

# Railroads, Reallocation, and the Rise of American Manufacturing\*

Richard Hornbeck  
University of Chicago

Martin Rotemberg  
NYU

September 2019

## Abstract

We examine the impacts of market integration on the development of American manufacturing, as railroads expanded through the latter half of the 19th century. Using new county-by-industry data from the Census of Manufacturers, we estimate substantial reduced-form impacts on manufacturing productivity from relative increases in county market access as the railroad network expanded. In general equilibrium, the railroads increased economic activity in marginally productive counties. We therefore estimate substantial aggregate economic gains from the railroads, much larger than previous estimates, by allowing for the presence of factor misallocation. 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 many colleagues and seminar participants at: Brown, Census, Chicago Booth, Chicago Federal Reserve, Clemson, Columbia, Columbia-NYU, Duke, Florida State, Harvard, Hunter, Iowa State, Indiana, LSE, New Economic School, Northwestern, NYU, OECD, Oxford, PERC, Princeton, Queens, SED, Sciences Po, Toronto, UCLA, Wharton, Williams, and Zurich. Andrea Cerrato, William Cockriel, and Julius Luetge 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.

Over the latter half of the 19th century, manufacturing grew substantially in the United States. We estimate impacts on manufacturing from the expansion of the railroad network, which integrated large domestic markets with vast land and commodity resources. The railroads represented a technological improvement in the transportation sector, with modest direct benefits, but we estimate that the railroads generated substantial indirect benefits through encouraging expansion in the manufacturing sector and other sectors. The railroads thereby generated much larger economic gains than included in previous estimates (e.g., Fogel, 1964; Donaldson and Hornbeck, 2016), which highlights how broadly-used technologies or infrastructure can substantially impact economic growth.

We digitize county-by-industry data from the US Census of Manufacturers to calculate the contribution of each county to US manufacturing productivity growth. Following Petrin and Levinsohn (2012), we decompose county manufacturing productivity growth into two components: growth in “technical efficiency,” which reflects increases in output for a given level of inputs; and growth in “reallocative efficiency,” which reflects increased input usage in counties where the value marginal product of inputs exceeds marginal cost (e.g., due to firm markups, capital constraints, or other “frictions”). Growth in reallocative efficiency is the main focus of our analysis.

We examine how the expansion of the railroad network impacted county manufacturing productivity, along with other measures of county manufacturing activity. As in Donaldson and Hornbeck (2016), we measure how changes in the national railroad network affected county “market access,” which captures how a county’s manufacturing establishments were affected by the railroads changing establishments’ access to consumers, workers, and material inputs. While railroad construction is potentially endogenous, and otherwise correlated with local growth in manufacturing, 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 a spreading railroad network complemented or substituted for the previously established waterway network.

In reduced-form regressions, we find that county manufacturing productivity increases substantially with relative increases in county market access. A one standard deviation greater increase in county market access increased county manufacturing productivity by 12.9% from 1860 to 1880. This increase in productivity is driven by gains in reallocative efficiency, as railroads enabled the substantial expansion of manufacturing activity in marginally productive areas (i.e., in areas with value marginal product greater than marginal cost). This expansion in economic activity generated substantial economic gains due to the presence of market inefficiencies, such as firm markups or credit constraints, but we do not estimate

that the railroads reduced these market inefficiencies. We estimate little productivity impact through increases in technical efficiency, though a caveat is that observed changes in manufacturing revenues and expenditures may understate the impacts of market access on technical efficiency. We do not find systematic changes in local manufacturing industries or a shift from agriculture to manufacturing; rather, increased market access lead to a general expansion of economic activity in counties.

We then explore aggregate productivity impacts from counterfactual transportation networks. New transportation infrastructure may shift economic activity across areas, with large relative effects on some counties but small aggregate effects on the United States. Our model builds on Donaldson and Hornbeck (2016), adding in market distortions that drive a wedge between firms' value marginal product and marginal cost, and allowing for spatially heterogeneous production functions and intermediate goods. We use our estimates from the manufacturing sector to discipline key parameters in the model, such as the county-specific gaps between the value marginal product of inputs and their marginal cost.

We estimate that US aggregate productivity would have been 25% lower in 1890, in the absence of the railroads, with an associated annual loss of \$3 billion or 25% of GDP. This annual economic loss, as a share of GDP, is much larger than previous estimates of 3.2% (Donaldson and Hornbeck, 2016) or 2.7% (Fogel, 1964). We estimate much larger economic losses because we allow for inputs to be productive on the margin, as measured in the manufacturing sector, whereas Donaldson and Hornbeck (2016) maintain Fogel's assumption that resources are allocated efficiently such that the value marginal product of inputs is equal to their marginal cost. When including our estimated impacts on annual productivity, we estimate a 45% annual social rate of return on the \$8 billion of capital invested in the railroads in 1890, and estimate that the railroads in 1890 privately captured 10% of this social return.

In the absence of the railroads, we estimate there would have been a less productive spatial allocation of resources in the United States. Holding the aggregate population of the United States fixed in 1890, we estimate that aggregate productivity would have been 5.3% lower. In this scenario, worker utility also declines by 34% due to decreasing labor productivity. If we hold worker utility fixed, then population declines by 68%. This is similar to the 58% population decline estimated by Donaldson and Hornbeck (2016), but in our model this generates much larger economic losses because the decline in workers leads to further declines in reallocative efficiency.

In a more moderate counterfactual scenario, we estimate that US aggregate productivity would have been 12% lower in 1890 if the railroad network had not expanded after 1860. Additional canals might have been constructed in the absence of the railroads, but we esti-

mate that replacing the railroad network with an extended canal network would have lowered aggregate productivity by 22%. The railroads had a central role in enabling the substantial growth of the United States economy, and would not have been easily replaced.

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; Ziebarth, 2013; Midrigan and Xu, 2014).<sup>1</sup> There are a variety of market distortions that drive a wedge between firms' value marginal product and their marginal cost, such firm markups, capital constraints, taxes and regulation, or imperfect enforcement of property rights.<sup>2</sup> If the railroads shift economic resources from one area to another, then aggregate productivity will increase or decrease depending on the relative degree of resource misallocation in those areas. Further, aggregate increases in resource use will increase aggregate productivity when the average marginal product is higher than the average marginal cost. Factor misallocation can thereby heighten the aggregate economic gains from new infrastructure or technologies that enable an expansion of economic activity. To quantify productivity impacts of the railroads, and market integration, we draw on methods from the reallocation literature (Hulten, 1978; Petrin and Levinsohn, 2012; Baqaee and Farhi, 2019*b*).

Our paper extends a literature on estimating the impacts of market access (Redding and Venables, 2004; Hanson, 2005; Redding and Sturm, 2007; Head and Mayer, 2011; Jaworski and Kitchens, 2016; Yang, 2018).<sup>3</sup> By allowing for resource misallocation, we demonstrate a quantitatively important additional channel through which increases in market access can generate economic gains. In doing so, our work relates to a literature that considers how the efficiency of resource allocation across firms is affected by policies such as trade liberalization, financial regulations, and taxes.<sup>4</sup> By bringing this research on resource misallocation into a model of economic geography, we can explore the spatial allocation of economic activity and especially how production expanded to use new resources and attract additional workers.<sup>5</sup>

---

<sup>1</sup>See Liu (2019) for a review of this literature, along with discussion of misallocation in environments with more complicated input-output linkages.

<sup>2</sup>We do not estimate the underlying sources of these wedges, and we do not find that increases in county market access have a systematic effect on county-level gaps between firms' value marginal product and their marginal cost.

<sup>3</sup>Redding and Turner (2015) and Redding and Rossi-Hansberg (2017) present reviews of the literature.

<sup>4</sup>See, for example, papers on trade liberalization (Khandelwal, Schott and Wei, 2013; Świecki, 2017; Bai, Jin and Lu, 2018; Berthou et al., 2018; Chung, 2018; Costa-Scottini, 2018; Singer, 2018; Tombe and Zhu, 2019), financial regulations (Blattner, Farinha and Rebelo, 2019; Rotemberg, Forthcoming), and taxes (Giroud and Rauh, 2019), and see Sraer and Thesmar (2019) for a review of this literature. Asturias, García-Santana and Ramos (2018) calibrate a model to consider how highway construction in India affects prices and the allocation of fixed aggregate inputs between firms with different markups.

<sup>5</sup>In contrast to previous work on resource misallocation, which tends to hold fixed aggregate inputs, an important feature of our analysis is how the railroads encouraged growth in aggregate inputs in the United States economy. By drawing on the economic geography literature, we can consider changes in aggregate inputs along with the reallocation of fixed inputs, which will be central to understanding the

Our paper also connects to a long tradition in economics, back to at least Harberger (1964), of simplifying welfare analysis by assuming there is no resource misallocation. Fogel (1964) implicitly adopts this assumption when calculating the economic gains from the railroads, and much of the critique by David (1969) can be seen as calling attention to Fogel’s assumption that value marginal product is equal to marginal cost throughout the economy.<sup>6</sup> Our estimated impacts of the railroads are a reminder that resource misallocation can magnify the economic gains from widely used infrastructure or technologies that encourage the expansion of otherwise productive activities across other sectors of the economy.

Understanding the economic impacts of the railroads speaks to the potential for market integration to drive economic growth and, more generally, for any isolated technological improvement to generate large wide-ranging gains throughout the economy. If the impacts of railroads were bounded above by the savings in water transportation costs, then Fogel’s larger conclusion is that any single technological advance is unlikely to generate transformative economic gains. Measured impacts on land values in the tradition of hedonic analyses, as in Donaldson and Hornbeck (2016), can similarly understate economic impacts dramatically. Resource misallocation greatly magnifies the impacts of technologies, such as transportation infrastructure, that enable and encourage other economic activities. This suggests an important role for transportation infrastructure in supporting economic growth and, more generally, wide-ranging impacts from the enabling technologies that underlie the economy. Indeed, we estimate that the railroads had a central role in encouraging the historical development of the American economy.

## I Data Construction

### I.A Data on Manufacturing

We use data from the US Census of Manufacturers (CMF), which we have digitized for each county and industry in 1860, 1870, and 1880 (see Data Appendix). Our analysis focuses on this 1860 to 1880 period, though we also extend our analysis to 1890 and 1900 using data aggregated by county (Haines, 2010).<sup>7</sup> Census enumerators were directed to collect

---

historical experience of the railroads and why the railroads had such large impacts on the development of the American economy.

<sup>6</sup>David (1969) focuses on increasing returns to scale, which would violate this assumption. In response, Fogel (1979) makes this assumption more explicit but disputes the empirical magnitude of increasing returns to scale. Crafts (2004) derives the Fogel upper-bound in partial equilibrium. Allen and Arkolakis (2019) derive Fogel (1964)’s social savings calculation in general equilibrium, and show how it can break down with departures from benchmark models (in their case, in the presence of agglomeration economies).

<sup>7</sup>For 1890 and 1900, only county-level tabulations of these data were published. For 1850, we have not found tabulations of all the variables needed, although the original surveys are available. Prior to 1850, there were greater concerns about the comprehensiveness of the data collection and the Census data collection was professionalized in 1850 (Atack and Bateman, 1999).

these data from each manufacturing establishment with more than \$500 in sales, which the census instructions refer to as capturing even basic manufacturing operations that might be run out of sheds or other part-time establishments. The Census then published aggregated statistics by county and industry, including county-by-industry cells that contain only one manufacturing establishment.<sup>8</sup>

The manufacturing data include the annual value of output, as well as the annual cost of labor and raw materials. The data also include the value of capital invested, which we multiply by a state-specific mortgage interest rate to estimate annual capital expenditures.<sup>9</sup>

Reported revenues and expenditures are intended to reflect “factory-gate” prices, based on Census instructions to enumerators. Transportation costs are included in establishment expenditures on materials, whereas revenue received by the manufacturing establishment does not include costs of shipping goods to customers. These factory-gate prices reflect what is directly paid or received by the manufacturing establishment, and are thereby the natural choice for tracking manufacturing productivity.

We have grouped the reported industries into 45 consistent industry groups (following Leontief (1936)), though we also report estimates using 159 industry categories or using aggregate county-level data.<sup>10</sup> Appendix Table 1 reports information on 10 large industry groups, aggregating further, along with their share of national manufacturing output in 1860, 1870, and 1880 (column 1). Columns 2 – 4 report industry cost shares in each decade, which is mostly stable over time and varies more across industries.

In supplemental analysis, we use data from the Census of Manufactures on the number of manufacturing establishments and the number of manufacturing workers. We also use data from the Census of Agriculture to consider relative changes in the manufacturing and agricultural sectors. The Census of Agriculture also includes data on the value of home manufactures, which we use to consider a potential shift from home manufacturing to more formal manufacturing production.

---

<sup>8</sup>Starting in 1870, the county-by-industry data do not list some “neighborhood industries” focused on local production, or industries with less than \$10,000 of output in total. We define a residual industry to capture the difference between county-level data and the summed county-by-industry data, and include this residual industry in our analysis. This residual “industry” includes around 10% of manufacturing output in 1870 and 1880. We also created an “other” industry, representing less than 1% of output.

<sup>9</sup>The state-specific mortgage interest rate varies between 5.5% and 11.4% , with an average value of 8% (Fogel, 1964).

<sup>10</sup>There are over one thousand different industries listed in the county-industry data, including “boots and shoes” separate from “boots and shoes, including custom work and repairing,” and the 159 industry groupings represent the largest number that seemed appropriate.

## I.B Data on Market Access

We measure how county-to-county freight transportation costs declined from an expanding railroad network. Figure 1, panel A, shows the network of waterway routes that includes navigable rivers, canals, lakes, and oceans. Panel B shows the railroad network constructed by 1860, which then expanded by 1870 (panel C) and 1880 (panel D).<sup>11</sup> 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.<sup>12</sup> 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 focus on measuring county-to-county transportation costs because we consider how the railroad network changes counties' access to markets, rather than considering impacts from the local presence of railroads themselves. This distinction is also particularly useful for empirical identification, as we can examine the impacts of changes in market access after controlling flexibly for changes in local railroad density.

We approximate the “market access” of county  $c$ , following Donaldson and Hornbeck (2016), as a sum over that county's cost of transporting goods ( $\tau$ ) to or from each other county  $d$  with population  $L$ :

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

County  $c$  has greater market access when it is cheaper to trade with other counties  $d$  that have greater population.<sup>13</sup>

This approximation for county market access is derived from a general equilibrium trade model (Eaton and Kortum, 2002; Donaldson and Hornbeck, 2016), which we extend in Section V to include market frictions. We derive this same approximation for county market

---

<sup>11</sup>Appendix Figure 1 shows the railroad network in 1890 and 1900.

<sup>12</sup>Following Donaldson and Hornbeck (2016), railroad rates are set at 0.63 cents per ton mile and waterway rates are set 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 concern the required length of wagon transportation and the number of transshipment points. Donaldson and Hornbeck (2016) provide further details on these cost calculations. In Section IV, we report the robustness of our estimates to alternative assumptions in calculating county-to-county transportation costs.

<sup>13</sup>We calculate a county's access to all other counties with reported population, including other counties that are excluded from our main regression sample because they do not report manufacturing data in each decade.

access, under more general conditions than in Donaldson and Hornbeck (2016), and we defer those details until Section V.

For measuring county market access, as defined in equation 1, we need to estimate  $\theta$  and  $\tau_{cd}$ . The parameter  $\theta$  reflects the “trade elasticity,” which varies across empirical contexts. The parameters  $\tau_{cd}$  represent “iceberg trade costs,” which normalize the measured county-to-county transportation costs  $t_{cd}$  by the average price of transported goods ( $\tau_{cd} = 1 + t_{cd}/\bar{P}$ ). In Section V, we use the model to estimate values for  $\theta$  (2.75) and  $\bar{P}$  (35.3).<sup>14</sup> We also report the robustness of our results to alternative parameter values.

Figures 2 and 3 show in darker shades those counties that have relatively greater increases in market access from 1860 to 1870 (Figure 2, panel A), from 1870 to 1880 (Figure 2, panel B), from 1880 to 1890 (Figure 3, panel A), and from 1890 to 1900 (Figure 3, panel B). Our baseline empirical specification will compare changes in darker shaded counties to changes in lighter shaded counties within the same state, and within similar latitude and longitude.<sup>15</sup> Comparing counties within nearby areas, there is substantial variation in changes in county market access. Further, looking across the four decades, it is often different counties that are experiencing relatively larger or smaller changes in market access.

Figures 2 and 3 map our main regression sample of 1,804 counties, which includes all counties that report manufacturing activity in 1860, 1870, and 1880.<sup>16</sup> We focus on manufacturing productivity growth along the intensive margin, in this balanced sample of counties, but we also consider the extensive margin growth of manufacturing activity into new counties.

## II Defining and Measuring Productivity Growth

We begin by considering the growth of total production surplus in the United States economy, summing across each sector  $s$  the growth in total value of output minus the growth in total value of inputs:

$$(2) \quad APG = \sum_s (dV_s^Q - dV_s^X).$$

---

<sup>14</sup>The estimated value of 35.3 for  $\bar{P}$  is very similar to the assumed value of 35 in Donaldson and Hornbeck (2016), following Fogel (1964). The estimated value of 2.75 for  $\theta$  is smaller than the estimated value of 8.22 in Donaldson and Hornbeck (2016), due to differences in the model and sample period, but the estimated impacts of the railroads are not sensitive to the value of  $\theta$ .

<sup>15</sup>We assign county “latitude” and “longitude” based on the x/y-coordinates of county centroids in these Figures, which are based on an Albers equal-area projection of the United States, such that displayed areas are comparable in geographic size and differences in the x/y-coordinates reflect consistent distances East-West and North-South.

<sup>16</sup>We have adjusted the data in each decade to maintain consistent geographic units (as in Hornbeck (2010)), which reflect county boundaries in 1890.

We follow previous literature in referring to this as “productivity growth” (Solow, 1957; Basu and Fernald, 2002; Petrin and Levinsohn, 2012; Baqaee and Farhi, 2017). This productivity growth includes “technological growth” in firm production efficiency, but also allows for the influence of firm markups or market frictions and thereby resource misallocation (e.g., Hall, 1988, 1990; Hsieh and Klenow, 2009). For example, when a firm’s output price is greater than its marginal cost, increased input usage by that firm will increase the total value of output by more than the total value of inputs and thereby result in productivity growth as defined above. Similarly, shifting inputs from low markup firms to high markup firms will increase productivity.

Equation 2 can be split into productivity growth in the transportation sector, in the manufacturing sector, and in other sectors:

$$(3) \quad APG = (dV_t^Q - dV_t^X) + (dV_m^Q - dV_m^X) + (dV_o^Q - dV_o^X).$$

The first term would include any direct benefit from the railroads through decreasing costs of transporting goods. The “social savings” calculation introduced by Fogel (1964) can be seen as an attempt to bound this direct benefit, in the transportation sector, as Fogel seeks to measure the increase in costs from providing the same transportation services using waterways instead of railroads. For that calculation to bound the economic gains from the railroads, however, it assumes that railroads generate no productivity growth in other sectors of the economy (David, 1969; Fogel, 1979; Allen and Arkolakis, 2019). Section VI discusses this further, in interpreting our results, connecting this old debate to the more recent literature on resource misallocation and its implications for welfare analysis.

## II.A Manufacturing Productivity Growth

We estimate impacts of the railroads on productivity growth in the manufacturing sector in Section IV, and in Section V we also consider impacts on other sectors. Productivity growth in the US manufacturing sector is a sum over productivity growth in each county  $c$  and industry  $i$ :

$$(4) \quad APG = \sum_c \sum_i P_{ci} dQ_{ci} - \sum_c \sum_i \sum_k W_{ci}^k dX_{ci}^k.$$

The growth in total output of county  $c$  and industry  $i$  ( $dQ_{ci}$ ) includes growth in final goods output ( $dY_{ci}$ ) and intermediate goods output ( $dM_{ci}$ ), and is valued at a constant price ( $P_{ci}$ ). From this growth in total output, we subtract the growth in  $k$  total inputs ( $dX_{ci}^k$ ) that includes intermediate goods inputs, labor inputs, and capital inputs each valued at a constant price ( $W_{cik}$ ). We later consider changes in input prices and output prices.

Expressed in growth rates, productivity growth in manufacturing is a weighted sum of productivity growth in each county and industry:

$$(5) \quad APG = \sum_c \sum_i D_{ci} \left[ d \ln Q_{ci} - \sum_k s_{ci}^k d \ln X_{ci}^k \right],$$

where  $s_{cik}$  is the revenue share of input  $k$  and  $D_{ci}$  are Domar weights (Domar, 1961; Hulten, 1978).<sup>17</sup> From equation 5, the productivity growth rate in each county and industry is given by the percent growth in output ( $d \ln Q_{ci}$ ) minus the percent growth in each input ( $d \ln X_{ci}^k$ ) multiplied by that input’s revenue share ( $s_{ci}^k$ ).

## II.B Decomposing Productivity Growth:

### Technical Efficiency Growth and Reallocative Efficiency Growth

The above productivity growth rate can be decomposed into changes in “technical efficiency” and changes in “reallocative efficiency” (Petrin and Levinsohn, 2012; Baqaee and Farhi, 2019a). Defining  $\epsilon_{cik}$  as the elasticity of output with respect to input  $k$ , adding and subtracting the expected change in output from an increase in inputs ( $\sum_k \epsilon_{ci}^k d \ln X_{ci}^k$ ), and rearranging equation 5 yields:

$$(6) \quad \begin{aligned} APG = & \sum_c \sum_i D_{ci} \left[ d \ln Q_{ci} - \sum_k \epsilon_{ci}^k d \ln X_{ci}^k \right] & \text{(TE)} \\ & + \sum_c \sum_i D_{ci} \left[ \sum_k (\epsilon_{ci}^k - s_{ci}^k) d \ln X_{ci}^k \right]. & \text{(RE)} \end{aligned}$$

The first term in equation 6 represents growth in technical efficiency (TE), or total factor productivity (TFP), which reflects increases in output that exceed the expected increase in output from the increases in input usage. This first term has been proposed as a summary measure of productivity growth (e.g., Solow, 1957), though under assumptions such that the second term in equation 6 is equal to zero.

The second term in equation 6 represents growth in reallocative efficiency (RE), which reflects gains from increased input usage ( $d \ln X_{ci}^k$ ) when the value marginal product of inputs exceeds their marginal cost ( $\epsilon_{ci}^k - s_{ci}^k > 0$ ). The value marginal product of inputs can exceed the marginal cost of inputs due to firm markups, capital constraints, taxes and regulations, or other market distortions.<sup>18</sup> Productivity increases when existing inputs are shifted to counties and industries with greater gaps between  $\epsilon$  and  $s$ . Further, productivity increases

<sup>17</sup>The Domar weight is defined as total output in that county and industry ( $P_{ci}Q_{ci}$ ), divided by the national value of final goods output ( $\sum_c \sum_i P_{ci}Y_{ci}$ ).

<sup>18</sup>In Section V, we model these gaps as arising due to input frictions ( $\psi$ ), which create a “wedge” between value marginal product and marginal cost, and can be measured using an input’s cost share divided by its revenue share. Firm markups can be represented as a common component in these frictions across inputs.

when new inputs are brought into usage in counties and industries with any positive gap between  $\epsilon$  and  $s$ .

For the decomposition of productivity growth into technical efficiency growth and reallocative efficiency growth, we assume Cobb-Douglas production with constant returns to scale and estimate output elasticities using industry cost shares. In Section IV, we report our estimates' robustness to alternative scale elasticities, estimation biases in the cost shares, and measurement error. We then aggregate equation 6 to the county level by summing over industries:<sup>19</sup>

$$(7) \quad \begin{aligned} APG &= \sum_c D_c [d \ln Q_c - \sum_k \alpha_c^k d \ln X_c^k] & (TE) \\ &+ \sum_c D_c [\sum_k (\alpha_c^k - s_c^k) d \ln X_c^k]. & (RE) \end{aligned}$$

We will examine how the railroad network (and changes in county market access) affected these measures of productivity, technical efficiency, and reallocative efficiency. The railroads may increase reallocative efficiency by increasing manufacturing inputs in counties with positive gaps. The railroads may increase technical efficiency if production practices improve along with increases in the extent of the market, for example.

## II.C Measured Productivity Growth

Our analysis focuses on quantifying changes in reallocative efficiency, rather than changes in technical efficiency, due to an important data limitation. The Census of Manufacturers reports revenues and expenditures that reflect local prices, whereas equation 7 is expressed in real values with fixed output prices and fixed input prices. Increases in county market access can affect input prices, however, which the model predicts leads to proportional changes in county output prices (Section V).<sup>20</sup>

We now express a county's measured change in reallocative efficiency ( $dRE_c$ ), allowing for changes in local input prices ( $d \ln W_c^k$ ):

$$(8) \quad d \ln RE_c = \sum_k (\alpha_c^k - s_c^k) d \ln W_c^k X_c^k.$$

---

<sup>19</sup>The Domar weight for each county ( $D_c$ ) is given by the sum of the Domar weights for each county and industry:  $D_c = \sum_i D_{ci}$ . To allow for industry entry and exit within counties, we define the change in log county output ( $d \ln Q_c$ ) as the change in the weighted sum of county-industry log output ( $d \sum_i \frac{P_{ci} Q_{ci}}{\sum_j P_{cj} Q_{cj}} \ln Q_{ci}$ ). Similarly, for inputs,  $d \ln X_c^k = d \sum_i \frac{P_{ci} Q_{ci}}{\sum_j P_{cj} Q_{cj}} \ln X_{ci}^k$ .

<sup>20</sup>Section V derives county market access and its predicted effects, extending the framework in Donaldson and Hornbeck (2016), which assumes a CES demand structure for county output and Cobb-Douglas production with constant returns to scale. Given these assumptions, changes in county input prices lead to a proportional change in county output prices.

We predict that increased county market access will increase the quantity of purchased inputs ( $d \ln X_c^k > 0$ ), but also decrease local input prices ( $d \ln W_c^k < 0$ ), such that the increase in input quantities will be understated by the measured increase in input expenditures ( $d \ln W_c^k X_c^k < d \ln X_c^k$ ). The measured growth in reallocative efficiency (from equation 8) will then be a downward-biased measure of the true growth in reallocative efficiency (from equation 7).<sup>21</sup>

We can also express a county's change in technical efficiency ( $d \ln TE_c$ ), allowing for changes in local input prices ( $d \ln W_c^k$ ) and local output prices ( $d \ln P_c$ ):

$$(9) \quad d \ln TE_c = d \ln P_c Q_c - \sum_k \alpha_c^k d \ln W_c^k X_c^k.$$

Decreases in local output prices will downward-bias the measured impacts of county market access on technical efficiency. Indeed, in a model with CES demand for county output and production with constant returns to scale, as in Section V, local output prices are predicted to decline proportionally with increases in technical efficiency (from equation 7) such that there is no change in measured technical efficiency (from equation 9).<sup>22</sup> In practice, however, changes in our measured technical efficiency (based on revenue data) can be correlated with changes in technical efficiency (based on quantity data) in settings when both are observed (Foster, Haltiwanger and Syverson, 2008; Haltiwanger, Kulick and Syverson, 2018).

In summary, we use data on manufacturing revenues and expenditures at the county-by-industry level to measure changes in county productivity ( $PR_c$ ), county technical efficiency ( $TE_c$ ), and county reallocative efficiency ( $RE_c$ ):

$$(10) \quad d \ln PR_c = d \ln P_c Q_c - \sum_k s_c^k d \ln W_c^k X_c^k,$$

$$(11) \quad d \ln TE_c = d \ln P_c Q_c - \sum_k \alpha_c^k d \ln W_c^k X_c^k,$$

$$(12) \quad d \ln RE_c = \sum_k (\alpha_c^k - s_c^k) d \ln W_c^k X_c^k.$$

Changes in county productivity correspond to the change in log revenue ( $d \ln P_c Q_c$ ) minus the revenue-share weighted changes in log input expenditures ( $\sum_k s_c^k d \ln W_c^k X_c^k$ ). Changes in county technical efficiency correspond to the change in log revenue ( $d \ln P_c Q_c$ ) minus the cost-share weighted changes in log input expenditures ( $\sum_k \alpha_c^k d \ln W_c^k X_c^k$ ). Changes in

<sup>21</sup>This discussion of downward-bias assumes that market access does not affect the gap between production function elasticities and revenue shares ( $\alpha_c^k - s_c^k$ ), which is consistent with our estimates in Section IV.

<sup>22</sup>There is no additional measurement bias in technical efficiency from decreases in local input prices, under the assumptions of CES demand and CRS, as lower input prices pass through proportionately to lower output prices and measured technical efficiency is unchanged.

county reallocative efficiency are given by the changes in log input expenditures ( $d \ln W_c^k X_c^k$ ), multiplied by the county-specific gap between the cost share and revenue share for each input ( $\alpha_c^k - s_c^k$ ).

### III Estimating Equation

We regress outcome  $Y$  in county  $c$  and year  $t$  on log market access ( $MA_{ct}$ ), county fixed effects ( $\gamma_c$ ), state-by-year fixed effects ( $\gamma_{st}$ ), and a cubic polynomial in county latitude and longitude interacted with year effects ( $\gamma_t f(x_c, y_c)$ ):<sup>23</sup>

$$(13) \quad Y_{ct} = \beta \ln(MA_{ct}) + \gamma_c + \gamma_{st} + \gamma_t f(x_c, y_c) + \varepsilon_{ct}.$$

The coefficient  $\beta$  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 after adjusting for changes associated flexibly with county latitude and longitude. The identification assumption is that counties with relative increases in market access would otherwise have changed similarly to nearby counties. This baseline empirical specification is the same as in Donaldson and Hornbeck (2016), and we also report robustness to alternative specifications.

The main identification concern is that railroad construction may have been targeted toward counties that would otherwise have grown more over this period. We estimate impacts of county market access, which depends in large part on more-distant changes in the railroad network and its interaction with the waterway network. We report estimates that control flexibly for changes in railroads within a county and within nearby areas. We also use variation in counties' access to more-distant counties.

The main regression sample is a balanced panel of 1,804 counties in 1860, 1870, and 1880 (shown in Figures 2 and 3). We report standard errors that are clustered by state to adjust for correlation in  $\varepsilon_{ct}$  over time and within states.<sup>24</sup>

---

<sup>23</sup>Note that we estimate equation 13 in levels, rather than in changes, but include county fixed effects that focus identification on county-level changes in market access. The productivity outcomes are thereby defined in levels, with revenue shares and cost shares assigned using data in that year, which allows the cost shares and revenue shares to change over time.

<sup>24</sup>If we instead adjust for spatial correlation across counties, following (Conley, 1999), we estimate standard errors for our baseline specification that are 3% lower for distance cutoffs of 200 miles or 300 miles, 12-21% lower for distance cutoffs between 400 miles and 700 miles, and 11-13% lower for distance cutoffs between 800 miles and 1000 miles. This method assumes that spatial correlation declines linearly up to the assumed distance cutoff and is zero thereafter.

## IV Estimated Impacts of County Market Access

### IV.A Estimated Impacts on Productivity

Table 1 presents results from estimating equation 13. Panel A reports impacts of market access on county productivity, and panels B and C decompose this effect into impacts on county reallocative efficiency (RE) and county technical efficiency (TE).

We estimate that county market access has a substantial and statistically significant impact on county manufacturing productivity. A one standard deviation greater increase in market access from 1860 to 1880 leads to a 12.9% increase in productivity (Table 1, panel A, column 1).<sup>25</sup>

This impact on county productivity is driven by increases in county reallocative efficiency (Table 1, panel B, column 1). Increases in county market access are increasing manufacturing input expenditures in counties where the value marginal product of those inputs exceeds their marginal cost. By contrast, increases in county market access have little estimated impact on county technical efficiency (Table 1, panel C, column 1).

Columns 2 and 3 report similar estimates when restricting the identifying variation in county market access. For column 2, we calculate market access in each period holding county populations fixed at 1860 levels, such that changes in county market access are only due to changes in county-to-county transportation costs. For column 3, we calculate counties' market access omitting other counties within 100 miles, such that changes in county market access only reflect more-distant economic influences.

Column 4 reports similar estimates when extending our sample through 1900, using the available manufacturing data aggregated at the county-level.<sup>26</sup> Our main estimates use county-by-industry data from 1860 to 1880, in 45 industry groups (Leontief, 1936), but columns 5 and 6 report similar estimates when using only county-level data from 1860 to 1880 or when defining 159 distinct industries.

### IV.B Endogeneity of Railroad Construction

A primary empirical concern in estimating the impacts of transportation infrastructure is that infrastructure investment is generally directed towards areas that might otherwise change differently over time. In particular, railroad construction may occur in counties that would otherwise have experienced relative increases in manufacturing activity.

---

<sup>25</sup>For ease of interpretation, we report the estimated impact of log market access in terms of standard deviations (calculating changes in log market access from 1860 to 1880 and taking the standard deviation of those changes across counties). A one standard deviation greater increase in market access corresponds to a 23% greater increase in market access from 1860 to 1880.

<sup>26</sup>For comparability to columns 1 – 3, we continue to report the estimated impact of a 23% increase in market access (i.e., a one standard deviation greater increase in market access from 1860 to 1880).

One of the primary 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 network database. Thus, we can estimate the impacts of county market access when controlling for railroad construction in the county and nearby areas. Local railroad construction might also directly impact local manufacturing activity, through increases in the demand for manufactured construction materials, and controlling for local railroad construction would absorb a potential direct effect on manufacturing.

Table 2 reports similar estimated impacts of county market access when controlling flexibly for local railroad construction. Column 1 reports baseline estimates from Table 1, as a basis for comparison. Column 2 controls for whether the county has any railroad. Column 3 also controls for the county’s length of railroad track, and column 4 controls for a cubic polynomial function of the county’s railroad track. Column 5 adds controls for a cubic polynomial function of railroad track within 10 miles of the county’s borders, and column 6 adds controls for separate cubic polynomial functions of railroad track within 20 miles, within 30 miles, and within 40 miles of the county’s borders. Local railroad construction predicts increases in county market access, but Table 2 reports that 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.

As an alternative empirical approach, we can further exploit 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 cheap access to markets through waterways should generally experience less increase in market access. We define counties’ “water market access” in 1860, which reflects their measured market access when excluding all railroads from the transportation network.

Table 3, 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 (and column 1 reports the first-stage results). Columns 2, 4, and 6 report the 2SLS estimated impacts of market access on productivity, reallocative efficiency, and technical efficiency, which are larger but less precise than the OLS impacts (columns 3, 5, and 7). We also estimate that counties with greater water market access in 1860 were not more or less likely to have any

railroad by 1870 or by 1880, as the railroad network expanded broadly throughout the country to places with or without waterway access. The IV identification assumption could still be violated if these counties with greater water market access would have changed differently from 1860 through 1880, even in the absence of differential changes in county market access, and so we focus on other robustness checks that directly control for local railroad construction or other regional shocks.

#### IV.C Robustness: Regional Shocks

Our baseline empirical specification estimates the impacts of market access, controlling for county fixed effects and state-by-year fixed effects, such that the identifying variation is within-state relative changes in counties' market access. In Appendix Table 2, rows 2 and 3, we report similar estimates when also controlling for region-by-year fixed effects (20 regions) or subregion-by-year fixed effects (106 subregions) that further restrict the analysis to relative changes in county market access within nearby economic regions that cut across state and county boundaries.<sup>27</sup>

Our baseline specification also adjusts for the general Westward expansion of the United States, even within states, by controlling for year-interacted cubic polynomial functions of counties' latitude and longitude. As alternative functional forms, Appendix Table 2 reports similar estimates when controlling for fifth-order polynomials (row 4) or linear functions of counties' latitude and longitude (row 5). We also estimate similar impacts of market access when excluding from our sample the Plains and West Coast regions of the United States.<sup>28</sup>

As the Civil War occurred within our sample period, and had substantially different implications for different areas of the country, we further explore the sensitivity of our results to adjusting for differential impacts of the Civil War and the abolition of slavery. In Appendix Table 2, we report similar estimates when excluding from our sample the Southern Census region (row 7), all slave states (row 8), or all Confederate states (row 9). We also report similar estimates when excluding all counties that had Civil War battles (row 10) or counties with battles that totaled more than 500 casualties (row 11).<sup>29</sup> For the whole sample, row 12 reports similar estimates when controlling for whether a county had a Civil War battle, the number of battles (cubic polynomial), and the number of casualties (cubic polynomial), all interacted with year fixed effects.

---

<sup>27</sup>We assign each county the share of its area in each region or subregion, and control for those shares interacted with year fixed effects.

<sup>28</sup>The West Coast sample states are California, Oregon, and Washington. The Plains sample states are Kansas, Nebraska, and Texas.

<sup>29</sup>Our data includes all battlefields identified as significant by the Civil War Sites Advisory Commission's Report on the Nations Civil War Battlefields.

#### IV.D Robustness: Measurement of Productivity

Our measurement of county productivity may be subject to various types of measurement error. As an initial check on the sensitivity of our results to extreme changes in county productivity, Appendix Table 3 reports similar impacts of market access when excluding sample counties with the largest and smallest changes in productivity from 1860 to 1880: row 2 excludes the top and bottom 1% of counties, and row 3 excludes the top and bottom 5% of counties.

Firm inputs may be under-reported in the Census of Manufacturers, which would overstate firms' productivity. For example, Census enumerators ask firms for their labor costs and these were intended to include in-kind boarding costs and labor supplied by firm owners working on their own account, but there has been debate about whether firm owner labor is fully reflected in these costs (Weeks, 1886; Atack, 1977; Sokoloff, 1984; Margo, 2014). Row 4 of Appendix Table 3 reports similar impacts of county market access, though, when inflating labor costs to reflect potentially omitted labor costs.<sup>30</sup> A similar concern is that firms' reported capital values may not reflect their entire capital stock, such that annual capital expenditures are understated when multiplying firms' capital value by an interest rate.<sup>31</sup> Indeed, from Appendix Table 1, column 5, we calculate that national manufacturing capital expenditures are only 15% of total labor costs and capital costs, which is below the typical range of a third (Kaldor, 1957). As an adjustment, we double the assumed interest rate and re-calculate capital expenditures and county productivity, and row 5 of Appendix Table 3 reports similar impacts of market access on productivity.

Our measurement of county productivity does not depend on an assumed production function, but we do assume Cobb-Douglas production with constant returns to scale for our baseline decomposition of county productivity into reallocative efficiency and technical efficiency. In Appendix Table 3, we assume decreasing returns to scale and re-scale the production function elasticities to add up to 0.90 (row 6) or 0.95 (row 7) or assume increasing returns to scale and re-scale the elasticities to add up to 1.05 (row 8) or 1.10 (row 9). Assuming decreasing returns to scale mechanically increases the impact on technical efficiency (column 3) and reduces the impact on reallocative efficiency (column 2), while leaving unchanged the impact on county productivity (column 1). There continues to be relatively little substantive impact on technical efficiency, however, as most of the impact on county productivity is driven by changes in reallocative efficiency.

---

<sup>30</sup>For this specification, we add to county-by-industry labor costs the number of establishments multiplied by the average wage in that county and industry.

<sup>31</sup>Firms' reported book values of capital may also not reflect depreciation, which would overstate their capital stock and calculated annual expenditure, but then depreciation should also be included in their annual capital expenditure.

A related issue arises in our use of input cost shares to measure production function output elasticities, when decomposing county productivity into reallocative efficiency and technical efficiency. If input frictions are consistently higher for one input than the others, then the measured cost shares will understate the output elasticity for that input relative to the others. To assess the sensitivity of our estimates, in row 10 of Appendix Table 3, we increase the cost share for labor by 5 percentage points and proportionally decrease the cost shares for materials and capital. Similarly, we increase the cost share for materials by 5 percentage points (in row 11) or for capital by 5 percentage points (in row 12). These adjustments to the cost shares have no effect on the measurement of county productivity (in column 1), mechanically, and in practice have little substantive effect on the estimated impact of market access on reallocative efficiency and technical efficiency (columns 2 and 3).

We measure manufacturing productivity using data from the Census of Manufacturers, though there may be additional manufacturing activity not included in the Census of Manufacturers. The Census of Manufacturers reports a butter and cheese industry with only two establishments in 1860 and \$13 thousand of output, but reports 1195 establishments in 1870 with \$16.5 million of output.<sup>32</sup> If we exclude the butter and cheese industry in each decade, for greater consistency in data coverage over time, we estimate a similar impact of market access on productivity (row 13).

The Census of Manufactures missed some manufacturing establishments, particularly smaller establishments (U.S. Census, 1870). To the extent that the Census coverage varies over time and geographic areas, this would be partly corrected for by our inclusion of state-by-year fixed effects and year-interacted controls for county longitude and latitude. The remaining concern is that changes in county market access might be systematically correlated with changes in Census data coverage. While the general concern is that smaller establishments are more likely to be missed by Census enumerators, we report below (in Table 5) that changes in market access are not associated with changes in firm size.

We can use data from the Census of Agriculture on the value of home manufactures to expand our analysis of manufacturing beyond the Census of Manufacturers.<sup>33</sup> The national value of home manufactures is only 1.2% of total manufacturing output in 1860 and 0.5% of total manufacturing output in 1870, summing output values from the Census of Agriculture and the Census of Manufacturing, though home manufactures are substantively important in

---

<sup>32</sup>In 1880, the Census of Manufacturers reports 3250 establishments with \$30.4 million of output. There were also large technological changes in dairy manufacturing during the time period (Boberg-Fazlic and Sharp, 2019). The Census of Agriculture reports quantities of butter and cheese produced, but does not report data on their values or associated inputs.

<sup>33</sup>For this analysis, we assume that home manufacturers are not already included in the Census of Manufacturers. Spot checks of the Census of Agriculture manuscripts show the values of home manufacturers tend to be substantially less than the \$500 threshold used by the Census of Manufactures.

some counties. Among our sample counties, the median output shares of home manufacturing are 3.5% in 1860 and 1.4% in 1870 and the average output shares of home manufacturing are 11.8% in 1860 and 6.6% in 1870.<sup>34</sup> The estimated impact of county market access on the value of manufacturing output is 0.196 (0.050), using only data from the Census of Manufacturers, and the effect declines to 0.133 (0.062) when adding the value of home manufacturers. Data on home manufactures also allow us to expand the balanced sample of counties to include 189 additional counties that do not report manufacturing activity in one decade of the Census of Manufacturers, but do report home manufacturers in that decade. The estimated impact of county market access on county manufacturing output increases to 0.296 (0.079) when including these 189 additional counties and adding the value of home manufactures to data from the Census of Manufactures. There appears to be some shift from home manufacturing to more formal manufacturing activity, though these counties with initial home manufacturing activity are also substantially increasing total manufacturing activity following increases in market access.

Indeed, the United States itself expanded substantially from 1860 to 1880. Our baseline estimates focus on a balanced sample of counties from 1860 to 1880, which include 68% of the counties that existed in 1880 along with 91% of population and 94% of manufacturing output in 1880. When focusing on this balanced panel, however, the analysis does not include impacts on the extensive margin of manufacturing growth in newly created counties. Over this period, from 1860 to 1880, we estimate that a one standard deviation increase in market access leads to a 6.5 percentage point increase in the probability that a county reports any manufacturing activity in the Census of Manufacturers. We cannot estimate what happened to manufacturing productivity in these counties, which is not measured in the earlier periods, but increases in county market access are leading to growth on the extensive margin along with our estimated productivity effects on the intensive margin.

#### **IV.E Robustness: Measurement of Market Access**

Our measurement of county market access makes a number of simplifying assumptions that can induce various sources of measurement error. We begin by verifying that the estimates are not sensitive to omitting counties with the largest and smallest changes in market access (Appendix Table 4, rows 2 and 3). The estimated impact of market access becomes moderately larger (in row 3), which is consistent with baseline attenuation bias from classical measurement error.

Measured changes in county market access reflect changes in county-to-county transportation costs, which are based on assumed rates for transporting goods using railroads,

---

<sup>34</sup>The Census of Agriculture stopped asking this question in 1880, and we assume for our analysis here that there was zero home manufacturing in 1880.

waterways, and wagons. Appendix Table 4, rows 4 and 5, report similar impacts of market access on county productivity when decreasing the costs of waterways and wagons to the lowest values considered by Fogel (1964). Row 6 reports similar estimates when removing the costs of transshipment within the waterway network. Rows 7 – 9 report similar estimates when increasing the cost of railroad transportation to reflect potential congestion, fragmented track ownership, or differences in gauges that would require transshipment of goods or more indirect routes within the railroad network.

In Appendix Table 4, we focus on the estimated impact of a one standard deviation greater increase in county market access (column 1). Market access is re-scaled when changing the transportation cost parameters, which mechanically changes the estimated impact of a 1% greater increase in county market access (column 2). Similarly, because market access is calculated using county-to-county transportation costs scaled by the average price of goods, assuming a higher average price of goods will mechanically decrease the variation in market access and increase its estimated effect. There is a similar estimated impact from a one standard deviation greater increase in market access, however, when assuming different average prices (rows 10 and 11). The normalized impact of a one standard deviation increase in market access is also similar when using alternative values of  $\theta$ , the trade elasticity (rows 12 to 14), which reflect the 95% confidence interval around our baseline estimate of  $\theta$  (1.81 – 3.40) and a larger value of  $\theta$  (8.22) from Donaldson and Hornbeck (2016). Smaller values of  $\theta$  compress the distribution of changes in market access, just as for larger average prices, such that 1% greater increases in market access have more impact on county productivity but there is a correspondingly smaller increase in market access over time as the railroad network expands.

Our baseline measure of county market access reflects counties’ access to all other counties’ population, though we estimate similar impacts from a one standard deviation increase in modified definitions of county market access. To incorporate the influence of access to international markets, we inflate the population in 11 counties with major international ports to reflect the value of imports and exports in each year divided by GDP per capita (Appendix Table 4, row 15). The Census of Population is known to have undercounted population, especially in the South in 1870, and so in row 16 we inflate counties’ population by year and region based on the estimated degree of undercounting (Hacker, 2013). Our baseline measure considers counties’ access to population, under the assumption that workers earn a homogeneous income, but we also replace counties’ population with counties’ wealth as an alternative proxy for counties’ market size (row 17). The estimates are also not sensitive to whether we include a county’s own population in its market access (row 18), or omit counties’ access to other counties within 5 miles, 50 miles, or 200 miles (rows 19 – 21).

## IV.F Sources of Growth in Reallocative Efficiency

County market access has a substantial and robust impact on county manufacturing productivity, which is driven by increases in county reallocative efficiency. In this section, we explore the sources of these estimated increases in reallocative efficiency. Overall, we estimate that increases in county market access lead to substantial increases in county production, but little systematic change in the structure of the county economy itself.

Table 4, column 1, reports that county market access has substantial impacts on manufacturing input expenditures. A one-standard deviation greater increase in county market access leads to similar increases in materials expenditures, labor expenditures, and capital expenditures. These increases in input expenditures are driving the increases in reallocative efficiency, given the measured county-specific gaps between each input’s value marginal product and its marginal cost. By contrast, column 2 of Table 4 reports that county market access had little impact on the gaps themselves. Column 3 reports little impact on counties’ input-specific “wedges,” defined as the input cost share divided by the input revenue share. These estimates indicate that increases in county market access are not systematically reducing input frictions or firm markups; rather, the increases in reallocative efficiency reflect increased manufacturing activity in the presence of input frictions or firm markups.

For context, Appendix Tables 5 and 6 report the average input “gaps” and “wedges,” by region and decade, where gaps greater than zero and wedges greater than one reflect a range of factors that lower input use below efficient levels (e.g., firm markups, capital constraints, taxes, insecure property rights).<sup>35</sup> The gaps and wedges are generally greater in the Plains and West Coast states in 1860, and decrease from 1860 through 1880.<sup>36</sup> We do not find impacts of market access on input gaps and wedges, within region, but it appears that more newly settled regions exhibit greater input gaps and wedges that decline with further settlement.

Increases in county market access lead to increases in county manufacturing activity, but we do not find that manufacturing production shifts systematically toward industries that are more materials-intensive, labor-intensive, or capital-intensive. Column 4 of Table 4 reports little impact of county market access on the input intensity of county manufacturing industries.<sup>37</sup>

---

<sup>35</sup>The calculated input gaps are equal to that input’s cost share minus its revenue share, and we report the sum of these gaps across inputs. The calculated input wedges are equal to that input’s cost share divided by its revenue share, and we report output-weighted averages of these county-level wedges by region and decade. The reported average input wedge reflects a simple unweighted average across the three inputs (materials, labor, capital).

<sup>36</sup>The labor wedge notably increases in the Southern United States from 1860 to 1870, following the abolition of slavery, and returns to 1860 levels by 1880.

<sup>37</sup>The input intensity of county manufacturing industries is measured as the county’s output share in

Increases in manufacturing activity appear to be a general expansion of existing economic activity in the county. Column 1 of Table 5 reports no 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 firms, measured as the average value of output per firm (column 2) or the average number of workers per firm (column 3). Instead, increases in county market access lead to a substantial increase in the number of manufacturing firms (column 4), which is driving the overall increases in output and inputs.

While manufacturing activity increased substantially with increases in county market access, we do not estimate that increases in county market access prompted an economic shift from the agricultural sector toward the manufacturing sector. Table 5, column 5, reports little impact of county market access on manufacturing output as a share of total manufacturing and agricultural output. Similarly, columns 6, 7, and 8 report little impact on manufacturing value-added, surplus, or employment as a share of the total in manufacturing and agriculture.<sup>38</sup>

Further, we do not estimate that increases in county market access encouraged counties to specialize in either manufacturing or agriculture. We define a county specialization index as the squared difference between a county’s manufacturing share and the national manufacturing share, added to the squared difference between a county’s agricultural share and the national agricultural share. This index increases as county production becomes more intensive in the manufacturing sector or agricultural sector, relative to the national shares in each decade. Table 6, panel A, reports little impact of market access on sectoral specialization in terms of output (column 1), value-added (column 2), surplus (column 3), or employment (column 4). Similarly, looking within the manufacturing sector, panel B reports little impact of market access on specialization across manufacturing industries. Using our county-by-industry data, however, we are unable to examine within-industry specialization across product varieties.

These estimates highlight the broad expansion of manufacturing activity (and agricultural activity) in counties that experience relatively greater increases in market access, and motivate several assumptions underlying the aggregate analysis in Section V. In particular, we will assume in Section V that increases in county market access do not affect the county’s

---

each industry multiplied by national industry cost shares. There is substantial industry entry and exit in these historical data, which makes it more difficult to interpret within- and between-industry analysis in a multi-sector model (Caliendo and Parro, 2015; Fajgelbaum and Redding, 2018), and which encourages us to model county-level production functions.

<sup>38</sup>For agriculture, we define value-added as 92% of output, reflecting an assumed 8% materials share in agriculture (Towne and Rasmussen, 1960), and we define surplus as the value of land multiplied by the state mortgage interest rate. For manufacturing, we define value-added as output minus materials expenditures, and we define surplus as output minus all input expenditures.

production function elasticities (as in Tables 4 and 5). We also assume that county market access does not directly affect county input frictions or firm markups (as in Table 4). These assumptions allow us to focus on how changes in county market access affect aggregate economic activity, in the presence of resource misallocation, and the resulting impacts on aggregate productivity.

## V Aggregate Counterfactual Analysis

We now examine aggregate economic gains from the railroads. The railroads increased economic activity in counties with positive “gaps,” where firms’ value marginal product of inputs exceeded their marginal cost, but some of this increase reflects the shifting of inputs from other counties that may also have positive gaps. The previous regressions estimate the relative impact on counties from relative increases in market access. By drawing further on the model’s structure, however, we can also consider aggregate economic impacts from the expansion of the railroad network and aggregate economic losses under counterfactual networks.

We extend a baseline model of economic geography (Eaton and Kortum, 2002; Donaldson and Hornbeck, 2016) to include market distortions. Each county is affected by changes in transportation costs due to expansion of the railroad network, through interacting goods markets and factor markets across all counties, and these effects are summarized by changes in a county’s market access. We show how the presence of market distortions causes there to be aggregate productivity gains from reductions in transportation costs, which are not captured by changes in land values or social savings calculations. We then estimate the aggregate productivity losses from counterfactual transportation networks, such as removing the railroad network or replacing the railroad network with an extended network of canals.

### V.A Model Setup

Firms in each county have a Cobb-Douglas production function for good variety  $j$ , which uses labor  $L$ , capital  $K$ , land  $T$ , and materials  $M$ . In county  $o$ , these inputs cost a wage  $w_o$ , interest rate  $r_o$ , rental rate  $q_o$ , and price index  $P_o$ ; alternatively, input  $k$  costs  $W_o^k$  in county  $o$ . Firms use a continuum of good varieties as materials, with a constant elasticity of substitution across varieties, and so  $P_o$  is the CES price index over good varieties.<sup>39</sup>

Firm production functions vary across counties in their Hicks-neutral technical efficiency  $z_o(j)$ . Following Eaton and Kortum (2002), each county has an exogenous technical efficiency level  $z_o(j)$ , for each variety  $j$ , drawn from a Fréchet distribution with CDF given by:  $F_o(z) =$

---

<sup>39</sup> $P_o \equiv [\int_0^n (p_o(j))^{1-\sigma} dj]^{1/(1-\sigma)}$ , where  $\sigma$  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  $o$ .

$1 - e^{-A_o z^{-\theta}}$ , with  $\theta > 1$ .<sup>40</sup> The parameter  $A_o$  captures average technical efficiency in county  $o$  (absolute advantage), while the parameter  $\theta$  captures the standard deviation of technical efficiency across varieties (scope for comparative advantage). A smaller  $\theta$  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. We also allow for firm production functions to vary across counties in their exogenous production function elasticities  $(\alpha_o^L, \alpha_o^K, \alpha_o^T, \alpha_o^M)$ . The marginal cost of producing variety  $j$  in county  $o$  is:

$$(14) \quad MC_o(j) = \frac{(w_o)^{\alpha_o^L} (r_o)^{\alpha_o^K} (q_o)^{\alpha_o^T} (P_o)^{\alpha_o^M}}{z_o(j)} = \frac{\prod_k (W_o^k)^{\alpha_o^k}}{z_o(j)}.$$

Marginal costs are lower when input costs are lower and when firm technical efficiency is higher.

The main addition in our model, as 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 ( $\psi^L$ ), capital ( $\psi^K$ ), land ( $\psi^T$ ), or materials ( $\psi^M$ ). Given any positive friction, firms in county  $o$  reduce production of variety  $j$  such that its price is greater than its marginal cost of production:

$$(15) \quad p_o(j) = \frac{\prod_k ((1 + \psi_o^k) W_o^k)^{\alpha_o^k}}{\beta_o(j)} > MC_o(j).$$

Firm markups, due to market power, can be represented as a common component in each of the  $\psi$  terms.<sup>41</sup> 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  $\psi$  terms.

We assume capital is mobile, such that interest rates are fixed exogenously, but this interest rate can vary across counties.<sup>42</sup> We assume land is fixed in each county (i.e., the total quantity of physical land in each county), and its price is endogenous.

Labor is supplied by workers, who consume good varieties  $j$  in the same manner that firms use these varieties in roundabout production, and so workers pay the same CES price index  $P_o$  as firms (Redding and Venables, 2004). A worker living in county  $o$ , and receiving

---

<sup>40</sup>This Fréchet distribution reflects firms receiving levels of technical efficiency from any distribution and discarding all but the best for each variety.

<sup>41</sup>While the input-specific frictions are modeled as constants, the assumed CES demand for firm output also implies constant firm markups.

<sup>42</sup>In practice, given data constraints, we allow this interest rate to vary across states according to the measured average mortgage interest rate in each state (Fogel, 1964).

a wage  $w_o$ , receives indirect utility  $V(P_o, w_o) = w_o/P_o$ . We assume workers are perfectly mobile across counties, such that worker utility is the same across counties at  $\bar{U}$ . The local wage rate is endogenous, however, and reflects local prices ( $w_o = \bar{U}P_o$ ).<sup>43</sup>

There is costly trade of good varieties across counties. Transporting goods from county  $o$  (origin) to county  $d$  (destination) incurs a proportional “iceberg” trade cost  $\tau_{od}$ .<sup>44</sup> In county  $d$ , the price of good variety  $j$  produced in county  $o$  is then given by:  $p_{od}(j) = \tau_{od}p_{oo}(j)$ , where  $p_{oo}(j)$  is the factory-gate price in county  $o$  and  $\tau_{od} > 1$ .

## 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 15) and the transportation cost ( $\tau_{od}$ ). 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  $\psi$ , the value of total exports from county  $o$  to county  $d$  is given by:<sup>45</sup>

$$(16) \quad E_{od} = \kappa_1 A_o \left( \Pi_k \left( (1 + \psi_o^k) W_o^k \right)^{\alpha_o^k} \right)^{-\theta} \tau_{od}^{-\theta} Y_d P_d^\theta.$$

County  $o$  sends more goods to county  $d$  when county  $o$  has higher technical efficiency ( $A_o$ ) or lower “effective costs”  $\left( \Pi_k \left( (1 + \psi_o^k) W_o^k \right)^{\alpha_o^k} \right)$ , where “effective costs” reflect input prices and input frictions. Higher input frictions have the same impact on trade flows as higher input prices or lower technical efficiency. County  $o$  also sends more goods to county  $d$  when bilateral transportation costs are lower ( $\tau_{od}$ ), when county  $d$  has higher income ( $Y_d$ ), and when county  $d$  has a higher goods price index ( $P_d$ ).

Consumer market access (CMA) in county  $d$  is an inverse transformation of the goods price index (Redding and Venables, 2004; Donaldson and Hornbeck, 2016):

$$(17) \quad CMA_d = P_d^{-\theta} = \kappa_1 \sum_o \tau_{od}^{-\theta} A_o \left( \Pi_k \left( (1 + \psi_o^k) W_o^k \right)^{\alpha_o^k} \right)^{-\theta}.$$

Consumer market access is higher in county  $d$  when it has access to cheaper goods: when there are lower costs of transporting goods from counties with higher technical efficiency and

<sup>43</sup>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 output.

<sup>44</sup>We measure the absolute cost of transporting goods between counties (Section I.B), and express this absolute cost in proportional terms using the average value of transported goods (estimated below, along with the estimation of the trade elasticity  $\theta$ ).

<sup>45</sup>Here,  $\kappa_1 = \left[ \Gamma \left( \frac{\theta+1-\sigma}{\theta} \right) \right]^{-\frac{\theta}{1-\sigma}}$ , where  $\Gamma(\cdot)$  is the  $\Gamma$  function defined by  $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx$ .

lower “effective costs.” Input frictions in county  $o$  lower consumer market access in county  $d$  because county  $d$  is not able to fully benefit from low marginal costs in county  $o$ .

Firm market access (FMA) in county  $o$  is a sum over firms’ access to all destination counties, adjusting for those destination counties’ access to other sources of goods:

$$(18) \quad FMA_o = \sum_d \tau_{od}^{-\theta} Y_d CMA_d^{-1}.$$

Firm market access is higher in county  $o$  when it has access to more product demand: when there are lower costs of transporting goods to counties with higher consumption, which have less access to other sources of goods ( $CMA_d$ ). We can also represent consumer market access in county  $d$  as a sum over consumers’ access to all origin counties:

$$(19) \quad CMA_d = \sum_o \tau_{od}^{-\theta} Y_o FMA_o^{-1}.$$

Similar to equation 18, consumer market access is higher in county  $d$  when it has access to more product supply: when there are lower costs of transporting goods from counties with higher production, which have less access to other destinations for goods ( $FMA_o$ ). Indeed, equations 18 and 19 imply that a county’s firm market access and consumer market access are exactly proportional:  $FMA_o = \rho CMA_o$ , where  $\rho > 0$ .<sup>46</sup> We therefore use a single measure of “market access” (MA), which reflects the ideas underlying both firm market access and consumer market access:  $MA_o \equiv FMA_o = \rho CMA_o$ . Given that workers receive a fixed share of output ( $\alpha^L$ ), we express market access in county  $o$  as a function of the endogenous number of workers in each other county  $d$ :<sup>47</sup>

$$(20) \quad MA_o = \kappa_2 \sum_d \tau_{od}^{-\theta} L_d MA_d^{\frac{-(1+\theta)}{\theta}} \frac{(1 + \psi_d^L)}{\alpha_d^L}.$$

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 zero labor input frictions and a homogeneous labor cost share.<sup>48</sup>

Changes in county market access summarize how each county is impacted by changes

---

<sup>46</sup>For this result, we assume that trade costs are symmetric ( $\tau_{od} = \tau_{do}$ ), as constructed in Section I.B.

<sup>47</sup>Here,  $\kappa_2 = \bar{U} \rho^{\frac{1+\theta}{\theta}}$

<sup>48</sup>Note that other input frictions also affect equilibrium behavior but, conditional on the distribution of population, market access can be expressed as a function of labor, labor input frictions, and labor cost shares.

in transportation costs. In the model, changing transportation costs between two counties has positive and negative spillover effects on other counties due to linked goods markets and factor markets. Changes in county market access summarize these impacts from changes in counties' connectedness and changes in overall market integration.<sup>49</sup> We solve for county market access in each county and each decade, using: the N-by-N system of equations for market access in each county (equation 20); data on county population and county-to-county transportation costs; and parameter choices discussed below in Section V.

The calculated changes in county market access, as derived from the full model, are highly correlated with the approximated changes in county market access used in the previous regression analysis (equation 20 vs. equation 1). From 1860 to 1880, changes in log market access derived from the model (equation 20) and changes in log market access based on our approximation (equation 1) have a correlation coefficient of 0.9995. When estimating equation 13, a one standard deviation greater increase in model-defined market access is estimated to increase county productivity by 0.130 (0.049), increase reallocative efficiency by 0.117 (0.045), and increase technical efficiency by 0.012 (0.008), which represent only rounding differences from our estimates in Table 1 for approximated market access.<sup>50</sup>

Figure 4 shows approximately log-linear impacts of market access, as predicted in Section V.C, on county productivity (panels A and B), reallocative efficiency (panels C and D), and technical efficiency (panels E and F).<sup>51</sup> Figure 4 shows the local polynomial relationship between residual productivity and residual market access, after partialling out the controls from estimating equation 13.

The approximated measure of county market access is sufficient for the regression analysis of relative changes in market access, but the full model-defined expression for market access is required for the aggregate counterfactual analysis. This is because we need to determine not only relative changes in county market access but the absolute changes in counties' market access under counterfactual scenarios.

---

<sup>49</sup>One other effect, discussed below in the counterfactual analysis, is that worker welfare can also change across all counties.

<sup>50</sup>A one standard deviation greater change in model-defined log market access is 0.223, from 1860 to 1880, which is similar to that for approximated market access (0.231).

<sup>51</sup>These figures are for approximated market access (panels A, C, E) and model-defined market access (panels B, D, F).

## V.C Predicted Impacts of Market Access

We now consider the predicted impacts of market access on productivity and other county-level outcomes. The impact of market access on productivity is given by:

$$(21) \quad \frac{d \ln PR_o}{d \ln MA_o} = \sum_k (\alpha_o^k - s_o^k) \frac{d \ln X_o^k}{d \ln MA_o}.$$

Market access increases productivity by increasing input usage ( $\frac{d \ln X_o^k}{d \ln MA_o}$ ), when the value marginal product of that input exceeds its marginal cost (i.e., when the cost share  $\alpha$  is greater than the revenue share  $s$ ).<sup>52</sup>

Building on equation 21, as described in the Appendix, we show that changes in market access have a log-linear impact on county productivity:

$$(22) \quad \begin{aligned} \frac{d \ln PR_o}{d \ln MA_o} = & (\alpha_o^L - s_o^L) \left( \frac{1}{\theta} + \frac{\alpha_o^M + \alpha_o^L + 1}{1 + \theta \alpha_o^T} \right) \\ & + (\alpha_o^M - s_o^M) \left( \frac{1}{\theta} + \frac{\alpha_o^M + \alpha_o^L + 1}{1 + \theta \alpha_o^T} \right) \\ & + (\alpha_o^K - s_o^K) \left( \frac{\alpha_o^M + \alpha_o^L + 1}{1 + \theta \alpha_o^T} \right). \end{aligned}$$

Market access increases capital usage by  $\left( \frac{\alpha_o^M + \alpha_o^L + 1}{1 + \theta \alpha_o^T} \right)$ , which then increases productivity by  $(\alpha_o^K - s_o^K) \left( \frac{\alpha_o^M + \alpha_o^L + 1}{1 + \theta \alpha_o^T} \right)$ . Market access has a larger impact on labor and materials usage, with the additional term  $\left( \frac{1}{\theta} \right)$  reflecting that market access also decreases materials costs and nominal wages in county  $o$ .<sup>53</sup>

This log-linear impact of market access on productivity can be re-written using the input frictions ( $\psi$ ):

$$(23) \quad \begin{aligned} \frac{d \ln PR_o}{d \ln MA_o} = & \alpha_o^L \left( \frac{\psi_o^L}{1 + \psi_o^L} \right) \left( \frac{1}{\theta} + \frac{\alpha_o^M + \alpha_o^L + 1}{1 + \theta \alpha_o^T} \right) \\ & + \alpha_o^M \left( \frac{\psi_o^M}{1 + \psi_o^M} \right) \left( \frac{1}{\theta} + \frac{\alpha_o^M + \alpha_o^L + 1}{1 + \theta \alpha_o^T} \right) \\ & + \alpha_o^K \left( \frac{\psi_o^K}{1 + \psi_o^K} \right) \left( \frac{\alpha_o^M + \alpha_o^L + 1}{1 + \theta \alpha_o^T} \right). \end{aligned}$$

---

<sup>52</sup>Note that county productivity is also increasing in its technical efficiency ( $A_o$ ), though in the model we have assumed that technical efficiency is exogenous and not directly affected by market access. As a consequence, we are only now considering productivity impacts of market access through increases in reallocated efficiency.

<sup>53</sup>These input price effects are why the previous reduced-form estimates understate the impact of market access on productivity, as the measured increase in input expenditures understates the increase in real input usage (as discussed in Section II.C).

There is no impact of market access on productivity when each input friction is equal to zero. Input frictions also have greater impact when applied to more important inputs (i.e., inputs with higher cost shares).

Market access increases productivity by increasing input usage, when input usage is otherwise below the efficient level due to input frictions. The Appendix derives the log-linear impact of market access on each input, as well as the log-linear impact of market access on each input price. An interesting result is that input frictions do not change the impact of market access on output or inputs, in percentage terms, similar to how the impact of market access does not depend on local technical efficiency. Input frictions do determine whether market access impacts productivity, however, as captured in equations 22 and 23.

Input frictions on land do not affect the impact of market access on productivity (equations 22 and 23), as the model assumes land is in fixed supply. Market access does increase land expenditure, however, as the price of land is log-linear in market access:

$$(24) \quad \frac{d \ln q_o}{d \ln MA_o} = \frac{\alpha_o^M + \alpha_o^L + 1}{1 + \theta \alpha_o^T}$$

Adding input frictions to the model does not affect the impact of market access on land value, which is the main estimated impact in Donaldson and Hornbeck (2016). We obtain the same predicted impact of market access on land value as Donaldson and Hornbeck (2016), when replacing our county-specific sum of the labor share and materials share ( $\alpha_o^M + \alpha_o^L$ ) with their average labor share of value-added ( $\alpha^L$ ).<sup>54</sup>

## V.D Model Interpretation and Discussion of Assumptions

Our model highlights the impact of market access on productivity, in the presence of input frictions, which generates a separate source of economic gains from the railroads that are in addition to the impacts on land values considered by Donaldson and Hornbeck (2016). In that model, all economic gains from increased market access are captured by the increase in land values.<sup>55</sup> 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 economic gains from the railroads are given by the increase in aggregate productivity in addition to the increase in land values.

Our model extends the approach in Donaldson and Hornbeck (2016), and is equivalent to that model when removing input frictions, removing intermediates, and constraining the

---

<sup>54</sup>Input frictions do affect land values in levels, just as unobserved technical efficiency affects land values in levels.

<sup>55</sup>There is also incidence on aggregate worker utility if there is not perfect labor mobility into the United States, which we discuss later.

production function elasticities to be constant across counties. The most important departure is the introduction of input frictions. We allow the elasticities to vary across counties so that we do not confound measured county-specific frictions with county variation in elasticities. By allowing for traded materials, similar to the Appendix in Donaldson and Hornbeck (2016), we adjust the chosen elasticities and are able to consider productivity gains from reallocation of materials.

We make some important assumptions to maintain tractability in this general equilibrium setting, while extending the model to include input frictions. In particular, we assume that several features of the economy are exogenous and not impacted by changes in market access. These assumptions are consistent, however, with the reduced-form evidence in Section IV.

First, we assume that county input frictions are exogenous. This assumption is consistent with estimates from Table 4, in which the measured “gaps” and “wedges” were not impacted by changes in market access.

Second, we assume that county production function elasticities are exogenous. This assumption is consistent with estimates from Table 4, which show little impact of market access on elasticities in manufacturing, and estimates from Table 4 that show similar percent changes in labor, materials, and capital expenditures. Further, Table 5 shows little impact of market access on the county manufacturing share of output, value-added, surplus, or employment.<sup>56</sup>

Third, we assume that county technical efficiency is exogenous. This assumption is consistent with estimates from Table 1 that showed little impact of market access on revenue-based technical efficiency, which is often correlated with technical efficiency in data when both are observed (Foster, Haltiwanger and Syverson, 2008; Haltiwanger, Kulick and Syverson, 2018).<sup>57</sup> Our counterfactual analysis then considers impacts on productivity only through changes in reallocative efficiency, under the conservative assumption that technical efficiency does not also decline in the absence of the railroads.

## V.E Estimating the Model

We have assumed that workers move across counties to equalize real wages, and so we use data on county population to infer an “amenity” value for each county. This “amenity” value reflects the quantity of fixed factors in each county, along with the county’s exoge-

---

<sup>56</sup>Note that the model allows for specialization in firms’ product varieties, within sectors or industries, as opposed to changes across sectors or industries that are associated with different elasticities. Table 6 reports no impacts of county market access on cross-sector specialization or cross-industry specialization, but our data do not allow for estimating impacts on product variety specialization within sectors or industries.

<sup>57</sup>Technical efficiency would not be correlated with revenue-based technical efficiency in our model, however, due to the assumptions of CES demand and constant returns to scale (which imply a constant markup of price over marginal cost).

nous technical efficiency.<sup>58</sup> Given our measured county-to-county transportation costs, the county population data imply a unique vector of county amenity values (see Appendix for details).<sup>59</sup> Holding those amenity values fixed, and using counterfactual county-to-county transportation costs, we can then calculate an implied counterfactual population in each county and associated counterfactual changes in capital and materials usage. Using our estimated county-level input frictions, and the associated gap between inputs’ value marginal product and marginal cost in each county, we then calculate the aggregate productivity impact from removing the railroad network (or from other counterfactual transportation networks).

The model describes overall economic activity in each county, which corresponds to the county’s total population and all goods consumption and production. We only have detailed data for manufacturing industries, however, on the value of output and value of inputs. We use these data to measure input frictions within the manufacturing sector in each county, as in Section II and IV, and we begin by assuming these input frictions are constant across sectors within each county.<sup>60</sup> We also estimate average production function elasticities for each county in the manufacturing sector only, and so we calculate county-level elasticities also using nationwide production function elasticities in agriculture.<sup>61</sup>

We jointly estimate values for the trade elasticity ( $\theta$ ) and the average price of transported goods ( $\bar{P}$ ).<sup>62</sup> For any given trade elasticity  $\theta$ , the model predicts a matrix of county-to-county railroad shipments (in tons) and county output (in dollars), which we use to generate a value for  $\bar{P}$  that jointly maximizes closeness of fit with the Census data on total manufacturing and agricultural output (in dollars), and total US railroad shipments in 1890 in tons (Adams,

---

<sup>58</sup>These two variables are not separately identified, but we only need their combined value for the counterfactual analysis.

<sup>59</sup>The vector of county amenity values is unique up to a proportional constant, whose value does not affect the analysis.

<sup>60</sup>As an alternative exercise, we consider counterfactual impacts when assuming there are input frictions only in the manufacturing sector.

<sup>61</sup>For the agricultural sector, 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 by taking the output-weighted average across its manufacturing industry elasticities in 1880. We then calculate county-level elasticities, weighting by the fraction of county output in agriculture or manufacturing in 1880: in 1880, manufacturing output is 30% of total output in the average county (and 70% of national output in manufacturing and agriculture). For 1880, Gallman (1960) and Gallman and Weiss (1969) find that manufacturing value added was 40% of national value added in manufacturing and agriculture, and that manufacturing value added was 20% of overall value added (GDP). For 234 counties in the counterfactual sample with missing county-by-industry data in 1880, which make up 0.72% of population in 1880, we use county-level manufacturing data from 1890.

<sup>62</sup>We use the average price of transported goods ( $\bar{P}$ ) to calculate the iceberg trade costs ( $\tau_{cd}$ ) based on the measured county-to-county transportation costs: ( $\tau_{cd} = 1 + t_{cd}/\bar{P}$ ).

1895). Given that average price per ton  $\bar{P}$  and trade elasticity  $\theta$ , we then estimate the impact of market access on agricultural land values (as in Donaldson and Hornbeck, 2016), under the assumption that land markets are integrated within counties. We choose the  $\theta$  (and corresponding  $\bar{P}$ ) that minimizes mean squared error of regressing the model-predicted change in land values (from Equation 24) on the observed change in land values. We estimate an average price of transported goods equal to 35.3, which is very close to the value of 35 assumed by Donaldson and Hornbeck (2016) following Fogel (1964). We estimate a value of  $\theta$  equal to 2.75 (95% confidence interval: 1.81 – 3.40), which is lower than the value of 8.22 estimated by Donaldson and Hornbeck (2016), but these estimates span the general range considered in this literature (Costinot and Rodríguez-Clare, 2014).<sup>63</sup>

## V.F Estimated Counterfactual Impacts

We estimate that aggregate productivity would have been 24.8% lower in the United States in 1890, if there were no railroad network and the economy had to rely on the existing waterway network and high-cost wagon transportation (Table 7, panel A, column 1).<sup>64</sup> This 25% decrease in aggregate productivity is worth annually 25% of GDP, or \$3 billion in 1890 dollars. We estimate substantially larger economic impacts of the railroads than previous estimates of 3.2% (Donaldson and Hornbeck, 2016) or 2.7% (Fogel, 1964).<sup>65</sup> Our estimated economic impact is in addition to either of those previous estimates, which we discuss further in Section VI. Our estimated decrease in aggregate productivity also reflects only decreases in reallocative efficiency, holding fixed counties’ technical efficiency.

Figure 5 maps the county-level counterfactual changes in productivity, where darker shaded counties represent larger counterfactual declines productivity in the absence of the railroads. These declines in county-level 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, as in Equation 23.

In estimating the decline in aggregate productivity, we consider several scenarios for counterfactual changes in the United States’ total population. Our baseline estimates reflect the counterfactual decline in total population that holds fixed worker utility. 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 holding fixed total population, we solve for counties’ population shares in the ab-

---

<sup>63</sup>Donaldson and Hornbeck (2016) estimate  $\theta$  using a similar approach, in a similar setting, though we extend this approach to jointly estimate  $\theta$  and  $\bar{P}$ , allow for heterogeneous gross output production functions, and expand the sample to 1860–1900.

<sup>64</sup>We focus on counterfactual estimates for 1890, following Fogel (1964).

<sup>65</sup>We also estimate counterfactual impacts on land values that contribute an additional loss equal to 3.6% of GDP, similar to the estimate of 3.2% from Donaldson and Hornbeck (2016).

sence of the railroad network. We then calculate corresponding changes in other production inputs, output, and the resulting change in aggregate productivity in the United States. Panel B reports that productivity is estimated to fall by 5.3% in 1890, in the absence of the railroad network, when maintaining the same total United States population. Population and other production inputs become condensed into limited geographic areas, decreasing labor productivity, and so worker utility falls by 33.6%.

When allowing for aggregate declines in population, the model predicts a substantial aggregate decline in population in the United States. For worker utility (or real wages) to be unchanged in the counterfactual, the model predicts that the United States population would need to be 68% 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. If we instead assumed that total United States population would be lower by 33% in the absence of the railroads, reflecting the native-born population of native-born parents in 1890, then we would estimate a 13% decline in productivity and a 24% decline in worker utility.<sup>66</sup> By comparison, the United States population was 37% lower in 1870 and 73% lower in 1840 than in 1890.

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 2.2% lower using only railroads that existed in 1880, 8.8% lower using railroads from 1870, 14.2% lower using railroads from 1860, or 20.4% lower using railroads from 1850 (Table 7, panel A, columns 2 – 5). In a related exercise, we can constrain the railroad network to what existed in 1860 and set aggregate population to what existed in 1860: we then capture 40% of the observed cross-county variation in 1860 county output.<sup>67</sup>

Additional canals might have been constructed to mitigate 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 Fogel’s proposed system of feasible canals that would have brought much of the settled United States within 40 miles of low-cost waterway transportation. We estimate that productivity would have been lower by 21.7% in 1890 when replacing the railroad network with this system of additional canals (Table 7,

---

<sup>66</sup>For this calculation, using the available Census tabulations, we reduce the total United States population by excluding foreign-born population in 1890 and the white native-born population in 1890 with foreign-born parents.

<sup>67</sup>That is, the R-squared is 0.40 when regressing actual 1860 log output on model-predicted 1860 log output. We would not expect to explain all variation in county output, given other unobserved shocks between 1860 and 1890 (e.g., as captured in our regression analysis by state-year fixed effects and controls for county latitude/longitude).

panel A, column 6), such that the additional canals would have mitigated only 13% of the losses from removing the railroad network.<sup>68</sup>

By contrast, the railroads would have been fairly cheap at twice the price. We estimate that productivity would have been lower by 8.0% in 1890 if railroad rates were double (Table 7, panel A, column 7). Compared to losing access to the railroad network entirely, using these more-expensive railroads would mitigate 68% of the productivity decline.

We would estimate zero impact of the railroads on productivity, mechanically, if the economy were efficient. Our baseline counterfactual assumes the measured frictions in the manufacturing sector also reflect frictions in the agricultural sector. Alternatively, we could assume 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.<sup>69</sup> In this case, we estimate that aggregate productivity would have declined by 15.4% in 1890 without the railroad network (Appendix Table 7, column 2).

Our estimated counterfactual impacts on productivity are not sensitive to the estimated value of  $\theta$ , the trade elasticity. Our point estimate for  $\theta$  is 2.75, with a 95% confidence interval between 1.81 and 3.40, which is smaller than the value of 8.22 estimated by Donaldson and Hornbeck (2016). For alternative values of  $\theta$ , we estimate that removing the railroad network would have lowered productivity in 1890 by 25.6% ( $\theta = 1.81$ ), 24.3% ( $\theta = 3.40$ ), or 21.5% ( $\theta = 8.22$ ). As in Donaldson and Hornbeck (2016), alternative values of  $\theta$  effectively re-scale market access but have little effect on the estimated impacts.<sup>70</sup> We also estimate an average price of traded goods,  $\bar{P} = 35.3$ , which is similar to the value of 35 assumed by Donaldson and Hornbeck (2016) following Fogel (1964). For alternative values of  $\bar{P}$ , we estimate counterfactual productivity declines of 32.3% ( $\bar{P} = 20$ ) and 20.4% ( $\bar{P} = 50$ ).<sup>71</sup>

## V.G Back-of-the-Envelope Magnitude Calculations

There are some useful back-of-the-envelope calculations to help understand the magnitude of the estimated counterfactual impact on productivity.

The substantial estimated impacts on productivity are driven by county-level gaps between the marginal value of inputs and their marginal cost, multiplied by counterfactual

---

<sup>68</sup>Productivity falls by less when holding total population fixed (Table 7, panel B, column 6), but then worker utility falls substantially when replacing the railroads with an extended canal network.

<sup>69</sup>We only measure output in manufacturing and agriculture, and so we are now effectively assuming that wedges in other sectors are equal to the weighted average wedges across manufacturing (from Section IV) and agriculture ( $\psi = 0$ , by assumption).

<sup>70</sup>This is because we use the observed distribution of population across counties to discipline the model, in estimating local amenities in each county, which effectively counteracts the change in  $\theta$ .

<sup>71</sup>Assuming alternative values of  $\bar{P}$  effectively re-scales the baseline transportation rates, whereby lower values of  $\bar{P}$  magnify differences between the factual and counterfactual scenarios and higher values of  $\bar{P}$  diminish differences between the factual and counterfactual scenarios.

changes in county-level production inputs. These gaps are measured as an input’s cost share minus its revenue share, and the national output-weighted average gaps are 0.071 (for materials), 0.041 (for labor), and 0.008 (for capital), such that the sum of these gaps is 0.120. If we assume that the counterfactual declines in production inputs were the same in each county (67% or 1.11 log points), then productivity in each county would decline by 0.133 log points (0.120 times 1.11). This corresponds to a national productivity decline of 18.9% (0.209 log points), once multiplying the productivity decline in each county by the sum of the county-level Domar weights (0.133 times 1.57).<sup>72</sup> Our calculated counterfactual decline in national productivity is larger (24.8%), but this also reflects cross-county heterogeneity in the gaps, population declines, and Domar weights.

Productivity declines in our counterfactual scenarios because of gaps between the value marginal product of inputs and their marginal cost, but our overall gap (0.12) is not especially large. For the United States, from 1958 to 2011, the manufacturing gap was larger and increased from approximately 0.20 to 0.30.<sup>73</sup> Thus, the substantial impacts of the railroads on aggregate productivity are not driven by especially large measured gaps in our data. The overall decline in productivity does reflect a substantial decline in aggregate population, in the absence of the railroads. We report more moderate scenarios, with smaller declines in aggregate population, but those productivity effects are also substantial and also associated with substantial declines in worker utility.

## VI Interpretation

We estimate that the railroads substantially increased the scale of the United States’ economy: increasing the production and use of materials, spurring increased capital investment, and encouraging population growth. The economic consequences of this expansion are substantially greater than previously considered because, in most counties, the value marginal products of materials, capital, and labor were greater than their marginal costs. We do not estimate that railroads reduced these market distortions, whether due to firm markups or input frictions, but the railroads generated substantial economic gains by encouraging the expansion of an economy with market distortions.

---

<sup>72</sup>To see why the Domar weights are the appropriate way to aggregate county-level productivity changes to the corresponding national productivity change, consider the following equivalence. The dollar loss in each county is given by multiplying each county’s productivity decline by that county’s output. Summing this loss across all counties, and dividing by national value-added or GDP, gives the percentage decline in national productivity. Equivalently, we can multiply the uniform decline in each county by the sum of the Domar weights, which is defined as the sum of county output divided by national value-added. Our counterfactual estimate assigns each county the average of its factual and counterfactual Domar weight, though in both scenarios the sum of the Domar weights is approximately 1.57.

<sup>73</sup>Using the NBER CES (Becker, Gray and Marvakov, 2013) database and assuming that the cost of capital is 8%

We calculate that aggregate productivity would have been 25% lower in 1890 in the absence of the railroads, through declines in reallocative efficiency alone. We assume that technical efficiency would have been unchanged in the counterfactuals, but this decline in reallocative efficiency is equivalent to a 25% decline in technical efficiency (or total factor productivity, TFP). It is challenging to estimate aggregate TFP growth, with the proper price deflators, but estimates suggest that annual TFP growth was approximately 0.37% from 1855 to 1890 and 1.24% from 1890 to 1927 (Abramovitz and David, 1973). That is, the railroads effectively contributed 31 years worth of technological innovation, by driving increases in reallocative efficiency.<sup>74</sup>

The railroads' 25% impact on productivity was worth 25% of GDP in 1890, or \$3 billion in 1890 dollars. As a comparison, the estimated cost of the railroad network in 1890 was \$8 billion (Adams, 1895). We estimate that the railroads generated an annual private return of 4.6% in 1890,<sup>75</sup> which increases to an annual social return of 8.6% – 9.4% once including estimates from Fogel (1964) or Donaldson and Hornbeck (2016) and increases to 45% when also including our estimated impact on productivity.<sup>76</sup> These estimates imply that the railroad sector was capturing roughly 10% of its social return in 1890.

Our estimated increases in productivity do not include the direct benefits of the railroads from decreasing resources spent on transportation. To see this, consider that we would mechanically estimate no impact on productivity from the railroads if there were no differences between value marginal product and marginal cost, whereas the economy would still benefit through decreases in transportation costs. In our model, those decreases in transportation costs are capitalized into higher land values.

Donaldson and Hornbeck (2016) estimate that agricultural land values would have fallen by 60% without the railroad network, which, multiplying by an interest rate, generates annual economic losses equal to 3.2% of GDP. The total loss of all agricultural land would only generate annual economic losses equal to 5.35% of GDP, so an analysis of agricultural land values could never find larger economic impacts.<sup>77</sup>

---

<sup>74</sup>For this calculation, we take the average growth rate of TFP from before 1890 and after 1890 (0.81%), from Abramovitz and David (1973).

<sup>75</sup>For this calculation, we sum the railroads' reported net income (\$145 million), debt interest payments (\$217 million), and net capital expenditure (\$5 million), and divide by the cost of the railroads including equipment (\$8.041 billion). 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.

<sup>76</sup>This annual social return of 45% reflects the annual private return (\$367 million), our estimated annualized increase in land value (\$431 million), and our estimated increase in annual productivity (\$2.796 billion), divided by the cost of the railroads including equipment (\$8.041 billion).

<sup>77</sup>In our models, land is fixed and labor and capital are both fully mobile. Thus, the increase in land values captures the entire increase in factor payments due to the railroads.

The crucial difference in our approaches is that Donaldson and Hornbeck (2016) assume an efficient economy, in which value marginal product is equal to marginal cost, such that all output value is paid to factors. By contrast, our estimated increases in reallocative efficiency reflect the creation of output value that is not paid to factors. In both of our analyses, the railroads increase the scale of the US economy, but because we allow for the marginal value of product to exceed marginal costs, this increase in economic activity generates surplus or “profit” that is reflected in productivity growth rather than increases in land values. Further, our estimated impacts on productivity do not include any economic gains reflected in increased factor payments, and so our estimated impact on productivity (25%) is in addition to impacts on land value (3.2%).

A general implication for measuring economic incidence is that factor payments do not include all economic gains when there are market distortions. More inelastically supplied factors will bear more economic incidence, but there are additional economic gains that are not captured by factor payments. We show that these additional economic gains can be substantively large, particularly when new infrastructure investment or new technologies are broadly used and encourage broad expansion of economic activity.

The additional economic gains, from decreasing resources spent on transportation, could also be measured directly by calculating the decreases in transportation costs using the railroads instead of the waterways. This is precisely the social savings calculation in Fogel (1964), which implies that our estimated economic gains (25%) are in addition to Fogel’s estimate of 2.7% of GDP.

In considering why Fogel’s estimates do not include our estimated economic gains, we highlight the importance of resource misallocation in welfare analysis more generally. Fogel (1964) proposes a social savings calculation to bound the economic gains from the railroads. Fogel focuses on the transportation sector, and looks to calculate the additional cost from using waterways to transport goods instead of the railroads. This calculation is closely related to aggregate productivity in the transportation sector, measured as revenue minus costs: for transporting the same quantity of goods (fixing revenue), calculating the increase in costs without the railroads.<sup>78</sup>

David (1969) critiques Fogel’s calculations on several grounds, but much of this critique is essentially calling attention to Fogel’s implicit assumption that value marginal product is equal to marginal cost.<sup>79</sup> This assumption is required for the increase in transportation costs

---

<sup>78</sup>In practice, Fogel modifies this social savings calculation using data on land values, drawing on a similar intuition as in Donaldson and Hornbeck (2016) and other hedonic-style analyses, that land values capitalize all economic value created in an area.

<sup>79</sup>Allen and Arkolakis (2019) derive Fogel (1964)’s social savings calculation in general equilibrium and show how it can break down with departures from the neoclassical model (in their case, in the presence of

without the railroads to equal the value lost from decreased production in non-transportation sectors. David (1969) proposes that this assumption would be violated by increasing returns to scale, and Fogel (1979) responds by disputing the empirical magnitude of increasing returns to scale.<sup>80</sup> Fogel (1979) also makes this assumption more explicit: that in non-transportation sectors, firms' value marginal product is equal to their marginal cost.

Our analysis relaxes this assumption, and estimates the economic consequences from a broader range of market distortions, which restates the above critique by David (1969). Rather than appealing to increasing returns to scale, we allow for a wide variety of distortions that can drive a wedge between the social benefit and private cost of firms expanding production (e.g., firm markups, credit constraints, taxes and regulation, imperfect property rights).<sup>81</sup> The railroads decrease transportation costs, effectively subsidizing the expansion of economic activities throughout the economy that have a positive social return (i.e., whose value marginal product exceeds their marginal cost).

Fogel (1964, 1979) emphasizes that assuming an inelastic demand for transportation provides an upper bound estimate on the railroads' impacts, for the social savings calculation, but the opposite is true in the presence of market distortions. A greater elasticity of demand for transportation magnifies the economic impacts of the railroads by yielding greater changes in activities whose value marginal product exceeds marginal cost across other sectors in the economy. Fogel does not consider the indirect losses in other sectors due to reductions in transported goods, because he aims to calculate the costs of maintaining the same levels of transportation, but it is precisely because transportation would fall that there are such large indirect losses in other sectors.<sup>82</sup>

More generally, there is an analogous need to consider resource misallocation in partial equilibrium welfare analysis. Harberger (1964) lays the foundation for much welfare analysis in economics, using the example of calculating the economic cost of a tax, making a powerful assumption that there are no other distortions in the economy. This assumption means that it is not necessary to consider how a marginal tax affects other activities, which reflect only small welfare "triangles," and the welfare effects of the tax are largely captured by the

---

agglomeration economies).

<sup>80</sup>Fogel (1964) also notes that any increasing returns to scale may be internal to the firm, rather than at the industry level, such that changes in industry size would occur through changes in the number of firms rather than changes in output per firm. Indeed, we estimate that increases in county market access increase total output through increases in the number of firms rather than output per firm (Table 5).

<sup>81</sup>Increasing returns to scale can also be represented as gaps between value marginal product and marginal cost, as the output elasticities sum to a value greater than one even if the revenue shares sum to one.

<sup>82</sup>When Fogel does consider indirect impacts on the manufacturing sector, it is to consider ways in which the manufacturing sector was affected by the construction and maintenance of the railroads themselves. He argues that these effects were small (as did Fishlow 1965), and we also set aside impacts on manufacturing from supplying railroads.

demand curve for the taxed activity.<sup>83</sup> Harberger (1964) makes this assumption clear, and notes that it probably has the effect of understating the true cost of a tax, but this assumption is often overlooked in applications due to its substantial analytical convenience.<sup>84</sup> In Fogel’s application, when analyzing the impacts of a higher transportation cost (similar to a higher tax), the demand curve for transported goods is used to capture the welfare effects, and the mistake is to not consider impacts from resulting changes in other activities.

Our estimated impacts of the railroads are a reminder that indirect effects on other economic activities can generate substantial economic benefits, which are missed in partial equilibrium welfare analysis. For example, in the case of taxes that indirectly increase or reduce other activities, it is necessary to consider the changes in those other activities when they are otherwise above or below their social optimum. When there is resource misallocation, such as due to firm markups or capital constraints, and those other activities are under-provided then there are first-order welfare gains from their encouragement. Only in a special case, when there are no market distortions and other economic activities are efficient, can we invoke the envelope theorem and consider only the direct economic effects.

## VII Conclusion

We estimate that the railroads caused substantial aggregate 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, substantially increasing production in counties where the value marginal product of new inputs was greater than their marginal cost.

Infrastructure investment is particularly impactful when there are market distortions in the economy, such that when production increases it is in places where the value marginal product of inputs is greater than their marginal cost. We find that the railroads generated large indirect economic gains by increasing marginally productive activities across other sectors of the economy. These indirect economic gains, from encouraging the expansion of other economic activities, were substantially larger than direct gains from decreased resources spent on transportation itself. The railroads generated large indirect gains because they were so broadly used in the economy, and this same mechanism can apply to a variety of other new technologies that become widely used and encourage the expansion of other marginally productive activities.

Our counterfactual analysis does not include impacts of the railroads on technical ef-

---

<sup>83</sup>In a similar spirit, Hulten (1978) presents a general motivation for focusing only on within-sector productivity improvements.

<sup>84</sup>A recent literature on production networks (Baqae and Farhi, 2017; Liu, 2019) has used general input-output models to move beyond this assumption.

efficiency, or ways in which technological innovation might respond to increases in market access. Our reduced-form analysis finds no substantial impact of county market access on measured technical efficiency, but our data on revenues and expenditures might understate increases in technical efficiency.

We also do not consider a variety of other mechanisms through which railroads impacted the United States economy. Our analysis does not consider how the construction and operation of the railroads may have directly affected the economy, such as through increased demand for iron (Fogel, 1964; Fishlow, 1965) or the development of improved management practices (Chandler, 1965). The railroads likely encouraged certain economic activities to agglomerate in major urban centers, particularly wholesale trade, and these effects are perhaps captured best by more qualitative analyses of individual major cities (e.g., Cronon, 1991). Our empirical analysis is better suited to examining the experience of a broader range of counties, who were induced to grow by the railroads and increases in market access, which contributed to substantial aggregate productivity gains.

## References

- Abramovitz, Moses, and Paul A David.** 1973. “Reinterpreting economic growth: parables and realities.” *The 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.** 2014. “Trade and the Topography of the Spatial Economy.” *The Quarterly Journal of Economics*, 129(3): 1085–1140.
- Allen, Treb, and Costas Arkolakis.** 2019. “The welfare effects of transportation infrastructure improvements.”
- Asturias, Jose, Manuel García-Santana, and Roberto Ramos.** 2018. “Competition and the welfare gains from transportation infrastructure: Evidence from the Golden Quadrilateral of India.” *Journal of the European Economic Association*.
- Atack, Jeremy.** 1977. “Returns to scale in antebellum United States manufacturing.” *Explorations in Economic History*, 14(4): 337.
- Atack, Jeremy, and Fred Bateman.** 1999. “Nineteenth-century US industrial development through the eyes of the Census of Manufactures.” *Historical Methods*, 32(4): 177–188.
- Bai, Yan, Keyu Jin, and Dan Lu.** 2018. “Misallocation under trade liberalization.”
- Baqae, David, and Emmanuel Farhi.** 2019a. “A Short Note on Aggregating Productivity.”
- Baqae, David Rezza, and Emmanuel Farhi.** 2017. “The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten’s Theorem.” *NBER WP*, 23145.
- Baqae, David Rezza, and Emmanuel Farhi.** 2019b. “The macroeconomic impact of microeconomic shocks: beyond Hulten’s Theorem.” *Econometrica*, 87(4): 1155–1203.
- Basu, Susanto, and John G Fernald.** 2002. “Aggregate productivity and aggregate technology.” *European Economic Review*, 46(6): 963–991.
- Becker, Randy, Wayne Gray, and Jordan Marvakov.** 2013. “NBER-CES manufacturing industry database: Technical notes.” *NBER WP*.
- Berthou, Antoine, Jong Hyun Chung, Kalina Manova, and Charlotte Sandoz Dit Bragard.** 2018. “Productivity, (mis)allocation and trade.”

- Blattner, Laura, Luisa Farinha, and Francisca Rebelo.** 2019. “When losses turn into loans: the cost of undercapitalized banks.” *ECB WP*.
- Boberg-Fazlic, Nina, and Paul Richard Sharp.** 2019. “Immigrant Communities and Knowledge Spillovers: Danish-Americans and the Development of the Dairy Industry in the United States.” *CEPR Discussion Paper No. DP13757*.
- Caliendo, Lorenzo, and Fernando Parro.** 2015. “Estimates of the Trade and Welfare Effects of NAFTA.” *The Review of Economic Studies*, 82(1): 1–44.
- 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.
- Chung, Jong Hyun.** 2018. “Firm heterogeneity, misallocation, and trade.”
- Conley, Timothy G.** 1999. “GMM estimation with cross sectional dependence.” *Journal of econometrics*, 92(1): 1–45.
- Costa-Scottini, Lucas.** 2018. “Firm-level distortions, trade, and international productivity differences.”
- Costinot, Arnaud, and Andrés Rodríguez-Clare.** 2014. “Trade theory with numbers: Quantifying the consequences of globalization.” In *Handbook of international economics*. Vol. 4, 197–261. Elsevier.
- Crafts, Nicholas.** 2004. “Social savings as a measure of the contribution of a new technology to economic growth.”
- Cronon, William.** 1991. *Nature’s metropolis: Chicago and the Great West*.
- 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.” *The Economic Journal*, 71(284): 709–729.
- Donaldson, Dave, and Richard Hornbeck.** 2016. “Railroads and American economic growth: A market access approach.” *The Quarterly Journal of Economics*, 131(2): 799–858.

- Eaton, Jonathan, and Samuel Kortum.** 2002. “Technology, Geography and Trade.” *Econometrica*, 70(5): 1741–1779.
- Fajgelbaum, Pablo, and Stephen Redding.** 2018. “Trade, Structural Transformation and Development: Evidence from Argentina 1869-1914.” *NBER WP*, 20217.
- Fishlow, Albert.** 1965. *American Railroads and the Transformation of the Ante-bellum Economy*. Vol. 127, Harvard University Press Cambridge, MA.
- Fogel, Robert W.** 1964. *Railroads and American Economic Growth: Essays in Econometric History*. Baltimore: Johns Hopkins University Press.
- Fogel, Robert W.** 1979. “Notes on the Social Saving Controversy.” *Journal of Economic History*, 39: 1–54.
- 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.
- Gallman, Robert E.** 1960. “Commodity Output, 1839-1899.” In *Trends in the American economy in the nineteenth century*. 13–72. Princeton University Press.
- Gallman, Robert E, and Thomas J Weiss.** 1969. “The service industries in the nineteenth century.” In *Production and productivity in the service industries*. 287–381. NBER.
- Giroud, Xavier, and Joshua Rauh.** 2019. “State taxation and the reallocation of business activity: Evidence from establishment-level data.” *Journal of Political Economy*, 127(3): 1262–1316.
- Hacker, J. David.** 2013. “New estimates of census coverage in the United States, 1850–1930.” *Social Science History*, 37(1): 71–101.
- Haines, Michael R.** 2010. “Historical, Demographic, Economic, and Social Data: The United States, 1790–2002.” *ICPSR*, 2896.
- Hall, Robert.** 1988. “The relation between price and marginal cost in US industry.” *Journal of Political Economy*, 96(5): 921–947.
- Hall, Robert.** 1990. “Invariance properties of Solows productivity residual.” *Growth/Productivity/Unemployment: Essays to Celebrate Bob Solows Birthday*.
- Haltiwanger, John, Robert Kulick, and Chad Syverson.** 2018. “Misallocation measures: The distortion that ate the residual.”

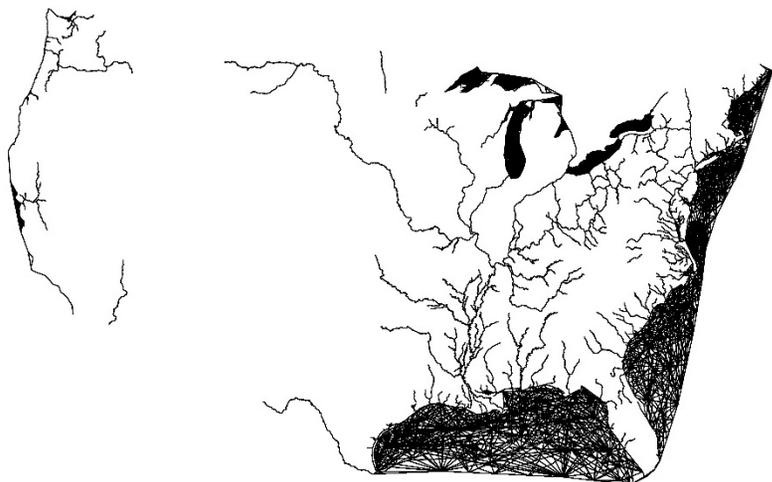
- Hanson, Gordon H.** 2005. "Market potential, increasing returns and geographic concentration." *Journal of International Economics*, 67(1): 1–24.
- Harberger, Arnold.** 1964. "Taxation, resource allocation, and welfare." In *The role of direct and indirect taxes in the Federal Reserve System*. 25–80. Princeton University Press.
- Head, Keith, and Thierry Mayer.** 2011. "Gravity, market potential and economic development." *Journal of Economic Geography*, 11(2): 281–294.
- Hornbeck, Richard.** 2010. "Barbed wire: Property rights and agricultural development." *The Quarterly Journal of Economics*, 125(2): 767–810.
- Hsieh, Chang-Tai, and Peter J Klenow.** 2009. "Misallocation and manufacturing TFP in China and India." *The Quarterly Journal of Economics*, 124(4): 1403–1448.
- Hulten, Charles R.** 1978. "Growth accounting with intermediate inputs." *The Review of Economic Studies*, 45(3): 511–518.
- Jaworski, Taylor, and Carl T. Kitchens.** 2016. "National Policy for Regional Development: Evidence from Appalachian Highways."
- Kaldor, Nicholas.** 1957. "A model of economic growth." *The economic journal*, 67(268): 591–624.
- 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.
- Leontief, Wassily W.** 1936. "Quantitative Input and Output Relations in the Economic Systems of the United States." *The Review of Economics and Statistics*, 18(3): 105–125.
- Leontief, Wassily W.** 1941. "The structure of American economy, 1919-1929: An empirical application of equilibrium analysis."
- Liu, Ernest.** 2019. "Industrial policies in production networks."
- 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." *The American Economic Review*, 104(2): 422–458.

- Petrin, Amil, and James Levinsohn.** 2012. “Measuring aggregate productivity growth using plant-level data.” *The RAND Journal of Economics*, 43(4): 705–725.
- 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.
- Redding, Stephen, and Esteban Rossi-Hansberg.** 2017. “Quantitative spatial economics.” *Annual Review of Economics*, 9: 21–58.
- Redding, Stephen, and Matthew Turner.** 2015. “Transportation costs and the spatial organization of economic activity.” In *Handbook of regional and urban economics*. Vol. 5, 1339–1398. Elsevier.
- Restuccia, Diego, and Richard Rogerson.** 2008. “Policy distortions and aggregate productivity with heterogeneous establishments.” *Review of Economic dynamics*, 11(4): 707–720.
- Rotemberg, Martin.** Forthcoming. “Equilibrium effects of firm subsidies.” *American Economic Review*.
- Singer, Gregor.** 2018. “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.” *The Review of Economics and Statistics*, 312–320.
- Sraer, David, and David Thesmar.** 2019. “A sufficient statistics approach for aggregating firm-level experiments.”
- Świecki, 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.
- U.S. Census.** 1860*a*. "Instructions to U.S. Marshals. Instructions to Assistants."
- U.S. Census.** 1860*b*. "Manufacturers of the United States in 1860."
- U.S. Census.** 1870. "Volume 3. The Statistics of Wealth and Industry of the United States."
- U.S. Census.** 1880. "Report on the Manufactures of the United States at the Tenth Census."
- Weeks, Joseph Dame.** 1886. *Report on the statistics of wages in manufacturing industries: with supplementary reports on the average retail prices of necessaries of life, and on trades societies, and strikes and lockouts*. Vol. 20, US GPO.
- Yang, Yang.** 2018. "Transport infrastructure, city productivity growth and sectoral reallocation: Evidence from China."
- 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

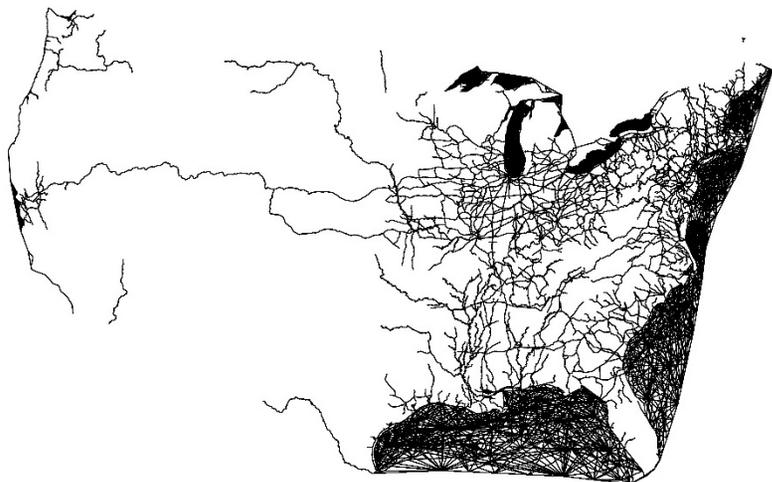
A. Waterways



B. Waterways and 1860 Railroads



C. Waterways and 1870 Railroads



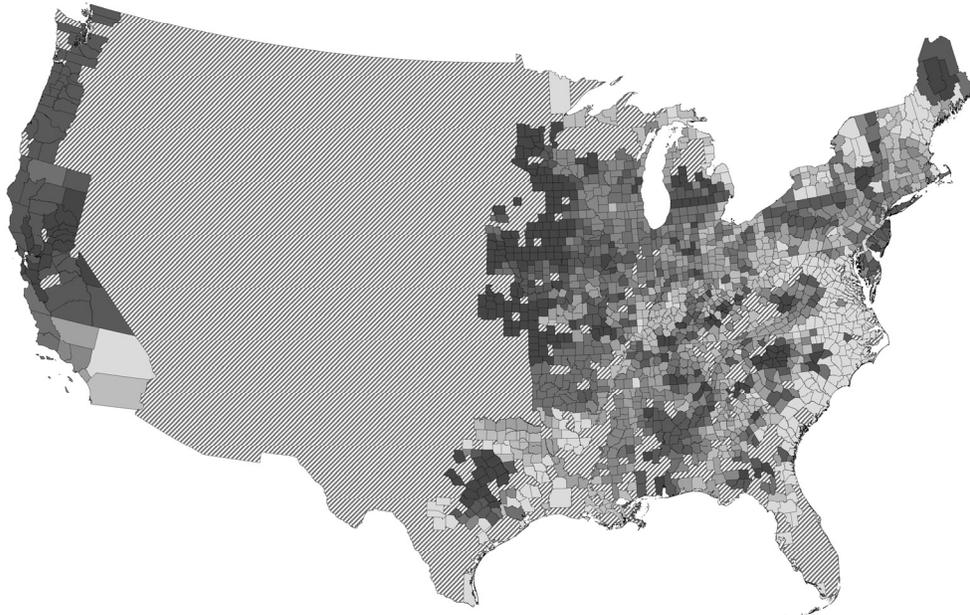
D. Waterways and 1880 Railroads



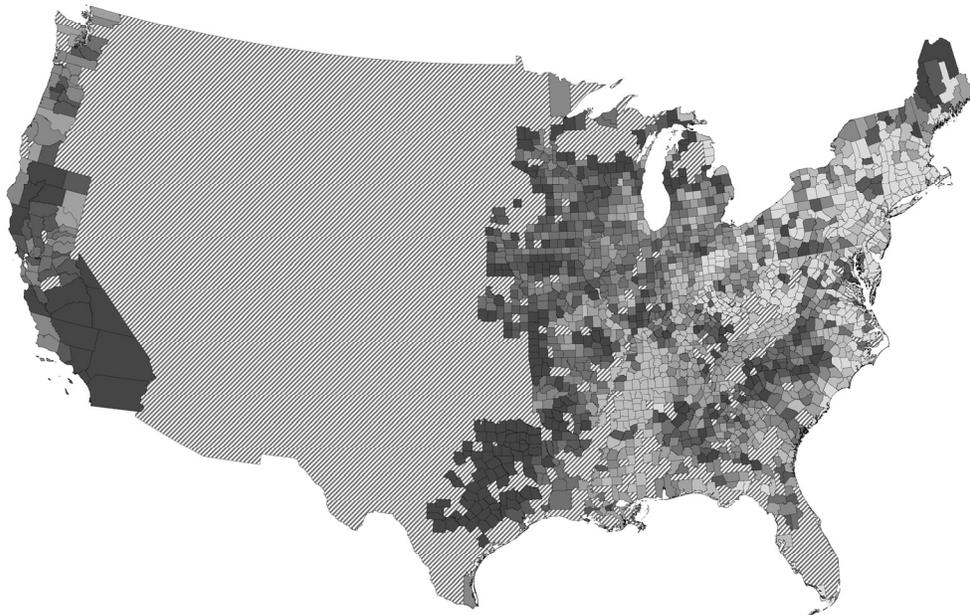
Notes: Panel A shows the waterway network: natural waterways (including navigable rivers, lakes, and oceans) and constructed canals. Panel B adds railroads constructed by 1860, Panel C adds railroads constructed between 1860 and 1870, and Panel D adds railroads constructed between 1870 and 1880.

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

A. From 1860 to 1870



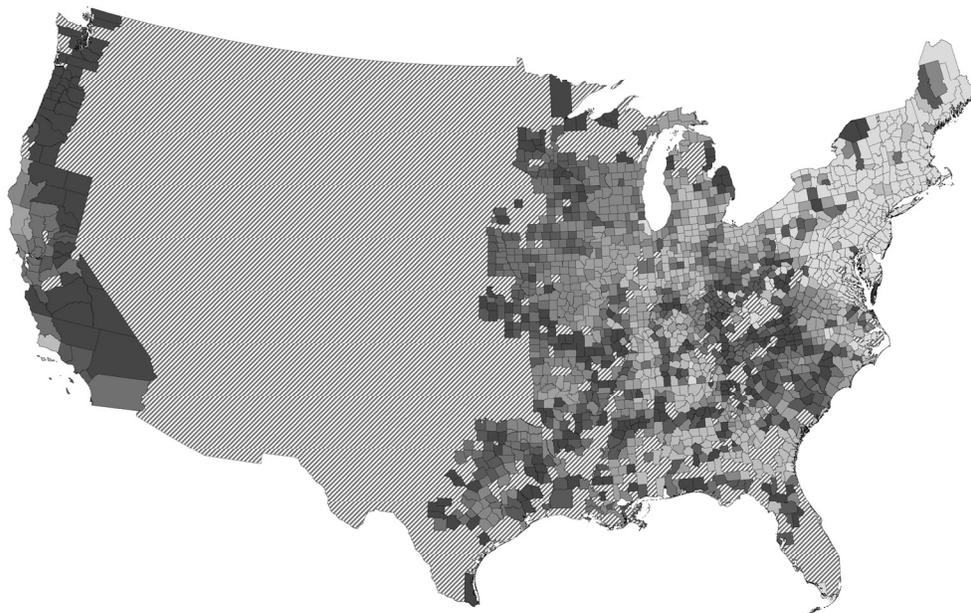
B. From 1870 to 1880



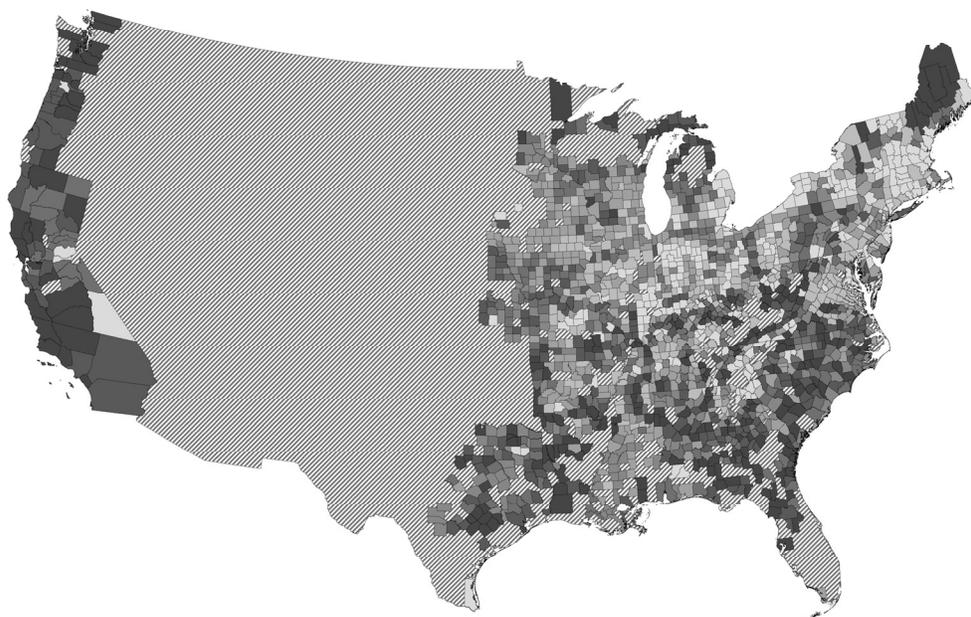
Notes: In each Panel, counties are shaded according to their calculated change in market access from 1860 to 1870 (Panel A) and from 1870 to 1880 (Panel B). Counties are divided into seven groups (with an equal number of counties per group), and darker shades denote larger increases in market access. These maps include the 1,804 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. Calculated Changes in Log Market Access, by County

A. From 1880 to 1890



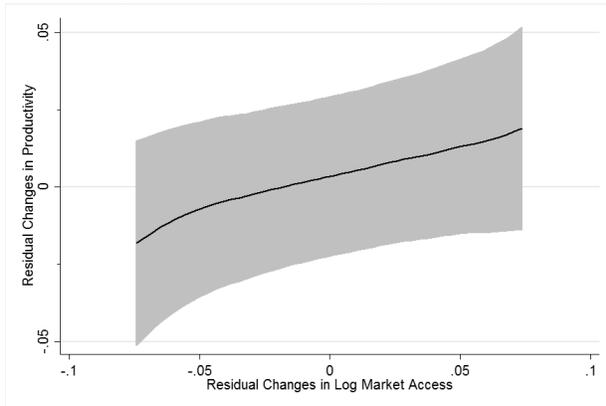
B. From 1890 to 1900



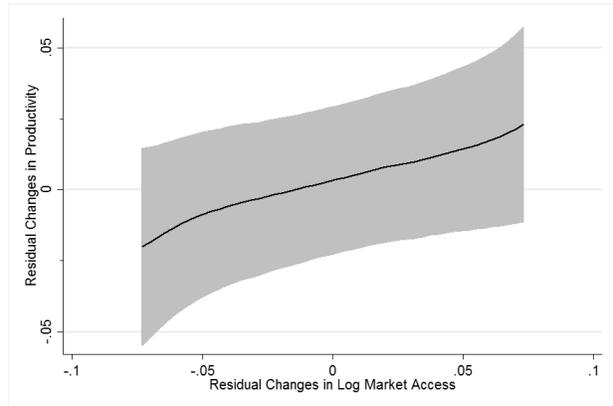
Notes: In each Panel, counties are shaded according to their calculated change in market access from 1880 to 1890 (Panel A) and from 1890 to 1900 (Panel B). Counties are divided into seven groups (with an equal number of counties per group), and darker shades denote larger increases in market access. These maps include the 1,804 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 4. Local Polynomial Relationships Between Productivity and Market Access, Using Approximated Market Access and Model-Defined Market Access**

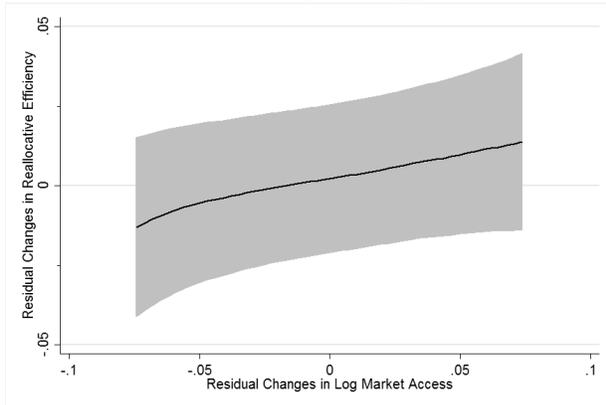
A. County Productivity and Approximated MA



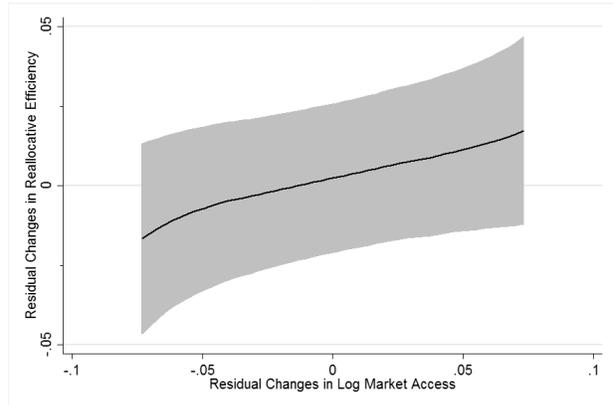
B. County Productivity and Model-Defined MA



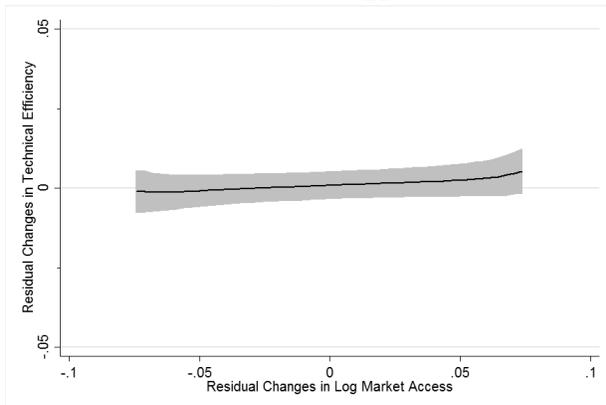
C. Reallocative Efficiency and Approximated MA



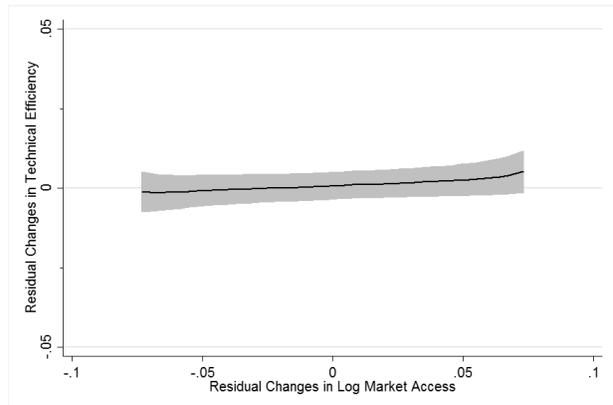
D. Reallocative Efficiency and Model-Defined MA



E. Technical Efficiency and Approximated MA

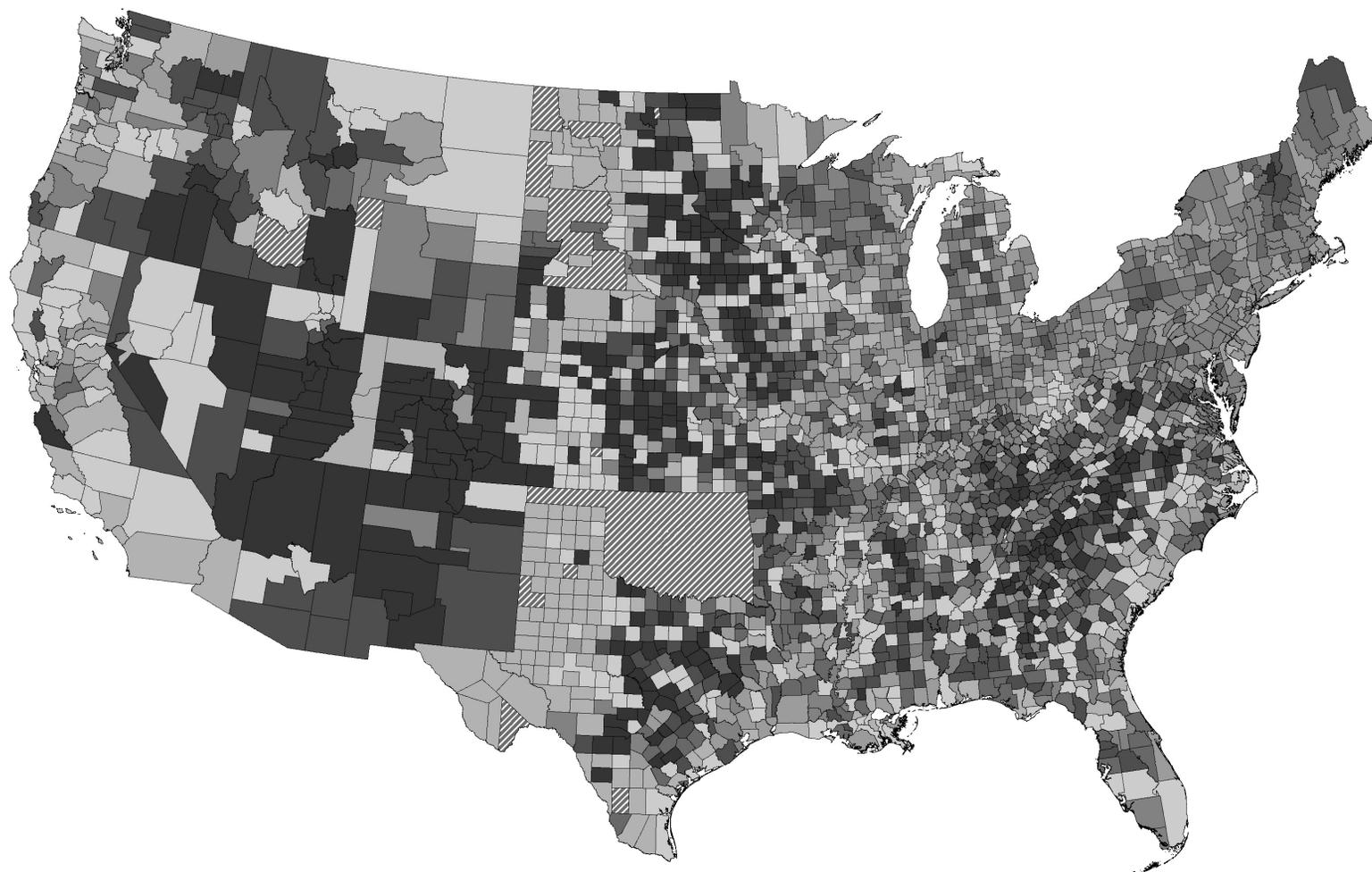


F. Technical Efficiency and Model-Defined MA



Notes: Each Panel plots the local polynomial relationship between residual productivity (y-axis) and residual market access (x-axis), where market access is based on our approximated measure (in Panels A, D, E) or based on our model-defined measure (in Panels B, C, and F). Residuals are calculated after controlling for county fixed effects, state-by-year fixed effects, and year-interacted cubic polynomial functions of county longitude and latitude. The local polynomial is based on an Epanechnikov kernel function with default bandwidth of 0.06. The shaded region reflects the 95% confidence interval.

**Figure 5. Counterfactual Changes in Productivity, by County**



Notes: This map shows the 2,760 counterfactual sample counties, which report population and output in 1890, shaded according to their change in productivity from 1890 to the baseline counterfactual scenario. Counties are divided into seven groups (with an equal number of counties per group), and darker shades denote larger declines in productivity. The excluded geographic areas are cross-hatched. County boundaries correspond to county boundaries in 1890.

**Table 1. Impacts on County Productivity, Technical Efficiency, and Reallocative Efficiency**

	Baseline Specification (1)	Fixed 1860 Population (2)	100-Mile Buffer Market Access (3)	Aggregate Data: 1860 to 1900    1860 to 1880 (4)                    (5)		Finest Detail Cost Shares (6)
<b>Panel A. County Productivity</b>						
Log Market Access	0.129 (0.050)	0.123 (0.048)	0.125 (0.048)	0.163 (0.041)	0.123 (0.051)	0.130 (0.049)
<b>Panel B. County Reallocative Efficiency (RE)</b>						
Log Market Access	0.117 (0.045)	0.111 (0.043)	0.113 (0.044)	0.160 (0.039)	0.118 (0.048)	0.112 (0.045)
<b>Panel C. County Technical Efficiency (TE)</b>						
Log Market Access	0.012 (0.008)	0.012 (0.008)	0.012 (0.008)	0.003 (0.005)	0.005 (0.007)	0.017 (0.007)
Number of Counties	1,804	1,804	1,804	1,804	1,804	1,804
County-Year Obs.	5,412	5,412	5,412	9,020	5,412	5,412

Notes: Column 1, panel A, reports estimates from equation 13 in the text: for a balanced panel of 1,804 counties in 1860, 1870, and 1880, county productivity (as defined in equation 10) is regressed on log market access (as defined in equation 1), county fixed effects, state-by-year fixed effects, and year-interacted cubic polynomials in county latitude and longitude. Panels B and C report estimated impacts on county reallocative efficiency and county technical efficiency, as defined in equations 12 and 11. 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 12.9% for counties with a one standard deviation greater change in market access from 1860 to 1880).

Column 2 reports estimates using a measure of counties' market access in each decade that holds counties' population levels fixed at 1860 levels. Column 4 uses a measure of counties' market access only to counties beyond 100 miles of a county. Columns 4 and 5 use county-level data, rather than county-by-industry data, to measure the outcome variables from 1860 through 1900 (in column 4) and from 1860 through 1880 (in column 5). Column 6 uses county-by-industry data based on 158 industry groups, rather than the 35 industry groups used in columns 1 to 3.

Robust standard errors clustered by state are reported in parentheses.

**Table 2. Impacts of Market Access, Controlling Flexibly for Local Railroad Construction**

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. County Productivity						
Log Market Access	0.129 (0.050)	0.147 (0.058)	0.152 (0.060)	0.140 (0.059)	0.131 (0.059)	0.105 (0.062)
Panel B. County Reallocative Efficiency (RE)						
Log Market Access	0.117 (0.045)	0.129 (0.053)	0.135 (0.055)	0.124 (0.055)	0.116 (0.054)	0.092 (0.057)
Panel C. County Technical Efficiency (TE)						
Log Market Access	0.012 (0.008)	0.018 (0.009)	0.017 (0.009)	0.016 (0.009)	0.015 (0.010)	0.013 (0.010)
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	1,804	1,804	1,804	1,804	1,804	1,804
County-Year Obs.	5,412	5,412	5,412	5,412	5,412	5,412

Notes: Column 1 reports the estimated impact of market access from the baseline specification (as in column 1 of Table 1). 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 polynomial 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. Robust standard errors clustered by state are reported in parentheses.

**Table 3. Impacts of Market Access, Instrumenting with Baseline Access through Waterways**

	Log Market Access		Productivity		Reallocative Efficiency		Technical Efficiency	
	OLS (1)	2SLS (2)	OLS (3)	2SLS (4)	OLS (5)	2SLS (6)	OLS (7)	
Log Market Access		0.281 (0.168)	0.129 (0.050)	0.248 (0.147)	0.117 (0.045)	0.033 (0.026)	0.012 (0.008)	
Instruments:								
Log Water Market Access in 1860 X year=1870	-0.240 (0.052)							
Log Water Market Access in 1860 X year=1880	-0.428 (0.060)							
Kleibergen-Paap F statistic		25.7		25.7		25.7		
Number of Counties	1,804	1,804	1,804	1,804	1,804	1,804	1,804	
County-Year Obs.	5,412	5,412	5,412	5,412	5,412	5,412	5,412	

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. Column 3 reports the baseline estimate for comparison (Table 1, panel A, column 1). Columns 4 and 5 report corresponding estimates for county reallocate efficiency, and columns 6 and 7 report corresponding estimates for county technical efficiency. For the 2SLS specifications, we report the Kleibergen-Paap F statistic from the first stage (in column 1).

All regressions include county fixed effects, state-by-year fixed effects, and year-interacted cubic polynomials 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. Robust standard errors clustered by state are reported in parentheses.

**Table 4. Impacts of Market Access on Input Expenditures, Gaps, Wedges, and Cost Shares**

	County Input Expenditure (1)	County Input Gap (2)	County Input Wedge (3)	County Input Cost Share (4)
Panel A. Materials				
Log Market Access	0.174 (0.051)	0.0093 (0.0056)	0.001 (0.037)	0.0007 (0.0025)
Panel B. Labor				
Log Market Access	0.197 (0.063)	-0.0020 (0.0045)	-0.057 (0.066)	-0.0008 (0.0023)
Panel C. Capital				
Log Market Access	0.159 (0.051)	0.0018 (0.0026)	0.034 (0.030)	0.0001 (0.0003)
Number of Counties	1,804	1,804	1,804	1,804
County/Year Obs.	5,412	5,412	5,412	5,412

Notes: For the indicated outcome variable, each column and panel reports the estimated impact of log market access from our baseline specification (Table 1, column 1): the indicated outcome variable is regressed on log market access, county fixed effects, state-by-year fixed effects, and year-interacted cubic polynomials in county latitude and longitude. In column 1, the outcome variable is log expenditure on materials (panel A), log expenditure on labor (panel B), and log expenditure on capital (panel C). In column 2, the outcome variable is the county-level input "gap," defined as that input's cost share minus its revenue share. In column 3, the outcome variable is the county-level input "wedge," defined as that input's cost share divided by its revenue share. In column 4, the outcome variable is the county-level cost share for that input, defined as the national industry-level cost shares multiplied by the share of county output in each industry.

The sample is our main balanced panel of 1,804 counties in 1860, 1870, and 1880. 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.

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

	Log Number of	Log Average Firm Size:		Log Number of	County Manufacturing Share of:			
	Industries	Output per Firm	Workers per Firm	Firms	Output	Value-Added	Surplus	Employment
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Market Access	0.005 (0.024)	0.025 (0.042)	0.025 (0.052)	0.172 (0.037)	0.0092 (0.0081)	0.0016 (0.0067)	0.0005 (0.0095)	0.0044 (0.0047)
Number of Counties	1,804	1,804	1,804	1,804	1,777	1,777	1,718	1,689
County/Year Obs.	5,412	5,412	5,412	5,412	5,331	5,331	5,154	5,067

Notes: For the indicated outcome variable, each column reports the estimated impact of log market access from our baseline specification (Table 1, column 1): the indicated outcome variable is regressed on log market access, county fixed effects, state-by-year fixed effects, and year-interacted cubic polynomials in county latitude and longitude. In column 1, the outcome variable is log number of manufacturing industries reporting positive output in the county. In columns 2 and 3, the outcome variables are log average manufacturing firm size in the county, based on output per firm (column 2) or workers per firm (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: output (column 5); value-added (column 6), which for manufacturing is defined as output minus materials expenditures and for agriculture is defined as 92% of output; surplus (column 7), which for manufacturing is defined as output 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 polynomials in county latitude and longitude. The samples are drawn from our main balanced panel of 1,804 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 1,804 counties. Robust standard errors clustered by state are reported in parentheses.

**Table 6. Impacts of Market Access on County Specialization**

	Output Shares (1)	Value-Added Shares (2)	Surplus Shares (3)	Employment Shares (4)
Panel A. Cross-Sector Specialization Index (Manufacturing vs. Agriculture)				
Log Market Access	-0.0122 (0.0113)	-0.0005 (0.0069)	-0.0047 (0.0121)	0.0013 (0.0052)
Number of Counties	1,777	1,777	1,718	1,689
County/Year Obs.	5,331	5,331	5,154	5,067
Panel B. Within-Manufacturing Specialization Index (Across Industries)				
Log Market Access	-0.0103 (0.0119)	-0.0467 (0.0404)	-0.0113 (0.0099)	-0.0016 (0.0111)
Number of Counties	1,804	1,804	1,804	1,804
County/Year Obs.	5,412	5,412	5,412	5,412

Notes: For the indicated outcome variable, each column and panel reports the estimated impact of log market access from our baseline specification (Table 1, column 1): the indicated outcome variable is regressed on log market access, county fixed effects, state-by-year fixed effects, and year-interacted cubic polynomials in county latitude and longitude. In panel A, the outcome variables reflect a cross-sector specialization index: the share of county value in manufacturing minus its national share (squared) plus the share of county value in agriculture minus its national share (squared), where those values are based on output (column 1), value-added (column 2), surplus (column 3), and employment (column 4) as defined in Table 5. In panel B, the outcome variables reflect a within-manufacturing specialization index: the share of county manufacturing value in each industry minus that industry's national manufacturing share (squared and summed across each industry), where the values for manufacturing are as defined in panel A.

All regressions include county fixed effects, state-by-year fixed effects, and year-interacted cubic polynomials in county latitude and longitude. The samples are drawn from our main balanced panel of 1,804 counties in 1860, 1870, and 1880, which are sometimes 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 1,804 counties. Robust standard errors clustered by state are reported in parentheses.

**Table 7. Counterfactual Impacts on Productivity**

	Baseline:	Restricted Railroad Networks:				No Railroads,	All Railroads,
	No Railroads	Only 1850 RRs	Only 1860 RRs	Only 1870 RRs	Only 1880 RRs	Extended Canals	Twice the Cost
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Counterfactual scenario, holding utility constant							
Change in Aggregate Productivity	-24.8%	-20.4%	-14.2%	-8.8%	-2.2%	-21.7%	-8.0%
Panel B. Counterfactual scenario, holding total population constant							
Change in Aggregate Productivity	-5.3%	-4.5%	-3.6%	-2.2%	-0.5%	-4.2%	-1.3%
Change in Utility	-33.6%	-27.9%	-18.8%	-11.6%	-2.9%	-29.8%	-11.4%

Notes: Each column reports the estimated change in 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 2,760 counties that report population data in 1890.

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.

# Appendices

## VIII Data Appendix

### VIII.A County-Industry Manufacturing Data

We have digitized manufacturing data, by county and industry, for 1860, 1870, and 1880 from the original published tabulations of the Census of Manufactures (U.S. Census, 1860*b*, 1870, 1880). The county-industry data report many industries in each decade, with some small variations, which we group together for our analysis. We homogenized industry names from each county to the list of industry names from US-industry tabulations in each decade: 331 names in 1880, 412 names in 1870, and 639 names in 1860. We then grouped these industries into 159 categories that were more consistent across decades, and further grouped these industries into 45 categories following Leontief (1941). Our estimates are not sensitive to these industry groupings (Table 1), but our goal was to balance industry-level details against statistical noise and to maintain comparability across decades and geographic areas.

These manufacturing data were collected by Census enumerators, who visited each manufacturing establishment to solicit responses. The 1860 Census instructions to enumerators discuss the data collection guidelines in useful detail, which we quote below, and there is similar language in the instructions for other decades.

Our main variables of interest, from the manufacturing data, are:

**Manufacturing Output.** Total value of products, by county and industry from 1860, 1870, and 1880. These products were valued at the factory gate, excluding transportation costs to customers: “In stating the value of the products, the value of the articles *at the place of manufacture* is to be given, exclusive of the cost of transportation to any market” (emphasis original, U.S. Census, 1860*a*).

**Manufacturing Materials.** Total value of materials, by county and industry from 1860, 1870, and 1880. These materials were valued at the factory gate, including transportation costs from suppliers: “this value is always to represent the cost of the article *at the place where it is used*” (emphasis original, U.S. Census, 1860*a*). Materials included fuel and “the articles used for the production of a manufacture,” which the instructions noted might be manufactured by another establishment. Unused materials (on June 1) were to be excluded.

**Manufacturing Wages.** Total amount paid in wages during the year, by county and industry from 1860, 1870, and 1880. Reported wages were intended to reflect total labor costs, including boarding costs paid in kind and the proprietor’s own labor. From the Census instructions: “In all cases when the employer boards the hands, the usual charge of board is to be added to the wages, so that *cost of labor* is always to mean the amount paid, whether in money or partly in money and partly in board...” (emphasis original) and to be included

was “the individual labor of a producer, working on his own account” (U.S. Census, 1860*a*).

**Manufacturing Capital Expenditure.** We impute annual capital expenditure by multiplying the reported total value of capital invested, in each county and industry (1860, 1870, 1880), by a state-specific interest rate from Fogel (1964). The firm’s capital value was directed to include “capital invested in real and personal estate in the business” (U.S. Census, 1860*a*).

**Manufacturing Establishment Counts.** The number of establishments in each county and industry (1860, 1870, 1880) with at least \$500 in annual sales. That is, the Census enumerators were instructed to survey every manufacturing establishment, except “household manufactures or small mechanical operations where the annual productions do not exceed five hundred dollars” (U.S. Census, 1860*a*). When multiple establishments were owned by the same party, and operated jointly, then Census enumerators were instructed to obtain separate detailed on the operations of each establishment. If this were impossible, particularly when one establishment manufactured the materials for the other establishment, then enumerators were instructed to “return the last manufacture, giving the raw materials for the first, and capital, fuel, and cost of labor, with the number of hands, in both” (U.S. Census, 1860*a*).

## **VIII.B Other County-Level Data**

For 1890 and 1900, when county-industry tabulations are unavailable, we use in some specifications the corresponding county-level data collected by (Haines, 2010).

Other county-level data are from the US Censuses of Population and Agriculture (Haines, 2010). Population is defined as the reported total population in each county. For one robustness check, we inflate these population data due to potential undercounting in the Census that varies by region and year: undercounting in the South by 7.6% in 1860, 8.8% in 1870, and 5.2% in 1880, and undercounting in the North by 5.6% in 1860, 6.0% in 1870, and 4.4% in 1880 (?). City population is defined as the county’s population living in cities of over 25,000 people, and urban population is defined as the county’s population living in urban areas of over 2,500 people. Agricultural land value is defined as the total value of land in farms, including the value of farm buildings and improvements.<sup>85</sup>

We adjust county-level data to maintain consistent county definitions in each decade. We adjust data from each decade to reflect county boundaries in 1890 following the procedure outlined by (Hornbeck, 2010). Using historical US county boundary files (from NHGIS), county borders in each decade are intersected with county borders in 1890. When counties in another decade fall within more than one 1890 county, data for each piece are calculated

---

<sup>85</sup>We follow Donaldson and Hornbeck (2016) in deflating these reported data, using Fogel’s state-level estimates of the value of agricultural land only (Fogel, 1964, pp. 82-83).

by multiplying that decade's county data by the share of its area in the 1890 county. For each other decade, each 1890 county is then assigned the sum of all pieces falling within its area. This procedure assumes that data are evenly distributed across county area, though for most counties in each decade there is little overlap with a second 1890 county. In three instances, we combine separately reported cities into a neighboring county for consistency: Baltimore City is combined into Baltimore County, St. Louis City is combined into St. Louis County, and Washington DC is combined into Montgomery County.

### VIII.C Sample Definition

For the regression analysis, the sample includes all 1,804 counties that report county-industry manufacturing data in 1860, 1870, and 1880. Figures 2 and 3 reflect this sample of counties included in the regression analysis. For the counterfactual analysis, note that we expand the sample to include 2,760 counties that report positive population and output in 1890, which is reflected in Figure 5.

## IX Theory Appendix

In this section, we provide some additional details on the model from Section V. These details relate to deriving the log-linear relationship between market access and productivity, and our estimation of counterfactuals.

### IX.A Market Access and Productivity

As described in equation 16, trade flows follow a gravity equation:

$$(25) \quad E_{od} = \kappa_1 A_o \left( \Pi_k \left( (1 + \psi_o^k) W_o^k \right)^{\alpha_o^k} \right)^{-\theta} \tau_{od}^{-\theta} Y_d P_d^\theta,$$

where the price index in equation 17 is:

$$(26) \quad CMA_d = P_d^{-\theta} = \kappa_1 \sum_o \tau_{od}^{-\theta} A_o \left( \Pi_k \left( (1 + \psi_o^k) W_o^k \right)^{\alpha_o^k} \right)^{-\theta}.$$

Goods markets clear in general equilibrium, whereby demand in each county is equal to supply. Production in each county is then equal to the sum of exports to all destinations (including itself). Summing equation 25 over all counties and taking logs gives:<sup>86</sup>

$$(27) \quad \ln(Y_o) = \kappa_1 + \varkappa_{1o} + (\alpha_{mo} + \alpha_{lo}) \ln(P_o^{-\theta}) - \theta \alpha_{to} \ln q_o + \ln \left( \sum_d \left( \frac{P_d}{\tau_{od}} \right)^\theta Y_d \right).$$

---

<sup>86</sup>  $\kappa_1 = \left( -\frac{\theta}{1-\sigma} \right) \ln \left( \Gamma \left( \frac{\theta+1-\sigma}{\theta} \right) \right)$  and  $\varkappa_{1o} = \ln(A_o) - \theta \alpha_{lo} \ln((1 + \psi_{lo}) \bar{U}) - \theta \alpha_{ko} \ln((1 + \psi_{ko}) r) - \theta \alpha_{mo} \ln(1 + \psi_{mo})$

Plugging  $P_d^{-\theta}$  from equation 26 into equation 27, and combining terms, gives:<sup>87</sup>

$$(28) \quad \ln Y_o = \kappa_1 + \varkappa_{2o} + \left( \frac{\alpha_{mo} + \alpha_{lo} + 1}{1 + \theta\alpha_{to}} \right) \ln (P_o^{-\theta}).$$

Once we have solved for output in each county, inputs in each county follow directly from our assumption that within-county revenue shares are constant.

For estimating changes in productivity, what matters is changes in real output and real inputs. Because markups are constant in this environment, we can convert from nominal output to real output using the changes in marginal costs. The price of capital is independent of market access. The price of land ( $q_o$ ) is endogenous to market access, as described in equation 24:

$$(29) \quad \frac{d \ln q_o}{d \ln MA_o} = \frac{\alpha_o^M + \alpha_o^L + 1}{1 + \theta\alpha_o^T}$$

The local prices for labor ( $w_o$ ) and materials ( $P_o$ ) are log-linear in market access, as described by equation 26:

$$(30) \quad \frac{d \ln w_o}{d \ln MA_o} = \frac{d \ln P_o}{d \ln MA_o} = -\frac{1}{\theta}.$$

The impact of market access on productivity, equation 22, is then given by the change in real output minus the revenue share weighted change in input quantities, and rearranging terms:

$$(31) \quad \begin{aligned} \frac{d \ln PR_o}{d \ln MA_o} = & (\alpha_o^L - s_o^L) \left( \frac{1}{\theta} + \frac{\alpha_o^M + \alpha_o^L + 1}{1 + \theta\alpha_o^T} \right) \\ & + (\alpha_o^M - s_o^M) \left( \frac{1}{\theta} + \frac{\alpha_o^M + \alpha_o^L + 1}{1 + \theta\alpha_o^T} \right) \\ & + (\alpha_o^K - s_o^K) \left( \frac{\alpha_o^M + \alpha_o^L + 1}{1 + \theta\alpha_o^T} \right). \end{aligned}$$

## IX.B Counterfactual Estimation and Uniqueness

To estimate counterfactuals, we solve the model in the cross section. Some of the primitives in the model, such as the production function elasticities, we estimate from our main county-by-industry data. Other primitives, such as the the trade elasticity ( $\theta$ ) and the average price of traded goods ( $\bar{P}$ ), we estimate using supplemental data.

Our general approach to estimating counterfactuals, following recent papers, is to ratio-

---

<sup>87</sup>  $\varkappa_{2o} = \frac{\varkappa_{1o} + \ln \xi_o - \theta\alpha_{to} \ln \frac{\alpha_{to}}{T_o}}{1 + \theta\alpha_{to}}$

nalize the observed distribution of population by estimating each county's productivity ( $A_o$ ) and quantity of fixed factors ( $T_o$ ). We start by noting that prices in any location can be expressed as a function of prices and population in all locations:

$$(32) \quad P_d^{-\theta} = \sum_o \frac{\tau_{od}^{-\theta} P_o \frac{(1+\psi_{lo})L_o}{\alpha_{lo}}}{\sum_i \tau_{oi}^{-\theta} P_i^{1+\theta} \frac{(1+\psi_{li})L_i}{\alpha_{li}}}.$$

This equation matches equation 15 in the appendix to Donaldson and Hornbeck (2016), with two changes: we allow for market distortions ( $\psi_{lo}$ ), and we allow variation across counties in the production function elasticity for labor ( $\alpha_{lo}$ ). As a result, many of the steps in our derivation match those in Donaldson and Hornbeck (2016), which in turn rely on results from Allen and Arkolakis (2014). Thus, we focus on describing where our new assumptions require a new approach. For example, the first step is the same: there is a steady state solution for prices (up to proportionality) that can be identified using the Fujimoto-Krause algorithm.

Having solved for prices, we then solve for the fixed characteristics of each county (i.e., its productivity and quantity of fixed factors). We define  $C_i$ , such that:<sup>88</sup>

$$(33) \quad C_i \equiv A_i T_i^{\theta \alpha_{ti}} r_i^{-\theta \alpha_{ki}} \alpha_{ti}^{-\theta \alpha_{ti}} \left[ \Gamma \left( \frac{\theta + 1 - \sigma}{\theta} \right) \right]^{-\frac{\theta}{1-\sigma}} (1 + \psi_{li})^{-\theta \alpha_{li}} (1 + \psi_{ki})^{-\theta \alpha_{ki}} (1 + \psi_{mi})^{-\theta \alpha_{mi}}.$$

We can then rewrite equation 16 as

$$(34) \quad C_o \sum_i \tau_{oi}^{-\theta} P_i^{1+\theta} \frac{(1 + \psi_{li}) L_i}{\alpha_{li}} = P_o^{1+\theta(\alpha_{lo}+\alpha_{to}+\alpha_{mo})} \left( \frac{(1 + \psi_{lo}) L_o}{\alpha_{lo}} \right)^{1+\theta \alpha_{to}} \bar{U}^{\theta(\alpha_{lo}+\alpha_{to})},$$

and re-write equation 17 as

$$(35) \quad P_o^{-\theta} = \sum_i C_i \bar{U}^{-\theta(\alpha_{li}+\alpha_{ti})} P_i^{-\theta(\alpha_{li}+\alpha_{ti}+\alpha_{mi})} \left( \frac{(1 + \psi_{li}) L_i}{\alpha_{li}} \right)^{-\theta \alpha_{ti}} \tau_{oi}^{-\theta}.$$

To solve this system above, we define:

$$(36) \quad \phi_o := \frac{(P_o^{-\theta} C_o)}{\left( \frac{(1+\psi_{lo})BL_o}{\alpha_{lo}} \right)^{1+\theta \alpha_{to}} \bar{U}^{\theta(\alpha_{lo}+\alpha_{to})} P_o^{1+\theta(\alpha_{lo}+\alpha_{to}+\alpha_{mo})}}.$$

We solve the model up to a proportional constant for labor, which we denote  $B$ , as described

---

<sup>88</sup>We will be able to solve for  $C_i$  using the factual data, our assumption on interest rates, and our estimated value of  $\theta$ .

below.

We now have the equations:

$$(37) \quad 1 = \sum_i \frac{C_o}{P_o^{1+\theta(\alpha_{lo}+\alpha_{to}+\alpha_{mo})} \left(\frac{(1+\psi_{lo})BL_o}{\alpha_{lo}}\right)^{1+\theta\alpha_{to}} \bar{U}^{\theta(\alpha_{lo}+\alpha_{to})}} \tau_{oi}^{-\theta} P_i^{1+\theta} \left(\frac{1+\psi_{li}}{\alpha_{li}}\right) BL_i$$

and

$$(38) \quad \phi_o = \sum_i \frac{C_o}{P_o^{1+\theta(\alpha_{lo}+\alpha_{to}+\alpha_{mo})} \left(\frac{(1+\psi_{lo})BL_o}{\alpha_{lo}}\right)^{1+\theta\alpha_{to}} \bar{U}^{\theta(\alpha_{lo}+\alpha_{to})}} \tau_{oi}^{-\theta} P_i^{1+\theta} \left(\frac{1+\psi_{li}}{\alpha_{li}}\right) BL_i \phi_i.$$

Following Allen and Arkolakis (2014), we use the Perron-Frobenius theorem. The Perron-Frobenius theorem says that for any matrix  $M$ , with all its elements positive, the equation  $\hat{\phi}\lambda = M\hat{\phi}$  has a unique eigenvalue  $\lambda$  and a unique (up to proportionality) eigenvector  $\hat{\phi}$  with all its elements positive. We define the matrix  $A$  with elements  $[A_{od}] = \left[ \frac{C_o}{P_o^{1+\theta(\alpha_{lo}+\alpha_{to}+\alpha_{mo})} L_o^{1+\theta\alpha_{to}} \tau_{od}^{-\theta} P_d^{1+\theta} \left(\frac{1+\psi_{ld}}{\alpha_{ld}}\right) L_d \right]$ . For any possible values of  $C_o$ ,  $P_o$ ,  $N_o$ , and  $\tau_{od}$ , the matrix  $A$  has all of its elements positive. Perron-Frobenius therefore implies that, for any values of  $P_o$ ,  $N_o$ ,  $\tau_{od}$ , the matrix  $A$  will have only one unique (up to proportionality) eigenvector with all elements positive. In other words, when our system of equations 37 and 38 holds, we have two eigenvectors. The two eigenvectors are  $\phi_o$  and 1, which therefore must be proportional since the eigenvector is unique (up to proportionality) and so  $\phi_o = \phi^{-1} * 1$  for some constant  $\phi^{-1}$ . Thus, we know that

$$(39) \quad P_o^{1+\theta(1+\alpha_{to}+\alpha_{lo}+\alpha_{mo})} = \phi L_o^{-(1+\theta\alpha_{to})} C_o.$$

We then solve for  $P_o$  in equation 39, and plug into equation 34 to get:

$$(40) \quad C_o \sum_i \tau_{oi}^{-\theta} \left( \phi \bar{U}^{-\theta(\alpha_{li}+\alpha_{ti})} \left(\frac{(1+\psi_{li})BL_i}{\alpha_{li}}\right)^{-(1+\theta\alpha_{ti})} C_i \right)^{\frac{1+\theta}{1+\theta(1+\alpha_{ti}+\alpha_{li}+\alpha_{mi})}} \frac{(1+\psi_{li})BL_i}{\alpha_{li}} = \left( \phi \bar{U}^{-\theta(\alpha_{lo}+\alpha_{to})} \left(\frac{(1+\psi_{lo})BL_o}{\alpha_{lo}}\right)^{-(1+\theta\alpha_{to})} C_o \right)^{\frac{1+\theta(\alpha_{lo}+\alpha_{to}+\alpha_{mo})}{1+\theta(1+\alpha_{lo}+\alpha_{to}+\alpha_{mo})}} \times \left(\frac{(1+\psi_{lo})BL_o}{\alpha_{lo}}\right)^{1+\theta\alpha_{to}} \bar{U}^{\theta(\alpha_{lo}+\alpha_{to})}$$

Rearranging and combining like terms, we get:

$$\begin{aligned}
& (\phi \bar{U}^{-\theta(\alpha_{lo} + \alpha_{to})})^{-\frac{1+\theta(\alpha_{lo} + \alpha_{to} + \alpha_{mo})}{1+\theta(1+\alpha_{lo} + \alpha_{to} + \alpha_{mo})}} \bar{U}^{-\theta(\alpha_{lo} + \alpha_{to})} C_o^{\frac{\theta}{1+\theta(1+\alpha_{lo} + \alpha_{to} + \alpha_{mo})}} * \\
& \sum_i \tau_{oi}^{-\theta} (\phi \bar{U}^{-\theta(\alpha_{li} + \alpha_{ti})} C_i)^{\frac{1+\theta}{1+\theta(1+\alpha_{ti} + \alpha_{li} + \alpha_{mi})}} \left( \frac{(1 + \psi_{li}) BL_i}{\alpha_{li}} \right)^{1 - \frac{(1+\theta\alpha_{ti})(1+\theta)}{1+\theta(1+\alpha_{ti} + \alpha_{li} + \alpha_{mi})}} \\
(41) \quad & = \left( \frac{(1 + \psi_{lo}) BL_o}{\alpha_{lo}} \right)^{(1+\theta\alpha_{to})} \left( 1 - \frac{1+\theta(\alpha_{lo} + \alpha_{to} + \alpha_{mo})}{1+\theta(1+\alpha_{lo} + \alpha_{to} + \alpha_{mo})} \right).
\end{aligned}$$

We know that there is a unique combination of  $\phi$  and  $B$  that solves equation 41. To find that solution, we grid-search over the parameter space. For each initial guess, we use the Fujimoto-Krause algorithm to solve for the distribution of population ( $L_i$ ) and we pick the parameters for which equation 41 holds. As in Donaldson and Hornbeck (2016), we do not separately identify  $A_o$  and  $T_o$ , but only their combined value is needed for estimating the counterfactuals.

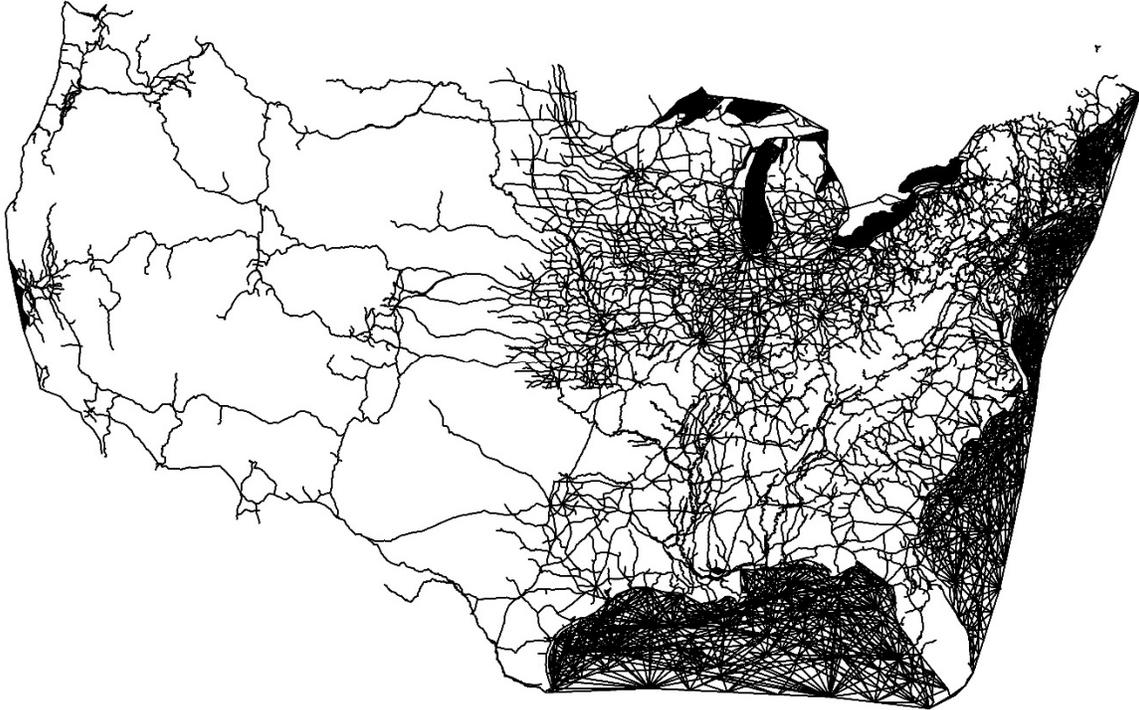
Given our estimated values of  $\theta$  and  $\bar{P}$ , we estimate how much production inputs would have changed in each county given a different vector of costs  $\tau_{od}$  (e.g., without the railroads), given a value of population (e.g., holding utility constant or holding population constant), as in Donaldson and Hornbeck (2016) and Fajgelbaum and Redding (2018).<sup>89</sup> We then calculate the associated productivity decline in each county from the estimated county-level gaps and the counterfactual declines in production inputs, and aggregate to the national productivity decline using the average of the factual Domar weights and counterfactual Domar weights.

---

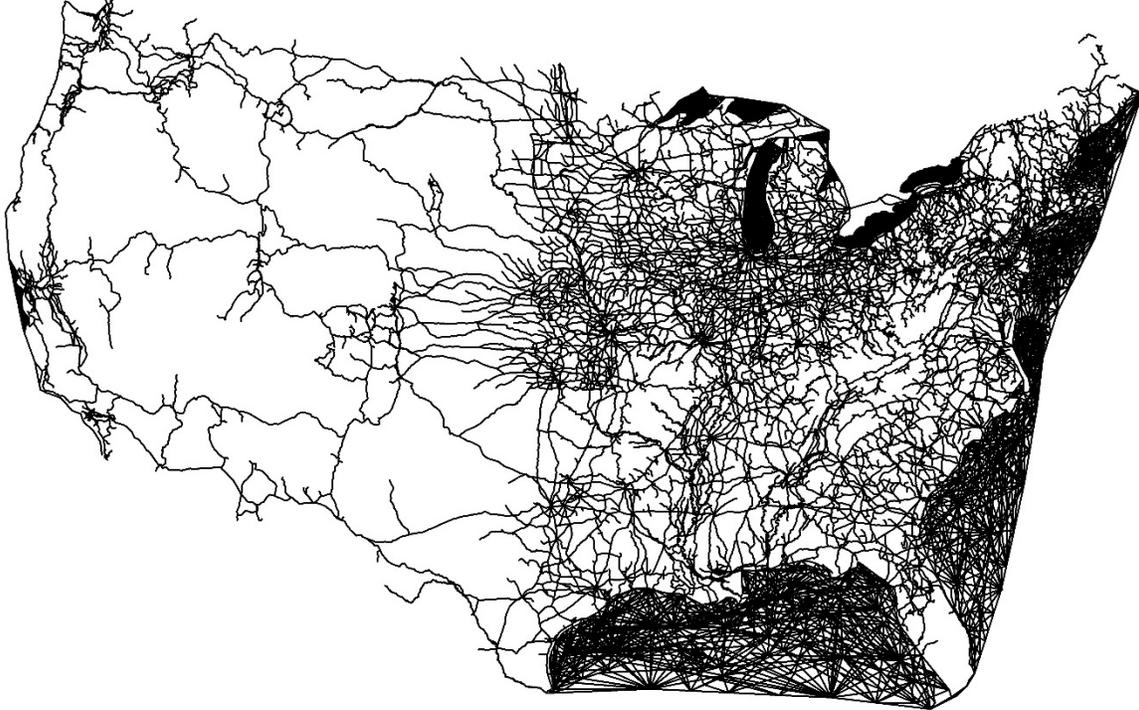
<sup>89</sup>Note that relative population levels are not independent of total population levels, unlike in Donaldson and Hornbeck (2016), because the production function elasticities vary over space but quantitatively this effect is relatively small.

**Appendix Figure 1. Waterways and Railroads, 1890 and 1900**

**A. Waterways and 1890 Railroads**



**B. Waterways and 1900 Railroads**



Notes: Similar to Figure 1, Panel A shows the railroads constructed by 1890, as well as the natural waterways (including navigable rivers, lakes, and oceans) and constructed canals. Panel B adds railroads constructed between 1890 and 1900.

**Appendix Table 1. Aggregated Industry Groups: Output Shares and Cost Shares, by Decade**

	Output Share (1)	Cost Shares:			(3)
		Materials (2)	Labor (3)	Capital (4)	(3) + (4) (5)
<b>Panel A. All Industries</b>					
Total, 1860	1	0.71	0.25	0.04	0.85
Total, 1870	1	0.73	0.23	0.04	0.85
Total, 1880	1	0.76	0.20	0.04	0.84
<b>Panel B. Aggregated Industry Groups</b>					
Clothing and Textiles, 1860	0.20	0.70	0.26	0.04	0.86
Clothing and Textiles, 1870	0.21	0.75	0.21	0.04	0.86
Clothing and Textiles, 1880	0.20	0.71	0.25	0.04	0.85
Food Products, 1860	0.20	0.91	0.06	0.03	0.71
Food Products, 1870	0.17	0.92	0.06	0.02	0.70
Food Products, 1880	0.22	0.93	0.05	0.02	0.71
Iron and Steel, 1860	0.11	0.55	0.39	0.06	0.86
Iron and Steel, 1870	0.16	0.65	0.31	0.05	0.87
Iron and Steel, 1880	0.13	0.68	0.27	0.05	0.84
Leather Products, 1860	0.10	0.68	0.30	0.03	0.91
Leather Products, 1870	0.09	0.76	0.22	0.02	0.90
Leather Products, 1880	0.09	0.78	0.19	0.02	0.90
Lumber and Timber Products, 1860	0.06	0.65	0.28	0.06	0.81
Lumber and Timber Products, 1870	0.06	0.70	0.24	0.05	0.82
Lumber and Timber Products, 1880	0.05	0.77	0.17	0.06	0.74
Chemicals, 1860	0.05	0.84	0.12	0.04	0.74
Chemicals, 1870	0.06	0.79	0.14	0.06	0.70
Chemicals, 1880	0.05	0.84	0.12	0.04	0.73
Beverages and Tobacco, 1860	0.05	0.80	0.16	0.04	0.80
Beverages and Tobacco, 1870	0.05	0.75	0.19	0.05	0.79
Beverages and Tobacco, 1880	0.05	0.76	0.19	0.05	0.79
Machinery and Electricity-related Products, 1860	0.03	0.51	0.43	0.06	0.88
Machinery and Electricity-related Products, 1870	0.04	0.52	0.43	0.05	0.89
Machinery and Electricity-related Products, 1880	0.05	0.58	0.37	0.05	0.88
Paper and Printing, 1860	0.03	0.65	0.30	0.05	0.85
Paper and Printing, 1870	0.04	0.65	0.30	0.05	0.86
Paper and Printing, 1880	0.04	0.61	0.34	0.05	0.87
Other Wood Products, 1860	0.03	0.55	0.41	0.04	0.91
Other Wood Products, 1870	0.03	0.59	0.37	0.04	0.90
Other Wood Products, 1880	0.03	0.65	0.31	0.04	0.89
Other Industries, 1860	0.14	0.58	0.37	0.05	0.88
Other Industries, 1870	0.10	0.61	0.35	0.05	0.88
Other Industries, 1880	0.09	0.63	0.33	0.04	0.89

Notes: Panel A reports aggregate statistics on manufacturing in the United States, by decade, from summing the county-by-industry data: annual expenditures on materials (column 2), labor (column 3), and capital (column 4) as a share of total annual expenditures; and, in column 5, annual expenditures on labor as a share of annual expenditures on labor and capital. Panel B reports these statistics for aggregated industry groups, along with that industry group's share of total output (column 1).

**Appendix Table 2. Impacts of Market Access, Robustness (Regional Shocks)**

	Estimated Impact of Market Access on:		
	Productivity	Reallocative Efficiency	Technical Efficiency
	(1)	(2)	(3)
1. Baseline Specification	0.129 (0.050)	0.117 (0.045)	0.012 (0.008)
2. Fixed Effects for 20 "resource regions"	0.153 (0.055)	0.138 (0.048)	0.015 (0.010)
3. Fixed Effects for 106 "resource subregions"	0.142 (0.079)	0.131 (0.068)	0.011 (0.014)
4. Fifth Order Polynomial	0.155 (0.046)	0.145 (0.041)	0.010 (0.009)
5. First Order Polynomial	0.109 (0.050)	0.104 (0.046)	0.005 (0.008)
6. Drop Western region (Plains and West Coast)	0.126 (0.060)	0.123 (0.052)	0.003 (0.012)
7. Drop Southern region	0.126 (0.055)	0.112 (0.049)	0.013 (0.010)
8. Drop Slave states	0.134 (0.089)	0.128 (0.081)	0.006 (0.015)
9. Drop Confederate states	0.136 (0.073)	0.124 (0.064)	0.012 (0.014)
10. Drop Counties with Civil War battles	0.116 (0.048)	0.105 (0.045)	0.011 (0.008)
11. Drop Counties with battles, > 500 casualties	0.137 (0.050)	0.125 (0.046)	0.011 (0.008)
12. Controls for Civil War battles and casualties	0.129 (0.050)	0.116 (0.045)	0.012 (0.008)

Notes: Row 1 reports our baseline estimates, from Table 1, for the impacts of market access on productivity (column 1), reallocative efficiency (column 2), and technical efficiency (column 3): for a balanced panel of 1,804 counties in 1860, 1870, and 1880, the indicated outcome variable is regressed on log market access, county fixed effects, state-by-year fixed effects, and year-interacted 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 in the full sample of 1,804 counties. Robust standard errors clustered by state are reported in parentheses.

Rows 2 to 12 report alternative estimates, which generally relate to controlling for other regional shocks. Row 2 controls for region-by-year fixed effects (20 regions), and row 3 controls for subregion-by-year fixed effects (106 subregions). Rows 4 and 5 modify the controls for county latitude and longitude to be a fifth-order polynomial or first-order polynomial, respectively. Row 6 excludes 201 counties in the Plains region and West Coast region of the sample, row 7 excludes 765 counties in the Southern region, row 8 excludes 980 counties in slave states, and row 9 excludes 745 counties in confederate states. Row 10 excludes 177 counties with recorded Civil War battles, and Row 11 excludes 99 counties with recorded Civil War battles that had more than 500 recorded casualties. Row 12 maintains the full balanced sample, but controls for whether a county had a Civil War battle, the number of battles (cubic polynomial), and the number of casualties (cubic polynomial), all interacted with year fixed effects. Note that each row represents separate modifications to our baseline specification, rather than cumulative changes.

**Appendix Table 3. Impacts of Market Access, Robustness (Measurement of Productivity)**

	Estimated Impact of Market Access on:		
	Productivity	Reallocative Efficiency	Technical Efficiency
	(1)	(2)	(3)
1. Baseline Specification	0.129 (0.050)	0.117 (0.045)	0.012 (0.008)
2. Dropping top/bottom centile, change in Productivity	0.152 (0.046)	0.140 (0.041)	0.012 (0.008)
3. Dropping top/bottom 5 centiles, change in Productivity	0.150 (0.044)	0.141 (0.040)	0.009 (0.007)
4. Inflating firm labor costs	0.126 (0.047)	0.112 (0.043)	0.014 (0.007)
5. Inflating firm capital costs	0.127 (0.062)	0.113 (0.059)	0.014 (0.008)
6. Decreasing returns to scale (0.90)	0.129 (0.050)	0.099 (0.045)	0.030 (0.009)
7. Decreasing returns to scale (0.95)	0.129 (0.050)	0.108 (0.045)	0.021 (0.008)
8. Increasing returns to scale (1.05)	0.129 (0.050)	0.126 (0.045)	0.003 (0.009)
9. Increasing returns to scale (1.10)	0.129 (0.050)	0.135 (0.045)	-0.006 (0.010)
10. Inflate labor cost share by 5 percentage points	0.129 (0.050)	0.118 (0.046)	0.011 (0.007)
11. Inflate materials cost share by 5 percentage points	0.129 (0.050)	0.116 (0.044)	0.013 (0.009)
12. Inflate capital cost share by 5 percentage points	0.129 (0.050)	0.116 (0.045)	0.014 (0.008)
13. Exclude butter and cheese industry	0.131 (0.049)	0.118 (0.045)	0.013 (0.008)

Notes: Row 1 reports our baseline estimates, from Table 1, for the impacts of market access on productivity (column 1), reallocative efficiency (column 2), and technical efficiency (column 3): for a balanced panel of 1,804 counties in 1860, 1870, and 1880, the indicated outcome variable is regressed on log market access, county fixed effects, state-by-year fixed effects, and year-interacted 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 in the full sample of 1,804 counties. Robust standard errors clustered by state are reported in parentheses.

Rows 2 to 13 report alternative estimates, which generally relate to adjusting our measurement of productivity. Rows 2 and 3 drop counties with the largest and smallest changes in productivity from 1860 to 1880: row 2 excludes the top and bottom 1% of counties, and row 3 excludes the top and bottom 5% of counties. Row 4 inflates firm labor costs, adding to county-by-industry labor costs the number of establishments multiplied by the average wage in that county and industry. Row 5 inflates firm capital costs, doubling the assumed interest rate to calculate annual capital expenditures. Rows 6 to 9 modify our baseline assumption of constant returns to scale, and re-scale the cost shares to add up to 0.90 (row 6), 0.95 (row 7), 1.05 (row 8), 1.10 (row 9). Rows 10 to 12 modify the relative cost shares for each factor, inflating by 5 percentage points the cost shares of labor (row 10), materials (row 11), and capital (row 12), and proportionally reducing the cost shares of the other factors. Row 13 excludes the butter and cheese industry from the analysis, for which coverage in the Census of Manufactures changes from 1860 to 1870.

**Appendix Table 4. Impacts of Market Access, Robustness (Measurement of Market Access)**

	Estimated Impact on Productivity from a:	
	One SD Increase in	1% Increase in
	Market Access	Market Access
	(1)	(2)
1. Baseline Specification	0.129 (0.049)	0.559 (0.215)
2. Dropping top/bottom centile, change in market access	0.130 (0.054)	0.565 (0.232)
3. Dropping top/bottom 5 centiles, change in market access	0.220 (0.080)	0.954 (0.346)
4. Reduces the cost of water to 0.198 cents per ton mile	0.144 (0.058)	0.632 (0.255)
5. Reduces the cost of wagons to 14 cents per ton mile	0.175 (0.065)	1.006 (0.375)
6. No transshipment costs between waterways	0.155 (0.060)	0.639 (0.248)
7. Include transshipment between Northern and Southern RRs	0.131 (0.050)	0.579 (0.219)
8. Raise railroad cost to 0.735 cents per ton mile	0.126 (0.049)	0.577 (0.224)
9. Raise railroad cost to 0.878 cents per ton mile	0.124 (0.049)	0.588 (0.233)
10. Average price of goods, $\bar{P}$ , set to 20	0.128 (0.048)	0.416 (0.156)
11. Average price of goods, $\bar{P}$ , set to 50	0.129 (0.050)	0.749 (0.292)
12. Trade elasticity, $\Theta$ , set to 1.81	0.129 (0.050)	0.849 (0.327)
13. Trade elasticity, $\Theta$ , set to 3.40	0.129 (0.050)	0.452 (0.174)
14. Trade elasticity, $\Theta$ , set to 8.22	0.132 (0.050)	0.190 (0.072)

15. Include access to international markets	0.128 (0.049)	0.570 (0.218)
16. Adjustment for Census undercounting	0.129 (0.050)	0.570 (0.219)
17. Measure access to county wealth	0.125 (0.048)	0.570 (0.216)
18. Include access to own market	0.129 (0.050)	0.571 (0.219)
19. Limit access to counties beyond 5 miles	0.129 (0.050)	0.572 (0.219)
20. Limit access to counties beyond 50 miles	0.128 (0.049)	0.571 (0.217)
21. Limit access to counties beyond 200 miles	0.116 (0.046)	0.543 (0.215)

Notes: Row 1 reports our baseline estimates, from Table 1, for the impacts of market access on productivity: for a balanced panel of 1,804 counties in 1860, 1870, and 1880, the indicated outcome variable is regressed on log market access, county fixed effects, state-by-year fixed effects, and year-interacted cubic polynomials in county latitude and longitude. In column 1, we report the estimated impact of a one standard deviation greater change in market access from 1860 to 1880 in the full sample of 1,804 counties. In column 2, we report the estimated impact of a 1% increase in market access. Robust standard errors clustered by state are reported in parentheses.

Rows 2 to 20 report alternative estimates, which generally relate to adjusting our measurement of market access. Rows 2 and 3 drop counties with the largest and smallest changes in market access from 1860 to 1880: row 2 excludes the top and bottom 1% of counties, and row 3 excludes the top and bottom 5% of counties. Row 4 reduces the cost of water transportation from 0.49 cents per ton mile to 0.198 cents per ton mile, and row 5 reduces the cost of wagon transportation from 23.1 cents per ton mile to 14 cents per ton mile. Row 6 removes transshipment costs (50 cents) when transferring goods within the waterway network. Row 7 adds transshipment costs between Northern and Southern railroads, and rows 8 and 9 raise to cost of railroad transportation (from 0.63 cents per ton mile to 0.735 cents or 0.878 cents) to reflect general congestion or indirect routes along the railroad network (as considered in Donaldson and Hornbeck 2016). Rows 10 and 11 replace our baseline estimated average price of transported goods (35.3) with alternative assumed values of 20 or 50. Rows 12 and 13 replace our baseline estimated trade elasticity (2.75) with alternative assumed values that reflect its estimated 95% confidence interval (1.81 to 3.40), and row 14 assumes a value of 8.22 from Donaldson and Hornbeck (2016). Row 15 adjusts our measurement of counties' market access to reflect access to international markets, inflating the population in counties with major international ports based on the value of imports and exports (scaled by GDP per capita). Row 16 adjusts counties' population for different under-enumerating rates in the Census of Population, by decade and region. Row 17 measures counties' market access based on their access to other counties' wealth, rather than other counties' population. Row 18 includes counties' own population in their market access, and Rows 19 to 21 measure counties' market access when excluding other counties within 5 miles, 50 miles, or 200 miles.

**Appendix Table 5. Measured Manufacturing Gaps, by Decade and Region**

	National	By Region:				
		Plains	West Coast	Midwest	Northeast	South
	(1)	(2)	(3)	(4)	(5)	(6)
<b>Panel A. Sum of Average Input Gaps</b>						
1860	0.22	0.35	0.26	0.20	0.22	0.20
	[0.09]	[0.19]	[0.13]	[0.09]	[0.05]	[0.11]
1870	0.20	0.24	0.22	0.22	0.17	0.22
	[0.08]	[0.10]	[0.14]	[0.08]	[0.07]	[0.09]
1880	0.15	0.15	0.15	0.15	0.16	0.16
	[0.08]	[0.08]	[0.03]	[0.05]	[0.09]	[0.07]
<b>Panel B. Average Materials Gap</b>						
1860	0.16	0.23	0.26	0.16	0.15	0.14
	[0.10]	[0.19]	[0.21]	[0.09]	[0.06]	[0.12]
1870	0.14	0.21	0.21	0.17	0.11	0.14
	[0.09]	[0.11]	[0.14]	[0.09]	[0.07]	[0.13]
1880	0.11	0.12	0.14	0.10	0.11	0.12
	[0.08]	[0.08]	[0.05]	[0.06]	[0.10]	[0.09]
<b>Panel C. Average Labor Gap</b>						
1860	0.05	0.11	0.00	0.04	0.06	0.07
	[0.07]	[0.11]	[0.16]	[0.07]	[0.04]	[0.08]
1870	0.05	0.04	0.02	0.05	0.05	0.08
	[0.05]	[0.09]	[0.07]	[0.05]	[0.03]	[0.08]
1880	0.04	0.02	0.01	0.03	0.04	0.04
	[0.04]	[0.06]	[0.03]	[0.04]	[0.04]	[0.07]
<b>Panel D. Average Capital Gap</b>						
1860	0.01	0.01	0.01	0.00	0.01	0.00
	[0.03]	[0.04]	[0.04]	[0.02]	[0.01]	[0.04]
1870	0.01	0.00	-0.01	0.00	0.01	0.00
	[0.02]	[0.02]	[0.03]	[0.01]	[0.01]	[0.03]
1880	0.01	0.00	0.00	0.01	0.01	0.00
	[0.02]	[0.05]	[0.02]	[0.02]	[0.01]	[0.03]

Notes: This Table reports measured gaps in the manufacturing sector, by decade, where the input gaps are equal to that input's cost share minus its revenue share. Column 1 reports these gaps at the national level, and columns 2 to 6 report these gaps by region, which weight county-level gaps by county output in that decade. Panel A reports the sum of these gaps across inputs, and panels B to D report gaps for materials, labor, and capital. Standard deviations are reported in brackets.

**Appendix Table 6. Measured Manufacturing Wedges, by Decade and Region**

	National (1)	By Region:				
		Plains (2)	West Coast (3)	Midwest (4)	Northeast (5)	South (6)
<b>Panel A. Average of Input Wedges</b>						
1860	1.42 [0.42]	2.09 [1.00]	2.00 [0.84]	1.37 [0.44]	1.37 [0.19]	1.41 [0.41]
1870	1.36 [0.40]	1.45 [0.50]	1.21 [0.26]	1.32 [0.27]	1.31 [0.16]	1.66 [0.85]
1880	1.29 [0.26]	1.31 [0.30]	1.13 [0.10]	1.30 [0.22]	1.29 [0.25]	1.37 [0.41]
<b>Panel B. Average Materials Wedge</b>						
1860	1.40 [0.82]	2.02 [1.86]	2.49 [2.31]	1.36 [0.73]	1.29 [0.22]	1.36 [0.77]
1870	1.26 [0.25]	1.44 [0.33]	1.53 [0.60]	1.32 [0.25]	1.20 [0.13]	1.29 [0.38]
1880	1.19 [0.13]	1.20 [0.17]	1.23 [0.11]	1.16 [0.10]	1.19 [0.12]	1.20 [0.21]
<b>Panel C. Average Labor Wedge</b>						
1860	1.44 [0.66]	2.33 [1.51]	1.69 [1.56]	1.42 [0.70]	1.36 [0.32]	1.55 [0.68]
1870	1.48 [1.01]	1.63 [1.29]	1.21 [0.45]	1.42 [0.63]	1.32 [0.28]	2.19 [2.29]
1880	1.29 [0.45]	1.37 [0.73]	1.09 [0.21]	1.26 [0.34]	1.28 [0.32]	1.54 [0.93]
<b>Panel D. Average Capital Wedge</b>						
1860	1.43 [0.61]	1.90 [1.24]	1.81 [1.21]	1.34 [0.64]	1.45 [0.38]	1.33 [0.70]
1870	1.34 [0.46]	1.28 [0.54]	0.89 [0.37]	1.22 [0.34]	1.40 [0.35]	1.50 [0.81]
1880	1.41 [0.55]	1.36 [0.50]	1.08 [0.24]	1.47 [0.48]	1.40 [0.59]	1.35 [0.65]

Notes: This Table reports measured wedges in the manufacturing sector, by decade, where the input wedges are equal to that input's cost share divided by its revenue share. Column 1 reports these wedges at the national level, and columns 2 to 6 report these wedges by region, which weight county-level wedges by county output in that decade. Panel A reports the unweighted average of these wedges across inputs, and panels B to D report wedges for materials, labor, and capital. Standard deviations are reported in brackets.

**Appendix Table 7. Counterfactual Impacts on Productivity, Robustness**

	Baseline Counterfactual (1)	Efficient Agriculture (2)	Trade Elasticity, $\Theta = 1.81$ (3)	Trade Elasticity, $\Theta = 3.40$ (4)	Trade Elasticity, $\Theta = 8.22$ (5)	Average Price, $\bar{P} = 20$ (6)	Average Price, $\bar{P} = 50$ (7)
Panel A. Fixed Utility							
Change in Aggregate Productivity	-24.8%	-15.4%	-25.6%	-24.3%	-21.5%	-32.3%	-20.4%
Panel B. Fixed Total Population							
Change in Aggregate Productivity	-5.3%	-4.1%	-5.4%	-5.3%	-4.2%	-7.2%	-4.3%
Change in Utility	-33.6%	-33.5%	-35.0%	-32.9%	-29.3%	-42.9%	-28.0%

Notes: Each column reports impacts under our baseline counterfactual scenario that removes all railroads in 1890, as in column 1 of Table 7. 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 2,760 counties that report population data in 1890.

Columns 2 to 7 report robustness of our baseline estimates (in column 1) under alternative parameters. In column 2, we reduce the estimated degree of resource misallocation 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. Our baseline estimates use our estimated value for  $\Theta$ , the trade elasticity, of 2.75 with a 95% confidence interval between 1.81 and 3.40. In columns 3 and 4, we alternatively impose values for  $\Theta$  of 1.81 or 3.40; in column 5, we impose a value of 8.22 from Donaldson and Hornbeck (2016). Our baseline estimates also use our estimated value for  $\bar{P}$  of 35.3, the average price of transported goods, which scales the assumed transportation costs into proportional costs. In Columns 5 and 6, we alternatively impose values for  $\bar{P}$  of 20 or 50.