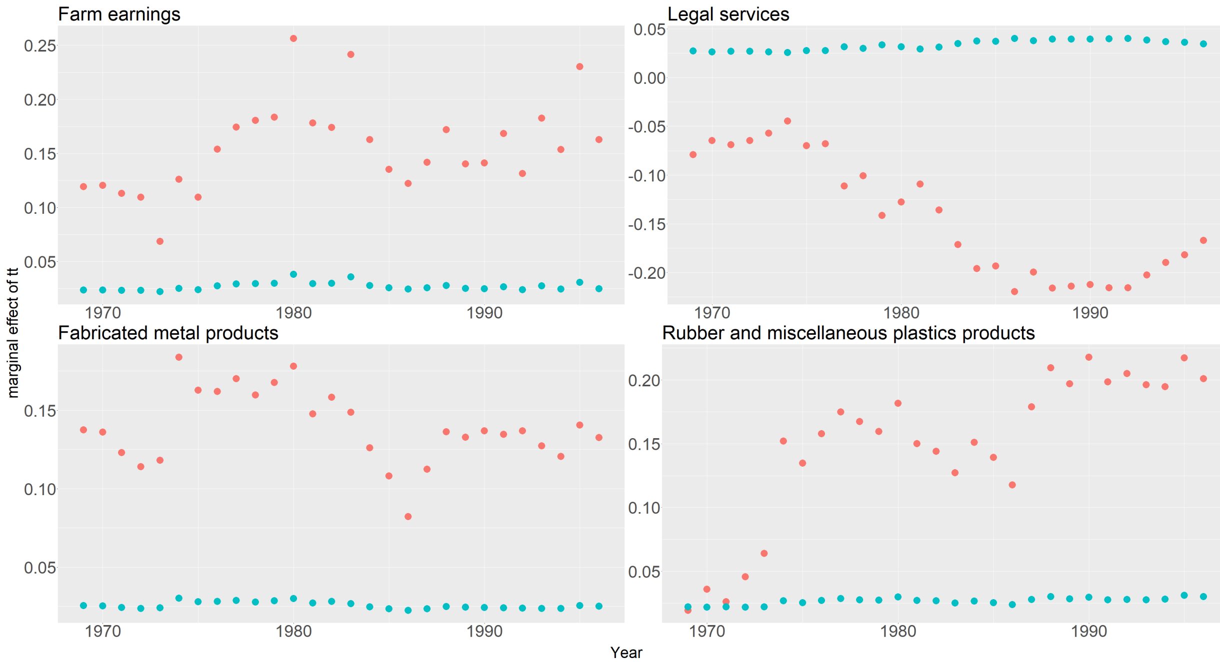


Figure 1.3: Marginal Effect of Travel Time on Spatial GINI and Standard Errors by Industry

of trucking share of inputs diminishes, suggesting that within regions these variables have slightly different importance.

1.5 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

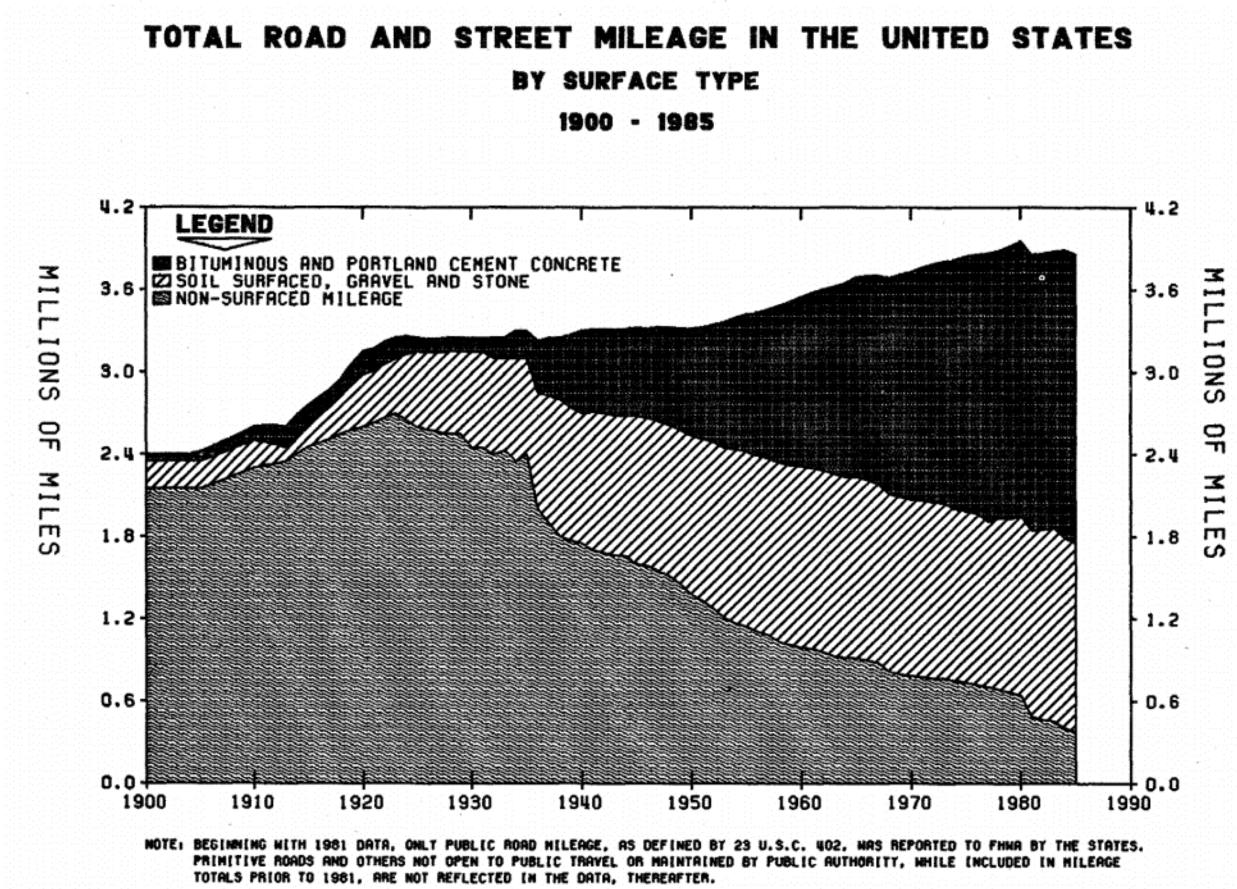


Figure 1.4: Surface Quality of US Roads, Source: US DOT 1985

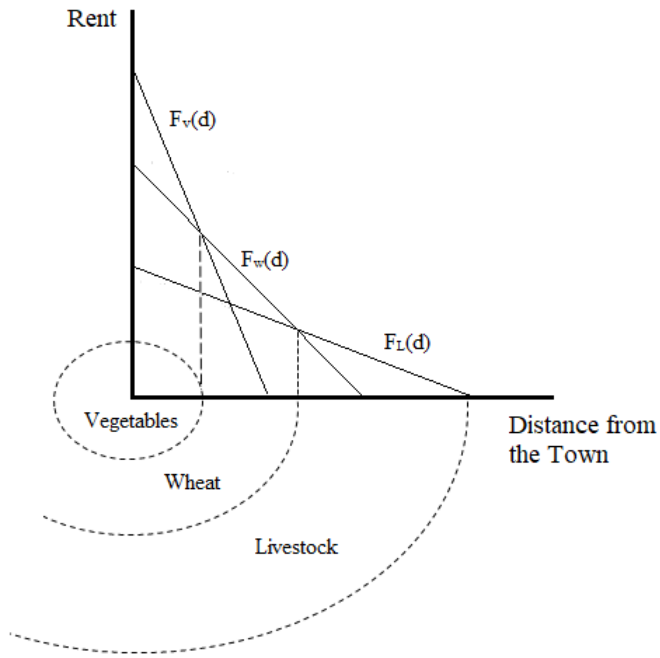


Figure 1.5: An Example Bid-Rent Curve

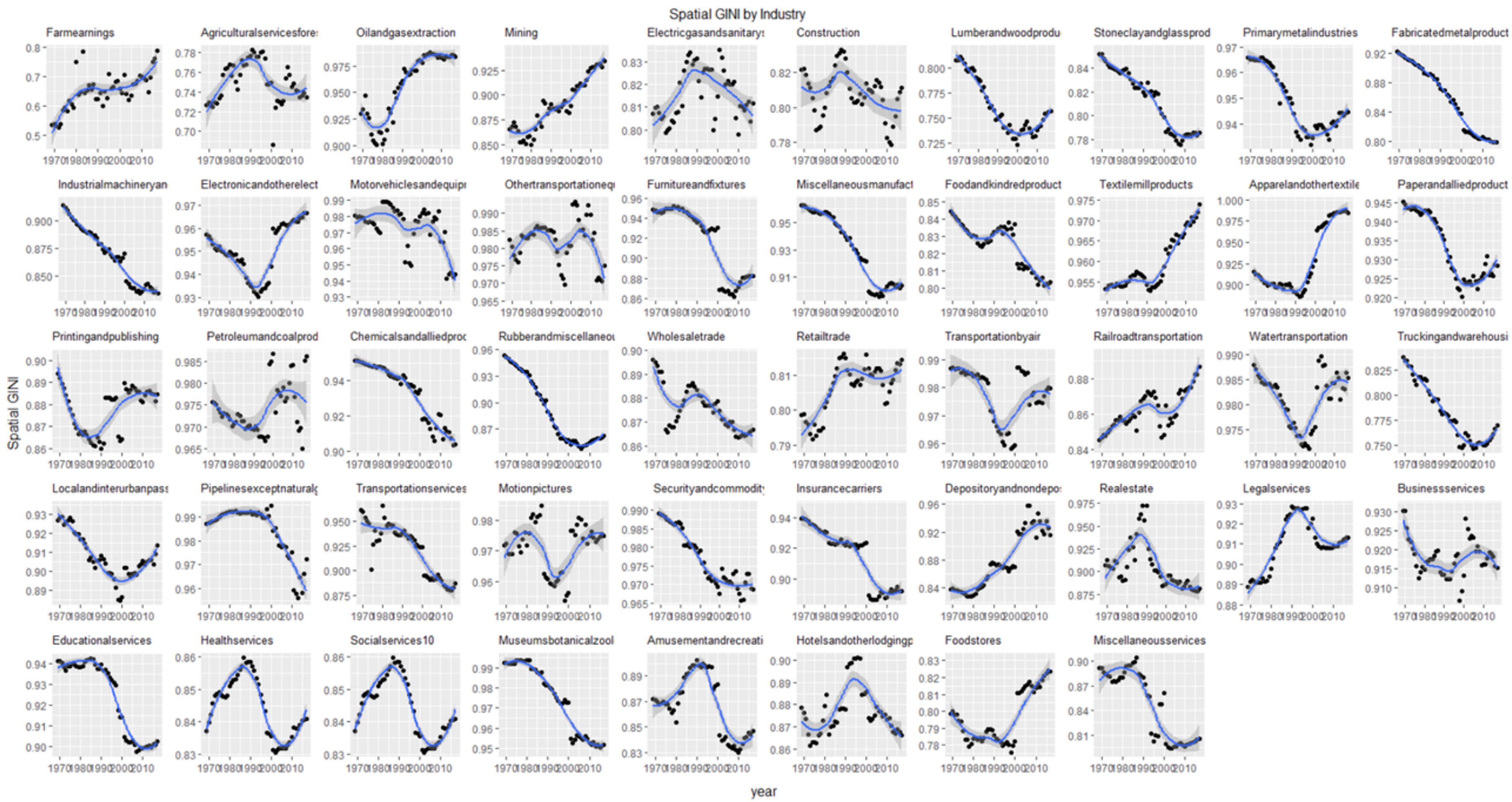


Figure 1.6: Spatial GINI by Industry before and after 2000

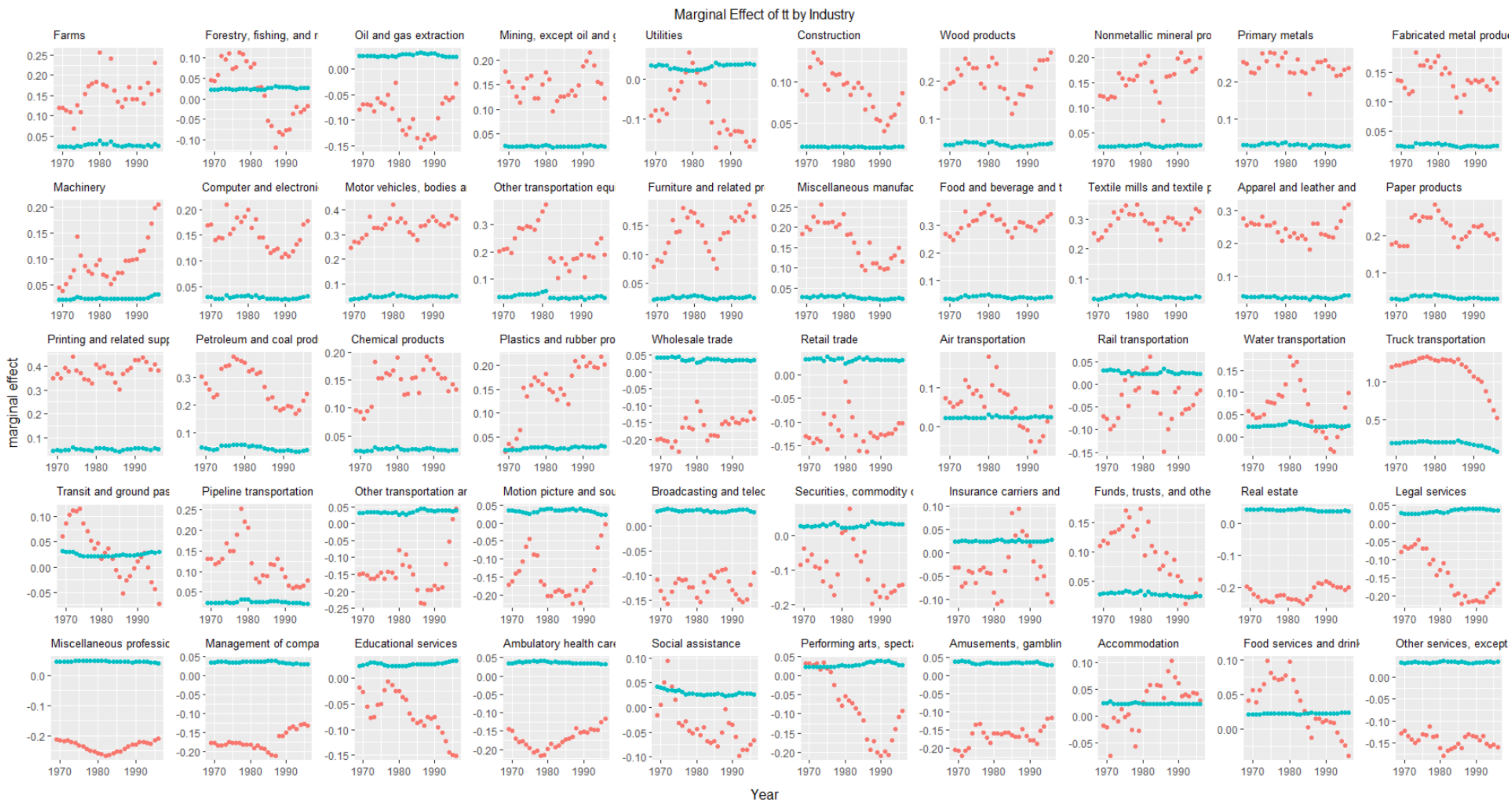


Figure 1.7: Marginal Effect of tt on Spatial GINI and Standard Errors by Industry

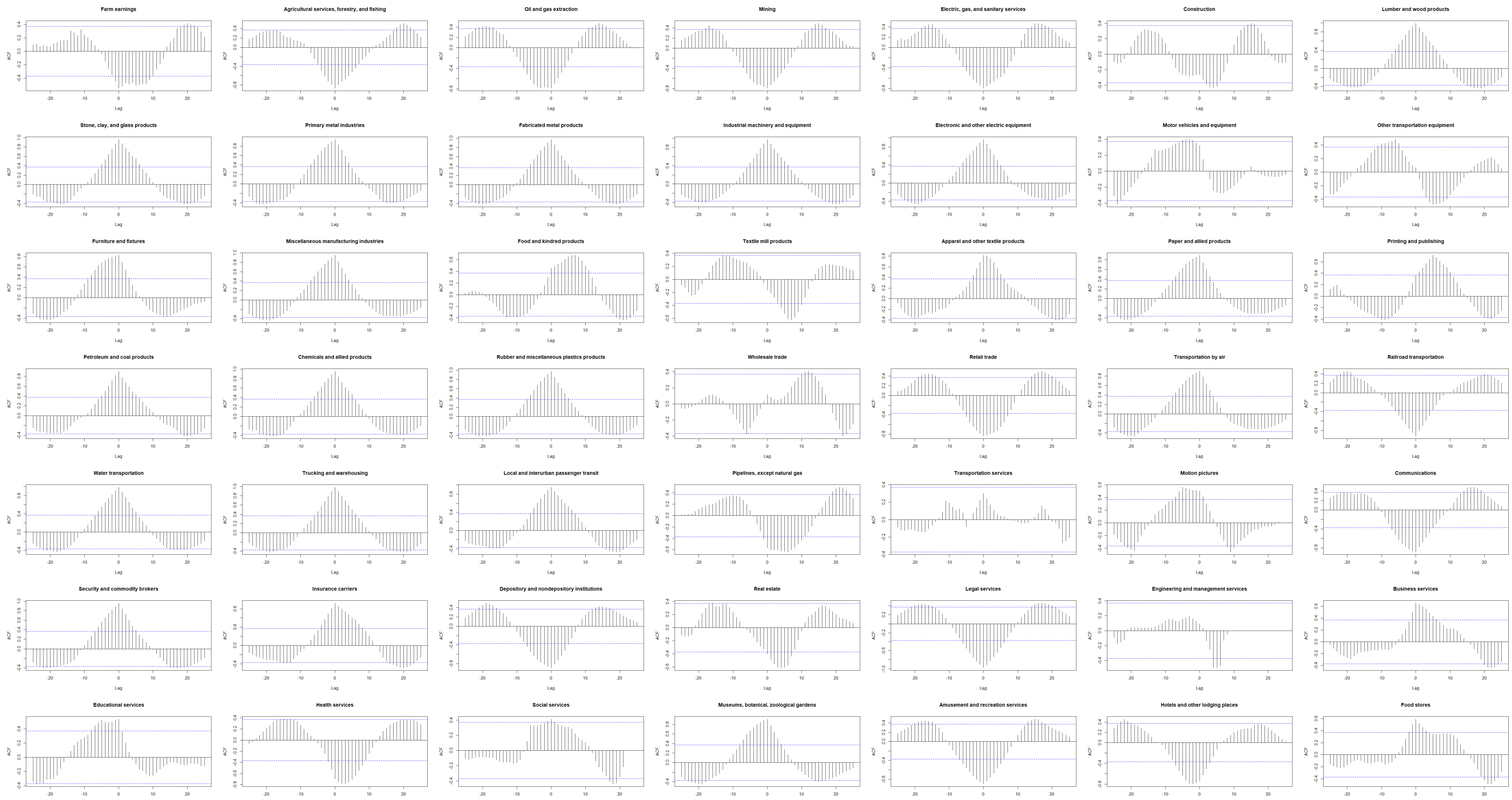


Figure 1.8: Cross Correlations for Lagged Values of Travel Time and Industry Spatial GINI

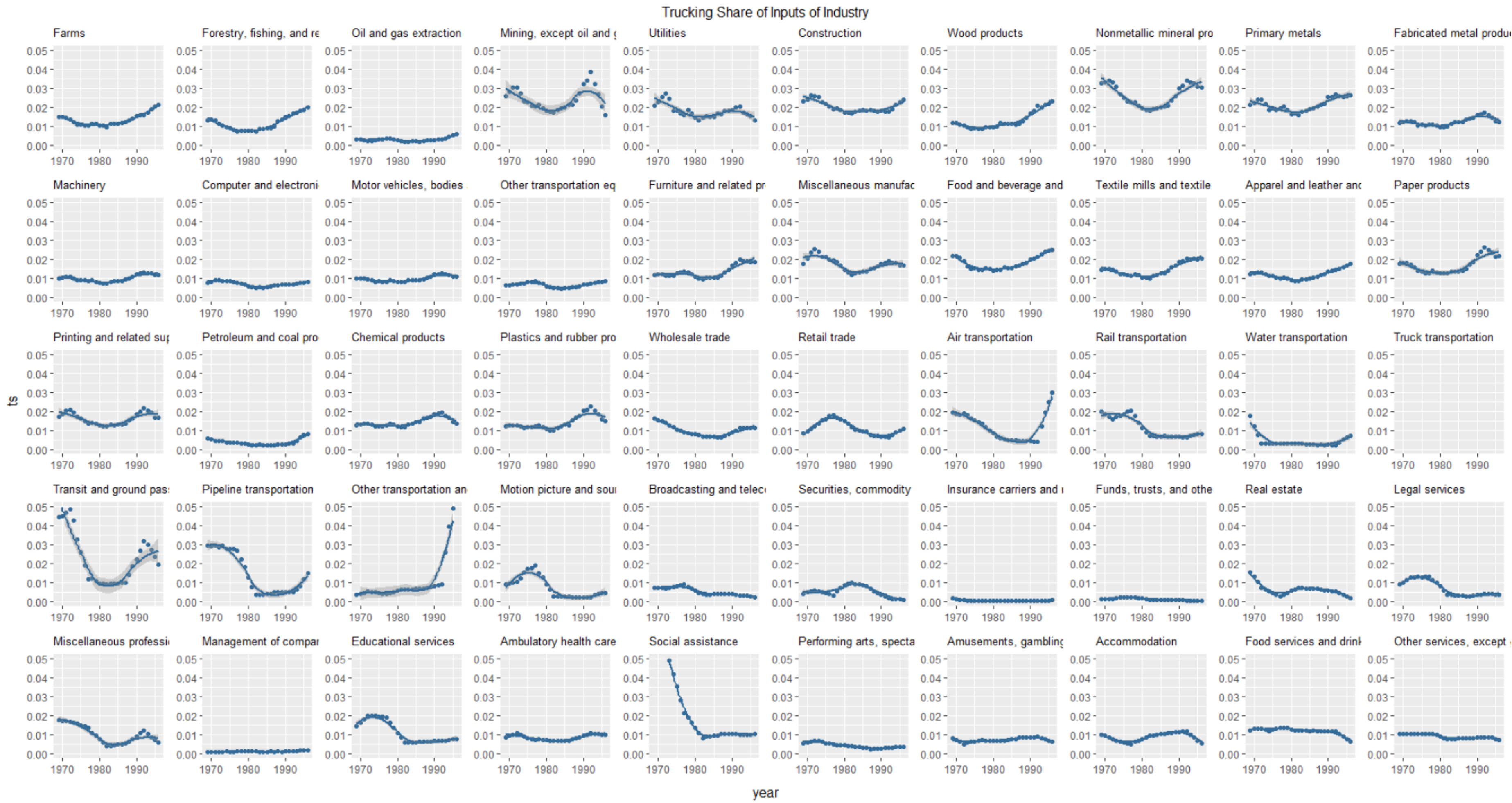


Figure 1.9: Trucking Share of Inputs by Industry

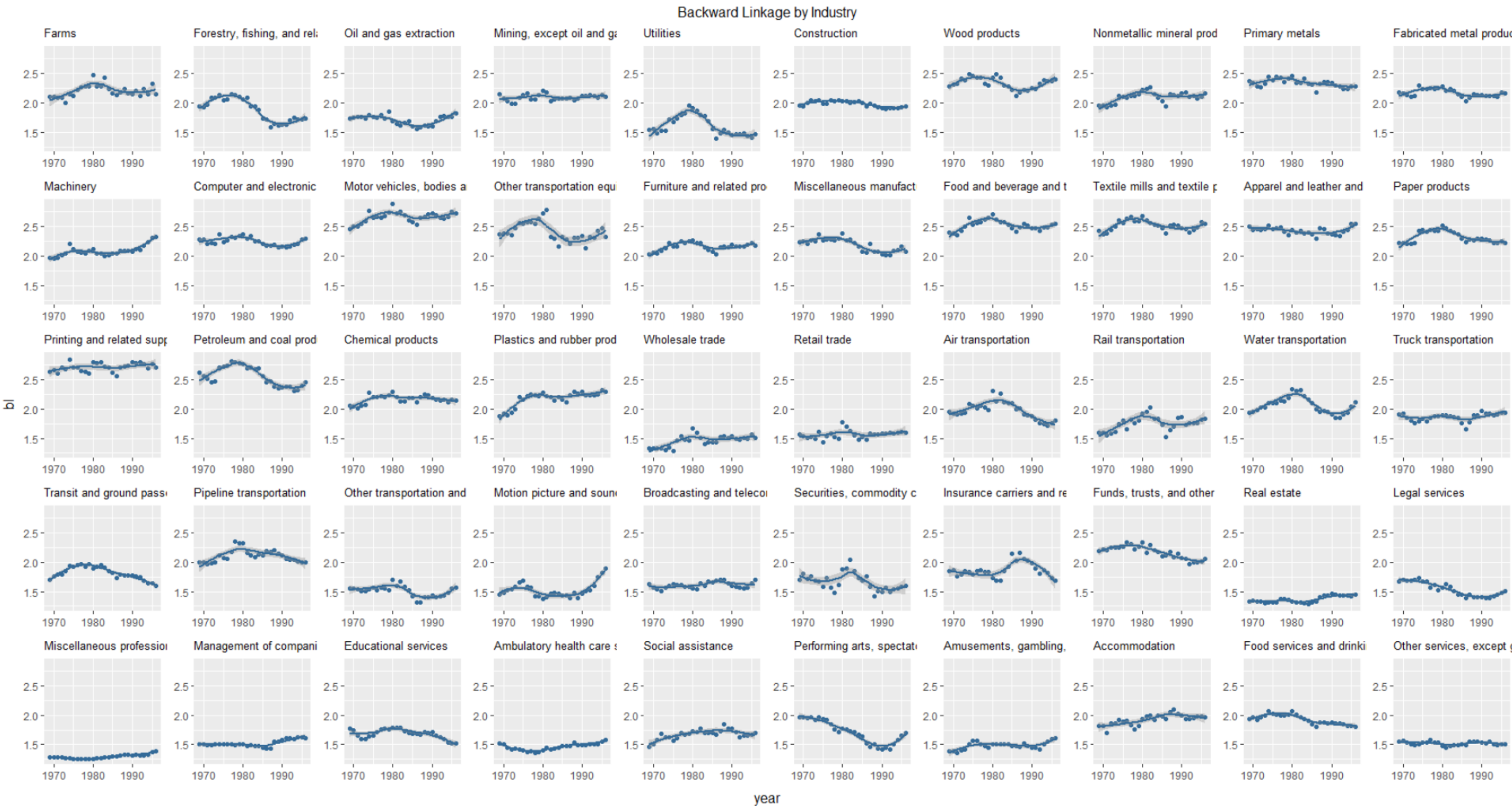


Figure 1.10: Rasmussen Measure of Backwards Linkages by Industry

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 1.4: 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 1.11: Simulation Results: Simple Trend

These results are a bit surprising, as the lagged independent variables are highly auto-

correlated, 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	ß_2	std.err_2	std.err_AF_2	ß_3	std.err_3	std.err_AF_3	ß_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	ß_5	std.err_5	std.err_AF_5	ß_6	std.err_6	std.err_AF_6	ß_7	std.err_7	std.err_AF_7	ß_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 1.12: 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 $\epsilon \sim$ N(0, σ^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 1.13: 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 1.14: 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 1.15: 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, 2012). 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 $\underline{\text{ts}}$ and $\underline{\text{bl}}$ 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 $\underline{\text{ts}}$ is .017, while the mean is .0018, both of which are considered to be not very dispersed. For $\underline{\text{bl}}$ 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 $\underline{\text{tt}}$, 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 $\underline{\text{tt}}$ entirely and replacing it with a time trend does not yield the same results, suggesting the movements in $\underline{\text{tt}}$ are meaningful beyond it's trend component. See Table 3 in the appendix for these regression results.

Chapter 2

Literature on Roads and Economy

2.1 Early Thoughts on Roads

Work on the importance of location and geographical factors like roads was scant in economic theory during the 19th century and earlier. As Cooley (1894) stated, "Since the work of Kohl, published in 1841, I know of no comprehensive and connected investigation of that branch of demography, or demographic sociology, that treats of the forces and laws that determine the territorial distribution of persons and wealth". Kohl's work, including *Traffic and Settlements of People with Regard to their Dependence on the Morphology of the Earth's Surface*, explored theoretical geography and the influence of man's relationship with nature on the transportation network and settlements (Peucker, 1968). Kohl develops a general theory for the functional relationship between movement and the development of towns and their mutual interdependence, exploring the physical basis needed for different modes of transportation, communications lines and networks, and eventually developing an ideal spherical city with skyscrapers and underground shopping centers (Peucker, 1968). Kohl even discusses the impact of agglomeration economies and of the uneven distribution of resources on the regional differentiation of towns, suggesting that the "more valuable the basic product is, the more difficult its extractions, the greater its concentration in a certain place or frequency in a certain area is, the more important the settlement which it creates will be" (Peucker, 1968). Cooley (1894) similarly explores the relationship between land transportation, physical geography, and the ideal spatial organization of society giving special attention to the compromise between the use of land for economy and transportation, suggesting a complexly branched system of transport lines connecting central points of distribution is ideal. Additionally, Cooley (1894) explores the significance of transport in determining rent, suggesting that even without varying soil fertility or productivity which generates land rent in the Ricardian theory, differences in the cost of transporting commodities to market give rise to advantages in locations and therefore land rent.

Cooley, like many others, missed the work of German land owner and economist von Thünen (1826) on the same subject. In his work *The Isolated State*, von Thünen explores the distribution of different agricultural commodities around a central marketplace. In von Thünen's model, whose parameters on transportation cost, wages, and market prices he collected from his own experiences farming, the cost of transporting each agricultural product, as well as the market price, demand for each good, cost of transporting inputs, soil yield, and labor intensity, determines where the product will be grown. In his theorized uniform plain concentric rings specialized in a particular good form around the town center as the types of goods that benefit the most and are therefore able to pay the highest rent settle into their optimal location given the other types of products. He even extends his model to consider transportation channels like rivers and roads and explores the implied distortions for his agricultural ring model. This work is considered a classic and although unnoticed for a relatively long period of time provided the foundation for later work in location theory and urban economics and is discussed heavily today.

Alfred Marshall (1919) discusses an interesting result of declining transport costs–that "an increase in the distance which goods can be carried at a given cost is likely to increase the trade in those goods in a greater ratio". Marshall terms this rule 'Lardner's Law of Squares in transport and trade', resulting from the simple fact that the area of a circle varies as the square of its radius. Cooley (1894) also explores this phenomenon, suggesting that as the efficiency of transport increases (given by speed or cheapness in cost) the radius of the habitable circle around a town center increases with the square. However, the extent of this law is limited by the coverage of the transportation network as seen in Figure 2.1 below showing the market area for a given transport cost.

As seen in the figures, for transport largely restricted to a road system the increase in market area in proportion to the reduction in transport costs is likely far less than the square. When transport costs are reduced by half the market area increases by little over double, although this increase is amplified the more saturated the road network is. This has more implications for improvements in vehicle efficiency than improvements in the road system, as individual roads are generally improved rather than the entire system, however the Interstate Highway System construction does represent an instance. Marshall points out 'Lardner's Law of Squares' has much more relevance for trading ports close to archipelagos or river deltas such as "Athens, Alexandria, Byzantium, Marseilles...Venice, the Hanseatic League, and Holland", as well as for travel over open water when routes are not restricted,

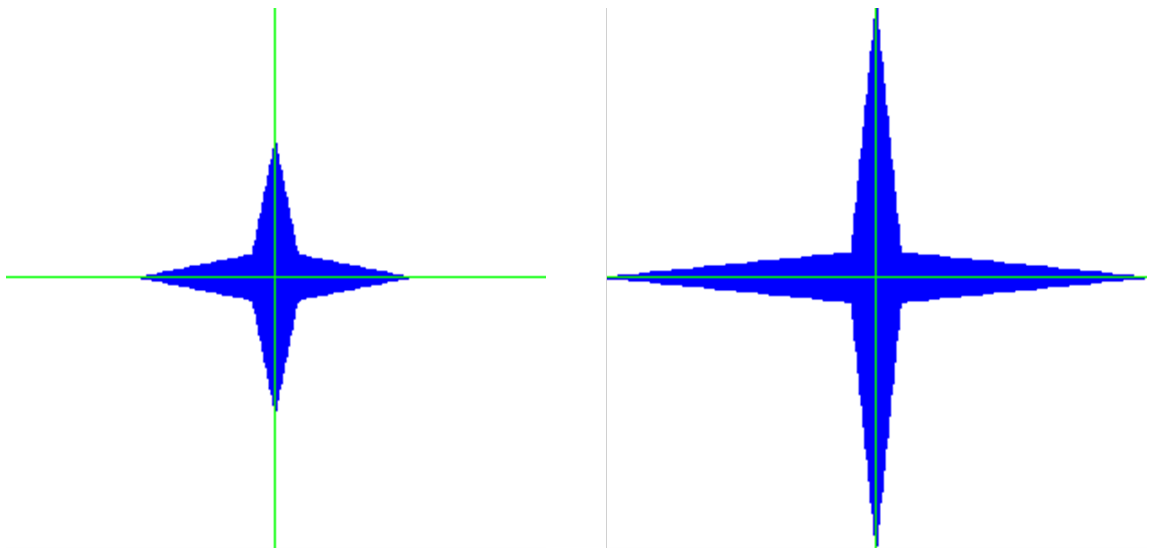
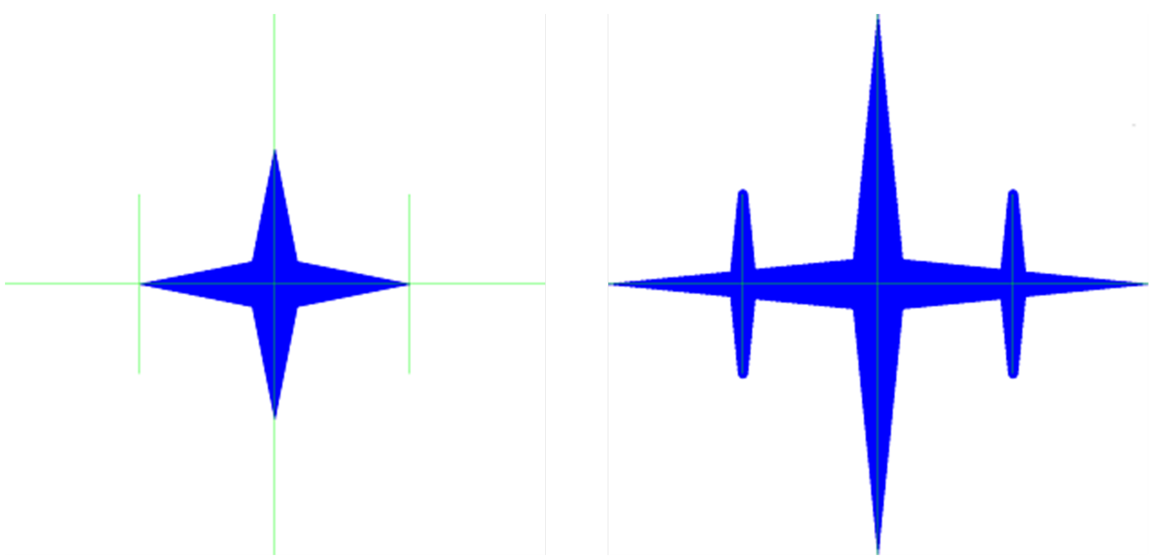


Figure 2.1: Lardner's Law of Squares in Transport and Trade for Road Systems

The blue shaded area shows the market area for a firm at the center–regions accessible for a given total transport cost. The green lines represent roads. The transport cost is t*distance. Along roads t = .2 for the left figures, and t = .1 in the right figures, for travel outside the roads t = 1 The ratio of market areas for the low to high transport cost is 2.20 for the case above, but 2.64 for the figure below with additional roads. As the road system is more saturated the increase in market area is larger and approaches Lardner's Law.



as evident in the rise of England as transoceanic travel became commonplace. Subsequently however, he notes her lesser benefit from the extending net of railways compared to large inland countries. Using this law, we can see the importance location and available modes of transport play in regional domination, which leads to interesting questions regarding the current state of communication technology and how market areas expand for digital goods.

In the case of the Interstate Highway, as the road system is improved the market area for a given firm expands, spurring economies of scale and generating competition between firms previously too far apart to interact. Marshall highlights the boundary of firms market areas will shift based on the proportion of transport cost reductions between them, a topic explored in detail in Chandra and Thompson (2000). As noted in Rothenberg (2011), this increase in market area tends to be larger for industries serving perishable goods.

Adam Smith (1776), in his discussion of expenses of the sovereign or commonwealth, reasons in favor of road tolls in proportion to the weight (and therefore the wear) of the vehicle, arguing "when high roads, bridges, canals, etc., are in this manner made and supported by the commerce which is carried on by means of them, they can be made only where that commerce requires them, and consequently where it is proper to make them". Roads financed this way by necessity must be located where it is economical for them to be, and are thus immune to political clout. This manner of road building may exclude socially beneficial roads if it's use generates positive externalities the users are not willing to pay for themselves.

Smith even suggests that "when the toll upon carriages of luxury upon coaches, post-chaises, etc., is made somewhat higher in proportion to their weight than upon carriages of necessary use, such as carts, waggons, etc., the indolence and vanity of the rich is made to contribute in a very easy manner to the relief of the poor, by rendering cheaper the transportation of heavy goods to all the different parts of the country". However, Smith pushes back against a weight based toll as a means to generating revenue, noting that cheap commodities tend to be coarse and bulky rather than precious and light, therefore tending to place the expense of the toll upon the poor who are the least able to supply it. He goes on to discuss the nuances of road management and the tradeoffs between private and public management of roads and navigable canals as they differ in maintenance requirements and impassability without repair, as well as the likelihood of governments to ignore small projects with utility, preferring large visible projects that generate political favor.

Given the logic of roads financed by tolls necessarily being placed in economically beneficial locations, why aren't more roads financed by tolls? Furthermore, do modern roads pay for themselves? Prior to the 20th century, most roads were privately built and financed by toll collection. In fact over 3000 private turnpike companies built between 30,000-52,000 miles of turnpike during the 19th century, nearly equivalent to the 48,440 miles of Interstate Highway (Samuel, 2007). As vehicles became faster, toll collection costs rose in proportion to the benefit of turnpikes, but toll roads were still prominent before 1950. Once the Interstate Highway was built (explicitly as a toll-less freeway), turnpikes suffered from the competition. Tolls also suffer from restrictive policies surrounding rate setting, a 2015 study finding that 80% of sampled toll operations in the U.S. have rates fixed by authority or contract (Beatty, 2015). Many operations do not even have a methodology for calculating the cost of collecting a toll, although the authors estimate costs around $.25-1 per transaction (Beatty, 2015). In the U.S. tolls are still used for many bridges and some turnpikes in the East, but there are currently only around 5000 miles of toll roads in relation to the approximately 3.4M miles of public roads.

The construction of the Interstate Highway was funded by a gasoline tax as it was thought it would not be able to sustain itself on tolls, a 1955 study estimating that less than 9000 of the proposed miles would have enough traffic to be able to pay for themselves (Time 1955). While some have claimed that the Interstate paid for itself through the gasoline tax others are critical that the funds came from traffic over the entire U.S. road system and not just the Interstate. Looking at expenditures on roads minus the revenue generated by the gas tax and other user revenues, the cumulative loss has only risen over time, reaching $600B in 2007 (Dutzik and Davis, 2011). Dutzik and Davis further estimate that user fees only pay for about half of the cost of building and maintaining the nation's network of highways, roads and streets. This is partially due to increases in fuel efficiency, changes in travel behavior, and the political difficulty of raising the gas tax. Furthermore, there are incentives for state governments to build roads over other forms of transportation infrastructure because of the Federal governments matching 80% of funds spend on highway expansion compared to 50% for transit related, many states' law restricting gasoline taxes to highway expenditures, and the 'use it or lose it' nature of federally matched funds. It's possible many of these roads do still provide a net benefit to society, particularly for their role in national security, emergency relief, and industry restructuring leading to agglomeration and productivity gains, but vehicle traffic has a long list of negative externalities as well, including congestion, accidents, pollution, and national security implications of protecting access to imported fossil fuels. Another study sampled seven highways in Texas, finding the percentage of costs paid for by user revenue ranged from 13-93% (The Highway Construction Equity Gap, 2008).

Early work discussing location theory and transportation networks covered much of the theoretical ground of later works. However, with the exception of von Thünen, much of this work was not unified with an economic framework, and even in the case of von Thünen, only offered examples of partial equilibrium. As economic theory itself developed over the next century, location theory and urban economics developed alongside it, incorporating modeling frameworks and working to endogenously explain observed variations in settlement patterns and densities.

2.2 Early Location Theory

Alfred Weber (1929) set out to develop a general theory of location establishing rules and factors influencing the location of manufacturing industries. He begins by distinguishing aspects of the production process that vary with location into general factors influencing all industries, including the costs of transporting material and product, labor, and the cost of land, from special factors influencing only some industries such as perishability, the influence of humidity on the manufacturing process, and the need for water as an input. In considering the general factors he initially abstracts away from varying costs of labor and land to focus on the influence of transportation costs, concluding that the predominant factors are the weight of the materials and the distances they must be transported.

He proposes that given a set of production requirements, uniform labor and land costs, uniform surfaces, and uniform transport rates, the optimal location for the plant is the one with the fewest ton-miles. Of essential consideration then is the material index—the ratio of the total weight of localized material inputs over the weight of the finished product. Localized material refers to inputs that can only be obtained from certain locations, in contrast to ubiquities such as air or certain commodities which are used as inputs but are equally available at all locations. For processes with a material index less than one, meaning the final product is heavier than the sum of the localized material weights, because of the higher cost of transporting the final product the optimal location is at the center of consumption. For weight losing processes, such that the material index is greater than one, the optimal location to minimize transport costs will be somewhere within the locational figure of the inputs and center of consumption, tending towards the points with heavier inputs.[1]

Using this theory, he predicts some interesting outcomes based on the changes in economic trends. First, that as development tends to lead to denser population concentrations, the increased demand makes ubiquities less available, potentially making them localized materials that must be acquired from an alternate location. This process would tend to drive industry location towards the source of inputs as it increases the material index. Second,

[1]for a visualization of Weber's Location Triangle see `https://transportgeography.org/contents/chapter2/transport-and-location/weber-location-triangle/`

development typically involves increasing control over nature, meaning increased extraction and refinement of materials, as well as increasing mechanization utilizing fuel and metal products. Both of these lead to an increasing material index from a larger weight loss in the production process, further driving industry location towards the sources of input. Additionally, he suggests that these processes, as well as other factors, contributed to the destruction of the crafts which presupposed that industrial location and consumption location coincided.

With the basic formulation of the general theory of location, Weber then relaxes the assumptions and discusses the implications. First, the reality that transport rates vary significantly based on vehicle strength requirements, excess capacity or thresholds, insurance rates, and other product specific considerations. His solution to accommodate these within the framework is to simply adjust the milage for each material in proportion to its rates, so goods with a higher rate are considered as if their distance traveled is farther, and vice versa. Second, the reality that surfaces are not uniform, but consist of a variety of terrain and that transport is often restricted to specific rail lines, waterways, or highways. The solution to this is that the locational theory used must necessarily be an approximation, and the true location for the plant should lie on the transportation network at the point closest to the optimal point if there was no restricting network. For differing rates between transport modes, the distance traveled along those lines must be adjusted accordingly as mentioned. For a firm considering multiple sources for the same inputs, locational figures should be constructed for each arrangement and the total transport cost should be compared between them. Third, the reality that labor costs for standardized tasks are not uniform, but rather vary across space as well as at specific points. Weber first points out that although wages vary significantly across space, citing union rates in Germany, often higher wages are associated with higher efficiency, and that the labor cost per unit of production may vary substantially less. Regardless, in the case of varying labor cost per unit across space, his accommodation must work under the assumption that local labor costs are fixed and are not limited in supply. Then, by constructing isodapanes, curves of equal total transport cost surrounding the minimum point, the question in considering a point with lower labor cost away from the minimum transport point is whether the rise in cost per ton for transport is less than the labor savings.

Weber's ideas are still frequently cited and taught. Criticisms suggest his work was too complicated and abstract, spending too much time on the geometrical figures which are not strictly necessary to comprehend the problems (Krzyanowski, 1927), and that the model is extremely partial "insofar as the levels, mixes, and market prices for both outputs and inputs (most variables of interest to economists) are assumed to be parametric" (Kilkenny and

Thisse, 1999). However, undoubtedly he showed the importance of the costs of transportation with respect to weight and volume in determining location.

2.3 Later Location Theory

Kilkenny and Thisse (1999) survey the economics of location discussing the literature's approach to how the problem changes as locations are restricted to a network, the input mix and location must be solved simultaneously, plant size and number must be considered, the problem is dynamic and inventory holdings must be considered, firms have local monopoly or monopsony power, there is uncertainty in factor prices and demand, small numbers of firms leads to strategic interaction, and when consumers and firms choose location, as well as commenting on central place theory. Incorporating these changes requires multiple solution stages, solving the optimal decision over each location and then choosing the best point among them. The general rules and outcomes vary across specifications, but two results consistent across many specifications are simultaneity—that the location problem and input mix must be solved simultaneously as transport alters relative factor prices, and the exclusion property—that the optimal site locations are at vertices of the network, ie market towns, resource towns, or cross roads. Some key takeaways from the fields include: the trade-off between fixed production costs and transportation costs is central to the spatial organization of economy, product differentiation fosters economic agglomeration by relaxing price competition, the choice of price policy (spatial discrimination vs mill pricing) influences the location of firms and welfare, transport infrastructure has a major impact on the spatial distribution of activity but it tends to spring up at nodes rather than along main lines at the micro and macro level and should strengthen the tendency for activity to agglomerate at existing centers of population (a point disputed by Dodson, 2021), and the location, production, and pricing decisions are often interdependent. They suggest future work in facility location analysis should integrate pricing and strategic competition into operational location models, perhaps by combining variational inequality techniques for solving oligopoly models with efficient algorithms for finding optimal location. A major difficulty is the potentially large number of calculations from considering the optimal decision at every location, which is why the exclusion property is so important.

Relaxing the assumption of price exogeneity is important because local market power is inherent to spatial differentiation. In reality, firms are not ubiquitous because fixed costs require economizing on the number of plants to serve each market. Monopolistic competition developed by Dixit and Stiglitz (1977), and used extensively in the new economic geography, is a useful framework but still suffers from an "inconsistency between the local labor supply limitation on the number of firms and the assumption that the number of firms is large enough to treat each firm as inframarginal with respect to prices and wages" (Kilkenny and Thisse, 1999). This leads to the authors suggestion to incorporate oligopoly models and game theory to understand the strategic interactions inherent in spatial settings.

Central place theory, developed by Christaller and furthered by Losch, concerns the size and distribution of central places and how they locate in relation to each other. The key idea is that regions form based on the marketable area for different goods. The range—the maximum distance consumers will travel to obtain a good, and the threshold—the minimum market area required to sustain production, determine the location of different production sites and generate a pattern of regional types. These regions overlap, with larger settlements distributed farther apart from each other than smaller settlements, distinguished by providing a wider variety of specialized goods and services that require larger market areas to sustain them. Higher order goods are assumed to be more durable, valuable, and variable.

Christaller (1933) assumes a uniform plane of constant population density and purchasing power with uniform linear transportation costs where consumers choose the nearest location offering their desired good to minimize cost. He assumes settlements would tend to form a hexagonal lattice as it is the most efficient pattern to serve all areas without any overlap, although the shapes would be distorted by geography and transportation networks. Depending on the sphere of influence of the central places, different hierarchical arrangements will result. Christaller identifies three arrangements based on different organizing principles: market, transport/traffic, and administrative.

Losch (1938) similarly works from the assumption of uniform population density on a plane, and derives the hexagonal distribution of producers from a monopolistic competition framework with entry, as it minimizes the overlap between market areas while deterring further entry to serve unsatisfied demand.

There are several criticisms against central place theory. For one, populations and resources are not uniformly distributed throughout space, and the theory does not address the observed tendency for clustering of population or feedback between firm and consumer location. Additionally, it does not address the economic benefits of agglomeration or provide any reason for industries to cluster in locations beyond distributional efficiency. There is little empirical support for the distribution patterns implied by the theory. Smith (1979) finds support of the central place theory predictions in the distribution of medical care centers and specialists, Haining (1980) estimates parameters from the Beckmann-McPherson model (an extension of the central place theory relating city sizes) and found significant estimates.

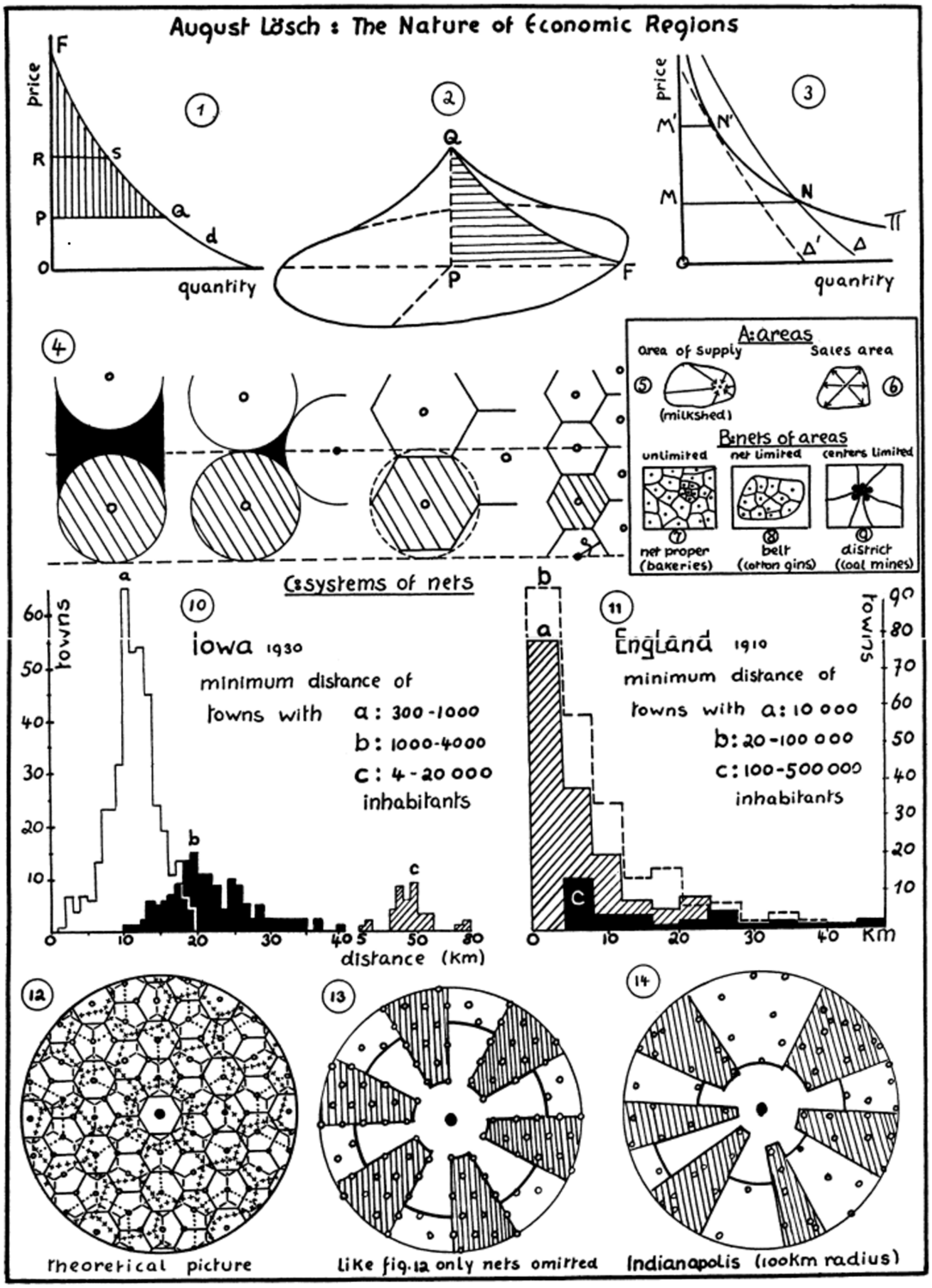


Figure 2.2: Visualization of Central Place Theory (Losch, 1938)

2.4 Urban Economics

The foundational work of urban economics emerged from two independent sources simultaneously, Alonso (1960) and Muth (1961). In their models of urban land and housing markets they examine the trade-off between access to a central point and competition via the price of land, laying the groundwork for future work on centripetal and centrifugal agglomerating forces.

Alonso (1960), in his stated to goal to form a "self-consistent explanatory theory ... which will shed light on some aspects of the internal structure of cities", is often credited with formalizing the bid-rent concept developed by von Thünen (1826) and expanding it to apply to modern central business districts. Under his assumptions of a plain with uniform transport costs, all land prices are known and taken as given, all employment is located at the center, and the population is fixed, he develops a model where households maximize utility in space stemming from a composite good, land, and a (dis)preference for commuting, subject to an exogenous income paying for the uniformly priced composite good, the land they use which varies in price with distance from the center, and costly commuting. One of his first conclusions from this setup is that if travel provides disutility then the price of land must decline with distance to the center. He then constructs residential bid price curves—the price consumers are willing to pay to live at each distance while maintaining a given level of utility. Crucially, the income effect of cheaper land counters the higher commute cost, as well as substituting the amount of land and composite good consumed. He concludes the paper with applications and suggestions for empirical research estimating the structural parameters of the model as well as extending the model to account for the effects of heterogenous incomes, urban transportation networks, and population growth on the observed heterogenous city shapes that depart from the circular city implied by his model. While the assumptions are highly restrictive, the model did not make the bid-price curves of various sectors consistent with each other, and there was no guaranteed clearing of the labor market, Alonso's work was an important step in developing microfounded economic models capable of generating something akin to observed spatial distributions of population and land prices.

Muth (1961), purportedly while locked indoors during a two-day snowstorm in DC, develops a model of two industries, housing services and agricultural commodity, utilizing labor and land as Cobb-Douglas inputs on a featureless plane where price falls exponentially from the point of production due to transportation costs. An advantage of Alonso (1961), is that Muth's model incorporates distributed labor involving commutes with wages varying across space. Relying on his assumption of the price distribution across space, he is able to show that the price of land and land use intensity decrease with distance from the center. Furthermore, he shows that an increase in the price of non-land inputs has an ambiguous effect depending upon the extent the price of each type of good depends on the relative demand elasticities. In his model, "For any pattern of residential location to be an equilibrium one, for each consumer at his optimal location the saving on housing costs from a small change in distance must exactly equal the change in transport costs" (Muth, 1961). Alongside Alonso (1960), Muth's was the first formal economic model in which population density endogenously declined exponentially with distance to the city center, and these were both an important part of the expansion of general equilibrium theory. See McDonald (2007) for an exposition of both of their models and a history of the related research.

While Alonso and Muth's models could endogenously explain spatial distributions around a city center, they could not explain why there would be a city center in the first place. Krugman (2011) summarizes some of the work of Masahisa Fujita as "escape from von Thünen", that is, creating a model that can account for the spontaneous formation of a city center as well as non-monocentric urban configurations. In 1982, Fujita and Ogawa do precisely this, in their model where positive spillovers between firms that fade with distance generate different clustering patterns depending on the parameters for firm externalities and commute cost. In a linear city with exogenous population, households consume land and an ubiquitous imported composite good using a wage in exchange for labor, and they must choose their residence and job site with costly transportation to maximize utility. The assumption that the composite good is imported avoids dealing with the product market clearing, and resembles a region exporting a specialized service while importing the basic commodities. Firms produce a good for export at a given price using a fixed amount of labor and land as input but receiving a positive externality from nearby firms. This externality, although ad hoc and opaque, generates the agglomerative benefit which the authors term location potential, while competition and the price of land acts as a dispersal force. The authors derive bid-rent equations showing the amount firms and consumers are willing to pay based on the parameters, and these determine how each parcel of land is allocated. Depending on the exogenous export price, the location potential parameter, and commute cost, the authors find several possible urban configurations including a monocentric city with an inner business district, an inner residential district, a completely mixed city as well as incompletely mixed city with mixed inner but residential outer, as well as a duocentric cities where firms cluster in two locations along the line with various configurations of business and residential mixes. They go on to show that tricentric cities are also possible, but do not extend the analysis far in this direction. The key insight is that both the location potential parameter and commute cost have medium values that encourage clustering, but that either force can send the city into a different configuration. As the population increases, it is less likely a monocentric city will form for any set of parameters, as the commute cost overrides the agglomeration benefit of a single cluster. Interestingly, the monocentric city generates the most land rent.

In 1988, Fujita addresses the previous shortcoming of an ad hoc and exogenously assumed benefit to agglomeration with a model of monopolistic competition where the external economies from the interaction of economies of scale with transport costs generate clustering behavior. In a similar linear city model where bid rent equations are derived for consumers and firms, but the key difference is that the positive externality from agglomeration stems from the preference for variety and desire to minimize transportation costs. Like Fujita and Ogawa (1982) several urban configurations are possible depending on the parameters for product differentiation and transportation cost. A higher degree of differentiation results in a higher agglomerative force, but like the effect of transportation costs whether this leads to a monocentric business district or multiple firm clusters depends on the interplay between the two.

Fujita's modeling approach is important because it facilitates a price for using land, endogenous agglomeration, and a framework where the equilibrium conditions can be analyzed for different parameter values to determine the spatial configuration. The spatial setting still must be reduced to a continuous line or discrete points in space, and even with that simplification the outcomes are complex potentially requiring numerical simulation to analyze the outcomes.

Urban economics today builds upon the foundations of these models, drawing on the use of preference for variety, differences in transportation costs, geographical differences in resources, and economies of scale to explain observed patterns within cities and regions. O'Sullivan (2003) suggests urban economics is divided into six related themes: market forces in the development of cities, land use within cities, urban transportation, urban problems and public policy, housing and public policy, and local government expenditures and taxes. Several fields overlap these topics, including urban economics, economic geography, the new economic geography, spatial economics, and regional science among others, and of course disputes about names and even some territorial behavior regarding ideas exist (Martin, 1999), but ultimately each of these fields benefit each other and expand the understanding of spatial relations in economic interactions. Storper (2010) suggests that while spatial economics has made progress in theorizing and measuring agglomeration effects the models still rely on questionable assumptions and do not sufficiently establish causality or account for spatial economic dynamics.

2.5 Cost Benefit Analysis

Assessing the benefits of roads to the economy is important for determining if the project should be supported with public funds. There are numerous direct and indirect effects of improved roads such as reducing travel times, expanding the range of suppliers and access to markets, facilitating specialization and trade, reducing required inventories and holdings, potentially improving safety, as well as providing direct consumption value. Benefits of roads for cost-benefit analysis are commonly counted from three sources 1) the value of time savings, 2) the reduction in vehicle operating costs (fuel), 3) the value generated by improvements in safety, and recently 4) the emissions reduction benefits (DOT, 2021). The benefit is the sum of the discounted future values stemming from these sources over the lifetime of the project, while the cost is based on the land acquisition, construction, projected maintenance, and relevant alternatives. Cost-benefit analysis studies distinguish economic benefits from economic impacts in part to avoid double counting and because of the uncertainty around what the final economic impacts will be (Appalachian, 2008; Forkenbrock and Weisbrod, 2001; Litman, 2009). The value of the road should theoretically include assessments of all outcomes that would not have happened without the road, but given the extent of influence and high degree of uncertainty there is not a universally agreed upon methodology. Recently, there has been push-back against cost-benefit analysis for its oversight on equity stemming from the focus on travel time savings over accessibility, the focus on vehicle travel demand which ignores many portions of the population, as well as the methodology and practice (Martens and Ciommo, 2017).

Many recent efforts to understand the economic impacts of roads date back to Auschauer (1989), who led the ongoing approach of viewing roads and other forms of public infrastructure as an input to the production function for the nation based on the macroeconomic models of the time. Using data on labor, capital inputs, and productivity from 1949-1989 he estimates the elasticity of national GDP with respect to the replacement value of the existing public capital stock, finding substantial positive values robust to assumptions of returns to scale and measures of utilization rates. He includes lagged values to address the uncertainty of response timing and finds these do not change the results. Sturm and de Haan (1995) later criticize that these findings are not robust to first differences, but such results are sensitive to the lag specification and do not rule out the hypothesized effect. Importantly, while these results suggest strong correlation they are not necessarily causal.

Auschauer's work initiated a heated empirical debate measuring the return to road investments (see footnote 2 in Chandra and Thompson (2000)), but convincing identification strategies were largely absent until Fernald (1999). He pointed out that if roads were truly productive, then industries utilizing roads more should see a larger increase in productivity, which he confirms by finding a disproportionate growth in vehicle intensive industries. Combining data on industry inputs, production, as well as a measure for congestion based on total miles driven per unit of road stock, he shows a convincing relationship between vehicle intensity and relative productivity performance. Interestingly, this approach was hinted at by Auschauer's (1989) isolating the trucking industry and finding a substantially higher return from the net stock of highways. Furthermore, Fernald repeats his estimates for multiple time frames, finding a much lower estimated return to roads after 1973 when much of the Interstate Highway had already been constructed. He interprets this as evidence that constructing a national highway system may act as a one-time boost to productivity, potentially explaining the observed productivity slowdown of the 80s and 90s, and that a second highway system should not be expected to yield similar results. While not explicitly accounting for what the roads are doing or how they lead to productivity changes, this is a step towards thinking about road system as a network that can be saturated.

Later strategies emerged that exploit features of the road systems as sources of identification. Baum-Snow (2007) examines the number of Interstate Highway rays emanating from US metropolitan statistical areas, arguing that this is related to how close they are to other population centers rather than the population of that area, to study the impact highways had on rates of suburbanization. Michaels (2008) exploits the incidental placement of US Interstate Highways in rural counties as well as the grid like nature of the system making highways more likely to be north, east, south, or west of major cities in order to test the effect of removing trade barriers on the demand for skilled labor. Leduc and Wilson (2012) construct a measure of government highway spending shocks capturing revisions in expectations about future government investment to examine the impact of federal highway funding shocks on local GDP. Pereira and Frutos (1999) focus on the time series nature of the data, getting around the issue of endogenous highways by focusing on the feedback between private production and public capital. Duranton and Turner (2012) construct instrumental variables using the original 1947 plan of the Interstate Highway System as well as a 1898 map of rail roads and maps of early explorations of the US in their examination of the impact on employment growth in cities.

These econometric developments aided in identifying the economic impact of roads, but the spatial distribution of the benefits from roads and how the roads generate these benefits were still aspects of the problem left unaddressed, partially due to the lack of regional data, but also inherent in using the value of the road stock without reference to where they actually are. Chandra and Thompson (2000) utilize the construction dates of Interstate Highway segments to compare the outcomes for counties that received a highway, counties adjacent to highway counties, and counties farther removed. They focus on non-metropolitan counties, arguing they received highways incidentally because of their position between metropolitan regions rather than their growth or economic makeup. Additionally, they test the assumption that having a highway is an exogenous event by regressing onto past growth, finding no statistical relationship and supporting their strategy of exploiting the variation in time and space of interstate construction as a source of identification. They address the issue of unknown lags and leads by using an 'age of highway' indicator with 5 years before and 24 years after the construction date. With county level earnings by industry and state and national effects, they find that generally the Interstate Highway raises the level of economy in counties it passes through, but reduces the activity of adjacent counties, ultimately having a net-zero effect on non-metropolitan regions. The results vary by industry, and consistent with their model suggest that the benefit to businesses in a highway improvement area depend on if they are relatively low cost, if they are located near the highway, and if they specialize in nationally traded goods or services.

These methods have focused on the impact of roads on the output of regions, but Haughwout (2002) argues that the dominant aggregate production and cost function approaches are limited, proposing instead a method based on spatial equilibrium accounting for the mediation of the infrastructure effects on firms and households through local prices. In a two-stage estimation, he estimates city and time effects on local wages and housing prices controlling for human capital and quality, and then examines if city infrastructure can account for the variance controlling for variance and tax or public debt conditions. He finds that infrastructure provision is positively associated with central city land prices and ambiguously with wages, suggesting a growing public capital stock is likely enjoyed primarily by households. Through calibrating his theoretical model of households and producers, he finds that although positive marginal benefits exist from public capital the aggregate city willingness to pay is less than their cost. His approach is limited to benefits that accrue to the central city.

The difficulties assessing the benefits of roads stem from their multiple avenues of potential impact. By reducing travel times roads generate savings based on the value of time, expanding the range of suppliers and access to markets facilitating specialization and trade, reducing required inventories and holdings, potentially improve safety, and providing direct consumption value. It is difficult to discover the marginal benefit of roads compared to what would have happened without them without an environment for controlled experiments, although identification strategies exploiting the details of road development are beneficial. The direct benefit and production function approaches have persisted in use for government estimates of the benefits of roads, while the literature seems to be shifting towards the market access approach using highly detailed spatial data as outlined in the next section.

2.6 New Economic Geography and Market Access

In a 2004 presentation Krugman outlines the key findings from his and other authors' work known as the new economic geography, utilizing simple stylized models designed for tractability rather than realism. The four main methodological tools used are the Dixit-Stiglitz framework of monopolistic competition, iceberg transportation costs where the value of the product 'melts' with distance traveled, evolution—the idea that history matters and multiple equilibria may exist depending on starting conditions, and the computer to generate numerical simulations as an "intuition pump" and gain insights from analytically intractable situations.

He suggests the new economic geography, in it's attempt to understand how economic interactions over space work, rely on four propositions. First, that transport costs shape international and interregional trade, that is, distance matters. He highlights the inability of gravity trade models, which imply the amount of trade between regions is proportional to the sizes of their economy, to explain why countries trade disproportionately with countries near them. Second, that the interaction of market size with increasing returns is important for determining location. This interaction, known as the "home-market effect", is generated by increasing returns incentivizing the concentration of production and costly transportation incentivizing proximity to larger markets, and implies that regions should tend to export goods subject to increasing returns with large local demand. Third, the cumulative process where large markets attract production generates a positive feedback loop leading to agglomeration and possibly multiple equilibria based on initial conditions. This proposition is difficult to support empirically, but bears out in many models. Lastly, that the same processes that shape economic geography within countries also shape international trade. However, borders are still significant, potentially representing between 1500-2500 miles. He goes on to discuss the persistence of distance despite technological developments in transportation and telecommunication, and the possibility that declining transportation costs have a U-shaped impact on agglomeration, first making it possible, but eventually making it unnecessary.

The new economic geography was initiated by Krugman's 1991 paper, where he developed a model of trade between two regions capable of endogenously generating a core-periphery pattern depending on the transportation cost, degree of economies of scale, and the share of manufacturing in national income. In a Dixit-Stiglitz monopolistic competition framework with increasing returns and preference for variety, workers/consumers are initially distributed between two regions, and work in either agriculture or manufacturing, at the population level proportional to the taste for each good. Manufacturing can take place at either region but transportation between the two is costly, and manufacturing workers will migrate to whichever location offers the higher real wage. Under a certain set of parameters, because of the increasing returns encouraging production in one location and the desire to minimize transportation costs by locating closer to the larger market, an initial difference in worker allocations will exacerbate the real wage difference and lead to a core-periphery pattern where all manufacturing is done in one region. Whether the region with the larger population will also have a larger real wage depends on the size of the home-market effect, as the degree of increasing returns to scale (in this model the degree of substitutability) and lower competition for wages in the periphery work against each other. The lower price level of manufactured goods in the core acts as an additional force for divergence, but Krugman shows the divergence depends on the relative values of transportation cost, substitutability, and size of the manufacturing sector.

This was an important development, for while many of the ideas of geographic dispersion were developed earlier, they had not been formalized in an economic model that could endogenously generate realistic location patterns without relying on realistic but difficult to measure benefits to production from agglomeration such as knowledge spillovers (see Fujita and Krugman, 2004 for a discussion on this). The work sparked a fruitful field of research and many extensions, including Krugman, Fujita, and Venables' (1999) three region model where a number of core-periphery patterns are possible depending on the parameters, and their ring model of continuous space where manufacturing cores emerge with different frequencies throughout the ring depending on the parameters. Extensions of the new economic geography model include Helpman (1998) and Tabuchi (1998) incorporating limited land availability and rent that increases with agglomeration acting as a dispersal force when transport costs fall below a threshold, and Venables (1996) incorporating vertical industry links finding a similar medium range of transport costs where agglomeration occurs.

A highly utilized concept formalized by the new economic geography is 'market access' or 'market potential'. The term, capturing the idea that an area is more marketable if it is close to other large markets because of the potential to trade with them, was actually created by

Harris (1954), who defined the market potential of an area as the sum of the size of other markets weighted by the distance between them. This was merely an abstract proposal, and Krugman (1992) was the first to capture the concept formally using the Dixit-Stiglitz framework. In his model, the market potential is reflected in the positive impact of purchasing power of other locations on the region's wage inversely weighted by the distance between them. Krugman's index also reflects the effect of competition from producers in other locations through the price index, which is missing from Harris' conception. See Figure 1 for a collection of the equations representing market access across authors. Similarly, Redding and Venables (2004) and Hanson (2005) derive market access terms from the monopolistic competition framework and utilize spatial variation in earning to identify their structural parameters, but Redding and Venables examine cross-country data and trade flows while Hanson examines cross-county data for a single country focusing on the spatial covariation in wages and purchasing power. Head and Mayer (2011) extend this work examining panel data on trade and income per capita for countries around the world from 1965-2003, finding a strong correlation between the market potential index and income per capita.

The market access concept has persisted and emerged from other economic frameworks as well. Anderson (1979) develops a gravity model based on constant elasticity of substitution preferences and regionally differentiated goods. Hummels (1995) examines the correlations between residuals from an augmented Solow-Swan neoclassical growth model and measures of geographical location. Eaton and Kortum (2002) develop a Ricardian trade model where the key difference between regions is production technology, which generates comparative advantage stimulating trade stifled by transportation costs and geography. They calibrate the model to explain observed trade flows between countries given their geographic barriers (distance). Duranton, Morrow and Turner (2014) extend the gravity model framework of Anderson (1979) to multiple industries findign that cities with more highways specialize in sectors producing heavy goods. Allen and Arkolakis (2014) extend the gravity model framework of Anderson (1979) to arbitrary geographies and labor mobility, providing conditions for the existence, uniqueness, and stability of spatial economic equilibriums as well as equations governing the relationship between economic activity and geography. In their work, geography is defined as a set of bilateral trade costs based on transportation between predefined regions. Donaldson and Hornbeck (2016) follow Eaton and Kortum's framework but modify the distance to account for the actual railroad network of the late 19th century U.S.. Similarly, Alder (2017) uses their framework but incorporates the distance between economic centers based on the highway network.

Market access has become a widely used framework for understanding how spatial relations interact with economic activity. Increasingly, this concept is combined with detailed spatial data on transportation infrastructure networks to accurately account for how regions are positioned in relation to each other. Fowler (2011) questions the necessity of general equilibrium in understanding the results of geographical economics, and particularly criticizes the core-periphery model's assumption of simultaneous decision making by firms and workers in response to changing conditions and the equilibrium constraint "that defines the number of firms in a city (or region) such that each firm produces its optimal output based on the labor present in that city...firms appear and disappear in cities based on the full employment of the workforce" which does not leave room for unemployment or pressure to move to places where jobs are available. Fowler (2011) proposes the use of an agent-based-model where firms and workers respond to shifting information, showing that economic systems based on the same foundation as the new economic geography, involving economies of scale and preferences for variety with transportation costs to establish patterns of agglomeration and dispersal, will often but not always lead to stable equilibria without the imposition of assumed conditions of general equilibrium.

2.7 Graph Theory and Roads

To study the interactions of roads and the economy it is necessary to represent the road system in a way that can be translated into relevant data. One of the increasingly common ways to do this is by thinking of the road as a network of nodes and edges. Depending on the problem under study there are several ways to do this. Major destinations, such as cities, town centers, or any economically significant point, can be represented as nodes and roads the edges between them. The edges can either be a direct representation of the roads using a set of nodes for points along the road, such as junctions, bends, or curves, connected by straight road segments, or the edges between points of interest can be reduced to the distance or travel time along the roads. The edges can be weighted by either distance, speed, topography, or travel time, and the nodes can be weighted by a single statistic or even a set of relevant data. Multiple lanes of one-way travel can be dealt with as an oriented multi-graph, and subgraphs and multi-line edges can be constructed for higher detail, although at the cost of larger data storage and longer computation time (Ting, Li, and Gong, 2000). Similarly, intersections with specific rules for directions of travel can be represented by sub-graphs with entrance points representing lanes (Ting, Li, and Gong, 2000). The level of detail should correspond to the scale of the question, for instance if one is studying traffic within an urban center or inter-city relationships.

By casting the roads as a system of nodes and edges techniques from graph theory can be used to study its characteristics, the impact of characteristics on economic outcomes, and to understand the relationships between points in space as well as the impact of changes in the network. For instance, the Dijkstra algorithm is a method for efficiently calculating the shortest path between nodes and can account for weighted edges. Network measures such as centrality, betweenness, diameter, density, cyclomatic number, alpha, beta, gamma, and pi reveal insights about the network (see Table 2.1 below)[2]. Xie and Levinson (2005) propose a measure of entropy, connection pattersn and continuity, showing "that the differentiated structures of road networks can be evaluated by the measure of entropy; predefined connection patterns of arterial roads can be identified and quantified by the measures of ringness, webness, beltness, circuitness, and treeness", as well as review measures useful for urban and transportation planning such as heterogeneity, connectivity, accessibility, and interconnectivity. Spanning trees, subgraphs connecting points of interest, can be used to determine the functional relevance of road segments (Thomson and Richardson, 1995).

Studies have utilized network representations of roads to examine many important topics. Singh et al (2018) develop a framework to examine the flood vulnerability of urban road networks by linking meteorological information, land use functions, a hydrodynamic model with speed and trip duration data. Sarkar et al (2020) examine the relationship between growth and development with connectivity and network accessibility of villages in the English Bazar municipality in India using a connectivity index and average shortest path length. Appert and Chapelon (2007) measure urban road network vulnerability of Montpellier to traffic blockages and congestion spillovers. Utilizing graph theory to represent roads offers significant opportunities for understanding how roads influence economic outcomes through accessibility and other network characteristics, as well as shedding light on important traffic channels for targeted expansion and maintenance.

[2] http://webspace.ship.edu/pgmarr/TransMeth/Lec%201-Network%20Measurements.pdf

Measure	**Equation**	**Use**
closeness	C(x) = $\frac{1}{\sum_y d(y,x)}$	yields how close node x is to other all other nodes. $d(x,y)$ is the shortest path distance
betweenness centrality	g(x) = $\sum_y \frac{\sigma_{yz}(x)}{\sigma_{yz}}$	σ_{yz} is the total number of shortest paths from node y to z and $\sigma_{yz}(x)$ is the number of shortest paths that pass through x. yields how often x is in the shortest path between nodes and how often it will be used
diameter	D = $\max_{xy} d(x,y)$	yields the maximum distance across the network, useful for other measures
density	$\frac{L}{S}$	L is total km of links, S is area of smallest bounding rectangle, yields how saturated the network is
cyclomatic number	e-n	useful measure of route redundancy and network resiliency. essentially the number of circuits
alpha	e - $\frac{n-1}{.5n(n-1)-(n-1)}$	ratio of existing circuits (path which ends at the node it began) to maximum number of circuits
beta	$\frac{e}{v}$	measure of connectivity, average number of edges per node
gamma	$\frac{e}{.5n(n-1)}$	measure of connectivity, percent existing routes to potential routes
pi	$\frac{L}{D}$	network distance per unit of diameter, indicator of the network shape

Table 2.1: Network Measures

e is the number of edges, n is the number of nodes

Chapter 3

Roads as Economic Environments

3.1 Motivation

The goal of the model is to represent an economy distributed over a two-dimensional space where firms enter and exit based on profitability in order to analyze patterns of clustering for different industries and how these patterns change after altering the road system. The setup is similar to the new economic geography models pioneered by Krugman (1991) and Fujita et al (1999) except for the explicit representation of space. In their models the simplifying assumptions avoid the issues presented by space, effectively dealing with a fixed number of condensed 'locations' operating as open economies in perfect competition trading with each other. With their setup analytical results can be derived for two-region models and even a ring-model with an arbitrary number of locations, however, even they approach setups where numerical simulations must be utilized to ascertain results. When extending to this two-dimensional spatial representation, the primary analytical difficulties are the market and monopsony power that stem from occupying a location in space, and that all locations face different sets of prices according to the distances between each agent and all of the other agents. The discrete space model is analytically untractable, so I instead develop a simulation to iterate with a computer. The downsides to this approach are a loss of closed form solution, potentially no equilibrium, sensitivity to parameters and starting conditions, and difficulty in interpretability. However, there is increased flexibility as any functional forms can be used, and arguably an additional degree of reality is gained by the explicit representation of space, both of which are important for the topic addressed here. Related research include Olner (2014), Epstein and Axtell (1996), Sasaki and Box (2003), and Olner, Evans, and Heppenstall (2015). This model is situated closely to three literatures: production networks, spatial equilibrium, and agent based modeling. While it explores themes related to the second, in mathematical form it is closer to the first, but in spirit it is closest to the third.

Production networks is about the interrelations between firms and how they shape outcomes and responses of the system to shocks. Carvalho and Tahbaz-Salehi (2019) provide a primer on production networks, focusing on the role of input-output linkages as a shock propagation channel and mechanism for transforming microeconomic shocks into macroeconomic fluctuations. The multisector general equilbrium models they discuss, developed by Long and Ploser (1983) and extended by Acemoglu et al (2012), are similar to the monopolistic competition framework with multiple industries and nested CES structure of composite goods used here. While many of these models operate under perfect competition within industries generating only downstream shock propagation (in part due to the assumed form of production) from changing input prices and possible input substitutions, papers such as Jones (2013), Bigio and La'O (2017) and Liu (2018) incorporate some form of exogenous distortions, wedges, or markups which still only produce downstream shock propagation. Papers such as Grassi (2017), who considers a production network with oligopolistic market structures, and Baqaee (2018), who endogenizes the mass of firms active in each industry with imperfect competition and external economies of scale due to firm entry and exit, show that frictions and market imperfections can generate additional upstream channels of shock propagation. This paper is situated closely to this later branch of the production network literature as the market power from occupying space acts as an endogenous source of friction capable of generating upstream shock propagations as the price and wage are not solely determined by marginal cost, although that is not the focus of this paper. Additionally, this model can explore how different production networks generate spatial relationships between industries and how systems respond to transportation shocks.

Spatial equilibrium models use the framework of general equilibrium to determine the locations of activities in relation to each other, differences between cities and regions, as well as the phenomenon of agglomeration. This literature is close to urban economics and the new economic geography in the frameworks utilized, location forces explored, and the assumptions and solution methods. Berliant and Wang (2019) suggest models typically fall under conventional Arrow-Debreu competitive equilibrium models, monopolistic competition models, and game theoretic models including search and matching setups. Glaeser and Gottlieb (2009) present a version of standard spatial equilibrium model utilizing a Solow style Cobb-Douglas aggregate production function for cities differentiated by consumer amenities, housing supplies, or productivity advantages that can generate urban concentration. The types of agglomeration forces considered vary, but as discussed in the 2nd chapter generally fall under the categories of trade, production, or knowledge transmission externalities.

Glaeser and Gottlieb (2009) also discuss various strands of the literature, where based on the data available unobserved exogenous differences in space attributed to productivity, amenities, and the construction sector drive differences in density, income, and home prices across space. Gollin et al (2017) search for a spatial equilibrium in the developing world, suggesting the higher consumption levels of urban areas are offset by lower non-monetary amenities showing how health, public goods, crime and pollution vary across space. Similarly, Ahlfeldt et al (2019) extend the standard spatial equilibrium framework to show how costly migration leads to spatial arbitrage, allowing them to solve for rather than assume a long-run spatial equilibrium. Spatial equilibrium models rely on the assumption that utility is equalized across space by migration, with prices and amenities accounting for differences in nominal wages. Glaeser and Gottlieb (2009) suggest "the urban emphasis on mobility implies that local poverty is more likely to reflect something good that an area is providing for the poor than a failure in local labor markets. Poor people are attracted to big cities because they offer access to public transportation and inexpensive rental housing", however a walk through different neighborhoods in an urban center like DC suggests these models are missing something. Partridge et al (2012) find that in the US people are not as mobile as many models suggest. The model developed here is situated closer to the monopolistic competition approach of the new economic geography. It does not assume any ad hoc benefit of agglomeration and the current version does not deal with migration or costly land use.

Section 2 sets up the model, section 3 discusses the results, and section 4 concludes.

3.2 Model Setup

The model is composed of households, which are fixed in space, and firms who enter and exit based on profitability. Firms are monopolistically competitive as in Dixit-Stiglitz (1977), producing distinct varieties, q_{ij}, of different types of goods that form composites for consumption as well as production. Different industries corresponding to types of goods are denoted with j, and unique firms i. σ is the degree of substitution within a type of good and is identitical for all industries. For each composite good there is a price index representing the cost of increasing the composite by one.

$$C_j = (\sum_i q_{ij}^{(\sigma-1)/\sigma})^{\sigma/(\sigma-1)}$$

$$P_j = (\sum_i p_{ij}^{(1-\sigma)})^{1/(1-\sigma)}$$

Households are immobile and maximize utility from consuming a variety of composite goods subject to the price index they face and the wage they receive for inelastically supplying their labor. In a two-stage optimization, consumers first choose how much of each composite to consume proportional to the price index and elasticity of each type of good.

$$\max U = \prod_j C_j^{\mu_j}$$

$$\text{s.t.} \sum_j P_j C_j = w$$

$$C_j = \frac{w\mu_j}{P_j}$$

Second, given the quantity for each composite good the cost minimizing bundle is chosen, yielding the demand for each firm's good as a function of their price and the price index.[1]

$$\min \sum_i p_{ij} q_{ij}$$

$$\text{s.t.} (\sum_i q_{ij}^{(\sigma-1)/\sigma})^{\sigma/(\sigma-1)} = C_j$$

$$q_{ij} = \frac{w\mu_j p_{ij}^{-\sigma}}{P_j^{1-\sigma}}$$

The consumer faces a unique set of prices and prices indexes based on their location. I deviate from the standard ice-berg assumption of transport costs so that the transport cost is not proportional to the value of the product, and simply depends on the distance and an industry specific term to represent differences in shipping cost. The difference between the firm's price and the price consumers pay is[2]

$$p_{ijc} = p_{ijf} + e^{\tau_j d_{fc}} - 1$$

Similarly, the wage each consumer receives from the firm depends on their distance from the firm, representing the cost of commuting. Each consumer inelastically supplies labor to whatever firm offers the best wage minus commute cost. If the effective wage falls below a threshold, the consumer works for themselves earning the threshold wage. This is necessary to generate a flow of money into the system, otherwise the firms would not be able to sustain

[1] Assuming $\sum_j \mu_j = 1$. For simplicity this is kept true and μ_j is identical across industries.

[2] the value of shipping is essentially lost in this framework, it is not earned as income

themselves from the wages they pay given the loss of value to transport.

$$w_c = w_f - e^{\tau_w d_{fc}} + 1$$

The space is a discrete two-dimensional grid (100x100 in the simulation) where each position is connected to the eight positions surrounding it with a distance of 1, unless the area is designated as road in which case the distance is lower through that position[3]. By constructing creating a network representation of the space where every point is a node connected to the surrounding nodes with a distance-weighted edge, the shortest path distance between every pair of locations, x, y, is calculated with the Djikstra algorithm and stored in a distance matrix with elements d_{xy}. This matrix can then be efficiently referred to when calculating the prices, wages, and price indices faced by each location as well as calculating features describing the space.

The firms take the prices and wages of all other firms as given and choose their own price and wage to maximize profit utilizing labor and intermediate composite goods produced by other firms for production. Similar to consumers, they first solve for how much of each composite they demand and then the cost minimizing bundle to determine their demand for every other firm's product.

$$\max \pi_{ij} = \sum_a p_{ij} q_{ija} - w_{ij} L_{ij} - \sum_k P_k C_{ijk}$$
$$\text{s.t.} \sum_a q_{ija} = Q_{ij} = A L_{ij}^{\gamma} \prod_k C_{ijk}^{\beta_{jk}}$$

The index a represents the quantity demanded by consumers and other firms, and k represents the other composite goods used by firm i in industry j. The output elasticities from each type of composite good can vary for each industry j, and later represent differences in input-output structure between the firms. A is general productivity and L is labor.

It is necessary for the firms to choose both price and wage because of the spatial market power every location holds. This creates an oligopoly type framework, where prices and wages should be chosen strategically, but with a large number of firms a Nash equilibrium becomes unfeasible to calculate. For simplicity, each firm naively chooses their price and wage from a range[4] by checking the market outcome given all of the prices and wages from

[3]this allows the transportation cost to vary with the distance, type of surface, and industry, while keeping the distance calculation simple for ease of computation

[4]The range is chosen based on half of the minimum and twice the maximum price/wage among all firms. Within that range ten values are checked in even intervals to reduce computation time.

the last period as if they were the only ones changing. Then, given the new prices and wages, firms and households make their decisions about how much to purchase, who to work for, and how much to produce[5]. If there is no profitable combination of price and wage the firm leaves the market. Under this framework it is possible no equilibriums or even multiple equilibriums occur from price and wage cycles of responses, but generally the prices and wages resulting from a set of firms is stable. Regardless, what we are interested in is the location distribution patterns that emerge, so small scale price-wage cycling is acceptable.

Because the space is discrete, the labor supply response to a change in the wage is a step function unique to every situation, so a closed form solution is not possible. Intuitively, as firms offer a higher wage the radius of workers they capture increases, leading to discrete jumps in labor. Given this, the solution is found as follows. First, the firms consider a set of prices which, given the distance matrix and prices of all other firms, maps to a total quantity demanded of their product, Q_{ij}. Second, the firms consider a set of wages which, given the distance matrix and wages of all other firms, maps to a set of laborers the firm would have. Third, for each combination of quantity demanded and labor supplied, the cost minimizing set of composite goods for use as intermediates is found.

$$\min w_{ij} L_{ij} + \sum_k P_k C_{ijk}$$

$$\text{s.t.} \sum_a q_{ija} = Q_{ij} = A L_{ij}^{\gamma} \prod_k C_{ijk}^{\beta_{jk}}$$

$$C_{ijm} = \left[\frac{Q_{ij}}{A L_{ij}^{\gamma} \prod_k (\frac{\beta_{jk}}{P_k})^{\beta_{jk}}} \right]^{\dfrac{1}{\sum_k \beta_{jk}}} \frac{\beta_{jm}}{P_m}$$

This results in a matrix of wage (labor) and price (quantity demanded) pairs where each pair dictates a cost minimizing bundle of composite goods, and therefore a total cost. This is combined with the total revenue from each price and therefore quantity demanded to form a profit matrix, and the highest element yields the optimal wage and price combination given the price and wage of every other firm. In many monopolistic competition frameworks the effect of an individual firm on the price index is negligible if there are a large number of firms, but in this case because every location has its own local price index and there are a

[5]If more workers show up to work for the firm than they planned they will hire all of them. If demand from a firm exceeds there supply then the demand is unmet

relatively small number of firms this effect must be considered.[6]

$$w = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \rightarrow L = \begin{bmatrix} L_1 \\ L_2 \\ L_3 \end{bmatrix}$$

$$p = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} \rightarrow Q = \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \end{bmatrix} \rightarrow TR = \begin{bmatrix} TR_1 & TR_2 & TR_3 \end{bmatrix}$$

$$\text{L , Q } \rightarrow L \underset{Q}{\begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix}} \rightarrow TC = \begin{bmatrix} TC_{11} & TC_{12} & TC_{13} \\ TC_{21} & TC_{22} & TC_{23} \\ TC_{31} & TC_{32} & TC_{33} \end{bmatrix}$$

$$\text{TR , TC } \rightarrow \pi = \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{bmatrix}$$

[6]This is the primary computationally expensive part of the simulation, because whenever a price change is considered the effect on the price index for several locations must be calculated.

The final component to the model is how firms choose to enter. It is too computationally expensive to consider every location, so instead each round the profitability of locations are checked randomly and if a profit greater than a fixed cost of entry is possible than a firm will enter there. This approach means that firms will not find every profitable point, but after enough iterations they will reasonably saturate the space such that further entry will drive out another firm because of the competition. The fixed cost of entry prevents the profit from being driven to zero, and profits from the firms are divided equally between all households as dividends as if there was common ownership of all firms. For every round of entry there are two rounds of competition where new prices and wages are chosen, which is generally enough to reach a new near-equilibrium for a set of firms and after entry.[7]

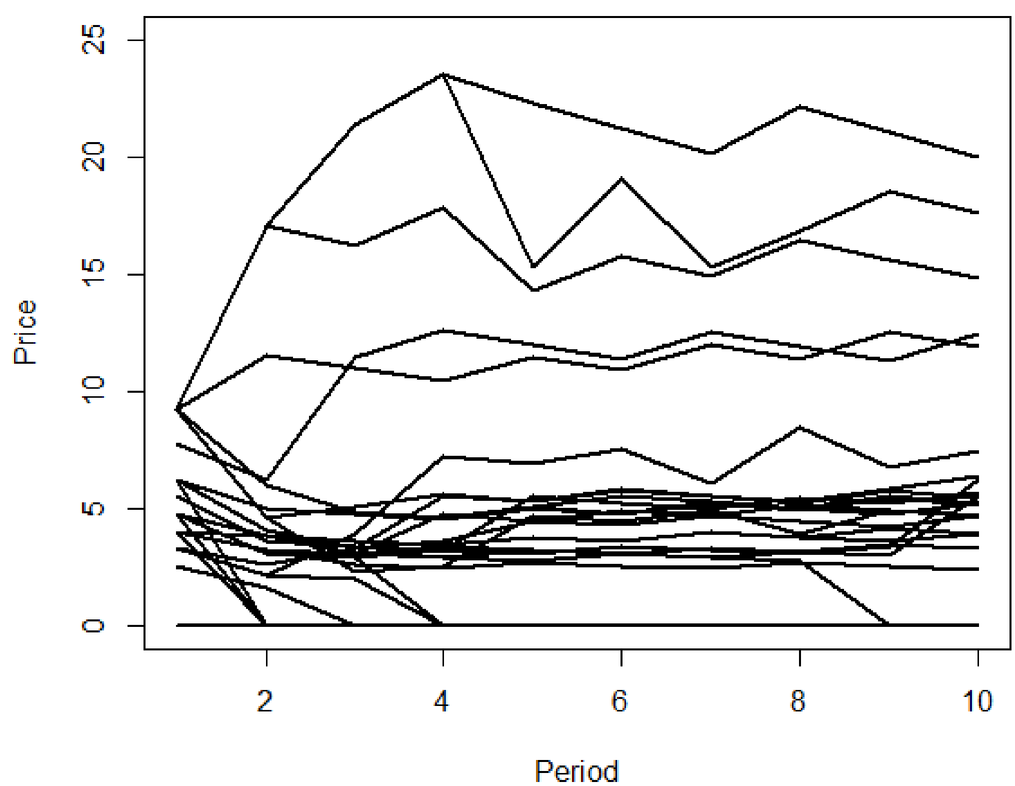


Figure 3.1: Example of Price Stabilization

[7]The figure below is from a simulation with 100 initial firms, input requirement variation, and no entry. Each line is a different firm's price over time

3.2.1 Definition of Clustering

There is not a consensus in the literature on a definition of clustering, particularly when the definition of space varies. To measure the change in agglomeration of different industries, similar to the first chapter I construct a spatial GINI based on the inequality of earnings between counties. To construct counties, I use a k-nearest neighbors algorithm to identify fifteen clusters based on the original household distribution based on euclidean distance, shown in the image below. This household distribution is then used for all of the simulations, and the counties used for the spatial Gini are based on resulting county borders.

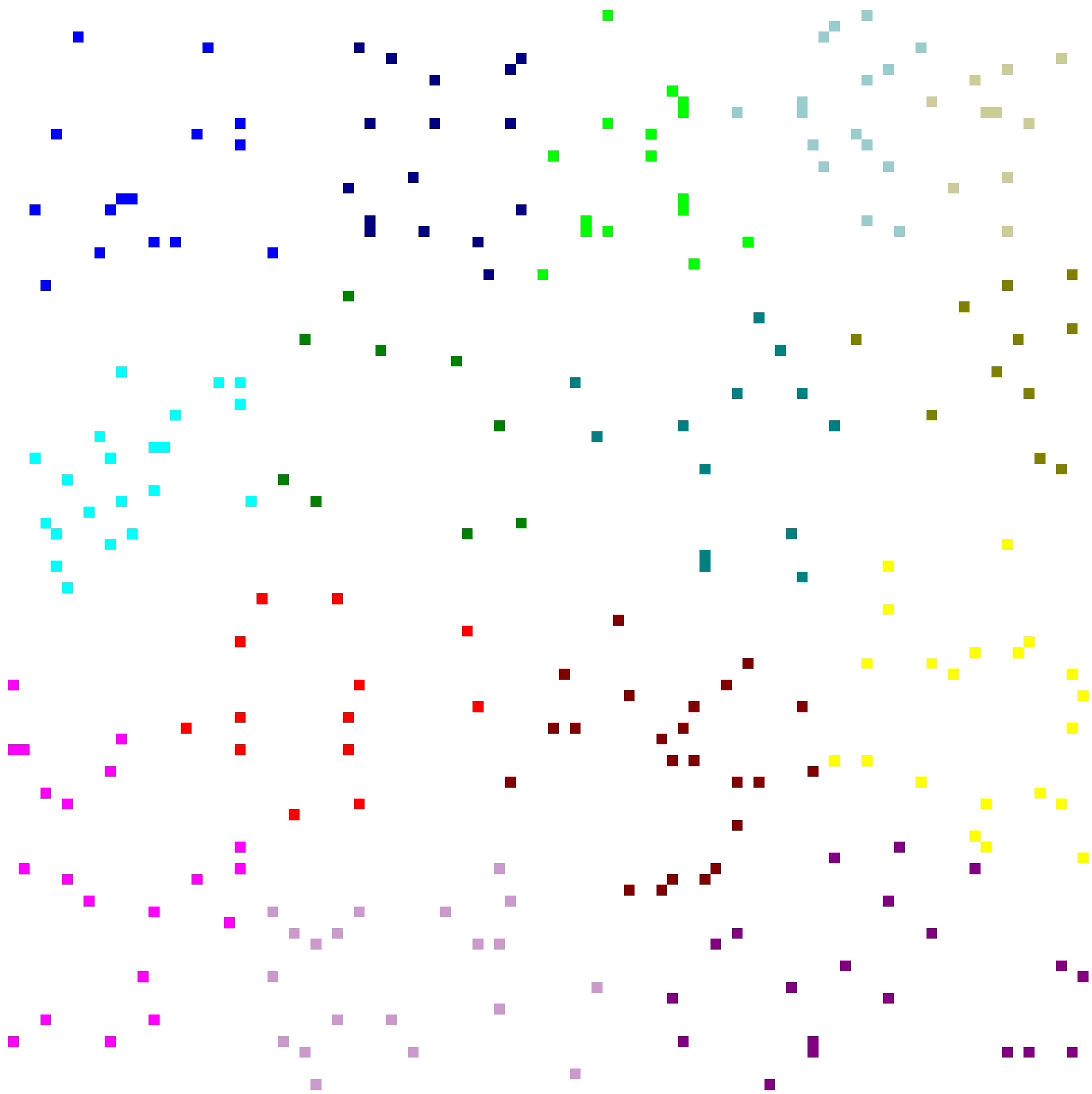


Figure 3.2: The Fifteen Counties Based on Household Clustering

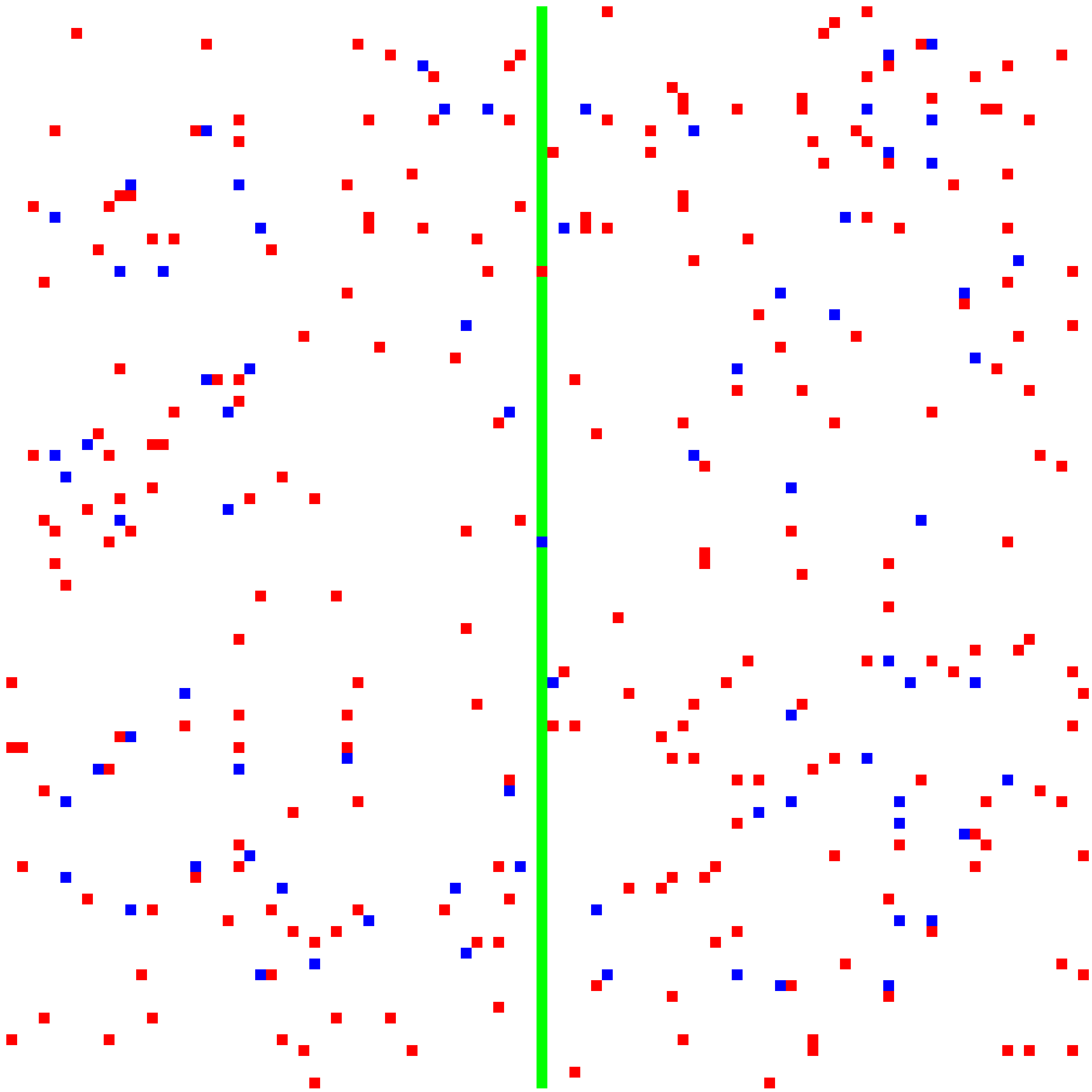


Figure 3.3: An Example Run of the Simulation. The red dots are households, the blue dots are firms, and the green line is the road.

3.3 Results

An appeal of this model is that any road structure can be utilized, and changes in the cost of movement along the road can be experimented with to see how firms with different characteristics respond. As in the first chapter, the spatial gini is calculated before and after an improvement in the road, hereby referred to as the improved road ($d_{xy} = .1$ compared to $d_{xy} = .2$ in the base case) to measure the varying response in clustering for two sets of firm characteristics. Then, the profitability of all locations is examined and regressed against several state-based features when industries vary in their input requirements. For all simulations the following sets of parameters are used: $\tau_w = .2$, $\gamma = .6$, $A = 6$, $w_{outside} = 3$, fixed cost for entry = 5, $d_{xy} = 1$ for non-road space, width = 100, height = 100, number of households = 300, and the road structure is as in Figure 3.3.

In the first experiment four industries vary in their input requirements, that is, the coefficients β_{jk}. The first industry does not require any inputs from other firms, akin to a raw material processor or agricultural commodity that is useful for other industries. The other industries vary in how many different industry's inputs they require, and thus the number of other industry locations they must consider for price comparison, with the sum of the coefficients equal to one in order to preserve the total amount of inputs required for a given output. The coefficient matrix is shown in Figure 3.4 below. The results below are from two simulation with thirty rounds of entry where two firms of each industry enter each round with three rounds of price and wage competition in between each. In the first iteration, $d_{xy} = .2$ for movement along the road and in the second iteration $d_{xy} = .1$, which results in substantially cheaper movement along the road. The industry requiring no inputs has a lower spatial gini for both, meaning it is more evenly spread out between the county clusters. This industry does not benefit from the lower prices from reduced transportation costs in agglomerations, and because it only requires labor from households firms in this industry can survive in more industry independent locations. The average value of the gini after ten periods, when the location shocks from firms entering has lessened, is .13 for the base road, and .29 for the improved road. The other three industries have a higher average gini after the initial ten periods, but the change in gini after the road improvement does not reveal a consistent pattern, the second industry moves from .40 to .36, the third industry moves from .35 to .43, and the fourth industry moves from .34 to .39. Intuitively, one might expect the industries requiring inputs from more industries to cluster in central positions to reduce the amount of shipping, but the underlying process is noisy and highly dependent on history as well as the method of firm entry.

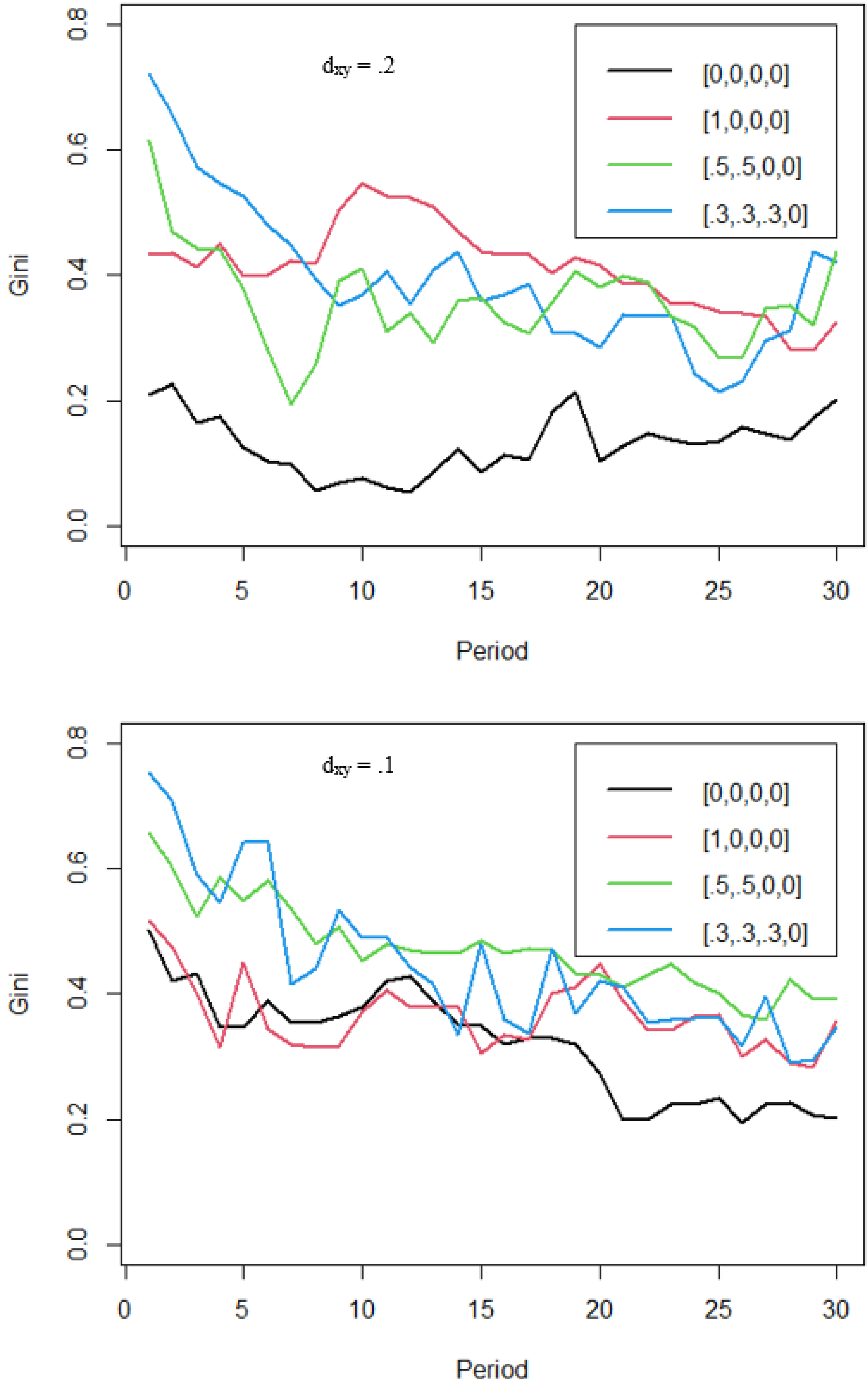


Figure 3.4: Spatial Gini for Industries with Varying Input Requirements

In the second experiment four industries vary in their transportation costs, that is, the coefficients τ_j. The first industry has a low rate of change in transportation cost as distance increases (τ_j=.1), and in even increments the rate of change with distance increases with each industry, with the fourth industry having the highest rate of change (τ_j=.4). The values are shown in the Figure 3.5 below. The results below are from two simulation with thirty rounds of entry where two firms of each industry enter each round with three rounds of price and wage competition in between each. In the first iteration, $d_{xy} = .2$ for movement along the road and in the second iteration $d_{xy} = .1$, which results in substantially cheaper movement along the road. A clear pattern does not emerge from differences in sensitivity to transportation costs. The average value of the gini for the low transport cost industry after ten periods, when the location shocks from firms entering has lessened, is .19 for the base road, and .19 for the improved road. The other three industries have a slightly lower average gini after the initial ten periods, but the change in gini after the road improvement does not reveal a consistent pattern, the second industry moves from .15 to .18, the third industry moves from .17 to .15, and the fourth industry moves from .19 to .18. Intuitively, one might expect the industries with a higher sensitivity to transportation costs to disperse more after an improvement in the road system since they are relatively less dependent on the road.

Overall, both experiments of the effect of the road improvement on the clustering behavior of differing industry characteristics suffer from the scale of agglomeration considered, a dependence on history, and a sensitivity to parameters and modeling assumptions. Agglomeration happens at multiple scales, and the definition of clustering based on differences in county production doesn't reveal the proximity of firms to one another or spatial relations between industries. Second, in this model where firms locate and where they are able to survive at any given moment is highly dependent on the positions of other firms, so many paths of spatial relations are possible from a given situation that could result in different stable states. Third, this model involves many parameters and assumptions that can significantly change a given outcome or the response to changes in other parameters. What is gained in flexibility comes at the cost of sensitivity, and tuning the model is an art requiring many iterations, experience, and judgement. In this model the production function parameters for the return to labor and general technology heavily influence the distribution of firms and the range of employment and profitability, the assumptions of random entry checks for profitability potentially lead to different arrangements than if the most profitable point was always selected until none remain, the mode of competition results in sub-optimal responses and potentially interferes with what could otherwise be optimal locations.

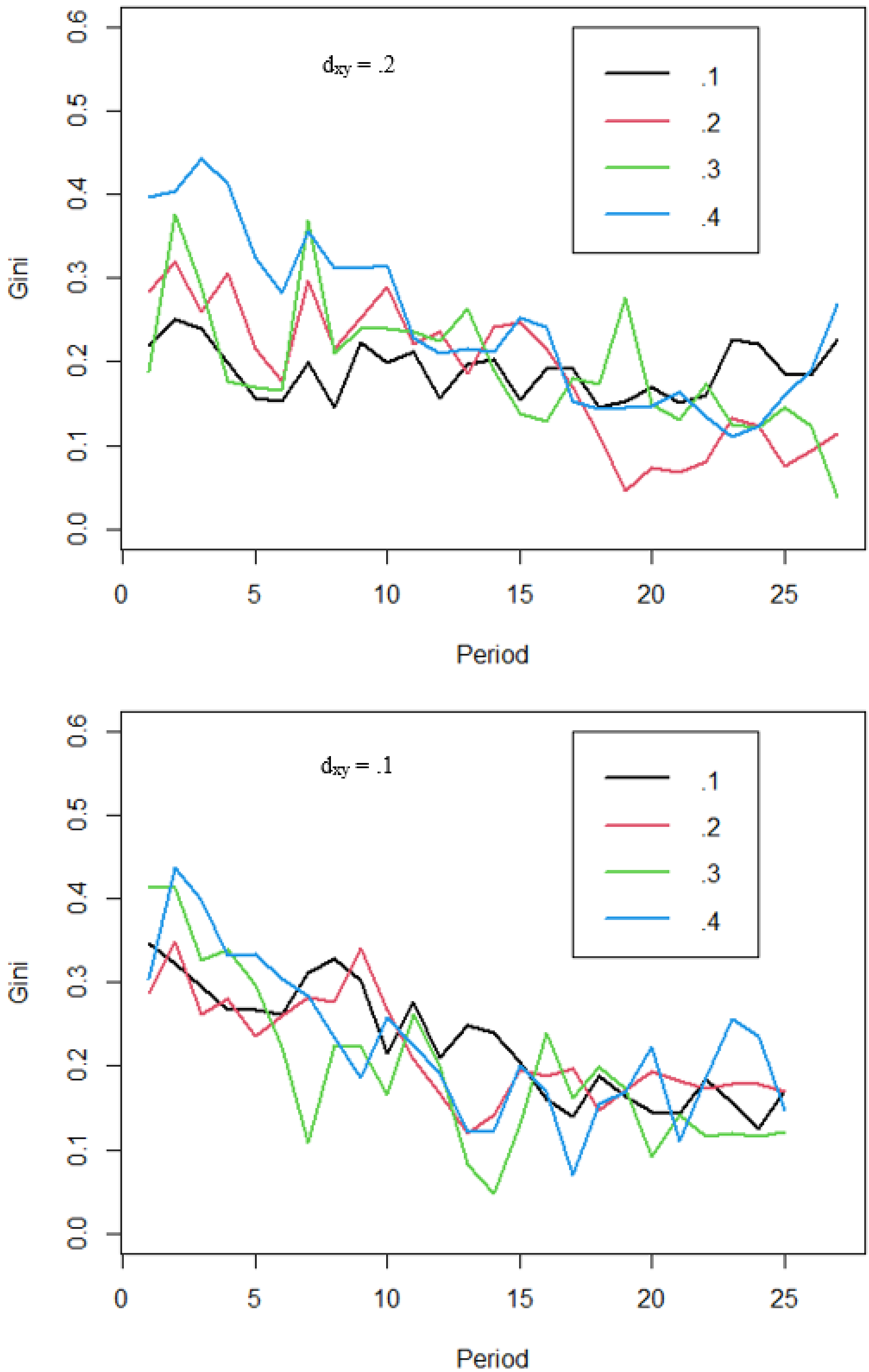


Figure 3.5: Spatial Gini for Industries with Varying Transportation Costs

Figure 3.6 below shows the profitability of locations sampled in the space before the market has been saturated by firms entering but after the prices and wages have stabilized for the firms randomly placed in the beginning. Clearly being near consumers increases profitability, and the road appears to have a positive effect. Figure 3.7 shows the results of a regression on the profitability of the sampled locations for industries 1 and 4 from the variation in input linkages on several features of the space. Industry 1 requires no inputs from other industries, while industry 4 requires inputs from all three other industries. The positive coefficient on own price index shows the deterring effect of price competition within industry, while the negative coefficient on the price index for goods used as inputs shows the cost reducing benefit of agglomeration. The positive coefficient on the maximum local wage is counter-intuitive, as we would expect firms to benefit from a low local wage, but this feature may be correlated with proximity to existing profitable firms, shown by profit_prox, as those two variables have alternating significance for the effect on the two industries. The coefficient on proximity to unemployed is very large and positive, which is intuitive as the labor is available for hire fairly cheaply and will also serve as a customer base. Note that the coefficient is less positive and insignificant for industry 4, which relies on inputs from other goods and cannot subsist purely off of labor. Proximity to the road has a positive effect on profitability for both industries, although this may be partially due to the central placement of the road. Lastly, we can see that industry 1 is in general more profitable than industry 3, likely because it does not need to purchase inputs.

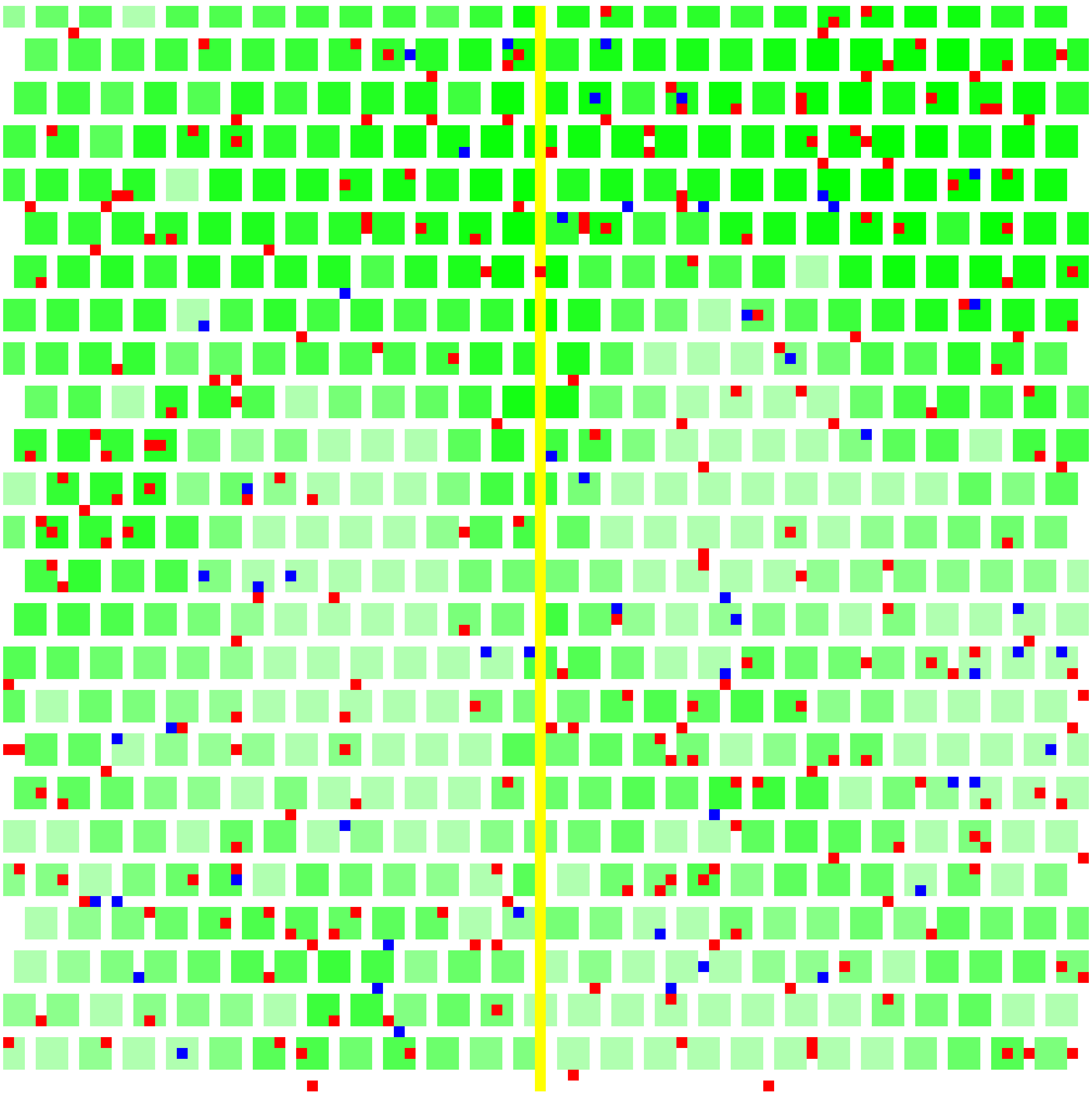


Figure 3.6: Profitability of Locations Before Market Saturation

	Dependent variable:	
	profit0	profit3
	(1)	(2)
P1	0.123***	-0.013**
	(0.010)	(0.005)
P2	0.006**	0.004***
	(0.003)	(0.002)
P3	0.003	-0.003**
	(0.002)	(0.001)
P4	0.109	1.046***
	(0.085)	(0.044)
wage	0.195***	-0.012
	(0.031)	(0.016)
cons_prox	3.139***	5.209***
	(1.192)	(0.613)
num_unemp_prox	3.918***	0.762
	(1.449)	(0.745)
profit_prox	-0.053	-0.059**
	(0.049)	(0.025)
road_prox	0.256***	0.242***
	(0.052)	(0.027)
Constant	-48.496***	-65.109***
	(5.870)	(3.017)
Observations	625	625
R^2	0.301	0.717
Adjusted R^2	0.291	0.713
Residual Std. Error (df = 615)	7.617	3.915
F Statistic (df = 9; 615)	29.401***	173.382***
Note:	*p<0.1; **p<0.05; ***p<0.01	

Figure 3.7: Regression of Profitability on Location Features

3.4 Conclusion

This paper developed an agent based model to simulate monopolistic competition in a two-dimensional plane to experiment with the effect of improving the speed of a roadway on the clustering behavior of industries with different characteristics. The outcomes of the model are sensitive to the parameters and modeling assumptions, but firms with more inter-industry links exhibited higher clustering and industries with higher sensitivity to transportation costs showed a slight dispersal in response to a road improvement.

The results showing the profitability for various locations are intuitive. Price competition deters profit, proximity to inputs reduces transpotation cost and improves profit, the effect of local wage is unclear due to its high correlation with the existing profitable locations much like the effect of cities, areas with high unemployment offer high profitability, and being near the roadway likely increases profits.

Future work for this model include incorporating the endogenous movement of households, explicit land use requirements for production and consumption, rent competition for households and industries, the use of physical capital by industries and investing behavior for households, an oligopolistic pricing framework with strategic interactions between firms, and of course improving the speed of the algorithms.

Bibliography

Acemoglu, Daron, Vasco Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. 2012. The Network Origins of Aggregate Fluctuations. *Econometrica*, 80:5, 1977-2016.

Ahlfeldt, Gabriel, Fabian Bald, and Duncan Roth. 2019. The Spatial Equilibrium with Migration Costs. Working Paper, London School of Economics, Universitat Duisburg-Essen, Institut Fur Arbeitsmarktforschung.

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

Allen, Treb and Costas Arkolakis. 2014. Trade and the Topography of the Spatial Economy. *The Quarterly Journal of Economics*, 129:3, 1085-1140.

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

Anderson, James. 1979. A Theoretical Foundation for Gravity Equation. *American Economic Review*, 69:1, 106-116.

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.

Appert, Manuel and Chapelon Laurent. 2007. Measuring urban road network vulnerability using graph theory: the case of Montpellier's road network. *HAL*.

Auschauer, David Alan. 1999. Is Public Expenditure Productive? *Journal of Monetary Economics*, 23, 177-200.

Baqae, David. 2018. Cascading Failures in Production Networks. *Econometrica*, 86:5, 1819-1838.

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

Beatty, Stephen. 2015. An evolution of tolling. KPMG International.

Bell, Andrew and Kelvyn Jones. 2015. Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data. *Political Science Research and Methods*, 3:1, 133-153.

Berliant, Marcus and Ping Wang. 2019. General Equilibrium Theories of Spatial Agglomeration. *Oxford Research Encyclopedia of Economics and Finance*.

Bigio, Saki and Jennifer La'O. 2016. Distortions in Production Networks. *National Bureau of Economic Research*. Cambridge, MA.

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. Understanding Interaction Models: Improving Empirical Analyses. *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.

Cambridge Systematics, Inc. 2008. The Highway Construction Equity Gap. Texas Department of Transportation.

Carvalho, Vasco and Alireza Tahbaz-Salehi. 2019. *Annual Review of Economics*, 11, 635-663.

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.

Christaller, Walter. 1933. *The Central Places of Southern Germany*. English translation. Carlisle W. Baskin. Englewood Cliffs, NJ: Prentice-Hall, 1966.

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.

Cooley, Charles Horton. 1894. The Theory of Transportation. *Sociological Theory and Social Research: Being Selected Papers of Charles Horton Cooley*, 17-118. Originally published in *Publications of the American Economic Association*, 9.

' Department of Transportation. 2021. Benefit-Cost Analysis Guidance for Discretionary Grant Programs.

Dixit, Avinash and Joseph Stiglitz. 1977. Monopolistic competition and optimum product diversity. *American Economic Review*, 67 297–308.

Dixon, P. M., J. Weiner, T. Mitchell-Olds, and R. Woodley. 1988. Erratum to 'Bootstrapping the Gini Coefficient of Inequality.' *Ecology*, 69, 1307.

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 and Matthew Turner. 2012. Urban Growth and Transportation. *The Review of Economic Studies*, 79:4, 1407-1440.

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

Dutzik, Tony and Benjamin Davis. 2011. Do Roads Pay for Themselves? U.S. PIRG Education Fund.

Eaton, Jonathon and Samuel Kortum. 2002. Technology, Geography, and Trade. *Econometrica*, 70:5, 1741-1779.

Epstein, Joshua and Robert Axtell. 1996. Growing Artificial Societies: Social Science from the Bottom Up. *The MIT Press.*

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.

Fernald, John G. 1999. Roads to Prosperity? Assessing the Link between Public Capital and Productivity. *The American Economic Review*, 89:3, 619-638.

Forkenbrock, David, Shauna Benshoff, and Glen Weisbrod. 2001. Assessing the Social and Economic Effects of Transportation Projects. Prepared for National Cooperative Highway Research Program. Transportation Reseach Board.

Fowler, Cristopher S. 2011. Finding equilibrium: how important is general equilibrium to the results of geographical economics? *Journal of Economic Geography*, 11, 457-480.

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.

Fujita, Masahisa and Hideaki Ogawa. 1982. Multiple equilibria and structural transition of non-monocentric urban configurations. *Regional Science and Urban Economics*, 12:2, 161-196.

Fujita, Masahisa, and Paul Krugman. 2004. The new economic geography: Past, present, and the future. *Regional Science*, 83, 139-164.

Glaeser, Edward and Joshua Gottlieb. 2009. The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States. *National Bureau of Economic Research*. Cambridge, MA.

Gollin, Douglas, Martina Kirchberger and David Lagakos. 2017. In Search of a Spatial Equilibrium in the Developing World. *National Bureau of Economic Research*. Cambridge, MA.

Grassi, Basile. 2017. IO in I-O: Size, Industrial Organization, and the Input-Output Network Make a Firm Structurally Important. Working Paper, Bocconi University, Milan, Italy.

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.

Hanson, Gordon H. 2005. Market potential, increasing returns and geographic concentration. *Journal of International Economics*, 67, 1-24.

Harris, Chauncy D. 1954. The market as a factor in the localization of industry in the United States. *Annals of the Association of American Geographers*, 64, 315–348

Haughwout, Andrew F. 2002. Public infrastructure investments, productivity and welfare in fixed geographic areas. *Journal of Public Economics*, 83, 405-428.

Head, Keith and Thierry Mayer. 2011. Gravity, Market Potential and Economic Development. *Journal of Economic Geography*, 11, 281-294

Helpman, Elhanan. 1998. The size of regions. *Topics in Public Economics: Theoretical and Applied Analysis*, ed. by D. Pines, E. Sadka, and I. Zilcha, Cambridge University Press: Cambridge, 33–54.

Hodge, Daniel, Daniel Beagan, Tyler Comings, Glen Weisbrod, and et al. 2008. Economic Impact Study of Completing the Appalachian Development Highway System. Prepared for the Appalachian Regional Commission. Cambridge Systematics.

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.

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.

Jones, Charles. 2013. Misallocation, Economic Growth, and Input-Output Economics. In *Proceedings of Econometric Society World Congress*, ed. D Acemoglu, M Arellano, E Dekel, pp. 419–55. Cambridge, UK: Cambridge Univ. Press.

Kilkenny, Maureen, and Jacques-Francois Thisse. 1999. Economics of location: A selective survey. *Computers and Operations Research*, 26, 1369-1394.

Krugman, Paul. 1991. Increasing returns and economic geography. *Journal of Political Economy*, 99, 483–499.

Krugman, Paul. 1992. A Dynamic Spatial Model. *National Bureau of Economic Research Working Papers*. Cambridge, MA.

Krzyzanowski, Witold. 1927. Review of the Literature of the Location of Industries. *Journal of Political Economy*, 35:2, 278-291.

Kula, Mehmet. 2008. Supply–Use and Input-Output Tables, Backward and Forward Linkages of the Turkish Economy. *The 16th Inforum World Conference in Northern Cyprus*.

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.

Lemoy, Rémi, Charles Raux, and Pablo Jensen. 2012. Exploring the polycentric city with an agent-based model. *HAL*.

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.

Liu, Ernest. 2019. Industrial Policies in Production Networks. *The Quarterly Journal of Economics*, 1883-1948.

Litman, Todd Alexander and Eric Doherty. 2009. Transportation Cost and Benefit Analysis: Techniques, Estimates and Implications. Second Edition. *Victoria Transport Policy Institute*.

Long, John and Charles Plosser. 1983. Real Business Cycles. *Journal of Political Economy*, 91, 39-69.

Losch, August. 1938. The Nature of Economic Regions. *Southern Economic Journal*, 5:1,71-78.

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

Marshall, Alfred. 1919. *Industry and Trade*. Macmillan, London.

Martens, Karel and Floridea Di Ciommo. 2017. Travel time savings, accessibility gains and equity effects in cost-benefit analysis. *Transport Reviews*.

McDonald, John F. 2007. William Alonso, Richard Muth, Resources for the Future, and the Founding of Urban Economics. *Journal of the History of Economic Thought*, 29:1, 67-84.

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.

Muth, Richard F. 1961. Economic Change and Rural-Urban Land Conversions. *Econometrica*, 29:1, 1-23.

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.

Olner, Dan. 2014. An agent-based modelling approach to spatial economic theory. Ph.D. thesis, University of Leeds, UK.

Olner, Dan, Andrew Evans and Alison Heppenstall. 2015. An agent model of urban economics: Digging into emergence. *Computers, Environment and Urban Systems*, 54, 414-

427.

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

O'Sullivan, Arthur. 2003. *Urban Economics*. Boston: McGraw-Hill/Irwin. 5th ed.

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.

Partridge, Mark, Dan Rickman, Rose Olfert, and Ying Tan. 2012. *Munich Personal RePEc Archive*, The Ohio State University, Oklahoma State University, University of Saskatchewan.

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

Pereira, Alfredo M. and Rafael Flores de Frutos. 1999. Public Capital Accumulation and Private Sector Performance. *Journal of Urban Economics*. 46, 300-322.

Peucker, Thomas K. 1968. Johann Georg Kohl, A Theotretical Geographer of the 19th Century. *The Professional Geographer*, 20:4, 247-250.

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

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

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.

Samuel, Peter. 2007. The Role of Tolls in Financing 21st Century Highways. Reason Foundation.

Sarkar, Trishna, Dababrata Sarkar, and Prolay Mondal. 2020. Road network accessibility analysis using graph theory and GIS technology: a study of the villages of English Bazar Block, India. *Spatial Information Research*, 29, 405-415.

Sasaki, Yuya and Paul Box. 2003. Agent-Based Verification of von Thünen's Location Theory. *Journal of Artificial Societies and Social Simulation*, 6:2.

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

Singh, Prasoon, Vinay Shankar Prasad Sinha, Ayushi Vijhani, and Neha Pahuja. 2018. Vulnerability assessment of urban road network from urban flood. *Iternational Journal of Disaster Risk Reduction*, 28, 237-250.

Smith, Adam. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations.* Edinburgh, London.

Smith, Margot W. 1979. A Guide to the Delineation of Medical Care Regions, Medical Trade Areas and Hospital Service Areas. *Public Health Reports*, 94:3, 247.

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.

Storper, Michael. Agglomeration, Trade, and Spatial Development: Bringing Dynamics Back in. 2010. *Journal of Regional Science*, 50:1, 313-342.

Sturm, Jan Egbert and Jakob de Haan. 1995. Is public expenditure really productive?: New evidence for the USA and The Netherlands. *Economic Modelling*, 12:1, 60-72.

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

Tabuchi, Takatoshi. 1998 Agglomeration and dispersion: A synthesis of Alonso and Krugman. *Journal of Urban Economics*, 44, 333–351.

Thomson, Robert and Dianne Richardson. 1995. A Graph Theory Approach to Road Network Generalization. *International Cartographic Conference*.

Time. 1995. Business: Private Toll Roads Show the Way.

Ting, Lei, Deren Li, and Jianya Gong. 2000. The Expression of Road Networks for Vehicle Navigation. *International Archives of Photogrammatery and Remote Sensing*, 33:B4, 567-571.

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, 44-67.

Venables, Anthony. 1996. Equilibrium locations of vertically linked industries. *International Economic Review*, 37, 341–359.

von Thünen, 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`.

Xie, Feng and David Levinson. 2005. Measuring the Structure of Road Networks. *Geographical Analysis*, 39, 336-356.