Railroad Barons and American Economic Growth:

A Modified Market Access Approach*

C Adam Pfander
University of Colorado, Boulder
Contact: charles.pfander@colorado.edu

Updated: May 24, 2024
Most recent version available here.

Abstract

This paper seeks to understand how freight market concentration changes the impact of domestic market access improvements throughout the late 19th and early 20th Century. I modify a multi-sector Ricardian trade model to allow for endogenous transport costs set by a profit-maximizing freight monopolist. This framework yields a modified market access term that accounts not only for exogenous bilateral frictions, but endogenous bilateral markups. Reduced-form analysis reveals that omitting endogenous freight prices from the calculation of market access reduces the estimated impact by as much as half. I further document marked convexity in this treatment effect by initial output: smaller, remote counties stand the most to gain from market access improvements, which also means these counties are the most at-risk from the exercise of freight market power. I show that, over time, the most-affected counties move west along the American Frontier.

*I would like to thank my adviser, Jeronimo Carballo, for his constant guidance, support, and patience. I also give thanks to the team behind the R Project for statistical computing (R Core Team, 2024), as well as the authors of the following packages: magrittr (Bache and Wickham, 2022), data.table (Barrett et al., 2024), doParallel (Microsoft and Weston, 2022a), foreach (Microsoft and Weston, 2022b), fixest (Bergé, 2018), fst (Klik, 2022), DescTools (Signorell, 2024), stringr (Wickham, 2023), haven (Wickham et al., 2023a), stringr (Wickham, 2023), and RPostgres (Wickham et al., 2023b). I also acknowledge the creators of and contributors to PostgreSQL. All remaining errors are my own.

1 Introduction

Throughout the late 19th Century, as the United States undergoes rapid economic growth and transformation, railroads web their way through the North American continent, becoming the dominant form of transport and uniting previously disparate locations with a newfound global marketplace. These railroads – along with the less cutting-edge wagon trains and inland canal network – are the gatekeepers to the international economy; however, anecdotal evidence from this time period suggests that freight markets are far from competitive. To what extent is U.S. economic growth hampered by a concentrated domestic freight market during this time period? This paper provides upper-bound estimates of total production losses and price increases stemming from the concentration of freight market power in the late-19th and early-20th Century.

I analyse county-level agricultural production in the United States from 1850 to 1910; this span is notable for – among other reasons – the rapid proliferation of railroads. Prior to the advent of the railroad, U.S. freight transportation relies primarily on wagons, which are relatively unreliable, slow, and costly; a limited Canal network is built in the early 19th Century to ease freight movement along major arteries, but these are geographically confined by proximity to water sources. The first commercial railroad opens in the U.S. in 1826; as of 1830, the national railroad network comprises only 61 miles of operable track (Atack, 2016). By 1850 – the start of my analysis – the rail network grows to a sparse patchwork totalling approximately 8,500 miles throughout the East coast and Midwest; by 1870, the rail network comprises about 35,000 miles, unites most major urban centers in the Northeast, Midwest, and upper South, and extends as far West as the Rocky Mountains. The total length of the rail network roughly doubles over each of the subsequent two decades, reaching approximately 152,000 miles by 1890, and includes multiple trans-continental routes uniting the East and West coasts. By the end of my analysis in 1910, the total length of the rail network is just shy of 200,000 miles (Atack, 2016).

Concurrent with this railroad boom comes rapid economic development. Figure 1 documents the growth of nominal per-capita U.S. manufacturing and agricultural output from 1850 to 1910, as well as the expansion of the transit network. Importantly, as of 1850 when the railroad is still in its infancy, manufacturing output is roughly level with agricultural production; though both sectors grow substantially over the subsequent decades, manufacturing significantly outpaces agriculture, such that the manufacturing output is roughly 3.5 times larger than agricultural output by 1910 (Michael R. Haines and ICPSR, 2010). The rail network follows a similar, rapid expansion, though the canal network stagnates. Previous analyses of this era have addressed to what extent the railroad boom enabled this rapid industrial expansion (Fogel, 1964; Donaldson and Hornbeck, 2016; Hornbeck and Rotemberg, 2024). I contribute to this literature by focusing on freight market power: I evaluate the extent to which horizontal concentration in the market for freight transportation shaped America's economic expansion.

My focus on freight market power is vital to my historical setting. In addition to rapid economic and infrastructure development, this period of history is also notable for the inception of U.S. antitrust policy: the landmark U.S. antitrust law, the Sherman Antitrust Act, passes Congress in 1890. One of the first cases brought under the Act, United States v. Trans-Missouri Freight Association, 166 U.S. 290 (1897), alleges that three railroad operators had entered into a price-fixing agreement for freight running West of the Mississippi and Missouri Rivers; the U.S. Supreme Court agrees and orders the dissolution of the cartel. Just one year later, the the Supreme Court disbands a similar price-fixing scheme among 31 carriers transporting freight between Chicago and the East Coast (United States v. Joint Traffic Association, 171 U.S. 505, 1898). These two legal cases underscore a lack of effective competition in freight markets during this time period, especially amongst the railroad; however, the economic literature has not yet explored the long-run consequences of this freight market concentration.

To evaluate the impact of non-competitive freight pricing, I modify a multi-sector Ricardian trade model to include endogenous transport costs. Specifically, I assume that a representative freight-service provider sets prices to maximize profit, subject to exogenous bilateral frictions (i.e., the distance between locations), as well as demand for transport services (i.e., trade flows). Trade flows, in turn, are a function of these freight prices. This endogenous freight-pricing framework contrasts with the approach typically taken in the trade literature, which assumes purely exogenous bilateral costs. My theoretical approach is attractive for two, key reasons: i) it permits a wide array of solution concepts, including the extremes of perfect competition and perfect oligopoly (monopoly) in the freight market; and ii) it is a generalization of a wide range of canonical trade models (Eaton and Kortum, 2002; Costinot et al., 2011), and perhaps most importantly, the models used in Donaldson and Hornbeck (2016) and Chan (2022) to analyse the historical impact of the railroads; hence, my estimates are immediately comparable to previous estimates in the literature.

I utilize my theoretical framework to develop a modified "market access" term. This term summarizes the demand for goods out of any one county, as determined by exogenous productivities and bilateral frictions. My formulation of market access not only reflects exogenous distances between locations, but also the endogenous markup set by a monopolistic transporter. This market access framework is used widely in the literature to understand the influence of trade frictions on local production (Donaldson and Hornbeck, 2016; Chan, 2022; Hornbeck and Rotemberg, 2024). I estimate that the effect of market access expansion is nearly double when accounting for freight market power. Hence, existing estimates in the literature constitute the lower-bound. Additionally, I confirm that market access expansions lead to substantial increases in the total volume of agricultural production; I further disentangle this estimate into a price and quantity effect, which have distinct welfare considerations. I also find substantial convexity in this treatment effect by initial output size; the smallest producers stood the most to gain from the expansion of market access in the 19th Century. However, this finding also means that these smaller producers are the most vulnerable to the exercise freight market power. I show that the distribution of losses from freight market concentration is highly skew right – those counties that are worst affected tend to be along the American Frontier.

This paper contributes to a long-standing literature estimating the contribution of railroads to American economic growth. First amongst this strand of literature is Fogel (1964), who utilizes a "social saving methodology" to argue that the impact of the railroads is actually quite limited; in the absence of railroad expansion, Fogel contends that the canal network would have provided comparable transport services at only slightly elevated prices. Responding to this conclusion some years later, Donaldson and Hornbeck (2016) utilize a "market access" approach – a reduced-form expression derived from equilibrium trade theory – to argue that railroads contributed significantly to U.S. agricultural output in the late 19th Century; these losses could not be overcome by other modes. Importantly, this market access methodology captures the effect of railroad development nationwide on any one county, even if it is not directly connected to the network; moreover, the estimates from this method are well-identified. I build upon this framework to estimate the causal impact of freight market concentrations on domestic agricultural production.

This notion of market access is key to my analysis and thus merits some exposition. In short, it is the weighted-sum of total expenditure across all available markets; the weights are inversely proportional to bilateral trade costs. Hence, market access reflects the global (or in my case, national) demand for goods from any one county. Importantly, this term captures both the direct and indirect effect of changes to trade frictions: a new rail line built between, e.g., Denver and Los Angeles will undoubtedly increase trade between these two cities; this trade will, in turn, make these two cities wealthier, and thus increase market access for all cities nationwide, even if they are not connected to this new rail line. This latter effect is critical to identifying the causal impact of changes to market access on agricultural output: changes to transit costs throughout the nation (due to, e.g., expansion of the railroad network or changes to the competitive landscape for freight services) will affect the market access, meaning changes to market access are largely exogenous to local economic conditions. Thus, the railroad boom of the 19th century offers a natural experiment to evaluate how changes in market access affect county-level agricultural production. In Section 4.2, I detail this identification argument more completely.

A number of papers utilize similar measures of market access to evaluate the economic impacts of rapid expansion of the railroad. Chan (2022) utilizes a market access framework to asses how the 19th Century railroad boom influenced U.S. agricultural production; they find that railroads sparked greater total agricultural output via increased agricultural land use and population growth, but had limited to null effects on productivity and the variety of crop output. Hornbeck and Rotemberg (2024) utilize a similar empirical approach to analyse how the railroad affected county-level manufacturing productivity; they find that the railroads greatly increased allocative efficiency, but found limited effects on total factor productivity. Hence, one of the primary mechanisms, through which the railroad increased economic output was by lessening input-output distortions.

I contribute to this literature by focusing on the role non-competitive freight pricing. I utilize a market access framework, but allow for transport costs to be set endogenously under familiar structural assumptions. This theoretical approach generates upper-bound estimates of how U.S. agricultural production was hampered by non-competitive freight pricing; no paper to-date examines the influence of freight market concentration on a national scale, nor in this historical context. Additionally, I build upon this market access methodology to separately identify the effects of changes in market access – brought about by either the expansion of the railroad or, as I focus on, counterfactual changes to the competitive landscape for freight services – on both agricultural prices and the quantity of agricultural output. Previous estimates in the literature focus on the total value of production (Chan, 2022; Hornbeck and Rotemberg, 2024), or on the value of agricultural land (Donaldson and Hornbeck, 2016). This paper is the first to separately identify these price and quantity effects, which have distinct implications for consumer welfare.

A number of other papers have explored the historical significance of transport and infrastructure. Among this literature, Jaworski et al. (2023) develop a Ricardian trade model with endogenous congestion and varying port efficiency to quantify the economic impacts of the U.S. interstate highway system. Taking a more reduced-form approach, Donaldson (2018) explore the economic impact of the British railway system in colonial India. Both of these papers highlight the importance of these national and international transport systems for economic integration, especially in rural areas. These papers highlight that transportation infrastructure is a critical tool for economic development and prosperity. However, these papers do not consider the role of imperfect competition in the freight services sector. The welfare impacts of these large infrastructure projects may be muted by non-competitive freight rates; this paper seeks to set an upper bound on these losses in the case of North America.

More broadly, the theoretical model posited in this paper contributes to an expanding literature that embed endogenous trade costs into canonical trade models. Recent examples of this literature include Allen and Arkolakis (2022), who embed a model of trade costs with endogenous congestion exeternalities into a neoclassical trade model, and Fuchs and Wong (2023), who then adapt this model to span multiple modes, and explore how congestion at ports spreads throughout the domestic transit network. I contribute to this strand of literature in two distinct ways: i) I focus on the role of horizontal concentration in the market for freight services, emphasizing the role that non-competitive freight rates play in determining domestic trade flows, and ii) I analyse production in the late 19th Century, a time of rapid economic transformation, as well as noted freight market concentration. This paper is also closely related Brancaccio et al. (2020), who examine the market for international sea shipping, and are the first to allow market power on behalf of freight carriers. In contrast to this model, I focus on the market for domestic freight spanning multiple modes, moreover I focus on this historical consequences of this freight market concentration, during a time when freight markets were far from competitive.

The rest of the paper proceeds as follows. Section 2 develops my theoretical model, which will inform a method to identify markups absent data on trade flows or freight prices. Section 3 outlines my data sources and develops a strategy to simulate bilateral trade flows and freight rates absent available data. Section 4 evaluates my strategy to identify the effect of changes in market access on output and prices; it also describes how counterfactual shifts in domesitc production. Section 5 concludes.

2 Theoretical Framework

I this section, I construct a theoretical framework to analyse the influence of freight market power on domestic production. First, I adapt a well-known multi-sector Ricardian trade model (Costinot et al., 2011) to allow endogenous freight pricing subject to exogenous bilateral frictions – namely, distance. Second, I model the transporter's pricing decision; utilizing long-standing insights from the empirical Industrial Organization literature, I develop a pricing rule that permits a continuum of solution concepts, including the extremes of perfect oligopoly (monopoly) and perfect competition.[1] This theoretical approach contrasts with standard practice in the trade literature, which typically assumes purely-exogenous bilateral frictions. I now delve into the details of the model.

2.1 Transport Demand: Equilibrium Trade Flows

The demand for freight transport between an origin and destination is given by the flow of goods. To analyse these trade flows, I utilize a multi-sector Ricardian trade model with no input-output linkages, as in Costinot et al. (2011).[2] This trade setting is critical to my analysis for two reasons: i) the model yields a log-linear expression for trade flows, which I will later exploit to derive an estimable equation; and ii) it provides a tractable structure to evaluate the impact of non-competitive freight pricing on total output, welfare, and population in each location (in my context, counties). Finally, let t index the year (decade) of observation. It is worth highlighting that, though my data span multiple time periods, I do not model savings or investment; each time period is thus completely independent.

I first establish some preliminaries. Let $\mathcal{K}$ denote a discrete, finite set of commodities (sectors), which will be indexed by $k \in \mathcal{K}$. Trade in these commodities occurs amongst a discrete, finite set of locations, $\mathcal{S}$; let $i, j \in \mathcal{S}$ index origins, destinations. Within each commodity, agents consume a continuum of varieties Ω^k; for each variety $\omega^k \in \Omega^k$, agents consume a quantity $q(\omega^k)$ to maximize

[1] I utilize much the same framework as Pfander (2024); however, I make the following modifications to make the model more appropriate to my empirical setting: i) I utilize a multi-sector trade model to capture the production and trade of distinct agricultural goods; ii) I simplify the routing structure such that I consider only the least-cost path between two locations along any one mode; and iii) I simplify the correlation structure between competing trade routes. These latter two simplifications are necessary due to a lack of bilateral trade data. See Section 3.1 for further detail.

[2] While the assumption of zero input-output linkages is strict, I contend that it is appropriate for my empirical setting. Specifically, I analyse the production of a distinct set of agricultural commodities, which do not generally exhibit any input-output structure. See Table 2 for detail.

utility given by

$$U_{jt} = \sum_{k \in \mathcal{K}} \left(\left(\int_{\omega^k \in \Omega^k} q(\omega^k)^{\frac{\sigma^k - 1}{\sigma^k}} d\omega \right)^{\frac{\sigma^k}{\sigma^k - 1}} \right)^{\zeta^k} \tag{1}$$

where $\sum_{k \in \mathcal{K}} \zeta^k = 1 \; \forall j$.

Locations vary in their productivity, creating incentive to trade. Let the marginal cost of producing variety $\omega^k \in \Omega^k$ in location i be given by

$$c_{it}(\omega^k) = \frac{c_{it}}{z^k_{it}(\omega^k)} \tag{2}$$

where c_i is a uniform input cost and $z_i(\omega^k)$ is a Hicks-neutral productivity shifter. I assume that $z^k_{it}(\omega^k) \sim$ Fréchet with location parameter A^k_{it} and shape parameter θ.[3] The former parameter represents location i's productivity and governs absolute advantage; the latter parameter dictates the dispersion of these productivities across locations, and it governs comparative advantage (Costinot et al., 2011). I further impose perfect competition in the goods market such that the price paid for ω^k in market j is equal to its marginal cost:

$$p_{jt}(\omega^k) = \min_{i \in \mathcal{S}} \left\{ c_{it}(\omega^k) \bar{\tau}_{ijt} \mu^k_{ijt} \right\} \tag{3}$$

where $\bar{\tau}_{ijt}$ denotes the expected, exogenous, ad-valorem cost of travelling from i to j, and μ^k_{ijt} is an endogenous markup set by a representative transporter. In Section 2.2, I further elaborate this markup term; for the purpose of deriving equilibrium transport demand (i.e., trade flows), I treat it as fixed. It follows that the share of commodity k travelling to j from i is given by

$$\pi^k_{ijt} = A^k_{it} \left(c_{it} \bar{\tau}_{ijt} \mu^k_{ijt} \right)^{-\theta} \left(\phi^k_{jt} \right)^{-1} \tag{4}$$

where $\phi^k_{jt} = \sum_{i' \in \mathcal{S}} A^k_{i't} \left(c_{i't} \bar{\tau}_{i'jt} \right)^{-\theta}$ governs the price index in location j.[4] Equation 4 constitutes the well-known gravity expression for trade flows between i and j (Costinot et al., 2011).

Exogenous bilateral trade costs, $\bar{\tau}_{ijt}$, are incurred while traversing the transit network. I follow what has become standard practice in the literature and assume a logit model of mode-choice

[3] As in Costinot et al. (2011), I require $\theta > \sigma^k - 1 \; \forall k$.

[4] Explicitly, the price index for a given commodity in market j is given by $P^k_{jt} = \Gamma \left(\frac{\theta + 1 - \sigma^k}{\theta} \right)^{\frac{1}{1 - \sigma^k}} \left(\phi^k_{jt} \right)^{-1/\theta}$. This formulation follows immediately from the utility assumption in Equation 1.

(Allen and Arkolakis, 2014; Donaldson and Hornbeck, 2016; Donaldson, 2018); agents choose the cheapest alternative among a set modes, whose cost is determined by the available infrastructure. A logit formulation for modal shares is attractive because it yields a reduced-form expression for expected trade costs that nests neatly into Equation 4. However, this model does impose some strict restrictions on demand for freight services – most notably the "independence of irrelevant alternatives" assumption, which requires that, upon the introduction of a new mode like the railroad, trade shares among existing modes remain proportional. This assumption likely fails, as railroads are more closely substitutable with canals than with roadways, as highlighted by (Fogel, 1964). I now expound the structure of this logit formulation.

Let $\mathcal{M}_{ijt}$ denote the set of modes serving a particular origin-destination pairing.[5] Let the cost of travelling between i and j via m be given by $\tau^m_{ijt} = \exp\left(Miles^m_{ijt}\beta^m + f^m + \epsilon^m_{ij}\right)$, where β^m denotes the cost per-mile, and f^m is a route-invariant, mode-specific cost. Intuitively, β^m captures the cost incurred en-route (e.g., fuel, labor, and time), while f^m captures "fixed" costs (e.g., loading, unloading, and maintenance).[6] The variable $Miles^m_{ij}$ denotes the distance of the least-cost path between i and j via m.[7] Further assume that $\epsilon^m_{ij} \sim$ Gumbel (Type-1 EV) with shape parameter ρ. From this setup, expected trade costs are

$$\bar{\tau}_{ijt} = \frac{1}{\rho}\Gamma\left(\frac{1}{\rho}\right) V_{ijt}^{-\frac{1}{\rho}}, \quad \text{where} \tag{5}$$

$$V_{ijt} = \sum_{m \in \mathcal{M}_{ijt}} \exp\big(-\rho(\beta^m Miles^m_{ijt} + f^m)\big) \tag{6}$$

and Γ denotes the gamma function.[8] Combining Equations 4 and 5 yields:

$$\pi^k_{ijt} = \kappa_1 A^k_{it}\left(c_{it}\mu^k_{ijt}\right)^{-\theta} V_{ijt}^{\theta/\rho}\left(\phi^k_{jt}\right)^{-1} \tag{7}$$

where $\kappa_1 = \left(\rho^{-1}\Gamma\left(\rho^{-1}\right)\right)^{-\theta}$. Thus, Equation 7 reformulates the gravity equation described in Equation 4 to be a function of available infrastructure at time t.

[5] Recall that not all locations are served by every mode; the canal network is geographically confined by the availability of major water sources; the rail network is only a loose patchwork at the start of my analysis (1850). See Figure 2 for further detail.

[6] A well-known feature of logit models is that they are only identified up to scale. I thus follow Donaldson and Hornbeck (2016) and Donaldson (2018) and normalize $f^{Road} = 0$; this normalization has intuitive appeal, as it is easy to imagine that road transport requires minimal loading, unloading, and maintenance relative to barges and/or rail.

[7] Details of generating these least-cost paths are reported in Section 3.2.

[8] Note that this formulation is identical to the expected trade costs posited in Allen and Arkolakis (2014).

2.2 Freight Prices: Transport Supply

In this section, I elaborate the endogenous, bilateral markup, μ^k_{ijt}. A representative transporter provides all freight transport services to a destination j.[9] This transporter's total profit is given by $\Pi_{jt} = \sum_{i\neq j}\sum_k \zeta^k B_{jt}\pi^k_{ijt}(\mu^k_{ijt} - 1)\bar{\tau}_{ijt}$, where B_{jt} is the total expenditure of market j, which the transporter treats as exogenous. It is evident from this formulation that the transporter makes zero profit on domestic trade; this normalization pins-down the scale of total freight profits. Maximizing profit with respect to μ^k_{ijt} yields the following pricing rule:

$$\mu^k_{ijt} = \frac{(1-\pi^k_{ijt})\bar{\tau}_{ijt} + \sum_{i'\neq i,j}(\mu^k_{i'jt} - 1)\bar{\tau}_{i'jt}\pi^k_{i'jt}}{(1-\pi^k_{ijt})\bar{\tau}_{ijt} - (\lambda/\theta)\bar{\tau}_{ijt}} \tag{8}$$

where $\lambda \in [0, 1]$ dictates the state of competition for freight services nationally. A detailed derivation of this pricing rule is presented in Appendix 1. This formulation is based on long-standing insights from the empirical Industrial Organization literature (Bresnahan, 1982); it is attractive because it does not assume a particular solution to the transporter's pricing game, but permits a continuum of equilibria, as determined by λ. This parameter is immediately interpretable at the extremes — as displayed by Equation 8, setting $\lambda = 1$ yields the monopoly pricing rule, while $\lambda = 0$ corresponds to perfect competition; intermediate values of λ correspond to intermediate solution concepts. Hence, I develop flexible pricing rule that permits a continuum of potential pricing regimes.

While this formulation is mathematically appealing, it is worth discussing its limitations. For one, it is agnostic about the source of market power – it cannot distinguish illegal price collusion from a natural monopoly. As stated in the Introduction, there is substantial anecdotal evidence of freight cartels during this time period; however, transport is also a high fixed-cost industry, naturally yielding some pricing power to existing carriers. The model is thus somewhat limited in terms of policy prescriptions.[10] Relatedly, I do not model long-term investments; the transporter's profit is completely static. However, infrastructure investments constitute a significant share of a transporter's total costs. Modelling network creation is a non-trivial exercise that is beyond the

[9]This setup is a minor departure from the markup model used in Pfander (2024); in that paper, I assume that the representative transporter is mode-specific. However, in this historical context, I cannot estimate the state of freight market competition due to a lack of bilateral flow data (see Section 3.1 for detail). I thus fix markups to their theoretical maximum to derive theoretically tractable upper-bound estimates.

[10]While the model is agnostic about the source of market power in its current formulation, it could easily be adapted to distinguish between illegal collusion and natural barriers; this more flexible formulation would require additional data on fixed costs and/or firm entry and exit to be of empirical value.

scope of this paper.[11] Expanding the model to include network creation is a promising avenue for future research; for now, I treat the network as exogenously given. Finally, the model assumes that prices are the only avenue of competition among freight service providers. In reality, transporters may compete along numerous dimensions – e.g., safety, reliability, speed, and/or volume, among numerous other potential dimensions. While these alternate forms of service are not explicitly modeled in my framework, they are captured to the extent that they are capitalized into aggregate trade costs.

3 From Theory to Empirics

A non-trivial challenge of my historical analysis is that I do not observe bilateral trade flows, which would otherwise enable estimation of the gravity equation (Equation 7). Importantly, absent this flow data, or any other data on freight prices, I do not observe freight markups. I do, however, observe county-level production of distinct commodities, as well as total farm output, manufacturing production, population, and the farm value (land plus equipment and buildings). I thus exploit the structure of the equilibrium trade model developed in Section 2.1 to derive an equation, which permits estimation of my model fundamentals. With these fundamentals, I may simulate trade flows, and in turn, markups subject to an assumption of freight market conduct. In this section, I discuss my data sources, elaborate my procedure to estimate model fundamentals, and finally, elaborate a procedure to simulate endogenous bilateral frictions in the absence of trade or freight price data.

3.1 Data

I utilize county-level agricultural production data from the U.S. Census of Agriculture from 1850 through 1910 (Michael R. Haines and ICPSR, 2010). These data provide snapshots of agricultural production in distinct commodities across every U.S. county at the start of each decade. The set of commodities included in the Census expands over time; I clean and harmonize production data for 15 of them, listed in Table 2. It is also important to note that U.S. counties are not static throughout my analysis: counties shift their borders, and the total number of counties grows as more counties and states incorporate. I follow Donaldson and Hornbeck (2016) to hold county bor-

[11] Fajgelbaum and Schaal (2020) tackle optimal network formation in the context of a neoclassical trade model; however, their notion of "optimal" focuses on consumer welfare. In reality, transport firms build networks in response to a profit motive, with likely limited attention paid to consumer welfare.

ders static to their 1880 limits; I adjust the population and acreage measurements accordingly.[12] Finally, in addition to these commodity-specific production series, I utilize three aggregate variables per county, which will be used to estimate production productivities and input costs: the total number of improved acres in farmland, the total (nominal) value of farm equipment, and the total value of farmland (inclusive of land, capital, and buildings).

In addition to this agricultural data, I incorporate population and output data for each county from the decennial Censuses. Specifically, I observe total population, urban population (defined as cities with greater than 25,000 citizens), the total (nominal) value of manufacturing output, and the total (nominal) value of farm output. These aggregate data will again inform county-level estimates of productivity and input costs in each time period.

My other main data source is historical records on the expansion of the rail and waterway networks in the mainland U.S. through the latter half of the 20th Century. Specifically, I utilize historical rail and inland waterway (canals and navigable rivers) shapefiles from Atack (2015, 2016, 2017). Importantly, these data record not only the location of rail lines and waterways, but also report their first year of operation, as well as their year of closing where relevant. From this data, I create a snapshot of the national transportation network for each decade reported in the Census of Agriculture. Historical records dating back to 1850 do not exist for the road network, so I connect all county centroids within a 300km radius with a straight line; these roads are assumed available every decade. Finally, I augment this inland transport network with Sea and Lake routes from Donaldson and Hornbeck (2016); as with the road network, I assume that these lines are available throughout my analysis period. Full detail of the transport network from 1850 through 1910 is provided in Figure 2.

3.2 Generating Distances in the Transport Network

As highlighted by Equation 6, a key input into the model is the distance along the least-cost path between two counties i and j via a given mode m. To generate these distances, I first calculate the (exogenous) cost of traversing the network. I follow Donaldson and Hornbeck (2016) and assume the relative cost structure displayed in Table 1. Note that the "fixed" cost of roadway travel is normalized to zero, so all f^m parameters are in reference to the roadway.[13] Moreover, the

[12]Specifically, I assume that improved farm acres and rural population are uniformly distributed across the county; the total adjustment to acres and population equals the total change in size of the county.

[13]While likely greater than zero, it seems reasonable to assume that roadway travel incurs the lowest "fixed" cost per mode: unlike rail, ships, or barges, which require travel to stations/ports and must be loaded completely before getting

per-mile cost of traversing the road is normalized to 1, so all cost parameters (both β^m and f^m) are in road-mile equivalents. Note that, absent bilateral trade data, I cannot estimate these cost parameters, and so hold them fixed. This parameterization further imposes that exogenous costs are symmetric.[14] I apply this cost structure to my historical snapshots of the transport network at the start of each decade; I thus observe the exogenous portion of trade costs.

Given this detailed topography of transit costs throughout the U.S., I calculate the least-cost distance between counties along each mode using Dijkstra's algorithm. Because I do not observe the distribution of economic activity within counties, I calculate the distance between county centroids.[15] All centroids are connected to the road network by construction (see Section 3.1 for detail); freight originating at any centroid within 50 kilometers (approximately 31 miles) of a rail or water line may immediately travel via these networks. Freight coming from farther afield must first travel via the road network to reach one of these more proximate counties. Fixed costs are incurred at the start and end of journey, and when switching between modal networks.[16] I thus generate ad-valorem, exogenous transit costs along every mode between every county pair at the start of each decade.

3.3 Transforming the Gravity Equation

I now develop a framework to simulate bilateral trade flows, and consequently, freight markups. The overall goal is to isolate an expression for trade flows, as well as aggregate production of a commodity k, in quantities as opposed to value; the necessity of this theoretical exposition stems from not observing commodity-specific prices or values in my data. Importantly, the structure elaborated here follows immediately from the trade model developed in Section 2.1; I do not impose any further assumptions.

From Equation 1, it follows that demand for variety ω^k in location j is given by

$$q_{jt}(\omega^k) = \zeta^k B_{jt}\ P_{jt}^{k\ (\sigma^k-1)}\ p_{jt}(\omega^k)^{-\sigma^k} \tag{9}$$

underway, wagons can be loaded at the factory gates relatively quickly.

[14]This assumption, while standard in the trade literature, is likely inaccurate when considering, e.g., asymmetric congestion (Allen and Arkolakis, 2022; Fuchs and Wong, 2023) or elevation (Sun, 2024). However, my model allows for asymmetric trade costs due to bilateral markups set by the transport sector; see Section 2.2 for detail.

[15]This methodology contrasts with the routing logic utilized Donaldson and Hornbeck (2016), who calculate routes from multiple uniformly-distributed points within each county to generate a mean routed distance for each origin-destination pairing. This approach is undoubtedly more accurate but significantly more cumbersome; my approach is appealing for both its simplicity and its computational feasibility.

[16]The only trans-shipment I allow is for freight to switch between the road and either the rail or water networks. I do not allow multi-modal movements that incorporate both rail and water.

where P_{jt}^k is the CES price index. Aggregating demand across varieties yields the following expression for the total quantity of k consumed in j:

$$\begin{aligned} Q_{jt}^k =& \int_{\Omega^k} q_{jt}(\omega^k) d\omega^k \\ =& \kappa_2^k B_{jt} \; \phi_{jt}^{k\ (1/\theta)} \end{aligned} \tag{10}$$

were $\kappa_2^k = \zeta^k \Gamma\left(\frac{\theta - \sigma^k}{\theta}\right) / \Gamma\left(\frac{\theta - \sigma^k + 1}{\theta}\right)$. It follows that the total quantity of k travelling from i to j is given by

$$\begin{aligned} Q_{ijt}^k =& Q_{jt}^k \pi_{ijt}^k \\ =& \kappa_3^k A_{it}^k \left(c_{it} \mu_{ijt}^k\right)^{-\theta} V_{ijt}^{\theta/\rho} \left(\phi_{jt}^k\right)^{\frac{1-\theta}{\theta}} B_{jt} \end{aligned} \tag{11}$$

Total production in i is thus

$$\begin{aligned} Q_{it}^k =& \sum_j Q_{ijt}^k \\ =& \kappa_3^k A_{it}^k c_{it}^{-\theta} \underbrace{\sum_j \left(\mu_{ijt}^k V_{ijt}^{-1/\rho}\right)^{-\theta} \left(\phi_{jt}^k\right)^{\frac{1-\theta}{\theta}} B_{jt}}_{\text{“Market Access”}} \end{aligned} \tag{12}$$

where $\kappa_3^k = \kappa_1 \kappa_2^k$. Note that the summation in Equation 12 constitutes a modified market access term; it captures the global (or in my case, national) demand for good k out of i. In contrast to other formulations of market access in the literature (Donaldson and Hornbeck, 2016; Chan, 2022; Hornbeck and Rotemberg, 2024), I measure demand in terms of quantity, not value. To understand market access intuitively, it is helpful to re-write Equation 12 as

$$Q_{it}^k = \kappa_4^k A_{it}^k c_{it}^{-\theta} \sum_j \left(\frac{\mu_{ijt}^k \bar{\tau}_{ijt}}{P_{jt}^k}\right)^{-\theta} \left(\frac{B_{jt}}{P_{jt}^k}\right) \tag{13}$$

where $\kappa_4^k = \zeta^k \Gamma\left(\frac{\theta - \sigma^k}{\theta}\right) \Gamma\left(\frac{\theta - \sigma^k + 1}{\theta}\right)^{\frac{1}{\theta}}$. It is evident from this equation that market access is a weighted sum of real expenditure across all available destinations; the weights are inversely related to bilateral trade costs (exogenous transit costs and markups), discounted by the price level in the destination county j.

Finally, it will prove expedient to define the share of exports from i travelling to j, which may be found by combining Equations 11 and 12:

$$\begin{aligned}\tilde{\pi}_{ijt}^{k} =& \frac{Q_{ijt}^{k}}{Q_{it}^{k}} \\ =& \frac{\left(\mu_{ijt}^{k} V_{ijt}^{-1/\rho}\right)^{-\theta} \left(\phi_{jt}^{k}\right)^{\frac{1-\theta}{\theta}} B_{jt}}{\sum_{j' \in \mathcal{S}} \left(\mu_{ij't}^{k} V_{ij't}^{-1/\rho}\right)^{-\theta} \left(\phi_{j't}^{k}\right)^{\frac{1-\theta}{\theta}} B_{j't}}\end{aligned} \tag{14}$$

Note that this trade share formulation differs from Equation 7, as this expresses the share of *exports* (in quantities) out of i that travel to j; in contrast, Equation 7 describes the share of *imports* (in value) into j that originate in i. I utilize the equations elaborated here to develop my estimation framework.

3.4 Simulating Trade Flows, Markups

The final task is to convert the structure expounded in Section 3.3 into an estimable equation. Recall that I observe aggregate production of commodity k in each county i, but do not observe bilateral frictions, trade flows, or commodity-specific prices. Thus, the market-access term in Equation 12 remains unobserved. It is particularly challenging to estimate bilateral markups μ_{ijt}^{k} absent data on freight prices or trade flows. I thus rely on a simulated method of moments strategy to estimate these fundamental parameters. In short, I simulate bilateral trade flows π_{ijt}^{k} to facilitate estimation of Equation 11, which is log-linear and thus provides a convenient framework to estimate my model fundamentals. The goal is to generate estimates of input costs c_{it} and productivities A_{it}^{k} that are consistent with the model structure derived in Section 2.

I emphasize that, absent bilateral data, I cannot estimate all model parameters via this strategy. Explicitly, all parameters that govern bilateral frictions – the modal elasticity parameter ρ and the conduct parameter λ – are not identified by my production data. In the case of the former, I rely on estimates of ρ from Donaldson and Hornbeck (2016), who – broadly speaking – examine the same time period, utilize similar data sources, and also employ a Ricardian trade model to identify the influence of the transportation sector on the domestic economy. The similarity of our approaches suggests that this estimate of the modal elasticity is appropriate. Regarding λ: due to the absence of comprehensive, national price data from the 19th Century, no comparable estimates of freight market conduct exist. Rather than calibrate this parameter from the literature, I simply set it to 1 at the outset. I thus present the upper-bound of markups assuming that freight markups were perfectly concentrated.

My estimation technique utilizes an Expectation-Maximization (EM) algorithm.[17] I utilize this framework to estimate the production parameters A_{it}^k and c_{it}, as well as bi-lateral markups. The process proceeds as follows.

1. From Equation 6, calculate $V_{ijt} = \sum_{m \in \mathcal{M}_{ijt}} \exp\big(-\rho(\beta^m Miles_{ijt}^m + f^m)\big)$.

2. Take an initial guess at the production parameters, $\{A_{it}^{k\,(0)}, c_{it}^{(0)}\}$. Recall that, throughout this analysis, I hold fixed the exogenous cost parameters β^m, f^m, the trade elasticity θ, as well as the modal elasticity parameter ρ and the conduct parameter λ.

3. Utilize a contraction mapping to calibrate trade shares, markups based on the current set of parameter estimates:

   (a) From Equation 7: $\pi_{ijt}^{k\,(0)} = A_{it}^{k\,(0)}\left(c_{it}^{(0)}\mu_{ijt}^{k\,(0)}\right)^{-\theta} V_{ijt}^{\left(\theta/\rho\right)}\left(\phi_{jt}^{k\,(0)}\right)^{-1}$, where the price parameter $\phi_{jt}^{k\,(0)} = \sum_{i'} A_{i't}^{k\,(0)}\left(c_{i't}^{(0)}\mu_{i'jt}^{k\,(0)}\right)^{-\theta} V_{i'jt}^{\left(\theta/\rho\right)}$.

   (b) From Equation 8: $\mu_{ijt}^{k\,(0)} = \frac{\left(1-\pi_{ijt}^{k\,(0)}\right)\bar{\tau}_{ijt} + \sum_{i' \neq i,j}\left(\mu_{i'jt}^{k\,(0)}-1\right)\bar{\tau}_{i'jt}\pi_{i'jt}^{k\,(0)}}{\left(1-\pi_{ijt}^{k\,(0)}\right)\bar{\tau}_{ijt} - (\lambda/\theta)\bar{\tau}_{ijt}}$.

4. Calculate the total quantity travelling between i and j utilizing Equation 14:

$$\begin{aligned} Q_{ijt}^{k\,(0)} =& Q_{it}^k \tilde{\pi}_{ijt}^{k\,(0)} \\ =& Q_{it}^k \left(\frac{\left(\mu_{ijt}^{k\,(0)}\right)^{-\theta} V_{ijt}^{(\theta/\rho)}\left(\phi_{jt}^{k\,(0)}\right)^{\frac{1-\theta}{\theta}} B_{jt}}{\sum_{j' \in \mathcal{S}}\left(\mu_{ij't}^{k\,(0)}\right)^{-\theta} V_{ij't}^{(\theta/\rho)}\left(\phi_{j't}^{k\,(0)}\right)^{\frac{1-\theta}{\theta}} B_{j't}}\right) \end{aligned}$$

   noting that Q_{it}^k and B_{jt} is observed.[18]

5. Utilize Equation 11 to inform a log-linear estimating equation:

$$\left(\text{asinh}\, Q_{ijt}^{k\,(0)} + \theta\, \text{asinh}\, \mu_{ijt}^{k\,(0)} - (\theta/\rho)\, \text{asinh}\, V_{ijt}\right) = \underbrace{\gamma_{s(i)t}^k + f(x_i, y_i)\gamma_t + X_{it}\alpha_t}_{\kappa_3^k + \ln A_{it}^k - \theta \ln c_{it}} + \underbrace{\gamma_{jt}^k}_{\frac{1-\theta}{\theta}\ln \phi_{jt}^k + \ln B_j} + \varepsilon_{ijt}^k \qquad (15)$$

   where asinh denotes the inverse hyperbolic sine transformation,[19] γ denotes the corresponding fixed-effects, $s(i)$ denotes the U.S. state of the originating county, $f(x_i, y_i)\gamma_t$ is a

[17] An EM process has been used to some extent in the economics literature – see, e.g., Bonadio (2021), as well as the parallel paper Pfander (2024)

[18] While I do not directly observe a county's total expenditure, I assume that B_{jt} is proportional to the sum of its manufacturing and agricultural output.

[19] This function, defined as $\text{asinh}\,(x) = \ln\left(x + \sqrt{x^2+1}\right)$ for some real-valued variable x, approximates the natural logarithm but has the advantage of being defined at zero.

cubic polynomial of the centroid's longitude and latitude interacted with a time fixed-effect, X_{it} is a vector of time-varying county characteristics, and finally, ε^k_{ijt} is a structural error term. I will provide further detail on this specification shortly.

Denote the estimated production parameters from this equation $\{A^{k\,(1)}_{it}, c^{(1)}_{it}\}$.

6. Return to step 3 with the new parameter estimates and repeat.

Under assumptions normally required for linear regression model, the parameter estimates from this process converge to the maximum-likelihood estimates, which I will denote $\hat{\pi}^k_{ijt}$, $\hat{\phi}^k_{ijt}$, and $\hat{\mu}^k_{ijt}$.[20]

The regression specification presented in Equation 15 merits some discussion. First, the variation that identifies productivities and input costs stems from differences in aggregate production across commodities, counties, and time. The bilateral variation (i.e., the flow of goods between i and j) is simulated for the sake of making OLS estimation feasible; it adheres strictly to the constructs of the model.[21] To isolate the non-simulated variation, I include destination-by-commodity-by-year fixed effects, γ^k_{jt}, which control for a wide array of demand determinants like preferences for a particular commodity in a particular location, as well as budget constraints and local price levels. I also normalize the quantity by two variables that capture bilateral frictions, $\ln V_{ijt}$ and $\ln \mu^{k\,(0)}_{ijt}$. These variables control for the remaining, simulated bilateral variation. The residual variation I attribute to the productivity and cost parameters A^k_{it} and c_{it}.

The portion of the fixed effects that vary by commodity and year are crucial to identifying the productivity parameters. These fixed-effects remove variation in aggregate production across commodities due to commodity- and year-specific idiosyncrasies, for example, differences in units of measurement[22] or for broad climate conditions that may affect the national production of a particular commodity. They also control for potential differences in bilateral trade costs across commodities. Though not explicitly modeled in the structure developed in Section 2, it is easy to imagine that different commodities incur different transit costs due to, e.g., weight, bulk, or spoilage rates. To the extent that these commodity-specific frictions are multiplicative of mean

[20] Pfander (2024) provides a detailed proof of convergence.

[21] A consequence of this simulated variation is that, were I to estimate a version of Equation 15 including an origin-by-commodity-by-year fixed effect, my regression would perfectly predict quantity flows; the estimated fixed-effect would not only capture productivity and input costs for a given county, but also market access, which is exactly what I aim to isolate via this EM procedure.

[22] To elaborate, production of livestock (cattle and sheep) are measured in individual animals; production of corn is measured in tons; and production of wine is measured in gallons. The commodity fixed-effects control for these differences in units.

trade costs, they are captured by these fixed effects. Hence, my fixed-effects control for a wide array of measurement and/or structural errors.

The parameters A_{it}^k and c_{it}, which collectively govern productivity and input costs in a given location, are jointly estimated by a bevy of time-varying county characteristics X_{it}, as well as time-specific cubic polynomials of the county centroid's longitude and latitude $f_t(x_i, y_i)$ and state-by-commodity-by-year fixed-effects $\gamma_{s(i)t}^k$. Explicitly, X_{it} includes the inverse hyperbolic sine transform of the county's total population, rural population (defined as all those living in areas with fewer than 25,000 residents), nominal dollar-value of farm equipment, total improved farm acreage, and nominal dollar-value of all farm output. Because these productivities and input costs are not directly observed, there is some concern regarding mis-specification of these production parameters. Explicitly, failure to accurately estimate these production parameters will result in error in my calibrated trade shares $\hat{\pi}_{ijt}$, which in turn will bias my estimates of bilateral markups $\hat{\mu}_{ijt}^k$. Absent additional, time-varying data on agricultural productivity or worker's wages, it is difficult to evaluate how well this paramterization fits the data. However, it is worth noting that a similar formulation is used by Chan (2022) to estimate agricultural productivity.

Recall that the goal of this section is to develop a methodology to estimate the fundamental parameters of my equilibrium trade model; with these fundamentals, I may simulate otherwise unobserved bilateral trade flows and the resultant markups. The EM process described above results in well-identified, robust estimates of A_{it}^k and c_{it}; subject to a calibrated modal elasticity and an assumed value of the conduct parameter, I simulate trade flows and freight markups according to the structure laid out in Section 2. I thus arrive at theoretically tractable estimates of freight markups, specified to distinct origin-destination pairs, throughout the latter-half of the 19th Century. Because I set the conduct parameter to 1, my estimates of markups represent the upper-bound. In the next section, I use these upper-bound estimates to assess the maximum impact of freight market concentration on domestic output during this time period.

4 Evaluating the Impact of Transport Pricing on Output

I adopt a relatively simple, reduced-form framework to evaluate the impact of freight market concentration on domestic output. The purpose of this analysis is to estimate how changes to market access affect agricultural production. Recall that, in this context, market access is a weighted-sum of real expenditure across every U.S. county; the weights are inversely proportional to transit costs, inclusive of freight markups. As a first step, I exploit plausibly-exogenous variation in market access across time to identify how changes in market access affect domestic production. With these causal estimate in hand, I estimate how increases in market access brought about by counterfactual reductions in freight market power cause corresponding shifts in production and price levels in each U.S. county throughout the mid 19th and early 20th Century.

4.1 Estimating the Effects of Market Access on the Level of Production

I first isolate the effect of market access on agricultural production. My estimating equation is:

$$\operatorname{asinh} Q_{it}^{k} = \gamma_{s(i)t}^{k} + \gamma_{i}^{k} + f(x_i, y_i)\gamma_t + \xi \operatorname{asinh} \underbrace{\mathrm{MA}_{it}^{k}}_{\sum_{j\neq i}\left(\hat{\mu}_{ijt}^{k} V_{ijt}^{-1/\rho}\right)^{-\theta}\left(\hat{\phi}_{jt}^{k}\right)^{\frac{1-\theta}{\theta}} B_{jt}} + u_{it}^{k} \tag{16}$$

where γ, as before, denotes the corresponding fixed effects, and $f(x_i, y_i)\gamma_t$ is a cubic polynomial of the county's latitude and longitude interacted with a time fixed-effect. A nearly-identical specification is used by both Donaldson and Hornbeck (2016) and Chan (2022) to evaluate the effect of market access on agricultural land values and the total value of agricultural production, respectively. The parameter ξ is the effect that I want to estimate: the causal impact of changes in market access on domestic quantity of production. I now describe why this relationship is causal.

The fixed-effects are crucial to my identification strategy. As in Equation 15, the parameter $\gamma_{s(i)t}^{k}$ controls for unobservable variation due to commodity, time, and state-specific idiosyncrasies like differences in units or measurement or statewide climate shocks; it also controls for state-specific policies that may influence agricultural output – for instance, settlement incentives. The parameter γ_{i}^{k} removes variation unique to the origin county and commodity; it controls for a county's overall suitability at growing k. Similarly, the polynomial of the county's latitude and longitude control for overall growing conditions. The residual variation that identifies ξ is thus the differential change in market access to specific counties over time within the state. This variation stems development of the transit network, the economic growth of destination markets (i.e., all other U.S. counties), and the resultant change in the bilateral markup.

An immediate concern is that these bilateral markups are endogenous to the total quantity produced. Counties that become relatively more productive over time will garner larger trade shares than their less-productive neighbors, thus earning higher bilateral markups and lower market access. Co-movement of total agricultural production and market access is thus at least partially determined by the endogenous pricing response of imperfectly competitive transporter.[23] To address this threat to identification, I utilize the perfectly-competitive version of market access as an instrument. Explicitly, my first-stage estimating equation is

$$\operatorname{asinh} \mathrm{MA}_{it}^{k} = \delta_{s(i)t}^{k} + \delta_{i}^{k} + f(x_i, y_i)\delta_t + \xi \operatorname{asinh} \underbrace{\widetilde{\mathrm{MA}}_{it}^{k}}_{\sum_{j \neq i} V_{ijt}^{\theta/\rho}\left(\hat{\phi}_{jt}^{k}\right)^{\frac{1-\theta}{\theta}} B_{jt}} + \upsilon_{it}^{k} \tag{17}$$

where δ now denotes fixed-effects. As emphasized by Donaldson and Hornbeck (2016), Chan (2022), and Hornbeck and Rotemberg (2024), the variation of this perfectly-competitive market access term over time within a particular county is plausibly exogenous to that county's agricultural production. The national scale is so large that local conditions (that is, economic and transit network development in your own and neighboring counties) matter relatively little in the calculation of the instrument. To further alleviate potential enodgeneity of market access to agricultural output, I omit a county's own expenditure from the market access calculation. The variation in perfectly-competitive market access over time thus stems from expansion of the transit network nationally.

Finally, there is the risk that growth in the destination's expenditure, B_{jt}, is endogenous to growth in agricultural output due to unobserved, bilateral factors. While this risk is perhaps greatest for own-county expenditure, which is removed from my calculation of market access, I further alleviate this concern by re-calculating both the endogenous market access term and the perfectly-competitive market access used as an instrument with alternate measures of market size. Specifically, I utilize: i) population, which is correlated with total expenditure but less obviously related to agricultural output;[24] ii) expenditure in 1850, which removes any variation in market size over time, such that all changes in market access are attributable solely to the development of the transit network and the consequent changes in markups;[25] and iii) population in 1850, which combines the previous two methods and presents the least endogeneity concerns, at the cost of

[23]It should be noted that my formulation does *not* suffer from endogeneity due to market entry. Counties that become more productive over time will garner higher trade volumes, thus attracting more freight service providers and, ideally, lowering freight rates. However, I observe neither trade flows nor freight prices – as explained in Section 3.4, the markups entering into equation 16 are simulated from the estimated productivity parameters and assume a perfect monopoly in the market for freight services. There is thus no scope for endogenous changes to the transporter's pricing conduct; rather, the estimates displayed here impose that freight markups are at the upper-bound.

[24]Explicitly, $\mathrm{MA}_{it}^{k\,Pop} = \sum_{j \neq i} \left(\hat{\mu}_{ijt}^{k} V_{ijt}^{-1/\rho}\right)^{-\theta} N_{jt}$, where N_{jt} denotes the population of market j at time t. The perfectly-competitive versions of this formulation (where bilateral markups mu_{ijt}^{k} are set to 1)is used by both Donaldson and Hornbeck (2016) and Chan (2022) as a primary measure of market access.

[25]Explicitly, $\mathrm{MA}_{it}^{k\,1850} = \sum_{j \neq i} \left(\hat{\mu}_{ijt}^{k} V_{ijt}^{-1/\rho}\right)^{-\theta} \left(\hat{\phi}_{jt}^{k}\right)^{\frac{1-\theta}{\theta}} B_{j\,1850}$.

potentially mis-measuring market access.[26]

My estimation sample comprises 1,572 counties and 15 commodities that are observed in all years of data. Utilizing the balanced panel mitigates concerns regarding endogenous county creation: new population centers are likely to develop in areas that stand to gain the most from the new transit network; I thus limit my primary specification to counties that exist throughout my analysis period.[27] Even amongst this balanced panel, small, remote counties will likely see bigger percentage gains relative to larger, established counties. To mitigate the effect of these small counties, I follow Donaldson and Hornbeck (2016) and Chan (2022), and weight my regressions by the value of agricultural output in 1850. Results of my primary specification, as well as those using alternative market access measures and an unweighted version are displayed in Panel A of Table 4. Reassuringly, my estimates are highly similar across all specifications, assuaging concerns of endogeneity.

4.2 Estimating the Effects of Market Access on Total Value

I also evaluate the effect of market access on the nominal value of total agricultural production. My estimating equation is as follows:

$$\text{asinh}\, Y_{it} = \gamma_i + \gamma_{s(i)t} + f(x_i, y_i)\gamma_t + \eta\, \text{asinh}\, \underbrace{\text{MMA}_{it}}_{\mathbf{E}_k[\text{MA}_{it}^k]} + u_{it} \tag{18}$$

where Y_{it} denotes the sum-total value of all agricultural production in i and MMA denotes "mean market access", the simple average of market access across commodities in i. The parameter η captures the effect of changes in market access on the total value of agricultural production. In contrast to ξ, which identified the effect of market access on the quantity of production, this identifies the effect of market access on the *value* of production. This distinction is important, as granular price data are scarcely available for this period – however, using my estimates of η and ξ, I may approximate the effect of market access on local agricultural prices, which I will elaborate shortly.

As before, there is concern of endogeneity of mean market access due to bilateral markups. I utilize a similar IV strategy where my instrument is the simple average of perfectly-competitive

[26] In this case, $\text{MA}_{it}^{k\, Pop1850} = \sum_{j \neq i} \left(\hat{\mu}_{ijt}^k V_{ijt}^{-1/\rho} \right)^{-\theta} N_{j\,1850}$

[27] As a robustness check, I re-run my analysis utilizing my full sample of 2,725 counties. Results from this unbalanced panel are presented in Table 6, and are generally larger than my primary specification.

market access across commodities. Explicitly, my first-stage estimating equation is given by:

$$\operatorname{asinh} \text{MMA}_{it} = \delta_i + \delta_{s(i)t} + f(x_i, y_i)\delta_t + \eta \operatorname{asinh} \underbrace{\widetilde{\text{MMA}}_{it}}_{\mathbf{E}_k[\widetilde{\text{MA}}_{it}^k]} + \upsilon_{it} \tag{19}$$

The identification argument for Equation 18 is very similar to the argument just laid out for Equation 16; I thus speak briefly to why my estimates of η are causal. Origin and time fixed effects respectively control for unobservable county-specific productivity – like the overall quality of farmland in the county – as well as national macroeconomic trends like inflation or technology growth. The state-by-time fixed effects controls for unobserved trends that would affect agricultural productivity statewide; similarly, the time-varying polynomial of a county's latitude and longitude control for the overall agricultural productivity of a particular county. As in Equation 16, the residual variation that identifies η is the change in market access within county over time, which is driven by the expansion of the transit network, the economic development of every other U.S. county, and the resultant change in freight markups. The IV strategy removes identification concerns stemming from the potential enodgeneity of these bilateral markups to aggregate agricultural production.

I reiterate that, because perfectly-competitive market access reflects the national distribution of economic activity, variation in the instrument is almost purely exogenous to local economic conditions – namely, agricultural output; this exogeneity is reinforced by the fact that I omit local expenditure from the calculation of market access. To evaluate the potential for endogeneity of expenditure nationwide to the total value of production in county i, I again re-calculate market access using a variety of measures for market size: contemporaneous population, expenditure in 1850, and population in 1850. These alternate measures reduce endogeneity problems, but potentially mis-measure market access.

My estimating sample consists of a balanced panel of 1,572 counties. I again omit counties that incorporate during my sample period to address concerns of endogenous entry. I further weight my regressions by the value of farm output in 1850, thus limiting the impact of small, isolated counties that would see large percentage gains from market access. As a robustness check, I again run an unweighted version. Results of this analysis are reported in Panel B of Table 4; the stability of my estimates across specifications again suggest that endogeneity is of relatively minor concern.

As stated previously, the combination of η and ξ allow me to estimate the effect of market access on local price levels. Explicitly, from the fact that $\ln Y_{it} = \ln Q_{it} + \ln P_{it}$, it follows that

$\eta = \xi + \kappa$, where κ is the effect of increased market access on local agricultural prices. Hence, with estimates of η and ξ from estimating Equations 16 and 18, I also estimate the price effect κ. Results of this exercise are reported in Panel C of Table 4.

4.3 Results

Table 3 presents summary statistics on the key variables in my regression analysis for my primary estimation sample. These include the total quantity of each commodity produced, the total nominal value of farm output, as well as market access and its perfectly-competitive analogue in each decade. In addition to providing context to the forthcoming regression analysis, these summary measures also reveal that bilateral markups at the upper-bound are non-trivial. Existing measures of market access in the literature, which exclude endogenous freight pricing, may thus drastically over-state the change in market access caused by the 19th-Century railroad boom. In this section, I utilize the reduced-form methodology expounded in Sections 4.1 and 4.2 to evaluate how my measure of market access – which includes upper-bound estimates of bilateral freight markups – affect U.S. agricultural production. Importantly, I utilize this same regression framework to quantify the extent of bias introduced by excluding endogenous freight pricing from my measures of market access.

I focus first on the causal impact of market access changes on agricultural production. Results of this analysis are displayed in Table 4; Panel A displays the effects of increased market access on the level of output; Panel B describes the effect of market access on the total value of production; Panel C isolates the effect on local agricultural prices, and is calculated by subtracting the estimated coefficient in Panel B from the estimate in Panel A.[28] Columns (2) through (4) utilize alternative formulations of market access to address potential sources of endogeneity, as discussed previously. Column (5) presents an unweighted version of my regression results.

The Table reveals substantial impacts of increased market access on agricultural output. Focusing first on Column (1), the estimate reported in Panel A reveals that a 1% increase in market access leads to an estimated 0.974% increase in the total level of production within a given county. Switching focus to Panel B, a 1% increase in market access leads to a 1.87% increase in the total value of agricultural production. In tandem, these two estimates imply that a 1% increase in market access leads to a 0.896% increase in mean agricultural prices, as displayed by Panel C. These effects are all statistically different from zero at the 1% level of confidence, if not greater.

[28]Standard errors in Panel C are pooled from the standard errors reported in Panels A and B.

Columns (2) through (4) report highly similar coefficients. The measures of market access using population tend to find slightly lower impacts on output, but highly stable price effects. The stability of my estimates across alternate measures of market access suggest that endogeneity of my instrument presents relatively little concern. Column (5) reports estimates from an unweighted regression analysis and finds much larger effects. This noted difference from my primary specification in Column (1) suggests that, as expected, the smallest counties see the biggest percentage increases. In other words, the treatment effect is likely convex by initial market size, biasing up the average treatment effect reported in Column (5); the weighting utilized in Columns (1) through (4) offsets this bias. However there remains concern that the weighting does not completely address this bias. To evaluate this further, I estimate heterogeneous effects by market size in Section 4.4.

A notable finding is that, across the weighted specifications, the price and quantity effects are roughly equivalent in magnitude and of the same sign. This finding is important for consumer welfare considerations: although the total quantity of production increases, the accompanying and equivalently-sized price effects offset potential welfare gains. This finding also suggests that, if the change to market access is purely a demand shock,[29] then the agricultural supply curve is approximately unit-elastic. The magnitude of welfare consequences of this change to market access thus depend entirely on the elasticity of demand for agricultural goods.

How does incorporating freight market power into the calculation of market access shape our understanding U.S. production growth? This question is central to the contribution of this paper; to address it directly, I re-estimate Equation 16 using perfectly-competitive market access (my instrument from the previous regression).[30] This version of market access more closely mirrors existing measures in the literature (Donaldson and Hornbeck, 2016; Chan, 2022); however, it imposes perfect competition in the freight-services sector, an assumption that likely fails during this historical period, which saw the rise and subsequent dissolution of numerous freight cartels. As displayed by Table 3, excluding endogenous freight pricing leads to drastic changes in the measure of market access. How does this mis-measurement impact estimates of the causal effect of market access changes on U.S. output?

Results of the analysis using the perfectly-competitive market access are displayed in Table 5.[31] Notably, the same patterns hold as in my primary specification: i) the price and quantity ef-

[29]This view is reinforced by findings in the literature that market access changes scale up the level of production but have limited effect on agricultural factor productivity (Chan, 2022).

[30]Explicitly, my regression specification is $\text{asinh}\, Q_{it}^k = \gamma_{s(i)t}^k + \gamma_i^k + f(x_i, y_i)\gamma_t + \tilde{\xi}\, \text{asinh}\, \widetilde{\text{MA}}_{it}^k + u_{it}^k$.

[31]Note that this versions does not require an instrumental variables strategy, as the endogenous portion of trade costs

fects are roughly similar in magnitude and of the same sign; ii) the findings are robust to alternate measures of market access that address potential endogeneity; and iii) the unweighted versions create much larger estimates, suggesting some convexity of the treatment effects by initial market size. However, the coefficients from this analysis are generally just over half the size of the coefficients reported in Table 4. This noted change in magnitude accords with expectation: the perfectly-competitive measure of market access omits potentially large bilateral frictions, generating positive bias.[32] This upward bias in the independent variable biases downward the estimated coefficient. Hence, existing estimates in the literature likely understate the true effect of market access improvements by omitting endogenous freight pricing; the extent of the bias can be substantial, nearly double the reported coefficient at the upper-bound.[33]

4.4 Heterogeneous Effects

The un-weighted regressions suggest convexity of the treatment effects by initial level of farm output. To evaluate this heterogeneity directly, I repeat my regression analysis allowing for heterogeneous effects of market access expansion by deciles of initial farm output. My regression specifications are as follows:

$$\text{asinh}\, Q_{it}^k = \gamma_{s(i)t}^k + \gamma_i^k + f(x_i, y_i)\gamma_t + \xi^d \Big(D_{i\,1850} \times \text{asinh}\, \text{MA}_{it}^k = D_{i\,1850} \times \text{asinh}\, \widetilde{\text{MA}}_{it}^k \Big) + u_{it}^k \tag{20}$$

$$\text{asinh}\, Y_{it} = \gamma_{s(i)t} + \gamma_i + f(x_i, y_i)\gamma_t + \eta^d \Big(D_{i\,1850} \times \text{asinh}\, \text{MMA}_{it} = D_{i\,1850} \times \text{asinh}\, \widetilde{\text{MMA}}_{it} \Big) + u_{it} \tag{21}$$

where $D_{i\,1850}$ is a vector of indicators denoting a county's decile of total farm output in 1850. As before, the endogenous market access term is instrumented by the perfectly-competitive market access formulation more common in the literature. I again back-out the price effects $\kappa^d = \eta^d - \xi^d$. Results of this analysis are presented in Figure 2; Panel A reports the effects on quantity, Panel B reports the effects on the total value, and Panel C reports effects on price.

As displayed by the Figure, the estimated effects on the quantity and the total value of farm output decrease monotonically in the county's initial size; the effects on local agricultural prices

is omitted.

[32]Table 3 summarizes the extent of this bias.

[33]Recall that I assume that freight markets are perfectly concentrated throughout this analysis. Hence, I present the upper-bound on bias stemming from this mis-measurement of market access.

are flat. As expected, the smallest counties see the largest gains: counties in the first decile of farm production in 1850 see an approximately 0.994% increase in the total quantity of production for every 1% increase in market access; this effect falls to 0.829% in the second decile and 0.710% in the third. The highest-producing counties in 1850 see an effect of 0.472%, just over half the size of the effects for the smallest counties. The patterns for total value are similar: the counties in the first decile of production see the total value of agricultural production increase 1.36% for every 1% increase in market access, while the highest-producing counties see an effect 0.853%. All of these effects are statistically significant at the 0.1% level; the difference in coefficients between the first and tenth decile are statistically significant for the quantity and value regressions at the 5% level of confidence. Finally, the average effects reported in Figure 2 are smaller than the average treatment effect reported in my main specification, suggesting that the weighting in my primary specification does not completely resolve bias due to the convexity of the treatment effect by market size.

This pattern of heterogeneity aligns with expectation. As stated previously, the literature has established that market access improvements have limited impacts on factor productivity (Chan, 2022; Hornbeck and Rotemberg, 2024). Larger, more initially-productive counties have less scope to scale-up production via increased land use, equipment, or population. Less-productive counties, in contrast, may simply have more (or at least, more easily-accessible) untapped resources to devote to agricultural production in response to market access shifts. Notably, no decile reports a null or negative effect, meaning market access improvements have non-trivial, positive effects on the level of agricultural production across the distribution of farm output.

Given this marked heterogeneity in the effects on quantity, it is notable that my estimated price effects are uniform. There are two potential explanations for this result. First, market access improvements – which may be thought of as an upward shift in the local demand curve – spark an equivalent, downward shift in the local agricultural supply curve; the uniform, $\sim 0.4\%$ effect that I find reflects the difference in elasticity across these two curves. This explanation seems unlikely, as Chan (2022) find that market access has relatively little impact on agricultural factor productivity; rather, the new output is almost entirely explained by an increase in inputs (namely, land, rural laborers, and equipment value) – hence, increases in quantity seem to stem from movement along, not a shift in, the supply curve. A second potential explanation is that agricultural prices are set on large, regional markets.[34] Because my specification includes state-by-commodity-by time fixed

[34] Craig (1993) emphasizes that agricultural prices are set in large, geographic regions.

effects, the residual variation that I capture is the national change in agricultural prices generally, which affects all counties equally.[35]. However, absent disaggregate, historical price data, it is difficult to confirm this hypothesis empirically.

4.5 Estimating the Effects of Freight Market Concentration

The final stage of my analysis is to understand which particular regions are affected by the exercise of freight market power. Towards this end, I conduct a simple exercise: utilizing the heterogeneous estimates from Section 4.4, I estimate how agricultural production would have shifted in response to the total elimination of freight market power. Explicitly, for every county in every decade, and for each commodity, I calculate the difference between my favored measure of market access and its perfectly-competitive counterpart:

$$\text{asinh}\,\widehat{\text{MA}}_{it}^{k} = \text{asinh}\,\underbrace{\widetilde{\text{MA}}_{it}^{k}}_{\sum_{j\neq i}\left(\bar{\tau}_{ijt}/P_{jt}^{k}\right)^{-\theta}\left(B_{jt}/P_{jt}^{k}\right)} - \text{asinh}\,\underbrace{\text{MA}_{it}^{k}}_{\sum_{j\neq i}\left(\mu_{ijt}^{k}\bar{\tau}_{ijt}/P_{jt}^{k}\right)^{-\theta}\left(B_{jt}/P_{jt}^{k}\right)} \tag{22}$$

I then utilize my causal estimates of the impact of market access changes to assess how this counterfactual elimination of freight market power would have affected agricultural production:

$$\text{asinh}\,\widehat{Y}_{it}^{k} = \eta^{d}\Big(D_{i\,1850} \times \text{asinh}\,\widehat{\text{MA}}_{it}^{k}\Big) \tag{23}$$

I then convert these counterfactual changes into percent changes from the observed production. The counties that see the largest gains from this counterfactual exercise are most vulnerable to the broad exercise of freight market power.[36]

I present the results of this analysis in two forms. Table 7 presents summary statistics on the distribution of these production changes each decade. This table reveals that these distributions are extremely skew right in every period. In 1910, 50% of observable counties[37] see production gains of approximately 0.20% or less as a result of the complete elimination of freight market power. The county at the 75th percentile see gains of just over 1%, while the most-affected county see gains of nearly 357%.[38] This same pattern holds for the preceding decades, with the period

[35] An implication of this theory is that local producers treat the demand curve as perfectly inelastic.

[36] What this analysis misses are the consequences of geographically heterogeneous exercise of freight market power. Throughout this paper, I assume that freight markets nationally are perfectly concentrated; absent data on bilateral trade and/or freight prices, I cannot identify the geographic distribution of freight market power.

[37] Note that not every county is included in this analysis. Rather, it is confined to those counties that appear in the Census of Agriculture in the given decade.

[38] This distribution omits the top 0.1% of affected counties each year to remove extreme outliers.

from 1860 through 1890 reporting the most extreme effects.

It is evident from these distributions that freight market power affects relatively few counties extremely. Where are these impacted counties? Figure 3 presents a heat map of counterfactual production changes each decade from 1950 through 1910. Dark red denotes more extreme reactions; lighter colors represent a more mild response. Generally, the worst-affected areas are newer, remote counties located along the American frontier, as well more-established counties in the Pacific Northwest, Appalachian mountain range, and along the Gulf of Mexico. Over time, these losses move west with American expansion into the Great Plains and Rocky Mountains; frontier counties along the plains generally exhibit more extreme losses than established population centers in the Midwest and along the East Coast.

This geographic pattern is at least partially explained by my earlier result that smaller, initially less-productive counties see the greatest impacts of market access; these coefficients input directly into this calculation of potential gains. However, the level of heterogeneity exhibited in this geographic distribution is not completely explained by differences in my estimated causal effects — rather, the residual difference is attributable to higher impacts of market power in these remote areas due to higher markups. By the end of my analysis period, the most extreme geographic heterogeneity subsides and losses become more uniformly distributed; a pattern which may be attributable to the proliferation of the railroad and the death of the American Frontier.

5 Conclusion

This paper modifies the well-known market access framework to generate upper-bound estimates of agricultural production losses due to the concentration of freight market power from 1850 to 1910. Utilizing long-standing insights from the empirical industrial organization literature, I modify a multi-sector Ricardian trade model to permit endogenous freight pricing; these bilateral freight markups respond to the demand for transport services along a given route, as well as the state of competition along that route. I employ this theoretical framework to simulate bilateral trade flows and freight prices across a panel of U.S. counties each decade from 1850 to 1910. My analysis assumes that freight markets are perfectly concentrated. I thus provide theoretically tractable, upper-bound estimates of bilateral freight markups across the U.S. throughout the latter-half of the 19th and early 20th century.

This theoretical model also generates modified market access term – a reduced-form expression that summarizes the demand for goods out of a particular location – that accounts for these endogenous markups. My formulation of market access contrasts with existing measures in the literature, which typically assume purely exogenous bilateral trade frictions. This modification has important implications for my reduced-form analysis, which estimates the causal effect of market access expansion brought about by the 19th Century railroad boom on local agricultural production and prices. I show that omitting endogenous freight pricing leads to drastically overstated market access, and thus, under-estimated effect sizes. Specifically, my estimated effects nearly double when accounting for endogenous freight pricing in the calculation of market access. Hence, existing estimates in the literature represent a lower-bound.

My reduced-form analysis yields a number of conclusions. First, I show that market access has substantial effects on the total value of agricultural production, which aligns with similar estimates in the literature (Donaldson and Hornbeck, 2016; Chan, 2022). Second, I decompose this value effect into a quantity and price effect, which are roughly equivalent in size across my favored specifications. Third, I estimate noted convexity of this treatment effect by initial level of output – counties in the first decile of farm output at the start of my analysis see quantity effects that are nearly double the size for the tenth decile. A similar pattern holds for effects on the quantity of production, but price effects are stable. Finally, I utilize these causal effects to estimate the geographic distribution of production losses due to the concentration of market power. I show that the national distribution of these losses is highly skew right, with the most-affected counties located along the American Frontier. Over time, these larger losses move West with American expansion into the Plains and Rocky Mountains. By the end of my analysis period at the turn of the Century, agricultural losses stemming from freight market power are much more uniform, due to the saturation of the American landscape with railroads.

The primary contribution of this paper is to highlight the role of non-competitive freight pricing to our understanding of market access. While the railroad boom of the 19th Century saw significant gains to market access nationally, these increases were offset by the concentration of freight market power. I present upper-bound estimates of these losses to agricultural production; in reality, freight markets are likely less-than-perfectly concentrated throughout this time period. However, there is strong evidence that, by 1890 at latest, freight traffic was dogged by a number of railroad cartels with explicit price-fixing arrangements. Hence, it also unrealistic to assume that transport costs are perfectly exogenous, as is standard in the literature. I provide a tractable theoretical and reduced-form framework to evaluate the impact of this endogenous freight pricing. A laudable goal of future research would be to estimate the geographic concentration of freight market power nationally; absent additional data on bilateral trade flows or freight prices, this would be a difficult exercise.

References

Allen, Treb and Costas Arkolakis, "Trade and the Topography of the Spatial Economy," *The Quarterly Journal of Economics*, 2014, *129* (3), 1085–1139.

__ **and** __ , "The Welfare Effects of Transportation Infrastructure Improvements," *Review of Economic Studies*, 2022, *89* (6), 2911–2957.

Atack, Jeremy, *Historical Geographic Information Systems (GIS) database of Steamboat-Navigated Rivers During the Nineteenth Century in the United States* September 2015.

__ , *Historical Geographic Information Systems (GIS) database of U.S. Railroads for 1850-1910* May 2016.

__ , *Historical Geographic Information Systems (GIS) database of Nineteenth Century U.S. Canals* March 2017.

Bache, Stefan Milton and Hadley Wickham, *magrittr: A Forward-Pipe Operator for R* 2022. R package version 2.0.3.

Barrett, Tyson, Matt Dowle, Arun Srinivasan, Jan Gorecki, Michael Chirico, and Toby Hocking, *data.table: Extension of 'data.frame'* 2024. R package version 1.15.4.

Bergé, Laurent, "Efficient estimation of maximum likelihood models with multiple fixed-effects: the R package FENmlm," *CREA Discussion Papers*, 2018, (13).

Bonadio, Barthélémy, "Ports vs. Roads: Infrastructure, Market Access and Regional Outcomes," 2021.

Brancaccio, Giulia, Myrto Kalouptsidi, and Theodore Papageorgiou, "Geography, transportation, and endogenous trade costs," *Econometrica*, 2020, *88* (2), 657–691.

Bresnahan, Timothy F., "The Oligopoly Solution Concept is Identified," *Economics Letters*, 1982, *10* (1–2), 87–92.

Chan, Jeff, "Farming output, concentration, and market access: Evidence from the 19th-century American railroad expansion," *Journal of Development Economics*, 2022, *157*, 102878.

Costinot, Arnaud, Dave Donaldson, and Ivana Komunjer, "What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas," *The Review of Economic Studies*, 09 2011, *79* (2), 581–608.

Craig, Lee A., *To Sow One Acre More: Childbearing and Farm Productivity in the Antebellum North*, The Johns Hopkins University Press, 1993.

Donaldson, Dave, "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*, 2018, *108* (4–5), 899–934.

__ **and Richard Hornbeck**, "Railroads and American Economic Growth: A "Market Access" Approach," *The Quarterly Journal of Economics*, 2016, *131* (2), 699–858.

Eaton, Jonathan and Samuel Kortum, "Technology, Geography, and Trade," *Econometrica*, 2002, *70* (5), 1741–1779.

Fajgelbaum, Pablo D. and Edouard Schaal, "Optimal Transport Networks in Spatial Equilibrium," *Econometrica*, 2020, *88* (4), 1411–1452.

Fogel, Robert W., *Railroads and American Economic Growth: Essays in Econometric History*, Baltimore, MD: Johns Hopkins University Press, 1964.

Fuchs, Simon and Woan Foong Wong, "Multimodal Transport Networks," Federal Reserve Bank of Atlanta Working Paper # 2022-13 2023.

Hornbeck, Richard and Martin Rotemberg, "Growth Off the Rails: Aggregate Productivity Growth in Distorted Economies," *Journal of Political Economy*, 2024.

Jaworski, Taylor, Carl Kitchens, and Sergey Nigai, "HIGHWAYS AND GLOBALIZATION," *International Economic Review*, 2023, *64* (4), 1615–1648.

Klik, Mark, *fst: Lightning Fast Serialization of Data Frames* 2022. R package version 0.9.8.

Michael R. Haines and ICPSR, *Historical, Demographic, Economic, and Social Data: The United States, 1790-2002* May 2010.

Microsoft and Steve Weston, *doParallel: Foreach Parallel Adaptor for the 'parallel' Package* 2022. R package version 1.0.17.

__ and __ , *foreach: Provides Foreach Looping Construct* 2022. R package version 1.5.2.

Pfander, Charles Adam, "Trade and the Competition for Transport: How (a Lack of) Competition in the Transportation Industry Aects Regional Trade Outcomes." PhD dissertation, University of Colorado, Boulder 2024.

R Core Team, *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing 2024.

Signorell, Andri, *DescTools: Tools for Descriptive Statistics* 2024. R package version 0.99.54.

Sun, Weizhao, "The Earth Is Not Flat: Bilateral Elevation as a Component of Trade." PhD dissertation, University of Colorado, Boulder 2024.

United States v. Joint Traffic Association, 171 U.S. 505, 1898.

United States v. Trans-Missouri Freight Association, 166 U.S. 290, 1897.

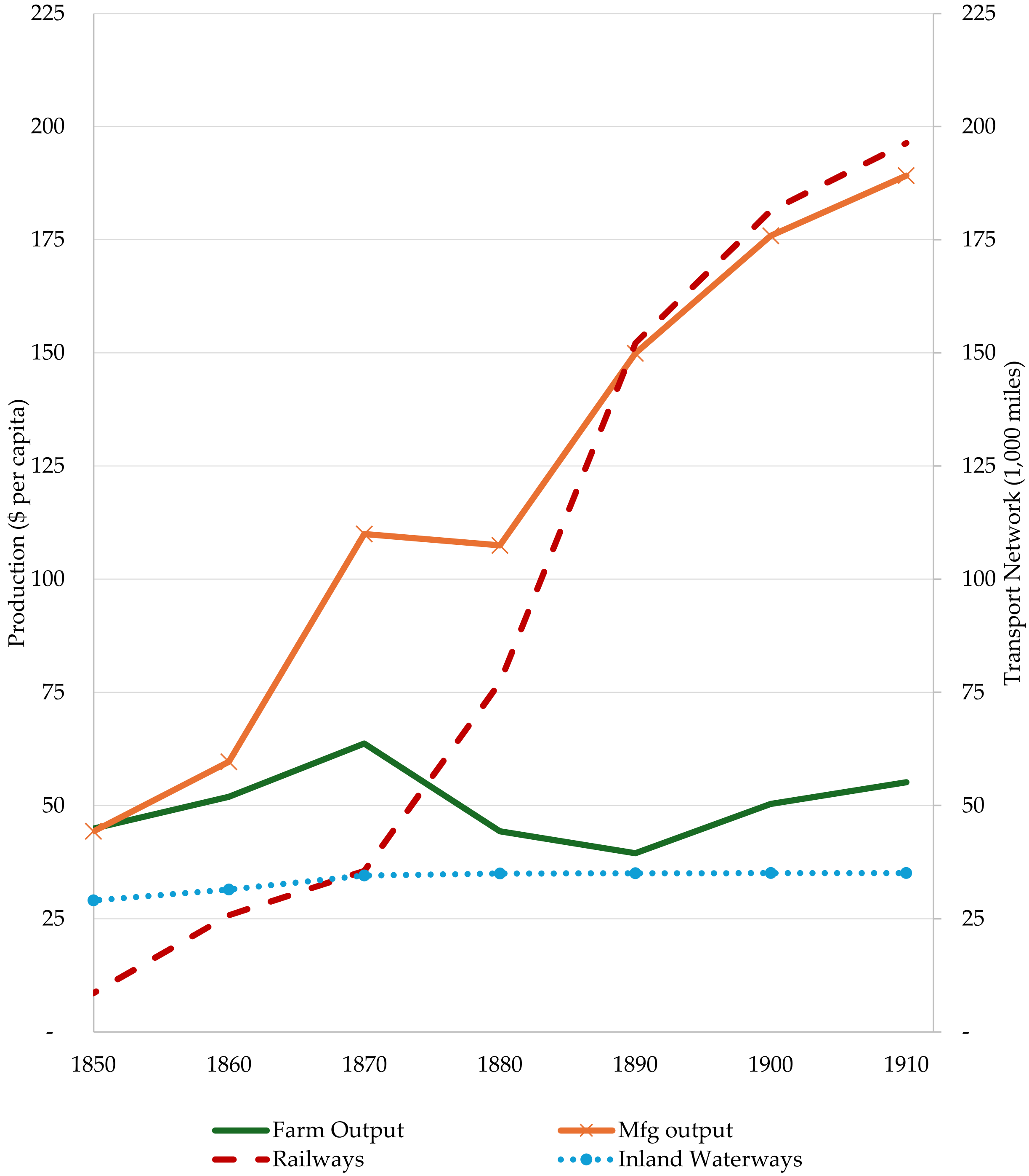
Wickham, Hadley, *stringr: Simple, Consistent Wrappers for Common String Operations* 2023. R package version 1.5.1.

__ , Evan Miller, and Danny Smith, *haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files* 2023. R package version 2.5.4.

__ , Jeroen Ooms, and Kirill Müller, *RPostgres: C++ Interface to PostgreSQL* 2023. R package version 1.4.6.

A Figures & Tables

Figure 1: Miles vs. Output



Notes: This figure displays the growth of nominal U.S. agricultural and manufcaturing output per-capita from 1850 through 1910. Also pictured is a growth in the total length of the rail and inland

Figure 2A: Transport Network - 1850

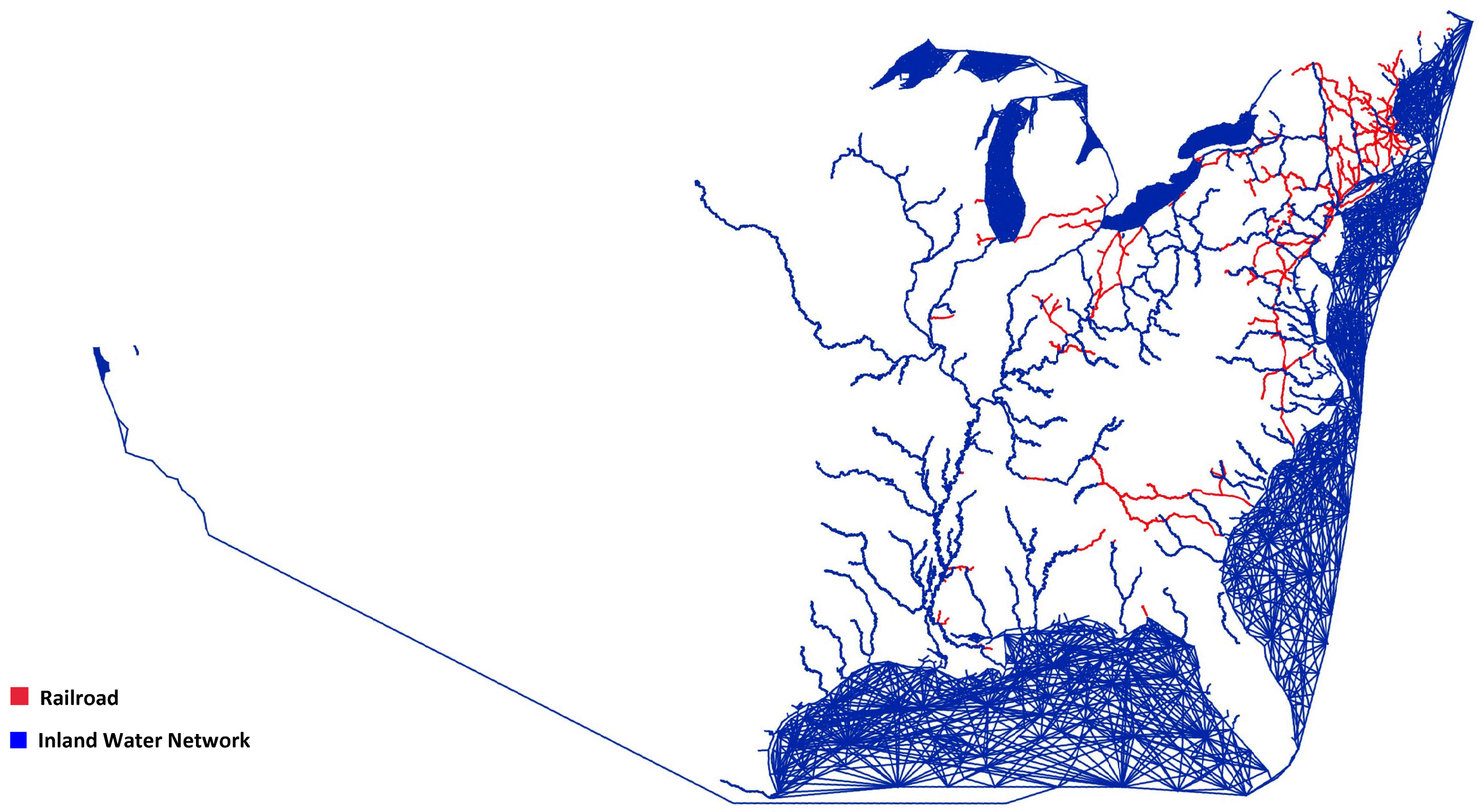


Figure 2B: Transport Network - 1860

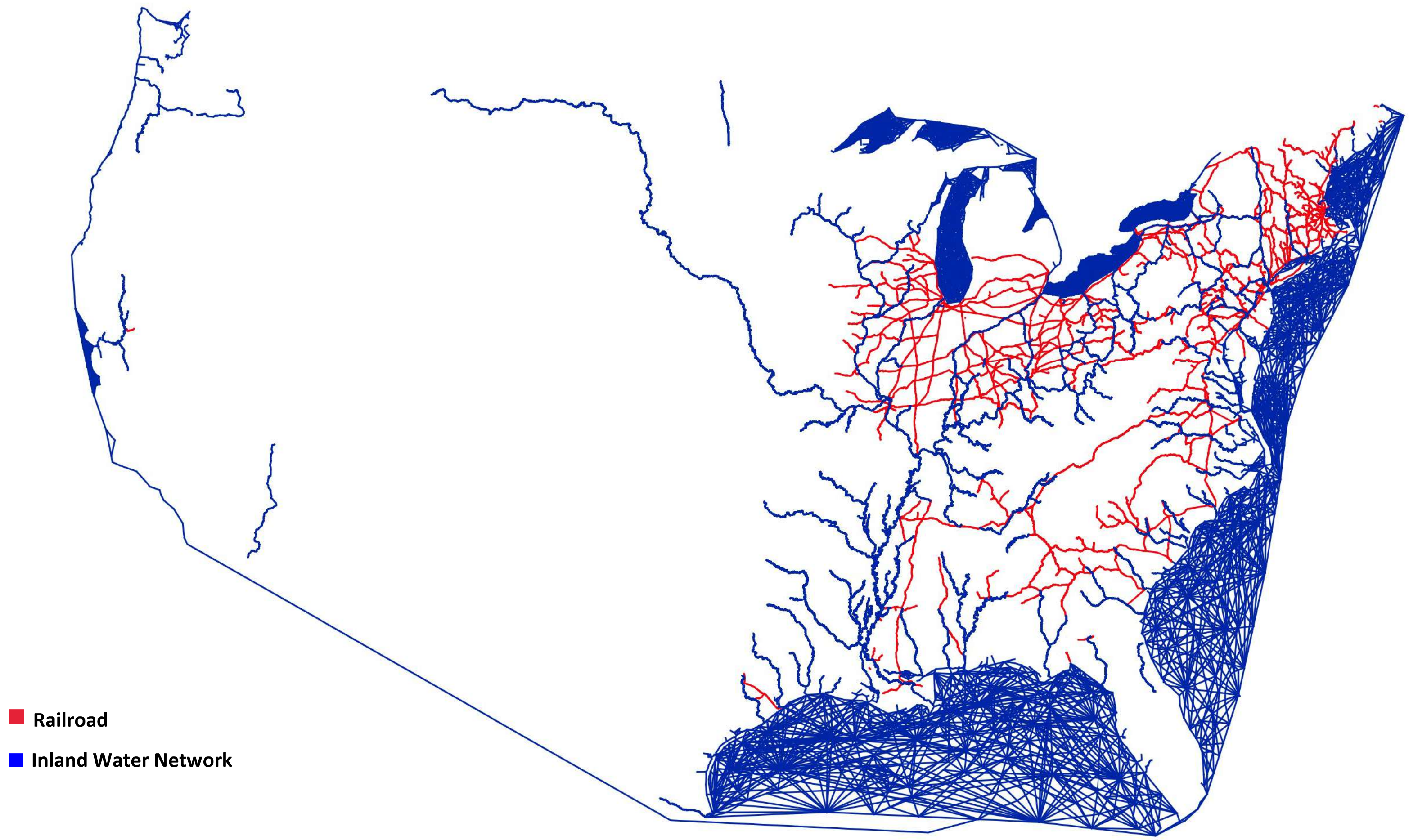


Figure 2C: Transport Network - 1870

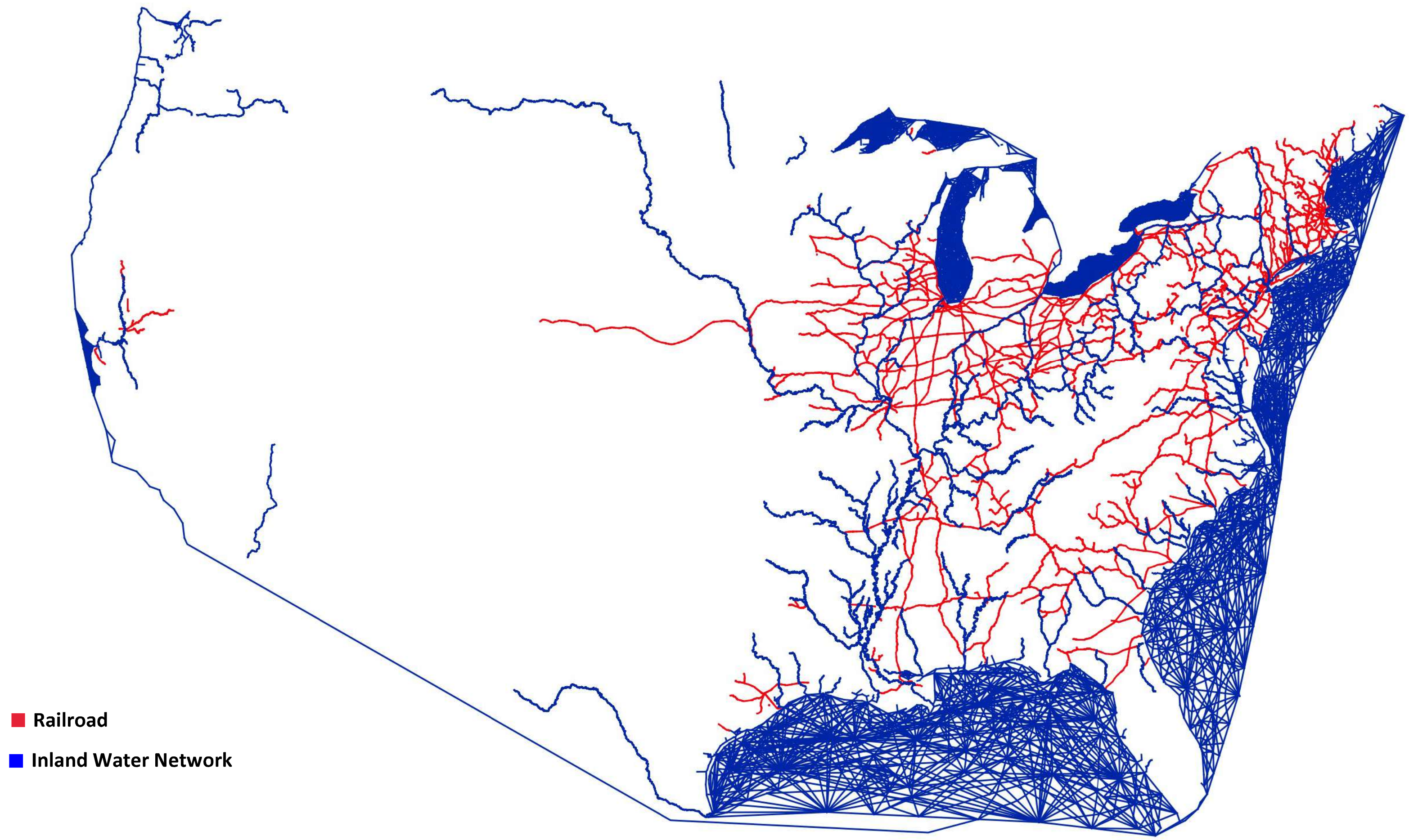


Figure 2D: Transport Network - 1880

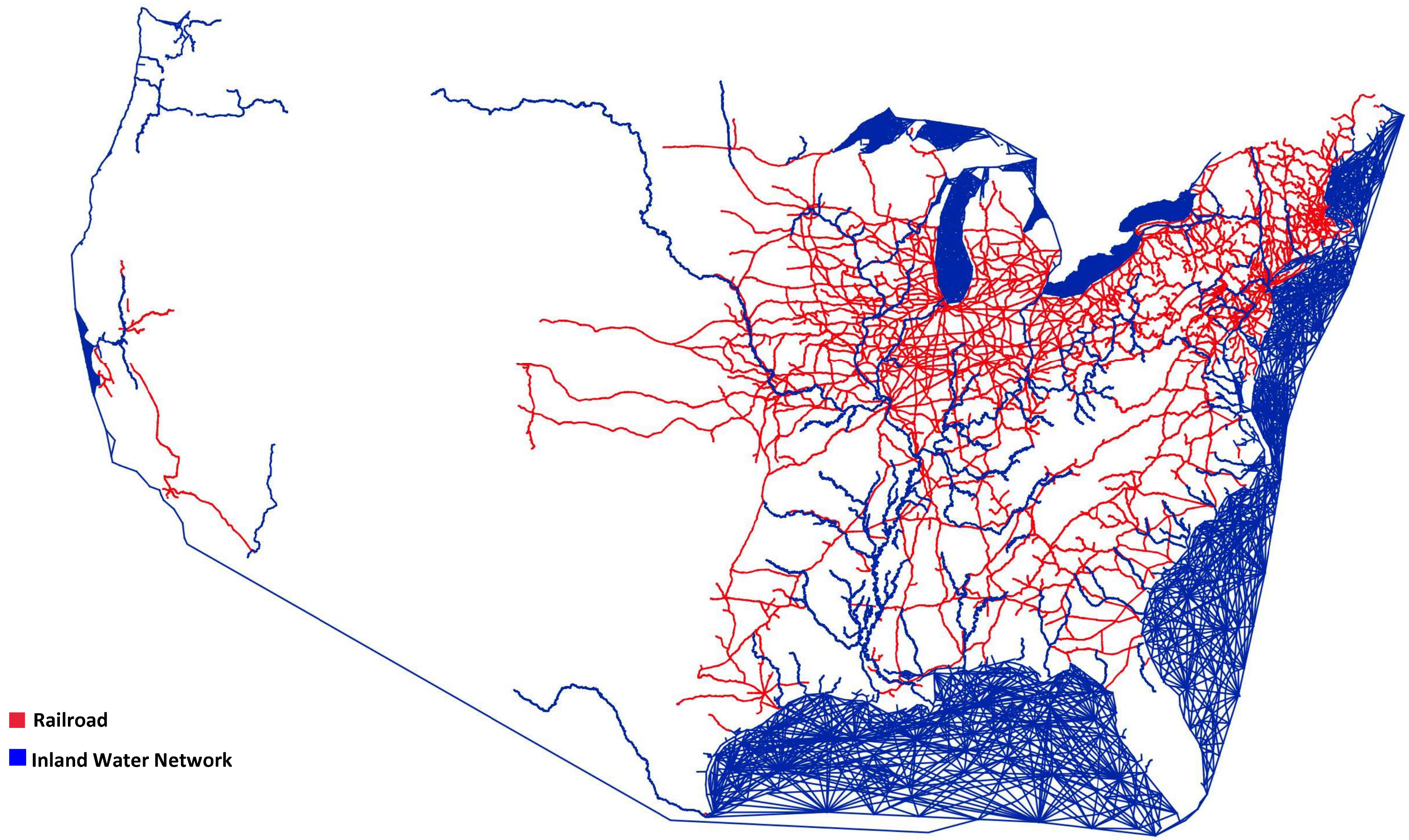


Figure 2E: Transport Network - 1890

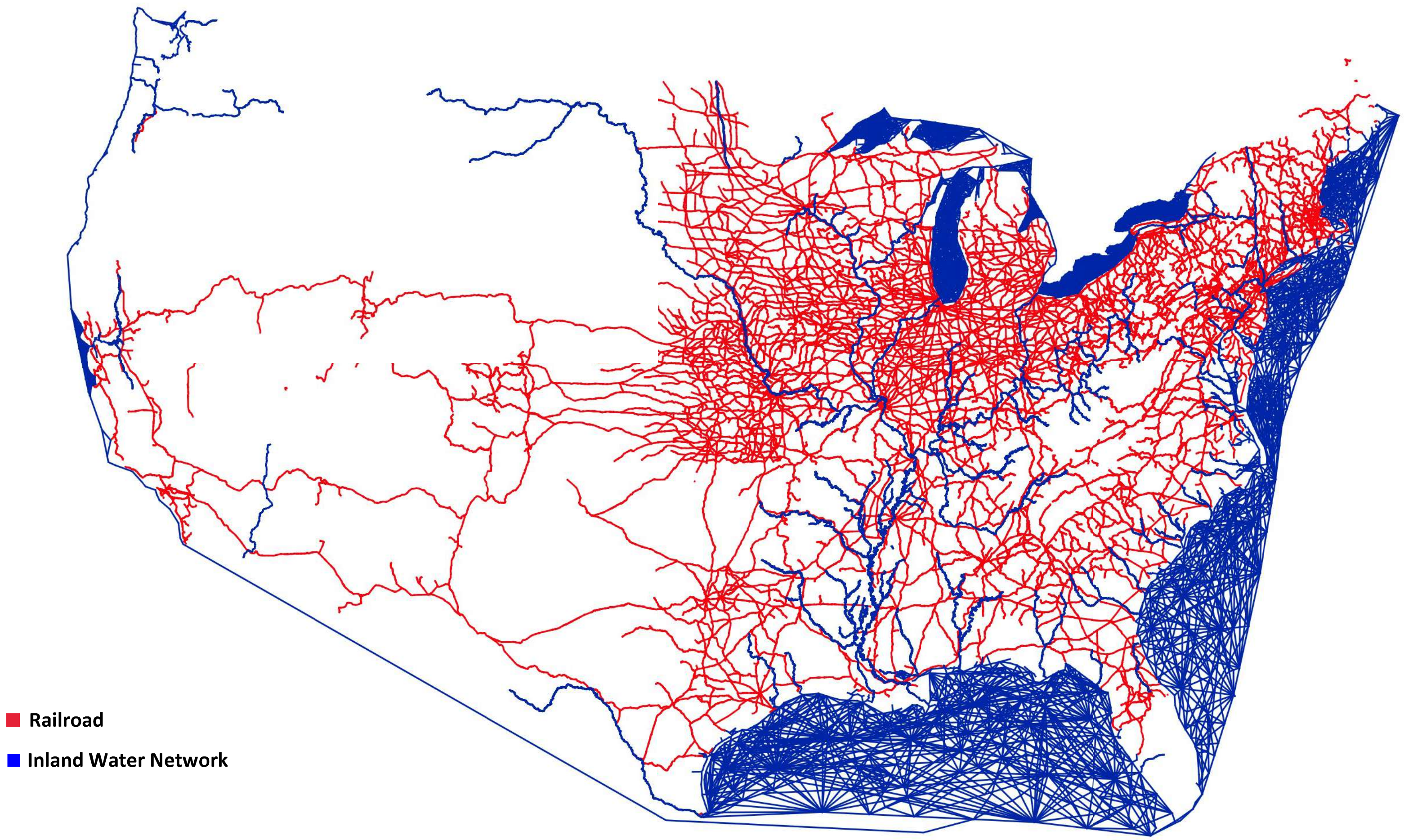


Figure 2F: Transport Network - 1900

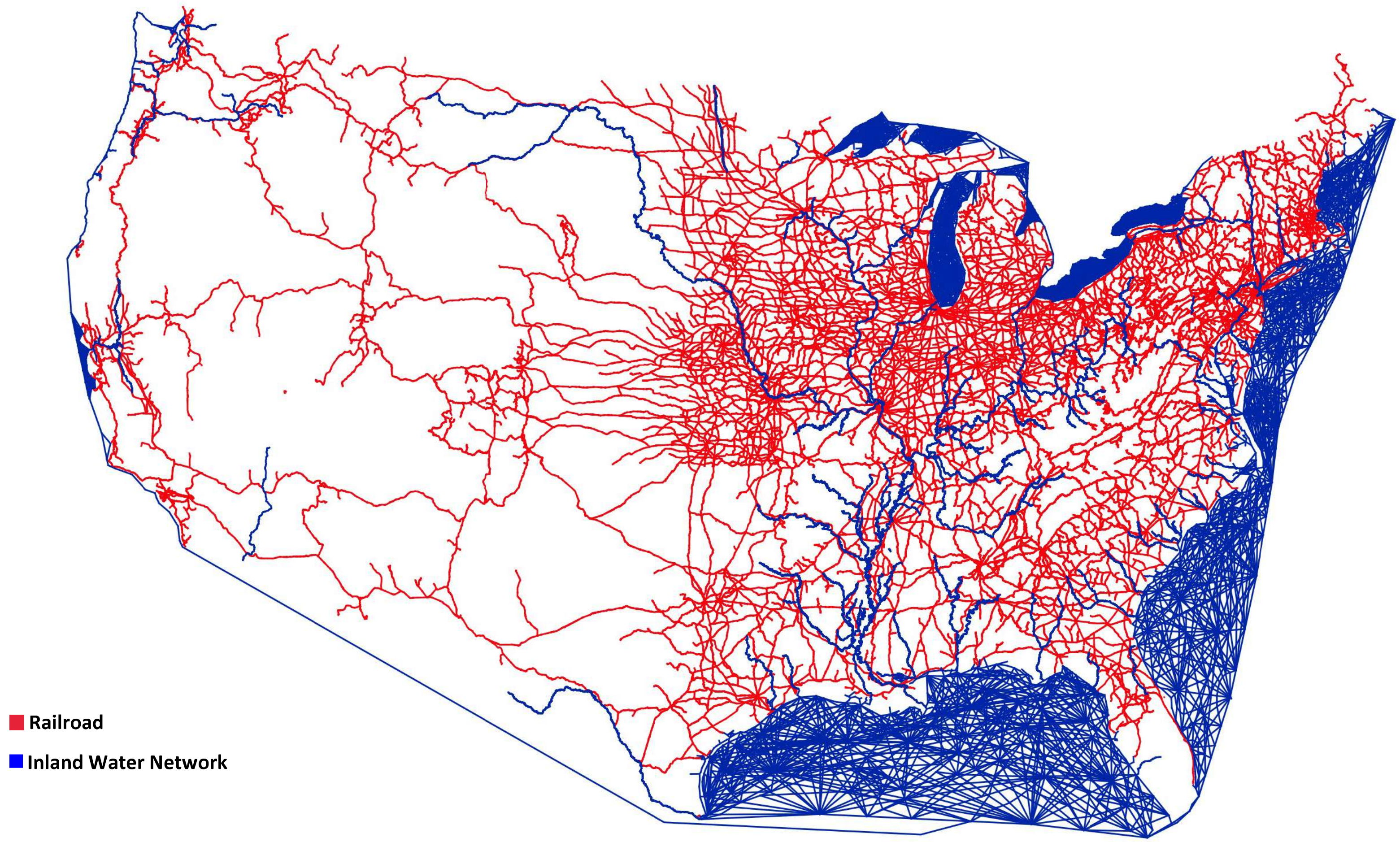


Figure 2G: Transport Network - 1910

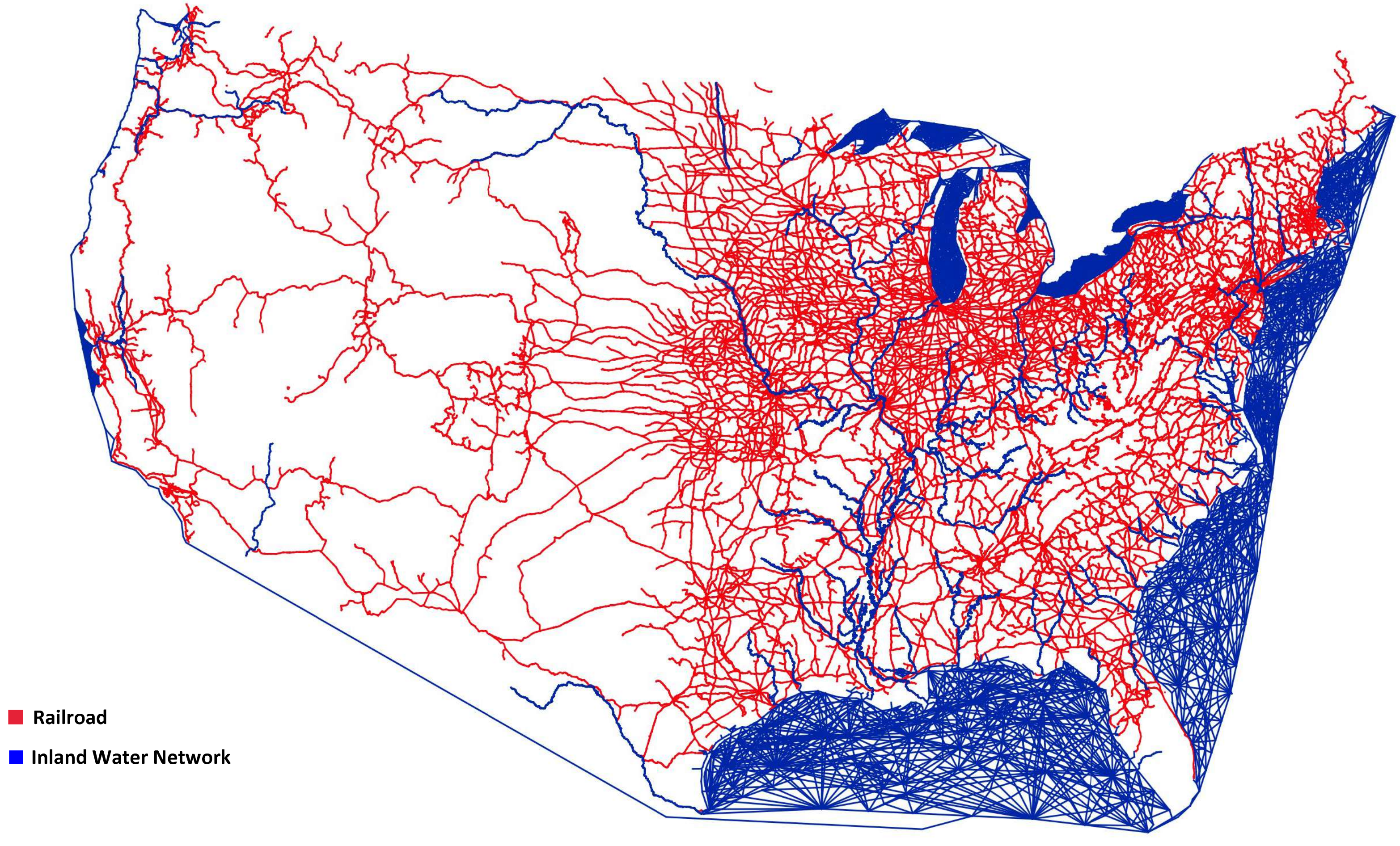


Table 1: Relative Modal Costs

β^m Coefficients (1000 miles)	
Road	1.0000
Rail	0.0272
Inland Water	0.0212
Sea/Lake Routes	0.0212
f^m Coefficeints (fixed)	
Road	0.0000
Rail	0.0022
Inland Water	0.0022
Sea/Lake Routes	0.0022
Coast-to-Coast	0.3463

Notes: This table displays the mode-specific cost structure that I utilize to generate least-cost paths and to estimate expected, exogenous transit costs. All cost parameters come from Donaldson and Hornbeck (2016); specifically, I take their dollar-value coefficients and divide them by the dollar-cost of road travel; this exercise yields the above relative cost structure.

Table 2: Commodities Included

Barley (bushels)	Cotton (400lb Bales)	Swine (#)
Buckwheat (bushels)	Cows (#)	Tobacco (lbs)
Butter (lbs)	Hops (lbs)	Wheat (bushles)
Cheese (lbs)	Rice (lbs)	Wine (gal)
Corn (bushels)	Rye (lbs)	Wool (lbs)

Notes: This table presents the list of commodities included in my analysis.

Table 3: Summary Statistics

	Mean	Std. Dev.	Min	25^{th} Pctile.	Median	75^{th} Pctile.	Max
Farm output (\$000, Nominal)	1,271.75	1,159.20	0.73	459.23	937.32	1,746.13	12,019.82
Quantity produced (000s)							
Barley (bushels)	15.81	99.89	0.00	0.00	0.06	1.37	3,439.30
Buckwheat (bushels)	7.31	29.38	0.00	0.00	0.06	1.60	667.36
Butter (lbs)	367.32	513.61	0.00	69.04	198.05	475.90	9,590.35
Cheese (lbs)	28.17	248.78	0.00	0.00	0.29	3.10	10,901.52
Corn (bushels)	736.01	1,079.88	0.00	195.29	402.37	804.30	16,001.36
Cotton (400lb Bales)	3.22	8.04	0.00	0.00	0.00	1.27	141.49
Cows (#)	12.72	12.43	0.00	5.14	9.05	15.71	198.79
Hops (lbs)	14.56	189.16	0.00	0.00	0.00	0.02	6,119.74
Rice (lbs)	77.88	1,212.87	0.00	0.00	0.00	0.00	55,805.39
Rye (lbs)	10.95	32.27	0.00	0.13	1.09	6.35	547.90
Swine (#)	21.52	20.66	0.00	8.55	15.96	27.08	247.76
Tobacco (lbs)	325.88	1,406.98	0.00	0.01	1.51	16.71	36,892.87
Wheat (bushles)	143.49	251.99	0.00	4.06	40.30	160.44	2,959.44
Wine (gal)	4.87	269.57	0.00	0.01	0.13	0.52	27,921.18
Wool (lbs)	53.02	116.14	0.00	6.74	17.09	48.01	2,635.95
asinh(Market Access)							
1850	8.06	0.73	3.65	7.50	8.17	8.61	9.34
1860	7.96	0.68	3.87	7.38	8.03	8.49	9.09
1870	9.85	0.59	6.45	9.55	9.94	10.24	10.88
1880	10.07	0.59	7.33	9.48	10.21	10.41	11.12
1890	10.52	0.59	8.65	10.04	10.67	10.96	11.64
1900	10.66	0.70	8.73	10.14	10.88	11.19	11.87
1910	10.73	0.84	8.37	10.34	11.01	11.31	12.16
asinh(Perfectly-Competitive Market Access)							
1850	15.38	0.90	9.86	14.78	15.28	15.98	17.56
1860	15.55	0.81	10.08	15.07	15.43	16.09	17.45
1870	17.45	0.69	12.78	17.10	17.37	17.90	18.92
1880	17.94	0.64	14.00	17.60	17.83	18.29	19.48
1890	18.62	0.61	16.39	18.24	18.47	18.91	20.43
1900	18.88	0.61	16.64	18.50	18.71	19.10	20.67
1910	19.06	0.60	16.76	18.69	18.96	19.27	20.93

Notes: This table presents summary statistics on the total value of farm output, the total quantity produced, the inverse hyperbolic sine transform of market access, as well as the inverse hyperbolic sine transform of perfectly-competitive market access. The summary statistics are calculated over the estimation sample – a balanced panel of 1,572 counties and 15 commodities from 1850-1910.

Table 4: Estimates of Market Access on Output, Prices

	Model-Derived Market Access	MA Calculated using Current Population	MA Calculated using 1850 Expenditure	MA Calculated using 1850 Population	Model-Derived Market Access (Unweighted)
	(1)	(2)	(3)	(4)	(5)
A. Effects on Quantity Produced					
asinh(MA)	**0.974*** **	**0.778** **	**0.894** **	**0.777** **	**2.09*** **
	(0.279)	**(0.250)**	**(0.279)**	**(0.251)**	**(0.186)**
Fixed-Effects					
Origin × Comm.	X	X	X	X	X
State × Comm. × Yr.	X	X	X	X	X
Lat, Lon polynominal	X	X	X	X	X
# of Obs	165,060	165,060	165,060	165,060	165,060
# of Counties	1,572	1,572	1,572	1,572	1,572
# of Commodities	15	15	15	15	15
First-Stage F stat	1,857,928	2,367,536	1,938,952	2,511,375	2,344,901
Adj. R^2	0.914	0.914	0.914	0.914	0.903
B. Effects on Agricultural Value					
asinh(MA)	**1.87*** **	**1.68*** **	**1.83*** **	**1.67*** **	**2.50*** **
	(0.184)	**(0.166)**	**(0.186)**	**(0.167)**	**(0.227)**
Fixed-Effects					
Origin	X	X	X	X	X
State × Yr.	X	X	X	X	X
Lat, Lon polynominal	X	X	X	X	X
# of Obs	11,004	11,004	11,004	11,004	11,004
# of Counties	1,572	1,572	1,572	1,572	1,572
First-Stage F stat	254,982	814,570	295,876	905,445	227,010
Adj. R^2	0.933	0.933	0.933	0.933	0.919
C. Effects on Average Agricultural Prices					
asinh(MA)	**0.896** **	**0.902*** **	**0.936*** **	**0.893*** **	**0.410* **
	(0.274)	**(0.246)**	**(0.274)**	**(0.247)**	**(0.189)**

Notes: This table presents impacts of the effect of increases in market access on agricultural output, value, and prices. Panel A reports the results on quantity produced. Panel B reports the results on the total value of agricultural output. Panel C is simply the coefficeint reported in B minus the coefficient reported in A. Columns (1) - (4) utilize different verisons of market access to address potential endogeneity concerns. Following Donaldson and Hornbeck (2016), the regressions in Columns (1) - (4) are weighted by a county's initial value of all farm output to minimize the influence of outliers. Column (5) utilizes the same market-access measure as in Column (1) but does not utilize any weighting. Standard errors in Panel A are clustered by origin-commodity pairing; standard errors in Panel B are clustered by origin county; standard errors in Panel C are pooled from the errors reported in Panels A and B. ***, **, and * denote signifcance at the 0.1%, 1% and 5% levels, respectively. The estimation sample is restricted to county-commodity combinations that are present in all years.

Table 5: Estimates of Perfectly-Competitive Market Access on Output, Prices

	Model-Derived Market Access	MA Calculated using Current Population	MA Calculated using 1850 Expenditure	MA Calculated using 1850 Population	Model-Derived Market Access (Unweighted)
	(1)	(2)	(3)	(4)	(5)
A. Effects on Quantity Produced					
asinh(MA)	**0.520*** **	**0.468** **	**0.476** **	**0.467** **	**1.17*** **
	(0.149)	**(0.150)**	**(0.149)**	**(0.151)**	**(0.103)**
Fixed-Effects					
Origin × Comm.	X	X	X	X	X
State × Comm. × Yr.	X	X	X	X	X
Lat, Lon polynominal	X	X	X	X	X
# of Obs	165,060	165,060	165,060	165,060	165,060
# of Counties	1,572	1,572	1,572	1,572	1,572
# of Commodities	15	15	15	15	15
Adj. R^2	0.914	0.914	0.914	0.914	0.903
B. Effects on Agricultural Value					
asinh(MA)	**0.999*** **	**1.02*** **	**0.975*** **	**1.01*** **	**1.40*** **
	(0.098)	**(0.100)**	**(0.099)**	**(0.100)**	**(0.124)**
Fixed-Effects					
Origin	X	X	X	X	X
State × Yr.	X	X	X	X	X
Lat, Lon polynominal	X	X	X	X	X
# of Obs	11,004	11,004	11,004	11,004	11,004
# of Counties	1,572	1,572	1,572	1,572	1,572
Adj. R^2	0.933	0.933	0.933	0.933	0.920
C. Effects on Average Agricultural Prices					
asinh(MA)	**0.479** **	**0.552*** **	**0.499*** **	**0.543*** **	**0.230* **
	(0.146)	**(0.147)**	**(0.146)**	**(0.148)**	**(0.104)**

Notes: This table presents impacts of the effect of increases in perfectly-competitive market access on agricultural output, value, and prices. In contrast to Table 4, bilateral freight markups are excluded from the calculation of market access. Panel A reports the results on quantity produced. Panel B reports the results on the total value of agricultural output. Panel C is simply the coefficeint reported in B minus the coefficient reported in A. Columns (1) - (4) utilize different verisons of market access to address potential endogeneity concerns. Following Donaldson and Hornbeck (2016), the regressions in Columns (1) - (4) are weighted by a county's initial value of all farm output to minimize the influence of outliers. Column (5) utilizes the same market-access measure as in Column (1) but does not utilize any weighting. Standard errors in Panel A are clustered by origin-commodity pairing; standard errors in Panel B are clustered by origin county; standard errors in Panel C are pooled from the errors reported in Panels A and B. ***, **, and * denote signifcance at the 0.1%, 1% and 5% levels, respectively. The estimation sample is restricted to county-commodity combinations that are present in all years.

Table 6: Estimates of Market Access on Output, Prices using Unbalanced Panel

	Model-Derived Market Access	MA Calculated using Current Population	MA Calculated using 1850 Expenditure	MA Calculated using 1850 Population	Model-Derived Market Access (Unweighted)
	(1)	(2)	(3)	(4)	(5)
A. Effects on Quantity Produced					
asinh(MA)	**1.02*** **	**0.829*** **	**0.804*** **	**0.783*** **	**1.44*** **
	(0.167)	**(0.149)**	**(0.169)**	**(0.149)**	**(0.128)**
Fixed-Effects					
Origin × Comm.	X	X	X	X	X
State × Comm. × Yr.	X	X	X	X	X
Lat, Lon polynominal	X	X	X	X	X
# of Obs	244,005	244,005	244,005	244,005	244,005
# of Counties	2,725	2,725	2,725	2,725	2,725
# of Commodities	15	15	15	15	15
First-Stage F stat	3,759,366	5,554,235	2,242,311	2,988,646	6,008,454
Adj. R^2	0.912	0.912	0.912	0.912	0.896
B. Effects on Agricultural Value					
asinh(MA)	**1.64*** **	**1.40*** **	**1.40*** **	**1.33*** **	**1.94*** **
	(0.174)	**(0.151)**	**(0.166)**	**(0.150)**	**(0.323)**
Fixed-Effects					
Origin	X	X	X	X	X
State × Yr.	X	X	X	X	X
Lat, Lon polynominal	X	X	X	X	X
# of Obs	16,267	16,267	16,267	16,267	16,267
# of Counties	2,725	2,725	2,725	2,725	2,725
First-Stage F stat	406,970	1,407,876	336,892	1,034,150	527,520
Adj. R^2	0.925	0.925	0.925	0.925	0.876
C. Effects on Agricultural Prices					
asinh(MA)	**0.620*** **	**0.571*** **	**0.596*** **	**0.547*** **	**0.500*** **
	(0.167)	**(0.149)**	**(0.169)**	**(0.149)**	**(0.148)**

Notes: This table presents impacts of the effect of increases in market access on agricultural output, value, and prices. Panel A reports the results on quantity produced. Panel B reports the results on the total value of agricultural output. Panel C is simply the coefficeint reported in B minus the coefficient reported in A. Columns (1) - (4) utilize different verisons of market access to address potential endogeneity concerns. Following Donaldson and Hornbeck (2016), the regressions in Columns (1) - (4) are weighted by a county's initial value of all farm output to minimize the influence of outliers. Column (5) utilizes the same market-access measure as in Column (1) but does not utilize any weighting. Standard errors in Panel A are clustered by origin-commodity pairing; standard errors in Panel B are clustered by origin county; standard errors in Panel C are pooled from the errors reported in Panels A and B. ***, **, and * denote signifcance at the 0.1%, 1% and 5% levels, respectively. Unlike Table 4, there are no restrictions placed on the sample.

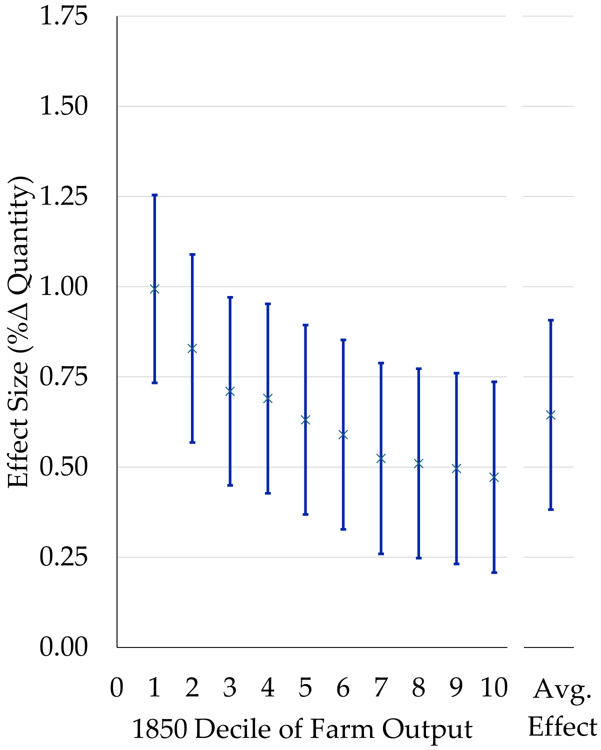
Table 7: Percentage Gains from the Elimination of Freight Market Power

	Mean	Std. Dev.	Min	25th Pctile.	Median	75th Pctile.	Max
1850	3.31	15.79	0.00	0.04	0.15	0.72	266.96
1860	10.01	65.86	0.00	0.05	0.20	1.26	1827.07
1870	11.53	93.24	0.00	0.04	0.16	1.05	1827.07
1880	21.51	147.21	0.00	0.06	0.29	1.95	1827.07
1890	14.88	109.38	0.01	0.08	0.37	2.20	1827.07
1900	2.77	15.54	0.00	0.07	0.25	1.26	525.32
1910	1.79	9.49	0.00	0.05	0.20	1.03	356.69

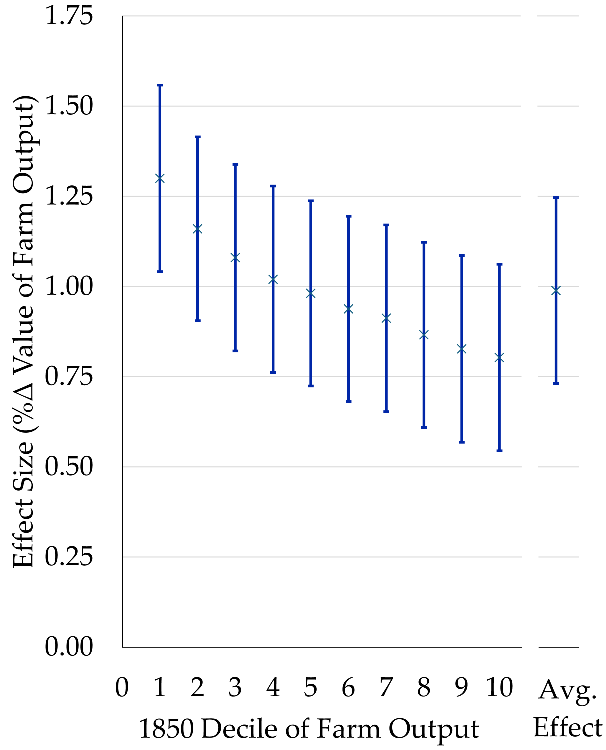
Notes: This table presents the distribution of counterfactual agricultural output gains stemming from the elimination of freight market power. All units are in percent-gains from baseline.

Figure 2: Heterogeneous Effects

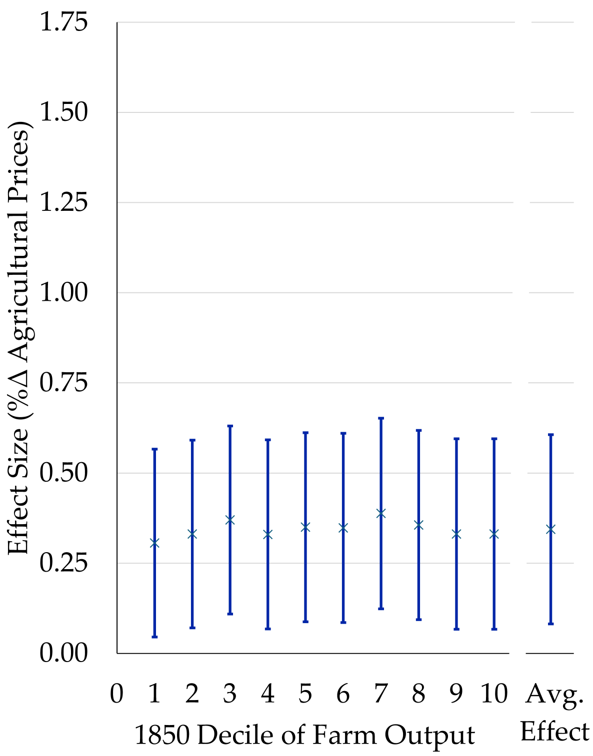
A. Effect on Agricultural Quantity by Initial Decile of Farm Output



B. Effect on Total Value of Farm Output by Initial Decile of Farm Output



C. Effect on Agricultural Prices by Initial Decile of Farm Output



Notes: This figure displays heterogeneous effects by decile of total farm output in 1850. Panel A displays the effects on the quantity of production; Panel B displays the effects on the total value of production; Panel C displays the effect on local agricultural prices. The point estimates are plotted along with a 95% confident interval. Additionally, I display an average effect, calculated as the simple mean of each plotted coefficient, on the far right of each figure. Controls in each regression are identical to those listed in Table 4. The sample comprises a balanced panel of 1,572 counties from 1850 to 1910.

Figure 3A: Geographic Gains from Elimination of Freight Market Power - 1850

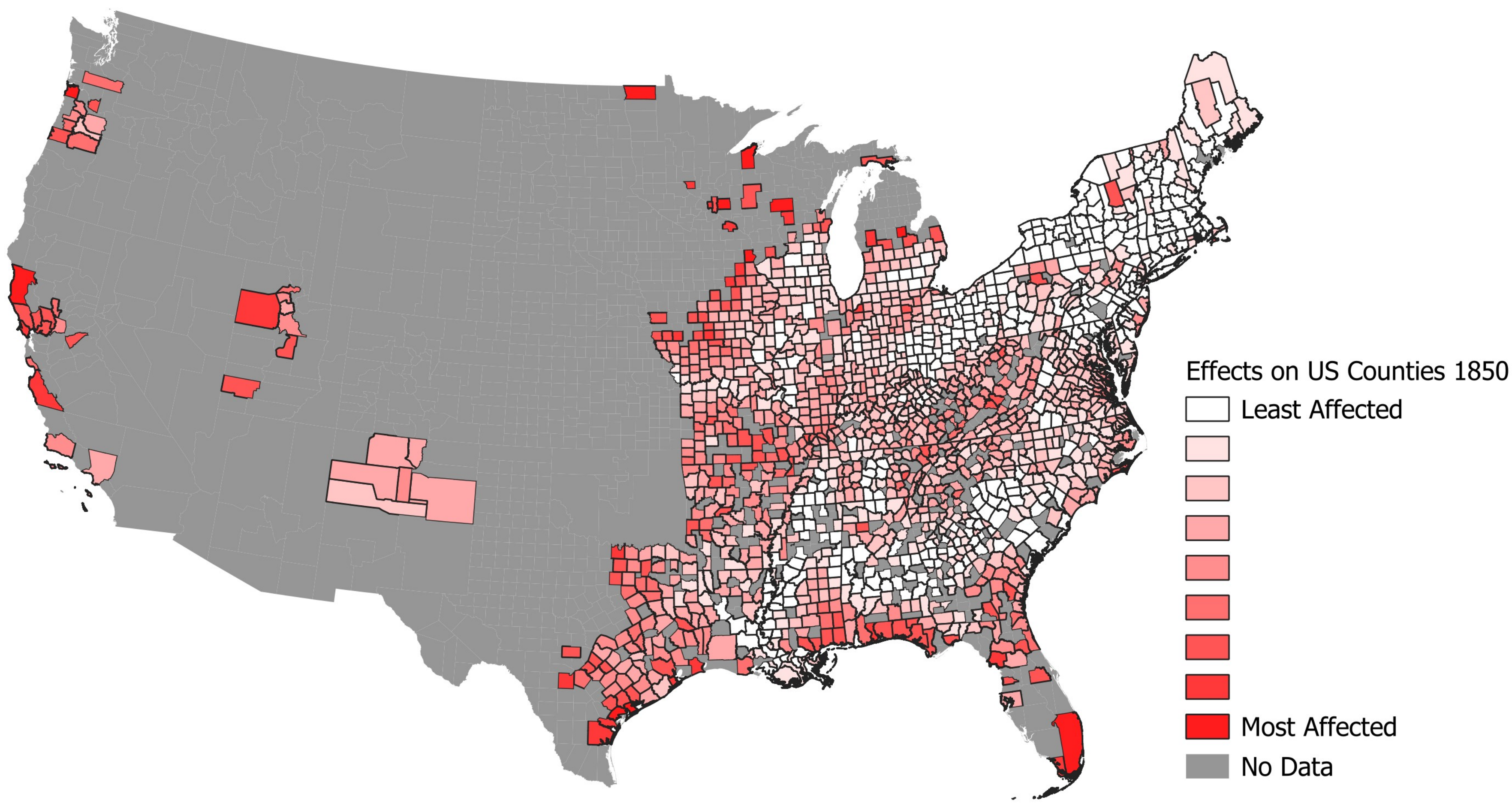


Figure 3B: Geographic Gains from Elimination of Freight Market Power - 1860

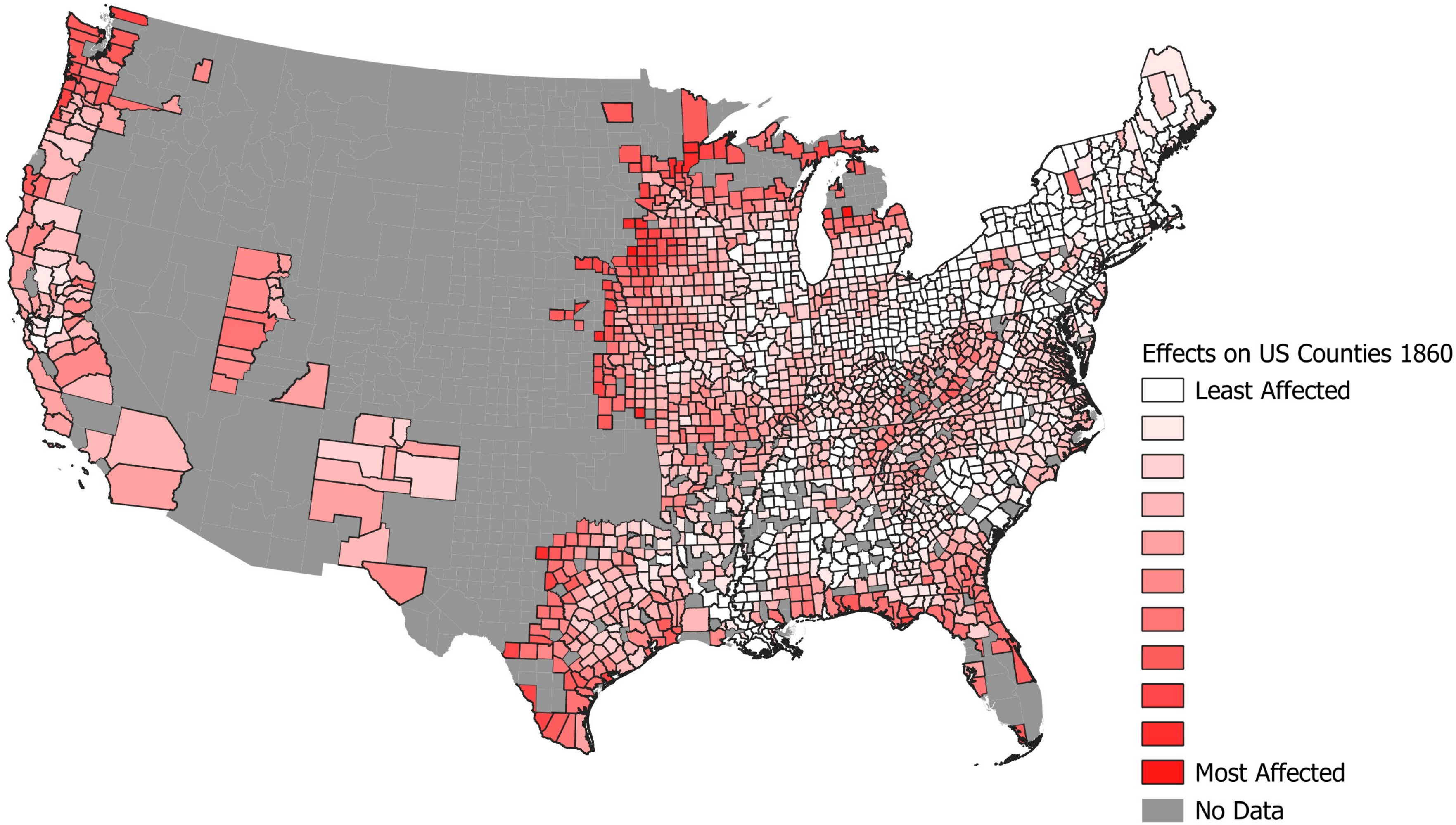


Figure 3C: Geographic Gains from Elimination of Freight Market Power - 1870

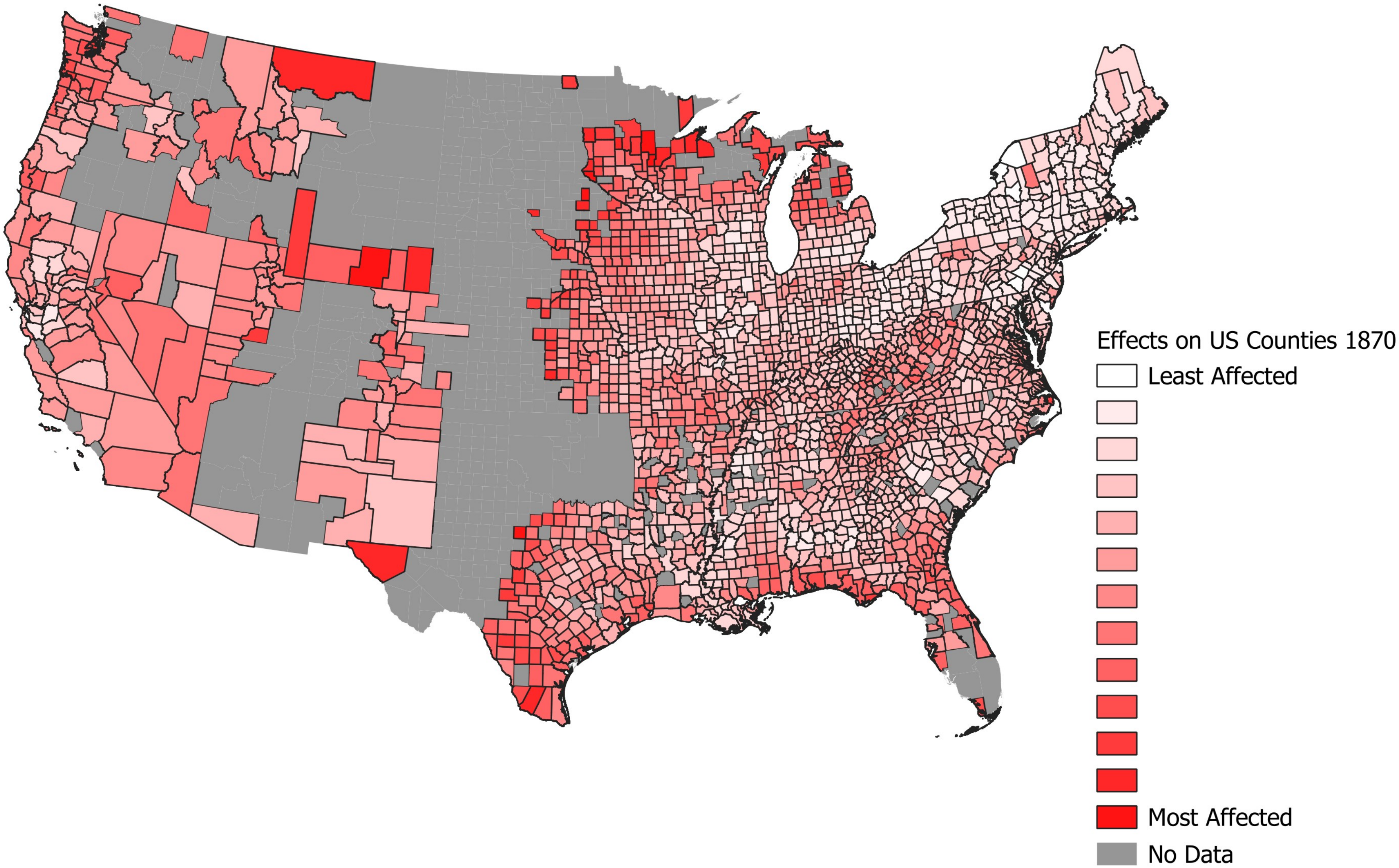


Figure 3D: Geographic Gains from Elimination of Freight Market Power - 1880

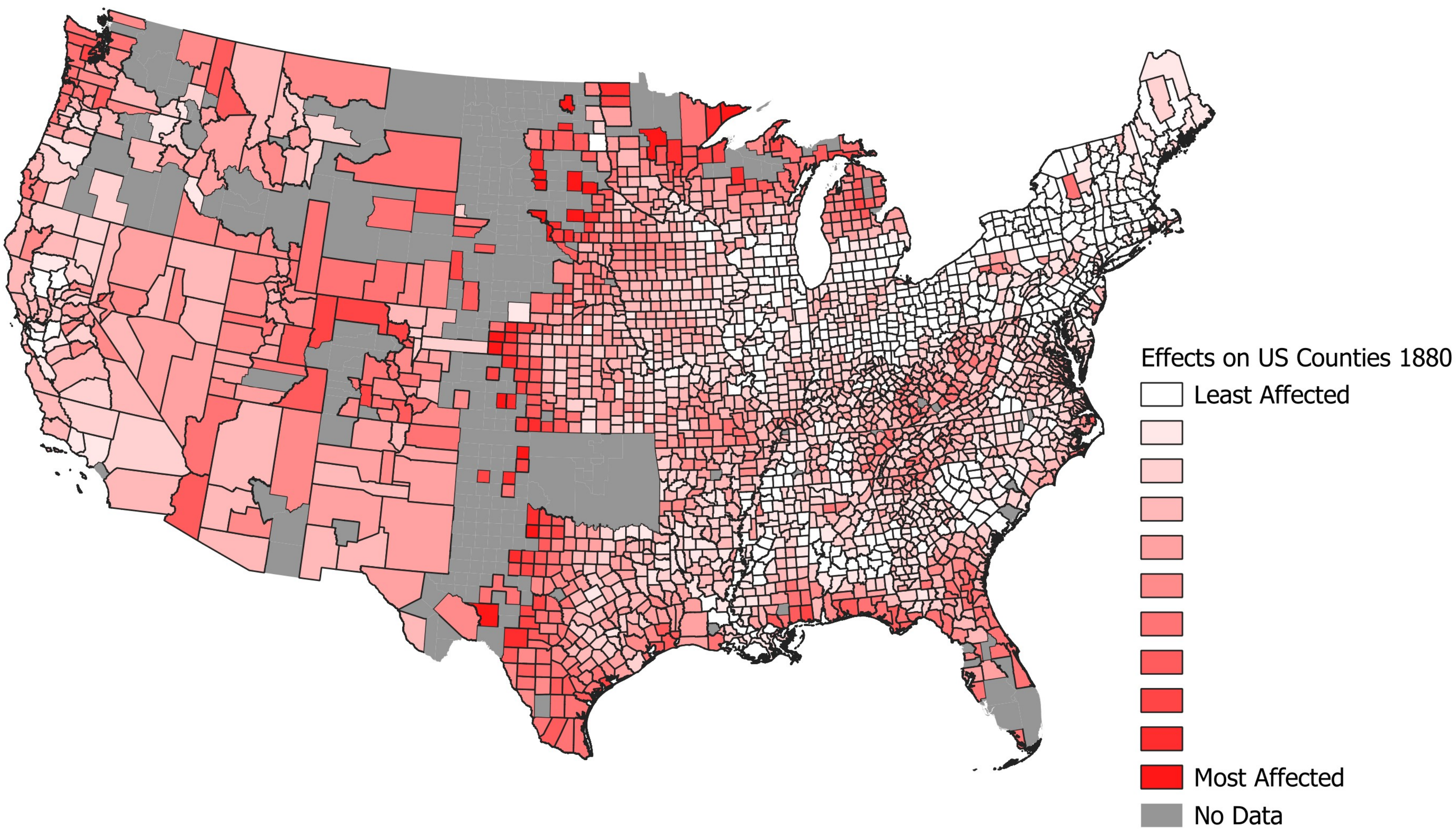


Figure 3E: Geographic Gains from Elimination of Freight Market Power - 1890

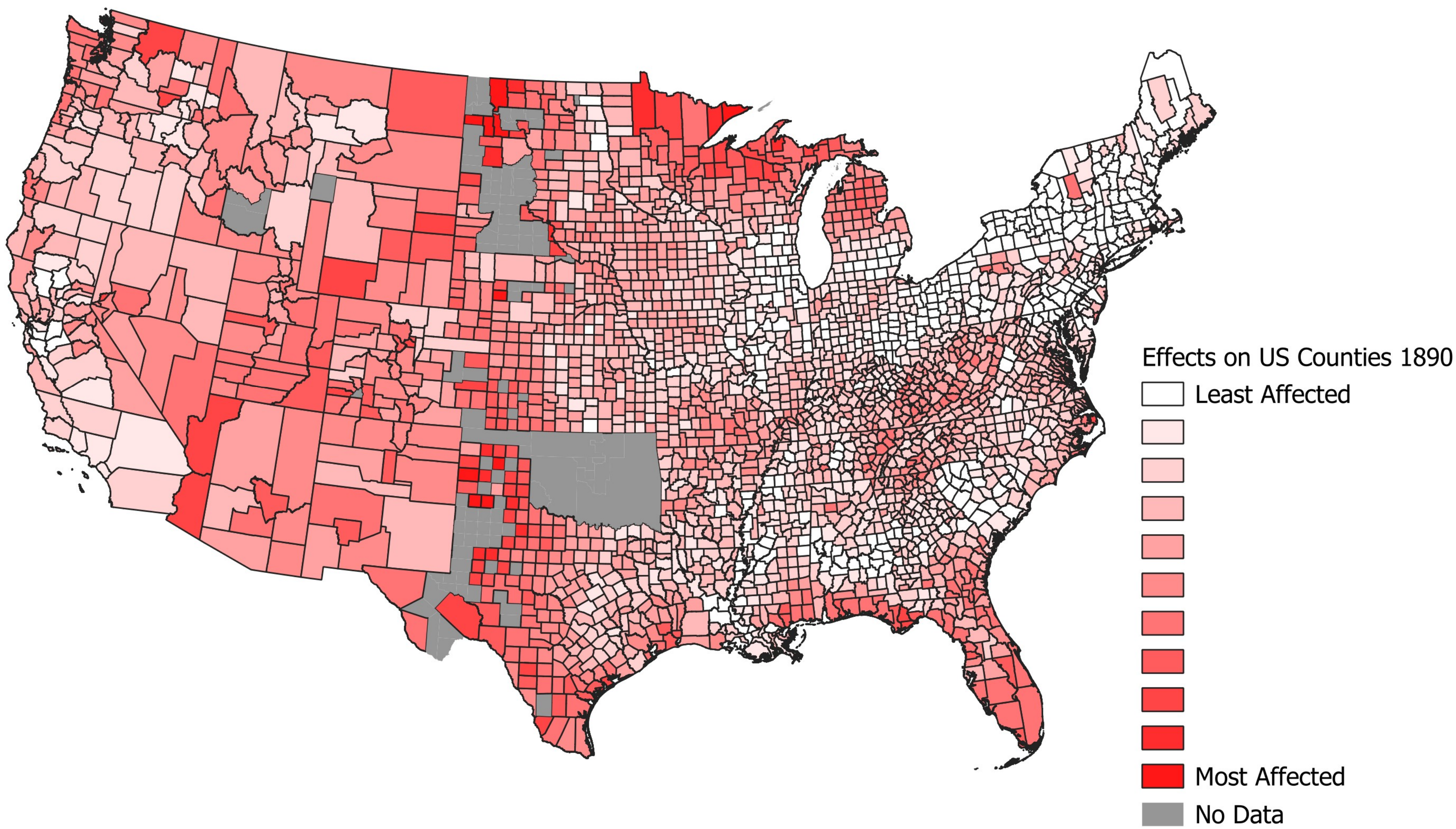


Figure 3F: Geographic Gains from Elimination of Freight Market Power - 1900

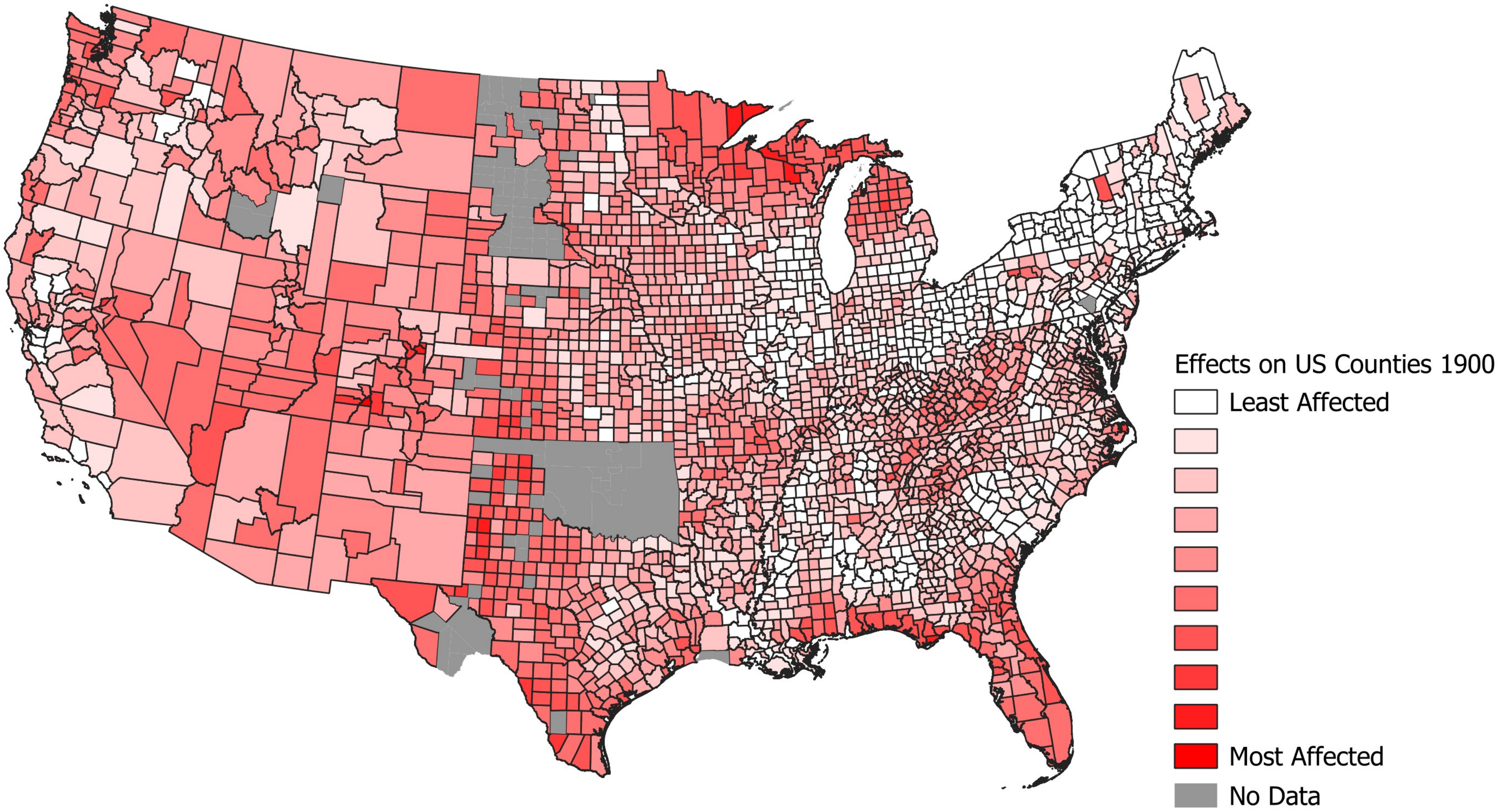
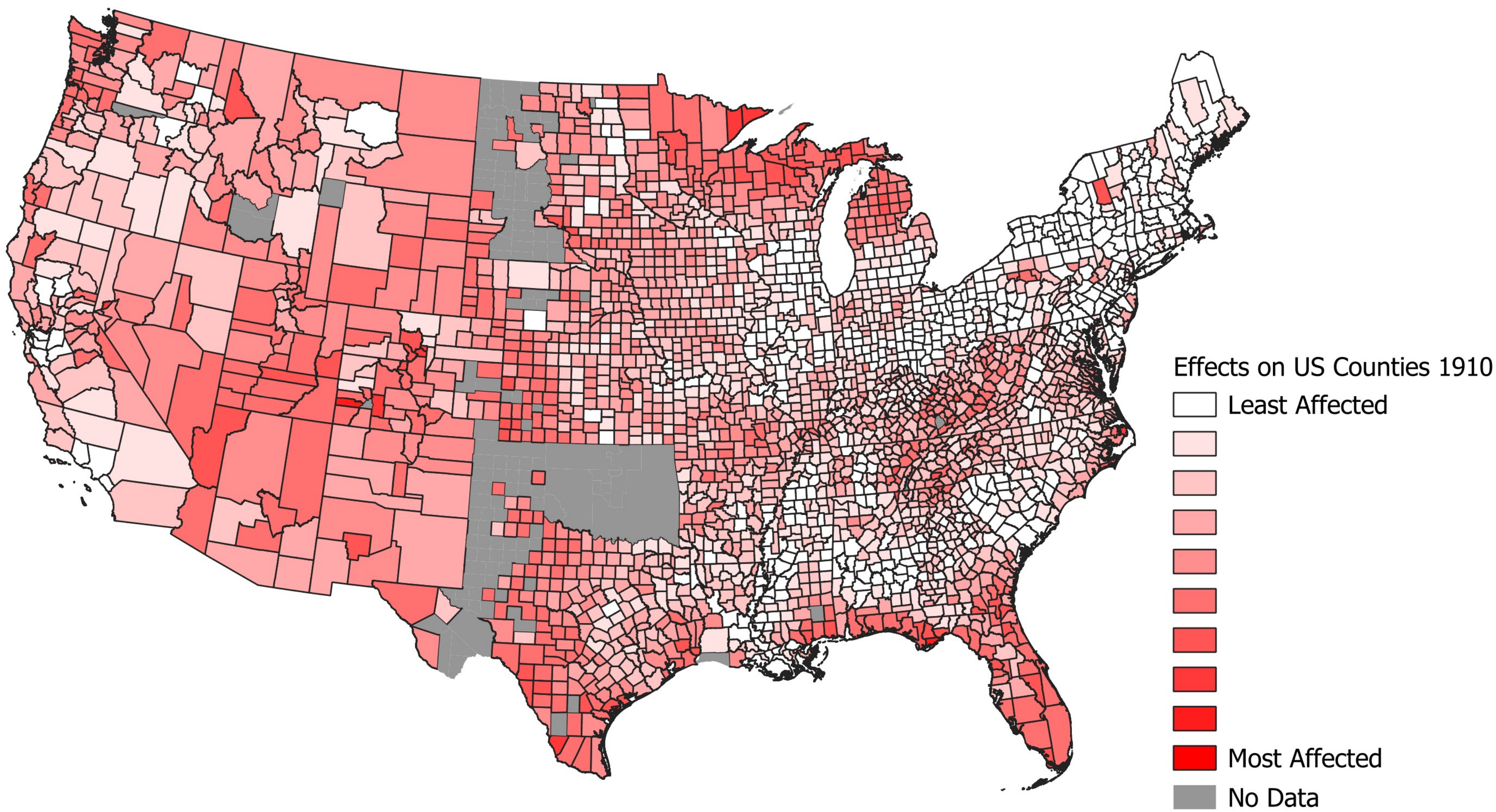


Figure 3G: Geographic Gains from Elimination of Freight Market Power - 1910



B Mathematical Appendix

1 Derivation of Pricing Rule

A representative transporter provides all transport services into location j at time t. The transporters total profit is given by

$$\Pi_{jt} = \sum_{i \neq j} \sum_{k} \zeta^k B_{jt} \pi^k_{ijt} (\mu^k_{ijt} - 1) \bar{\tau}_{ijt}.$$

where ζ^k is the constant consumption share of commodity k, B_{jt} is the (exogenous) expenditure of the location, π^k_{ijt} is the total share of trade in k to j that comes from i, μ^k_{ijt} is the markup set by the transporter, and $\bar{\tau}_{ijt}$ denotes expected trade costs between an origin i and the destination j. Taking first-order conditions yields:

$$\begin{aligned}
0 =& \left(\partial \Pi_{jt} / \partial \mu^k_{ijt}\right) \\
0 =& \left(\partial / \partial \mu^k_{ijt}\right) \sum_{i' \neq j} \sum_{k} \zeta^k B_{jt} \pi^k_{i'jt} (\mu^k_{i'jt} - 1) \bar{\tau}_{i'jt} \\
0 =& \zeta^k B_{jt} \left(\lambda \pi^k_{ijt} \bar{\tau}_{ijt} + \sum_{i' \neq j} \left(\partial \pi^k_{i'jt} / \partial \mu^k_{ijt}\right) (\mu^k_{i'jt} - 1) \bar{\tau}_{i'jt} \right) \\
0 =& \lambda \pi^k_{ijt} \bar{\tau}_{ijt} + \sum_{i' \neq j} \left(\partial \pi^k_{i'jt} / \partial \mu^k_{ijt}\right) (\mu^k_{i'jt} - 1) \bar{\tau}_{i'jt}. \qquad (24)
\end{aligned}$$

where $\lambda \in [0, 1]$ captures the state of freight market competition. Note that this formulation imposes independence of trade flows across commodities, which stems from the assumption of zero input-output structure as well as exogeneity of the budget.

It will prove expedient to define the following derivatives:

$$\begin{aligned}
\text{If } i' = i, \text{ then } \left(\partial \pi^k_{i'jt} / \partial \mu^k_{ijt}\right) =& - \theta A^k_{it} \left(c_{it} \bar{\tau}_{ijt}\right)^{-\theta} \left(\mu^k_{ijt}\right)^{-\theta - 1} \left(\phi^k_{jt}\right)^{-1} \\
& + \theta A^k_{it} \left(c_{it} \bar{\tau}_{ijt} \mu^k_{ijt}\right)^{-\theta} \left(\phi^k_{jt}\right)^{-2} A^k_{it} \left(c_{it} \bar{\tau}_{ijt}\right)^{-\theta} \left(\mu^k_{ijt}\right)^{-\theta - 1} \\
=& - \theta \pi^k_{ijt} \left(1 - \pi^k_{ijt}\right) \left(\mu^k_{ijt}\right)^{-1}.
\end{aligned}$$

$$\text{If } i' \neq i, \text{ then } \left(\partial \pi^k_{i'jt} / \partial \mu^k_{ijt}\right) = \theta A^k_{i't} \left(c_{i't} \bar{\tau}_{i'jt} \mu^k_{i'jt}\right)^{-\theta} \left(\phi^k_{jt}\right)^{-2} A^k_{it} \left(c_{it} \bar{\tau}_{ijt}\right)^{-\theta} \left(\mu^k_{ijt}\right)^{-\theta - 1}$$

$$=\theta \pi^k_{i'jt}\pi^k_{ijt}\big(\mu^k_{ijt}\big)^{-1}.$$

Plugging these expressions into Equation 24 yields:

$$\begin{aligned}
0 =& \lambda \pi^k_{ijt}\bar{\tau}_{ijt} - \theta \pi^k_{ijt}\big(1-\pi^k_{ijt}\big)\big(\mu^k_{ijt}\big)^{-1}(\mu^k_{ijt}-1)\bar{\tau}_{ijt} + \sum_{i'\neq j,i} \theta \pi^k_{i'jt}\pi^k_{ijt}\big(\mu^k_{ijt}\big)^{-1}(\mu^k_{i'jt}-1)\bar{\tau}_{i'jt} \\
=& \big(\lambda/\theta\big)\bar{\tau}_{ijt} - \big(1-\pi^k_{ijt}\big)\big(\mu^k_{ijt}\big)^{-1}(\mu^k_{ijt}-1)\bar{\tau}_{ijt} + \big(\mu^k_{ijt}\big)^{-1}\sum_{i'\neq j,i}\pi^k_{i'jt}(\mu^k_{i'jt}-1)\bar{\tau}_{i'jt} \\
=& \big(\lambda/\theta\big)\bar{\tau}_{ijt} - \big(1-\pi^k_{ijt}\big)\bar{\tau}_{ijt} + \big(1-\pi^k_{ijt}\big)\big(\mu^k_{ijt}\big)^{-1}\bar{\tau}_{ijt} + \big(\mu^k_{ijt}\big)^{-1}\sum_{i'\neq j,i}\pi^k_{i'jt}(\mu^k_{i'jt}-1)\bar{\tau}_{i'jt}.
\end{aligned}$$

Some re-organization of this equation yields:

$$\begin{aligned}
\big(1-\pi^k_{ijt}\big)\big(\mu^k_{ijt}\big)^{-1}\bar{\tau}_{ijt} + \big(\mu^k_{ijt}\big)^{-1}\sum_{i'\neq j,i}\pi^k_{i'jt}(\mu^k_{i'jt}-1)\bar{\tau}_{i'jt} =& \big(1-\pi^k_{ijt}\big)\bar{\tau}_{ijt} - \big(\lambda/\theta\big)\bar{\tau}_{ijt} \\
\big(\mu^k_{ijt}\big)^{-1}\left(\big(1-\pi^k_{ijt}\big)\bar{\tau}_{ijt} - \sum_{i'\neq j,i}\pi^k_{i'jt}(\mu^k_{i'jt}-1)\bar{\tau}_{i'jt}\right) =& \big(1-\pi^k_{ijt}\big)\bar{\tau}_{ijt} - \big(\lambda/\theta\big)\bar{\tau}_{ijt}.
\end{aligned}$$

Isolating the markup term yields:

$$\mu^k_{ijt} = \frac{(1-\pi^k_{ijt})\bar{\tau}_{ijt} + \sum_{i'\neq j,i}(\mu^k_{i'jt}-1)\bar{\tau}_{i'jt}\pi^k_{i'jt}}{(1-\pi^k_{ijt})\bar{\tau}_{ijt} - (\lambda/\theta)\bar{\tau}_{ijt}}.$$