Infrastructure Investment, Self-Employment, and Structural Change in the US Labor Market*

R. Benjamin Rodriguez†

October 2025

Abstract

This paper studies the effects of infrastructure investment on the labor market. Between 1920 and 1950, the US government began constructing its first interstate highways, the Numbered Highway System. Regions gaining access to the Numbered Highway System experienced significant population growth alongside a shift from self-employment to salaried jobs. To identify the causal effect of this highway network on local labor markets, I use a novel instrumental variable approach exploiting a hypothetical set of highways proposed by the US Army for national defense. I further interpret these causal effects through a spatial equilibrium model in which highway construction induces households to select out of self-employment as local agglomeration—manifested in higher land costs and denser markets—increases the fixed cost of production near highways. I find that investment in the Numbered Highway System can account for one fifth of the decline in the aggregate self-employment rate over this period, underscoring how infrastructure investment fosters regional integration and structural change.

Keywords: Transportation, Occupational Choice, Structural Transformation
JEL Classification: R40, O18, J21, O14, R11, N72

*I am grateful to John Haltiwanger, John Shea, and Jose Asturias for their guidance, encouragement, and thoughtful suggestions. Daniel Agness, Emek Basker, Thomas Drechsel, Matthew Gentzkow, Richard Hornbeck, Ethan Kaplan, Guido Kuersteiner, Simone Lenzu, Nuno Limão, Philipp Ludwig, Tyler Pike, Alex Rodrigue, Cory Smith, Luminita Stevens, Ludwig Straub, Jeffrey Traczynski, and participants at multiple seminars and conferences provided useful comments that significantly improved the paper.

†Department of Economics, University of Maryland - College Park; Email: rbenrod@umd.edu.

1 Introduction

As economies grow, the share of workers in low-productivity self-employment typically declines, giving way to larger average firm size and more efficient resource allocation (Kuznets and Murphy 1966). The role of self-employment itself evolves with development, shifting from a dominant source of income to a residual or entrepreneurial margin as wage jobs expands. Yet, systematically studying these patterns of structural change remains difficult today, as modern developing economies often rely on limited, survey-based employment data rather than comprehensive administrative records.

During the early 20th century, the United States experienced a similar structural change: self-employment fell by roughly 27 percent between 1920 and 1950 as economic activity shifted from small unincorporated enterprises to larger, multi-worker firms. While this broad transition is well documented, its underlying causes also remain incompletely understood (Fairlie and Meyer 2000). This paper provides new evidence that federal investment in highway infrastructure was a key driver of the shift toward salaried employment by lowering transportation costs, fostering agglomeration, and reallocating labor toward wage work in denser markets.

Using a database that merges newly digitized full-count Census data and historical geographic information on local highway construction, I identify the causal effect of infrastructure investment on local labor markets with a new instrumental variables approach and measure its aggregate effect using a spatial equilibrium model. I find that local areas that benefited from infrastructure investment experienced significant population growth alongside a marked decline in self-employment, indicating that the decline in self-employment is partly driven by increasing agglomeration forces. Embedding these local effects in the model, I find that investment in US highways caused at least a two percentage point decline in the national rate or explains roughly one fifth of the observed decline. Together, these findings highlight a previously unrecognized channel through which infrastructure investment drives structural change.

The first part of the paper establishes this mechanism empirically by exploiting a historic natural experiment: the construction of the Numbered Highway System (NHS), the first federally funded highway network in the US. Built between 1920 and 1950, the NHS was constructed when both automobiles and modern paving methods were still in their infancy, with the explicit goal of stimulating economic development by connecting economic and population hubs.[1] As shown in Figure 1, the NHS was a vast network of rural

[1]Using the NHS as a natural experiment is well-timed from an identification standpoint. The design of the NHS began in 1921 and was finalized in 1926 prior to the onset of the Great Depression.The construction

Figure 1: The Numbered Highway System - 1926 Finalized Plan

interstate highways which spanned all 48 states.

To identify exogenous variation in the layout of the NHS, I construct a new instrumental variable for local highway access that combines the historic Pershing Map, a hypothetical network used in the design phase of the NHS, with the recentering approach proposed by Borusyak and Hull (2023). The Pershing Map originates from the US Army's 1922 proposal for a national highway aimed at enhancing military logistics and national defense capabilities. The empirical appeal of the Pershing Map lies in its timing and straight-line network layout. A key limitation of the Pershing Map is that — even with a defense-oriented focus — its route design is not entirely orthogonal to local economic fundamentals. To address this concern, I create a recentered version of the instrument that leverages the Pershing Map's geometric simplicity while accounting for the fact that the US Army's route choices at major intersections may have been endogenously aligned with the level of local economic development or growth potential. This recentered instrument is a strong predictor of the eventual NHS layout and is uncorrelated with key pre-treatment characteristics such as railroad access, manufacturing value added, and farmland value, supporting its validity for causal inference.

Using detailed data and plausibly exogenous variation in highway access, I document a series of causal effects of infrastructure investment on local labor markets. First, be-

of the NHS was also nearly completed by 1940, allowing me to study its long-run impact on the US labor market prior to the construction of the Interstate Highway Network beginning in 1956.

tween 1920 and 1950, local access to the NHS increased county-level population by 44 log points, indicating that highways substantially increased the attractiveness of connected areas and promoted a migration toward areas with highway access. Second, highway access led to a 4.3 percentage point decline in self-employment amongst prime-working age men over the same period, suggesting that lower transportation costs and improved market integration encouraged workers to transition from self-employment to wage jobs. Third, this decline was not driven by shifts across sectors but occurred mainly within high self-employment sectors such as agriculture and services, reflecting changes in the type of work performed rather than the sector of employment. Moreover, I find no evidence that this decline was driven by reduced search frictions for wage jobs, suggesting that highways affected occupational choice primarily through changes in local market conditions rather than easier job matching. Finally, using within-individual variation, I show that local market density largely explains why individuals exited self-employment: those who moved to areas that were ex ante denser—and therefore more developed—experienced the sharpest reductions in self-employment, consistent with rising fixed costs of production in increasingly dense local markets.[2]

While these causal estimates help explain how the NHS lowered self-employment in relative terms, the second part of the paper studies how infrastructure investment affected the aggregate economy. To answer this question, I develop a novel spatial equilibrium model with endogenous selection into self-employment, mobile labor, and trade of goods to estimate the effect of the NHS on the national self-employment rate. In the model, the construction of the NHS alters trade costs between regions. In the long run, these changes in relative costs increase market access in areas connected by the NHS and by extension increase population density as households flock towards opportunities presented by expanding local markets. Within each local labor market, households choose between working as salaried employees or earning business income with self-employment. To capture the observed negative relationship between population density and self-employment, the model assumes that self-employed households must purchase a parcel of land as a fixed input to begin production.[3] As a result, when households migrate to areas that experi-

[2]Additional analysis shows that among counties located near breaks in highway construction during the early 1930s, those that had just gained highway access experienced a significant increase in automobile registrations per capita in the following year relative to nearby control counties. This finding helps explain why local populations expanded with highway access, as automobile travel was in high demand at the time. It also supports the validity of my identification strategy, indicating that road construction was not driven by prior automobile utilization.

[3]The assumption that fixed costs are denominated in land units follows Behrens et al. (2014), who suggest this mechanism explains regional variation in self-employment. Because land is an inelastically supplied local good, increases in population density raise land prices and, consequently, the fixed cost of pro-

ence reductions in trade costs following the construction of the NHS, the cost of owning a business increases and the local self-employment rate decline.

Using the model, I study the counterfactual change in the national self-employment rate that comes solely from infrastructure investment between 1920 and 1950. To do so, I measure local changes in market access due to the construction of the NHS and estimate its effect on local population size and self-employment. Taking the state of the 1920 US economy as given, changes in local market access solely reflect the completion of the NHS by 1950. Quantitatively, the spatial equilibrium model predicts that investment in road infrastructure led to a 1.7 percentage relative decline in the self-employment rate, compared to 9.2 percentage point decline observed in the data. This implies that the model explains approximately one fifth of the overall decline in self-employment.

My findings make clear how infrastructure investment can foster both local and aggregate economic development by reallocating resources away from less productive self-employed businesses (Aschauer 1989). This mechanism parallels the selection effects observed in other periods focused on free trade and regional integration, where lowering market frictions enables more productive firms to expand and less productive ones to exit (Melitz and Trefler 2012). Importantly, highway infrastructure investment was a result of well-timed and targeted policy: investing in highway construction reduced transportation costs at a moment when demand for automobile travel was rising rapidly but still constrained, amplifying the economic impact of its investment (Eli et al. 2025).

Literature Review. This paper contributes to three strands of literature: selection into self-employment, evidence on spatial distribution of structural change, and the benefits of infrastructure investment. The historical context of the NHS sheds light on the policies and mechanisms that lead to the decline in self-employment as economies grow. Broadly speaking, self-employed individuals can be classified as either entrepreneurs or sole proprietors.[4] These two groups tend to be distinct (Levine and Rubinstein 2017). Entrepreneurs play a central role as drivers of economic growth in the US (Decker et al. 2014; Walsh 2023) while poorer and less able individuals are more likely to engage in forms of self-employment with lower productivity (Herreño and Ocampo 2023).

This paper shows how infrastructure policy aimed at promoting mobility contributed to the long-run decline in self-employment by strengthening agglomeration forces. These findings offer guidance for policymakers in low- and middle-income countries that are

duction.

[4]Sole proprietors, also known as owners of non-employer or unincorporated businesses, are often used as a proxy for informal labor in the development literature. See La Porta and Shleifer 2014 for a discussion on informal firms in developing economies.

now undergoing rapid urbanization and considering large-scale infrastructure investments (Khor 2022). While the historical effects reflect the land-intensive production environment of the early 20th century US, the broader mechanism still remains relevant for high-income economies today. However, the context has changed over the course of the century in the digital age: businesses may rely less on physical capital such as land but face new costs associated with digital access, supply chain disruptions, and the rise of the gig economy (Abraham et al. 2024). Overall, policies that seek to influence self-employment should therefore target the underlying costs of doing business—whether physical or digital—rather than simply encouraging a shift toward wage employment.

The NHS's impact on the spatial distribution of economic activity provides evidence on the spatial dimension of structural change, in which local economic development coincided with a decline in self-employment and growth in average firm size (Kuznets and Murphy 1966, Gollin 2008, Bick et al. 2022). Notably, Fairlie and Meyer (2000) originally find US self-employment was declining in the early 20th century within as opposed to between industries or sectors. Guida-Johnson (2022) and Perra et al. (2024) provide reduced-form evidence in a development context that infrastructure investment reduces the size of the informal sector, which is predominantly occupied by unincorporated businesses and serves as a common proxy for the incidence of self-employment in developing countries.[5]

Beyond these studies, I offer both empirical evidence, exploiting rich administrative data, and a formal model that explains the mechanism for how infrastructure investment led to a decline in self-employment. Applying this model to my empirical setting, I quantify the aggregate effect that the NHS had on self-employment, not just the relative local effect. This counterfactual exercise provides the economic significance on the role of infrastructure investment in driving structural change at a national level.

Finally, this paper contributes new evidence on the benefits of infrastructure investment during a transformative but understudied period on transportation infrastructure. Within the US context, influential work such as Donaldson and Hornbeck (2016) and Hornbeck and Rotemberg (2024) document the economic benefits of railroad expansion preceding the NHS, while studies including Michaels (2008) and Allen and Arkolakis (2014) examine impacts of the Interstate Highway System that followed the NHS. Research on the evolution of the automobile industry during the early 20th century for example by Cooper and Haltiwanger (1993) and Eli et al. (2025) provides important com-

[5] Economic geography research has documented other patterns of structural change that are often distributed unevenly within countries and depend on the reduction of frictions such as trade costs. These include a declining agricultural sector (Adamopoulos 2011; Eckert and Peters 2022), the rise of services (Faber and Gaubert 2019; Fajgelbaum and Redding 2022), urbanization effects (Michaels et al. 2012), and increasing labor skill sophistication (Michaels et al. 2019).

plementary insights for this time period, but there is little research on the effects of road construction during this time due to the lack of available data and clean identification.

The closest paper to this study is Michaels et al. (2019) (MRR), which shares the same empirical context but focuses on the evolution of tasks performed in production.[6] MRR documents how the construction of federal highways and telephone lines lowered trade costs for tasks, resulting in a spatial concentration of more interactive tasks in urban areas.[7] However, this work does not focus on how infrastructure investment affected the selection into self-employment, which is a distinctly different margin in the labor market.

Complementary to MRR, I find evidence that infrastructure investment would lead to a marked increase in more efficient resource allocation for production as those exiting self-employment transitioned to salaried employment. Moreover, I apply new methods from Borusyak and Hull (2023) to estimate the causal effects of the NHS without the concern that both local access to the NHS and changes in local labor markets were attributed to baseline differences in local economic development.

Roadmap. The rest of this paper is structured as follows: Section 2 describes this paper's empirical context and data sources. Section 3 presents the identification strategy and estimated causal effects of the NHS. Section 4 describes the theoretical framework for a general equilibrium analysis of the NHS's effect on self-employment. Section 5 presents the counterfactual analysis based on the theoretical framework. Section 6 concludes.

2 Background and Data

2.1 Historical Background

In this section, I provide information on the empirical context for both the federal government's investment in the NHS and the economic geography of self-employment during the first half of the 20th Century.[8]

---

[6] To my knowledge, Morin and Swisher (2016) is the only other paper to have studied road construction in the early 20th century; however, neither it nor MRR address changes in population density or self-employment, nor provide estimates for the aggregate effect of the NHS.

[7] I argue there is clear empirical evidence showing the mechanism I study differs from MRR. For example, MRR find that urban areas initially had relatively less interactive tasks but gained a comparative advantage post-NHS construction. In contrast, I document that self-employment rates were initially lower in denser areas and negatively correlated with contemporaneous increases in population density. Unlike finding in MRR of an urban comparative advantage emerging after the NHS, my findings highlight how agglomeration forces continuously raised fixed costs of production, amplifying the decline in self-employment through infrastructure investment.

[8] The section on infrastructure investment is in large part thanks to the historical account in Swift (2011), which describes the construction of US roads from its infancy in the early 20th Century through the con-

2.1.1 Investment in the NHS

To situate the NHS within its historical trajectory, it is helpful to first recall the developments that preceded its creation. Automobile enthusiasts began lobbying for the development of rural highways in the late 1900s and through the 1910s to promote automobile travel outside of major metropolitan areas. During this phase, individual interstate highways such as the Lincoln Highway, which would later become part of the NHS, provided a proof-of-concept for the value of interstate travel. Moreover, engineers experimented with the materials used to fortify roads. Because roads were initially composed of dirt or macadam (a mixture of crushed stone and a bituminous binding agent), their structure would weaken with use, leading to the use of tarmacadam or asphalt for paving improved roads. The Federal Aid Road Act of 1916 was the first attempt by the federal government to fund road construction. While it spurred an initial attempt at construction, it suffered from a lack-of-coordination across different levels of government and the US entering into World War I.

Congress revisited constructing an interstate highway network in the early 1920s with the Federal Highway Act of 1921. The Act introduced the first formula-based allocation of federal funds: a 50–50 federal–state match that could be applied to no more than seven percent of a state's roads designated for interstate travel. States contributed input on the interstate road layout, while the Bureau of Public Roads also relied on Census data—such as local population counts and indicators on economic activity in manufacturing, agriculture, and mining—to guide its planning decisions. The overall goal of the Bureau of Public Roads was to create a network of 'through roads,' which would lead to a crowding-in effect for state and local governments to build onto the NHS using local roads and state highways. This process led to the design of the NHS, as shown in Figure 1, which was presented to Congress in November 1926.

The 1926 plan for the NHS consisted of 75 thousand miles of roads and spanned all 48 states in the US. Using digitized state atlases, which I describe in more detail in the next subsection, I track the completion of the NHS over time and across space. For example, Figure 2 shows both the overall fraction of paved NHS roads and the binned county-level averages, where counties are grouped into deciles by planned road length. In aggregate, 31 percent of the NHS was already paved by 1926, likely reflecting earlier private highways such as the Lincoln Highway and federal road investments dating to 1916. The construction of the NHS during the 1920s and 1930s was extensive with 68 percent of roads being paved by 1931 and 93 percent of roads being paved by 1939. Lastly, there is

struction of the Interstate Highway System.

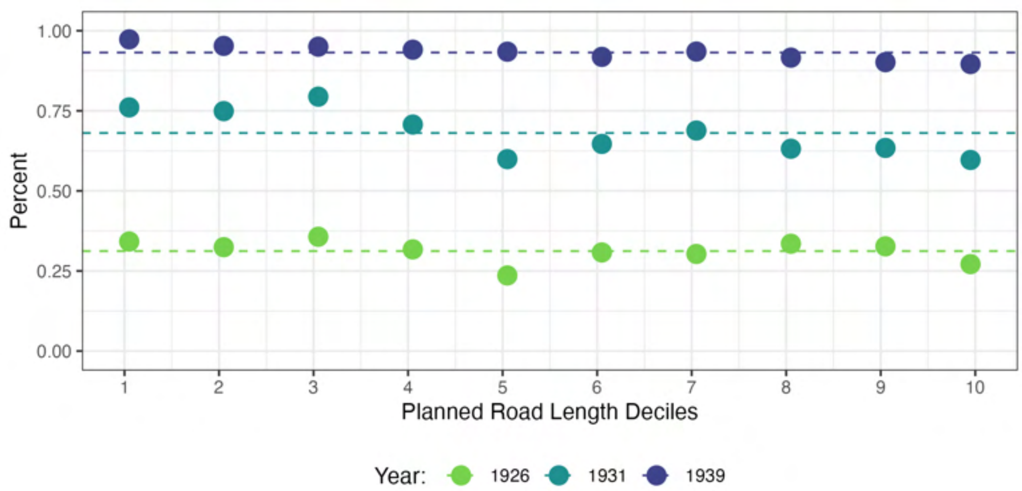


Figure 2: Completed Sections of the NHS over Time

Notes: This figure plots the length of paved roads as a fraction of the total length of NHS roads for 1926, 1931, and 1939. Counties are binned into deciles based on the intended length of NHS roads per county. The horizontal dashed line plots the fraction of paved NHS roads across counties.

little variation in the completed portions of the NHS across deciles.

2.1.2 Self-Employment

Between 1920 and 1950, the self-employment rate among prime-age men fell nine percentage points—from 33 to 24 percent. In this section, I provide a high-level overview of the data to illustrate how highway infrastructure investment may have contributed to this long-run decline.

A large portion of the decline in self-employment could, in principle, reflect the rise of manufacturing, given the substantial cross-sector differences in self-employment rates. In 1920, self-employment was 51 percent in agriculture, 23 percent in services, and only 6 percent in manufacturing, implying that a shift of labor toward manufacturing would mechanically reduce overall self-employment. To quantify the contribution of intersectoral labor reallocation to the aggregate decline, I apply a simple accounting decomposition that separates within-sector changes in self-employment from changes in employment shares across sectors.[9] Within-sector declines account for roughly five per-

[9] The exact decomposition for the change in national self-employment rate between 1920 and 1950 is:

$$\Delta\ \text{Self-Emp}^{20-50} = \sum \omega_k^{20} \Delta\text{Self-Emp}_k^{20-50} + \sum \Delta\ \omega_k^{20-50}\text{Self-Emp}_k^{20} + \sum \Delta\ \omega_k^{20-50} \Delta\ \text{Self-Emp}_k^{20-50}$$

where ω is the fraction of the labor force in sector $k \in$ {Agriculture, Services, Manufacturing} The first term captures within-sector changes in self-employment, the second reflects the effect of labor reallocation

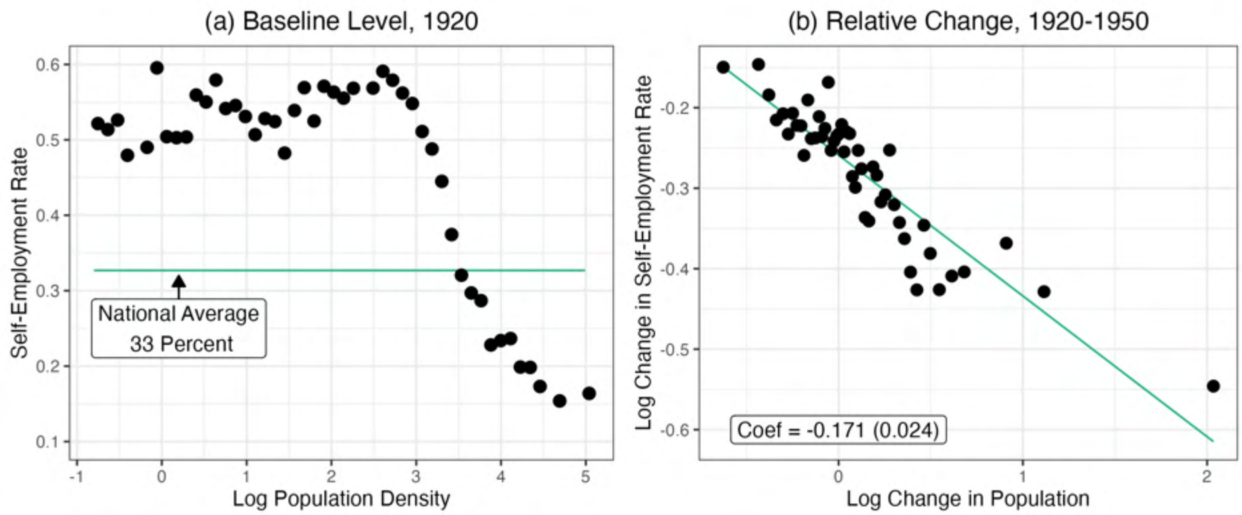


Figure 3: Economic Geography of Self-Employment

Notes: This figure plots the binned scatterplot, relating local population density and self-employment. Panel A plots the relationship between level of population density and self-employment for 1920, the last year of data before highway construction began. Panel B plots the change in local population density and self-employment between 1920 and 1950. The OLS coefficient and robust standard error from regressing log changes in self-employment on log changes in population density are included in Panel B.

centage points of the total reduction in self-employment, while labor reallocation toward low–self-employment sectors explains another five points. These effects are partially offset by a one–percentage–point increase as manufacturing both expanded and experienced a modest rise in its own self-employment rate. Overall, about half of the aggregate decline can be attributed to structural reallocation toward manufacturing, yet a substantial share remains unexplained by the standard structural transformation narrative.[10]

To explore how geography shaped the long-run decline in self-employment, I next examine its spatial distribution. Figure 3 shows a strong negative relationship with population density. In 1920, the average self-employment rate in low-density counties ranged from 50 to 60 percent—well above the national average—and declined steadily as density increased. Moreover, changes in local population density and self-employment are strongly negatively correlated. In the next section, I quantify how much of these shifts in population and local self-employment can be attributed to the construction of the NHS.

across sectors, and the final term represents their covariance.

[10] Despite the rich microdata available from the Decennial Census, few studies document changes in self-employment during the early twentieth century. An exception is Fairlie and Meyer (2000), who analyze a random sample of non-agricultural male workers from 1910–1970 and find that self-employment declined primarily within industries. Using full-count Census microdata for prime-working-age men from 1920–1950, I confirm similar within-sector patterns across all industries. The key distinction is that my decomposition leverages complete population data rather than a sampled subset.

To motivate the mechanisms through which infrastructure investment contributed to the decline in self-employment, I focus on three channels. First, the NHS may have affected self-employment through either within-sector adjustments or intersectoral labor reallocation. The latter channel would be consistent with the NHS reinforcing the United States' comparative advantage in manufacturing. Second, improved local connectivity could have reduced the cost of searching for wage employment, encouraging transitions out of self-employment. Third, rising local development may have increased the fixed costs of operating small businesses, prompting less productive owners to exit self-employment and seek wage work instead.

2.2 Data

In this section, I briefly describe the variety of sources used to compile the main datasets for the analysis, which include historical state road atlases, microdata from the 1910 through 1950 Decennial Censuses, and various sources of county-level information. Additional information on the sources and digitization of historical data is provided in the Appendix A.

On the construction of the NHS, I create a digitized database of geographic information system (GIS) shapefiles for the set of roads included in the final plan presented to Congress in 1926 and identify the location of completed paved portions of the NHS in 1926, 1931, and 1939. The final plan of the NHS is provided by the University of Texas at Arlington. The completed portions of the NHS comes from historical state road atlases published by Rand McNally. While additional roads are added to the NHS over time after 1926, I focus on the original set of roads due to greater concerns of endogeneity of later additions. The Pershing Map is also available through the NBER and digitized in a similar fashion to the NHS plans. The various highway shapefiles are merged with county shapefiles available through IPUMS NHGIS to identify the set of counties that had access to the NHS according to the finalized plan or subsequently in a given year (Manson et al. 2024). I additionally observe the length of the NHS within each county and calculate the straight-line distance to the nearest county with access to an NHS road based on each county's centroid. I repeat this process as well for the Pershing Map.

I use the individual-level responses to the Decennial Census from 1910 to 1950, available through IPUMS USA, to study the employment dynamics for the prime-working-age male population (ages 25-54) who work outside of the public sector (Ruggles et al. 2024). The microdata include information on age, race, occupation, associated industry of employment, county-of-residence and the class of a worker to identify individuals as either self-employed or wage workers across all years in my sample. I use the IPUMS Multigenerational Longitudinal Panel to link individual responses across Censuses on a pairwise basis. Labor force participation (i.e. being gainfully employed) is available for all years except for 1920, and an indicator for whether or not someone, who is self-employed, is an employer or non-employer is available for all years except 1950. Labor income is available in 1940 and 1950, and business income is available only in 1950. The main dataset used in the analysis includes all prime-working-age males with a county-of-residence included in my sample.

Lastly, I draw on a variety of sources to construct county-level measures of population density and indicators of local economic development for the baseline years in my sample. I use data from the Censuses of Agriculture, Manufacturing, and Population from 1910 to 1950 to measure county-level aggregates such as total population, manufacturing value added, and average farmland value, all available through IPUMS NHGIS (Manson et al. 2024). Population density is measured as the ratio of county population counts from the Census of Population to county geographic area, with area derived from county shapefiles available through IPUMS NHGIS. I incorporate information on the 1916 US railroad network to identify counties with rail access before construction of the NHS (Atack et al. 2010). I also digitize the locations of domestic military bases and camps from World War I using historical US Army publications, available here. Counties that were part of a major metropolitan area during the sample period are identified using historical US Census Bureau data.[11] For each county, I calculate the straight-line distance to the nearest major metropolitan area and World War I military base based on the county's centroid. Following Saiz (2010), I measure the fraction of each county that is uninhabitable using US Geological Survey shapefiles to identify surface water and raster files available through Amazon Web Services Terrain Tiles and the Open Topography global datasets API to measure the fraction of land with a slope greater than 15 percent.

3 Empirical Analysis

In this section, I estimate the causal effect of the NHS on long-run changes in population and self-employment using an instrumental variable strategy based on the Pershing Map.

[11]The definition of metropolitan areas is based on the 1930 Decennial Census and included in historical documents here . To account for growth in metropolitan boundaries over the sample period, I use the list of counties included in each metro area according to the 1950 Decennial Census here.

3.1 Population

Through the lens of a spatial equilibrium with mobile labor, households should respond to relative changes in long-run local economic conditions by moving towards areas that prosper. In this context, areas that are more amenable to automobile travel should see an increase in local population as automobiles provided a new low cost form of travel and trade.[12] Since the construction of the NHS infrastructure took several decades, I analyze regional differences in local labor market conditions between 1910 and 1950.

3.1.1 Identification Strategy

To test whether or not the local effects of the NHS results in an increase in local population, I build on the regression specification from Fajgelbaum and Redding (2022) and estimate the effect of the highway's presence using the following regression specification:

$$\Delta \ln \text{Pop}_c = \alpha + \beta \, \mathbb{1}[c \in \text{NHS}] + X_c' \, \Omega + \epsilon_c \tag{1}$$

where c indexes counties in the US. The variable, $\mathbb{1}[c \in \text{NHS}]$, is an indicator equal to one if a NHS road goes through county c. The vector X_c includes an initial set of controls such as baseline population, area, and the fraction of a county that is covered in water or has above a 15 percent incline in elevation.[13] In turn, a county's baseline population controls for potential heterogeneity in the baseline level of economic development. A county's size controls for variation in the likelihood of a highway line going through a given county. Lastly, the fraction of a county that is covered in water or has above a drastic incline controls for a county's suitability for automobile travel. These baseline controls allow for time-varying effects of initial county-level differences. To account for correlated errors in nearby counties, I cluster the error term ϵ_c at the state level.

Estimating the regression specification in equation (1) by OLS would produce an unbiased estimate for the local effects of the NHS on population growth if gaining access

[12] In Appendix Section B, I provide evidence showing the effect that gaining access to the NHS had on short-run changes in automobile utilization using a spatial regression discontinuity approach. These additional results support the notion that the NHS spurred the local adoption of automobiles, which would entice households to migrate to be closer to the NHS.

[13] Controlling for the fraction of a county that is covered in water or has above a 15 percent incline in elevation is in line with the definition used in Saiz (2010) to define land that is undevelopable. These covariates differ slightly from the original specification in Fajgelbaum and Redding (2022) who use latitude and longitude to control for alternative explanations of geographic variation population growth that are unrelated to the construction of associated transportation network. While not reported in the text, I find this alternative set of controls provide a sharper point estimate than controlling for latitude and longitude. I also show in Table C3 that additionally controlling for latitude and longitude does not significantly change the estimated coefficient.

to the NHS was quasi-random. However, this assumption would go against the intentions of the NHS to connect economics hubs within and across states.[14] To account for this potential endogeneity, I construct an instrumental variable using a highway network that was designed by the US Army following World War I in recommendation for the planned NHS. This highway network is commonly referred to as the Pershing Map as it was designed under the guidance of General John Pershing and has been used in previous studies to study the effects of the Interstate Highway.[15]

The validity of the instrumenting the NHS indicator using the placement of roads in the Pershing Map to produce a relatively unbiased coefficient estimate relies on the policy objective of the US Army to design a highway network for national defense and logistics as opposed to local economic development. This argument for the unadjusted instrument's validity is plausible as the Pershing map exhibits straight line connections between cities along the US border and interior, which is an indication that the US Army did not intend to include unnecessary locations in the highway's design. However, the inclusion of straight line highways does not directly address the concern that local interests of the US Army are potentially correlated with local economic development. In particular, these concerns are likely most prevalent at the location of nodes in the Pershing map where potential highway lines would intersect. For example, four Pershing lines connect in Newport News, VA, which housed the production of warships such as the USS Enterprise, a prominent Navy carrier in World War II. The manufacturing of such vessels provides local fiscal stimulus, would be of primary concern for national defense, but remains unobservable in my database. To account for non-random placement of nodes in the network, I transform the Pershing indicator to produce a novel version of the instrument in line with Borusyak and Hull (2023):

$$Z_c = \mathbb{1}\left[c \in \text{Pershing}\right] - \frac{1}{N_c}\sum_{k=1}^{N_c} \mathbb{1}\left[k \in \text{Pershing} \mid \text{neighboring county to } c\right] \tag{2}$$

which recenters a binary indicator equal to one if a county would receive highway access

[14]When developing the NHS, the Bureau of Public Roads worked with state and local governments to identify areas that would receive a rural highway, allowing for lobbying by the lower levels of government. Moreover, the Bureau of Public Roads used county-level Census data to identify counties within each state that were largest in term of population and sectoral output.

[15]This hypothetical network and later revisions have been used in studies to instrument for the layout of the Interstate Highway System on an unadjusted basis (e.g. Baum-Snow 2007, Michaels 2008, Duranton and Turner 2012, and Frye 2024) While not always explicitly referred to as the Pershing Map, some studies construct instruments based the map of the US Army's recommended layout for a highway network in the 1940s. This recommendation in the 1940s mirrors the recommendation made in the 1920s. Note also that the Pershing Map is not the sole instrument employed in each of these works.

according to the Pershing Network by differencing out the average access of the N_c adjacent counties. In doing so, I remove the variation at and around nodes that is systematic given an increased density of potential highways.

Figure 4 plots the differences between the unadjusted and recentered Pershing instrument. The effective treatment and control groups based on the unadjusted Pershing instrument are plotted as light- and dark-shaded counties in Panel A of Figure 4. The effect of recentering is apparent based on the contrasting variation across counties in Panel B of Figure 4 (indicated by change in high and low recentered highway access). Relative to Panel A, the recentered instrument no longer places network nodes in the effective treatment group since their treatment is expected given the increased density of roads. Similarly, coastal (interior) counties that are clustered around (away) from any Pershing line are no longer in the effective treatment (control) group. Instead, the effective treatment and controls groups are generally rural counties that are distant from dense clusters of Pershing lines where local interests of the US Army are potentially least likely to be correlated with local economic development.

The identification strategy to estimate any local effects from the NHS requires that my proposed instrument is a strong predictor of the finalized NHS and does not effect outcomes through alternative channels conditional baseline controls. To provide graphical evidence for the predictive strength of the instrument, Appendix Figure C6 displays a binscatter plot of the first stage relationship along with its corresponding regression coefficient. To further probe for a potential violation of the exclusion restriction, I regress the recentered Pershing instrument on 1920 county characteristics controlling again for baseline controls, and examine the size and significance of the coefficient estimates.[16] The primary concern in this test is that a significant correlation with observables gives credence to the concern the unexpected inclusion in the Pershing network is correlated with local economic characteristics.

To show the relative strength of recentering the Pershing indicator, I also regress the NHS indicator, the unadjusted Pershing indicator, and the average Pershing exposure for neighboring counties on the same set of characteristics from equation (1). One should expect the NHS indicator is correlated with local economic indicators based on the Bureau

[16]By conditioning on baseline controls, primarily initial differences in population density, I am testing whether or not highway networks favored larger economic hubs holding population density constant. Both networks potentially favored intersecting at economic hubs, implying there is a likely conditional correlation with local economic indicators. The goal of recentering is to purge the Pershing network from any conditional correlation that comes from the location of network nodes. To control for population density, I include a county's population and size as separate controls in each regression. The preferred specification uses the 1920 population counts, but results are robust to instead controlling for 1910 population counts, which were likely population counts that were available when the NHS design began.

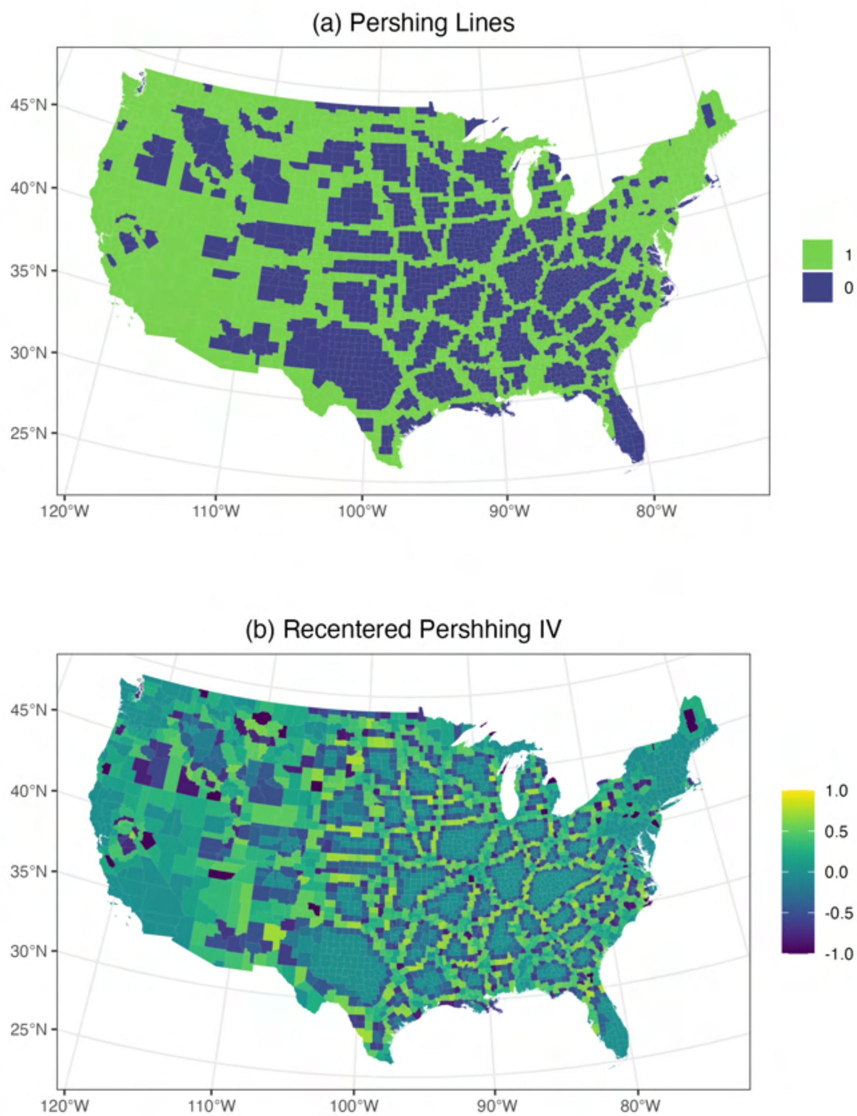


Figure 4: Highway Access with the Pershing Map

Note: This figure plots the values for the Pershing IV. Panel A plots the unadjusted Pershing IV, where a value of one indicates areas that would have received highway access according to the Pershing Map. Panel B plots the Pershing IV after recentering according to equation (2).

.

of Public Roads' intention to connect economic hubs. The unadjusted Pershing indicator can also be correlated with local economic indicators if the local interest of the US Army is correlated with local economic development as I discussed in the motivation for recentering. Lastly, I also include the measure for the expected inclusion in the Pershing network to mechanically show how recentering effects the Pershing indicator and produces a more robust instrument.

Table 1 reports conditional correlations between the Pershing instrument and baseline county characteristics. Columns (1) and (2) show that both the endogenous NHS indicator and the unadjusted Pershing indicator are significantly correlated with variables such as local railroad access and manufacturing output, indicating that the exclusion restriction is violated for the unadjusted instrument. However, these correlations are explained by counties' systematic inclusion in the Pershing Map, as reflected in the similarly significant correlations in column (3). As a result, the recentered Pershing instrument in column (4) exhibits no significant correlation with baseline characteristics. Consistent with this reduction in explanatory power, the adjusted R-squared falls relative to the specification using the unadjusted Pershing indicator.

The weaker correlation between baseline characteristics and the recentered Pershing instrument, relative to the NHS and unadjusted Pershing indicators, shows that recentering lowers the concern of any endogenous network effects. The recentered instrument therefore provides quasi-random variation for estimating the local effects of the NHS. By extension, I assume that unexpected inclusion in the Pershing network is arguably exogenous to baseline local economic conditions, conditional on controls.

3.1.2 Empirical Results

I now turn to an examination of whether access to the NHS led to changes in a county's local population in the long-run. I begin by focusing on counties outside major metropolitan areas to ensure that estimated effects are not driven by agglomeration forces around large US cities. I then conduct robustness checks to show that the long-run effects are not an artifact of this sample restriction. My primary specification estimates equation (1) using the recentered Pershing indicator, defined in equation (2), as an instrument for the NHS indicator.

Table 2 reports OLS, reduced-form, and IV estimates of the NHS's effect on local population, varying the set of control variables. Baseline controls, described in the previous subsection, include 1920 population (log), county area (log), and the fraction of unsuitable land.[17] I add state fixed effects to capture unobserved regional variation in geography and state-funded highways. The full set of controls further includes distance to the nearest metro area (log), indicators for local railroad access and the presence of a military base, and 1920 average farmland value (log). Reporting point estimates as these controls

[17] Baseline controls account for initial county characteristics that could influence highway placement. Specifically, population size proxies for initial economic development, land area captures the likelihood of highway routing, and water coverage or steep terrain reflects geographic suitability for automobile travel.

Table 1: Correlation with Baseline Characteristics, 1920

	Dependent variable:			
	NHS	Pershing	Neighbors	IV
	(1)	(2)	(3)	(4)
Local Access to Railroads	0.227***	0.130**	0.113***	0.026
	(0.051)	(0.062)	(0.043)	(0.045)
Local Military Base	0.007	−0.374***	−0.345***	−0.029
	(0.052)	(0.068)	(0.059)	(0.057)
Dist to Military Base (logs)	−0.008	−0.124***	−0.110***	−0.014
	(0.012)	(0.017)	(0.014)	(0.010)
Local Metro Area	−0.063	0.095	0.023	0.060
	(0.097)	(0.104)	(0.094)	(0.055)
Dist to Metro Area (logs)	−0.021	−0.005	−0.020	0.013
	(0.021)	(0.023)	(0.022)	(0.010)
Manufacturing VA (logs)	0.030***	0.039***	0.036***	0.004
	(0.008)	(0.009)	(0.010)	(0.005)
Farmland Value (logs)	0.074***	−0.005	0.006	−0.010
	(0.017)	(0.023)	(0.022)	(0.010)
Baseline Covariates	Y	Y	Y	Y
Adjusted R^2	0.110	0.157	0.219	0.032
N	2,755	2,755	2,755	2,755

Note: This table displays the conditional correlation between county-level indicators for access to either the NHS, the network on the Pershing Map, the anticipated inclusion to the network on the Pershing Map, or the Pershing IV as defined in equation (2). *p<0.1; **p<0.05; ***p<0.01.

are sequentially added documents the extent to which variation in local highway access may be correlated with potential confounders.

Focusing on results included in panels A and C, two important patterns emerge. IV point estimates indicate that local access to the NHS increased a county's local population by 0.44 to 0.48 log points, which are larger in magnitude than OLS estimates of 0.08 to 0.11 log points.[18] Across either specification, the inclusion of additional control variables leads to more positive point estimates for the coefficient of interest, although IV estimates

[18]Given the magnitude of point estimates presented in Panel B, IV estimates are larger in magnitude to their OLS counterpart as a result of the first-stage relationship between the NHS indicator and the recentered Pershing instrument. Fajgelbaum and Redding (2022) find results of a similar magnitude for population growth in reference to the effect of local railroad access in Argentina between 1869-1914. Taking their point estimate of 0.65 log points over a 45 year period from Table 2 and rescaling this value to a 30 year period, the implied point estimate is approximately 0.43 log points.

Table 2: Specification Tests for the Local Effect of the NHS, 1920-1950

	Dependent variable:			
	Change in Population (logs)			
	(1)	(2)	(3)	(4)
Panel A: OLS Estimates				
NHS	0.071^{*}	0.087^{***}	0.118^{***}	0.114^{**}
	(0.037)	(0.027)	(0.043)	(0.042)
Panel B: Reduced Form Estimates				
Pershing IV	0.057^{***}	0.066^{*}	0.075^{**}	0.074^{**}
	(0.015)	(0.035)	(0.036)	(0.035)
Panel C: IV Estimates				
$\widehat{\text{NHS}}$	0.284^{***}	0.444^{**}	0.483^{**}	0.477^{**}
	(0.067)	(0.221)	(0.209)	(0.202)
F statistic	48	28	28	27
Baseline Covariates	N	Y	Y	Y
State Fixed Effects	N	N	Y	Y
Full Set of Covariates	N	N	N	Y
N	2,733	2,733	2,733	2,719

Note: This table displays the 2SLS coefficient estimates from equation (1) using the Pershing IV as defined in equation (2). Additional OLS and reduced-form estimates are included in panels A and B, respectively. Baseline covariates include 1920 population (logs), county size (logs), the fraction of a county covered by water, and the fraction of a county with an average change in elevation of 15 percent or more. The full-set of controls include: distance to nearest military base and metro area, binary indicators for a local military base or metro area, average farm land value (logs), and a binary indicator for local access to railroads (1916). *p<0.1; **p<0.05; ***p<0.01.

increase only slightly relative to the baseline specification. These findings suggest that variation in local highway access is at least partially correlated with pre-existing county characteristics, which raises the concern that estimated effects remain biased due to omitted unobserved differences that are correlated with the recenter Pershing instrument.

To assess concerns of potential pre-trends, I regress log changes in local population on local highway access for periods preceding NHS construction and find no statistically or economically significant relationship. Table 3 presents OLS and IV estimates for changes in local population from 1910 to 1950 relative to baseline year of 1920 (i.e. changes in population from 1910 to 1920, 1920 to 1930, etc.). The effect of local highway access in the pre-construction period of 1910 to 1920 is negative and smaller in magnitude for the

Table 3: Local Effect on Population Growth Before and After Construction Began

	Census year:			
	1910	1930	1940	1950
	(1)	(2)	(3)	(4)
Panel A: OLS Estimates				
NHS	−0.013	0.017	0.033*	0.087***
	(0.014)	(0.013)	(0.018)	(0.027)
Panel B: IV Estimates				
$\widehat{\text{NHS}}$	0.001	0.220	0.266*	0.444**
	(0.069)	(0.145)	(0.156)	(0.221)
Baseline Covariates	Y	Y	Y	Y
N	2,733	2,733	2,733	2,733

Note: This table displays the 2SLS coefficient estimates from equation (1) using the Pershing IV as defined in equation (2). OLS estimates are additionally included in Panel A. The county-level changes in population is measured as the log difference in population relative to the population counts in 1920. The year for the corresponding population counts is listed as the header of each column. *p<0.1; **p<0.05; ***p<0.01

OLS estimates but is a tightly estimated null effect for the IV estimates. In contrast, local highway access has a positive and statistically significant effect only after the construction of the NHS began. Furthermore, the effect of local highway access increases in size and significance as construction the NHS nears completion. These findings indicate that results based on exogenous variation from the recentered Pershing IV do not capture pre-existing population trends nor anticipatory road placement and instead identifies population growth driven by the actual rollout of the highway network, reinforcing the credibility of the identification strategy.

In addition to the results reported in this section, Tables C3 and C4 provides additional robustness checks concerning the point estimates measuring the effect of local highway access on population growth. The additional results in Table C3 address the following potential concerns: (1) including counties within metropolitan areas in the sample, (2) using the inverse hyperbolic sine function instead of a log function to measure the change in local population, and (3) controlling for a county's relative location in the country based on its latitude and longitude. The additional results in Table C4 account for factors related to the Great Depression, where I include additional controls for New Deal spending and the decline in the manufacturing sector and local credit markets between 1929 and 1931.[19] Estimates across each robustness check are similar to the IV estimates in Tables 2 in terms of sign, magnitude, and statistical significance.

3.2 Self-Employment

Having established evidence that the NHS led to an increase in local population for the set of treated counties, I now look at how local labor markets evolved in response. In particular, I focus on the decision individuals face between working for oneself versus working for a salary. Using a similar regression specification to equation (1), I isolate long-run changes in self-employment that are in response to the construction of the NHS.

3.2.1 Long-Run Changes in Self-Employment

To estimate the local effects of the NHS on self-employment using the available Census microdata, I use the following regression specification:

$$Y_{it} = \alpha_{c(i)} + \gamma_t + \beta\,\mathbb{1}\,[c(i) \in \text{NHS}] \times \mathbb{1}\,[t > 1920] + X_{it}'\,\Omega + \epsilon_{it} \tag{3}$$

where i indexes individuals in the Census microdata, $c(i)$ indexes an individual's county-of-residence, and t indexes years in my sample period. The variable, Y_{it}, is a binary indicator that signifies if an individual is self-employed or works in a specific sector (i.e. agriculture, manufacturing, or services). I additionally add binary controls for an individual's occupation and sector of work, which are interacted with the post 1920 indicator, to control for time-varying differences across occupation and sectors. Otherwise, the specification in equation (3) is relatively unchanged to the county-level specification in equation (1), and by extension, the identifying assumptions are the same as well.

IV estimates presented in Table 4 indicate that highway access reduced self-employment by approximately 4 percentage points between 1920 and 1950. Relative to the pre-construction average of 46 percent among rural counties that were eventually connected to the NHS, this effect corresponds to roughly a 9 percent decline in self-employment, calculated as the ratio of the estimated change to the baseline rate. Similar

[19] New Deal spending is the county-level amount of spending per capita in logs. Data on New Deal spending comes from Wallis (1987). To control for the decline in manufacturing, I use a shift-share variable equal to the local employment weighted to the national decline in manufacturing value added between 1929 and 1931 based on reported changes in manufacturing value added from Fabricant (1940). For both variables, I rely on population and employment data from the 1930 Decennial Census. Lastly, I use the log change in county-level deposits from 1929 to 1931 from FDIC (1992).

Table 4: Local Effect on Self-Employment, 1910-1950

	Dependent variable: Self-Employment					
	1910-1920			1920-1950		
	(1)	(2)	(3)	(4)	(5)	(6)
NHS	0.001			0.001		
	(0.002)			(0.002)		
Pershing IV		-0.001			-0.006^{***}	
		(0.001)			(0.002)	
$\widehat{\text{NHS}}$			-0.011			-0.043^{**}
			(0.010)			(0.014)
NHS Mean			0.483			0.456
N	19,589,336	19,589,336	19,589,336	21,652,300	21,652,300	21,652,300
Baseline Covariates	Y	Y	Y	Y	Y	Y
County Fixed Effects	Y	Y	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y	Y	Y

Note: This table displays the 2SLS coefficient estimates from equation (3) using the Pershing IV as defined in equation (2) where the dependent variable is a binary indicator for being self-employed. Additional OLS and reduced-form estimates are included in columns 1, 2, 4, and 5, respectively. See Table 2 for a description of baseline covariates. The non-NHS mean refers to the average self-employment rate for individuals in counties without access to the NHS. The sample is a subset of counties outside of metro areas. *p<0.1; **p<0.05; ***p<0.01

to Table 3, I report OLS, reduced-form, and IV estimates for changes in self-employment across two periods: the pre-construction decade (1910–1920) and the treatment period (1920–1950). The estimated effect during the pre-construction decade is small—a statistically insignificant one percentage point decline—further supporting the validity of the identification strategy.

While the Decennial Census in 1910 and 1930 through 1950 included questions regarding an individual's labor force participation, the 1920 Decennial Census excluded any such question. For this reason, estimates on the local effects on self-employment, such as Table 4, are measured for all available individuals in my sample including those that are currently unemployed or outside of the labor force. As a robustness check, Table C5 presents the local effect on self-employment from 1910 to 1950 for all individuals and the subset of those employed. Reassuringly, the estimated effect on self-employment across both cuts of data differ by only 0.3 percent. For additional robustness, Table C6 displays the effect of the NHS on self-employment using county-level data. Overall, the direction and statistical significance of the coefficient estimates using county-level data are consistent with the results using individual-level data in Table 4.

Table 5: Exiting Self-Employment versus Sectoral Reallocation, 1920-1950

	Sector:			
	Overall	Agriculture	Manufacturing	Services
	(1)	(2)	(3)	(4)
Panel A: Self-Employment				
$\widehat{\text{NHS}}$	-0.033^{***}	-0.037^{**}	-0.014	-0.036^{*}
	(0.011)	(0.018)	(0.018)	(0.019)
N	21,652,300	10,819,196	4,160,303	6,672,801
Panel B: Sectoral Reallocation				
$\widehat{\text{NHS}}$		-0.009	-0.009	0.018
		(0.015)	(0.026)	(0.022)
N		21,652,300	21,652,300	21,652,300
Baseline Covariates	Y	Y	Y	Y
County Fixed Effects	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y

Note: This table displays the 2SLS coefficient estimates from equation (3) using the Pershing IV as defined in equation (2). The overall effect of the NHS on self-employment differs from the 2SLS estimate from Table 4 for the 1920-1950 period as observations are reweighted to measure the sector-level average. The other coefficient estimates in Panel A correspond to for the same regression specification for individuals within specific sectors of the economy. Consistent with the IPUMS variable `IND1950`, the agriculture sector corresponds to values 1-299, the manufacturing corresponds to values 300-499, and the service sector corresponds to 500-899. The dependent variable for panel B corresponds to binary indicators for being employed in either the agriculture, manufacture, and service sectors as defined in this table description. *p<0.1; **p<0.05; ***p<0.01

Lastly, Table C7 shows the local effect of the NHS does not impact other margins of the labor force participation outside of being self-employed versus working for a wage. Specifically, I look at long-run changes labor force participation and overall employment. For this robustness test, I continue to use individuals in the 1910 and 1950 Decennial Census to take advantage of the relevant questions on labor force participation. Relative to the approximate 5.6 percentage point decline in self-employment in columns 4 and 5 of Table C7, the estimated effect on labor force participation and overall employment are an order of magnitude smaller and statistically insignificant.

3.2.2 Within-Sector Declines versus Sectoral Reallocation

During my sample period, the US economy experienced a dramatic shift away from agricultural production in favor of manufacturing. Hence, the decline in self-employment is in part a result of the shifting labor from one sector with a high self-employment rate to a low self-employment rate sector.[20] To assess whether this mechanism explains the NHS effect on self-employment, Table 5 presents the local effects of the NHS on self-employment by sector and the reallocation of labor across sectors between 1920 and 1950. Relative to estimates presented in Table 4, I weight individual observations to measure the average effect across sectors as opposed to the population weighted change.[21] Panel A indicates that the effect of the NHS on self-employment was similar in sign across sectors but varying in magnitude and significance. There is a 3.3 percentage point decline in self-employment for agriculture and service sectors, which are both statistically significant, whereas the manufacturing sector experienced a smaller and insignificant decline in self-employment. Panel B also shows that the effects of the NHS did not cause a significant shift across sectors but does suggest a slight reallocation of labor towards the service sector. Taken together, the negative effect of the NHS on self-employment came as a result of within sector changes in self-employment rather than a reallocation of labor across sectors, which is consistent with FM as discussed in the previous section.

3.3 Analysis of Mechanisms with Within-Individual Analysis

Having ruled out that sectoral reallocation was a driving force behind the negative effect of the NHS on self-employment, I propose an alternative mechanism based on the process of urbanization, where the fixed cost of production increases as agglomeration forces strengthen near highways. I propose that areas that were connected by the NHS experienced a significant increase in local population density. As land supply is presumably inelastic, the increase in population density would result in an increase in the cost of land, meaning if business revenues were in part spent on leasing land (e.g. a storefront, manufacturing plant, or farm land), the increase in land prices would lower business income for the self-employed, and as a result, less productive business owners would transition

[20] Fairlie and Meyer (2000) find that the decline in self-employment in occurred primarily within industries. Their sample focused on non-agricultural industries. Relative to their work, I test whether or not the NHS caused a decline in self-employment by reallocating individuals from high self-employment intensive sectors such as agriculture and services to low self-employment intensive sectors such as manufacturing. As noted in Section 2, sectoral reallocation played a significant role in the overall decline in self-employment.

[21] For the regression estimates included in Table 4, I inversely re-weight observations based on the number of individuals who work in each sector in 1920 and 1950. In doing so, the overall effect of the NHS on self-employment reflects the sector-level average.

out of self-employment.

3.3.1 Identification Strategy

To test the explanatory power of this proposed mechanism, I look at the local effect of the NHS on self-employment within individuals. To do so, I link individuals across adjacent Census years and identify the effect of the NHS based on individuals that move between counties, controlling for unobservables at the individual level. I am then able to interact the treatment variable of changing highway access with indicators signifying if an individual moves to a more or less urban area using base period county characteristics.[22]

The main regression specification examines the relationship between local access to the NHS and self-employment within individuals:

$$\mathbb{1}\left[\text{Self-Emp}_{it}\right] = \alpha_i + \gamma_t + \beta\,\mathbb{1}\left[c(it) \in \text{NHS}\right] + X_{it}'\,\Omega + \epsilon_{it} \tag{4}$$

where individuals i are linked across adjacent Census years (e.g. 1910-1920, 1920-1930, etc.) and included in the sample if they move between counties during the interim years.[23] In this specification, the treatment indicator, $\mathbb{1}[c(it) \in \text{NHS}]$, changes based on whether or not an individual's county-of-residence has access to an NHS road before or after their move. This indicator is instrumented using a similarly defined version of the Pershing instrument. To account for county-level characteristics that are important for cross-sectional identification with the Pershing instrument, I also include time-varying measures of baseline county characteristics in the vector of controls X_{it}, which are consistent with the set of controls included in equations (1) and (3). The specification includes individual fixed effects α_i and year fixed effects γ_t, where the inclusion of individual fixed effects ensures that the effect of the NHS is identified from within-individual variation. The error term is clustered at the individual's pre-move state-of-residence to account for

[22]Using a within-individual specification is advantageous in my analysis given the available data on the value of land at the county level. The Census of Agriculture provides average value of farm land across decades in my sample period, but this measure likely provides only partial information regarding the influence of population density on the overall value of land. Focusing on home values as a more informative alternative, the earliest collection of average home values is in the 1920 Census of Housing for 273 cities providing little variation in treatment for my base period (Fishback and Kollmann 2014). Instead by using a within-individual specification and the sample of movers between 1940 and 1950, I can interact the treatment effect from the NHS with indicators for increasing land values using the average home values from the 1940 Decennial Census. This particular specification measures heterogeneous effects of the NHS due to changes in the costliness of a given location.

[23]Because I am identifying the effect of the NHS within individuals, I include individuals that originally reside in a metro counties in the base period or move between rural and metro counties in my preferred specification.

Table 6: Exiting Self-Employment and Population Density, 1940-1950

	Dependent variable:			
	Self-Employment			
	(1)	(2)	(3)	(4)
$\widehat{\text{NHS}}$	-0.057^{**}	-0.024	-0.021	0.027
	(0.027)	(0.028)	(0.031)	(0.031)
$\widehat{\text{NHS}}$ × Closer to Metros		-0.079^{**}		
		(0.031)		
$\widehat{\text{NHS}}$ × Higher Density			-0.111^{***}	
			(0.038)	
$\widehat{\text{NHS}}$ × More Expensive				-0.128^{***}
				(0.035)
Baseline Covariates	Y	Y	Y	Y
Individual Fixed Effects	Y	Y	Y	Y
Occupation × Year Fixed Effects	Y	Y	Y	Y
N	3,518,814	3,518,814	3,518,814	3,518,814

Note: This table displays the 2SLS coefficient estimates from equation (4) using the Pershing IV as defined in equation (2). The interaction term in column 2 is a binary indicator equal to one if an individual's post-move county-of-residence is part of a metro area. The interaction term in column 3 is a binary indicator equal to one if an individual's post-move county-of-residence has a greater population density as of 1940. The interaction term in column 4 is a binary indicator equal to one if an individual's post-move county-of-residence has a higher average home value as of 1940. For columns 2 through 4, the interaction term is additionally interacted with a year fixed effect to account for time-varying effects across Census years. *p<0.1; **p<0.05; ***p<0.01

spatially correlated errors. Lastly, I construct sample weights to account for the non-random set of individuals, who are connected across Census years, so observations in all regressions are weighted accordingly.[24]

3.3.2 The Rising Fixed Cost of Production

Table 6 examines whether the patterns of heterogeneity in the data are consistent with the hypothesis that increasing value of land from rising population density led to a decline in self-employment. Column 1 presents the baseline results for the sample of movers

[24]The method to construct population weights follows the Census Linking Project. Weights are constructed based on a GLM regression where the dependent variable is equal to one to indicate movers and the independent variables include indicators for occupation, self-employment, and region. Regions correspond to Census regions as defined in the link here.

between 1940 and 1950 and shows that local access to the NHS led to a six percentage point decline in self-employment after accounting for unobservable differences across individuals. Column 2 examines whether effects are larger for individuals that move closer to major metro areas, where the local self-employment rate is generally lower than the national average. Similar to the hypothesis that labor was shifting towards the manufacturing sector, migrating towards urban areas that have a lower baseline self-employment rate may mechanically induce individuals to exit self-employment and is in line with the long-run increase in urbanization within the US (Boone and Wilse-Samson 2023). I interact the NHS treatment indicator with an indicator equal to one if an individual's county-of-residence in 1950 is relatively closer to a metro area than their 1940 county-of-residence. The effect of the NHS on self-employment lowers in magnitude and significance for individuals that did not move closer to a major metro area and becomes significantly more negative for individuals that move closer to a major metro area.

To show instead that local population density and by extension increasing land values are material drivers for the decline in self-employment, Columns 3 and 4 focus on pre-move differences in density and costs as opposed to the relative to distance to other markets. Column 3 examines the effect of the NHS when an individual moves to a county that has an initially higher population density, where the treatment indicator is interacted with an indicator equal to one if an individual's move is towards a higher density county as of 1940. Point estimates suggest that individuals who do not move to a more densely populated county did not experience a statistically significant decline in self-employment but the negative effect on self-employment increases significantly for the individuals who do move to a more densely populated county. Column 4 directly tests whether the cost of land is driving the negative effect of the NHS on self-employment using the 1940 county level median home value. For individuals who did not move to a more expensive county, the effect of the NHS on self-employment is again statistically insignificant, whereas the NHS only causes a decline in self-employment for the individuals that moved to a relatively more expensive county.[25]

[25] Table C8 provides additional robustness tests for results presented in Table 6 with respect to the sample selection and regression specification, and results are broadly similar across all specifications in terms of direction, magnitude, and statistical significance. These robustness tests include: (1) excluding individuals who are employers according to the 1940 Decennial Census, (2) excluding individuals who are unemployed or non-labor force participants, and (3) including all individuals connected between the 1940 and 1950 Decennial Census.

3.3.3 An Alternative Decline in the Fixed Cost of Wage Work

An alternative explanation for the decline in self-employment is the falling fixed cost of wage work proposed by Bick et al. (2022). In their framework, economic development lowers the cost of entering salaried employment rather than raising the fixed cost of operating a business. In my context, this mechanism would operate through increases in local road density that reduce search costs, thereby encouraging transitions into wage work.[26] A key implication is that the intensive margin of labor increases with development, as transitions into wage employment are associated with longer working hours and more weeks worked per year (Bick et al. 2022). These predictions provide testable implications to assess whether this alternative mechanism explains the observed decline in self-employment. To evaluate this, I interact the instrumented NHS treatment with measures capturing local road density and labor market intensity—specifically, the number of hours worked in the previous week and the number of weeks worked in the previous year.

Table C9 shows these factors do not have a significant effect on the decline in self-employment, implying the improved access to search for a job locally did not act as a significant pull factor. To test this alternative mechanism, I first measure differences local transportation costs decline as the density of local roads as of 1939.[27] These measures account for both the county-level total length of roads constructed and density of roads. Neither measure accurately explains the significant decline in self-employment although they appear to at least capture some explanatory power for the effect of the NHS. Second, I interact the treatment indicator with two indicators to account for moves to a more developed labor market. These include the log difference in the average weeks worked in 1939 and the average number of hours worked between March 24th and 30th in 1940 between origin and destination counties. Again, both measures appear to enhance increase the effect of the NHS but are statistically insignificant. These results suggest that declining fixed costs of wage work, as modeled by Bick et al. (2022), do not account for the observed reduction in self-employment.

[26]Under this mechanism, the rise in local population and the decline in self-employment could occur independently, since areas that gained the most market access from the NHS may also have experienced a disproportionately greater expansion of local roads.

[27]I use the length of NHS roads as of 1939 to account for only pre-existing portions of the NHS as opposed to sections that would be constructed concurrently with the timing of individuals movements across counties.

3.3.4 Additional Robustness for Within-Individual Analysis

For my sample of movers, I compare changes in self-employment over a ten-year interval across four potential forms of treatment: staying near the NHS, staying away from the NHS, moving to the NHS, and moving away from the NHS. The identifying assumptions for this series of treatment follow a difference-in-difference design, which assumes cross-sectional variation in treatment is quasi-random and changes in employment did not development prior to an individual's move. I establish validity in treatment using a falsification test to show self-employment did not systematically differ in the pre-move year based on their exposure to the NHS and the effect of the NHS is insignificant prior to its construction for the 1910-1920 linked sample. For this falsification test, I provide cross-sectional estimates on the effect of the NHS before and after an individuals move for each pair of linked Census years in my sample period. To account for characteristics that are inevitably swept out in the mover estimate with individual fixed effects, I again include occupation and sector fixed effects in line with the regression specification in equation (3). Table C10 presents the estimates for this falsification test. Across each set of paired Census years as shown in Panel A, there are no significant effects from the NHS prior to an individual's move. Similarly in Panel C, there are only significant effects from the NHS after its construction began starting with the 1920-1930 linked sample.[28]

Two additional robustness tests further clarify the mechanism. First, Table C11 shows that the NHS had no significant effect on the number of employers between 1910 and 1940, implying that the program primarily reduced the prevalence of sole proprietorships rather than employer-operated firms. This test cannot be extended to the 1940–1950 sample of movers because it relies on a question in the Decennial Census asking whether individuals were working on their own account or as employers.[29] For the 1940–1950 mover sample, however, I estimate a similar regression using a binary indicator for self-employment as the dependent variable and include an interaction term identifying employers. The results confirm that highway access primarily affected unicorporated

[28]Additionally, Figure C7 presents the estimated treatment effect based on equation (4) by an individual's pre-move occupation. Occupation groups include: professionals, farmers, managers, clerical workers, sales workers, craftsmen, operatives, service workers, farm laborers, and laborers. Results indicate once construction began the NHS led to a decline in self-employment amongst managers, sales workers, craftsmen, farmers, and farm laborers between 1920 and 1950. These results are in line with those presented in Table 5, which showed that the decline in self-employment was largely concentrated within agriculture and service sectors.

[29]This question was excluded from the 1950 and 1960 Decennial Censuses and replaced in 1970 by a question on business incorporation. Levine and Rubinstein (2017) use a similar question in the CPS to distinguish between one-person businesses and more productive forms of entrepreneurship.

rather than employer businesses.[30] Second, Table C12 examines the robustness of results in Panel C of Table C10 based on the choice of controls, particularly with respect to the set of time fixed effects. For the specification test, Panels A through C substitute in state-by-year, sector-by-year, and occupation-by-year fixed effects. Across all three panels, the effect of the NHS on self-employment is insignificant for the 1910-1920 linked sample and increases in magnitude and significance for subsequent samples. In particular, occupation-by-year fixed effects provide the most robust point estimates across the linked samples. When using occupation-by-year fixed effects, the treatment effect for the 1910-1920 sample is an accurately measured null result up to the second decimal place. Moreover, the growth in magnitude and significance across subsequent samples is most in line with the effect of the NHS on local population growth in Table 3.

4 Theoretical Framework

This section develops a spatial equilibrium model connecting self-employment and infrastructure investment, which is consistent with the empirical results presented in the previous section. The key mechanism is that highway construction reduces transportation costs, which attracts residents and workers, increasing local population over time. Higher population density, in turn, raises land prices, thereby increasing the fixed cost of production. The structure of the model allows me to calibrate parameters and quantify the aggregate impact of the NHS on the national self-employment rate in Section 5.

4.1 Model Set-Up

Following the baseline model presented in Redding (2016), I consider a static spatial equilibrium in which households choose between working for themselves versus working for a wage. The economy has a population $\overline{L}$ and consists of $|\mathbb{N}|$ counties indexed by $l,\ n,\ k\ \in \mathbb{N}$. Counties differ from each other in terms of their land supply, productivity, and relative location within the country.

Household preferences are a Cobb-Douglas combination of housing (i.e. land) and consumption goods. The latter uses a CES aggregator to combine traded varieties, which

[30] Although these results are not shown in a separate table, the key coefficients can be summarized as follows. For the 1940–1950 sample of movers, the NHS caused a six percentage–point decline in self-employment among non-employers, significant at the 5 percent level and consistent with the results in Table 6. In contrast, the NHS had an insignificant but slightly positive effect on employers, consistent with the findings in Table C11. Together, these estimates reinforce the conclusion that the NHS primarily reduced the prevalence of sole proprietorships rather than employer-operated businesses.

are subject to a quasi-symmetric iceberg trade costs τ such that $\tau_{ln} = \tau_{nl}$ and $\tau_{ln} = 1$ if and only if $l = n$.

Households inelastically supply one unit of labor and are endowed with a level of managerial productivity ϕ that is used in the production of intermediates if a given household chooses self-employment. Following Behrens et al. (2014), I assume that productivity is drawn from a Pareto distribution. Before production begins, the self-employed must purchase ρ units of land at a price r. Households optimally choose between working for a competitive salary w or being self-employed based on their level of productivity ϕ and earning profits π. Given the fixed cost of production, regions exhibit an increasing return to scale, and each county produces an endogenous measure of varieties M equal to the number of self-employed households.

4.1.1 Preferences

Households are perfectly mobile and are indifferent between locations in the long-run. Conditional on residing in county l, utility is a function of total nominal income v_l, the aggregate price of tradeable goods P_l, and the price of land r_l:

$$\overline{U}_l = \frac{v_l}{P_l^{\zeta} r_l^{1-\zeta}} \tag{5}$$

where $\overline{U}_l$ represents the real earnings of a household in county l. Consumption of tradeables follows a CES aggregator, which is defined over varieties that are traded across all counties:

$$C_l = \left[\sum_{n \in \mathbb{N}} \int_{M_n} c_{ln}(j)^{\frac{\sigma-1}{\sigma}} dj \right]^{\frac{\sigma}{\sigma-1}} \tag{6}$$

where M_n is the measure of varieties produced in county n and σ represents the elasticity of substitution across both firms and counties. The local price of tradeable goods P_l follows the standard CES price index formula:

$$P_l = \left[\sum_{n \in \mathbb{N}} \left(\tau_{ln} p_n\right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}} \text{ s.t. } p_n = \left[\int_{M_n} p_n(j)^{1-\sigma} dj \right]^{\frac{1}{1-\sigma}} \tag{7}$$

which is a product of an index of factory-gate prices p_n from an origin county n and the trade costs associated with transporting goods across regions.

4.1.2 Production and Occupational Choice

Production of q units of variety j in county n combines managerial productivity ϕ and salaried labor input N such that:

$$q_n(j) = \phi_n(j)N_n(j) \tag{8}$$

I suppose a household's level of managerial productivity is drawn independently from a Pareto distribution:

$$G(\phi) = 1 - \left[\frac{\beta}{\phi}\right]^{\alpha} \tag{9}$$

where the parameters β and α denote the minimum value on the support of G and the degree of dispersion in the distribution, respectively. From this draw, households determine the optimal division of labor between salaried and self-employment. Lastly, I assume firms compete in a monopolistically competitive market such that the price of a given variety is the CES constant markup over marginal cost:

$$p_n(j) = \left[\frac{\sigma}{\sigma - 1}\right]\left[\frac{w_n}{\phi_n(j)}\right] \tag{10}$$

where w_n denotes the market wage for salaried employees.

Profits are non-zero and are a function of the degree of substitutability between goods and regions, dispersion in productivity, revenues, and the fixed cost of production:

$$\pi_n(\phi) = \left[\frac{1}{\sigma}\right]\left[\frac{\phi}{\Phi_n}\right]^{\sigma-1} Y_n - r_n\rho_n \text{ s.t. } \Phi_n = \left[\int_{M_n} \phi_n(j)^{\sigma-1} dj\right]^{\frac{1}{\sigma-1}} \tag{11}$$

where r_n denotes the price of land and ρ_n denotes the efficient units of land. Moreover, profits are increasing in managerial productivity, where an individual chooses salaried-employment when $\pi_n(\phi) < w_n$, self-employment when $\pi_n(\phi) > w_n$, and otherwise indifferent between either occupation. Based on the indifference point between self- and salaried-employment, the break-even level of productivity for active firms $\underline{\phi}_n$ determines the local self-employment rate E_l:

$$E_n = \left[1 + \left(\frac{\alpha(\sigma-1)}{\alpha - \sigma + 1}\right)\left(1 + \frac{r_n\rho_n}{w_n}\right)\right]^{-1} \text{ s.t. } E_n = 1 - G(\underline{\phi}_n) \tag{12}$$

where the value of ρ_l denotes the efficient units of land occupied by each firm. By extension, the number of varieties produced in a given county is equal to the number of self-employed households: $M_n = E_n L_n$.

4.1.3 Local Land Use

A county l is endowed with $\overline{H}_l$ units of land, where a fraction θ is designated for commercial use and the rest is enjoyed by households. Rents are paid to an absentee landlord and are rebated lump-sum back to households.

4.1.4 Market Clearing Conditions

Household income v_l is proportional to the sum of wages, profits, and expenditures on commercial land:

$$v_l L_l = \frac{\pi_l E_l L_l + w_l \left[1 - E_l\right] L_l + r_l \theta \overline{H}_l}{\zeta} \tag{13}$$

where π_l is defined in Appendix section D.1. The land market clears when land prices represent the unit value of residential property:

$$r_l = \left[\frac{1-\zeta}{1-\theta}\right] \left[\frac{v_l L_l}{\overline{H}_l}\right] \tag{14}$$

which states that price of land is increasing in income and population density. The labor market clearing condition is satisfied when the labor share of income to a constant fraction of total income adjusted by the local self-employment rates:

$$w_l = \left[\frac{\sigma-1}{\sigma}\right] \left[\frac{v_l}{1-E_l}\right] \tag{15}$$

The good market clearing condition is satisfied when total income from production is equal to total expenditures on tradeable goods:

$$v_l L_l = \sum_{k \in \mathbb{N}} s_{kl} v_k L_k \tag{16}$$

where s_{kl} are expenditure shares defined under the standard CES preference assumption.

4.2 General Equilibrium

In a spatial equilibrium, the transportation costs (τ_{ln}), location characteristics (ρ_l), and the amount of land available in a given county ($\overline{H}_l$) are taken as given. Given these exogenous variables, the general equilibrium of the model can be referenced by the following endogenous variables: (1) the local price of tradable goods and land ($P_l,\ r_l$), (2) the population residing in a given county (L_l), (3) the labor income for salaried employees and business income for the self-employed ($w_l,\ \pi_l$), (4) the share of expenditures for tradable goods across counties ($\{s_{ln}\}_{n\in\mathbb{N}}$), (5) the level of household income (v_l), and (6) the local self-employment rate (E_l). To close the model, the set of endogenous variables satisfy the following conditions in each county:

1. Households choose $\{s_{ln}\}_{n\in\mathbb{N}}$ and C_l to maximize utility
2. Firms choose p_l and N_l to maximize profits
3. Households choose E_l to satisfy the free entry condition
4. Households are indifferent to moving (i.e. $\overline{U}_l = \overline{U}_k \ \forall\ l, k \in \mathbb{N}$)
5. Land, labor, and goods markets clear

4.3 The Local Effect of Infrastructure Investment

The spatial equilibrium provides structure to connect changes in transportation costs from road construction to changes in the local self-employment rate. I follow Donaldson and Hornbeck (2016) and use a "market access" approach to determine this long-run effect. Overall, areas that experience an increase in market access will also experience an increase in population density and by extension a decline in self-employment.

4.3.1 Market Access and Residential Choice

Local population increases with respect to market access (denoted as MA), fraction of salaried employment, productivity, and county size. Local market access represents the ability of consumers and firms to reach other markets in the economy taking into account the cost of trade:

$$MA_l \propto \sum_{k\in\mathbb{N}} \tau_{lk}^{1-\sigma} MA_k^{\frac{\sigma}{1-\sigma}} L_k \tag{17}$$

which is proportional to the trade cost weighted market access to other counties of varying size in the economy. Using the structure of the model, local population is written as a log-linear function of market access and the fraction of salaried employment:

$$\ln L_l = \alpha + \beta \ \ln MA_l + \phi \ \ln[1 - E_l] + \eta_l + \epsilon_l \tag{18}$$

where α represents the change in aggregate utility, the fixed effect η_l represents the effect of county size, and the error term represents unobserved changes in local productivity. The equilibrium effect of changes in market access, holding utility and productivity constant, on population is:

$$\frac{\partial \ln L_l}{\partial \ln MA_l} = \left[1 + \phi \left(\frac{E_l}{1 - E_l}\right) \frac{\partial \ln E_l}{\partial \ln L_l}\right]^{-1} \beta \tag{19}$$

which accounts for both the direct change in market access and the indirect change in population due to endogeneous changes in salaried employment.

4.3.2 Population Growth and Self-Employment

To summarize the effect of changes in local population on self-employment, the model predicts that local self-employment declines as population rises because higher land prices raise the fixed cost of production, pushing less productive households out of self-employment. Formally, the elasticity of the local self-employment rate with respect to local population is given by:

$$\frac{\partial \ln E_l}{\partial \ln L_l} = -\left[\frac{r_l \rho_l}{w_l + r_l \rho_l}\right] \in (-1, 0), \tag{20}$$

where $r_l \rho_l$ denotes the fixed cost of production and w_l is the wage from salaried employment. The magnitude of this elasticity depends on how large fixed costs are relative to the total opportunity cost of being self-employed, equal to the sum of forgone wage income and fixed business costs.

Because land prices and wages are not observable at the beginning of my sample period, I simplify the model to connect it directly to available data. I assume that the efficiency units of land used in production are proportional to the share of commercial land across firms, such that $\rho_n = \theta \overline{H}_n / (E_n L_n)$. Under this assumption, the response of self-employment to population changes depends on the observed ratio of self- to salaried employment and a set of estimable parameters:

$$\frac{\partial \ln E_l}{\partial \ln L_l} = -\left[1 + \left(\frac{\sigma - 1}{\sigma}\right)\left(\frac{\zeta}{1-\zeta}\right)\left(\frac{1-\theta}{\theta}\right)\left(\frac{E_l}{1-E_l}\right)\right]^{-1}. \tag{21}$$

Here, σ governs the elasticity of substitution across varieties of goods, ζ is the expenditure share on tradable goods, and θ represents the share of commercial land. In my counterfactual analysis, I use these baseline parameter values from 1920 to compute the implied decline in local self-employment, holding the parcel of land per firm constant over time.

5 Aggregate Implications

In this section, I use the model predictions from Section 4.3 to quantify how much of the national decline in self-employment observed between 1920 and 1950 can be attributed to the construction of the NHS. To do so, I calculate the portion of the aggregate decline in self-employment that would arise solely from changes in market access due to the construction of the NHS, taking the 1920 US economy as given. This counterfactual estimate differs from the causal estimates in Section 3 because: (1) it captures how local market access responds not only to nearby construction but also to distant changes in the NHS network, (2) it provides a mechanism to explain how market access lowers self-employment through the spatial redistribution of the US population, and (3) it allows for aggregation to the national level.[31] Measuring this counterfactual change highlights the economic significance of the structural change set in motion by the NHS and provides a policy-relevant gauge of the aggregate role of infrastructure in reshaping labor markets.

5.1 Calibration of Parameter Values

Before performing any counterfactual, I must first calibrate the spatial equilibrium model to the 1920-1950 sample period, when the NHS was under construction. My calibration strategy is as follows. The model is characterized by (i) the transportation cost of traveling across counties over the NHS, (ii) the elasticity of substitution between firms and regions, (iii) the change in market access that comes from the construction of the NHS, (iv) the local effect of changes in market access on county-level population, and (v) the fraction of expenditures of housing and local land available for residential housing. In

[31] The closest reduced-form benchmark is Table 4, which shows that individual self-employment fell by 4.3 percentage points in counties connected to the NHS between 1920 and 1950, relative to contemporaneous trends.

each step, I calibrate the relevant set of parameters using external data. This methodological approach follows other spatial work, in which each have a historical empirical context, such as Donaldson and Hornbeck (2016), Hausman et al. (2019), and Fajgelbaum and Redding (2022).

Using structural equations from the model, I first estimate transportation costs and the elasticity of substitution. I next use these estimates to measure changes in market access due to the construction of the NHS, and then regress local changes in population on this measure of market access to estimate the corresponding elasticity. I use historical survey data to measure the fraction of expenditures on housing and the fraction of residential land.[32]

5.1.1 Relative Cost of Traveling on the NHS

Transportation costs in 1920 and 1950 are calculated using the least-cost route between origin and destination counties. As there were portions of the NHS that were paved prior to its construction, I measure the transportation costs in the base period accounting for the paved sections of the NHS as of 1926, when the NHS design was finalized and the first year with available digitized road conditions. Accounting for baseline differences in highway access as of 1926 will provide a more accurate estimate for the transportation costs in areas that were initially more developed and had already invested in a local highway infrastructure. Based on the least-cost route between counties, I express the transportation cost as follows:

$$\tau_{lkt} = \begin{cases} \left[\left(1 - \overline{\text{NHS}}_{lk}\right) + \kappa\, \overline{\text{NHS}}_{lk}\right] \times \text{Distance}_{lkt} & t = 1950 \\ \left[\left(1 - \overline{\text{Paved}}_{lk}^{1926}\right) + \kappa\, \overline{\text{Paved}}_{lk}^{1926}\right] \times \text{Distance}_{lkt} & t = 1920 \end{cases} \tag{22}$$

where the parameter κ represents the relative cost of traveling with an available paved road network, $\overline{\text{NHS}}_{lk}$ is the distance weighted average length of a route traveled on the NHS following its construction, $\overline{\text{Paved}}_{lk}^{1926}$ is the distanced weighted average length of a route traveled on a paved section of the NHS as of 1926, and Distance_{lkt} is the distance of the least cost path between counties l and k in year t.

To account for the relative cost of traveling over the NHS, I calculate the value of the parameter κ based on available data for the cost of a bus ticket versus train ticket between

[32] These surveys include the BLS Consumer Expenditure Survey (see Davis and Ortalo-Magné 2011) and a 1950 survey of residential versus commercial land use conducted by the American Society of Planning Officials (see Lovelace 1949)

the same origin and destination pair as of 1939 (Landon 1945).[33] The estimate provides a measure on the relative cost of transportation between the two most popular modes of long-distant transport during my sample period. This estimate is not the ideal as it does not provide a direct estimate for the average change in transportation costs. Instead, it reflects the cost-savings that automobiles provided relative to trains. Using data from Landon (1945), I set this parameter equal to 0.732, which is equal to the median relative price of a bus ticket to a train ticket for the small sample of origin and destination pairs. This parameter value corresponds to a 12 percent decline in the average transportation costs across origin-destination pairs. To account for the narrow scope of data used to calibrate this parameter, I additionally include counterfactual changes based on the interquartile range of relative prices.[34]

5.1.2 Elasticity of Substitution

To estimate the elasticity of substitution, I use the labor market clearing condition, which states that labor share of income is a constant fraction of total production income.[35] As noted in Gollin (2002), the labor share of income is difficult to measure when labor income is concentrated in many small firms or amongst the self-employed. Luckily, I calculate the labor share of income using data from the 1950 Decennial Census, which is the first year that asks respondents for information on labor, business, and other income.[36] This approach produces an estimate of 3.229 for the elasticity of substitution and corresponds to a labor share of income of 69 percent, which is within the range of estimates for the US based on calculations in Gollin (2002) (66-77 percent) and close to the estimate of 67 percent in Piketty and Zucman (2014). My estimation approach also differs from the international trade literature's gravity or CES demand estimations. However, my estimated elasticity lies within empirical ranges from Simonovska and Waugh (2014) (2.47–4.42), which supports the external validity of the estimates used in the model's quantification.

[33] I find that the NHS was over 90 percent complete by 1939, and these relative prices likely reflect the short-run pass-through of changes in highway access had during my sample period.

[34] For reference, the full sample of prices used to determine the value of κ are included in Table C13. The interquartile range of relative price of bus to train tickets is 0.682 and 0.813, and the average relative price is 0.749. With respect to the average decline in transportation costs, it magnitude is similar to Morten and Oliveira (2024), which looks at the effects of highway construction in Brazil on trade and migration.

[35] In particular, the elasticity of substitution is a non-linear function of the labor share of income, where the CES markup (i.e. $(\sigma - 1)/\sigma$) determines this share. I use this model relationship to estimate this particular parameter.

[36] Each category of income is observed separately in the data, and each are top coded at $10,000 in 1950 dollars, which corresponds to approximately $140,000 in 2025 dollars. I consider total income as the sum of these three sources of income.

5.1.3 Estimating Changes in Market Access

I define county-level market access based on the population weighted average transportation costs between origin and destination locations, and measure its impact on long-run changes in local population based on equation (18). I use the following first-order approximation to define local market access with the estimated value of transportation costs, elasticity of substitution, and 1920 population counts as inputs:

$$MA_{lt} \approx \sum_{k \neq l} \left[\frac{\lambda_{k,1920}}{1 - \lambda_{l,1920}} \right] \tau_{lkt}^{1-\sigma} \quad \text{s.t. } \lambda_{k,1920} = \frac{L_{k,1920}}{\sum_{k \in \mathbb{N}} L_{k,1920}} \tag{23}$$

which equals the population-weighted average trade cost present in delivered prices.[37] Using only 1920 population counts allows for market access to vary over time only as a result of transportation costs between counties as opposed to changes in local population that responded endogenously to the construction of the NHS. Figure 5 plots the estimated change in market access due to the construction of the NHS. While the NHS was intended to link all 48 states with a paved highway network, areas with paved highway access as of 1926 were concentrated in the Northeast and along the West Coast. By accounting for these previously paved sections of the NHS when calculating transportation costs, market access was expected to increase the most in the Southeast, Midwest, and Western Rocky Mountains.

With a county-level measure of market access, I next estimate the causal effect of changes in market access on changes in local population. I use the following regression specification to estimate this effect between 1910 and 1950 for a balanced panel of counties:

$$\ln L_{lt} = \alpha_l + \rho_t + \beta \, \Delta \ln MA_{lt} \times \mathbb{1}(t > 1920) + X_{lt}' \Omega + \epsilon_{lt} \tag{24}$$

where α_l and ρ_t are county and year fixed effects. To identify the causal effect of market access, I construct an instrument equal to the degree of county-level market access based the layout of the Pershing Map.[38] To account for the systematic inclusion in the the Pershing Map, I include the expected exposure to the hypothetical network in the Pershing Map, based on neighboring counties as defined in equation (2), as a control variable.

[37] Note that population shares are normalized to exclude county l as the average is calculated for all other counties in the country.

[38] To construct this instrument, I measure transportation costs between US counties using the hypothetical highway network in the Pershing Map and estimated value of κ. I calculate county-level market access according to these transportation costs.

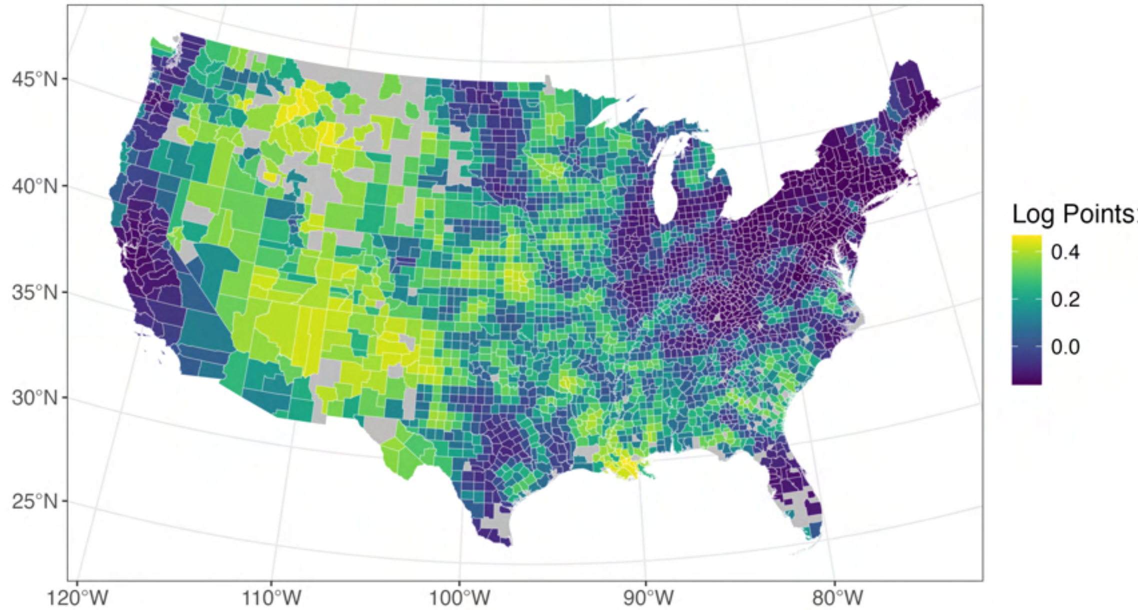


Figure 5: Changes in Market Access with the NHS

Notes: This figure plots the change in market access following the construction of the NHS. The change in market access is normalized by the county average and accounts for the paved sections of the NHS as of 1926.

Other control variables include the distance to the nearest military base, a binary indicator for a local military base, a binary indicator for local railway access, average farm land value as of 1920, and average manufacturing value added as of 1920. These additional controls account for potential confounding factors that in part determined the layout of the Pershing Map and are interacted with a time-dummy for the construction phase of the NHS to allow for time-varying effects. Lastly, standard errors are double-clustered at the county and the state-by-year level.

Based on the definition for market access in equation (23) and the estimated parameter values, the change in market access caused an positive increase local population following the construction of the NHS. Figure 6 displays the coefficient estimates from the regression specification in equation (24). It is important to note that market access only began impacting local population after the construction of the NHS began following 1920.[39] For robustness, Table C14 shows the estimated effect of changes in market access is robust to additional geographic controls including state fixed effects, the relative location of major

[39]Relative to the estimates from Table 3, the effect of market access on county-level population between 1930 and 1940 are larger in magnitude and have a greater statistical significance. Note that between these years, the construction of the NHS was between 60 and 90 percent complete. Because the completion rate was so high, the local and distant effects included in the measure of market access would be more significant in these early years of the sample period.

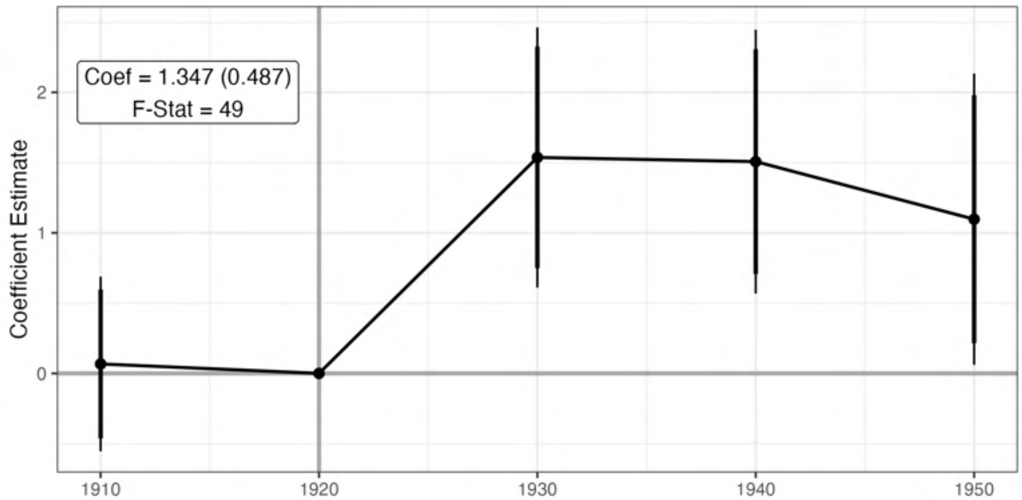


Figure 6: Local Effects of Market Access on Population Growth, 1910-1950

Note: This figure displays the 2SLS coefficient estimates from equation (24) and an associated event study. For both regressions, I use an instrument equal to the measure of market access, as defined in 5.1.3, using the transportation costs calculated based on the layout of the Pershing Map. I include the F statistic for the change in market access for the regression specification from equation (24).

metro areas, controls for latitude and longitude, and the fraction of undevelopable land.[40] Additionally, Figure C8 displays the coefficient estimates when calculating market access with updated population shares. When population shares change endogenously with changes in transportation costs, the estimated effect of market access remains positive and statistically significant but decreases in magnitude.

Relative to equation (18), the present specification excludes the contemporaneous changes in the fraction of salaried employment as a control variable. The model predicts that local population and share of salaried employment are positively correlated, consistent with the local increase in productivity when self-employment declines. However, these contemporaneous changes occur primarily as a downstream consequence of the initial population increase triggered by improved highway access and require a separate instrument.[41] Omitting this channel means that the estimated effect of market access on population and the resulting counterfactual on the true aggregate impact should be

[40] All additional control variables are interacted with period fixed effects and allow for time-varying effects. Controls for the fraction of undevelopable land include the fraction of a county that is covered in water or has above a 15 percent incline in elevation and correspond to baseline controls include in the regression specification in equation (1).

[41] Including the change in salaried employment as a control variable without an instrument does not impact the direction, magnitude, nor the statistical significance of the estimates included in Figure 6. Developing an instrument that provides exogenous variation and that is uncorrelated with the instrument for market access is potential future work.

interpreted as a conservative lower bound.

5.1.4 Remaining Parameters Values and Changes in Self-Employment

The remaining parameters needed to estimate the counterfactual change in self-employment are based on the assumption that self-employed households must purchase a parcel of land to begin production. These parameters include the fraction of expenditures spent on housing $(1-\zeta)$ and the fraction of residential land $(1-\theta)$.[42]

For both parameters, I use historical survey data to estimate their values. Using the Consumer Expenditure Survey available between 1918 and 1950, I set the expenditure shares for housing to 0.275, which is consistent with the estimate of 0.25 from Davis and Ortalo-Magné (2011). To calculate this value, I take the simple average of nationally representative expenditure share on housing from three rounds of consumer expenditure survey from the early 20th century: 0.233 from the 1918-1919 survey, 0.320 for the 1934-1936 survey, and 0.272 from the 1950 survey. Lastly, I set the fraction of residential land equal to 0.270 based on survey estimates from Lovelace (1949). This particular survey was conducted by the American Society of Planning Officials to measure current land-use for locations across the US and parts of Canada as of 1950. I calculate the geo-mean of residential land use for the 38 US locations to estimate this parameter value.[43]

Taken together, the counterfactual change in local self-employment depends on the effect of changes in market access on local population, the initial county-level self-employment rate, and estimated parameter values:

$$\widehat{\Delta \ln E_l} = -\left[1 + \underbrace{\left(\frac{\sigma-1}{\sigma}\right)}_{0.690}\underbrace{\left(\frac{\zeta}{1-\zeta}\right)}_{2.636}\underbrace{\left(\frac{1-\theta}{\theta}\right)}_{0.379}\left(\frac{E_{l,1920}}{1-E_{l,1920}}\right)\right]^{-1}\underbrace{\beta}_{1.347}\,\Delta \ln MA_l \tag{25}$$

Based on the estimates for the fraction of expenditures spent on housing $(1-\zeta)$ and the fraction of residential land $(1-\theta)$, these two parameters cancel one another out when calculating the counterfactual change in self-employment. As a result, the counterfactual change relies on an accurate measure of the labor share of income and the local effect of market access, and the assumption that the fixed cost is denominated in efficiency units of land plays a limited role.

[42]These parameters have an offsetting effect with respect to the counterfactual change based on equation (21). If households were to spend more on residential land, increasing the size of $(1-\zeta)$ and by extension land prices, then the number of households exiting self-employment would increase under these harsher conditions. The opposite is true when there is relatively more residential land with a higher value of $(1-\theta)$.

[43]The share of residential land use for the 38 US locations in Lovelace (1949) are included in Table C15.

5.2 Quantifying the Aggregate Implications of the NHS

To estimate the aggregate change in self-employment resulting from the construction of the NHS, I use the elasticities for changes in population and self-employment with respect to market access to construct a first-order approximation of county-level changes in population weights and local self-employment rates implied solely by improvements in market access derived from the structural model in Section 4. I then aggregate these county-level changes to obtain a counterfactual national self-employment rate.

Before presenting the aggregate result, I first describe the geographic variation in the predicted county-level changes. Figure C9 plots a binned scatterplot of county population shares and local self-employment rates in 1950 against their 1920 values. The model prediction captures, for example, the large population growth of Chicago, IL during this period. Other changes are smaller in magnitude, but the predicted and observed 1950 population shares closely track each other. The model also reproduces a substantial fraction of the observed cross-sectional decline in self-employment, although the counterfactual distribution does not perfectly align with the data, implying that the NHS explains only part of the aggregate change. This outcome is to be expected as the effects of the NHS were largely orthogonal to sectoral changes such as the rise of manufacturing. Overall, the model successfully captures the geographic variation underlying the national decline in self-employment.[44]

Having established that the model reproduces the spatial pattern of change, I next turn to the aggregate effect with Table 7. First, between 1920 and 1950, the aggregate self-employment rate declined by 9.2 percentage points.[45] The counterfactual exercise however predicts that the construction of the NHS accounted for a 1.7 percentage-point decline, or roughly one-fifth of the observed aggregate change. In levels, this corresponds to approximately 170,000 additional wage employees in the 1950 labor force.[46] Because this estimate depends on the parameter κ, which is based on a limited sample of relative prices, I also report bounds using the interquartile range of transportation costs. The implied effect varies between one-tenth and one-quarter of the observed decline. Overall,

[44] Figure C10 maps the geographic variation in the predicted and observed distributions of county-level population shares and self-employment rates in 1950. The spatial pattern of the model prediction mirrors the observed distribution, confirming that the model captures a meaningful fraction of the aggregate decline in self-employment attributable to the NHS.

[45] This value differs slightly from the population-weighted county-level rate, which declined by 11%. Because county self-employment rates are identical across both calculations, the discrepancy arises from the weighting scheme. To maintain consistency, I compute counterfactual weights using the 1920 baseline population of prime-working-age men and the predicted change in population.

[46] The estimated number of reallocated workers is calculated by multiplying the 1.7 percentage-point decline by the 1950 number of self-employed persons from FRED Series `A4501C0A173NBEA`.

Table 7: Counterfactual Change in Aggregate Self-Employment Rate, 1920–1950

	Percentage Point Change	Fraction Explained
Counterfactual Estimate	-0.017	0.189
	$[-0.022, -0.011]$	[0.112, 0.240]
Observed Decline in Self-Employment	-0.092	1.000

Note: This table reports the counterfactual change in the aggregate self-employment rate predicted by the change in market access due to the construction of the NHS. Bracketed estimates correspond to alternative values of κ set to the 25th and 75th percentiles of the relative price of travel on the NHS (see Table C13). These values are 0.682 and 0.813, respectively.

the construction of the NHS had an economically significant effect on structural change in the US labor market.

Comparing this counterfactual to the causal estimates in Section 3, the aggregate decline is smaller than the roughly 4.3 percentage-point decline presented in Table 4. This difference is expected for two reasons. First, the regression estimate measures the relative change between counties with and without highway access, without accounting for the relative size of treated and control groups. Second, the regression does not capture indirect effects of highway construction in distant counties arising from trade linkages. The market access approach of Donaldson and Hornbeck (2016) addresses this latter issues by providing a continuous measure of treatment intensity across space. Accounting for these channels naturally yields a smaller but more comprehensive estimate of the aggregate effect.

6 Conclusion

This paper proposes a novel mechanism linking infrastructure investment to the aggregate decline in self-employment during the early 20th century. I show that the construction of the first federally funded highways in the US strengthened agglomeration forces by lowering transportation costs, raising local land prices, and increasing the fixed cost of operating small-scale businesses. Using a new instrumental variable based on the planned but unrealized Pershing Map, I provide causal evidence that counties with highway access experienced large and persistent population gains, while individuals in these subsequently denser, highway-connected counties were more likely to exit self-employment and take up wage jobs.

To quantify the aggregate implications of this mechanism, I design a spatial equilibrium model with endogenous occupational choice, where the fixed costs of production rise with population growth. The model predicts that the construction of the Numbered Highway System accounts for roughly one-fifth of the observed decline in self-employment between 1920 and 1950, equivalent to about 170,000 additional wage employees. These results reveal that large-scale infrastructure investment not only shaped the spatial distribution of population but also accelerated the transition from self-employment to salaried employment, which is a key dimension of modern economic development.

While the analysis focuses on a historical episode, the mechanism uncovered here has broader relevance. The same forces that linked transportation access to occupational reallocation in the early 20th century continue to operate today as governments expand both physical and digital infrastructure. In developing economies, new road networks and urban transportation projects may similarly influence the composition of the formal versus informal labor by altering the overhead cost of an incorporated business. In high-income countries, emerging forms of infrastructure, such as broadband access, can affect entrepreneurship through parallel channels, changing the nature rather than the magnitude of these costs.

Looking ahead, future work will explore two complementary channels through which infrastructure investment shaped economic development during this period. First, I will examine whether the construction of the NHS induced a crowding-in effect, encouraging additional investment by state and local governments in complementary road networks. Although the federal program was designed to provide through roads for interstate travel, its success depended on local governments building connecting roads that expanded increases in market access to smaller towns. Second, I plan to study whether improved infrastructure accelerated technological adoption, particularly of automobiles. At the time, cars and trucks were in high demand but often underutilized in regions with unreliable roads; reducing the uncertainty related to reliable automobile travel may have spurred both vehicle ownership and complementary investments in maintenance and services. Together, these projects aim to broaden our understanding of how public infrastructure programs catalyze further investment and technology diffusion.

References

Abraham, K. G., Haltiwanger, J. C., Hou, C., Sandusky, K., and Spletzer, J. (2024). *Driving the Gig Economy*. US Census Bureau, Center for Economic Studies.

Adamopoulos, T. (2011). Transportation Costs, Agricultural Productivity, and Cross-Country Income Differences. *International Economic Review*, 52(2):489–521.

Allen, T. and Arkolakis, C. (2014). Trade and the Topography of the Spatial Economy. *The Quarterly Journal of Economics*, 129(3):1085–1140.

Aschauer, D. A. (1989). Is public expenditure productive? *Journal of monetary economics*, 23(2):177–200.

Atack, J., Bateman, F., Haines, M., and Margo, R. A. (2010). Did Railroads Induce or Follow Economic Growth?: Urbanization and Population Growth in the American Midwest, 1850–1860. *Social Science History*, 34(2):171–197.

Baum-Snow, N. (2007). Did Highways Cause Suburbanization? *The Quarterly Journal of Economics*, 122(2):775–805.

Behrens, K., Duranton, G., and Robert-Nicoud, F. (2014). Productive Cities: Sorting, Selection, and Agglomeration. *Journal of Political Economy*, 122(3):507–553.

Bick, A., Fuchs-Schündeln, N., Lagakos, D., and Tsujiyama, H. (2022). Structural Change in Labor Supply and Cross-Country Differences in Hours Worked. *Journal of Monetary Economics*, 130:68–85.

Boone, C. D. and Wilse-Samson, L. (2023). Structural Change and Internal Labor Migration: Evidence from the Great Depression. *Review of Economics and Statistics*, 105(4):962–981.

Borusyak, K. and Hull, P. (2023). Nonrandom Exposure to Exogenous Shocks. *Econometrica*, 91(6):2155–2185.

Brooks, W. and Donovan, K. (2020). Eliminating Uncertainty in Market Access: The Impact of New Bridges in Rural Nicaragua. *Econometrica*, 88(5):1965–1997.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica*, 82(6):2295–2326.

Cooper, R. and Haltiwanger, J. (1993). Automobiles and the National Industrial Recovery Act: Evidence on Industry Complementarities. *The Quarterly Journal of Economics*, 108(4):1043–1071.

Davis, M. A. and Ortalo-Magné, F. (2011). Household Expenditures, Wages, Rents. *Review of Economic Dynamics*, 14(2):248–261.

Decker, R., Haltiwanger, J., Jarmin, R., and Miranda, J. (2014). The Role of Entrepreneurship in US Job Creation and Economic Dynamism. *Journal of Economic Perspectives*, 28(3):3–24.

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

Duranton, G. and Turner, M. A. (2012). Urban Growth and Transportation. *Review of Economic Studies*, 79(4):1407–1440.

Eckert, F. and Peters, M. (2022). Spatial Structural Change. Technical report, National Bureau of Economic Research.

Eli, S., Hausman, J. K., and Rhode, P. W. (2025). The Model T. *The Journal of Economic History*, 85(1):110–151.

Faber, B. and Gaubert, C. (2019). Tourism and Economic Development: Evidence from Mexico's Coastline. *American Economic Review*, 109(6):2245–2293.

Fabricant, S. (1940). Manufacturing output, 1929-1937. In *Manufacturing Output, 1929-1937*, pages 1–32. NBER.

Fairlie, R. W. and Meyer, B. D. (2000). Trends in Self-Employment among White and Black Men during the Twentieth Century. *Journal of Human Resources*, pages 643–669.

Fajgelbaum, P. and Redding, S. J. (2022). Trade, Structural Transformation, and Development: Evidence from Argentina 1869–1914. *Journal of Political Economy*, 130(5):1249–1318.

FDIC (1992). Federal Deposit Insurance Corporation Data on Banks in the United States, 1920–1936. Version V1, distributed 1992-02-16.

Fishback, P. and Kollmann, T. (2014). New Multicity Estimates of the Changes in Home Values, 1920-1940. In *Housing and Mortgage Markets in Historical Perspective*, pages 203–244. University of Chicago Press.

Frye, D. (2024). Transportation Networks and the Geographic Concentration of Employment. *Review of Economics and Statistics*, pages 1–34.

Gelman, A. and Imbens, G. (2019). Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs. *Journal of Business & Economic Statistics*, 37(3):447–456.

Gollin, D. (2002). Getting Income Shares Right. *Journal of Political Economy*, 110(2):458–474.

Gollin, D. (2008). Nobody's Business but My Own: Self-Employment and Small Enterprise in Economic Development. *Journal of Monetary Economics*, 55(2):219–233.

Guida-Johnson, N. (2022). Informality and Development: The Role of Transportation Infrastructure in Reducing Self-Employment. *Mimeograph, Boston University*.

Hausman, J. K., Rhode, P. W., and Wieland, J. F. (2019). Recovery from the great depression: The farm channel in spring 1933. *American Economic Review*, 109(2):427–472.

Herreño, J. and Ocampo, S. (2023). The Macroeconomic Consequences of Subsistence Self-Employment. *Journal of Monetary Economics*, 136:91–106.

Hornbeck, R. and Rotemberg, M. (2024). Growth off the Rails: Aggregate Productivity Growth in Distorted Economies. *Journal of Political Economy*, 132(11):3547–3602.

Imbens, G. and Kalyanaraman, K. (2012). Optimal Bandwidth Choice for the Regression Discontinuity Estimator. *The Review of Economic Studies*, 79(3):933–959.

Khor, N. J. K. (2022). *World Cities Report 2022: Envisaging the Future of Cities*. United Nations Human Settlements Programme (UN-Habitat).

Kuznets, S. and Murphy, J. T. (1966). *Modern Economic Growth: Rate, Structure, and Spread*, volume 2. Yale University Press New Haven.

La Porta, R. and Shleifer, A. (2014). Informality and Development. *Journal of Economic Perspectives*, 28(3):109–126.

Landon, E. C. (1945). The National Traffic Pattern. `https://catalog.hathitrust.org/Record/000967871/Home`. [Online; accessed 01-August-2025].

Levine, R. and Rubinstein, Y. (2017). Smart and Illicit: Who Becomes an Entrepreneur and Do They Earn More? *The Quarterly Journal of Economics*, 132(2):963–1018.

Lovelace, E. (1949). Urban Land Use—1949. *Journal of the American Institute of Planners*, 15(2):26–32.

Manson, S., Schroeder, J., Van Riper, D., Knowles, K., Kugler, T., Roberts, F., and Ruggles, S. (2024). Ipums USA: Version 19.0 [dataset]. *Minneapolis, MN: IPUMS, 2024.*

Melitz, M. J. and Trefler, D. (2012). Gains from trade when firms matter. *Journal of Economic Perspectives*, 26(2):91–118.

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

Michaels, G., Rauch, F., and Redding, S. J. (2012). Urbanization and Structural Transformation. *The Quarterly Journal of Economics*, 127(2):535–586.

Michaels, G., Rauch, F., and Redding, S. J. (2019). Task Specialization in US Cities from 1880 to 2000. *Journal of the European Economic Association*, 17(3):754–798.

Morin, M. and Swisher, S. N. (2016). The Effect of Infrastructure Investment in a Low-Growth Environment: Evidence from the Great Depression. Technical report, Working Paper, University of Cambridge.

Morten, M. and Oliveira, J. (2024). The effects of roads on trade and migration: Evidence from a planned capital city. *American Economic Journal: Applied Economics*, 16(2):389–421.

Perra, E., Sanfilippo, M., and Sundaram, A. (2024). Roads, Competition, and the Informal Sector. *Journal of Development Economics*, 171:103339.

Piketty, T. and Zucman, G. (2014). Capital is back: Wealth-income ratios in rich countries 1700–2010. *The Quarterly journal of economics*, 129(3):1255–1310.

Redding, S. J. (2016). Goods Trade, Factor Mobility and Welfare. *Journal of International Economics*, 101:148–167.

Ruggles, S., Nelson, M. A., Sobek, M., Fitch, C. A., Goeken, R., Hacker, J. D., Roberts, E., and Warren, J. R. (2024). IPUMS Acenstry Full Count Data: Version 4.0 [dataset]. *Minneapolis, MN: IPUMS, 2024.*

Saiz, A. (2010). The Geographic Determinants of Housing Supply. *The Quarterly Journal of Economics*, 125(3):1253–1296.

Simonovska, I. and Waugh, M. E. (2014). The elasticity of trade: Estimates and evidence. *Journal of international Economics*, 92(1):34–50.

Swift, E. (2011). *The Big Roads: The Untold Story of the Engineers, Visionaries, and Trailblazers who Created the American Superhighways*. HMH.

Wallis, J. J. (1987). Employment, politics, and economic recovery during the great depression. *The Review of Economics and Statistics*, pages 516–520.

Walsh, C. (2023). The Startup Multiplier. *Available at SSRN*.

A Data Appendix

Construction of NHS Network: I digitize state road atlases to track the progression of the NHS construction from 1926 to 1939. Atlases include:

- 1926 Rand McNally Auto Road Atlas of the United States
- 1931 Rand McNally Special Auto Road Atlas of the United States and Eastern Canada
- 1939 Rand McNally Auto Road Atlas of the United States, Canada, and Mexico

NHS network is:

- United States System of Highways : Adopted for Uniform Marking by the American Association of State Highway Officials, November 11, 1926.

The link for state atlases is here, the link for NHS map is available here, and the link for the Pershing map is available here. I initially digitized the NHS map to determine the placement for the planned highway lines. I used the 1926-1939 state atlases to track the completion of each highway line and record the lengths of highway that are paved in each state. Road conditions are identified as:

- Unimproved, improved, or paved in 1926
- Under construction, dirt, gravel, graded, improved, and paved in 1931
- Under construction, dirt, graded, improved, and paved in 1939

Because road conditions are recorded in slightly different fashions across atlases, I focus on completely paved sections of road in my analysis. I also record the entire length of highways lines regardless of road conditions according to the 1939 atlas to produce a "completed" version of the NHS map as well. To produce a dataset with a uniform set of roads across years, I remove portions of the "completed" map that are not paved in a given year. There are lines in the NHS Map that appear to have been rerouted in its constructed or not constructed. For this reason, I note the impacted areas:

- Area between Phoenix and Tuscon, AZ is rearranged to accommodate rerouted lanes
- Area between Albany and New York, NY is consolidated to a single road

- Area between Dunnellon and Ocala, FL is moved north to High Springs, FL to accommodate rerouted lanes
- Area between Shoshone and Challis, ID is moved west to accommodate rerouted lane
- Area between Oakaloosa and Des Moines, IO is rerouted north through Malcolm, IO
- Area between Marquette, IO and Viroqua, WI is rerouted east
- Area between Camdenton and Nevada, MO is moved north through Weaubleau, MO
- Area between Albany and Newport, OR is not constructured
- Area southwest of Klamath Falls, OR is moved south
- Lanes east of Seattle are rerouted and appear closer than they are drawn in the original NHS map

Instruction across Census Years to Define Self-Employment:

1910. Whether employer, employee, or working on own account. For one employing persons, other than domestic servants, in transacting his own business, write "Emp" (for employer). For a person who works for wages or a salary, write "W" (for wage earner). For a gainful worker who is neither an employer nor an employee, write "OA" (for own account). For all persons returned as having no occupation, leave the column blank.

- Employer: An employee is one who employs helpers, other than domestic servants, in transacting his own business. The term employer does not include the superintendent, agent, manager, or other person employed to manage an establishment or business, and it does not include the foreman of a room, the boss of a gang, or the coal miner who hires his helper. All such should be returned as employees, for, while any one of these may employ persons, none of them does so in transacting his own business. Thus no individual working for a corporation either as an officer or otherwise should be returned as an employer. A person employing domestic servants in his own home but not employing any helpers in his business should not be returned as an employer. But, on the other hand, a persons who is the proprietor of a hotel or boarding or lodging house and employs servants in running that hotel or boarding or lodging house should be returned as an employer, because he employs these servants in his business.

- Employee: Any person who works for wages or a salary and is subject to the control and direction of an employer, is an employee, whether he be president of a large corporation or only a day laborer, whether he be paid in money or in kind, and whether he be employed by his own parent or by another. The term employee does not include lawyers, doctors, and others who render professional service for fees, and who, in their work, are not subject to the control and direction of those whom they serve. It does include actors, professors, and others who are engaged to render professional service for wages or salaries. A domestic servant should always be returned as an employee even though, as previously explained, the person employing a domestic servant is not always returned as an employer.
- Working on own account: Persons who have a gainful occupation and are neither employers are considered to be working on their own account. They are the independent workers. They neither pay nor receive salaries or regular wages. Examples of this class are: Farmers and the owners of small establishments who do not employ helpers; professional men who work for fees and employ no helpers; and, generally speaking, hucksters, peddlers, newsboys, bootblacks, etc., although it not infrequently happens that persons in these pursuits are employed by others and are working for wages, and in such case should, of course, be returned as employees.

1930. For an employer - that is, one who employs helpers other than domestic servants in transacting his own business - write in column 27 "E"; for a wage or salary worker write "W"; for a person working on his own account write "O"; for an unpaid family worker - that is, a member of the family employed without pay on work which contributes to the family income - write "NP". For all persons returned as having no gainful occupation, leave column 27 blank.

- Employer ("E") - An employer is one who employs helpers, other than domestic servants, in transacting his own business. The term "employer" does not include the superintendent, agent, manager, or other person employed to manage an establishment or business; and it does not include the foreman of a room, the boss of a gang, or the coal miner who hires his helper. All such should be returned as wage or salary workers, for, while any one of these may employ persons, none of them does so in transacting his own business. In short, no person who himself works for wages or a salary is to be returned as an employer.

- Wage or salary worker ("W") - Any person who works for wages or salary, at piece rates, or on commission, and is subject to the control and direction of an employer, is to be considered a wage or salary worker. This classification will include the president of the bank or the manager of the factory as well as the clerks and the laborers who may be also employed by the bank or the factory.

- Working on own account ("O") - A person who has a gainful occupation and is neither an employer, nor a wage or salary worker, nor an unpaid family worker, is considered to be working on his own account; such persons are the independent workers. They neither pay nor receive salaries or regular wages. Examples of this class are: Farmers and the owners of small establishments who do not employ helpers; professional men who work for fees and employ no helpers; and generally speaking, hucksters, peddlers, newsboys, bootblacks, etc.

- Unpaid family worker ("NP") - A wife, son, daughter, or other relative of the head of the family who works regularly and without wages or salary on the family's farm, in a shop or store from which the family obtains its support, or on other work that contributes to the family's income (not including housework or incidental chores) is to be returned as an unpaid family worker. Examples are: A son working regularly and without wages on his father's farm; a wife working regularly without salary in her husband's store or office; a girl assisting her mother regularly without wages on sewing done in the home for a clothing factory.

1940. For each persons for whom an occupation is entered in col. 28, enter in col. 30 a symbol for class of worker applying to that occupation as indicated below:

- For a wage or salary worker in private work - PW

- For a wage or salary worker in Government work (including public emergency work) - GW

- For an employer, that is, a persons who employed one or more workers other than unpaid family workers (or domestic servants) in conducting his own business - E

- For a person who worked in his own account - OA

- For an unpaid family worker, that is, a related member of the family who was employed without wages or salary on work (other than housework or incidental chores) that contributed to the family income - NP

- For a person who followed more than one class of work in his occupation, enter the symbol for that class of work at which he worked the longest during the week of March 24-30. For example, for a carpenter who worked as an employee two days of the week and on his own account without employees for three days of the week, enter "OA" for "own account."

1950. There must be an entry of one of the four codes shown in the heading of item 20c for each person with occupation and industry entries. You frequently will not have to ask a specific question before making the entry because the correct answer will be obvious from the preceding conversation. If you have any doubts, however, ask for class of worker specifically. The class-of-worker code should refer to the same job or business as the occupation and industry entries for the person. Definition of class-of-worker codes:

- P - Work for a PRIVATE employer for wages, salary, commission tips, piece-rates or pay in kind; this applies regardless of the occupation at which the employee worked, whether general manager, file clerk, or porter. It includes veterans working for a private employer and receiving Federal GI subsistence payments. It includes also persons working for settlement houses, churches, unions, and other private nonprofit organizations.

- G - Work for any branch of GOVERNMENT Federal, State, city, county, etc.; this includes public schools and government-owned bus lines, government-owned electric power companies, etc. It includes persons who were elected to paid offices and civilian employees of the armed forces. Enter "G" also for employees of international organizations such as United Nations and for employees of foreign governments such as persons employed by the British Embassy or by the French Purchasing Commission; this rule applies only to those persons already listed in accordance with the instructions on whom to enumerate. Persons employed by such private organizations as the American Red Cross and the U S. Chamber of Commerce are not government employees and should be reported as "P."

- O - Work for profit or fees in OWN business, farm, shop, office, etc.; this does not include superintendents, foremen, managers, or other executives hired to manage a business or farm, salesmen working for commission, or officers of corporations.

- NP - Work WITHOUT PAY on a farm or business operated by a member of the household to whom the person is related. Note that room and board and a cash allowance are not counted as pay for these family workers; however, if the worker receives money which is definitely considered to be wages for work performed, he should be reported as "P."

Special points on class-of-worker code.-The following are special points which may be useful in certain problem cases:

- Corporation employees - All employees of an incorporated business, regardless of the particular occupation at which they work) should be reported as "P" (or, in some few cases, "G"). They are not to be reported as "O" even though they own part or all of the stock of the incorporated business.
- Domestic work in other persons' homes - This should be reported as "P" for example, "Maid, Private family, P."
- Partnerships - Persons who operate a business in partnership with one or more people should be reported as "in OWN business." The word "OWN" is not limited to single ownership.
- Work far pay in kind - Pay in kind includes room, board, supplies, and food, such as eggs or poultry on a farm. This is considered pay except in the case of the unpaid family worker.
- Work on an odd-job or casual basis - This should be reported as "P."
- Clergymen - Preachers, ministers, priests, rabbis, and other clergymen are to be reported as "P" in class of worker, except in the following two cases:
  - Enter "G" for a clergyman, such as a prison chaplain, working in a civilian government job;
  - Enter "O" for a clergyman who is not attached to one particular church or congregation but who conducts religious services in various places on a fee basis.

B Highway Access and Automobile Utilization

In this Appendix section, I show how the utilization of automobiles changes after portions of the NHS are completed. Given the high demand for automobiles during the early 20th century as documented in Eli et al. (2025), individuals were likely restrained in their adoption of cars or trucks based on the quality of local roads and would increase utilization once nearby roads were fully paved.[47]

B.1 Identification Strategy

To test this hypothesis, I combine county-level automobile registration data with interim breaks in highway construction to measure the change in the number of newly registered automobiles per capita when counties only just gain highway access. I use the following regression specification to measure this effect controlling for a county's distance to the nearest completed portion of the NHS:

$$\begin{aligned} \ln \text{Regs}_c^{1932} = \alpha + \beta\, \mathbb{1}\left[c \in \text{Highway Access}^{1931}\right] + f\left(\text{distance}_c\right) \\ + \mathbb{1}\left[c \in \text{Highway Access}^{1931}\right] \times f\left(\text{distance}_c\right) + X_c'\Omega + e_c \end{aligned} \tag{B1}$$

where the running variable measures the minimum straight-line distance to the nearest break in NHS construction as of 1931, the vector X_c controls for latitude, longitude, and their interaction to compare counties that are nearby one another, and the error term e_c is clustered at the state level.[48] Taken together, the coefficient β measures the percent change in new automobile registrations per capita around the most recently completed portions of the NHS.

This test follows a regression-discontinuity design where the coefficient β measures the effect of marginally gaining highway access if the highway's completion status is arguably exogenous to a given county near the discontinuity. The identifying assumption behind this design is that baseline characteristics are continuous across breaks in construction, that is, counties only differ in terms of the status of highway construction as of

[47] For example, dirt roads that predated the NHS would become impassable mud after a heavy rain and were regularly referred to as "gumbo" in Oklahoma given its thick brown consistency (Swift 2011). Brooks and Donovan (2020) find that eliminating uncertainty in market access increases the effects of lowering transportation costs. In the context of this study, a completed section of the NHS is paved and includes drainage to reduce erosion over time. In doing so, a paved highway lowers the uncertainty around the use of automobiles.

[48] Because the break in construction occurs within a given county, I drop this set of counties from the sample for robustness. In Figure C3, I include the average number of new automobile registrations for this set of counties, which is in line with the nearby counties that have recently gained highway access.

1931.[49] While I argue in the next section that the placement of the NHS is endogenous to local economic development, I focus on changes in automobile utilization amongst counties that are set to receive access to the NHS, and use only variation in construction within this set of counties to estimate the coefficient β. So while there may be characteristics that determine the placement of the NHS, I only require that these characteristics do not change discretely around a break in construction to interpret the coefficient estimate as causal. Because I estimate the coefficient of interest based on the set of counties that will be connected according to the finalized NHS network, I interpret these coefficients as an intent-to-treat effect.

For implementation, I use a linear polynomial fit in equation (B1) following Gelman and Imbens (2019). I present primary coefficient estimates using both a fixed bandwidth of 120 kilometers (approximately 75 miles) and an optimal bandwidth based on Imbens and Kalyanaraman (2012).[50] I use a uniform kernel to estimate the percent change in automobile registrations around breaks in NHS construction. The main results also remain similar in sign and magnitude across alternative specification choices based on polynomial fit, bandwidth selection, and kernel weight, as shown in Table C1.[51] Lastly, I provide additional evidence that characteristics do not systematically vary around the break in NHS construction using a set of covariates that are available in 1930 at the county-level including the total number of automobile registrations per capita as of 1929.[52] Figure C2 plots the binned means for characteristics around the breaks in NHS construction within the fixed bandwidth of 1.2. For robustness, Table C1 displays the coefficient estimates for the regression specification in equation (B1) using baseline characteristics. Taken together, there is no visual evidence that characteristics change discontinuous around breaks in NHS construction, and all coefficient estimates are statistically indistiguishable from zero.

[49] Figure C1 plots portions of the NHS that are fully paved and the location of breaks in construction as of 1931.

[50] The running variable measured in hundreds of kilometers, meaning that a bandwidth of 1 is equivalent to 100 kilometers or approximately 62 miles.

[51] I include a variety of coefficient estimates using the optimal bandwidth based on Calonico et al. (2014). These estimates are similar in sign but larger in magnitude relative to other estimates. Because the bandwidth is significantly smaller, confidence intervals are also significantly noisier. In addition, Figure C5 shows the stability of coefficient estimates and confidence intervals based on the size of the bandwidth. Coefficient values stay relatively stable for most bandwidth values, but confidence intervals expand significantly as the bandwidth decreases passed 0.80 (or approximately 50 miles from the closest break in NHS construction).

[52] Other characteristics include: indicators for local railroad access, local military base, local metro area, distance to nearest metro area (logs), manufacturing value added (logs), and farm land value (logs).

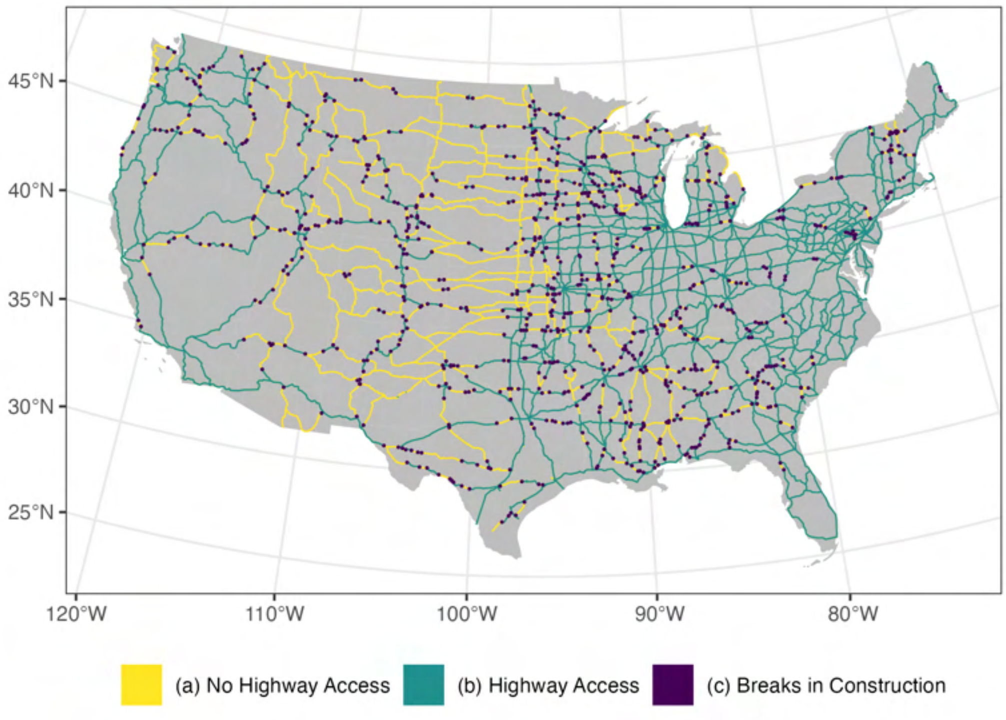
B.2 Empirical Results

Figure C3 provides graphical evidence of a discontinuous jump in new automobile registrations at the break in NHS construction. The plot overlays the predicted linear polynomial from equation (B1) with the scatter plot of new automobile registrations in 1932 collapsed to evenly spaced bins of approximately 1.25 miles in size. Using the selected bandwidth and uniforming weighting, I find that new automobile registrations increase by approximately 0.35 log points at the break in construction. The average number of new automobile registrations for the control group is approximately 50 automobiles per 1,000 individuals, implying an increase of 20 additional automobiles per 1,000 individuals.

Figure C4 shows that this increase in automobile utilization can be explained by a greater density of NHS roads, which local households could have used to commute locally within their county-of-residence. To estimate this effect, I categorize counties in my sample into two separate bins based on whether or not the length of planned NHS roads within a county is above average. I interact the treatment variable in equation (B1) with this indicator to measure the percentage difference in new automobile registrations at breaks in construction for these two groups of counties. Figure C4 presents the binned means plot for automobile registrations and the length of planned NHS roads per county for robustness. In Panel A, I show that the length of planned NHS roads did not systematically differ at the break in construction for either group. However in Panel B, I show the differences in new automobile registrations across these two groups is significant. I find that first counties with an above average number of NHS roads that have yet to be constructed trended similarly to those with a below average number of roads through the break in construction, whereas counties that had highway access as of 1931 with an above average number of roads have a significantly higher adoption rate than any other group. Accounting for differences at the break in construction, I find that highway access increased new automobile registrations by 0.95 log points for counties with an above average length of planned NHS roads.

C Additional Figures and Tables

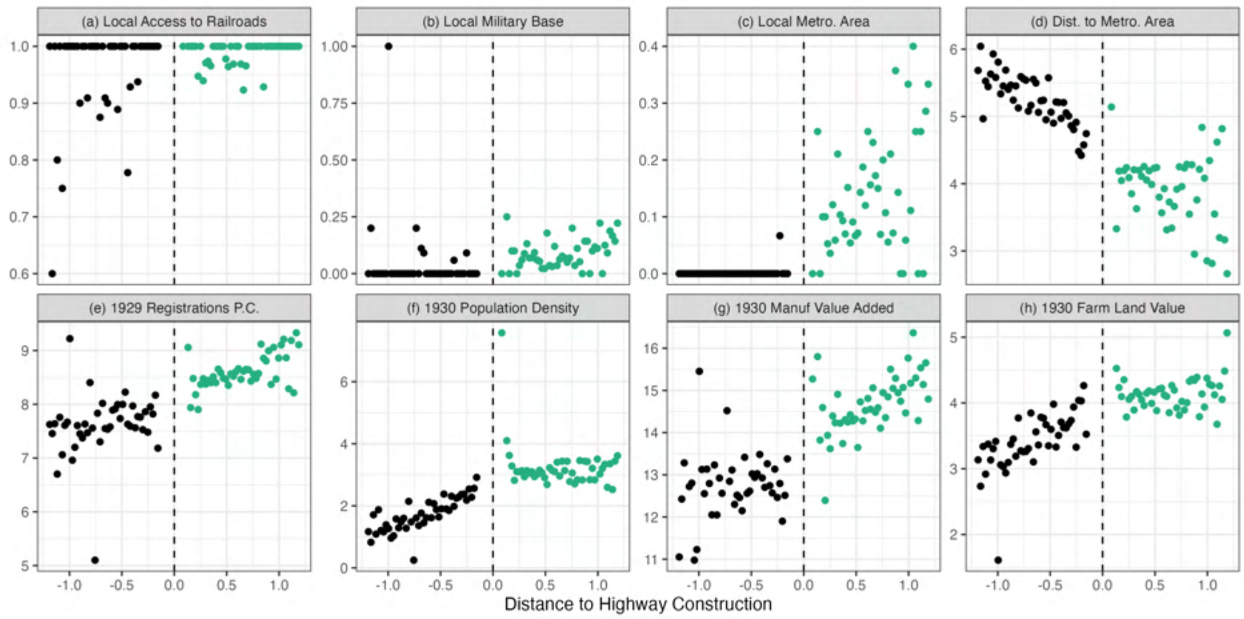
Figure C1: Location of Breaks in NHS Construction



Note: This figure plots the status of highway construction for the NHS as of 1931. Sections of the NHS in yellow correspond to incomplete sections of highway that were dirt roads. Sections of the NHS in green correspond to paved sections. Areas with purple dots indicate where the end of a paved portion of the NHS and a dirt road begins.

.

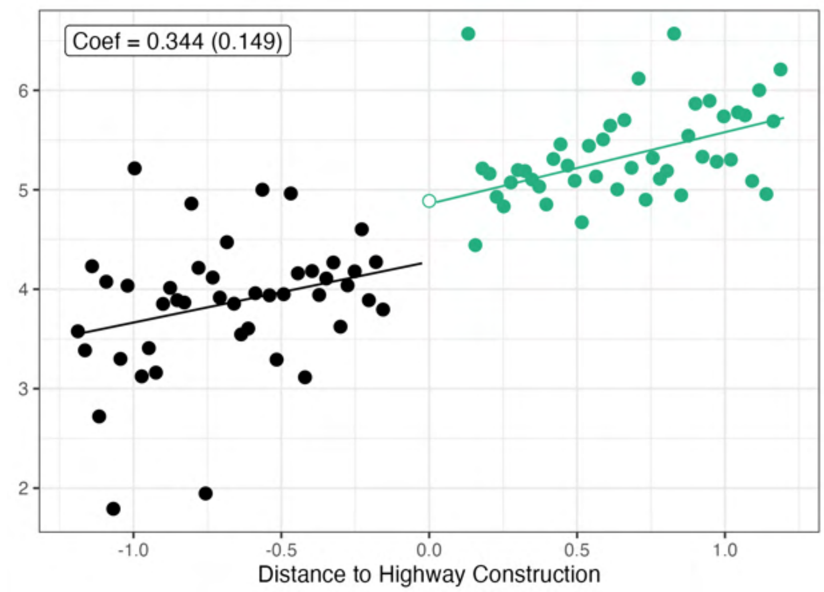
Figure C2: Graphical Evidence of Balanced Characteristics



Note: This figure plots the binned scatterplot for 1930 baseline characteristics. The x-axis is the straight-line distance between a county and the nearest break in highway construction in 1931, where the negative value on the x-axis corresponds to the set of counties that do not have access to a completed portion of the NHS in 1931. Baseline characteristics include: a binary indicator for local railroad access, a binary indicator for a local military base, a binary indicator for being part of a local metro area, distance to the nearest metro area (logs), total number of automobile registrations per capita as of 1929 (logs), population density (log), manufacturing value added (logs), and average farm land value (logs).

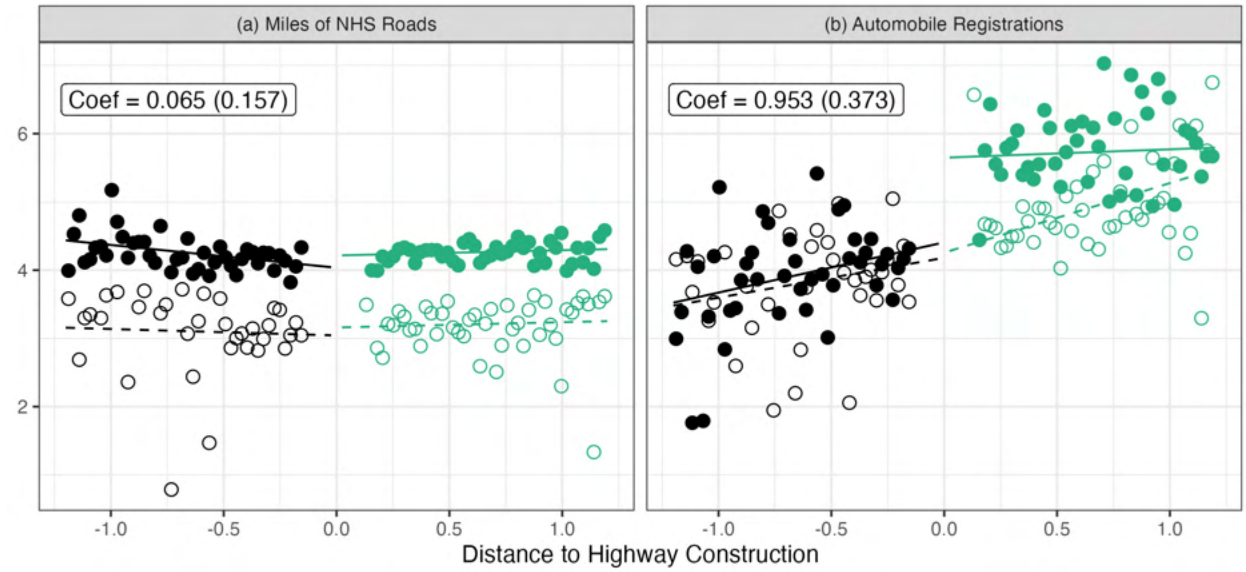
.

Figure C3: New Automobile Registrations around Breaks in NHS Construction, 1932



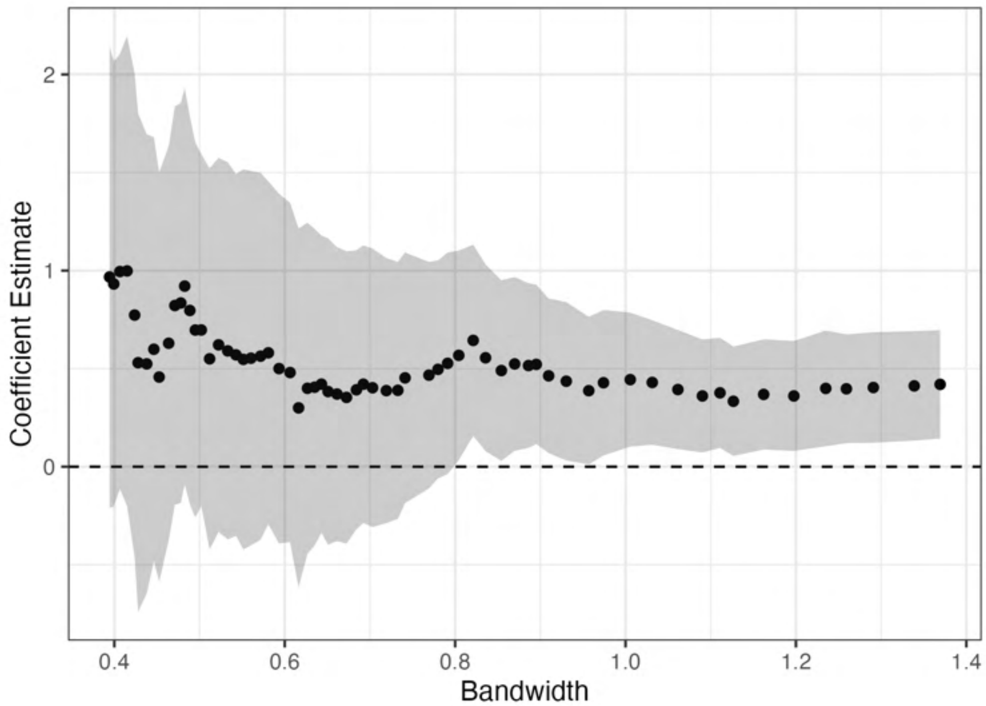
Note: This figure plots the binned scatterplot of new automobile registrations per capita (logs) at the county level in 1932. The x-axis is the straight-line distance between a county and the nearest break in highway construction in 1931, where the negative value on the x-axis corresponds to the set of counties that do not have access to a completed portion of the NHS in 1931. The coefficient estimate and robust standard error from equation (B1) is included in the figure.

Figure C4: Heterogeneity in New Automobile Registrations, 1932



Note: This figure plots the binned scatterplot for two outcome variables: miles of planned NHS roads (logs) and new automobile registrations per capita (logs). The x-axis is the straight-line distance between a county and the nearest break in highway construction in 1931. Solid (hollow) points correspond to set of counties with an above (below) average length of planned NHS roads. The coefficient estimate and robust standard error from equation (B1) is included in the figure.

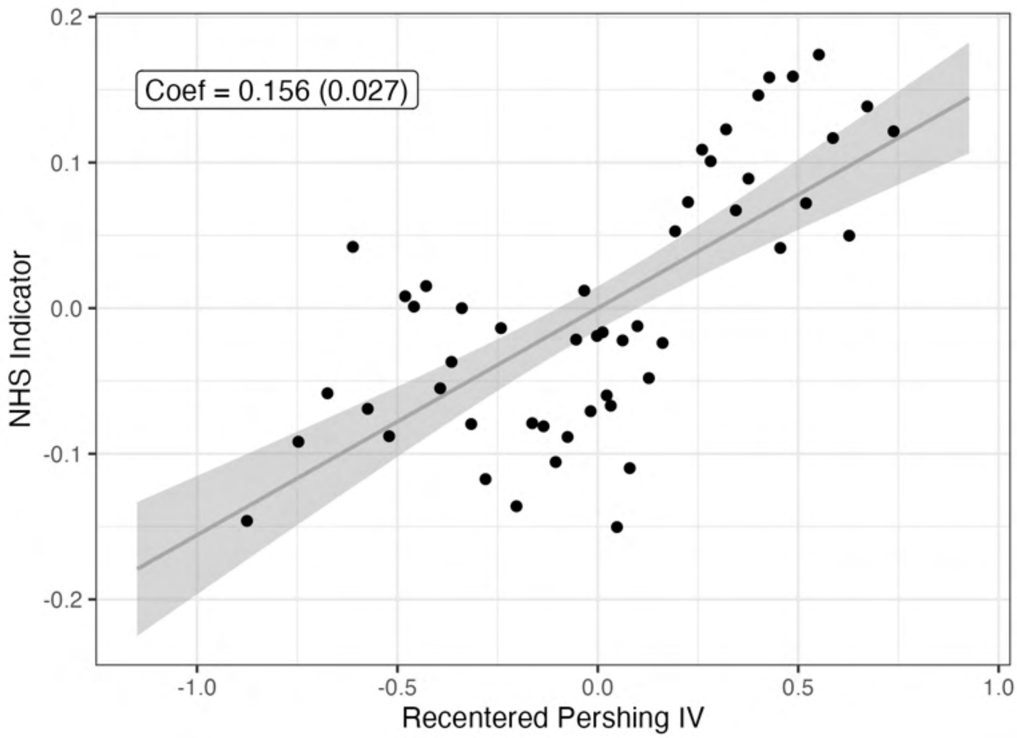
Figure C5: Effect of Breaks in NHS Construction with Varying Bandwidth



Note: This figure plots coefficient estimates and 95 percent confidence intervals for the regression specification presented in equation (B1) when the effective bandwidth varies between 0.4 and 1.4. These values correspond to a range of approximately 25 and 87 miles from the nearest break in construction.

.

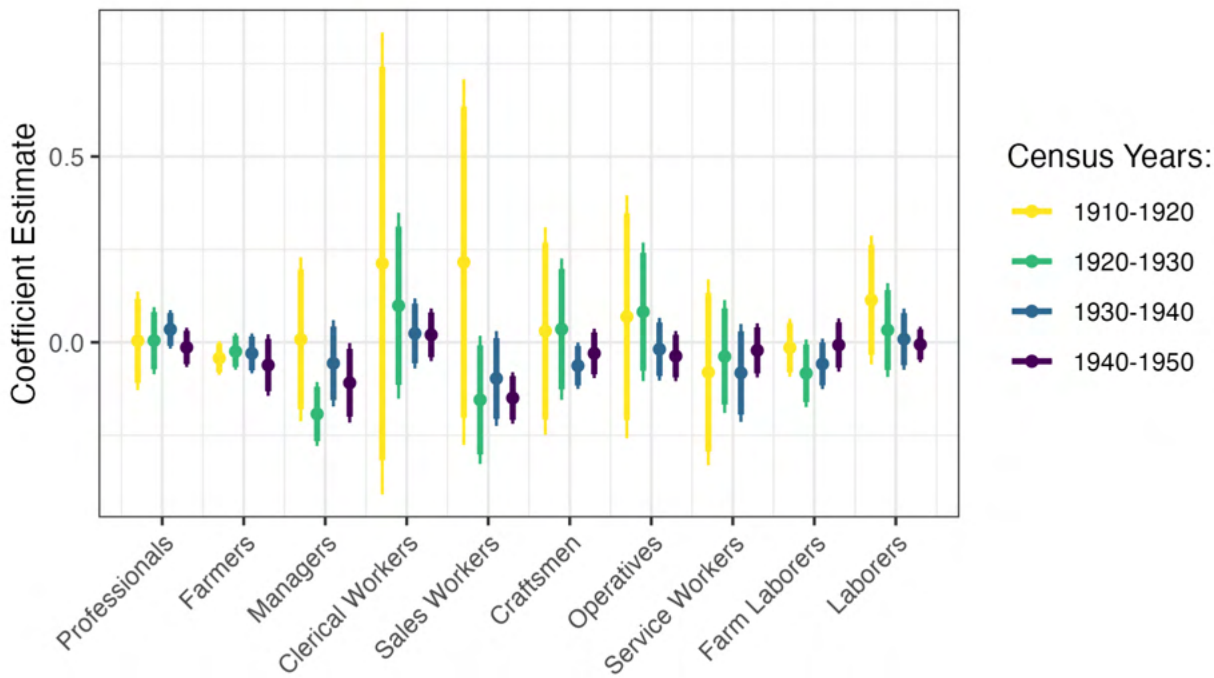
Figure C6: First Stage Relationship



Note: This figure plots the binned scatterplot for the first stage relationship between the NHS indicator and the Pershing IV. Each variable is residualized after being regressed on the baseline covariates such as 1920 population (logs), county size (log), and the fraction of a county covered by water or areas with at least a 15 percent incline in elevation.

.

Figure C7: Within-Individual Effects by Occupation



Note: This figure plots the coefficient estimates and 90/95 percent confidence intervals for the regression specification presented in equation (4) when the sample of movers are stratified by Census Years and occupation. The definitions for occupation are based on the broad categories for `OCC1950`, which spans all observed occupations in the microdata, are available on the IPUMS USA website.

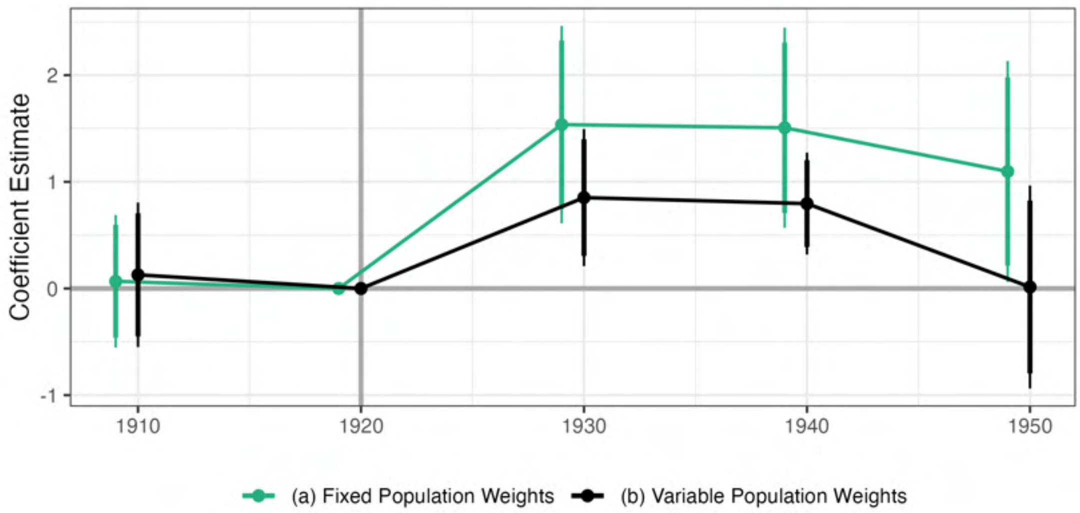
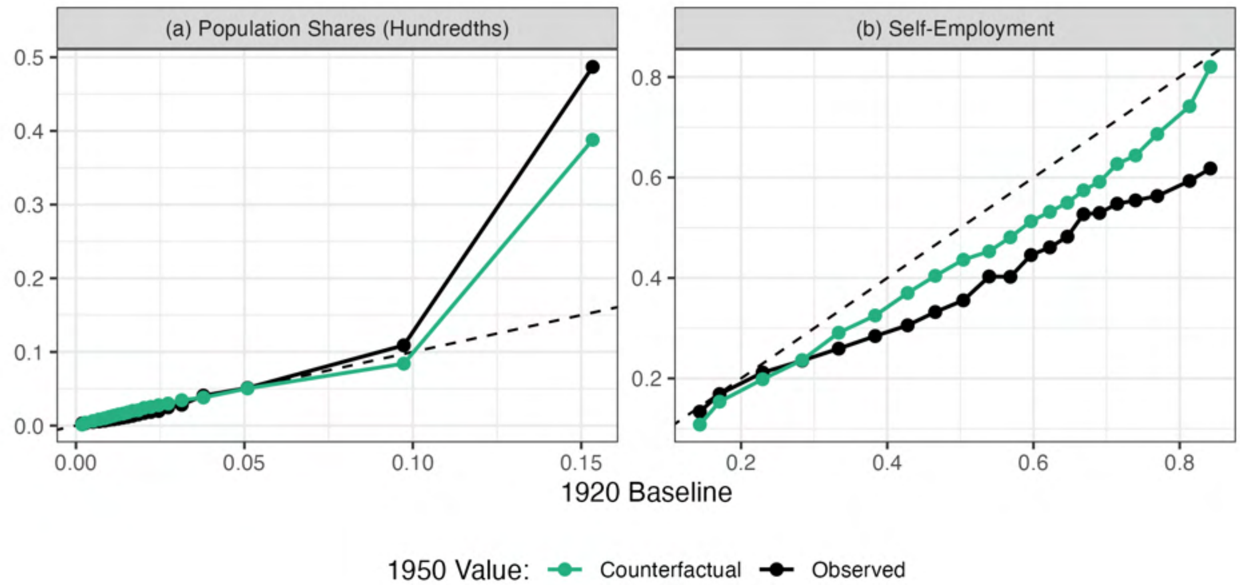


Figure C8: Local Effects of Market Access with Updated Population Shares

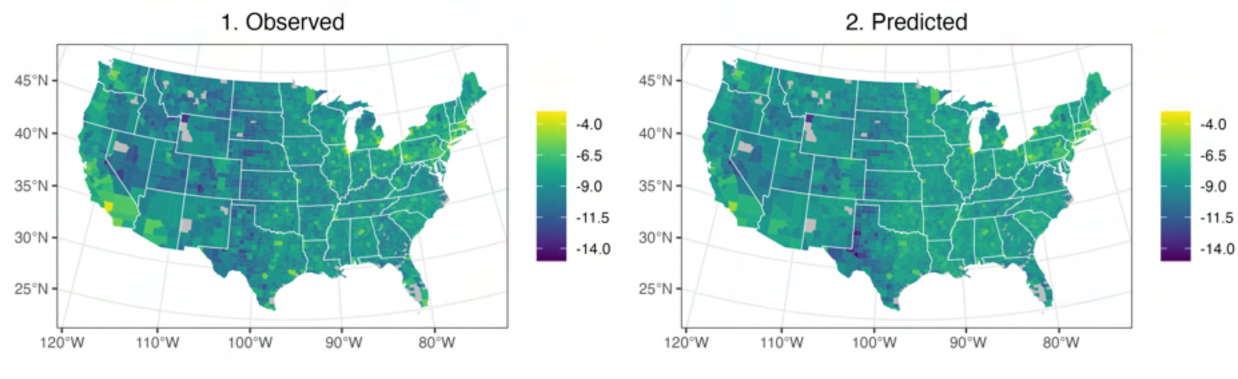
Note: This figure displays the 2SLS coefficient estimates from equation (24) and an associated event study. The estimates from Figure 6 are included as the series named Figure Population Weights, whereas the series titled, Variable Population Weights, updates population shares when calculating market access. In particular, I use 1920 population shares to calculate market access for 1910 and 1920, and I use 1950 population shares for the following decades.

Figure C9: Matching 1950 Population Shares and Self-Employment Rates

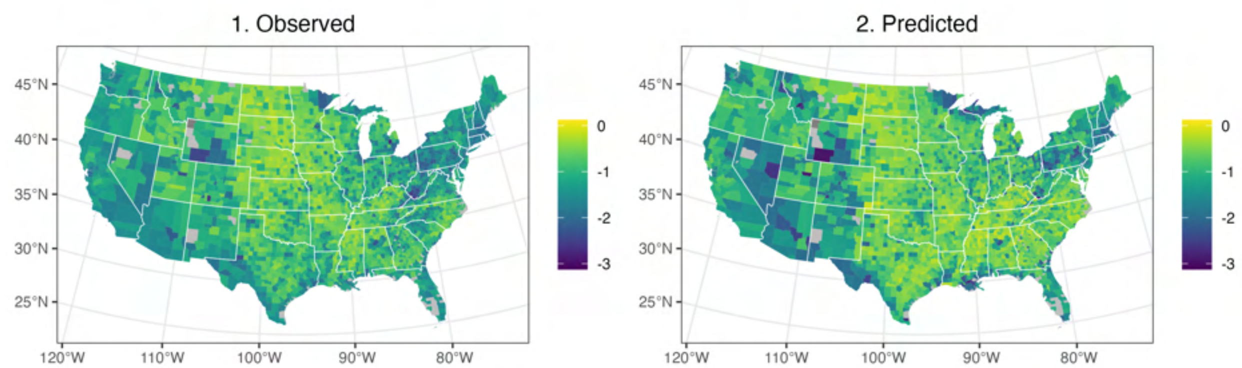


Note: This figure plots the binned scatterplot of county-level population shares and the local self-employment rate in 1950 against their 1920 value. Data points plotted in green correspond to the model predictions of the 1950 population shares and local self-employment rates based on the theoretical framework. Data points in black correspond to their observed 1950 values. The 1920 distribution included on the x-axis are winsorized at the 2.5 and 97.5 percentile.

Figure C10: Geographic Variation in 1950 Population Shares and Self-Employment Rates



(a) Population Shares (logs)



(b) Self-Employment Rate (logs)

Note: This figure plots the county-level population shares and the local self-employment rate in 1950 using their observed values on the left hand side and the model prediction on the right.

Table C1: Robustness Checks for Breaks in NHS Construction, 1932

	Kernel weight:					
	Uniform		Triangular		Epanechnikov	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Fixed Bandwidth						
Highway Access	0.344**	0.722	0.441**	0.722	0.420**	0.704
	(0.149)	(0.499)	(0.200)	(0.595)	(0.179)	(0.584)
Bandwidth	1.200	1.200	1.200	1.200	1.200	1.200
N	961	961	961	961	961	961
Panel B: Imbens-Kalyanaraman Bandwidth						
Highway Access	0.398***	0.528	0.524	0.818	0.478	0.968
	(0.141)	(0.393)	(0.340)	(0.723)	(0.362)	(0.720)
Bandwidth	1.325	1.325	0.843	0.843	0.785	0.785
N	993	993	781	781	731	731
Panel C: Calonico-Cattaneo-Titiunik Bandwidth						
Highway Access	0.922	0.896	1.178	1.408	1.063	1.304
	(0.806)	(0.877)	(0.704)	(0.866)	(0.684)	(0.856)
Bandwidth	0.333	0.694	0.404	0.696	0.406	0.698
N	184	663	297	665	300	667
Polynomial Fit:	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic

Note: This table displays the coefficient estimates from equation (B1). Coefficient estimates vary based on the selection of the bandwidth and kernel estimate used to determine the effect sample and weighting scheme of observations. Below each coefficient estimate, the effective bandwidth and sample are provided.
*p<0.1; **p<0.05; ***p<0.01

Table C2: Balance of Baseline Characteristics, 1930

	Dependent variable:							
	Railroads	Military Base	Metro Area	Dist to Metro	Registrations	Pop Density	Value Added	Land Value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Fixed Bandwidth								
Highway Access	−0.049	0.038	−0.253	0.033	0.132	0.172	0.546	−0.116
	(0.034)	(0.025)	(0.166)	(0.030)	(0.132)	(0.120)	(0.334)	(0.122)
N	1,321	1,321	1,321	1,321	1,320	1,321	1,154	1,320
Panel B: Imbens-Kalyanaraman Bandwidth								
Highway Access	−0.051	0.032	−0.046	0.040	0.095	0.189	0.495	−0.089
	(0.034)	(0.023)	(0.215)	(0.033)	(0.126)	(0.189)	(0.328)	(0.119)
Bandwidth	1.340	1.164	0.951	0.773	1.140	0.654	1.127	1.170
N	1,381	1,305	1,174	998	1,289	859	1,128	1,307
Panel C: Calonico-Cattaneo-Titiunik Bandwidth								
Highway Access	0.000	0.067	−0.014	0.116	−0.045	0.257	0.901	−0.205
	(0.000)	(0.042)	(0.045)	(0.257)	(0.482)	(0.231)	(0.469)	(0.240)
Bandwidth	0.217	0.631	0.707	0.664	0.451	0.556	0.807	0.555
N	45	815	932	874	506	705	923	703

Note: This table displays the coefficient estimates from equation (B1) using 1930 baseline covariates. Coefficient estimates vary based on the selection of the bandwidth and kernel estimate used to determine the effect sample and weighting scheme of observations. Below each coefficient estimate, the effective bandwidth and sample are provided. The set of covariates include: indicators for a local military base or metro area, the distance to the nearest metro area, the total number of automobile registrations per captia, population density, manufacturing value added, and the average farm land value. *p<0.1; **p<0.05; ***p<0.01

Table C3: Additional Specification Checks for Local Effects on Population Growth

	Dependent variable:			
	Change in Population (logs)			
	(1)	(2)	(3)	(4)
Panel A: IV Estimates with Metro Areas				
$\widehat{\text{NHS}}$	0.385***	0.464**	0.475**	0.493**
	(0.067)	(0.230)	(0.218)	(0.217)
F statistic	54	32	32	31
N	2,919	2,919	2,919	2,919
Panel B: IV Estimates using Inverse Hyperbolic Sine Function				
$\widehat{\text{NHS}}$	0.303***	0.446**	0.465**	0.477**
	(0.068)	(0.217)	(0.205)	(0.202)
F statistic	47	29	28	27
N	2,719	2,719	2,719	2,719
Panel C: IV Estimates controlling for Latitude and Longitude				
$\widehat{\text{NHS}}$	0.280***	0.450**	0.434**	0.454**
	(0.066)	(0.197)	(0.190)	(0.195)
F statistic	47	28	28	27
N	2,719	2,719	2,719	2,719
Baseline Covariates	N	Y	Y	Y
State Fixed Effects	N	N	Y	Y
Full Set of Covariates	N	N	N	Y

Note: This table displays the 2SLS coefficient estimates from equation (1) using the Pershing IV as defined in equation (2). Panel A increases the sample size to include counties that are part of a major metro area. Panel B calculates the change in population using an inverse hyperbolic sine function as opposed to a log transformation. Lastly, Panel C includes controls based on a county centroid. These controls consist of a second order polynomial for latitude and longitude. See Table 2 for additional details. *p<0.1; **p<0.05; ***p<0.01.

Table C4: Local Effect on Population Growth with Controls for Great Depression

	Dependent variable:			
	Change in Population (logs)			
	(1)	(2)	(3)	(4)
$\widehat{\text{NHS}}$	0.299***	0.473**	0.466**	0.473**
	(0.070)	(0.238)	(0.226)	(0.232)
F statistic	42	25	28	27
N	2,680	2,680	2,680	2,680
Baseline Covariates	N	Y	Y	Y
State Fixed Effects	N	N	Y	Y
Full Set of Covariates	N	N	N	Y
Add. Great Depression Covariates	Y	Y	Y	Y

Note: This table displays the 2SLS coefficient estimates from equation (1) using the Pershing IV as defined in equation (2). Additional controls accounting for differential exposure to the Great Depression include additional controls for New Deal spending and the decline in the manufacturing sector and local credit markets between 1929 and 1931. New Deal spending is the county-level amount of spending per capita in logs. Data on New Deal spending comes from Wallis (1987). To control for the decline in manufacturing, I use a shift-share variable equal to the local employment weighted to the national decline in manufacturing value added between 1929 and 1931 based on reported changes in manufacturing value added from Fabricant (1940). For both variables, I rely on population and employment data from the 1930 Decennial Census. Lastly, I use the log change in county-level deposits from 1929 to 1931 from FDIC (1992). See Table 2 for additional details. *p<0.1; **p<0.05; ***p<0.01.

Table C5: Local Effects on Self-Employment for the Employed, 1910-1950

	Dependent variable:		
	Self-Employment		
	(1)	(2)	(3)
Panel A: All Individuals			
NHS	0.002		
	(0.003)		
Pershing IV		-0.008^{***}	
		(0.002)	
$\widehat{\text{NHS}}$			-0.056^{***}
			(0.018)
F statistic			31
Non-NHS Mean			0.484
N	20,724,570	20,724,570	20,724,570
Panel B: Employed Individuals			
NHS	0.002		
	(0.003)		
Pershing IV		-0.008^{***}	
		(0.002)	
$\widehat{\text{NHS}}$			-0.058^{***}
			(0.018)
F statistic			31
Non-NHS Mean			0.487
N	20,075,109	20,075,109	20,075,109
Baseline Covariates	Y	Y	Y
County Fixed Effects	Y	Y	Y
Year Fixed Effects	Y	Y	Y

Note: This table displays the 2SLS coefficient estimates from equation (3) using the Pershing IV as defined in equation (2) where the dependent variable is a binary indicator for being self-employed. The sample includes employed individuals in 1910 and 1950. Note that a question on labor force participation is not included in the 1920 Decennial Census. See Table 4 for additional details. *p<0.1; **p<0.05; ***p<0.01

Table C6: Local Effect on Self-Employment using County-Level Data

	Dependent variable:				
	Local Self-Employment Rate (logs)				
	(1)	(2)	(3)	(4)	(5)
$\widehat{\text{NHS}}$	0.138	-0.572^{**}	-0.352^{*}	-0.273^{**}	-0.248^{**}
	(0.130)	(0.256)	(0.187)	(0.121)	(0.118)
Baseline Covariates	N	Y	Y	Y	Y
Controls for Latitude and Longitude	N	N	Y	Y	Y
State Fixed Effects	N	N	N	Y	Y
Full Set of Covariates	N	N	N	N	Y
N	2,940	2,937	2,937	2,937	2,917

Note: This table displays the 2SLS coefficient estimates from equation (3) using the Pershing IV as defined in equation (2) where the dependent variable is a binary indicator for being self-employed. The unit of observation for data included in these regressions is the county-level as opposed to the individual-level. See Table 4 for additional details. *p<0.1; **p<0.05; ***p<0.01

Table C7: Local Effect on Other Margins of Labor Force Participation, 1910-1950

	Dependent variable:				
	Labor Force Participant	Employed		Self-Employment	
	(1)	(2)	(3)	(4)	(5)
$\widehat{\text{NHS}}$	-0.004	0.003	0.007	-0.057^{**}	-0.056^{***}
	(0.003)	(0.013)	(0.013)	(0.018)	(0.018)
Labor Force Participants	N	N	Y	N	Y
Baseline Covariates	Y	Y	Y	Y	Y
County Effects	Y	Y	Y	Y	Y
Year Effects	Y	Y	Y	Y	Y
N	20,896,548	20,896,548	20,724,570	20,896,548	20,724,570

Note: This table displays the 2SLS coefficient estimates from equation (3) using the Pershing IV as defined in equation (2). The dependent variables included in these regressions include binary indicators for whether someone was employed or actively looking for a job. Observations vary in columns 3 and 5, which removes individuals who were out of the labor force. See Table 4 for additional details. *p<0.1; **p<0.05; ***p<0.01

Table C8: Robustness Tests for Within-Individual Design

	Dependent variable:			
	Self-Employment			
	(1)	(2)	(3)	(4)
Panel A: Excluding Employers according to 1940 Decennial Census				
$\widehat{\text{NHS}}$	-0.058^{**}	-0.028	-0.006	0.025
	(0.027)	(0.028)	(0.031)	(0.031)
$\widehat{\text{NHS}} \times$ Closer to Metros		-0.073^{**}		
		(0.031)		
$\widehat{\text{NHS}} \times$ Higher Density			-0.105^{***}	
			(0.038)	
$\widehat{\text{NHS}} \times$ More Expensive				-0.127^{***}
				(0.036)
N	3,449,914	3,449,914	3,449,914	3,449,914
Panel B: Subsetting to Employed Individuals				
$\widehat{\text{NHS}}$	-0.062^{**}	-0.026	-0.002	0.028
	(0.028)	(0.029)	(0.031)	(0.032)
$\widehat{\text{NHS}} \times$ Closer to Metros		-0.084^{**}		
		(0.033)		
$\widehat{\text{NHS}} \times$ Higher Density			-0.122^{***}	
			(0.039)	
$\widehat{\text{NHS}} \times$ More Expensive				-0.140^{***}
				(0.036)
N	3,181,560	3,181,560	3,181,560	3,181,560
Panel C: Including All Connected Individuals				
$\widehat{\text{NHS}}$	-0.083^{***}	-0.054^{*}	-0.029	-0.037
	(0.027)	(0.027)	(0.024)	(0.026)
$\widehat{\text{NHS}} \times$ Closer to Metros		-0.064^{**}		
		(0.031)		
$\widehat{\text{NHS}} \times$ Higher Density			-0.101^{***}	
			(0.026)	
$\widehat{\text{NHS}} \times$ More Expensive				-0.074^{***}
				(0.026)
N	8,272,133	8,272,133	8,272,133	8,272,133
Baseline Covariates	Y	Y	Y	Y
Individual Fixed Effects	Y	Y	Y	Y
Occupation $\times$ Year Fixed Effects	Y	Y	Y	Y

Note: This table displays the 2SLS coefficient estimates from equation (4) using the Pershing IV as defined in equation (2). The estimates in Panel A are based on a sample of movers which exclude employers based on responses in the 1940 Decennial Census. The estimates in Panel B are based on a sample of movers which excludes unemployed or non-labor force participants according to the 1940 and 1950 Decennial Census. Lastly, Panel C additionally includes all connected individuals between the 1940 and 1950 Decennial Census. Sample weights are re-estimated for the observations included in Panel C. See Table 6 for additional details. *p<0.1; **p<0.05; ***p<0.01

Table C9: Testing Alternative Mechanism, 1940-1950

	Dependent variable:			
	Self-Employment			
	(1)	(2)	(3)	(4)
$\widehat{\text{NHS}}$	-0.029	-0.030	-0.057^{**}	-0.057^{**}
	(0.044)	(0.046)	(0.026)	(0.026)
$\widehat{\text{NHS}} \times$ More Roads	-0.081			
	(0.053)			
$\widehat{\text{NHS}} \times$ Denser Roads		-0.079		
		(0.052)		
$\widehat{\text{NHS}} \times$ Weeks (logs)			-0.149	
			(0.106)	
$\widehat{\text{NHS}} \times$ Hours (logs)				-0.149
				(0.106)
Baseline Covariates	Y	Y	Y	Y
Individual Fixed Effects	Y	Y	Y	Y
Occupation $\times$ Year Fixed Effects	Y	Y	Y	Y
N	3,518,814	3,518,814	3,518,814	3,518,814

Note: This table displays the 2SLS coefficient estimates from equation (4) using the Pershing IV as defined in equation (2). The interaction term in column 1 is a binary indicator equal to one if an individual's post-move county-of-residence has a greater length of NHS roads as of 1939. The interaction term in column 2 is a binary indicator equal to one if an individual's post-move county-of-residence has a greater density of NHS roads as of 1939. The interaction term in column 3 is relative difference in average number of weeks worked as of 1939 between an individual's county-of-residence before and after their move. Lastly, the interaction term in column 4 relative difference in average number of hours worked as of March 24th through 30th in 1940 between an individual's county-of-residence before and after their move. The interaction term is additionally interacted with a year fixed effect to account for time-varying effects across Census years. *p<0.1; **p<0.05; ***p<0.01

Table C10: Falsification Test for Within-Individual Design, 1910-1950

	Linked Census years:			
	1910-1920	1920-1930	1930-1940	1940-1950
	(1)	(2)	(3)	(4)
Panel A: Pre-Move Differences				
$\widehat{\text{NHS}}$	0.003	-0.018	-0.026	-0.030
	(0.022)	(0.021)	(0.018)	(0.030)
Baseline Covariates	Y	Y	Y	Y
N	1,296,850	1,572,290	1,419,334	1,759,407
Panel B: Post-Move Differences				
$\widehat{\text{NHS}}$	-0.058	-0.142^{***}	-0.120^{***}	-0.206^{***}
	(0.074)	(0.054)	(0.035)	(0.033)
Baseline Covariates	Y	Y	Y	Y
N	1,296,850	1,572,290	1,419,334	1,759,407
Panel C: Difference-in-Differences				
$\widehat{\text{NHS}}$	-0.050	-0.103^{**}	-0.071^{*}	-0.120^{***}
	(0.052)	(0.046)	(0.039)	(0.038)
Baseline Covariates	Y	Y	Y	Y
Individual Fixed Effects	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y
N	2,593,700	3,144,580	2,838,668	3,518,814
Baseline Covariates	Y	Y	Y	Y

Note: This table displays the 2SLS coefficient estimates from equation (4) using the Pershing IV as defined in equation (2). Panels A and B reflect the cross-sectional differences in individual outcomes before and after moving between counties. Panel C displays the full difference-in-difference estimate accounting for individual and year fixed effects. Each column corresponds to adjacent years in the Decennial Census. See Table 6 for additional details. *p<0.1; **p<0.05; ***p<0.01

Table C11: Within-Individual Design with Employer as an Outcome, 1910-1940

	Linked Census Years:		
	1910-1920	1920-1930	1930-1940
	(1)	(2)	(3)
$\widehat{\text{NHS}}$	0.031	0.012	0.013
	(0.031)	(0.016)	(0.010)
Baseline Covariates	Y	Y	Y
Individual Fixed Effects	Y	Y	Y
Year Fixed Effects	Y	Y	Y
N	2,593,700	3,144,580	2,838,668

Note: This table displays the 2SLS coefficient estimates from equation (4) using the Pershing IV as defined in equation (2). The outcome variable is a binary indicator for being an employer. Each column corresponds to adjacent years in the Decennial Census. Note the question on being an employer is excluded from the 1950 Decennial Census. See Table 6 for additional details. *p<0.1; **p<0.05; ***p<0.01

Table C12: Specification Test for Within-Individual Design, 1910-1950

	Linked Census Years:			
	1910-1920	1920-1930	1930-1940	1940-1950
	(1)	(2)	(3)	(4)
Panel A: State × Year Fixed Effects				
$\widehat{\text{NHS}}$	-0.050	-0.093^{*}	-0.072^{*}	-0.107^{***}
	(0.048)	(0.052)	(0.039)	(0.039)
Panel B: Sector × Year Fixed Effects				
$\widehat{\text{NHS}}$	-0.011	-0.052	-0.055	-0.093^{**}
	(0.066)	(0.057)	(0.038)	(0.035)
Panel C: Occupation × Year Fixed Effects				
$\widehat{\text{NHS}}$	0.002	-0.008	-0.045	-0.057^{**}
	(0.072)	(0.064)	(0.037)	(0.027)
Baseline Covariates	Y	Y	Y	Y
Individual Fixed Effects	Y	Y	Y	Y
N	2,593,700	3,144,580	2,838,668	3,518,814

Note: This table displays the 2SLS coefficient estimates from equation (4) using the Pershing IV as defined in equation (2). Panels A, B, and C differ based on the time fixed effects used in the regression specification. In Panel B, sectors correspond to agriculture, manufacturing, and services. Consistent with the IPUMS variable `IND1950`, the agriculture sector corresponds to values 1-299, the manufacturing corresponds to values 300-499, and the service sector corresponds to 500-899. In Panel C, occupation groups correspond to the IPUMS variable `OCC1950` and include: professionals, farmers, managers, clerical workers, sales workers, craftsmen, operatives, service workers, farm laborers, and laborers. See Table 6 for additional details. *p<0.1; **p<0.05; ***p<0.01

Table C13: Bus and Rail Fares between Selected Points, 1939

Origin	Destination	Bus	Rail
Atlanta, GA	Dallas, TX	12.55	15.52
Atlanta, GA	Los Angeles, CA	34.95	62.90
Boston, MA	Detroit, MI	14.00	19.84
Chicago, IL	Amarillo, TX	15.00	20.30
Chicago, IL	Denver, CO	16.75	20.71
Cleveland, OH	Nashville, TN	7.15	12.90
Memphis, TN	Minneapolis, MN	13.60	17.92
Minneapolis, MN	Seattle, WA	28.00	33.00
New York, NY	Atlanta, GA	10.30	15.17
New York, NY	Chicago, IL	12.70	22.69
New York, NY	Miami, FL	15.90	23.57
Norfolk, VA	Nashville, TN	10.55	14.60
Omaha, NE	Salt Lake City, UT	17.10	20.64
Pittsburgh, PA	St. Louis, MO	11.20	15.44
Salt Lake City, UT	San Francisco, CA	16.00	16.29
San Francisco, CA	Seattle, WA	12.00	17.50
St. Louis, MO	Amarillo, TX	12.50	16.41
St. Louis, MO	Los Angeles, CA	31.25	36.08
St. Louis, MO	New Orleans, LA	9.50	13.90
Washington, DC	Chicago, IL	12.95	19.27

Note: This table displays prices for bus and train tickets for the same origin and destination pairs and are taken from Landon (1945).

Table C14: Local Effects of Market Access with Additional Geographic Controls

	Dependent variable:				
	Population (logs)				
	(1)	(2)	(3)	(4)	(5)
$\Delta \ln \widehat{MA} \times \mathbb{1}(t > 1920)$	1.347***	1.892**	2.503**	1.597**	1.902**
	(0.487)	(0.834)	(1.134)	(0.741)	(0.889)
State-Year Fixed Effect	N	Y	Y	Y	Y
Metro-Area Covariates	N	N	Y	Y	Y
Geo-Coordinates Covariates	N	N	N	Y	Y
Undevelopable Covariates	N	N	N	N	Y
County Fixed Effect	Y	Y	Y	Y	Y
Year Fixed Effect	Y	N	N	N	N
Counties	2,941	2,941	2,941	2,941	2,941

Note: This figure displays the 2SLS coefficient estimates from equation (24). The coefficient estimate in column 1 corresponds to the baseline estimate from Figure 6. Columns 2-5 additionally controls for state-year fixed effects. Columns 3-5 include controls for a local metro area and distance to the nearest metro area. Columns 4 and 5 include a second-order polynomial based on the latitude and longitude of a county centroid. Column 5 includes controls for the fraction of land covered in water or has above a 15 percent incline in elevation consistent with controls included in equation (1). See Figure 6 for additional details. *p<0.1; **p<0.05; ***p<0.01

Table C15: Residential Land-Use by Location

Location	Percent
Albert Lea, Minnesota	0.367
Charleston, West Virginia	0.193
Decatur, Alabama	0.109
Des Moines, Iowa	0.212
Detroit, Michigan	0.302
Duluth, Minnesota	0.090
Fairfield, Connecticut	0.103
Fort Wayne, Indiana	0.334
Fort Worth, Texas	0.203
Greensboro, North Carolina	0.097
Greenwich, Connecticut	0.500
Harrison, New York	0.643
Kansas City, Missouri	0.262
Little Rock, Arkansas	0.244
Los Angeles, California	0.298
Manchester, Massachusetts	0.978
Maplewood, New Jersey	0.567
Mason City, Iowa	0.118
Meridian, Mississippi	0.257
Minneapolis, Minnesota	0.368
Montclair, New Jersey	0.546
Norfolk, Virginia	0.178
Oklahoma City, Oklahoma	0.380
Omaha, Nebraska	0.309
Patchogue, New York	0.370
Petaluma, California	0.490
Petersburg, Virginia	0.343
Port Huron, Michigan	0.269
Portland, Oregon	0.229
Providence, Rhode Island	0.248
Quincy, Illinois	0.420
Richmond, Virginia	0.214
Schenectady, New York	0.281
Seattle, Washington	0.230
St. Louis, Missouri	0.296
Stockton, California	0.236
Tacoma, Washington	0.183
Waterloo, Iowa	0.253

Note: This table displays the fraction of residential land use for locations in the US and are taken from Lovelace (1949).

D Technical Appendix for Theoretical Framework

D.1 Profits and Self-Employment

To derive an expression for profits and self-employment in terms of entrepreneurial productivity, I first write out the local PPI as a function of productivity:

$$
\begin{aligned}
P_n &= \left(\int_{M_n} p_n(j)^{1-\sigma} dj \right)^{\frac{1}{1-\sigma}} \\
&= \left(\frac{\sigma}{\sigma - 1} \right) \frac{w_n}{\left(\int_{M_n} \phi(j)^{\sigma-1} dj \right)^{\frac{1}{\sigma-1}}} \\
&= \left(\frac{\sigma}{\sigma - 1} \right) \frac{w_n}{\Phi_n}
\end{aligned}
$$

where I define $\Phi_n = \left[\int_{M_n} \phi(j)^{\sigma-1} dj \right]^{\frac{1}{\sigma-1}}$. Using the equation for profits π_n of a firm with productivity ϕ:

$$
\begin{aligned}
\pi_n(\phi) &= (p_n - mc_n) q_n - r_n \rho_n \\
&= (p_n - mc_n) \left(\frac{p_n}{P_n} \right)^{-\sigma} \frac{Y_n}{P_n} - r_n \rho_n \\
&= \left[\left(\frac{\sigma}{\sigma - 1} \right) \frac{w_n}{\phi_n} - \frac{w_n}{\phi_n} \right] \left(\frac{\left(\frac{\sigma}{\sigma-1} \right) \frac{w_n}{\phi_n}}{\left(\frac{\sigma}{\sigma-1} \right) \frac{w_n}{\Phi_n}} \right)^{-\sigma} \frac{Y_n}{\left(\frac{\sigma}{\sigma-1} \right) \frac{w_n}{\Phi_n}} - r_n \rho_n \\
&= \left[\frac{1}{\sigma} \right] \left[\frac{\phi}{\Phi_n} \right]^{\sigma-1} Y_n - r_n \rho_n
\end{aligned}
$$

where mc_n denotes the marginal cost of production.

The break even level of productivity is when an individual is indifferent between self and salaried employment (i.e. $\pi(\underline{\phi}_n) = w_n$). To begin, I write out an expression for Φ_n in terms of $\underline{\phi}_n$:

$$
\begin{aligned}
\Phi_n &= \left[\int_{M_n} \phi_n(j)^{\sigma-1} dj \right]^{\frac{1}{\sigma-1}} \\
&= \left[L_n \int_{\underline{\phi}_n}^{\infty} \phi^{\sigma-1} dG(\phi) \right]^{\frac{1}{\sigma-1}} \\
&= \left[L_n \int_{\underline{\phi}_n}^{\infty} \phi^{\sigma-1} \left(\frac{\alpha \beta^{\alpha}}{\phi^{1+\alpha}} \right) d\phi \right]^{\frac{1}{\sigma-1}}
\end{aligned}
$$

$$= \left[\left(\frac{\alpha \beta^{\alpha}}{\alpha - \sigma + 1} \right) L_n \underline{\phi}_n^{\sigma - \alpha + 1} \right]^{\frac{1}{\sigma - 1}}$$

where $\alpha > \sigma - 1$. Now, I invert the indifference point between profits and wages:

$$\begin{aligned}
\left[\frac{1}{\sigma} \right] \left[\frac{\underline{\phi}_n}{\Phi_n} \right]^{\sigma-1} Y_n - r_n \rho_n &= w_n \\
\left[\frac{\phi}{\Phi_n} \right]^{\sigma-1} &= \sigma \left[\frac{w_n + r_n \rho_n}{Y_n} \right] \\
\frac{\underline{\phi}_n^{\sigma-1}}{\left(\frac{\alpha \beta^{\alpha}}{\alpha - \sigma + 1} \right) L_n \underline{\phi}_n^{\sigma - \alpha + 1}} &= \sigma \left[\frac{w_n + r_n \rho_n}{Y_n} \right] \\
\underline{\phi}_n = \left[\left(\frac{\alpha \beta^{\alpha} \sigma}{\alpha - \sigma + 1} \right) \left(\frac{(w_n + r_n \rho_n) \, L_n}{Y_n} \right) \right]^{\frac{1}{\alpha}}
\end{aligned}$$

where break even productivity is a function of opportunity costs to nominal revenues per capita.

After solving for the break even level of productivity, I use the Pareto assumption to write out the self-employment rate:

$$\begin{aligned}
1 - G(\underline{\phi}_n) &= \left(\frac{\beta}{\underline{\phi}_n} \right)^{\alpha} \\
&= \frac{\beta^{\alpha}}{\left(\frac{\alpha \beta^{\alpha} \sigma}{\alpha - \sigma + 1} \right) \left(\frac{(w_n + r_n \rho_n) \, L_n}{Y_n} \right)} \\
&= \left(\frac{\alpha - \sigma + 1}{\alpha \sigma} \right) \left(\frac{Y_n}{(w_n + r_n \rho_n) \, L_n} \right)
\end{aligned}$$

With the labor market clearing condition such that $w_n G(\underline{\phi}_n) L_n = \left(\frac{\sigma - 1}{\sigma} \right) Y_n$, the self-employment rate simplifies to the following expression:

$$E_n = \left[1 + \left(\frac{\alpha(\sigma - 1)}{\alpha - \sigma + 1} \right) \left(1 + \frac{r_n \rho_n}{w_n} \right) \right]^{-1} \quad \text{s.t. } E_n = 1 - G(\underline{\phi}_n)$$

which is declining if the fixed cost of land rises rises faster than the marginal cost of labor.

Lastly, profits are:

$$\begin{aligned}
\pi_n [1 - G(\underline{\phi}_n)] L_n &= \int_{M_n} \left(\left[\frac{1}{\sigma} \right] \left[\frac{\phi(j)}{\Phi_n} \right]^{\sigma-1} Y_n - r_n \rho_n \right) dj \\
&= \left[\frac{Y_n}{\sigma} \right] - r_n \rho_n L_n \int_{\underline{\phi}_n}^{\infty} dG(\phi)
\end{aligned}$$

$$= \left[\frac{Y_n}{\sigma}\right] - \left[1 - G(\underline{\phi}_n)\right] r_n \rho_n L_n$$

$$= \left[\frac{Y_n}{\sigma}\right] \left[1 - \left(1 - \left[\frac{\sigma - 1}{\alpha}\right]\right) \left(\frac{r_n \rho_n}{r_n \rho_n + w_n}\right)\right]$$

$$= \frac{\left(\frac{\sigma-1}{\sigma}\right) Y_n}{\left(\left[\frac{w_n}{r_n \rho_n + w_n}\right] (\sigma - 1)^{-1} + \left[\frac{r_n \rho_n}{r_n \rho_n + w_n}\right] \alpha^{-1}\right)^{-1}}$$

With the labor market clearing condition such that $w_n G(\underline{\phi}_n) L_n = \left(\frac{\sigma-1}{\sigma}\right) Y_n$, profits simplifies to the following expression:

$$\pi_n = \left[\frac{1 - E_n}{E_n}\right] \left[\frac{w_n}{\left(\left[\frac{w_n}{r_n \rho_n + w_n}\right] (\sigma - 1)^{-1} + \left[\frac{r_n \rho_n}{r_n \rho_n + w_n}\right] \alpha^{-1}\right)^{-1}}\right]$$

where profits are a rescaled value of wages accounting for the local self-employment rate and the balance between firm-level competition and productivity.

D.2 Defining Market Access and Residential Choice

In this section, I describe the market access approach to describing population shares in the economy. To begin, expenditure shares under CES preferences are:

$$s_{ln} = \frac{\left[\tau_{ln} v_n / (1 - E_n) \Phi_n\right]^{1-\sigma}}{\sum_{k \in \mathbb{N}} \left[\tau_{lk} v_k / (1 - E_n) \Phi_n\right]^{1-\sigma}}$$

I write consumer market access CMA_l as follows:

$$CMA_l \equiv P_l^{1-\sigma} = \sum_{k \in \mathbb{N}} \left[\tau_{lk} v_k / (1 - E_n) \Phi_n\right]^{1-\sigma}$$

which represents the trade cost weighted availability of cheap goods. Using the CES expression for expenditure shares, local output is a function of local marginal cost of production and firm market access FMA_n:

$$v_n L_n = \left[v_n / (1 - E_n) \Phi_n\right]^{1-\sigma} FMA_n$$

where firm market access accounts for the trade cost weighted access to other markets accounting for the degree of competition:

$$FMA_n = \sum_{l \in \mathbb{N}} v_l L_l \, \tau_{ln}^{1-\sigma} CMA_l^{-1}$$

Given trade costs are quasi-symmetric, consumer market access can be similarly written as a function of firm market access:

$$CMA_l = \sum_{n \in \mathbb{N}} v_n L_n \tau_{nl}^{1-\sigma} FMA_n^{-1}$$

where consumer and firm market access are proportional to one another as discussed in Donaldson and Hornbeck (2016). By extension, I write a general equation for market access MA_l:

$$MA_l = \overline{U} \times \sum_{k \in \mathbb{N}} \tau_{lk}^{1-\sigma} MA_k^{\frac{\sigma}{1-\sigma}} L_k$$

where market access is a function of the expected utility across counties and the trade cost weighted market access to other counties in the economy.

Using the CES expression for expenditure shares, local output can be expressed as a function of local market access:

$$v_l L_l = \left[v_l / (1 - E_l) \Phi_l \right]^{1-\sigma} MA_l$$

Using the expression for indirect utility and the land market clearing condition, the expression total income simplifies:

$$v_l = \left[\left(\frac{1-\zeta}{1-\theta} \right) \left(\frac{L_l}{\overline{H}_l} \right) \right]^{\frac{1-\zeta}{\zeta}} \overline{U}^{\frac{1}{\zeta}} \, MA_l^{\frac{1}{1-\sigma}}$$

Taken together, the expression for local population can be rewritten as follows:

$$\begin{aligned} \ln L_l = & \left[\frac{(2\sigma - 1)\zeta}{(\sigma - 1)(2\zeta - 1)} \right] \ln MA_l + \left[\frac{(\sigma - 1)\zeta}{2\zeta - 1} \right] \ln[1 - E_l] + \left[\frac{(\sigma - 1)\zeta}{2\zeta - 1} \right] \ln \Phi_l \\ & + [\sigma(1 - \zeta)(2\zeta - 1)] \ln \overline{H}_l + \ln \left[\left(\frac{1-\zeta}{1-\theta} \right)^{\frac{1-\zeta}{\zeta}} \overline{U}^{\frac{1}{\zeta}} \right] \end{aligned}$$

where the size of counties are held constant. The effect of changes in market access, holding utility and productivity constant, on population is:

$$\frac{\partial \ln L_l}{\partial \ln MA_l} = \left[\frac{1}{1 + \left[\frac{(\sigma-1)\zeta}{2\zeta-1}\right]\left[\frac{E_l}{1-E_l}\right]\frac{\partial \ln E_l}{\partial \ln L_l}}\right]\left[\frac{(2\sigma-1)\zeta}{(\sigma-1)(2\zeta-1)}\right]$$

Alternatively, taking changes between equilibria:

$$\Delta \ln L_l = \left[\frac{(2\sigma-1)\zeta}{(\sigma-1)(2\zeta-1)}\right]\Delta \ln MA_l + \left[\frac{(\sigma-1)\zeta}{2\zeta-1}\right]\Delta \ln[1-E_l] + \left[\frac{(\sigma-1)\zeta}{2\zeta-1}\right]\Delta \ln \Phi_l + \left[\frac{1}{\zeta}\right]\Delta \ln \overline{U}$$

where county size is held constant and is differenced out. By extension, a regression to identify log changes in population can be written in the following manner:

$$\Delta \ln L_l = \alpha + \beta\, \Delta \ln MA_l + \phi\, \Delta \ln[1-E_l] + \epsilon_l$$

where α represents the change in aggregate utility and the error term represents unobserved changes in local productivity.

D.3 Population Density and Self-Employment

The elasticity of the self-employment to salaried employment with respect to the relative price of land is:

$$\frac{\partial \ln\left[\frac{E_n}{1-E_n}\right]}{\partial \ln L_n/\overline{H}_n} = -\left[\frac{r_n\rho_n}{w_n + r_n\rho_n}\right]\left[\frac{\partial \ln r_n/w_n}{\partial \ln L_n/\overline{H}_n}\right]$$

Using the land and labor market clearing conditions, the relative price of land is:

$$\frac{r_n}{w_n} = \left[\frac{1-\zeta}{\zeta}\right]\left[\frac{\sigma}{\sigma-1}\right]\left[\frac{1}{\theta}\right]\left[\frac{(1-E_n)L_n}{\overline{H}_n}\right]$$

The elasticity with respect to population density:

$$\frac{\partial \ln r_n/w_n}{\partial \ln L_n/\overline{H}_n} = 1 - \left[\frac{E_n}{1-E_n}\right]\left[\frac{\partial \ln E_n}{\partial \ln L_n/\overline{H}_n}\right]$$

where I relate the elasticity of salaried employment to the original expression:

$$\frac{\partial \ln\left[\frac{E_n}{1-E_n}\right]}{\partial \ln L_n/\overline{H}_n} = \left[\frac{1}{1-E_n}\right]\frac{\partial \ln E_n}{\partial \ln L_n/\overline{H}_n}$$

The elasticity of the self-employment rate simplifies to the following expression:

$$\frac{\partial \ln E_n}{\partial \ln L_n/\overline{H}_n} = -\left[\frac{r_n \rho_n}{w_n + r_n \rho_n}\right]$$

Using the assumption $\rho_l = \theta \overline{H}_l / E_l L_l$, the elasticity simplifies further:

$$\begin{aligned}
\frac{\partial \ln E_n}{\partial \ln L_n/\overline{H}_n} &= -\left[\frac{r_n \rho_n}{w_n + r_n \rho_n}\right] \\
&= -\left[1 + \left(\frac{E_l}{1 - E_l}\right)\left(\frac{\zeta v_l L_l}{r_l \theta \overline{H}_l}\right)\right]^{-1} \\
&= -\left[1 + \left(\frac{\sigma - 1}{\sigma}\right)\left(\frac{\zeta}{1 - \zeta}\right)\left(\frac{1 - \theta}{\theta}\right)\left(\frac{E_l}{1 - E_l}\right)\right]^{-1}
\end{aligned}$$