Growth Off the Rails: Aggregate Productivity Growth in Distorted Economies*

Richard Hornbeck
University of Chicago

Martin Rotemberg
NYU

December 2022

Abstract

We examine impacts on aggregate productivity growth in the United States, as the railroad network expanded in the 19th century. Using data from the Census of Manufactures, we estimate relative increases in county aggregate productivity from relative increases in county market access. In general equilibrium, we find that the railroads substantially increased national aggregate productivity. By accounting for input distortions, we estimate much larger aggregate economic gains from the railroads than previous estimates. Our estimates highlight how broadly-used infrastructure or technologies can have much larger economic impacts when there are inefficiencies in the economy.

*For helpful comments and suggestions, we thank the editor, referees, and many colleagues and seminar participants at: Berkeley, Brown, U.S. Census Bureau, Chicago Booth, Chicago Federal Reserve, Clemson, Columbia, Columbia-NYU, Duke, EHA, Florida State, Harvard, Hunter, Iowa State, Indiana, LSE, Montreal, MSU, NBER DAE, NBER DEV, NBER EG, New Economic School, Northwestern, NYU, OECD, Oxford, PERC, Princeton, Queens, Sciences Po, SED, SMU, Toronto, UCLA, UPF, Wharton, Williams, and Zurich. We are grateful to Matt Jaremski and Chenzi Xu for providing banking data and guidance on the financial institutional context. Andrea Cerrato, William Cockriel, and Julius Luettge provided extensive research assistance. This research was funded in part by the Initiative on Global Markets at the University of Chicago Booth School of Business, the Neubauer Family Faculty Fellowship, NBER Innovation Policy grant program, and PERC. This material is based upon work supported by the National Science Foundation under Grant Number SES-1757050/1757051. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

We estimate impacts on aggregate productivity from the expansion of the railroad network, which integrated large domestic markets with vast land and commodity resources in the United States over the latter half of the 19th century. The railroads represented a technological improvement in the transportation sector, with modest direct benefits through decreased resources spent on transportation. However, we estimate that the railroads generated substantial indirect benefits through encouraging expansion in manufacturing and other sectors that were below efficient production levels. The railroads thereby generated much larger economic gains than considered by previous estimates (e.g., Fogel, 1964; Donaldson and Hornbeck, 2016), which assume zero input distortions, highlighting how broadly-used technologies or infrastructure can more-substantially impact aggregate economic growth in distorted economies.

Using newly digitized county-by-industry data from the US Census of Manufactures, we measure counties' manufacturing revenue and costs for materials, labor, and capital. We define "county aggregate productivity" or "county productivity" as the aggregate surplus each county generates (county revenues minus county costs), which sums to national aggregate productivity and which Jorgenson and Griliches (1967) describe as the "conventional" definition of productivity. In our main estimates, we focus on growth in counties' revenues, costs, and productivity. A key feature of these data is that we can use the detailed industry-level data to measure county-specific production functions, as counties produced different manufactured goods.

The manufacturing data allow us to decompose county productivity growth into two sources. The first source is increased revenues per input expenditures, called revenue total factor productivity or TFPR. TFPR growth, or the *rate* at which inputs become output, is equal to county productivity when markets are efficient (Solow, 1957). However, we find that value marginal product of inputs tends to be larger than their marginal costs, which we attribute to distortions such as markups (Hall, 1988) or input "frictions" (Hsieh and Klenow, 2009). As a consequence, changes in input-use also lead to changes in county productivity. This second source of county productivity growth reflects growth from allocative efficiency or AE: the gap between input marginal products and marginal costs means that increasing county inputs then increases county output by more (in dollars) and thereby increases county productivity (Petrin and Levinsohn, 2012; Baqaee and Farhi, 2020).

An expanding railroad network substantially decreased some counties' freight transportation costs, increasing manufacturing establishments' access to consumers, workers, and material inputs. The railroads had less benefit for counties on navigable waterways, and even could undercut those counties' access to previously captive consumers and inputs. We develop a general equilibrium model that summarizes these effects through changes in county "market access," building on Donaldson and Hornbeck (2016).

We estimate that increases in counties' market access led to substantial increases in county manufacturing activity and, because input-use in these counties was generally inefficiently low, this increase in manufacturing activity generated larger increases in revenues than costs (i.e., growth in county productivity). A one standard deviation greater increase in county market access, from 1860 to 1880, led to a 20% increase in county productivity, with similar percent impacts on county revenue and county expenditures on materials, labor, and capital. We decompose the increase in county productivity, finding impacts mostly driven by gains in allocative efficiency (AE) rather than changes in county TFPR.

Increases in county market access led to a general expansion of county economic activity, rather than systematic changes in local manufacturing industry concentration or a shift from agriculture to manufacturing. Similarly, we do not find that increases in county market access directly affected county gaps between the value marginal product of inputs and their marginal costs. Increases in county market access did not make counties more efficient; rather, it encouraged the expansion of economic activity in otherwise distorted counties that, as a consequence, led to increases in county aggregate productivity.

County market access is a function of the entire transportation network, which allows us to explore various sources of reduced form identification. While local railroad construction is potentially endogenous, and otherwise correlated with local growth, the estimated impacts from changes in county market access are robust to controlling flexibly for local railroad construction. The estimated impacts of county market access are thereby identified from more-distant changes in the railroad network, and how the spreading railroad network complemented or substituted for the previous transportation network that relied on navigable waterways for low-cost freight transportation. Places with high initial access to markets through waterways benefit less from expansion of the national railroad network, which we exploit in an instrumental variables approach that yields similar estimates to our baseline approach.

Our empirical specifications estimate relative growth in county aggregate productivity from relative increases in county market access, comparing counties that experience differential growth in market access. These estimated relative effects are not sufficient to estimate how the railroads affected *national* aggregate productivity, however, because an expanding railroad network (1) shifted production inputs between counties and (2) increased aggregate production inputs in the United States. Cross-county differences in input distortions matter for (1), but for (2) the average level of input distortions also matters, and this second channel has been particularly under-emphasized in the literature relative to its quantitative importance in our setting.

To quantify impacts of the railroads on national aggregate productivity, we extend a benchmark quantitative spatial model (Eaton and Kortum, 2002; Donaldson and Hornbeck, 2016) to include market distortions that drive a wedge between firms' value marginal product of inputs and their marginal cost. These wedges create a gap between the revenue elasticity of each input and expenditure on that input as a share of overall revenue. We use our data from the manufacturing sector to calculate key parameters of the model, including county-level input wedges. In the model, as in the data, changes in county market access do not affect the wedges. While the data only measure dollar revenues and expenditures, we can also use the model to map dollars to quantities and find that the real effects of market access are similar to the nominal effects.

Holding fixed the total US population in 1890, we estimate that national aggregate productivity would have been lower by 5.5% in 1890 in the absence of the railroads. This aggregate productivity loss without the railroads is in addition to aggregate impacts capitalized in land values (Donaldson and Hornbeck 2016, who find 3.2% losses) or differences in transportation costs (Fogel 1964, who finds 2.7% losses), so this estimated aggregate loss without the railroads is roughly triple that of previous estimates. We estimate larger aggregate economic impacts of the railroads because we allow for changes in county input-use to affect county productivity, due to county-level distortions in input-use, rather than assuming that the value marginal product of inputs is equal to their marginal cost in all counties.

This 5.5% impact on national aggregate productivity reflects changes in the allocation of economic activity across counties without the railroads, holding fixed the total US population, and we also estimate that worker utility (real wages) declines by 33% in this counterfactual scenario. As an alternative counterfactual assumption, given the substantial immigration to the United States in the 19th century, we hold fixed worker utility and allow the total US population to be lower in 1890 without the railroad network. We estimate that US population would have been 66% lower in 1890 without the railroads, which assumes frictionless international migration but is consistent with our estimated long-run intranational response of workers to changes in county market access.

We estimate that national aggregate productivity would have been 27% lower in 1890, allowing for lower aggregate population and holding fixed worker utility, with an associated annual loss of $3.2 billion (27% of GDP). For a 51% lower US population, based on the short-run intranational response of workers to changes in county market access, we estimate a 20% decline in aggregate productivity and 13% decline in worker welfare. If construction of the railroad network had stopped in 1860, as a more moderate counterfactual scenario, national aggregate productivity would have been 16% lower (with 42% lower population, holding fixed worker utility, which reflects 84% of total population growth between 1860 and

1890).

The railroads had a central role in enabling the substantial growth of the US economy, and would not have been easily replaced. We estimate a 48% annual social rate of return on the $8 billion of capital invested in the railroads in 1890 (in 1890 dollars), and estimate that the railroads privately captured only 7% of this social return. Additional canals might have been constructed in the absence of the railroads (Fogel, 1964), but we estimate that replacing the railroad network with this extended canal network would have mitigated only 12% of the aggregate losses from removing the railroad network.

Our paper extends a literature on estimating the impacts of market access to highlight the quantitative importance of market distortions (Redding and Venables, 2004; Hanson, 2005; Redding and Sturm, 2007; Head and Mayer, 2011; Duranton, Morrow and Turner, 2014; Donaldson and Hornbeck, 2016; Yang, 2018; Jaworski and Kitchens, 2019; Heblich, Redding and Sturm, 2020; Balboni, 2021). We find that input distortions create a quantitatively important additional channel through which increases in market access can generate economic gains (or losses, in principle). In doing so, our work relates to a literature that considers how the efficiency of resource allocation is affected by policies such as trade liberalization, financial regulations, and taxes (Khandelwal, Schott and Wei, 2013; Święcki, 2017; Singer, 2019; Tombe and Zhu, 2019; Bai, Jin and Lu, 2020; Berthou et al., 2020). In contrast to previous work on resource misallocation, which generally holds aggregate inputs fixed and considers the gains or losses from their reallocation (Asturias, García-Santana and Ramos, 2019; Firth, 2019; Zárate, 2021), an important feature of our analysis is how the railroads encouraged growth in aggregate inputs in the economy. By bringing this research on resource misallocation into a model of economic geography, we can explore both (1) the spatial allocation of economic activity and (2) how production expanded to use additional workers and new resources.

Our paper draws on a large literature that highlights the presence of resource misallocation in generating income differences across countries (Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009; Midrigan and Xu, 2014; Baqaee and Farhi, 2019*a*; Liu, 2019). We focus on the growth opportunities created by the variety of market distortions that can drive a wedge between firms' value marginal product and their marginal cost. Roughly one-fifth of the railroads' impact on aggregate productivity came from shifting resources across counties with heterogeneous distortions. The remaining impact on aggregate productivity came from the railroads increasing aggregate inputs in the US economy when, on average, all counties used inputs below their efficient levels. We draw on a framework that allows for changes in aggregate productivity from increased input-use without changes to the production technology itself or changes in input distortions (Hulten, 1978; Petrin and Levinsohn, 2012; Baqaee and Farhi, 2020).

Our paper highlights an important limitation underlying a long tradition in economics, back to at least Harberger (1964), of simplifying welfare analysis by assuming there are no distortions in secondary sectors or locations. David (1969) critiques Fogel (1964) on related grounds, emphasizing the potential for increasing returns to scale, while Allen and Arkolakis (2020) show how the rationale for Fogel's social savings calculation can break down in the presence of agglomeration economies. There is a persistent appeal to economic analysis, in the style of Fogel, that assigns value to some technology based on the cost of accommodating its absence. We highlight a problem with this intuition if other activities have positive social returns, and those activities would decline in the absence of the technology. Measured impacts on land values in the tradition of hedonic analyses, as in Donaldson and Hornbeck (2016), can similarly understate economic impacts dramatically because substantial economic surplus may not be paid out to land (or other factors) when there are market distortions.

Understanding the local and aggregate economic impacts of the railroads speaks to the potential for market integration to drive economic growth and, more generally, for single technological advances to generate large economic gains throughout the economy. Market distortions magnify the impacts of technologies or infrastructure that encourage other economic activities that are *marginally* productive and thereby increase the value of output by more than the increased cost of inputs. The resulting economic gains are largest when the economy is most inefficient; that is, with great problems come great possibilities.

I Data Construction

I.A Manufacturing Data and Descriptive Statistics

We use data from the US Census of Manufactures (CMF), which published county-level totals for 1850, 1860, 1870, 1880, 1890, and 1900 (Haines, 2010). These manufacturing data include total annual revenue, total cost of raw materials, total wages paid, and total value of capital invested. Revenues and materials costs reflect "factory-gate" prices, based on Census instructions to enumerators: transportation costs were included in establishment expenditures on materials, whereas revenue received by the manufacturing establishment did not include costs of shipping goods to customers. We measure annual capital expenditures by multiplying the total value of capital invested by a state-specific mortgage interest rate that varies between 5.5% and 11.4% , with an average value of 8% (Fogel, 1964). See Appendix A for more discussion of data measurement.

We digitized county-by-industry totals, published from 1860 to 1880. Our main analysis groups the reported industries into 31 industry groups, though we also report outcomes using 193 more-detailed industry categories. We assume each industry has its own Cobb-

Douglas production function. Our baseline regressions are at the county level, as industry entry and exit within counties makes it difficult to interpret percent growth at the county-industry level. To mitigate this concern, for a county-industry level analysis, we aggregate industries to five more-consistently present categories: clothing, textiles, and leather; food and beverage; lumber and wood products; metals and metal products; and other industries.

For each county-industry-year, we observe factory-gate revenue (R_{cit}) and expenditure on each input k (E^k_{cit}). As in Hsieh and Klenow (2009), properties of Cobb-Douglas production functions rationalize using inputs' cost shares to infer production function elasticities, and the relationship between producers' cost and revenue shares to infer distortions. The general view is that historical manufacturing firm returns to scale were roughly constant (Atack, 1977; Sokoloff, 1984; Margo, 2014), as in modern manufacturing (Blackwood et al., 2021), so we assume constant returns to scale in our main specification. We use national cost shares to measure industry production function elasticities: $\alpha^k_{it} = \frac{\sum_c E^k_{cit}}{\sum_c \sum_k E^k_{cit}}$. We then calculate county production functions, computing the revenue-share weighted average of the cost shares of the industries in the county and averaging across 1860-1880: $\alpha^k_c = \frac{1}{3}\sum_t \sum_i \alpha^k_{it} \frac{R_{cit}}{\sum_i R_{cit}}$. Materials are the most important input, with a production function elasticity of 0.71 in 1860, followed by labor (0.25) and capital (0.04), shown in Appendix Table 1.

County input expenditures tend to be smaller than revenue, which we rationalize with the presence of county input-specific distortions that cause "wedges" between the value marginal product of inputs and their marginal cost (ψ^k_c). As in Hsieh and Klenow (2009), we calculate $\psi^k_c = \frac{\alpha^k_c - s^k_c}{s^k_c}$, where s^k_c is input k's average share of county revenue ($s^k_c = \frac{1}{3}\sum_t \frac{\sum_i E^k_{cit}}{\sum_i R_{cit}}$). Appendix B discusses alternative methods for calculating county production function elasticities and wedges. The ψ^k_c and α^k_c terms are determined by the manufacturing data and firm cost minimization, which are independent of how we later model consumer demand and trade.

It is also useful to consider the "gap" between the value marginal product of inputs and their marginal cost. Positive gaps arise from the input wedges, since $\alpha^k_c - s^k_c = \psi^k_c s^k_c$. The "gaps" and "wedges" are closely related, but highlight different features: wedges reflect the perspective of producer optimization, since they cause producers to use inefficiently too few inputs; gaps reflect the perspective of aggregate productivity growth, since the size of the gap determines the size of potential allocative efficiency gains.

Appendix Figure 1 maps the cross-county variation in average input wedges, which are similar across inputs but moderately smaller for materials (Appendix Table 2). Average input wedges are roughly one-fifth to one-third, which is similar to measured input wedges for the modern United States (Rotemberg and White, 2021). Input wedges are initially higher in the Plains region, with associated larger input gaps (Appendix Table 3). These regional differences in the wedges are largely driven by differences in revenue shares (Appendix Table 4), rather than differences in output elasticities (Appendix Table 1). Average input wedges and gaps declined over this period, particularly in the Plains region, with the notable exception of a sharp temporary increase in labor wedges in the South in 1870 following the emancipation of enslaved people and a substantial restructuring of labor markets.

Appendix Figure 2 shows the cross-county dispersion in wedges by decade. While average input wedges declined moderately over time, there were not systematic changes in the dispersion in wedges: dispersion declined for the materials wedge, increased for the labor wedge, and remained similar for the capital wedge. In the cross section, the materials wedge has the least dispersion.

Input wedges can reflect a variety of market distortions, including markups and borrowing constraints. Producer cartels and insider lending may have contributed to misallocation in the 19th century United States (Lamoreaux, 1996; Ziebarth, 2013). County-level bank capital is itself endogenous, but we estimate that county input wedges are often lower in counties with more national-chartered banking activity and find more limited effects of state-chartered banks (Appendix Table 5). This is consistent with literature that national-chartered banks were more relevant for local manufacturing activity than state-chartered banks (Pope, 1914; Jaremski, 2014; Jaremski and Fishback, 2018; Carlson, Correia and Luck, 2022; Xu and Yang, 2022).

The correlation of the wedges and production function elasticities ranges from 0.3 (materials) to 0.4 (capital), and Appendix Figure 3 shows their joint distribution. If markups were the only source of wedges between marginal products and costs, then the correlation would be zero. The positive correlation is consistent with additional input-specific distortions (e.g., borrowing constraints) that bind more on industries that use that input more (as in Rajan and Zingales, 1998).

Appendix Table 6 reports information on 5 large industry groups, aggregating the reported industries further, along with their share of national manufacturing revenue in 1860, 1870, and 1880 (column 1). Columns 2 – 4 report industry cost shares in each decade, which are mostly stable over time and vary more across industries. Appendix Tables 7 and 8 report that average wedges and gaps are similar across industry groups.

In supplemental analysis, we use data from the Census of Manufactures on the number of manufacturing establishments and workers. We also use data from the Census of Agriculture and Census of Population, which include county-level data on the total value of home manufactures, agricultural land value, and population.

I.B Market Access Data and County-level Changes

An expanding railroad network lowered county-to-county freight transportation costs. Figure 1, panel A, shows the network of waterway routes that includes canals, navigable rivers, lakes, and oceans. Panel B shows the railroad network constructed by 1860, which then expanded by 1870 (panel C) and 1880 (panel D). Appendix Figure 4 shows the railroad network in 1890 and 1900.

Railroads and waterways both provided low-cost freight transportation routes, but the comparatively sparse waterway network required more wagon transportation that was much more expensive per ton mile. We calculate freight transportation costs between each pair of counties using the available transportation routes in each decade.[1] We also calculate transportation costs under counterfactual scenarios that remove the railroad network or replace the railroad network with an expanded canal network proposed by Fogel (1964).

We approximate the "market access" of origin county o, summing over that county's cost of transporting goods (τ) to or from each other county d with population L:

$$MA_c = \sum_{d \neq c} (\tau_{od})^{-\theta} L_d. \tag{1}$$

County o has greater market access when it is cheaper to trade with other counties d that have greater population. Changes in counties' market access summarize how changes in transportation costs affect counties through interacting goods markets and factor markets across all counties. In Section V we derive this approximation for county market access in a general equilibrium trade model with input distortions. This same approximation for market access arises in a more-restricted model without input distortions (Donaldson and Hornbeck, 2016).

For measuring county market access, as defined in equation 1, we need estimates of θ and τ_{od}. The parameter θ reflects the "trade elasticity," which varies across empirical contexts. The parameters τ_{od} represent "iceberg trade costs," which normalize the measured per ton county-to-county transportation costs t_{od} by the average price per ton of transported goods ($\tau_{od} = 1 + t_{od}/\overline{P}$).

[1] Following Donaldson and Hornbeck (2016), our main specifications set railroad rates at 0.63 cents per ton mile and waterway rates at 0.49 cents per ton mile. Transshipment costs 50 cents per ton, incurred whenever transferring goods to/from a railroad car, river boat, canal barge, or ocean liner. Wagon transportation costs 23.1 cents per ton mile, defined as the straight line distance between two points. Due to the wide dispersion in travel costs by transportation method, the key features of the transportation network in this setting concern the required length of wagon transportation and the number of transshipment points. These assumptions abstract from price variation within transportation method, for instance due to competition. See Atkin and Donaldson (2015) for discussion of a setting where markups in the transportation sector affect the incidence of decreasing trade barriers.

In Section V.C, we jointly estimate values for θ (3.05) and $\overline{P}$ (38.7). The estimated value of 38.7 for $\overline{P}$ is close to the value of 35 assumed by Donaldson and Hornbeck (2016) based on commodity price data from Fogel (1964). The estimated value of 3.05 for θ is smaller than the estimated value of 8.22 in Donaldson and Hornbeck (2016), due to differences in the model and sample period. Estimated counterfactual impacts in our framework are not sensitive to the value of θ, though, unlike in other methods of estimating the gains from trade (Arkolakis, Costinot and Rodríguez-Clare, 2012).

Figure 2 shows in darker shades those counties that have relatively greater increases in market access from 1860 to 1870 and from 1870 to 1880 (see Appendix Figure 5 for 1880 to 1890 and 1890 to 1900). Our empirical specifications compare changes in darker shaded counties to changes in lighter shaded counties. Comparing counties within nearby areas, there is substantial variation in changes in county market access. Further, comparing across decades, it is often different counties that are experiencing relatively larger or smaller changes in market access; which means that the estimated impacts of county market access do not only reflect particular counties growing relatively over the entire sample period.

Figure 2 also shows our main regression sample of 1,802 counties, which includes all counties that report manufacturing revenues and input expenditures in 1860, 1870, and 1880. We calculate county market access to all other counties with reported population, including other counties that are excluded from the regression sample because they do not report manufacturing data in each decade. We adjust the data in each decade to maintain consistent geographic units, as in Hornbeck (2010), which reflect county boundaries in 1890 and match the network database from Donaldson and Hornbeck (2016).

II Defining and Decomposing County Productivity

Our main outcome variable for county c is "county aggregate productivity" or "county productivity" in the manufacturing sector: total revenue minus total input costs, $Pr_c \equiv R_c - \sum_k E_c^k$ (Solow, 1957). We focus on dollar revenue and expenditures in the regression analysis, though we use the model in Section V to discuss how market access affects both real and nominal outcomes, since market access can affect input prices.

For considering why county productivity increases with county market access, it is useful to re-write the impact of log market access on log productivity as a function of the impacts of log market access on log revenue (R_c) and log expenditures on k inputs (E_c^k):[2]

$$\frac{\partial \ln Pr_c}{\partial \ln MA_c} = \frac{R_c}{Pr_c}\left[\frac{\partial \ln R_c}{\partial \ln MA_c} - \sum_k s_c^k \frac{\partial \ln E_c^k}{\partial \ln MA_c}\right] \tag{2}$$

$$= \nu_c\left[\frac{\partial \ln R_c}{\partial \ln MA_c} - \sum_k \alpha_c^k \frac{\partial \ln E_c^k}{\partial \ln MA_c}\right] \qquad \text{(TFPR)} \tag{3}$$

$$+\nu_c\left[\sum_k (\alpha_c^k - s_c^k)\frac{\partial \ln E_c^k}{\partial \ln MA_c}\right], \qquad \text{(AE)}$$

where $s_c^k = \frac{E_c^k}{R_c}$ or the revenue share of input k (see Appendix C.2). In equation 2, the term in brackets represents the percent impact of market access on revenue minus the (revenue share weighted) percent impact of market access on input expenditures. This term in brackets is scaled up by the ratio of county revenue to county productivity, $\nu_c = \frac{R_c}{Pr_c}$, which re-scales percent growth in county revenue into percent growth in county productivity.[3] In equation 2, we measure the relationship between county productivity and market access regardless of the production technology.

Equation 3 decomposes the impact of market access on productivity, adding and subtracting $\sum_k \alpha_c^k \frac{\partial \ln E_c^k}{\partial \ln MA_c}$, where assuming production function elasticities for α_c^k gives a useful meaning to the decomposition (Petrin and Levinsohn, 2012).

The first row reflects the impact of county market access on county TFPR (revenue total factor productivity). Increases in TFPR represent increases in revenue above-and-beyond the expected increases in revenue from increases in input expenditures. When prices are fixed, increases in county TFPR represent increases in county "technical efficiency" or physical productivity (TFPQ). Increases in county TFPR are equal to increases in county productivity if resources are allocated efficiently and the value marginal product of inputs is equal to their marginal cost.

The second row reflects the impact of county market access on county AE (allocative efficiency). AE increases when input-use increases in counties with positive "gaps" between production function elasticities and revenue shares ($\alpha_c^k - s_c^k > 0$, when $\psi_c^k > 0$). National AE increases when inputs are reallocated from low-gap to high-gap counties, or if new inputs become used in counties with positive gaps.

[2]Revenue is equal to physical output of county c (Q_c) times price P_c. Expenditure on each input is equal to physical inputs (X_c^k) times its price W_c^k.

[3]This scaling factor approaches infinity as productivity approaches zero; in practice, we use the average county scaling factor across 1860-1880 (5.1) and discuss robustness to alternative calculations in Appendix B. This scaling factor is similar to that in Hulten (1978) and Petrin and Levinsohn (2012), though it is not a Domar weight (Domar, 1961) because we are reporting the percent impact on productivity rather than the percent impact on value-added.

We use data from the Census of Manufactures to define several outcome variables for county c in year t. We start by showing effects on log revenue ($\ln(R_{ct})$), log materials expenditures ($\ln(E^M_{ct})$), log labor expenditures ($\ln(E^L_{ct})$), and log capital expenditures ($\ln(E^K_{ct})$), defined as the total values for county c in year t. Our main outcome variable is log productivity, in county c and year t, which we define as $\nu_c \left[\ln R_{ct} - \sum_k s^k_c \ln E^k_{ct}\right]$. We then define two additional outcome variables that decompose the impacts of market access on county productivity into the impacts through county TFPR ($\nu_c \left[\ln R_{ct} - \sum_k \alpha^k_c \ln E^k_{ct}\right]$) and the impacts through county AE ($\nu_c \left[\sum_k (\alpha^k_c - s^k_c) \ln E^k_{ct}\right]$). County revenue and county input expenditure vary by year, whereas fixed over time are the county revenue shares, county production function elasticities, and the scaling factor. Appendix A provides a reference for these formulas, along with further information on the underlying data from the Census of Manufactures.

III Estimating Equation

We regress outcome Y in county c and year t on log market access (MA_{ct}), county fixed effects (γ_c), state-by-year fixed effects (γ_{st}), and a cubic polynomial in county latitude and longitude interacted with year effects ($\gamma_t f(y_c)$ and $\gamma_t f(x_c)$).[4] Following the specification in Donaldson and Hornbeck (2016):

$$Y_{ct} = \beta \ln(MA_{ct}) + \gamma_c + \gamma_{st} + \gamma_t f(y_c) + \gamma_t f(x_c) + \varepsilon_{ct}. \tag{4}$$

The coefficient β reports the impact of county market access on outcome Y, comparing changes in counties with relative increases in market access to other counties within the same state and adjusting for changes associated flexibly with county latitude and longitude. We report standard errors that are clustered by state to adjust for correlation in ε_{ct} over time and within states.

The identification assumption is that counties with relative increases in market access would otherwise have changed similarly to nearby counties. In Appendix B, we show that our estimates are robust to a variety of controls for alternative sources of differential county growth, and that counties had similar growth prior to relative increases in market access.

When estimating impacts of railroads, the main identification concern is generally that railroad construction may have been directed toward counties that would have otherwise grown more for reasons other than the new railroads, as discussed by Atack et al. (2010). We estimate impacts of county market access, which depends in large part on more-distant

[4] We assign county "latitude" (y_c) and county "longitude" (x_c) using the y-coordinate and x-coordinate of the county centroid, based on an Albers equal-area projection of the United States, whose coordinates reflect consistent distances North-South and East-West.

changes in the railroad network and its interaction with the existing transportation network. We also report estimates that control flexibly for changes in railroads within a county and within nearby areas. We also exploit the interaction between railroads and pre-existing low-cost waterway transportation, whereby some counties inherently benefited less from the national railroad network, to instrument for growth in county market access.

IV Estimated Impacts of County Market Access

IV.A Estimated Impacts on Productivity

Table 1 presents results from estimating equation 4. We estimate that county market access has a substantial and statistically significant impact on county manufacturing revenue and input expenditures. Column 1 reports that a one standard deviation greater increase in market access from 1860 to 1880 leads to a 19.2% increase in revenue (panel A), 18.3% increase in materials expenditure, 19.6% increase in labor expenditure, and 15.8% increase in capital expenditure. This one standard deviation greater increase in market access corresponds to a 24% greater increase in market access from 1860 to 1880 for our baseline definition of market access.

We estimate substantial increases in county productivity from increases in county market access, as Panel E reports a 20.4% increase in log productivity. As county market access increases, there is an increase in total county revenue that substantively exceeds the increase in total county input expenditures; that is, there is increasingly more revenue produced in excess of the value of inputs used. Column 2 reports moderately larger estimates when extending the sample period to 1900, using county-aggregate manufacturing data. Column 3 reports similar impacts on county revenue and capital expenditures, when extending the sample with the available county-level data for 1850.

Table 2 shows that most of the estimated impact of market access on log productivity (panel A) is driven by growth in county AE (panel C) with an insignificant contribution from county TFPR growth (panel B). This decomposition depends on the county production function elasticities, described in Section I.A, but column 2 reports similar estimates when using a more-detailed industry classification for this calculation and column 3 reports similar estimates when assuming one production function for all manufacturing industries (from 1860 to 1880). Because market access has similar percent impacts on revenue and each input (Table 1), this implies little effect on county TFPR when production exhibits constant returns to scale. There would be more contribution from county TFPR growth under decreasing returns to scale, and less contribution under increasing returns to scale, but the impacts on county productivity continue to be driven more by county AE growth under moderate decreasing returns to scale (Appendix B). Column 4 reports larger impacts on county productivity and county AE when extending the sample through 1900 using the available aggregated county-level data only.

IV.B Sources of Growth in County Allocative Efficiency (AE)

We estimate that county AE growth is driven by increases in input expenditures, in places where distortions lead to gaps between the value marginal product of inputs and their marginal cost, but those gaps do not themselves decrease with county market access. We estimate little systematic change in the structure of the county economy itself; rather, there was a general expansion of county economic activity from increases in counties' market access.

Table 3, column 1, reports that the estimated increase in county AE (from Table 2) is largely driven by increases in materials (panel C), followed by labor (panel B), with little change from capital (panel A). Table 1 showed similar percent effects on expenditures for each input, but average county gaps are largest for materials, followed by labor and then capital (Appendix Table 3). This is because materials expenditures are the largest share of total input expenditures (Appendix Table 1), rather than input wedges being higher for materials than labor or capital (Appendix Table 2).

Table 3, columns 2 and 3, report that county input gaps and wedges do not themselves change systematically from increases in county market access. This is consistent with firm markups contributing to the measured county-level distortions, as markups would not vary with market access under CES demand and Cobb-Douglas production with constant returns to scale. There is some indication of market access decreasing labor wedges in the South and increasing input gaps in Western areas, but we do not estimate systematic impacts of market access on gaps or wedges in "frontier areas" (Bazzi, Fiszbein and Gebresilasse, 2020) and, overall, there is little systematic impact of market access on wedges and gaps within region (Appendix Table 9).

Table 3, column 4, reports that increased market access did not shift county production toward industries that are more capital-intensive, labor-intensive, or materials-intensive. Column 5 reports little change in the standard deviation of wedges across industries within a county, which suggests that inputs did not shift from more-distorted industries to less-distorted industries within counties. These estimates suggest that within-county reallocation of inputs, from marginally less-productive industries to marginally more-productive industries, is not changing county TFPR.

To further explore the within-industry impacts of market access, we run a county-industry analysis that extends our baseline specification to include county-industry fixed effects and state-year-industry fixed effects. Column 1 of Table 4 reports estimated average impacts of market access on county-industry productivity, county-industry AE, and county-industry

TFPR that are similar to our county-level estimates from Table 2, which suggests that across-industry within-county reallocation of inputs is not driving our baseline county-level estimates. Column 2 reports similar average impacts when weighting county-industries by their 1860 share of county revenue. There is some variation in industry-specific effects of market access, in columns 3 – 6, but no industry-specific effect is statistically different than the average of the other industries. Market access has little impact on the food sector, which was indeed a more local industry during our sample period: the refrigerated railcar was not used widely until the 1880s, and George Smith's patent for grinding less-perishable flour was filed in 1882 (Cronon, 2009). Appendix Table 10 reports little systematic effect of market access on gaps and wedges within these industry groups.

Increases in county manufacturing activity appear to be a general expansion of existing economic activity in the county. Column 1 of Table 5 reports little impact of county market access on the number of industries in a county. Table 5 also reports little impact of market access on the average size of establishments, measured as average revenue per establishment (column 2) or average number of workers per establishment (column 3). Instead, increases in county market access lead to a substantial increase in the number of manufacturing establishments (column 4), which is driving the overall increases in revenue and expenditures.

While county manufacturing activity increased substantially with increases in market access, we do not estimate that increases in market access prompted an economic shift from the agricultural sector toward the manufacturing sector within counties. Table 5, column 5, reports little impact of market access on county manufacturing revenue as a share of total manufacturing and agricultural revenue in the county. Similarly, columns 6, 7, and 8 report little impact on a county's manufacturing share of value-added, surplus, or employment. We also do not estimate that increases in market access encouraged economic activity in counties to become specialized in either manufacturing or agriculture (Appendix Table 11, panel A). Similarly, looking within the manufacturing sector, panel B reports little impact of market access on specialization across manufacturing industries.

IV.C Endogeneity of Railroad Construction

A main empirical concern when estimating the impacts of transportation infrastructure is that infrastructure investment is generally directed toward areas that might otherwise change differently over time. Local railroad construction might also directly impact local manufacturing activity through increases in the demand for manufactured construction materials (Fishlow, 1965). One of the advantages of analyzing changes in county market access, rather than directly estimating impacts of local railroad construction, is that much variation in counties' market access is due to changes elsewhere in the railroad network and how the railroad network interacts with other components of the transportation network (Donaldson and Hornbeck, 2016).

Table 6 reports similar impacts of county market access on county productivity when controlling flexibly for local railroad construction. Local railroad construction predicts increases in county market access, but the estimated impacts of county market access are similar when identified from more-distant changes in the railroad network and how railroad construction complemented or substituted for the previously established waterway network of rivers, canals, lakes, and oceans. Random variation in railroad construction could be used to identify impacts of market access (Borusyak and Hull, 2021), though that approach requires plausibly random construction of particular railroad lines. That approach also requires a less dense network than in the United States, such that there would be substantive variation in county market access from those particular railroad lines.

As an alternative empirical approach, we can focus on changes in county market access that are driven by how the existing waterway network interacts with changes in the railroad network. As the railroad network expands throughout the country, counties with pre-existing cheap access to markets through waterways would generally experience less increase in market access. We define county "water market access" in 1860, which reflects its measured market access when excluding all railroads from the transportation network.

Table 7, column 1, reports that counties with greater water market access in 1860 experienced less increase in market access from 1860 to 1870 and from 1870 to 1880. Under the identification assumption that counties with greater water market access would have otherwise experienced similar changes in manufacturing productivity, we can instrument for changes in county market access using county water market access in 1860. Columns 2, 4, and 6 report the 2SLS estimated effects of market access on county productivity, county TFPR, and county AE, which are less precise but similar in magnitude to the OLS estimates (columns 3, 5, and 7).

The waterway IV identification assumption would be violated if these counties with greater water market access would have changed differently from 1860 through 1880, however, and a different empirical approach would estimate the effects of county market access controlling for counties' water market access in 1860 (interacted with decade, allowing these places to change differently over time). Appendix Table 12 reports these estimates, which are similar to those in Table 2, and also reports similar estimates when controlling for counties' 1860 market access (interacted with decade). Appendix B reports that our estimates are generally not sensitive to controlling for other sources of differential growth (Appendix Table 12), using alternative methods of measuring productivity (Appendix Table 13), and using alternative measures of market access (Appendix Table 14).

V Aggregate Counterfactual Analysis

We now examine how national aggregate productivity was affected by changes in county market access due to the railroads. The regressions estimate that relative increases in county productivity were mostly driven by relative increases in input-use, in counties with positive gaps where the value marginal product of inputs exceeded their marginal cost, but some of this relative increase reflects shifting inputs from other counties that also have positive gaps. An expanding railroad network may also increase aggregate inputs in the US economy. By drawing further on the model's structure, we can consider national aggregate economic impacts from the expansion of the railroad network and national aggregate economic losses under counterfactual transportation networks.

We add input distortions to a baseline model of economic geography (Eaton and Kortum, 2002; Donaldson and Hornbeck, 2016). Workers maximize utility and, in long-run equilibrium, are indifferent between counties. Firms maximize profits. Goods markets clear when total production in a county equals total consumption in that county, net of transportation costs. Each county is impacted by changes in transportation costs across the country, through linked goods markets and factor markets, and the effects on each county are summarized by changes in a county's market access.

We show how the presence of input distortions causes there to be national aggregate productivity gains from reductions in transportation costs that are not captured by changes in land values (as in Donaldson and Hornbeck, 2016) or transportation cost savings (as in Fogel, 1964). We estimate the national aggregate productivity losses from counterfactual transportation networks, such as removing the railroad network or replacing the railroad network with an extended network of canals, along with impacts on worker welfare under alternative assumptions about international labor mobility.

V.A Model Setup

Following Eaton and Kortum (2002), each origin county o has an exogenous Hicks-neutral technical efficiency level $z_o(j)$, for each variety j, drawn from a Fréchet distribution with CDF given by: $F_o(z) = 1 - e^{-A_o z^{-\theta}}$, with $\theta > 1$. The parameter A_o captures average technical efficiency in county o (absolute advantage), while the parameter θ captures the standard deviation of technical efficiency across varieties (scope for comparative advantage). A smaller θ or "trade elasticity" is associated with more-dispersed technical efficiency across varieties, larger incentives to trade across counties, and a less elastic response of cross-county trade flows to trade costs.

Firms in each county have a Cobb-Douglas production function for good variety j, $z_o(j) \prod_{k \in \{L,K,T,M\}} {X_o^k}^{\alpha_o^k}$, using labor L, capital K, land T, and materials M. Firms use a continuum of good varieties as materials, with a constant elasticity of substitution across varieties, and so W_o^M is the CES price index over good varieties in county o.[5] The marginal cost of producing variety j in county o is:

$$MC_o\left(j\right) = \frac{\prod_k \left(W_o^k\right)^{\alpha_o^k}}{z_o\left(j\right)}, \tag{5}$$

where W_o^k and α_o^k are the price and production function elasticities, respectively, for input k in county o.

The main addition in our model, compared to Donaldson and Hornbeck (2016), is that firms face input-specific frictions. These frictions are exogenous and represent market inefficiencies that discourage further use of labor (ψ^L), capital (ψ^K), land (ψ^T), or materials (ψ^M). Given positive frictions, firms in county o reduce production of variety j such that its price is greater than its marginal cost of production:

$$p_o\left(j\right) = \frac{\prod_k \left(\left(1+\psi_o^k\right) W_o^k\right)^{\alpha_o^k}}{z_o\left(j\right)} > MC_o\left(j\right). \tag{6}$$

Firm markups, due to market power, can be represented as a common component in each of the ψ terms. Given a positive firm markup, or positive input-specific frictions, the value created by firms increasing production would be greater than the value of resources used. Firms are unable or unwilling to use more inputs, however, as reflected in the ψ terms.

County input prices reflect factor mobility. We assume capital is mobile, such that interest rates are fixed exogenously, but this interest rate varies across counties (as in Fogel, 1964). As in Donaldson and Hornbeck (2016), we assume that the United States faces a perfectly elastic supply of capital and that the nominal price of capital relative to the price index in New York City is fixed. County land prices are endogenous, and we assume the total quantity of physical land is fixed in each county. Labor is supplied by workers, who consume good varieties j in the same manner that firms use these varieties in roundabout production; the CES price index workers pay for their consumption basket, P_o, is the same price index W_o^M paid by local firms for their inputs (Redding and Venables, 2004). We assume workers spend their income in their home county and that payments to land and capital, along with any profits, are also spent in each county in proportion to its share of aggregate revenues. Workers' indirect utility in county o is $V(P_o, W_o^L) = W_o^L/P_o$. We assume workers are mobile across counties, focusing on a long-run equilibrium in which workers can arbitrage real wage

[5] $W_o^k \equiv \left[\int_0^n (p_o(j))^{1-\sigma} dj\right]^{1/(1-\sigma)}$, where σ is the elasticity of substitution across varieties j, n is the exogenous measure of varieties, and $p_o(j)$ is the price of variety j in county o.

differences, such that worker utility is constant across counties ($\bar{U}$). County wage rates are endogenous and reflect local prices, so nominal wages are higher in counties with higher goods prices:

$$W_o^L = \bar{U} P_o \tag{7}$$

There is costly trade of good varieties across counties, for both final goods and intermediate goods (materials). Transporting goods from county o (origin) to county d (destination) incurs a proportional "iceberg" trade cost $\tau_{od} > 1$. In county d, the price of good variety j produced in county o is given by: $p_{od}(j) = \tau_{od} p_o(j)$, where $p_o(j)$ is the price in county o.

V.B Solving for Market Access

We begin by deriving the gravity equation for cross-county trade flows, allowing for input frictions. When firms sell good varieties from county o in county d, the offered price in county d reflects the factory-gate price in county o (equation 6) and the transportation cost (τ_{od}). Consumers buy good varieties from their cheapest source, where "consumers" includes workers buying goods and firms buying materials. Following Eaton and Kortum (2002), with the addition of input frictions ψ, the value of total exports from county o to county d is given by:[6]

$$\text{Exports}_{od} = \kappa_1 A_o \left(\prod_k \left(\left(1 + \psi_o^k\right) W_o^k \right)^{\alpha_o^k} \right)^{-\theta} \tau_{od}^{-\theta} Y_d P_d^{\theta}. \tag{8}$$

County o sends more goods to county d when county o has higher technical efficiency (A_o) or lower "effective costs" $\left(\prod_k \left(\left(1 + \psi_o^k\right) W_o^k \right)^{\alpha_o^k} \right)$, where "effective costs" reflect input prices and distortions. County o also sends more goods to county d when bilateral transportation costs are lower (τ_{od}), and when county d has higher income (Y_d).

Producers in a county sell more to destinations with a high price index P_d, whereas producers and consumers purchase more from origins with a low price index P_o (Appendix Section C.1). These forces are proportional, so we can express "market access" in county o as a function of the endogenous number of workers in each other county d:[7]

$$MA_o = \kappa_2 \sum_d \tau_{od}^{-\theta} L_d MA_d^{\frac{-(1+\theta)}{\theta}} \frac{\left(1 + \psi_d^L\right)}{\alpha_d^L}. \tag{9}$$

[6] Here, $\kappa_1 = \left[\Gamma\left(\frac{\theta+1-\sigma}{\theta}\right)\right]^{-\frac{\theta}{1-\sigma}}$, where $\Gamma(\cdot)$ is the Γ function defined by $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx$.
[7] Here, $\kappa_2 = \bar{U} \rho^{\frac{1+\theta}{\theta}}$

Market access is higher in county o when it is cheaper to trade with more-populated counties that have less access to other markets, a greater labor cost share, and a higher labor input friction. This equation for market access simplifies to the corresponding equation 9 in Donaldson and Hornbeck (2016), when there are no labor input frictions and a homogeneous labor production function elasticity.

We solve for market access in each county and each decade (and counterfactual scenario). The equilibrium values of market access are the solutions to the N-by-N system of equations (equation 9) with N unknowns (market access in each county n). County population (L_d) comes from the Census of Population in each decade, and we discuss below how we measure the additional components of equation 9 and related parameters for estimating counterfactual changes in county population. Given the demand function, county production function parameters, distortions, transportation costs, national capital price, and total population, an equilibrium is a matrix of wages (w_o), prices (P_o), trade flows (Exports_{od}), and populations (L_o) such that Equations 6, 7, 8, and 9 hold for all origin counties o and destination counties d.

V.C Estimating Parameters

For measuring the origin county input frictions, ψ_o^k, we use the input frictions described in Section I.A from the manufacturing sector in each county. Our baseline approach assumes that county input frictions in the agricultural sector and other sectors are the same as in the manufacturing sector in that county, due to the absence of detailed data on input expenditures outside the manufacturing sector.[8]

For the county production function elasticities, α_o^k, we start with the county manufacturing sector elasticities from Section I.A. For the agricultural sector, we assume national elasticities based on values from Caselli and Coleman (2006) and Towne and Rasmussen (1960).[9] We calculate county-level elasticities in 1890 as the weighted average of county manufacturing sector elasticities and agricultural sector elasticities, where the weights are

[8]The counterfactual sample is 2,722 counties with positive population and positive manufacturing or agricultural revenue in 1890, of which 309 sparsely-populated counties do not report manufacturing data in 1880 (which we assume reflects zero manufacturing revenue). We use the 1880 county-by-industry manufacturing data to measure ψ_o^k, as in Section I.A, when these data are available. We use 1890 county-level manufacturing data for 176 counties with no manufacturing data in 1880 (4% of US population in 1890), then use 1900 county-level manufacturing data for 69 counties with no manufacturing data in 1890 (1.5% of US population in 1890), and then use state median values of ψ_o^k in 1880 for the remaining 64 counties (0.2% of US population in 1890).

[9]We use value-added elasticities from Caselli and Coleman (2006) and the materials input share from Towne and Rasmussen (1960), giving us production function elasticities of 0.552 for labor, 0.1932 for capital, 0.1748 for land, and 0.08 for materials. For economic activity outside the agricultural sector, we continue to assume a "fixed factor" share of 0.1748 and assign county-specific elasticities for materials, labor, and capital from the manufacturing sector in 1880 from Section I.A.

each sector's share of total county revenue. Given county production function elasticities α_o^k, the county input frictions ψ_o^k then imply county input revenue shares s_o^k.

For the origin-to-destination trade costs, τ_{od}, we use the calculated transportation costs from Section I.B. The network database calculates transportation cost per ton (t_{od}), whereas trade costs in the model have a proportional "iceberg" form (τ_{od}). We reconcile these by estimating an average price per ton of goods shipped in the economy ($\overline{P}$), such that: $\tau_{od} = 1 + t_{od}/\overline{P}$.

For the county-level fundamentals — average productivity (A_o) and the quantity of fixed factors (T_o) — we solve for the values that rationalize the observed distribution of population in 1890 (see Appendix C.3).

Appendix C.3 describes how we jointly estimate $\overline{P}$ and the trade elasticity θ, using data on total railroad shipments, aggregate revenue, and county land values. Broadly, we iterate to find the best values: first find the value of θ that minimizes the residual sum of squares between the model-predicted relationship between land values and market access and the corresponding relationship in the data (conditional on $\overline{P}$), as shown in Appendix Figure 6. Given θ, we then find the $\overline{P}$ that minimizes the difference between actual and model-implied total railroad shipments, as shown in Appendix Figure 7. This procedure allows us to estimate $\overline{P}$ in each county up to a proportional constant γ, which we estimate by minimizing the distance between nominal output in the model and in the data. We repeat the process until the values converge on estimates $\overline{P} = 38.7$ and $\theta = 3.05$. For these values of $\overline{P}$ and θ, the estimated impact of market access on county land value is 0.286 (0.037) from estimating equation 4.

V.D Predicted Impacts of Market Access

We now describe how market access affects county productivity, and how we aggregate from impacts on county productivity to national aggregate productivity. The effect of market access on productivity in county o is given by:

$$\frac{d \ln Pr_o}{d \ln MA_o} = \nu_o \sum_k \left(\alpha_o^k - s_o^k\right) \frac{d \ln X_o^k}{d \ln MA_o}. \tag{10}$$

Market access increases county productivity by increasing real input usage ($\frac{d \ln X_o^k}{d \ln MA_o}$), when the value marginal product of that input exceeds its marginal cost (when $\alpha_o^k > s_o^k$ or $\psi_o^k > 0$).

Appendix C derives the log-linear impact of market access on each input ($\frac{d \ln X_o^k}{d \ln MA_o}$). Market access increases capital usage by $\frac{\alpha_o^M + \alpha_o^L + 1}{1 + \theta \alpha_o^T}$ percent. Market access has a larger impact on labor and materials usage ($\frac{\alpha_o^M + \alpha_o^L + 1}{1 + \theta \alpha_o^T} + \frac{1}{\theta}$ percent), where there is an additional term $\left(\frac{1}{\theta}\right)$ because market access also decreases materials costs and nominal wages in county o.

To measure national aggregate productivity growth, expressed in percent terms relative to national value-added (GDP), we sum the growth in county productivity:[10]

$$APG \quad = \quad \sum_{o} D_o \, \Sigma_k \left(\alpha_o^k - s_o^k\right) d \ln X_o^k. \tag{11}$$

D_o is the Domar (1961) weight for county o (county revenue divided by national value-added). Our counterfactual analysis assigns each county the average of its factual and counterfactual Domar weight, which in both scenarios sum to 1.6. These Domar weights sum to more than 1, and are the appropriate way to aggregate county-level changes in settings with intermediate goods (Hulten, 1978) and distortions that generate a gap between the value marginal product of inputs and their marginal cost (Petrin and Levinsohn, 2012; Baqaee and Farhi, 2019*b*).

V.E Model Interpretation and Discussion of Assumptions

Our model highlights the impact of market access on county productivity and national aggregate productivity, which arises due to input distortions and generates economic gains from the railroads that are in addition to impacts on land values considered by Donaldson and Hornbeck (2016). In that model without distortions, all economic gains from increased market access are captured by the increase in land values (i.e., capitalized in the price of the fixed factor when there is frictionless international migration and worker utility is fixed). Increased land values reflect greater factor input payments, but increased productivity reflects gains in output that are not paid to inputs. Thus, in our model, the aggregate economic gains from the railroads are given by the increase in national aggregate productivity in addition to the increase in land values.

We make some important assumptions to maintain tractability in this general equilibrium setting, while extending the model to include input frictions. First, we assume that county input frictions are exogenous, which is consistent with our estimates that measured wedges and gaps were not impacted by changes in market access (Table 3). Second, we assume that county production function elasticities are exogenous, which is consistent with our estimates that county market access growth did not change county manufacturing cost shares (Table 3) or the county manufacturing share (Table 5). Third, we assume that county technical efficiency is exogenous, which is consistent with estimates from Table 2 that showed little impact of market access on county TFPR. Impacts of market access on TFPR may understate impacts on county technical efficiency (TFPQ) due to lower output prices, though TFPR is

[10] To go from equation 10 to equation 11, we multiply by county productivity (Pr_o), sum the level increase across counties, and express that sum relative to national value-added (GDP). Petrin and Levinsohn (2012) and Baqaee and Farhi (2019*b*) provide alternative derivations that simplify to equation 11 for settings, like ours, in which technical efficiency is constant.

often correlated with technical efficiency in settings when both are measured (Foster, Haltiwanger and Syverson, 2008; Haltiwanger, Kulick and Syverson, 2018). Our counterfactual analysis considers impacts on aggregate productivity only through changes in allocative efficiency, under the assumption that technical efficiency does not also decline in the absence of the railroads. We use the model to compare impacts on real and nominal productivity, and report counterfactual impacts on real productivity.

V.F Model Validation Exercises

We undertake a few exercises to validate implications of the model. First, we use the model to generate predicted county-level outcomes in 1860, 1870, and 1880, where the only primitive allowed to vary over "time" is the railroad network and the model parameters are estimated in 1890. Table 8, Column 2, reports similar impacts of market access on nominal outcomes in this simulated data as in our reduced-form results. The model replicates the reduced-form results despite our not disciplining model parameters with the estimated relationship between market access and manufacturing revenues or expenditures. This similarity reflects two countervailing forces, though: (1) the model reflects a somewhat larger response of real inputs to market access, which pushes up the effect on productivity; but (2) in the model market access has no effect on TFPR, while in the data market access leads to a small (insignificant) increase in TFPR.

We can measure quantities and prices separately in the model-generated data, and Column 3 of Table 8 reports the model-generated relationship between market access and real values. The prices of labor and materials decrease with market access, which decreases output prices, so the effect of market access on real productivity is slightly larger than its impact on nominal productivity. This implies that the previous regression estimates understate the impact of market access on real county AE because measured increases in input expenditures understate the increase in real input usage. Nevertheless, the real and nominal values of productivity are highly correlated: for each counterfactual scenario, their cross-county correlation is above 0.99.

One reason the model generates a larger productivity response than in the data is that workers are instantaneously mobile across counties in the model, whereas in the data workers might respond over time (Allen and Donaldson, 2022). Appendix Table 15 shows the estimated impact of market access and lagged market access on log manufacturing employment. Over a 20 year period, a one standard deviation increase in market access leads to a 0.270 log point increase in manufacturing employment, which is similar to the model value of 0.259 from Table 8, Column 2. Of this total response, 77% of the migration response comes in the first decade.

The model predicts that the effect of market access on revenue is similar within a county, across industries, regardless of the industries' initial gaps. This contrasts with the predicted effect on productivity, where industries with higher initial gaps should see larger productivity growth from the same change in revenue. Appendix Table 16 shows estimates consistent with this theoretical result. In the county-industry data, the impact on revenue from market access does not vary with the counties' average wedge (Panel A), but this interaction effect does predict a greater increase in productivity (Panel B). This result comes from adding the interaction to our baseline regression (in Column 1), or using county-by-year fixed effects to absorb the direct effects of market access and identify the interaction coefficient on variation in wedges across industries within each county (in Column 2).

The model predicts log-linear impacts of market access on county productivity and, indeed, Appendix Figure 8 shows approximately log-linear impacts of market access on county productivity (panels A and B) that are driven by county AE growth (panels C and D) with little change in county TFPR (panels E and F). This pattern holds for model-derived market access (equation 9) and the first-order approximation of market access (equation 1).

The model-derived changes in county market access, from 1860 to 1880, have a correlation coefficient above 0.99 with the first-order approximated changes in county market access used in the regression analysis. These measures need not be so highly correlated in other empirical settings, for instance if large destinations experience very different changes in market access, but this appears to be rare and applications of this method generally use the first-order approximated formula. While the reduced-form analyses can rely on the approximation in equation 1, the full expression for market access is required for aggregate counterfactual analysis because we need to determine not only relative changes in county market access but the absolute changes in counties' market access under counterfactual scenarios.

V.G Estimated Counterfactual Impacts

We now estimate the decline in national aggregate productivity from removing the railroad network or from other counterfactual transportation networks. To benchmark the magnitudes, US aggregate productivity growth in manufacturing was about 2.2% annually from 1860 to 1900 in our data, with roughly a quarter coming from technical efficiency growth. Similarly, for the whole economy in this era, estimates indicate around a 0.5% annual growth in technical efficiency (Abramovitz and David, 1973).

Figures 3 and 4 map the county-level declines in market access and productivity when removing the railroad network, such that county-to-county freight transportation must rely on the existing waterway network and high-cost wagon transportation. Darker-shaded counties represent larger counterfactual declines in market access and productivity in the absence of the railroads, as economic activity shifts toward the waterway network. The declines in county productivity reflect counterfactual declines in market access and production inputs, interacted with county-level "gaps" between the value marginal product of inputs and their marginal cost (Appendix Figure 9), as in equation 10. National aggregate productivity declines as aggregate inputs decline, given positive average gaps, and additionally as inputs shift from counties with larger gaps to counties with smaller gaps.

We estimate that national aggregate productivity would have been 26.7% lower in the United States in 1890, if there were no railroad network (Table 9, panel A, column 1). This 26.7% decline in national aggregate productivity reflects only decreases in allocative efficiency, with no decline in county technical efficiency. This is equivalent to roughly 12 years of aggregate productivity growth at the rate of growth in manufacturing.

This 26.7% decrease in national aggregate productivity is worth 26.7% of GDP annually, or $3.2 billion in 1890 dollars. As a comparison, the estimated cost of the railroad network in 1890 was $8 billion (Adams, 1895). We estimate that this investment in the railroads generated an annual social return of 48%, and that the railroad sector privately captured only 7% of its social return in 1890.[11]

Rather than removing the entire railroad network, we can also consider the counterfactual economic losses if the railroad network had stopped expanding. For example, in 1890, we estimate that productivity would have been 2.6% lower using only railroads that existed in 1880, 9.9% lower using railroads from 1870, 15.6% lower using railroads from 1860, or 22.3% lower using railroads from 1850 (Table 9, panel A, columns 2 – 5). Expansion of the railroad network after 1860 contributed roughly one-fifth of total manufacturing productivity growth, given the observed annual growth rate of 2.2%.

Additional canals might have been constructed to mitigate national productivity losses, in the absence of the railroad network, but we find that these canals would have been an ineffective substitute for the railroad network. We evaluate the system of feasible canals proposed by Fogel (1964), estimating that productivity would have been lower by 23.5% in 1890 when replacing the railroad network with these additional canals (Table 9, panel A,

[11] We estimate that the railroads generated an annual private return of 3.5% in 1890. For this calculation, based on numbers from Adams (1895), we sum the railroads' reported net income ($145 million), debt interest payments ($217 million), net capital expenditure ($5 million), and subtract losses not otherwise reflected from some companies ($30 million) along with subtracting income from other sources ($52 million). We then divide $285 million by the cost of the railroads including equipment ($8.041 billion) and value of land ($80 million). Much of the railroads' reported transportation expenses were maintenance costs (39% or $271 million), and we interpret the reported "permanent improvements" of $5 million as total capital expenditure minus depreciation. To calculate the annual social return, we sum the annual private return ($285 million), our estimated annualized increase in agricultural land value ($414 million), and our estimated increase in annual productivity ($3.204 billion), divided by the cost of the railroads including equipment and land ($8.121 billion).

column 6). That is, the additional canals would have mitigated only 12% of the national aggregate productivity loss from removing the railroad network.

By contrast, the railroads would have been "cheap at twice the price." We estimate that productivity would have been lower by 8.7% in 1890 if railroad rates were double (Table 9, panel A, column 7). Compared to losing access to the railroad network entirely, using these more-expensive railroads would mitigate 67% of the national productivity decline.

In estimating the decline in aggregate productivity, we consider several scenarios for counterfactual changes in US total population. Our baseline estimates reflect the counterfactual decline in total population that holds fixed worker utility (real wages). We also consider a scenario that holds fixed total population, and calculate the associated decline in worker utility, along with scenarios that reflect intermediate declines in total population.

When allowing for aggregate declines in population, the model predicts a substantial decline in population in the United States. For worker utility to be unchanged in the counterfactual, the model predicts that the US population would need to be 66% lower in 1890. This is similar to the 58% counterfactual decline in population estimated by Donaldson and Hornbeck (2016), but this population decline has much greater economic impact in our analysis because the marginal product of inputs is allowed to be greater than their marginal cost. By comparison, the US population was 39% lower in 1870 and 73% lower in 1840 than in 1890 (United States Census Bureau, 1975). If we replace the 1890 railroad network with the 1860 railroad network, the model explains 84% of total population growth between 1860 and 1890.

When holding fixed total population, we solve for counties' population shares in the absence of the railroad network. We then calculate corresponding changes in other production inputs, revenue, and the resulting change in aggregate productivity in the United States. Panel B reports that national aggregate productivity is estimated to fall by 5.5% in 1890, in the absence of the railroad network, when maintaining the same total US population. Population and other production inputs become condensed into limited geographic areas, decreasing labor productivity due to an increase in the land-labor ratio and increasing goods prices, such that worker utility falls by 32.7%. Intuitively, the incidence of economic gains from the railroads falls more on workers when their mobility is restricted.

These two counterfactual scenarios highlight the relative contributions from the average level of distortions in the economy, as compared to variation in the distortions across counties. We can decompose the the 26.7% aggregate productivity decline, without the railroads, into an average component from county inputs changing given positive average national gaps

$(\overline{\alpha_o^k - s_o^k})$ and a residual term that reflects idiosyncratic county gaps:

$$APG = \sum_o D_o \, \Sigma_k \overline{(\alpha_o^k - s_o^k)} d\ln X_o^k + \sum_o D_o \, \Sigma_k \left((\alpha_o^k - s_o^k) - \overline{(\alpha_o^k - s_o^k)} \right) d\ln X_o^k. \tag{12}$$

We calculate a 20.7% decline in aggregate productivity from the average component, with one-fifth of the total effect driven by the residual component that itself is similar to the estimated loss in overall aggregate productivity when holding aggregate population fixed. The contribution from the average component is consistent with multiplying the sum of the revenue-weighted average input gaps for the whole economy (0.132) by the average counterfactual decline in county inputs (1.11 log points) and by the sum of the county Domar weights (1.6) to get a national aggregate productivity decline of 0.234 log points or 21%. The sum of the average input gaps (0.131) largely reflects the gap for materials (0.073), whereas the gap for capital is 0.012, so the aggregate productivity decline is not sensitive to capital distortions that are more subject to measurement error.

The railroads and an expanding US economy encouraged immigration and aggregate population growth, but the true counterfactual response in aggregate population is likely somewhere between our extreme scenarios of a complete migration response (holding utility fixed) and no response (holding aggregate population fixed). We cannot directly estimate the impact of county market access on immigration and aggregate population growth in the US, but can provide a benchmark using relative worker movement within the US. The within-US response of workers to market access is 77% of the model-predicted full migration response within the first decade (Appendix Table 15). Appendix Table 17 shows a range of counterfactual estimates, for alternative assumptions on aggregate population declines. If we assume the aggregate population decline is 77% of the population decline predicted from a full migration response, which corresponds to a 51% decline in total population, then counterfactual aggregate productivity falls by 20% (and utility falls by 13%). If we instead assume that total US population would be lower by 33% in the absence of the railroads, which excludes the foreign-born population in 1890 and white native-born children of foreign-born parents, then we estimate a 14% decline in productivity and a 22% decline in worker utility.

National aggregate productivity falls in the counterfactual scenarios because of gaps between the value marginal product of inputs and their marginal cost, but manufacturing gaps in our data are not large in comparison to other eras. For the United States, in 1997, the manufacturing gap is around 0.3 using the NBER-CES database (Becker, Gray and Marvakov, 2013; Jaumandreu, 2022). Thus, the substantial impacts of the railroads on national aggregate productivity are not driven by especially large measured gaps in the historical data; rather, the effects are driven by moderately-sized gaps and the substantial impacts of the railroads on both the relative allocation of inputs across counties and aggregate inputs in the United States.

We would estimate zero impact of the railroads on national aggregate productivity, mechanically, if we assumed zero gaps between inputs' value marginal product and marginal cost. Our baseline counterfactual assumes the measured frictions in the manufacturing sector also reflect frictions in the agricultural sector. If we assume no distortions outside of the manufacturing sector, we estimate an aggregate productivity loss of 16.5% in 1890 without the railroad network (Table 10, column 2). The estimated counterfactual impacts also become moderately smaller if we adjust counties' measured input expenditures using counties' measured materials wedges as a proxy for capital wedges or labor wedges (Table 10, columns 3 and 4). The measurement of capital expenditures is particularly subject to measurement error, but capital expenditures are a small share of total input expenditures and so assuming zero misallocation in capital only moderately reduces the aggregate productivity impact to 20.2% (column 5). If we decrease the dispersion in capital wedges (or all input wedges) wedges by 5, 10, or 25 percent, the counterfactual estimates are within one percentage point of our baseline estimate.

Our estimated counterfactual impacts on national aggregate productivity vary moderately with the estimated value of $\overline{P}$ (average price per ton of traded goods) and are not sensitive to the estimated value of θ (the trade elasticity) in columns 6 – 10 of Table 10. Alternative values of θ re-scale market access but have little effect on the estimated impacts because we use the observed distribution of population across counties to discipline the model. As a result, changes in estimated county-level parameters (e.g., A_o) counteract the change in θ. The estimated counterfactual impacts are more sensitive to the estimated value of $\overline{P}$ because higher values of $\overline{P}$ effectively re-scale the baseline transportation cost parameters and diminish differences between the factual and counterfactual scenarios.

VI Interpretation

We estimate substantially larger economic gains from the railroads, as a share of GDP, than previous estimates of 3.2% (Donaldson and Hornbeck, 2016) or 2.7% (Fogel, 1964). Our estimated impact on national aggregate productivity also supplements those previous estimates: we would estimate no impact on national aggregate productivity if there were no differences between counties' value marginal product of inputs and their marginal cost (as assumed by Fogel (1964) and Donaldson and Hornbeck (2016)), whereas the economy would still benefit from the railroads decreasing resources spent on transportation (as in

Fogel 1964) or economic gains capitalized in land values (as in Donaldson and Hornbeck, 2016).

Our analysis starts with the manufacturing sector and extends this analysis to the broader economy, whereas Fogel (1964) and Donaldson and Hornbeck (2016) start with the agricultural sector and extend their analyses to the broader economy. In considering impacts on the broader economy, the key difference in our approaches is where those economic gains will appear: for Fogel (1964), the benefits from railroads are confined to the transportation sector through savings in transportation costs; for Donaldson and Hornbeck (2016), the aggregate impacts are capitalized in land values. In our model that allows for market distortions, the difference between output value and input costs is not capitalized in land values and so there can also be impacts of the railroads on national aggregate productivity that are not captured by changes in total land value.

One general implication for measuring the economic incidence of new infrastructure or new technologies is that increased payments to land (or labor or capital) do not include all economic gains when there are market distortions. We show that these additional economic gains can be substantively large, particularly when new infrastructure or new technologies are broadly used and encourage substantial expansion of economic activity. As in Baqaee and Farhi (2020), TFP growth in one sector (transportation) can increase production in other sectors that were inefficiently small and thereby generate larger aggregate productivity gains than implied by the Domar-weighted increase in transportation sector TFP.

The railroads decreased transportation costs, effectively subsidizing the expansion of economic activities throughout the economy that had a positive social return (i.e., activities whose value marginal product exceeded their marginal cost). The more that economic activity expands in response to decreased transportation costs, the greater the aggregate economic gains, which is opposite to the intuition of Fogel (1964, 1979) in which the railroads' impacts were supposed to be bounded above by assuming an inelastic demand for transportation.

We do not find that railroads reduced market distortions, whether due to firm markups or input frictions, but the railroads generated substantial national aggregate productivity gains by encouraging the expansion of an economy with market distortions. There would also be large potential gains from reducing distortions (as in Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009): estimating a counterfactual that removes all input distortions, and maintains the railroad network in 1890, we estimate that national aggregate productivity would increase by 110% (holding worker utility fixed) or by 30% (holding total population fixed). But market integration need not decrease market distortions; indeed, estimated distortions in modern US data are similar to this historical era, such that there continue to be large potential aggregate productivity gains from further increases in input-use.

These same aggregate productivity gains accrue whether market distortions are due to input frictions or market power. Aggregate productivity increases as inputs are reallocated to firms with more market power, even if that increases the average markup in the economy. Even if high-markup firms are not especially productive, in a physical sense, their high price means that consumers on the margin would value more of that good than the goods being produced with those inputs by lower-markup firms. Further, aggregate productivity increases when inputs move to higher marginal productivity firms even if those firms are less productive on average.

We do not find that the railroads increased county TFPR, and we hold counties' technical efficiency fixed in our counterfactual estimates. Some specifications indicate a larger impact of market access on county TFPR, and our estimated impacts on county TFPR may also understate impacts on physical productivity, and future research can explore impacts of market access on firm-level production decisions and physical productivity. Our estimated increases in county-level production are associated with substantial increases in the number of establishments, with little change in average establishment size, which relates to a literature highlighting the role of entry in aggregate productivity growth (Foster, Haltiwanger and Krizan, 2001; Foster, Haltiwanger and Syverson, 2008). The railroads were also associated with increased patenting activity, though in part through encouraging the filing of lower-quality patents (Perlman, 2017).

Increases in national aggregate productivity are not synonymous with increases in welfare, given any social welfare function, but increases in the difference between total output value and total input costs (aggregate productivity) represent additional resources that society may consume and so are closely associated with increases in welfare (Solow, 1957; Weitzman, 1976; Basu and Fernald, 2002). There is additional surplus in society when the value of output increases by more than the cost of inputs, but we do not consider the distribution of that surplus across people and how that might be weighted. We report substantial losses in national aggregate productivity without the railroads, holding fixed worker utility (real wages), but we also report substantial losses in worker welfare when total population is held fixed or partially restricted in the counterfactual.

VII Conclusion

We estimate that the railroads drove substantial productivity growth in the United States, playing a central role in the economy's growth through the latter half of the 19th century. The railroads integrated domestic markets within the United States, shifting economic activity across counties and increasing aggregate economic activity. We estimate that increases in county aggregate productivity were mostly driven by increases in county AE (allocative efficiency): input-use increased substantially in counties where the value marginal product of inputs was greater than their marginal cost, increasing the value of output more than the value of inputs even if holding fixed county TFPR (revenue total factor productivity).

We emphasize that new technologies or new infrastructure can be particularly impactful when there are market distortions in the economy, such that economic activity increases in places where the value marginal product of inputs is greater than their marginal cost. These potential economic gains are largest when the economy is most inefficient; that is, with great problems come great possibilities.

We find that the railroads generated large indirect economic gains, outside the transportation sector, by increasing marginally productive activities in other sectors. These indirect economic gains were substantially larger than the direct gains from decreased resources spent on transportation itself (i.e., the "social savings" proposed by Fogel 1964) or gains capitalized in land values (as in Donaldson and Hornbeck 2016). The railroads generated large indirect gains because they encouraged a substantial expansion of economic activity in the United States, and this same mechanism would apply to a variety of new technologies or infrastructure investments that encourage the substantial expansion of other activities that have value marginal product greater than marginal cost.

Our counterfactual analysis does not include impacts of the railroads on physical productivity (technical efficiency), or ways in which local or aggregate technological innovation might respond to increases in market access. Further research could use more-detailed firm-level data to explore impacts of market access on technical efficiency, firm-level specialization, technology adoption, and other ways in which market integration could further increase local and aggregate productivity.

We also do not consider a variety of other mechanisms through which railroads may have impacted the US economy. Our analysis does not consider how the construction and operation of the railroads may have directly affected the economy, such as through the development of improved management practices (Chandler, 1965). We also do not consider how the railroads may have impacted worker mobility, both across counties and within urban areas. The railroads encouraged certain economic activities to agglomerate in major urban centers, with potential benefits from urbanization (Haines and Margo, 2008) and particular gains in major cities (Cronon, 2009). Our empirical analysis complements city histories, examining how a broad range of counties were induced to grow by the railroads and increases in market access. The railroads shifted economic activity from some counties to others, along with increasing aggregate economic activity in the United States, and these effects combined to generate both local productivity gains and substantial national aggregate productivity gains.

References

Abramovitz, Moses, and Paul A David. 1973. "Reinterpreting economic growth: parables and realities." *American Economic Review*, 63(2): 428–439.

Adams, Henry C. 1895. *Report on transportation business in the United States at the eleventh census: 1890.* Vol. 1, US GPO.

Allen, Treb, and Costas Arkolakis. 2020. "The welfare effects of transportation infrastructure improvements."

Allen, Treb, and Dave Donaldson. 2022. "Persistence and path dependence in the spatial economy." *Working Paper.*

Andrews, Isaiah. 2018. "Valid two-step identification-robust confidence sets for GMM." *Review of Economics and Statistics*, 100(2): 337–348.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare. 2012. "New trade models, same old gains?" *American Economic Review*, 102(1): 94–130.

Asturias, Jose, Manuel García-Santana, and Roberto Ramos. 2019. "Competition and the welfare gains from transportation infrastructure: Evidence from the Golden Quadrilateral of India." *Journal of the European Economic Association*, 17(6): 1881–1940.

Atack, Jeremy. 1977. "Returns to scale in antebellum United States manufacturing." *Explorations in Economic History*, 14(4): 337.

Atack, Jeremy, Fred Bateman, Michael Haines, and Robert A Margo. 2010. "Did railroads induce or follow economic growth? Urbanization and population growth in the American Midwest, 1850–1860." *Social Science History*, 34(2): 171–197.

Atkin, David, and Dave Donaldson. 2015. "Who's Getting Globalized? The Size and Implications of Intra-national Trade Costs."

Bai, Yan, Keyu Jin, and Dan Lu. 2020. "Misallocation under trade liberalization." *Working Paper.*

Balboni, Clare. 2021. "In Harm's Way? Infrastructure Investments and the Persistence of Coastal Cities." *Working Paper.*

Baqaee, David, and Emmanuel Farhi. 2019*a*. "The macroeconomic impact of microeconomic shocks: beyond Hulten's Theorem." *Econometrica*, 87(4): 1155–1203.

Baqaee, David, and Emmanuel Farhi. 2019*b*. "A Short Note on Aggregating Productivity."

Baqaee, David, and Emmanuel Farhi. 2020. "Productivity and misallocation in general equilibrium." *Quarterly Journal of Economics*, 135(1): 105–163.

Basu, Susanto, and John G Fernald. 2002. "Aggregate productivity and aggregate technology." *European Economic Review*, 46(6): 963–991.

Bazzi, Samuel, Martin Fiszbein, and Mesay Gebresilasse. 2020. "Frontier culture: The roots and persistence of "rugged individualism" in the United States." *Econometrica*, 88(6): 2329–2368.

Becker, Randy, Wayne Gray, and Jordan Marvakov. 2013. "NBER-CES manufacturing industry database: Technical notes." *NBER Working Paper*, 5809.

Berthou, Antoine, Jong Hyun Chung, Kalina Manova, and Charlotte Sandoz Dit Bragard. 2020. "Productivity, (mis)allocation and trade."

Blackwood, G Jacob, Lucia S Foster, Cheryl A Grim, John Haltiwanger, and Zoltan Wolf. 2021. "Macro and Micro Dynamics of Productivity: From Devilish Details to Insights." *American Economic Journal: Macroeconomics*, 13(3): 142–72.

Borusyak, Kirill, and Peter Hull. 2021. "Non-Random Exposure to Exogenous Shocks: Theory and Applications." *Working Paper*.

Carlson, Mark, Sergio Correia, and Stephan Luck. 2022. "The effects of banking competition on growth and financial stability: Evidence from the national banking era." *Journal of Political Economy*, 130(2): 462–520.

Caselli, Francesco, and Wilbur John Coleman. 2006. "The world technology frontier." *American Economic Review*, 96(3): 499–522.

Chandler, Alfred D. 1965. "The railroads: pioneers in modern corporate management." *Business History Review*, 39(1): 16–40.

Cronon, William. 2009. *Nature's metropolis: Chicago and the Great West.* WW Norton & Company.

David, Paul A. 1969. "Transport Innovation and Economic Growth: Professor Fogel on and off the Rails." *Economic History Review*, 22(3): 506–524.

Domar, Evsey D. 1961. "On the measurement of technological change." *Economic Journal*, 71(284): 709–729.

Donaldson, Dave, and Richard Hornbeck. 2016. "Railroads and American economic growth: A "market access" approach." *Quarterly Journal of Economics*, 131(2): 799–858.

Duranton, Gilles, Peter M Morrow, and Matthew A Turner. 2014. "Roads and Trade: Evidence from the US." *Review of Economic Studies*, 81(2): 681–724.

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

Firth, John. 2019. "I've Been Waiting on the Railroad: The Effects of Congestion on Firm Production." *Working Paper*.

Fishlow, Albert. 1965. *American Railroads and the Transformation of the Ante-bellum Economy.* Vol. 127, Harvard University Press.

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

Fogel, Robert W. 1979. "Notes on the Social Saving Controversy." *Journal of Economic History*, 39: 1–54.

Foster, Lucia, John C Haltiwanger, and Cornell John Krizan. 2001. "Aggregate productivity growth: lessons from microeconomic evidence." In *New developments in productivity analysis*. 303–372. University of Chicago Press.

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008. "Reallocation, firm turnover, and efficiency: selection on productivity or profitability?" *American Economic Review*, 98(1): 394–425.

Haines, Michael R. 2010. "Historical, Demographic, Economic, and Social Data: The United States, 1790–2002." *ICPSR*, 2896.

Haines, Michael R, and Robert A Margo. 2008. "Railroads and local economic development." *Quantitative Economic History: The good of counting*.

Hall, Robert. 1988. "The relation between price and marginal cost in US industry." *Journal of Political Economy*, 96(5): 921–947.

Haltiwanger, John, Robert Kulick, and Chad Syverson. 2018. "Misallocation measures: The distortion that ate the residual."

Hanson, Gordon H. 2005. "Market potential, increasing returns and geographic concentration." *Journal of International Economics*, 67(1): 1–24.

Harberger, Arnold. 1964. "Taxation, resource allocation, and welfare." In *The role of direct and indirect taxes in the Federal Reserve System*. 25–80. Princeton University Press.

Head, Keith, and Thierry Mayer. 2011. "Gravity, market potential and economic development." *Journal of Economic Geography*, 11(2): 281–294.

Heblich, Stephan, Stephen J Redding, and Daniel M Sturm. 2020. "The making of the modern metropolis: evidence from London." *Quarterly Journal of Economics*, 135(4): 2059–2133.

Hornbeck, Richard. 2010. "Barbed wire: Property rights and agricultural development." *Quarterly Journal of Economics*, 125(2): 767–810.

Hsieh, Chang-Tai, and Peter J Klenow. 2009. "Misallocation and manufacturing TFP in China and India." *Quarterly Journal of Economics*, 124(4): 1403–1448.

Hulten, Charles R. 1978. "Growth accounting with intermediate inputs." *Review of Economic Studies*, 45(3): 511–518.

Jaremski, Matthew. 2014. "National Banking's role in US industrialization, 1850–1900." *Journal of Economic History*, 74(1): 109–140.

Jaremski, Matthew, and Price V Fishback. 2018. "Did inequality in farm sizes lead to suppression of banking and credit in the late nineteenth century?" *Journal of Economic History*, 78(1): 155–195.

Jaumandreu, Jordi. 2022. "The Remarkable Stability of the US Manufacturing Markups." *Working Paper*.

Jaworski, Taylor, and Carl T. Kitchens. 2019. "National Policy for Regional Development: Historical Evidence from Appalachian Highways." *Review of Economics and Statistics*, 101(5): 777–790.

Jorgenson, Dale Weldeau, and Zvi Griliches. 1967. "The Explanation of Productivity Change." *Review of Economic Studies*, 34(3): 249–283.

Khandelwal, Amit K, Peter K Schott, and Shang-Jin Wei. 2013. "Trade liberalization and embedded institutional reform: Evidence from Chinese exporters." *American Economic Review*, 103(6): 2169–95.

Lamoreaux, Naomi R. 1996. *Insider lending: Banks, personal connections, and economic development in industrial New England.* Cambridge University Press.

Liu, Ernest. 2019. "Industrial policies in production networks." *Quarterly Journal of Economics*, 134(4): 1883–1948.

Margo, Robert A. 2014. "Economies of Scale in Nineteenth-Century American Manufacturing Revisited: A Resolution of the Entrepreneurial Labor Input Problem." In *Enterprising America: Businesses, Banks, and Credit Markets in Historical Perspective*. 215–244. University of Chicago Press.

Midrigan, Virgiliu, and Daniel Yi Xu. 2014. "Finance and misallocation: Evidence from plant-level data." *American Economic Review*, 104(2): 422–458.

Perlman, Elisabeth. 2017. "Dense enough to be brilliant: Patents, Urbanization, and Transportation in Nineteenth Century America." *Working Paper*.

Petrin, Amil, and James Levinsohn. 2012. "Measuring aggregate productivity growth using plant-level data." *RAND Journal of Economics*, 43(4): 705–725.

Pope, Jesse E. 1914. "Agricultural credit in the United States." *Quarterly Journal of Economics*, 28(4): 701–746.

Rajan, Raghuram G, and Luigi Zingales. 1998. "Financial Dependence and Growth." *American Economic Review*, 88(3): 559–586.

Redding, Stephen, and Anthony Venables. 2004. "Economic geography and international inequality." *Journal of International Economics*, 62(1): 53–82.

Redding, Stephen, and Daniel M. Sturm. 2007. "The Cost of Remoteness: Evidence from German Division and Reunification." *American Economic Review*, 98(5): 1766–1797.

Restuccia, Diego, and Richard Rogerson. 2008. "Policy distortions and aggregate productivity with heterogeneous establishments." *Review of Economic dynamics*, 11(4): 707–720.

Rotemberg, Martin, and T Kirk White. 2021. "Plant-to-Table(s and Figures): Processed Manufacturing Data and Measured Misallocation." *Working Paper*.

Singer, Gregor. 2019. "Endogenous markups, input misallocation and geographical supplier access."

Sokoloff, Kenneth L. 1984. "Was the transition from the artisanal shop to the nonmechanized factory associated with gains in efficiency? Evidence from the US Manufacturing censuses of 1820 and 1850." *Explorations in Economic History*, 21(4): 351–382.

Solow, Robert M. 1957. "Technical change and the aggregate production function." *Review* *of Economics and Statistics*, 312–320.

Sun, Liyang. 2018. "Implementing valid two-step identification-robust confidence sets for linear instrumental-variables models." *The Stata Journal*, 18(4): 803–825.

Święcki, Tomasz. 2017. "Intersectoral distortions and the welfare gains from trade." *Journal of International Economics*, 104: 138–156.

Tombe, Trevor, and Xiaodong Zhu. 2019. "Trade, migration, and productivity: A quantitative analysis of china." *American Economic Review*, 109(5): 1843–72.

Towne, Marvin, and Wayne Rasmussen. 1960. "Farm Gross Product and Gross Investment in the Nineteenth Century." In *Trends in the American Economy in the Nineteenth Century*. 255–316. NBER.

United States Census Bureau. 1975. *Historical Statistics of the United States, Colonial Times to 1970.* US Department of Commerce, Bureau of the Census.

Weitzman, Martin L. 1976. "On the welfare significance of national product in a dynamic economy." *Quarterly Journal of Economics*, 90(1): 156–162.

Xu, Chenzi, and He Yang. 2022. "Real effects of stabilizing private money creation." *CEPR Discussion Papers*, 2(17164).

Yang, Yang. 2018. "Transport infrastructure, city productivity growth and sectoral reallocation: Evidence from China." *Working Paper*, 18(276).

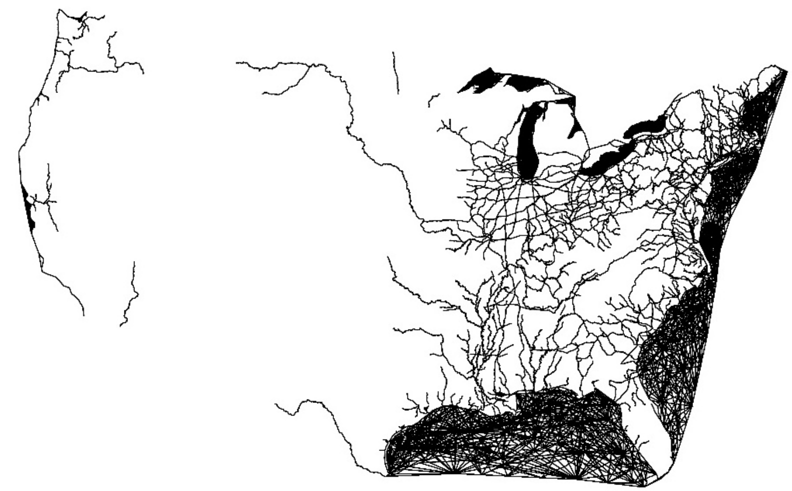
Zárate, Román David. 2021. "Factor Allocation, Informality and Transit Improvements: Evidence from Mexico City." *Working Paper*.

Ziebarth, Nicolas L. 2013. "Are China and India backward? Evidence from the 19th century US Census of Manufactures." *Review of Economic Dynamics*, 16(1): 86–99.

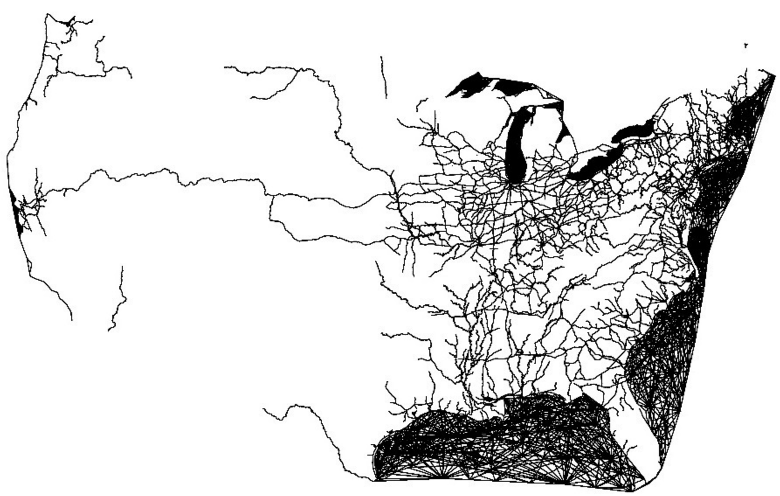
Figure 1. Waterways and Railroads, by Decade

A. Waterways

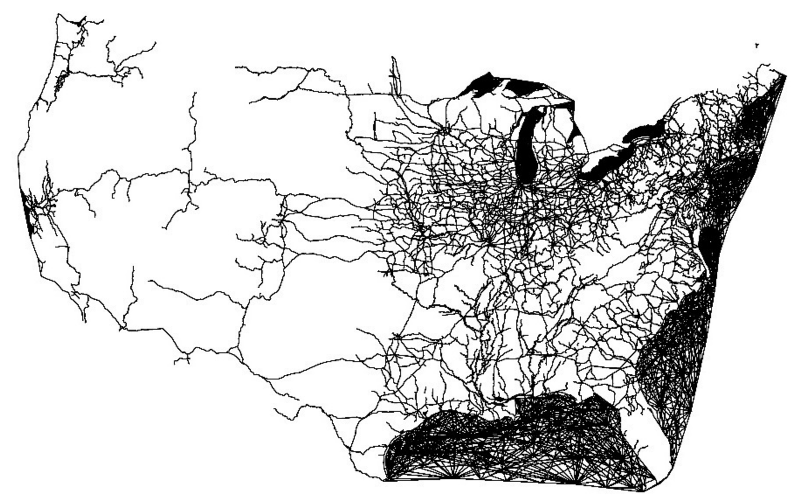
B. Waterways and 1860 Railroads



C. Waterways and 1870 Railroads



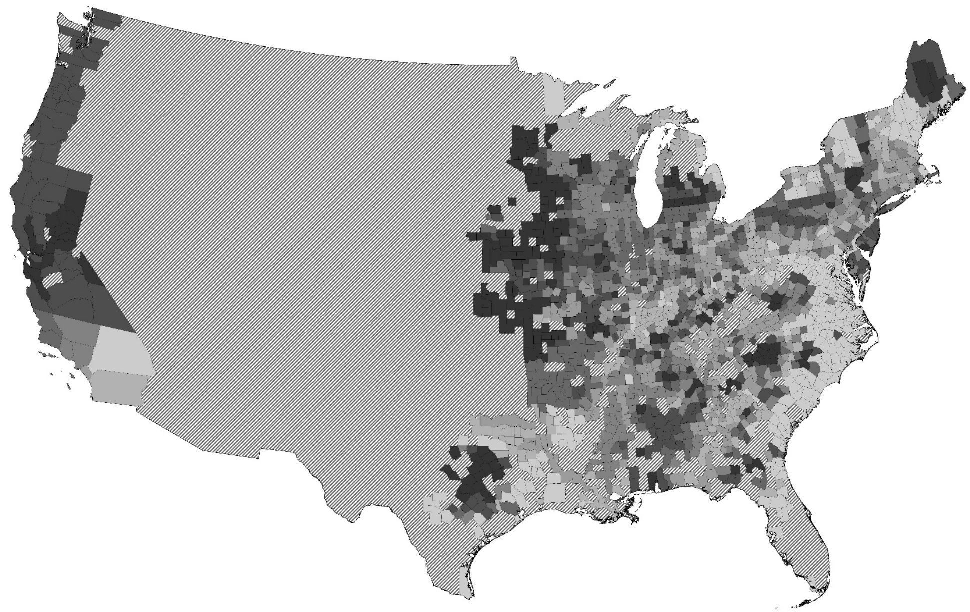
D. Waterways and 1880 Railroads



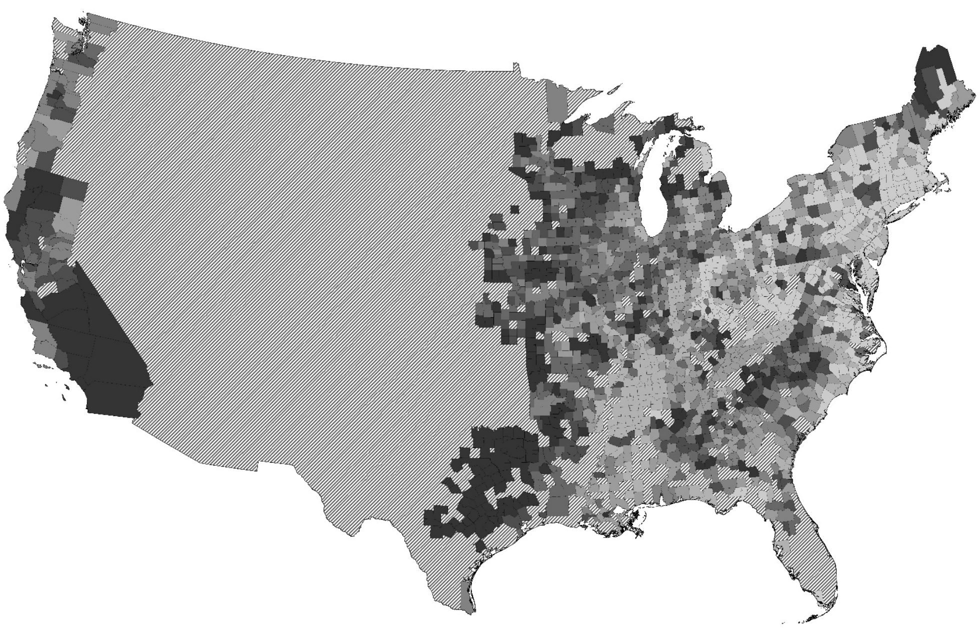
Notes: Panel A shows the waterway network: natural waterways (including navigable rivers, lakes, and oceans) and constructed canals. Panel B adds railroads constructed by 1860, Panel C adds railroads constructed between 1860 and 1870, and Panel D adds railroads constructed between 1870 and 1880.

Figure 2. Calculated Changes in Log Market Access, by County

A. From 1860 to 1870

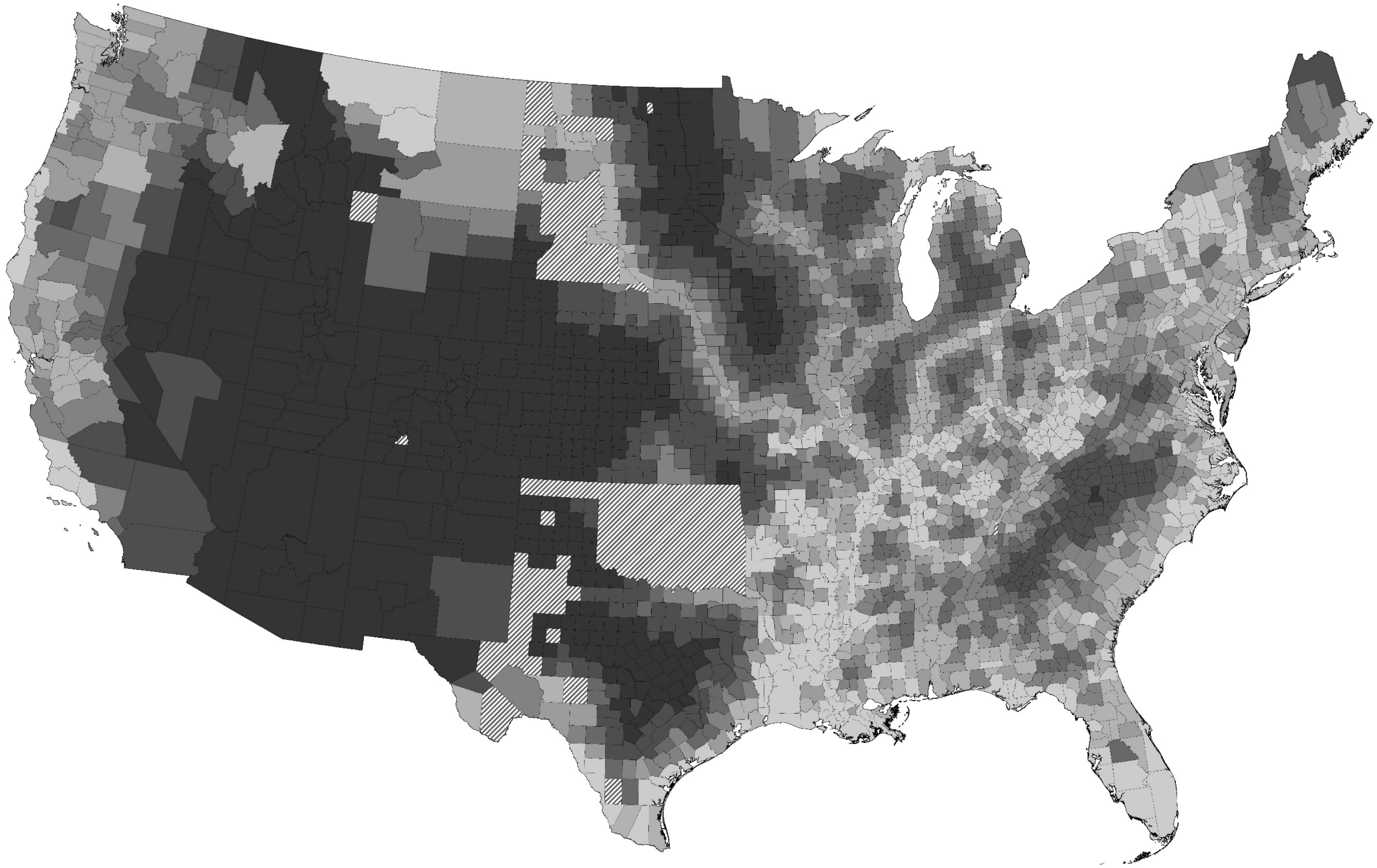


B. From 1870 to 1880



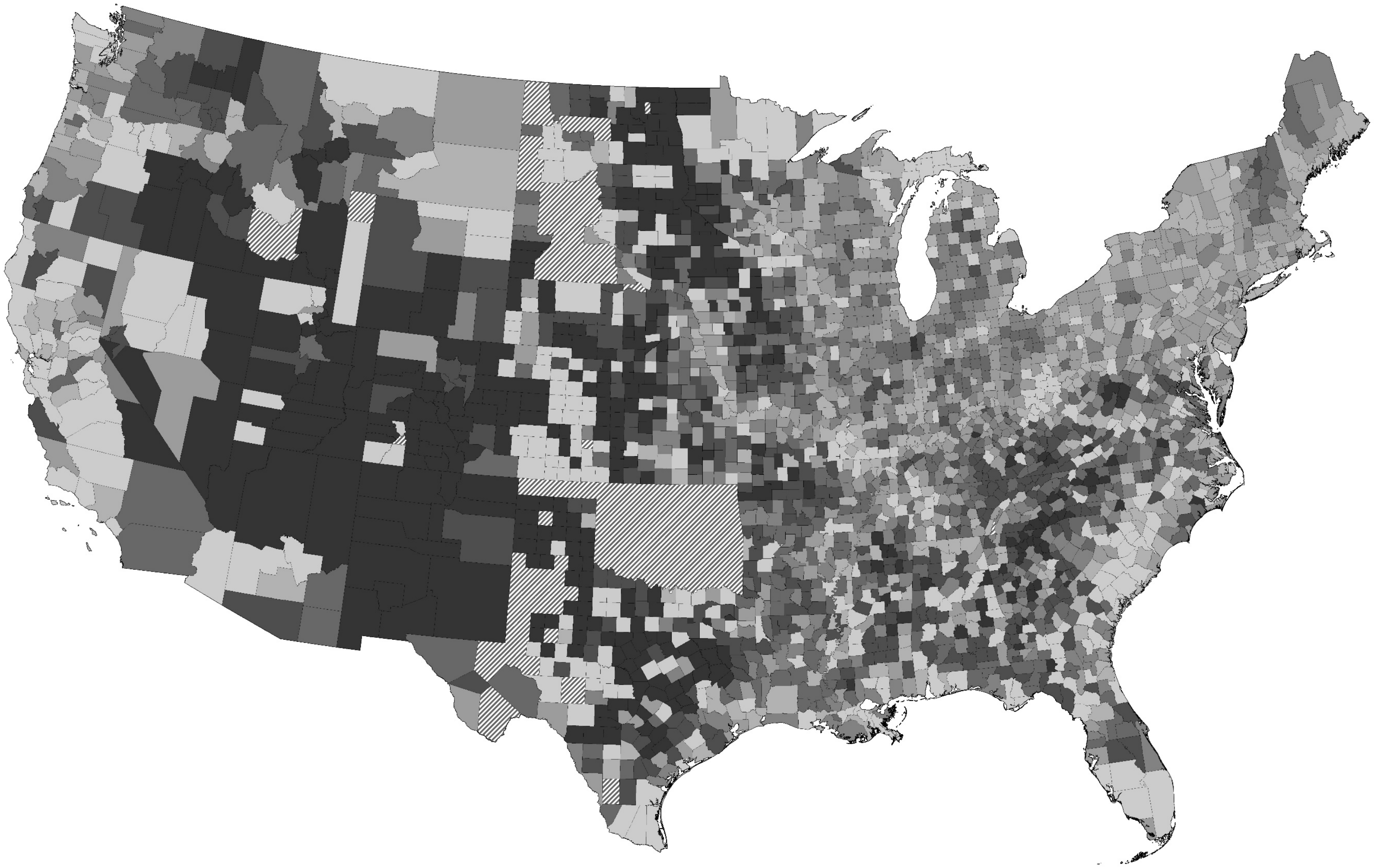
Notes: In each Panel, counties are shaded according to their calculated change in market access from 1860 to 1870 (Panel A) and from 1870 to 1880 (Panel B). Counties are divided into seven groups (with an equal number of counties per group), and darker shades denote larger increases in market access. These maps include the 1,802 sample counties in the regression analysis, which are all counties that report non-zero manufacturing activity from 1860, 1870, and 1880. The excluded geographic areas are cross-hashed. County boundaries correspond to county boundaries in 1890.

Figure 3. Counterfactual Changes in Market Access, by County



Notes: This map shows counties shaded according to their change in market access from 1890 to the baseline counterfactual scenario without railroads: darker shades denote larger declines in market access, and counties are divided into seven equal groups. This counterfactual sample includes all 2,722 counties that in 1890 report positive population and positive revenue (agriculture and/or manufacturing). The excluded geographic areas are cross-hashed. County boundaries correspond to county boundaries in 1890.

Figure 4. Counterfactual Changes in Productivity, by County



Notes: This map shows counties shaded according to their change in productivity from 1890 to the baseline counterfactual scenario without railroads: darker shades denote larger declines in productivity, and counties are divided into seven equal groups. This counterfactual sample includes all 2,722 counties that in 1890 report positive population and positive revenue (agriculture and/or manufacturing). The excluded geographic areas are cross-hashed. County boundaries correspond to county boundaries in 1890.

Table 1. Impacts of Market Access on County Revenue, Input Expenditure, and Productivity

	Baseline Specification	County-level Data Only: 1860 to 1900	County-level Data Only: 1850 to 1900
	(1)	(2)	(3)
Panel A. Log Revenue			
Log Market Access	0.192	0.257	0.235
	(0.049)	(0.061)	(0.056)
Panel B. Log Capital Expenditure			
Log Market Access	0.158	0.225	0.208
	(0.053)	(0.060)	(0.055)
Panel C. Log Labor Expenditure			
Log Market Access	0.196	0.292	
	(0.061)	(0.068)	
Panel D. Log Materials Expenditure			
Log Market Access	0.183	0.242	
	(0.050)	(0.062)	
Panel E. Log Productivity			
Log Market Access	0.204	0.279	
	(0.051)	(0.057)	
Number of Counties	1,802	1,802	1,437
County-Year Obs.	5,406	9,010	8,622

Notes: Column 1 reports estimates from equation 4: for the indicated outcome variable in each panel, we report the estimated impacts of log market access (as defined in equation 1), controlling for county fixed effects, state-by-year fixed effects, and year-interacted cubic polymials in county latitude and longitude. The sample is a balanced panel of 1,802 counties (1860, 1870, 1880) in columns 1 and 2, and a balanced panel of 1,437 counties in column 3.

The outcome variables are: the log of total county manufacturing annual revenue (Panel A); the log of total county manufacturing annual expenditures on capital, labor, and materials (Panels B, C, D); and the log of total county manufacturing revenue minus the weighted logs of total county manufacturing expenditures on capital, labor, and materials (where those weights are the county's average revenue share for that input, and the variable is scaled by the ratio of average county revenue to average county productivity, as defined in equation 2).

In each column, we report the estimated impact of a one standard deviation greater change in market access from 1860 to 1880 (e.g., the coefficient in column 1, panel E, can be interpreted as a relative productivity increase of 20.4% for counties with a one standard deviation greater change in market access from 1860 to 1880). This one standard deviation greater increase in market access corresponds to a 24% greater increase in market access from 1860 to 1880 for our baseline definition of market access.

Columns 2 and 3 use county-level data only, rather than county-by-industry data, which only affects the definition of Log Productivity in Panel E. Column 2 reports estimates for the 1860 to 1900 period, and Column 3 reports estimates for the 1850 to 1900 period using available data on county revenue and county capital expenditures in 1850.

Robust standard errors clustered by state are reported in parentheses.

Table 2. Impacts on County Productivity, Decomposed into TFPR and AE

	Baseline Specification	Detailed Industry Groups	County-level Data Only	
			1860 to 1880	1860 to 1900
	(1)	(2)	(3)	(4)
Panel A. Log County Productivity				
Log Market Access	0.204	0.204	0.204	0.279
	(0.051)	(0.051)	(0.051)	(0.057)
Panel B. County TFPR (Revenue Total Factor Productivity)				
Log Market Access	0.036	0.038	0.038	0.020
	(0.025)	(0.025)	(0.026)	(0.026)
Panel C. County AE (Allocative Efficiency)				
Log Market Access	0.168	0.166	0.166	0.258
	(0.051)	(0.052)	(0.054)	(0.067)
Number of Counties	1,802	1,802	1,802	1,802
County-Year Obs.	5,406	5,406	5,406	9,010

Notes: Column 1, panel A, corresponds to the estimate reported in Panel E of Column 1 in Table 1. Column 1 reports estimates from equation 4: for the indicated outcome variable in each panel, we report the estimated impacts of log market access (as defined in equation 1), controlling for county fixed effects, state-by-year fixed effects, and year-interacted cubic polymials in county latitude and longitude. The sample is a balanced panel of 1,802 counties (1860, 1870, 1880).

Panel A reports the estimated impacts on log county productivity (as in Panel E of Table 1), and Panels B and C report the impacts on productivity through changes in county TFPR (revenue total factor productivity) and through changes in county AE (allocative efficiency) as defined in equation 3.

In each column, we report the estimated impact of a one standard deviation greater change in market access from 1860 to 1880 (e.g., the coefficient in column 1, panel A, can be interpreted as a relative productivity increase of 20.4% for counties with a one standard deviation greater change in market access from 1860 to 1880). The coefficients in panel B and panel C imply 3.6% county productivity growth through increases in county TFPR and 16.8% county productivity growth through increases in county AE.

Column 2 calculates the outcome variables using county-by-industry data based on 193 industry categories, rather than the 31 industry categories used in column 1. Columns 3 and 4 calculate the outcome variables using county-level data, rather than county-by-industry data, for the same period from 1860 to 1880 (in column 3) and an extended period from 1860 through 1900 (in column 4).

Robust standard errors clustered by state are reported in parentheses.

Table 3. Sources of Growth in County Allocative Efficiency (AE)

	Allocative Efficiency by Input	County Input Gap	County Input Wedge	County Input Cost Share	County Std. Dev. of Wedges
	(1)	(2)	(3)	(4)	(5)
Panel A. Capital					
Log Market Access	-0.004	0.001	0.022	-0.0004	-0.015
	(0.009)	(0.002)	(0.036)	(0.0006)	(0.044)
Panel B. Labor					
Log Market Access	0.066	-0.001	-0.048	-0.0012	-0.022
	(0.015)	(0.005)	(0.068)	(0.0034)	(0.044)
Panel C. Materials					
Log Market Access	0.107	0.012	-0.028	0.0016	0.032
	(0.049)	(0.006)	(0.040)	(0.0038)	(0.054)
Number of Counties	1,802	1,802	1,802	1,802	1,802
County/Year Obs.	5,406	5,406	5,406	5,406	5,406

Notes: For the indicated outcome variable, each column and panel reports the estimated impact of log market access from our baseline specification (as in column 1 of Table 1). Column 1 reports impacts on county productivity through changes in county allocative efficiency (as in Table 2, panel C, column 1) through changes in capital (panel A), labor (panel B), and materials (panel C). Column 2 reports impacts on county-level input gaps (defined as the input's cost share minus its revenue share in that decade), Column 3 reports impacts on county-level input wedges (defined as the input's cost share divided by its revenue share, minus one, in that decade), and Column 4 reports impacts on county-level cost shares (defined as the national industry-level cost shares in each decade multiplied by the share of county revenue in each industry in that decade). Column 5 reports impacts on counties' standard deviation of input wedges across industries in that county and decade.

All regressions include county fixed effects, state-by-year fixed effects, and year-specific cubic polynomials in county latitude and longitude. We continue to report the estimated impact of a one standard deviation greater change in market access from 1860 to 1880. The sample is a balanced panel of 1,802 counties (1860, 1870, 1880). Robust standard errors clustered by state are reported in parentheses.

Table 4. Impacts of Market Access, County-by-Industry Level Regressions

			By Industry Group:			
	Pooled Specification	Weighted by 1860 Revenue Share	Clothing, Textiles, Leather	Food and Beverage	Lumber and Wood Products	Metals and Metal Products
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Log County-Industry Productivity						
Log Market Access	0.189	0.160	0.186	0.007	0.220	0.303
	(0.058)	(0.052)	(0.096)	(0.056)	(0.149)	(0.159)
Panel B. County-Industry TFPR (Revenue Total Factor Productivity)						
Log Market Access	0.056	0.044	0.059	-0.012	0.075	0.054
	(0.023)	(0.025)	(0.073)	(0.023)	(0.038)	(0.123)
Panel C. County-Industry AE (Allocative Efficiency)						
Log Market Access	0.133	0.116	0.127	0.019	0.145	0.249
	(0.058)	(0.043)	(0.089)	(0.063)	(0.139)	(0.119)
Number of Counties	1,800	1,800	994	1,338	1,480	709
County-Year Obs.	5,400	5,400	2,640	3,665	3,984	1,860

Notes: this table reports estimates from regressions at the county-by-industry level, after aggregating the more-detailed industries to five industry groups: clothing, textiles, leather; food and beverage; lumber and wood products; metals and metal products; and other industries. We extend our baseline estimating equation 4 to include county-industry fixed effects and state-year-industry fixed effects. The sample is drawn from our main balanced panel of 1,802 counties in 1860, 1870, and 1880, though each industry group is not reported in each county and decade. We omit county-industries that appear only once, but do not restrict the sample to county-industries that appear all three years. We continue to report the estimated impact of a one standard deviation greater change in market access from 1860 to 1880. Robust standard errors clustered by state are reported in parentheses.

Column 1 reports estimated average impacts of market access on county-industry productivity, county-industry TFPR, and county-industry AE. Column 2 reports estimates when weighting county-industries by their 1860 share of county revenue. Columns 3 to 6 report industry-specific effects of market access from separate regressions for each consistent aggregated industry group.

Table 5. Impacts of Market Access on County Industries, Establishments, and Sector Shares

	Log Number of Industries	Log Average Estab. Size: Revenue / Estab.	Log Average Estab. Size: Workers / Estab.	Log Number of Establishments	County Manufacturing Share of: Revenue	County Manufacturing Share of: Value-Added	County Manufacturing Share of: Surplus	County Manufacturing Share of: Employment
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Market Access	0.028	0.021	0.024	0.172	0.0076	0.0001	-0.0019	0.0045
	(0.021)	(0.030)	(0.043)	(0.034)	(0.0078)	(0.0059)	(0.0090)	(0.0048)
Number of Counties	1,802	1,802	1,802	1,802	1,774	1,774	1,713	1,687
County/Year Obs.	5,406	5,406	5,406	5,406	5,322	5,322	5,139	5,061

Notes: For the indicated outcome variable, each column reports the estimated impact of log market access from our baseline specification (as in column 1 of Table 1). In column 1, the outcome variable is the log number of manufacturing industries reporting positive output in the county. In columns 2 and 3, the outcome variables are log average manufacturing establishment size in the county, based on revenue per establishment (column 2) or workers per establishment (column 3). In column 4, the outcome variable is the log number of manufacturing establishments in the county. In columns 5 to 8, the outcome variables are the county's manufacturing share of total values for manufacturing and agriculture: revenue (column 5); value-added (column 6), which for manufacturing is defined as revenue minus materials expenditures and for agriculture is defined as 92% of revenue; surplus (column 7), which for manufacturing is defined as revenue minus all input expenditures and for agriculture is defined as the value of land multiplied by the state mortgage interest rate; and employment (column 8).

All regressions include county fixed effects, state-by-year fixed effects, and year-interacted cubic polymials in county latitude and longitude. The samples are drawn from our main balanced panel of 1,802 counties in 1860, 1870, and 1880, which for columns 5 to 8 is smaller due to missing data for some counties in some years. We continue to report the estimated impact of a one standard deviation greater change in market access from 1860 to 1880 in the full sample of 1,802 counties. Robust standard errors clustered by state are reported in parentheses.

Table 6. Impacts of Market Access, Controlling Flexibly for Local Railroad Construction

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Log County Productivity						
Log Market Access	0.204	0.218	0.198	0.197	0.175	0.142
	(0.051)	(0.056)	(0.057)	(0.058)	(0.058)	(0.056)
Panel B. County TFPR (Revenue Total Factor Productivity)						
Log Market Access	0.036	0.049	0.047	0.041	0.032	0.013
	(0.025)	(0.030)	(0.030)	(0.030)	(0.032)	(0.035)
Panel C. County AE (Allocative Efficiency)						
Log Market Access	0.168	0.169	0.152	0.156	0.143	0.129
	(0.051)	(0.055)	(0.056)	(0.055)	(0.055)	(0.054)
Additional Controls for:						
Any Railroad	No	Yes	Yes	Yes	Yes	Yes
Railroad Length	No	No	Yes	Yes	Yes	Yes
Railroad Length Polynomial	No	No	No	Yes	Yes	Yes
Railroads in Nearby Buffer	No	No	No	No	Yes	Yes
Railroads in Further Buffers	No	No	No	No	No	Yes
Number of Counties	1,802	1,802	1,802	1,802	1,802	1,802
County-Year Obs.	5,406	5,406	5,406	5,406	5,406	5,406

Notes: Column 1 reports the estimated impact of market access from the baseline specification (as in column 1 of Table 2). Column 2 includes an additional control for whether a county contains any railroad track. Column 3 also controls for the length of railroad track in the county, and column 4 controls for a cubic polynomial function of the railroad track mileage in a county. Column 5 includes additional controls for whether a county contains any railroad track within 10 miles of the county boundary, and a cubic polynomial function of the railroad track mileage within 10 miles of the county boundary. Column 6 adds controls for separate cubic polymial functions of railroad track within 20 miles, within 30 miles, and within 40 miles of the county.

All regressions include county fixed effects, state-by-year fixed effects, and year-specific cubic polynomials in county latitude and longitude. We continue to report the estimated impact of a one standard deviation greater change in market access from 1860 to 1880. The sample is a balanced panel of 1,802 counties (1860, 1870, 1880). Robust standard errors clustered by state are reported in parentheses.

Table 7. Impacts of Market Access, Instrumenting with Baseline Access through Waterways

	Log Market Access	County Productivity		County TFPR Revenue Total Factor Productivity		County AE Allocative Efficiency	
	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log Market Access		0.221	0.204	0.040	0.036	0.193	0.168
		(0.029,0.435)	(0.051)	(-0.111,0.315)	(0.025)	(0.001,0.367)	(0.051)
Instruments:							
Log Water Market Access in 1860 X year=1870	-1.003						
	(0.212)						
Log Water Market Access in 1860 X year=1880	-1.780						
	(0.245)						
Kleibergen-Paap F statistic		27.2		27.2		27.2	
Number of Counties	1,802	1,802	1,802	1,802	1,802	1,802	1,802
County-Year Obs.	5,406	5,406	5,406	5,406	5,406	5,406	5,406

Notes: Column 1 reports the impact of log water market access in 1860 on changes in log market access from 1860 to 1870 and changes in log market access from 1870 to 1880: log market access is regressed on log water market access in 1860, interacted with year fixed effects for 1870 and 1880. Column 2 reports the estimated impact of log market access on county productivity, instrumenting for log market access using the first-stage relationships reported in column 1. Columns 3, 5, and 7 reports the baseline estimates for comparison (from column 1 of Table 2). Columns 4 and 6 report corresponding 2SLS estimates for county TFPR and county AE. For the 2SLS specifications, we report the Andrews (2018) and Sun (2018) two-step weak-instrument-robust confidence sets, as well as the first stage F statistic.

All regressions include county fixed effects, state-by-year fixed effects, and year-interacted cubic polymials in county latitude and longitude. In columns 2 to 7, we continue to report the estimated impact of a one standard deviation greater change in market access from 1860 to 1880. The sample is a balanced panel of 1,802 counties (1860, 1870, 1880). Robust standard errors clustered by state are reported in parentheses.

Table 8. Impacts of Market Access on Model-Implied Values

	Baseline Specification	Model-Implied Values: Nominal	Real
	(1)	(2)	(3)
Panel A. Log Revenue			
Log Market Access	0.192	0.259	0.333
	(0.049)		
Panel B. Log Capital Expenditure			
Log Market Access	0.158	0.259	0.259
	(0.053)		
Panel C. Log Labor Expenditure			
Log Market Access	0.196	0.259	0.337
	(0.061)		
Panel D. Log Materials Expenditure			
Log Market Access	0.183	0.259	0.337
	(0.050)		
Panel E. Log Productivity			
Log Market Access	0.204	0.197	0.257
	(0.051)		
Number of Counties	1,802	1,802	1,802
County-Year Obs.	5,406	5,406	5,406

Notes: Column 1 reports estimates from column 1 of Table 1: for the indicated outcome variable in each panel, we report the estimated impacts of log market access (as defined in equation 1), controlling for county fixed effects, state-by-year fixed effects, and year-interacted cubic polymials in county latitude and longitude. The sample is a balanced panel of 1,802 counties (1860, 1870, 1880). We report the estimated impact of a one standard deviation greater change in market access from 1860 to 1880. Robust standard errors clustered by state are reported in parentheses.

Columns 2 and 3 show the relationship between log market access and model-predicted values in 1860, 1870, and 1880, where the only primitive allowed to vary over "time" is the railroad network and the model parameters are estimated in 1890. Column 2 reports impacts on nominal values for the outcome variables, and column 3 reports impacts on real values for the outcome variables.

Table 9. Counterfactual Impacts on National Aggregate Productivity

	Baseline: No Railroads	Restricted Railroad Networks: Only 1850 RRs	Only 1860 RRs	Only 1870 RRs	Only 1880 RRs	No Railroads, Extended Canals	All Railroads, Twice the Cost
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Counterfactual scenario, holding worker utility constant							
Change in Aggregate Productivity	-26.7%	-22.3%	-15.6%	-9.9%	-2.6%	-23.5%	-8.7%
Panel B. Counterfactual scenario, holding total population constant							
Change in Aggregate Productivity	-5.5%	-5.0%	-4.1%	-2.7%	-0.7%	-4.4%	-1.4%
Change in Utility	-32.7%	-27.2%	-18.3%	-11.4%	-2.9%	-29.0%	-11.1%

Notes: Each column reports the estimated change in national aggregate productivity from counterfactual changes in the transportation network. Panel A reports estimates from our baseline scenario, which holds worker utility constant in the counterfactual and allows for declines in total population. Panel B reports estimates from an alternative scenario, which holds total population fixed, and so we also report the associated decline in worker utility. In all scenarios, population is allowed to relocate endogenously within the country. The sample includes all 2,722 counties that in 1890 report positive population and positive revenue (agriculture and/or manufacturing).

Column 1 reports impacts under our baseline counterfactual scenario, which removes all railroads in 1890. Columns 2 to 5 report impacts under more moderate counterfactual scenarios, which restrict the railroad network to those railroads that had been constructed by 1850 (column 2), by 1860 (column 3), by 1870 (column 4), or by 1880 (column 5). Column 6 reports impacts from replacing the railroads with feasible extensions to the canal network, as proposed by Fogel (1964). Column 7 reports impacts from maintaining the 1890 railroad network, but doubling the cost of transportation along all railroads.

Table 10. Counterfactual Impacts on National Aggregate Productivity, Robustness

			Use Materials Wedge for:			Alternative Trade Elasticities:			Alternative Average Prices:	
	Baseline	Efficient Agriculture	Capital	Capital and Labor	Efficient Capital	$\Theta = 2.0$	$\Theta = 3.9$	$\Theta = 8.2$	$\bar{P} = 20$	$\bar{P} = 50$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A. Fixed Worker Utility										
Change in Aggregate Productivity	-26.7%	-16.5%	-23.4%	-24.3%	-20.2%	-27.6%	-26.1%	-24.5%	-36.1%	-23.2%
Panel B. Fixed Total Population										
Change in Aggregate Productivity	-5.5%	-4.6%	-4.1%	-4.6%	-6.1%	-5.7%	-5.4%	-4.1%	-7.6%	-4.7%
Change in Utility	-32.7%	-32.9%	-32.7%	-32.7%	-33.0%	-34.1%	-31.9%	-30.0%	-43.5%	-28.6%

Notes: Each column reports impacts under our baseline counterfactual scenario that removes all railroads in 1890, as in column 1 of Table 9. Panel A reports estimates from our baseline scenario, which holds worker utility constant in the counterfactual and allows for declines in total population. Panel B reports estimates from an alternative scenario, which holds total population fixed, and so we also report the associated decline in worker utility. In all scenarios, population is allows to relocate endogenously within the country. The sample includes all 2,722 counties that in 1890 report positive population and positive revenue (agriculture and/or manufacturing).

Columns 2 to 10 report robustness of our baseline estimates (in column 1) under alternative parameters. Our baseline estimates use our estimated value for Θ, the trade elasticity, of 3.05 with a 95% confidence interval between 1.95 and 3.925. In column 2, we reduce the estimated degree of input distortions in each county by assuming that the agricultural sector is efficient, and only apply our estimated manufacturing wedges to the county's manufacturing share of combined output across manufacturing and agriculture. Columns 3 and 4 use the estimated materials wedge in each county to assign counties' capital wedge (Column 3) or capital wedge and labor wedge (Column 4). Column 5 assumes that capital-use is efficient, such that there is zero gap for capital (or a wedge of 0). In columns 6 and 7, we alternatively impose values for Θ of 1.95 or 3.925; in column 8, we impose a value of 8.22 from Donaldson and Hornbeck (2016). Our baseline estimates also use our estimated value for $\bar{P}$ of 38.7, the average price of transported goods, which scales the assumed transportation costs into proportional costs. In Columns 9 and 10, we alternatively impose values for $\bar{P}$ of 20 or 50.