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

Richard Hornbeck
University of Chicago

Martin Rotemberg
NYU

December 2021

Abstract

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

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

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

Using newly digitized data from the US Census of Manufactures, we measure counties' manufacturing revenue and costs for materials, labor, and capital. We define "county aggregate productivity" or "county productivity" as the aggregate surplus each county generates (county revenues minus county costs), which sums to national aggregate productivity and which Jorgenson and Griliches (1967) describe as the "conventional" definition of productivity. In our main estimates, we focus on growth in counties' revenues, costs, and productivity.

To understand the sources of growth in county productivity, we decompose county productivity growth into two components (following Petrin and Levinsohn (2012)): growth in county revenue total factor productivity or TFPR (increased revenues per input expenditures) and growth in county allocative efficiency or AE (the residual increases in productivity not driven by TFPR). TFPR growth, or the *rate* at which inputs become output, is equal to county aggregate productivity when markets are efficient (Solow, 1957). However, it is only part of what determines growth in county aggregate productivity when the value marginal product of inputs does not equal their marginal costs. In distorted economies, potentially due to firm markups (Hall, 1988) or input "frictions" such as credit constraints or insecure property rights (Hsieh and Klenow, 2009),[1] changes in input-use also can lead to changes in aggregate productivity.

We measure how changes in the national railroad network affected county "market access,"

[1]See Gavazza and Lizzeri 2021 for a recent survey. For example, firms with market power generally produce too little, such that further increases in input-use would increase revenues by more than input expenditures. If decreased transportation costs were to encourage those firms to expand production, increased input usage by those firms would then increase the total value of output by more than the total value of inputs (and thereby raise aggregate productivity). Shifting inputs from low-markup firms to high-markup firms would increase aggregate productivity, as would bringing new inputs into use by average-markup firms or even low-markup firms. Note that transportation costs themselves are not "frictions" that would generate a gap between the value marginal product of inputs and marginal cost; rather, transportation is a component of costs.

building on Donaldson and Hornbeck (2016), which captures how a county's manufacturing establishments were affected by the railroads changing establishments' access to consumers, workers, and material inputs. While local 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 identified from more-distant changes in the railroad network, and how the spreading railroad network complemented or substituted for the previous transportation network that relied on navigable waterways for low-cost freight transportation.

Increases in counties' market access led to substantial increases in county economic activity and, because economic activity in these counties was generally inefficiently low due to market distortions (e.g., firm markups or input frictions), this increase in economic activity generated substantial increases in county aggregate productivity. A one standard deviation greater increase in county market access, from 1860 to 1880, led to similar percent increases in county revenue (19%) and county expenditures on materials (18%), labor (20%), and capital (16%). This increase in county market access increased county aggregate productivity in manufacturing by 20%.

We decompose this increase in county aggregate productivity, finding little effect on counties' revenue total factor productivity (TFPR) and that the estimated increase in county productivity was driven by gains in allocative efficiency (AE). This is because increases in county market access increased input-use in counties where the value marginal product of inputs was greater than their marginal cost, on average, such that increases in input expenditures led to larger dollar increases in revenues than costs.[2] This growth in county productivity reflects the railroads both (1) shifting production inputs across counties and (2) increasing aggregate production inputs in the US economy, which we quantify in the aggregate counterfactual analysis. Cross-county differences in input distortions matter for (1), but for (2) the average level of input distortions also matters, and this second channel has been particularly under-emphasized in the literature relative to its quantitative importance in our setting.

Measured differences between marginal costs and marginal products may reflect real distortions as well as measurement error (Rotemberg and White, 2021). There are several challenges in measuring county-specific gaps in the decomposition of county productivity

[2] Our estimated impacts on county productivity do not assume a particular production function, but the decomposition of county productivity growth into TFPR growth and AE growth does depend on the production function and, in particular, the output elasticities of inputs. Our benchmark decomposition assumes Cobb-Douglas production and constant returns to scale, and we report the estimates' robustness to alternative methods of estimating these output elasticities for the decomposition.

growth, particularly in the measurement of capital expenditures, which may generate spurious measurement of misallocation.[3] We estimate that growth in county allocative efficiency is predominately driven by growth in materials inputs, however, which are more easily measured. We also undertake a variety of empirical exercises to assess the role of measurement error in our results, including: estimating production functions at different levels of aggregation, imposing that capital expenditure is efficient, or shrinking the cross-county dispersion in gaps. Across these specifications, we continue to find that estimated increases in county productivity are largely driven by increases in county allocative efficiency (AE) rather than increases in county revenue total factor productivity (TFPR).[4]

Increases in county market access led to a general expansion of county economic activity, rather than systematic changes in local manufacturing industry concentration or a shift from agriculture to manufacturing. Similarly, we do not find that increases in county market access directly affected county gaps between the value marginal product of inputs and their marginal costs.

Our baseline empirical specifications estimate relative growth in county aggregate productivity from relative increases in county market access, comparing counties that experience differential growth in market access. These estimated relative effects are not sufficient to estimate how the railroads affected national aggregate productivity, however, because an expanding railroad network (1) shifted production inputs between counties and (2) increased aggregate production inputs in the United States.

To quantify impacts of the railroads on national aggregate productivity, we extend a benchmark quantitative spatial model (Eaton and Kortum, 2002; Donaldson and Hornbeck, 2016) to include market distortions that drive a wedge between firms' value marginal product of inputs and their marginal cost. We use our estimates from the manufacturing sector to discipline the key parameters of the model, including county-level wedges and the trade elasticity. In the model, as in the data, changes in county market access do not affect county-level gaps between marginal products and marginal costs.

Holding fixed the total US population in 1890, we estimate that national aggregate productivity would have been lower by 5.8% in 1890 in the absence of the railroads. This aggregate productivity impact of the railroads is in addition to aggregate impacts capital-

[3]For instance, firms make forward-looking investment decisions (Solomon, 1970; Fisher and McGowan, 1983; Fisher, 1987; Caplin and Leahy, 2010), such that apparent market distortions can reflect dynamically efficient input decisions (Asker, Collard-Wexler and De Loecker, 2014).

[4]Intuitively, given that county revenues and county input expenditures change by similar percent amounts in response to increases in county market access, there is little scope for increases in county TFPR under constant returns to scale. Increases in county productivity could be driven by increases in county TFPR if there were sufficiently low returns to scale, but these would be below even low estimates of returns to scale in the empirical literature.

ized in land values (Donaldson and Hornbeck 2016, who find 3.2% losses) or differences in transportation costs (Fogel 1964, who finds 2.7% losses), so this estimated aggregate economic gain from the railroads is roughly triple that of previous estimates.[5]

We estimate larger aggregate economic impacts of the railroads because we allow for changes in county input-use to affect county productivity, due to county-level distortions in input-use, rather than assuming zero input distortions.[6] The importance of distortions is not necessarily context specific, as the estimated distortions in input-use are not larger in our historical data, on average, than in data for the modern United States.

This 5.8% impact on national aggregate productivity reflects changes in the allocation of economic activity across counties without the railroads, holding fixed the total US population in 1890, and we also estimate that worker utility (real wages) declines by 34% in this counterfactual scenario where total US population remains at its 1890 level without the railroads. As an alternative counterfactual assumption, given the substantial immigration to the United States in the 19th century, we hold fixed worker utility and allow the total US population to be lower in 1890 without the railroad network. We estimate that US population would have been 68% lower, similar to the 58% population decline estimated by Donaldson and Hornbeck (2016). This change in population has greater economic consequences in our framework, however, because workers and other production inputs were, on average, paid less than their marginal product.

Allowing for population decline in the US, we estimate that national aggregate productivity would have been 28% lower in 1890, with an associated annual loss of $3.4 billion (28% of GDP). We also estimate substantial national aggregate productivity losses without the railroads (17.4%) when assuming that the agricultural sector is efficient.

The railroads had a central role in enabling the substantial growth of the US economy, and would not have been easily replaced. We estimate a 50% annual social rate of return on the $8 billion of capital invested in the railroads in 1890 (in 1890 dollars), and estimate that the railroads in 1890 privately captured 7% of this social return. Additional canals might have been constructed in the absence of the railroads (Fogel, 1964), but we estimate that replacing the railroad network with this extended canal network would have mitigated only 12% of the aggregate losses from removing the railroad network. If construction of the railroad

[5] For comparability to those estimates, we report the aggregate economic gains relative to Fogel's measure of GNP in 1890 ($12 billion).

[6] Fogel (1964) and Donaldson and Hornbeck (2016) begin their analysis with the agricultural sector, whereas we begin with the manufacturing sector, but our analyses then consider implications for the broader economy. The key conceptual difference in our approaches is that we allow for distortions in input-use, as we measure in the manufacturing sector, whereas Fogel (1964) and Donaldson and Hornbeck (2016) assume that input-use is efficient (such that in all counties the value marginal product of inputs is equal to their marginal cost).

network had stopped in 1860, as a more moderate counterfactual scenario, national aggregate productivity would still have been 16% lower in 1890 (equivalent to roughly 30 years of lost technological innovation in this era). Across our counterfactuals, we find aggregate gains substantially higher than the benchmark measured gains from trade (Arkolakis, Costinot and Rodríguez-Clare, 2012), due to impacts through allocative efficiency.

Our paper extends a literature on estimating the impacts of market access to highlight the quantitative importance of market distortions (Redding and Venables, 2004; Hanson, 2005; Redding and Sturm, 2007; Head and Mayer, 2011; Donaldson and Hornbeck, 2016; Yang, 2018; Balboni, 2019; Jaworski and Kitchens, 2019).[7] We find that input distortions create a quantitatively important additional channel through which increases in market access can generate economic gains (or losses, in principle). In doing so, our work relates to a literature that considers how the efficiency of resource allocation is affected by policies such as trade liberalization, financial regulations, and taxes.[8] By bringing this research on resource misallocation into a model of economic geography, we can explore both (1) the spatial allocation of economic activity and (2) how production expanded to use new resources and attract additional workers from outside the United States, increasing aggregate inputs in the US economy. Our estimated increases in production are associated with substantial increases in the number of establishments, with little change in average establishment size, which relates to a literature highlighting the role of entry in aggregate productivity growth (Foster, Haltiwanger and Krizan, 2001; Foster, Haltiwanger and Syverson, 2008).

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).[9] A variety of market distortions can drive a wedge between firms' value marginal product and their marginal cost. By shifting resources across counties, the railroads affect aggregate productivity when there are heterogeneous distortions across space. Further, when there are average distortions across

[7] Redding and Turner (2015) and Redding and Rossi-Hansberg (2017) review this literature.

[8] See, for example, papers on trade liberalization (Khandelwal, Schott and Wei, 2013; Święcki, 2017; Bai, Jin and Lu, 2020; Berthou et al., 2020; Singer, 2019; Tombe and Zhu, 2019) and Sraer and Thesmar (2019) for a broader review of this literature. Asturias, García-Santana and Ramos (2019) calibrate a model to consider how highway construction in India affects prices and the allocation of fixed aggregate inputs between firms with different markups. Firth (2019) and Zárate (2021) study the role of congestion in allocating factors to firms with potentially different marginal products. In contrast to previous work on resource misallocation, which generally holds aggregate inputs fixed and considers the gains or losses from their reallocation, an important feature of our analysis is how the railroads encouraged growth in aggregate inputs in the economy. By drawing on the economic geography literature, we can consider changes in aggregate inputs along with the reallocation of fixed inputs, which is central to understanding the historical experience of the railroads and why the railroads had such large impacts on the development of the US economy.

[9] See Baqaee and Farhi (2019*b*) and Liu (2019) for reviews of this literature, along with discussion of misallocation in environments with more complicated input-output linkages.

all counties that reduce inputs below their efficient levels, the railroads affect aggregate productivity by increasing aggregate inputs in the US economy. We draw on a framework that generalizes changes in productivity (Hulten, 1978; Petrin and Levinsohn, 2012; Baqaee and Farhi, 2020), which allows for changes in aggregate productivity from increased input-use without changes to the production technology itself or changes in input distortions. We then show the substantive historical importance of an additional channel through which market integration can increase aggregate productivity: when the average marginal product is higher than the average marginal cost, aggregate increases in input-use can substantially increase aggregate productivity (raising the value of output by more than the cost of inputs).

Our paper highlights an important limitation underlying a long tradition in economics, back to at least Harberger (1964), of simplifying welfare analysis by assuming there are no distortions in secondary sectors or locations.[10] There is a persistent appeal to economic analysis that assigns value to some technology based on the cost of accommodating its absence. We highlight a problem with this intuition when marginal activities have positive social returns, and those activities would decline in the absence of the technology. Measured impacts on land values in the tradition of hedonic analyses, as in Donaldson and Hornbeck (2016), can similarly understate economic impacts dramatically because substantial economic surplus may not be paid out to land (or other factors).

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

I Data Construction

I.A Data on Manufacturing

We use data from the US Census of Manufactures (CMF), which published county-level totals for 1860, 1870, 1880, 1890, and 1900 (Haines, 2010). We digitized county-by-industry

[10]Fogel (1964) implicitly adopts this assumption in the "social savings" approach to calculating the economic gains from the railroads, and much of the critique by David (1969), in "Transport Innovation and Economic Growth: Professor Fogel on and off the Rails," can be seen as calling attention to Fogel's implicit assumption that the social value marginal product of inputs is equal to their marginal cost throughout the economy. David (1969) focuses on increasing returns to scale, which would violate this assumption, and Fogel (1979) responds by making this assumption more explicit and disputing that there are increasing returns to scale at the economy-level. Crafts (2004) derives the Fogel upper-bound in partial equilibrium. Allen and Arkolakis (2020) derive Fogel'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).

totals, published from 1860 to 1880 (see Appendix A for more information on the data construction).[11]

These manufacturing data include total annual revenue, total cost of raw materials, total wages paid, and total value of capital invested.[12] Revenues and materials costs are the most easily measured variables, and least subject to adjustment costs and other dynamic considerations (De Loecker and Warzynski, 2012).[13] The measurement of labor costs raises more challenges, particularly in the treatment of owner-operator labor (Weeks, 1886), and we show the robustness of our results to inflating measured labor costs to account for potential under-measurement of owners' labor.

The measurement of capital is the most challenging. The 1880 Census of Manufactures describes some difficulties with measuring capital, including distinguishing between nominal and resale values and potential non-reporting of rented land and equipment, since the question specifically asked for the value of "capital invested in real and personal estate in the business" (United States Census Bureau, 1860*a*).[14] Our baseline approach calculates annual capital expenditures by multiplying the total value of capital invested by a state-specific mortgage interest rates that varies between 5.5% and 11.4% , with an average value of 8% (Fogel, 1964).[15]

We undertake a variety of alternative approaches to understand the importance of capital measurement error, described in Appendix B, including: inflating capital expenditures, assuming capital is allocated more efficiently than is measured in the data, or assuming that actual capital misallocation is the same as misallocation for materials.[16] The empirical re-

[11] Census enumerators were directed to collect 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, including county-by-industry cells that contain only one manufacturing establishment (in 1860, 1870, 1880).

[12] For 1850, only some of the variables we need were aggregated and published (revenue and capital), although the original establishment-level manuscripts mostly exist. Prior to 1850, there are greater concerns about the comprehensiveness of the data collection and the Census data collection was professionalized in 1850 (Atack and Bateman, 1999).

[13] Revenues and materials costs were intended to reflect "factory-gate" prices, based on Census instructions to enumerators: transportation costs were to be included in establishment expenditures on materials, whereas revenue received by the manufacturing establishment would 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 the natural choice for tracking manufacturing productivity.

[14] Rented-in capital has only been irregularly collected in the modern Annual Survey of Manufactures, but Cunningham et al. 2021 report that it is a small share of total capital in years when measured.

[15] The mortgage interest rates are similar to the antebellum returns to equity collected by Bodenhorn and Rockoff (1992). They are also around the implied interest rate currently used by the BLS to convert capital stocks to the flow value of capital services, when only considering assets that existed in the 19th century such as buildings, land, and steam equipment (Cunningham et al., 2021).

[16] The goal of this latter exercise is both to take advantage of the relative ease in measuring materials expenditures and, because materials expenditures are flexible across time periods, deviations between ma-

sults are not sensitive to these alternative approaches, however, in part because the annual cost of capital is substantially smaller than labor and materials expenditures and because the estimated percent impacts on capital expenditures are similar to the estimated percent impacts on labor and material expenditures.

The empirical analysis is generally at the county-level, given county-level variation in market access and the absence of a balanced panel at the county-by-industry level.[17] For later estimates that decompose county aggregate productivity growth, we need measures of county production function elasticities (i.e., the expected increase in output from increases in inputs). We use the county-industry data to estimate production functions at the industry level: we assume that production is Cobb-Douglas, with constant returns to scale, whereby cost minimization implies that each input's output elasticity is equal to its cost share.[18] Our main analysis groups the reported industries into 31 industry categories, though we also calculate production function elasticities using 193 more-detailed industry categories.[19] We then calculate an aggregate county production function by computing the revenue-share weighted average of the cost shares of the industries in the county.[20]

Our baseline estimated impacts on county aggregate productivity do not require assumptions on the production function. For the estimates that decompose county productivity growth into changes in county TFPR and changes in county allocative efficiency, Appendix B discusses alternative methods for calculating production function elasticities (e.g., using only data from counties with relatively low measured frictions, computing the weighted average using county cost shares instead of revenue shares, alternative assumptions on the returns to scale, etc).

Appendix Table 1 reports information on 5 large industry groups, aggregating the reported industries 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 are mostly stable over time and vary more across industries.

In supplemental analysis, we use data from the Census of Manufactures on the number

terials' value marginal product and marginal cost may reflect markups or static input frictions that equally distort each input (De Loecker and Warzynski, 2012).

[17] We do run some analysis at the county-industry level, using the unbalanced panel.

[18] Given heterogeneous input distortions, the cost share within a given county-industry will not only reflect the production function, so we calculate each industry's cost shares at the national level.

[19] Starting in 1870, the county-by-industry data do not list some "neighborhood industries" such as blacksmithing (Atack and Margo, 2019) or additional 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 less than 5% of manufacturing output in 1870 and 1880. We also created an "other" industry, representing less than 1% of output, reflecting named but small industries.

[20] In our main analysis, we use the 1860-1880 average for both industry production functions and county revenue shares.

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 total value of home manufactures, which we use to consider a potential shift from home manufacturing to more formal manufacturing production. The Census of Population reports total population of each county.

I.B Data on Market Access

An expanding railroad network lowered county-to-county freight transportation costs. Figure 1, panel A, shows the network of waterway routes that includes canals, navigable rivers, lakes, and oceans. Panel B shows the railroad network constructed by 1860, which then expanded by 1870 (panel C) and 1880 (panel D).[21] 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.[22] We also calculate transportation costs under counterfactual scenarios that remove the railroad network or replace the railroad network with an expanded canal network proposed by Fogel (1964).

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

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

County c has greater market access when it is cheaper to trade with other counties d that have greater population.[23] Changes in counties' market access summarize how changes in transportation costs affect counties through interacting goods markets and factor markets across all counties. In Section V we derive this approximation for county market access in a general equilibrium trade model with input distortions. This same approximation for market

[21] Appendix Figure 1 shows the railroad network in 1890 and 1900.

[22] 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 B.3, we report the robustness of our estimates to alternative assumptions in calculating county-to-county transportation costs.

[23] We calculate a county's access to all other counties with reported population, including other counties that are excluded from the regression sample (for instance, because they do not report manufacturing data in each decade).

access arises in a more-restricted model without input distortions (Donaldson and Hornbeck, 2016).

There are two main advantages in measuring changes in counties' market access, rather than focusing directly on counties' proximity to railroads (e.g, whether a county has a railroad or its mileage of railroad track). First, because changes in counties' market access summarize how counties are impacted by the railroad network in a general equilibrium framework, there is a natural transition from estimating relative impacts of county market access to considering aggregate impacts of changes in county market access under counterfactual scenarios. Second, we can identify the impacts of changes in market access using identifying variation from more-distant changes in the railroad network and how changes in the railroad network interact with existing railroads and navigable waterways.

For measuring county market access, as defined in equation 1, we need estimates of θ and τ_{cd}. The parameter θ reflects the "trade elasticity," which varies across empirical contexts. The parameters τ_{cd} represent "iceberg trade costs," which normalize the measured county-to-county transportation costs t_{cd} by the average price per ton of transported goods ($\tau_{cd} = 1 + t_{cd}/\overline{P}$). In Section V.E, we jointly estimate values for θ (2.79) and $\overline{P}$ (35.7).[24] We also report the robustness of our results to alternative parameter values.

Figure 2 shows in darker shades those counties that have relatively greater increases in market access from 1860 to 1870 (panel A) and from 1870 to 1880 (panel B).[25] Our baseline empirical specification compares changes in darker shaded counties to changes in lighter shaded counties within the same state, and within similar latitude and longitude.[26] Comparing counties within nearby areas, there is substantial variation in changes in county market access. Further, comparing across decades, it is often different counties that are experiencing relatively larger or smaller changes in market access; which means that the estimated impacts of county market access do not only reflect particular counties growing relatively over the entire sample period.

Figure 2 maps our main regression sample of 1,802 counties, which includes all counties that report manufacturing revenues and input expenditures in 1860, 1870, and 1880.[27] We

[24] The estimated value of 35.7 for $\overline{P}$ is very close to the value of 35 assumed by Donaldson and Hornbeck (2016) based on commodity price data from Fogel (1964). The estimated value of 2.79 for θ is smaller than the estimated value of 8.22 in Donaldson and Hornbeck (2016), due to differences in the model and sample period. Estimated counterfactual impacts in our framework are not very sensitive to the value of θ, though, unlike in other methods of estimating gains from trade (Arkolakis, Costinot and Rodríguez-Clare, 2012).

[25] Appendix Figure 2 maps relative changes in market access from 1880 to 1890 (panel A) and from 1890 to 1900 (panel B).

[26] 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.

[27] We have adjusted the data in each decade to maintain consistent geographic units (as in Hornbeck (2010)), which reflect county boundaries in 1890 and match to the network database of transportation

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 Decomposing County Productivity

Our main outcome variable is "county aggregate productivity" or "county productivity" in the manufacturing sector: total output value minus total input costs (Solow, 1957). Basu and Fernald (2002) show this measure is a first-order approximation to welfare in the presence of distortions.

We estimate impacts of market access on county revenues and county input expenditures, along with associated impacts on county productivity. Section V derives a predicted log-linear relationship between county productivity growth and increases in county market access, which we see in the data.[28] Section V.E makes further use of the general equilibrium model to estimate impacts on national aggregate productivity, using reduced-form results to pin down relevant parameters.

In this section, we show how to decompose county aggregate productivity growth into two components (Hulten, 1978; Petrin and Levinsohn, 2012; Baqaee and Farhi, 2020), which each also have a log-linear relationship with market access. The first component (growth in county TFPR) measures growth in how efficiently manufacturers in a county turn input expenditures into revenue. The second component (growth in county AE or "allocative efficiency") reflects growth in county input-use and how far the county is from efficient input levels given its production technology. Changes in allocative efficiency reflect relative input distortions (Hsieh and Klenow, 2009) as well as absolute gaps (for instance, markups would imply that the value marginal product of every input is greater than its marginal cost, as in Hall 1988, 1990).

For considering why log county productivity increases with log county market access, it is useful to re-write the impact of log market access on log productivity as a function of the impacts of log market access on log revenue ($P_c Q_c$) and log expenditures on k inputs

routes.

[28]County output prices and county input prices can change with county market access. Endogenous declines in local prices would understate the reduced-form impacts of market access on measured county aggregate productivity. While the regression analysis is limited to data on values, our counterfactual analysis draws on the model to separate changes in quantities from changes in prices and reports impacts on real national aggregate productivity.

$(W_c^k X_c^k)$):[29]

$$\text{(2)} \qquad \frac{\partial \ln Pr_c}{\partial \ln MA_c} \equiv \frac{\partial \ln(P_cQ_c - \sum_k W_c^k X_c^k)}{\partial \ln MA_c}$$
$$= \frac{1}{Pr_c}\left[\frac{\partial P_cQ_c}{\partial \ln MA_c} - \sum_k \frac{\partial W_c^k X_c^k}{\partial \ln MA_c}\right]$$
$$= \frac{1}{Pr_c}\left[P_cQ_c\frac{\partial \ln P_cQ_c}{\partial \ln MA_c} - \sum_k W_c^k X_c^k \frac{\partial \ln W_c^k X_c^k}{\partial \ln MA_c}\right]$$
$$\text{(3)} \qquad \frac{\partial \ln Pr_c}{\partial \ln MA_c} = \frac{P_cQ_c}{Pr_c}\left[\frac{\partial \ln P_cQ_c}{\partial \ln MA_c} - \sum_k s_c^k \frac{\partial \ln W_c^k X_c^k}{\partial \ln MA_c}\right],$$

where $s_c^k = \frac{W_c^k X_c^k}{P_cQ_c}$ or the revenue share of input k. In equation 3, the term in brackets represents the percent impact of market access on revenue minus the (revenue share weighted) percent impact of market access on input expenditures.[30] The term in brackets is scaled up by the ratio of county revenue to county productivity, re-scaling percent growth in county revenue into percent growth in county productivity.[31] From equation 3, we can estimate the impact of log county market access on log county productivity by defining the outcome variable: log revenue minus log input expenditures, weighting log input expenditures by the county's fixed revenue shares and scaling by a fixed ratio of county revenue to county productivity.[32] Notably, measuring the relationship between county aggregate productivity and market access does not require any assumptions on production functions.

Equation 3 can be further decomposed, only now using estimates of production function elasticities (α_c^k). We add and subtract the growth in "expected output" caused by the changes in input expenditures from changes in market access: the sum over the growth rate of each

[29]Productivity is given by total revenue minus total expenditures on k inputs: $Pr_c = P_cQ_c - \sum_k W_c^k X_c^k$. Physical output of county c (Q_c) includes final goods output (Y_c) and intermediate goods output (M_c), and is valued at price P_c. Physical inputs of county c (X_c^k) include intermediate goods inputs (materials), labor inputs, and capital inputs each valued at price W_c^k.

[30]The revenue shares appear in equation 3 because of the distinction between level comparisons and log comparisons (or percent comparisons). For example, if total county revenue is greater than total county input costs, then a 1% increase in revenue is greater, in dollars, than a 1% increase in each input cost.

[31]For example, if input costs were unchanged, a 1% increase in county revenue would correspond to a larger than 1% percent increase in county productivity.

[32]Conceptually, as productivity approaches zero (P_cQ_c/Pr_c) the scaling factor approaches infinity. In practice, we use the average county scaling factor across 1860-1880 (5.1) and discuss robustness to alternative calculations in Appendix B.

input multiplied by its respective output elasticity (α_c^k).[33] Rearranging terms:

$$\begin{aligned}
(4) \qquad \frac{\partial \ln Pr_c}{\partial \ln MA_c} &= \frac{P_c Q_c}{Pr_c}\left[\frac{\partial \ln P_c Q_c}{\partial \ln MA_c} - \sum_k \alpha_c^k \frac{\partial \ln W_c^k X_c^k}{\partial \ln MA_c}\right] && \text{(TFPR)} \\
&+ \frac{P_c Q_c}{Pr_c}\left[\sum_k (\alpha_c^k - s_c^k)\frac{\partial \ln W_c^k X_c^k}{\partial \ln MA_c}\right]. && \text{(AE)}
\end{aligned}$$

The first term in brackets is the impact of county market access on county TFPR (revenue total factor productivity). There are a number of considerations for interpreting TFPR, discussed below, but increases in TFPR represent increases in revenue above-and-beyond the expected increases in revenue from increases in input expenditures. For instance, holding prices fixed, TFPR would increase if establishments become able to make more physical output for the same inputs.[34] Increases in county TFPR are equal to increases in county productivity, as defined above, if resources are allocated efficiently and the value marginal product of inputs is equal to their marginal cost (e.g., as assumed by Solow, 1957; Donaldson and Hornbeck, 2016).

However, markups (Hall, 1988, 1990) or input frictions (Hsieh and Klenow, 2009) drive a "wedge" between the value marginal product of inputs and their marginal costs ($\alpha_c^k / s_c^k > 1$). In the decomposition, this wedge matters for county productivity because it leads to a "gap" between production function elasticities and revenue shares ($\alpha_c^k - s_c^k > 0$), such that increases in input-use would increase revenue by more than the increase in input expenditures (and thereby increase county aggregate productivity holding fixed county TFPR). Increases in county allocative efficiency (AE) do not reflect increased "efficiency" in the sense of reductions in markups or input frictions themselves. Instead, allocative efficiency increases most when input-use increases in counties with high markups or input frictions because those counties use too few inputs from an efficiency perspective.[35] Increases in county allocative

[33] The output elasticity represents the expected log increase in physical output (Q_c) from a log increase in physical input k (X_c^k), which is distinct from the relationship between log revenue and log input expenditure, with implications that we discuss below.

[34] In the model, impacts of market access on TFPR would understate impacts on TFPQ because decreases in transportation costs for output and inputs lower local output prices (with CES demand and Cobb-Douglas production with CRS), though TFPR is often correlated with TFPQ in settings when both are measured (Foster, Haltiwanger and Syverson, 2008; Haltiwanger, Kulick and Syverson, 2018). Reallocation of inputs across industries or establishments, within a county, would also affect county TFPR without underlying changes in establishment TFPQ. In the model, impacts of market access on county AE would also understate real productivity gains, because declining local input prices would cause the increases in input expenditures to understate increases in real inputs (though some local input prices could increase with county market access, particularly for non-traded inputs like land).

[35] Regardless of the underlying source of inefficiency, whether due to markups or input frictions, if value marginal products are below marginal costs then input-use is inefficiently low and "allocative efficiency" increases when input-use increases in that county.

efficiency (AE) do not just reflect "reallocation" of the same inputs across counties, but also reflect growth in aggregate inputs in the United States. When there are positive gaps in a county, county AE increases with higher input-use and the implications for national aggregate productivity depend on the cross-county variation in gaps and both relative and aggregate changes in county inputs.

The second term in brackets represents this change in allocative efficiency (AE), an additional potential source of growth in county productivity even when county TFPR is fixed.[36] County productivity increases when county inputs increase if there is a positive "gap" ($\alpha_c^k - s_c^k > 0$) between measured production function elasticities (which reflect marginal products) and measured revenue shares (which reflect marginal costs, as in Hall 1988). With a positive gap, increases in (dollar) input use lead to larger increases in (dollar) output, thereby increasing county productivity (even holding fixed TFPR).

Our regressions estimate relative changes in county productivity from relative changes in county market access, but these estimated relative effects are not sufficient to estimate national aggregate productivity effects of the railroads because of induced cross-county movement in production inputs and induced changes in aggregate inputs. Section V.E makes further use of a general equilibrium model to estimate these aggregate effects.

In summary, we use data from the Census of Manufactures to define several outcome variables for county c in year t. We start by showing effects on log revenue ($\ln(P_{ct}Q_{ct})$), log materials expenditures ($\ln(W_{ct}^M X_{ct}^M)$), log labor expenditures ($\ln(W_{ct}^L X_{ct}^L)$), and log capital expenditures ($\ln(W_{ct}^K X_{ct}^K)$), defined as the total values for county c in year t. Our main outcome variable is log productivity, in county c and year t, as discussed above:

$$\frac{P_c Q_c}{Pr_c}\left[\ln P_{ct}Q_{ct} - \sum_k s_c^k \ln W_{ct}^k X_{ct}^k\right]. \tag{5}$$

We then define two additional outcome variables that decompose the impacts of market access on county productivity into the impacts through county TFPR growth

$$\frac{P_c Q_c}{Pr_c}\left[\ln P_{ct}Q_{ct} - \sum_k \alpha_c^k \ln W_{ct}^k X_{ct}^k\right] \tag{6}$$

[36]This term has other names in the literature: Petrin and Levinsohn (2012) call it "reallocative efficiency," as did previous drafts of this paper. However, given that the full gains from market access are both due to reallocation over space and increases in aggregate inputs, the term "allocative efficiency" may be clearer in this context.

and the impacts through county AE growth

$$(7) \qquad \frac{P_c Q_c}{Pr_c}\left[\sum_k (\alpha_c^k - s_c^k) \ln W_{ct}^k X_{ct}^k\right].$$

Appendix A.2 provides a reference for these formulas, along with further information on the underlying data from the Census of Manufactures.[37]

III Estimating Equation

We regress outcome Y in county c and year t on log market access (MA_{ct}), county fixed effects (γ_c), state-by-year fixed effects (γ_{st}), and a cubic polynomial in county latitude and longitude interacted with year effects ($\gamma_t f(x_c, y_c)$):[38]

$$(8) \qquad Y_{ct} = \beta \ln(MA_{ct}) + \gamma_c + \gamma_{st} + \gamma_t f(x_c, y_c) + \varepsilon_{ct}.$$

The coefficient β reports the impact of county market access on outcome Y, comparing changes in counties with relative increases in market access to other counties within the same state and adjusting for changes associated flexibly with county latitude and longitude. The identification assumption is that counties with relative increases in market access would otherwise have changed similarly to nearby counties. Our empirical estimates are robust to controlling for other sources of differential growth over this period, and counties experienced similar growth prior to increases in market access.

When estimating impacts of railroads, the main identification concern is generally that railroad construction may have been directed toward counties that would otherwise have grown more over the sample 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 existing transportation network. We also report estimates that control flexibly for changes in railroads within a county and within nearby areas. Borusyak and Hull (2021) discuss how random variation in railroad construction could be used to identify impacts of market access, though that approach requires plausibly random construction of particular railroad lines (and a less dense network than in the United States, such that there would be substantive variation in county market access from those particular railroad lines). A

[37]Our main estimates use the county's average input revenue shares from 1860 to 1880 (as in Petrin and Levinsohn 2012's Törnqvist-Divisia approximation), though we also report similar estimates using the county's input revenue shares from 1860 (and output elasticities from 1860, in the decomposition of county productivity growth).

[38]Note that we estimate equation 8 in levels, rather than in changes, but include county fixed effects that focus identification on county-level changes in market access. We also report separate estimates for pairs of sample periods, where the estimation of equation 8 is equivalent in changes or in levels with county fixed effects.

key feature of our context is that, after controlling flexibly for local railroad construction, there remains substantial residual variation in county market access. We also exploit the interaction between railroads and pre-existing low-cost waterway transportation, whereby some counties inherently benefited less from the national railroad network, to instrument for growth in county market access. These specifications follow those in (Donaldson and Hornbeck, 2016), along with other robustness checks such as using variation in counties' access to more-distant counties.

The main regression sample is a balanced panel of 1,802 counties in 1860, 1870, and 1880 (Figure 2). We report standard errors that are clustered by state to adjust for correlation in ε_{ct} over time and within states.[39]

IV Estimated Impacts of County Market Access

IV.A Estimated Impacts on Productivity

Table 1 presents results from estimating equation 8. We estimate that county market access has a substantial and statistically significant impact on county manufacturing revenue and input expenditures. Column 1 reports that a one standard deviation greater increase in market access from 1860 to 1880 leads to a 19.2% increase in revenue (panel A), 18.3% increase in materials expenditure, 19.6% increase in labor expenditure, and 15.8% increase in capital expenditure.[40] These estimated percent effects are similar in magnitude, which imply little effect on county TFPR, but market access could still increase the total value of county output by more than the total value of county inputs if the value marginal product of inputs is greater than their marginal cost.

Indeed, we estimate substantial increases in county productivity from increases in county market access. Panel E reports a 20.4% increase in log productivity, as defined in Section II.[41] As county market access increases, there is an increase in total county revenue that substantively exceeds the increase in total county input expenditures; that is, there is increasingly more value of output produced in excess of the value of inputs used.

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

[39] If we instead adjust for spatial correlation across counties, following Conley (1999), we estimate smaller standard errors for our baseline specification. The standard errors are 3-5% lower for distance cutoffs of 200 miles or 300 miles, 12-19% lower for distance cutoffs between 400 miles and 700 miles, and 19-30% 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.

[40] 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 for our baseline definition of market access.

[41] Appendix A.2 provides a reference on the formulas for these outcome variables.

populations fixed at 1860 levels, such that changes in county market access are only due to changes in county-to-county transportation costs.[42] 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 moderately larger estimates from our baseline specification when extending the sample period, using available county-level manufacturing data in 1890 and 1900.[43] Column 5 reports similar estimated impacts on county revenue and capital expenditures, using available county-level data in 1850 and restricting the sample to be a balanced panel back to 1850.

Table 2 shows that the above estimated impact of market access on log productivity (panel A) is driven by growth in county AE (panel C) with less effect through growth in county TFPR (panel B). This decomposition depends on production function assumptions (i.e., the calculation of counties' output elasticities for each input, α_c^k). Column 1 reports a decomposition when calculating separate production functions for 31 industries, whereas Column 2 uses 193 industries. Note that different production function elasticities mechanically do not affect the measurement of county productivity in panel A. Column 3 reports similar estimates when estimating one production function for all of manufacturing from 1860 to 1880, and Column 4 reports moderately larger estimates when extending the sample through 1900 using the available aggregated county-level data only.

For county productivity growth to be driven by increases in county allocative efficiency (AE), these estimates imply that county market access is increasing input expenditures in counties where the value marginal product of inputs is greater than their marginal costs. If the value marginal product of inputs were equal to their marginal cost, then county productivity would only increase with increases in county TFPR.

Figure 3 shows approximately log-linear impacts of market access, as predicted in equation 4, on county productivity (panels A and B) and the contribution to county productivity from county AE (panels C and D) and county TFPR (panels E and F).[44] This figure shows the local polynomial relationship between residual productivity and residual market access, after partialling out the controls in estimating equation 8.

The estimates from this decomposition raise questions about potential mismeasurement

[42] Our preferred specifications use actual market access, rather than measuring market access with fixed populations, because railroads also potentially affect market access through changes in the population distribution. Actual market access is highly correlated with population-fixed market access, however, and so our estimates are effectively unchanged whether we use actual market-access, market access with fixed populations, or if we instrument the former with the latter.

[43] 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).

[44] These figures are for approximated market access (panels A, C, E) and model-defined market access derived in Section V.C (panels B, D, F), which are highly correlated with each other.

of input expenditures, which might lead to spurious gaps between the value marginal product of inputs and their marginal cost that would overstate increases in county AE. These measurement concerns are particularly acute for capital (Fisher and McGowan, 1983; Fisher, 1987; Hulten, 1991), though we report later in Table 5 that estimated growth in county AE is largely driven by materials with little contribution from changes in capital expenditure. In Appendix Table 2, we also show that capital mismeasurement does not appear to substantively affect our estimates.[45] In Rows 2 and 3, we double and triple the baseline measured values of capital and find similar effects of market access on productivity and allocative efficiency. This adjustment also addresses concerns that the flow rate of capital services should be larger than what the mortgage rate would imply, for example because of higher depreciation.[46] Row 4 uses national interest rates instead of state-specific ones.

Appendix B discusses a variety of additional specifications, in Appendix Table 2, that adjust for various types of measurement error that might also vary across counties. There is greater scope for county productivity growth when inputs' production function elasticities are larger than their respective marginal cost, as measured by the revenue share. This measured gap may be too large if we under-measure inputs or if we impose incorrect production function elasticities.[47] Appendix Table 2 explores the estimates' sensitivity to the measurement of $\left(\alpha_c^k - s_c^k\right)$, both considering heterogeneous measurement error across space as well as systematic error. In Appendix Table 2, we: use the materials wedge to proxy for both the capital and labor wedges; shrink the dispersion in cross-county wedges; adjust for potential under-counting of labor expenditures; define county productivity based on 1860 county wedges (rather than average county wedges over the 1860 to 1880 period); and decompose productivity growth into TFPR growth and AE growth using alternative output elasticities and returns to scale. In Appendix B, we also explore a potential shift from home manufacturing (in the Census of Agriculture) to more formal manufacturing (in the Census of Manufactures) and impacts on the extensive margin of manufacturing growth in new counties.

Appendix Table 3 considers whether counties experiencing relative growth in market

[45] United States Census Bureau (1880) describes concerns that capital may be under-measured; indeed, Appendix Table 1 shows that national manufacturing capital expenditures are 15% of total labor costs and capital costs, which is below typical values of roughly one-third.

[46] Estimated depreciation rates for equipment in this historical era are around 6% (Davis and Gallman, 2019).

[47] The general view in the literature is that historical manufacturing firm returns to scale were roughly constant, with evidence for decreasing returns or increasing returns depending on adjustments to measured inputs (see, e.g., Atack, 1977; Sokoloff, 1984; Margo, 2014), which is also the view for modern manufacturing (Blackwood et al., Forthcoming). Lafortune et al. (2021) estimate returns to scale around 0.95 for the late 19th century, and we show our results are robust to their findings, as well as to instead assuming returns to scale of 1.05.

access might otherwise have changed differently: estimating similar pre-trends in counties prior to their relative growth in market access; controlling for time-varying effects of county characteristics in 1860, including counties' 1860 input wedges and input gaps; and adjusting for potential differential effects of the Civil War. For example, Appendix Table 3 reports similar estimates when controlling for the share of counties' 1860 revenue in each industry, interacted with year, given the potential for relative changes in industry output prices or other industry-specific shocks to differentially impact counties' growth.

Appendix Table 4 focuses on the measurement of county market access: calculating county-to-county transportation costs under alternative assumptions; using different values for the trade elasticity θ or the average price per ton of transported goods $\overline{P}$; and adjusting for the influence of international trade and mismeasurement of population in the 1870 Census. See Appendix B for further discussion.

IV.B Endogeneity of Railroad Construction

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

One of the advantages of analyzing changes in county market access, rather than directly estimating impacts of local railroad construction, is that much variation in counties' market access is due to changes elsewhere in the railroad network and how the railroad network interacts with other components of the transportation network. We can estimate the impacts of county market access, controlling flexibly for railroad construction in the county and nearby areas, which focuses the identifying variation on changes in county market access from more-distant railroad construction and how railroad construction interacts with waterways and the existing transportation network. Local railroad construction might also directly impact local manufacturing activity, through increases in the demand for manufactured construction materials (Fishlow, 1965), and so these controls also adjust for potential direct effects on manufacturing from local railroad construction itself.[48]

[48] Other approaches sometimes focus on isolating some portion of new infrastructure placement that may be more exogenous: because it followed historical plans or paths, because lines were built between particular locations and inadvertently affected intermediate places, or because some lines were built for reasons less associated with local economic demand (and see Borusyak and Hull (2021) for discussion of how to use some exogenous portion of infrastructure placement to identify its effects, modeling treatment effect spillovers through changes in market access). Our empirical approach is designed for contexts in which infrastructure placement is potentially endogenous to local economic demand, and the transportation network is quite dense, such that there is not substantive variation in areas' market access from particular known exogenously-placed connections. Places with growing market access may have otherwise changed differently, for instance because of potentially changing usefulness of their initial geographic location. We explore the scope for this to be driving our results in Appendix Table 3, and broadly find that our results are quantitatively similar even

Table 3 reports similar estimated impacts of county market access when controlling flexibly for local railroad construction. Column 1 reports baseline estimates from Table 2, 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 3 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 of rivers, canals, lakes, and oceans.

As an alternative empirical approach, we can otherwise 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 benefit less from the new railroad network and should generally experience smaller increases in market access. We define county "water market access" in 1860, which reflects its measured market access when excluding all railroads from the transportation network.

Table 4, 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,[49] we can instrument for changes in county market access using county water market access in 1860 (and column 1 reports the first-stage results).[50] Columns 2, 4, and 6 report the 2SLS estimated effects of market access on county productivity, county TFPR, and county AE,[51] which are less precise but similar in magnitude to the OLS estimates (columns 3, 5, and 7).

The waterway IV identification assumption would be violated if these counties with greater water market access would have changed differently from 1860 through 1880, however, and a different empirical approach would estimate the effects of county market access controlling for counties' water market access in 1860 (interacted with decade, allowing these places to change differently over time). Appendix Table 3 reports these estimates, which are

after controlling for a variety of potentially time-varying effects of initial conditions.

[49]Empirically, counties with better waterway access are not differentially likely to gain a connection to the railroad network by 1870 or by 1880.

[50]The first stage F-statistic is larger than the relevant critical value from Olea and Pflueger (2013).

[51]We report point estimates and confidence sets following Andrews (2018) and Sun (2018).

similar to those in Table 2, and also reports similar estimates when controlling for counties' 1860 market access (interacted with decade).

An expanding national railroad network affected different counties' market access from 1860 to 1870 and from 1870 to 1880, and over each decade there were similar effects of market access on county productivity. Splitting our baseline analysis by decade pair (1860 and 1870, 1870 and 1880, 1860 and 1880), increases in county market access lead to substantial increases in county productivity that are more often driven by increases in county AE than by increases in county TFPR (Appendix Table 3).[52] We also estimate little serial correlation in county market access, regressing changes in log market access from 1870 to 1880 on changes in log market access from 1860 to 1870.[53]

We also also find that growth in county market access was not associated with differential pre-trends in county manufacturing activity. Estimating the effects of county market access and counties' future market access, controlling for contemporaneous railroads and future railroads, Appendix Table 3 reports that contemporaneous county market access has significant effects on county productivity (driven by growth in county AE) but future market access does not predict county productivity growth (i.e., these outcomes were changing similarly prior to growth in counties' market access).[54]

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

The estimated impact of market access on county productivity is largely driven by estimated growth in county allocative efficiency (AE), with little estimated change in county TFPR, and this section explores sources of growth in county AE along with other potentially associated changes in county economic activity. We estimate that county AE growth is driven by increases in input expenditures, in places where distortions lead to "gaps" between the value marginal product of inputs and their marginal cost, but those gaps do not themselves decrease with county market access. Across a range of outcomes, we estimate little systematic change in the structure of the county economy itself; rather, there was a general expansion of county economic activity from increases in counties' market access.

Table 5, column 1, reports that the estimated increase in county AE (from Table 2)

[52] For the 1870 to 1880 period, there is more indication of productivity gains driven by TFPR growth. Pooling the 1870 to 1900 period, however, changes in county productivity are driven by county AE growth and in pooled models this difference for the 1870-1880 period is not statistically significant.

[53] Controlling for state fixed effects and latitude/longitude the point estimate (on a one percent increase in market access from 1860 to 1870) is -0.02, with a standard error of 0.04. The estimate is 0.001 (0.04) when additionally controlling for contemporaneous and future growth in whether a county has any railroad and the length of its railroads.

[54] This specification controls for contemporaneous and future values for whether a county has any railroad and the length of its railroads, and the estimates are similar with additional cubic polynomial controls for contemporaneous and future railroad length in the county and nearby areas.

is largely driven by increases in materials (panel C), followed by labor (panel B), with no change from capital (panel A). Table 1 showed similar percent effects on expenditures for each input, but average county gaps are largest for materials, followed by labor and then capital (Appendix Table 5). Average county wedges are more similar across inputs, and even moderately smaller for materials (Appendix Table 6), but materials expenditures are by far the largest share of total input expenditures (Appendix Table 1). Input wedges represent the degree of distortion (its output elasticity divided by its revenue share), but the impacts on county AE from increased input expenditures depend on that input's importance and the resulting gap (its output elasticity minus its revenue share).

There is variation in input wedges across regions (Appendix Table 6), with associated differences in input gaps across regions (Appendix Table 5).[55] Gaps greater than zero and wedges greater than one may reflect a range of factors that would lower input-use below efficient levels.[56] These regional differences in gaps and wedges are largely driven by differences in revenue shares (Appendix Table 7), rather than differences in output elasticities (Appendix Table 8), which means that the mix of industries across regions is not systematically more intensive in particular inputs.

Cross-county variation in input gaps matters for how much national aggregate productivity would decline without the railroads. County aggregate productivity increases as that county receives more inputs, if that county's input gaps are positive, but the impacts on national aggregate productivity depend on how much that county input growth reflects inputs being shifted from another county whose input gaps are relatively smaller or larger and, also, how much it reflects aggregate input growth. Section V.F reports these effects on national aggregate productivity, based on the cross-county variation in input gaps and the relative and aggregate changes in input-use under counterfactual scenarios.[57]

Table 5, columns 2 and 3, report that county input gaps and wedges do not themselves change systematically from increases in county market access. These estimates suggest that increases in county market access are not systematically reducing input frictions or firm

[55] Appendix Figure 3 shows the cross-county dispersion in wedges by decade, where there is not much of a systematic pattern: the dispersion in the materials wedge decreased over time, increased for the labor wedge, and for the capital wedge stayed roughly the same.

[56] 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).

[57] Section V clarifies that the percent impact of market access on input expenditures does not vary with county-level gaps, and even the percent impact on county productivity does not depend on county-level gaps, but the impact on national aggregate productivity depends on county-level gaps and how inputs change across counties in the absence of the railroad network.

markups in the county. We cannot measure how much of the wedge is driven by markups or particular types of frictions, but the potential sources of wedges need not decline with county market access. Market access need not affect a variety of market distortions (e.g., borrowing constraints, implicit or explicit taxes, contracting failures, imperfect property rights enforcement). There could also be competing effects on input frictions, which cancel out in aggregate: borrowing or inventory management may become easier when there is more frequent and less costly movement of goods or managers (Cronon, 1991)), but input frictions may become more severe if local sources of funds are slow to expand and meet increased demand.[58] In Section V, we write down a model where the gaps between marginal products and marginal costs of inputs are not themselves a function of market access, since in this context they appear not to have been.[59]

We have focused on county-level changes in input-use, though changes in county productivity could also reflect within-county reallocation of inputs across industries. In our decomposition of county productivity growth, within-county reallocation of inputs across county-industries would appear in county TFPR: if, within a county, inputs shift from marginally less-productive industries to marginally more-productive industries, then county revenue would increase holding county input expenditures fixed. Our previous estimates indicated little change in county TFPR, and column 4 of Table 5 reports that county production did not systematically shift toward industries that are more capital-intensive, labor-intensive, or materials-intensive. Column 5 reports there was little decline in the standard deviation of wedges across industries within a county, as county market access increased, which suggests that inputs did not shift from more-distorted industries to less-distorted industries within counties.

Our main regressions are at the county level, aggregating county-industry data while allowing for cross-county variation in output elasticities due to variation in industry-level production functions, because there is substantial industry entry and exit within counties

[58] Table 6 shows that average establishment size is not changing with increases in market access, which may be related to this absence of systematic change in county wedges.

[59] Average gaps and wedges are declining over this time period, particularly in the Plains region, but we also estimate little systematic effect of market access on gaps or wedges within counties (Appendix Table 9). One notable exception is a substantial decline in labor wedges with market access in the Southern United States (Appendix Table 9), along with a sharp overall increase in labor wedges in the South in 1870 following the emancipation of enslaved people (Appendix Table 6), which may be related. Our baseline estimates are not sensitive, though, to omitting the South or other robustness checks associated with the Civil War (Appendix Table 3). There is some indication of market access increasing input gaps in Western areas, but these areas make up a small share of the overall sample and the baseline estimates are also not sensitive to omitting Western areas. We do not estimate systematic impacts of market access on gaps or wedges in "frontier areas" (Bazzi, Fiszbein and Gebresilasse, 2020), defined as counties with between two and six people per square mile in 1860 and that are within 100km of the boundary where population density fell below two people per square mile in 1860.

that makes it difficult to interpret percent growth at the county-industry level. Appendix Table 10 reports estimates from regressions at the county-industry level, after aggregating industries to five more-consistently present categories: clothing, textiles, and leather; food and beverage; lumber and wood products; metals and metal products; and other industries. We extend our baseline specification by interacting the fixed effects with industry.[60] Column 1 reports estimated average impacts of market access on county-industry productivity, county-industry AE, and county-industry TFPR that are similar to our county-level estimates from Table 2, which is consistent with little cross-industry reallocation within counties that would appear as county TFPR growth in Table 2. Column 2 reports similar average impacts when weighting county-industries by their 1860 share of county revenue. Columns 3 – 6 report industry-specific effects of market access for each consistent industry group, with some variation in effects across industries but no industry-specific effect is statistically different than the average over the other industries. Appendix Tables 11 and 12 report average gaps and wedges by industry group, and Appendix Table 13 reports little systematic effect of market access on gaps and wedges within these industry groups.

Increases in county manufacturing activity appear to be a general expansion of existing economic activity in the county. Column 1 of Table 6 reports little impact of county market access on the number of industries in a county. Table 6 also reports little impact of market access on the average size of establishments, measured as average revenue per establishment (column 2) or average number of workers per establishment (column 3). Instead, increases in county market access lead to a substantial increase in the number of manufacturing establishments (column 4), which is driving the overall increases in revenue and expenditures.

While county manufacturing activity increased substantially with increases in market access, we do not estimate that increases in market access prompted an economic shift from the agricultural sector toward the manufacturing sector within counties. Table 6, column 5, reports little impact of market access on county manufacturing revenue as a share of total manufacturing and agricultural revenue in the county. Similarly, columns 6, 7, and 8 report little impact on county manufacturing value-added, surplus, or employment as a share of the county total in manufacturing and agriculture.[61] Further, we do not estimate that increases in market access encouraged economic activity in counties to become specialized in either

[60]The specification includes county-industry fixed effects and state-year-industry fixed effects. We omit county-industries that appear only once, but do not restrict the sample to county-industries that appear all three years.

[61]For agriculture, we define value-added as 92% of revenue, 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 revenue minus materials expenditures, and we define surplus as revenue minus all input expenditures.

manufacturing or agriculture.[62] Appendix Table 14, panel A, reports little impact of market access on sectoral specialization in terms of revenue (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. There may be within-industry specialization across product varieties, though, which does not appear in our county-by-industry data.

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 production function elasticities (as in Tables 5 and 6). We also assume that county market access does not directly affect county gaps or wedges (as in Table 5). The mechanism driving productivity growth is the change in counties' input-use, combined with county-specific gaps between the value marginal product of inputs and their marginal cost. Whereas the regression analysis estimates relative growth in county productivity from relative increases in county market access, Section V.F reports the national aggregate productivity effects from aggregate changes in each counties' input-use under counterfactual scenarios.

V Aggregate Counterfactual Analysis

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

We add input distortions to a baseline model of economic geography (Eaton and Kortum, 2002; Donaldson and Hornbeck, 2016). 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 the effects on each county are summarized by changes in

[62] 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. The estimated effects are similar for a Herfindahl index that reflects sector concentration (or industry concentration), whereas this specialization index reflects movement toward or away from "typical" production shares.

a county's market access.[63] We show how the presence of input distortions causes there to be national aggregate productivity gains from reductions in transportation costs, which are not captured by changes in land values (as in Donaldson and Hornbeck, 2016) or social savings calculations (as in Fogel, 1964). We then estimate the national aggregate productivity losses from counterfactual transportation networks, such as removing the railroad network or replacing the railroad network with an extended network of canals.

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.[64] 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.[65]

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: $F_o(z) = 1 - e^{-A_o z^{-\theta}}$, with $\theta > 1$.[66] The parameter A_o captures average technical efficiency in county o (absolute advantage), while the parameter θ captures the standard deviation of technical efficiency across varieties (scope for comparative advantage). A smaller θ or "trade elasticity" is associated with more-dispersed technical efficiency across varieties, larger incentives to trade across counties, and a less elastic response of cross-county trade flows to trade costs. 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:

$$MC_o(j) = \frac{(w_o)^{\alpha_o^L}(r_o)^{\alpha_o^K}(q_o)^{\alpha_o^T}(P_o)^{\alpha_o^M}}{z_o(j)} = \frac{\prod_k \left(W_o^k\right)^{\alpha_o^k}}{z_o(j)}. \tag{9}$$

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

[63] Workers maximize utility and, in equilibrium, are indifferent between locations; firms maximize profits. Ultimately, goods markets clear as total production in a location equals total consumption in that location, net of transportation costs.

[64] We now refer to "firms" for convenience, though note that the Census enumeration is at the establishment level (activity is recorded where it takes place, not at headquarters) and so this refers to single-establishment "firms."

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

[66] This Fréchet distribution reflects firms receiving levels of technical efficiency from various potential distributions and discarding all but the best for each variety.

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 (ψ^L), capital (ψ^K), land (ψ^T), or materials (ψ^M). Given positive frictions, firms in county o reduce production of variety j such that its price is greater than its marginal cost of production:

$$p_o(j) = \frac{\prod_k \left(\left(1+\psi_o^k\right) W_o^k\right)^{\alpha_o^k}}{z_o(j)} > MC_o(j) \,. \tag{10}$$

Firm markups, due to market power, can be represented as a common component in each of the ψ terms.[67] Given a positive firm markup, or positive input-specific frictions, the value created by firms increasing production would be greater than the value of resources used. Firms are unable or unwilling to use more inputs, however, as reflected in the ψ terms.

We assume capital is mobile, such that interest rates are fixed exogenously, but this interest rate can vary across counties.[68] 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 a wage w_o, receives indirect utility $V(P_o, w_o) = w_o/P_o$.[69] We assume workers are perfectly mobile across counties, such that worker utility is the same across counties at $\bar{U}$.[70] The local wage rate is endogenous, however, and reflects local prices ($w_o = \bar{U}P_o$) such that nominal wages are higher in counties with higher goods prices. In assuming that workers are perfectly mobile across counties, we focus on a long-run equilibrium in which workers can arbitrage real wage differences over a decade.

There is costly trade of good varieties across counties, for both final goods and intermediate goods (materials). Transporting goods from county o (origin) to county d (destination) incurs a proportional "iceberg" trade cost τ_{od}.[71] In county d, the price of good variety j

[67] While the input-specific frictions are modeled as constants, the assumed CES demand for firm output also implies that firms would choose constant markups.

[68] 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).

[69] We assume workers spend their income in their home county, and that payments to land and capital along with any profits are also spent in each county in proportion to its share of aggregate revenues.

[70] In this context, population movements are generated both from the migration of American-born workers (Rusanov, 2021) and foreign immigration (Sequeira, Nunn and Qian, 2020). In the counterfactual analysis we report estimates for different scenarios that assume a perfectly elastic supply of immigrants from abroad, a perfectly inelastic supply of immigrants from abroad, as well as more intermediate cases.

[71] We measure the cost per ton 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 θ).

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 10) and the transportation cost (τ_{od}). Consumers buy good varieties from their cheapest source, where "consumers" includes workers buying goods and firms buying materials. Following Eaton and Kortum (2002), with the addition of input frictions ψ, the value of total exports from county o to county d is given by:[72]

$$E_{od} = \kappa_1 A_o \left(\Pi_k \left(\left(1 + \psi_o^k\right) W_o^k \right)^{\alpha_o^k} \right)^{-\theta} \tau_{od}^{-\theta} Y_d P_d^{\theta}. \tag{11}$$

County o sends more goods to county d when county o has higher technical efficiency (A_o) or lower "effective costs" $\left(\Pi_k \left(\left(1 + \psi_o^k\right) W_o^k \right)^{\alpha_o^k} \right)$, 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 (τ_{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):

$$CMA_d = P_d^{-\theta} = \kappa_1 \sum_o \tau_{od}^{-\theta} A_o \left(\Pi_k \left(\left(1 + \psi_o^k\right) W_o^k \right)^{\alpha_o^k} \right)^{-\theta}. \tag{12}$$

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 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:

$$FMA_o = \sum_d \tau_{od}^{-\theta} Y_d CMA_d^{-1}. \tag{13}$$

Firm market access is higher in county o when it has access to more product demand: when

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

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:

$$CMA_d = \sum_o \tau_{od}^{-\theta} Y_o FMA_o^{-1}. \tag{14}$$

Similar to equation 13, 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 13 and 14 imply that a county's firm market access and consumer market access are exactly proportional: $FMA_o = \rho CMA_o$, where $\rho > 0$.[73] 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 revenue $\left(\frac{\alpha_d^L}{(1+\psi_d^L)}\right)$, we can then express market access in county o as a function of the endogenous number of workers in each other county d:[74]

$$MA_o = \kappa_2 \sum_d \tau_{od}^{-\theta} L_d MA_d^{\frac{-(1+\theta)}{\theta}} \frac{\left(1+\psi_d^L\right)}{\alpha_d^L}. \tag{15}$$

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.[75]

Changes in county market access summarize how each county is impacted by changes 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, capturing both "demand-side" and "supply-side" channels through which counties are affected by changes in overall market integration. 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 15); data on county population and county-to-county

[73] For this result, we assume that trade costs are symmetric ($\tau_{od} = \tau_{do}$), as constructed in Section I.B.

[74] Here, $\kappa_2 = \bar{U} \rho^{\frac{1+\theta}{\theta}}$

[75] 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 only labor, labor input frictions, and labor cost shares.

transportation costs; and parameter choices discussed below.

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 15 vs. equation 1). From 1860 to 1880, changes in log market access derived from the model (equation 15) and changes in log market access based on our approximation (equation 1) have a correlation coefficient of 0.9998. When estimating equation 8, a one standard deviation greater increase in model-defined log market access is estimated to increase log county productivity by 0.205 (s.e. of 0.051), which is driven by increases in county AE (0.169 , s.e. of 0.051) rather than increases in county TFPR (0.036 , s.e. of 0.025). These estimates represent rounding differences from our estimates in Table 2 for approximated market access.[76]

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.

V.C Predicted Impacts of Market Access

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

$$\frac{d\ln PR_o}{d\ln MA_o} = \frac{P_o Q_o}{Pr_o} \sum_k \left(\alpha_o^k - s_o^k\right) \frac{d\ln X_o^k}{d\ln MA_o}. \tag{16}$$

Market access increases county productivity by increasing real input usage ($\frac{d\ln X_o^k}{d\ln MA_o}$), when the value marginal product of that input exceeds its marginal cost (i.e., when the production function elasticity α is greater than the revenue share s).[77]

Building on equation 16, Appendix C derives the log-linear impact of market access on each input and each input price. Changes in market access have a log-linear impact on

[76] A one standard deviation greater change in model-defined log market access is 0.225, from 1860 to 1880, which is similar to that for approximated market access (0.231).

[77] 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 now only considering productivity impacts of market access through increases in allocative efficiency.

county productivity:

$$
\frac{d\ln PR_o}{d\ln MA_o} = \frac{P_oQ_o}{Pr_o}\left[\alpha_o^L\left(\frac{\psi_o^L}{1+\psi_o^L}\right)\left(\frac{1}{\theta}+\frac{\alpha_o^M+\alpha_o^L+1}{1+\theta\alpha_o^T}\right) + \alpha_o^M\left(\frac{\psi_o^M}{1+\psi_o^M}\right)\left(\frac{1}{\theta}+\frac{\alpha_o^M+\alpha_o^L+1}{1+\theta\alpha_o^T}\right) + \alpha_o^K\left(\frac{\psi_o^K}{1+\psi_o^K}\right)\left(\frac{\alpha_o^M+\alpha_o^L+1}{1+\theta\alpha_o^T}\right)\right]. \tag{17}
$$

Market access increases productivity by increasing input usage, when input usage is otherwise below the efficient level due to input frictions ($\psi > 0$). Market access increases capital usage by $\frac{\alpha_o^M+\alpha_o^L+1}{1+\theta\alpha_o^T}$ percent, which then increases county productivity by $\frac{P_oQ_o}{Pr_o}\left(\frac{\alpha_o^K\psi_o^K}{1+\psi_o^K}\right)\frac{\alpha_o^M+\alpha_o^L+1}{1+\theta\alpha_o^T}$ percent. Market access has a larger impact on labor and materials usage (in real values, but not in nominal values), with the additional term $\left(\frac{1}{\theta}\right)$ reflecting that market access also decreases materials costs and nominal wages in county o.[78]

An interesting result is that input frictions do not change the impact of market access on revenue or input expenditures, in percentage terms, similar to how the impact of market access does not depend on initial local technical efficiency. Input frictions do determine whether market access impacts productivity, however, as captured in equation 17: the effect of market access on productivity, in levels, goes to zero as the frictions go to zero.[79]

Input frictions on land do not affect the impact of market access on productivity (equation 17), as the model assumes land is in fixed supply and so the quantity of land used is not below its efficient level (nor does it respond to market access).[80] Market access does increase land expenditure, however, as the price of land is log-linear in market access:

$$
\frac{d\ln q_o}{d\ln MA_o} = \frac{\alpha_o^M+\alpha_o^L+1}{1+\theta\alpha_o^T} \tag{18}
$$

Adding input frictions to the model does not affect the impact of market access on county 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 ($\alpha_o^M + \alpha_o^L$)

[78] These input price effects are why the previous regression estimates are predicted to understate the impact of market access on real county AE (as discussed in Section II), as the measured increase in input expenditures is predicted to understate the increase in real input usage.

[79] The impact of market access on productivity, in levels, is given by multiplying equation 17 by the level of productivity (Pr_o). The resulting impact, in dollar terms, can be summed across counties. That sum, expressed relative to GDP, is equal to the Domar-weighted sum of percent changes in county productivity (see equations 19 and 20).

[80] Note that this fixed supply of land corresponds to the total physical land in each county, where some unused portion of county land reflects a lower value of total physical land in that county.

with their average labor share of value-added (α^L).[81]

We define national aggregate productivity growth as the total growth in county aggregate productivity, expressed in percent terms relative to national value-added (GDP):

$$
\begin{aligned}
(19) \qquad APG &= \frac{1}{GDP}\sum_o d\,PR_o \\
&= \frac{1}{GDP}\sum_o Pr_o\, d\ln PR_o \\
&= \frac{1}{GDP}\sum_o P_oQ_o\,\Sigma_k\left(\alpha_o^k - s_o^k\right) d\ln X_o^k \\
(20) \qquad APG &= \sum_o D_o\,\Sigma_k\left(\alpha_o^k - s_o^k\right) d\ln X_o^k,
\end{aligned}
$$

where D_o is the Domar (1961) weight for county o (county revenue divided by national value-added).[82] These Domar weights sum to more than 1, and are the appropriate way to aggregate county-level changes in settings with intermediate goods (Hulten, 1978) and distortions that generate a gap between the value marginal product of inputs and their marginal cost (Petrin and Levinsohn, 2012; Baqaee and Farhi, 2019*a*).[83]

V.D Model Interpretation and Discussion of Assumptions

Our model highlights the impact of market access on county aggregate productivity and national aggregate productivity, in the presence of input distortions, 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 without distortions, all economic gains from increased market access are captured by the increase in land values.[84] Increased land values reflect greater factor input payments, but increased productivity reflects gains in output that are not paid to inputs. Thus, in our model, the aggregate economic gains from the railroads are given by the increase in national aggregate productivity in addition to the total increase in land values.

Our model extends the counterfactual analysis in Donaldson and Hornbeck (2016), and

[81] Input frictions do affect land values in levels, just as unobserved technical efficiency affects land values in levels.

[82] Our counterfactual analysis assigns each county the average of its factual and counterfactual Domar weight. In both scenarios the sum of the Domar weights is around 1.6.

[83] Petrin and Levinsohn (2012) and Baqaee and Farhi (2019*a*) provide alternative derivations that simplify to equation 20 for settings like ours in which technical efficiency is constant.

[84] There is also incidence on aggregate worker utility if there is not perfect labor mobility into the United States, which we discuss later. In principle, there could be gains to local workers (given some inelasticity to local labor supply) or gains to capital owners (given some inelasticity to capital supply), but we assume those factors are supplied elastically and land is the only fixed factor and so it bears the total incidence.

is equivalent to that model when removing input frictions, removing intermediates, and constraining the production function elasticities to be constant across counties. The most important departure is the introduction of input frictions. We allow the output elasticities to vary across counties so that we do not confound measured county-specific frictions with county variation in output elasticities (e.g., cross-county variation in industry revenue shares is also driven by different input intensities in production).

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 estimates in Section IV.

First, we assume that county input frictions are exogenous. This assumption is consistent with estimates from Table 5, in which the measured "gaps" and "wedges" were not impacted by changes in market access.[85]

Second, we assume that county production function elasticities are exogenous. This assumption is consistent with estimates from Table 5, which show little impact of market access on elasticities in manufacturing, and estimates from Table 1 that show similar percent changes in expenditures on materials, labor, and capital. Further, Table 6 shows little impact of market access on the county manufacturing share of revenue, value-added, surplus, or employment.[86]

Third, we assume that county technical efficiency is exogenous. This assumption is consistent with estimates from Table 2 that showed little impact of market access on county TFPR, which is often correlated with technical efficiency in data when both are observed (Foster, Haltiwanger and Syverson, 2008; Haltiwanger, Kulick and Syverson, 2018). Our counterfactual analysis then considers impacts on aggregate productivity only through changes in allocative efficiency, under the empirically-motivated assumption that technical efficiency does not also decline in the absence of the railroads.

V.E Estimating the Model

We use data on county population to infer an "amenity" value for each county, drawing on our assumption that workers move across counties to equalize real wages. This "amenity"

[85]We do not identify the underlying sources of input frictions, but various sources of input frictions need not vary with market access (e.g., borrowing constraints, contracting failures, implicit or explicit taxes or regulations, etc). To the extent that county input frictions reflect firm markups, the model's assumptions (CES demand and Cobb-Douglas production with CRS) imply that firms choose constant markups that do not vary with market access.

[86]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. Appendix Table 14 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.

value reflects the quantity of fixed factors in each county, along with the county's exogenous technical efficiency.[87] Given our measured county-to-county transportation costs, the county population data imply a unique vector of county amenity values (see Appendix for details).[88] 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 national aggregate productivity impact from removing the railroad network (or from other counterfactual transportation networks).

The model describes total 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.[89] We also estimate 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.[90]

We jointly estimate values for the trade elasticity (θ) and the average price per ton of transported goods ($\overline{P}$).[91] For any given trade elasticity θ, the model predicts a matrix of

[87] These two variables are not separately identified, but we only need their combined value for the counterfactual analysis.

[88] The vector of county amenity values is unique up to a proportional constant, whose value does not affect the analysis.

[89] As an alternative exercise, we consider counterfactual impacts when assuming there are lower input frictions outside the manufacturing sector.

[90] 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. We calculate county-level elasticities in 1890 by weighting these agricultural sector elasticities by the fraction of 1890 county revenue in agriculture (or manufacturing). 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 revenue-weighted average of its manufacturing industry elasticities in 1880. The counterfactual sample is 2,722 counties with positive population and positive agricultural or manufacturing revenue in 1890, of which 309 sparsely-populated counties do not report manufacturing data in 1880 (which we assume reflects zero manufacturing revenue). We use the 1880 county-by-industry data to calculate county-level wedges, as in Section II, when these data are available. We use 1890 county-level data to calculate wedges and elasticities for 176 counties with no manufacturing data in 1880 (4% of US population in 1890), then use 1900 data to calculate wedges for 69 counties with no manufacturing data in 1890 (1.5% of US population in 1890), and then use state median wedges in 1880 for the remaining 64 counties (0.2% of US population in 1890). We then divide the county-level elasticities by the county-level wedges to calculate county-level revenue shares (and then calculate county-level gaps, defined as the difference between county-level elasticities and county-level revenue shares).

[91] We use this average price per ton of transported goods ($\overline{P}$) to calculate the iceberg trade costs (τ_{cd})

county-to-county railroad shipments (in tons) and county output (in dollars), which we use to generate a value for $\overline{P}$ 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, 1895). Given that average price per ton $\overline{P}$ and trade elasticity θ, 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 across sectors within counties. We choose the θ and $\overline{P}$ that minimizes the residual sum of squares from regressing the model-predicted change in land values (from equation 18) on the observed change in land values. We estimate an average price of transported goods equal to 35.7.[92] We estimate a value of θ equal to 2.788 (95% confidence interval: 1.815 – 3.556), which is moderately lower than (but within the general range of) estimates in the literature (Costinot and Rodríguez-Clare, 2014) and we report the robustness of our counterfactual estimates to assuming alternative parameter values.

V.F Estimated Counterfactual Impacts

We estimate that national aggregate productivity would have been 28.0% 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).[93] This 28% decline in national aggregate productivity reflects only decreases in allocative efficiency, with no decline in county technical efficiency or underlying physical productivity of firms, but this decline in allocative efficiency is equivalent to a 28% decline in aggregate real TFP or roughly 50 years worth of technological innovation in this era.[94]

This 28% decrease in national aggregate productivity is worth 28% of GDP annually, or \$3.4 billion in 1890 dollars. As a comparison, the estimated cost of the railroad network in 1890 was \$8 billion (Adams, 1895). We estimate that this investment in the railroads generated an annual social return of 50%, and that the railroad sector privately captured only 7% of its social return in 1890.[95]

based on the measured county-to-county transportation costs: ($\tau_{cd} = 1 + t_{cd}/\overline{P}$).

[92]This estimated value of 35.7 for $\overline{P}$ is very close to the value of 35 assumed by Donaldson and Hornbeck (2016) based on commodity price data from Fogel (1964).

[93]We focus on counterfactual estimates for 1890, following Fogel (1964) and Donaldson and Hornbeck (2016).

[94]For the 1855 to 1905 period, Abramovitz and David (1973) report an average of 0.50% annual growth in real value-added TFP.

[95]We estimate that the railroads generated an annual private return of 3.5% in 1890. For this calculation, based on numbers from Adams (1895), we sum the railroads' reported net income (\$145 million), debt interest payments (\$217 million), net capital expenditure (\$5 million), and subtract losses not otherwise reflected from some companies (\$30 million) along with subtracting income from other sources (\$52 million). We then divide \$285 million by the cost of the railroads including equipment (\$8.041 billion) and value of land (\$80 million). Much of the railroads' reported transportation expenses were maintenance costs (39% or \$271

Figure 4 maps the county-level counterfactual changes in productivity, where darker shaded counties represent larger counterfactual declines in 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 **??**. Appendix Figure 4 maps the county-level counterfactual declines in market access, and Appendix Figure 5 maps the county-level gaps.

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

When holding fixed total population, we solve for counties' population shares in the absence of the railroad network. We then calculate corresponding changes in other production inputs, output, and the resulting change in aggregate productivity in the United States. Panel B reports that national aggregate productivity is estimated to fall by 5.8% in 1890, in the absence of the railroad network, when maintaining the same total US population. Population and other production inputs become condensed into limited geographic areas, decreasing labor productivity, and so worker utility falls by 34.4%.

When allowing for aggregate declines in population, the model predicts a substantial decline in population in the United States. For worker utility to be unchanged in the counterfactual, the model predicts that the US population would need to be 68% lower in 1890. This is moderately larger than 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 US 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 14% decline in productivity and a 24% decline in worker utility.[96] By comparison, the US population was 37% lower in 1870 and 73% lower in 1840 than in 1890 (United States Census Bureau, 1975).

million), and we interpret the reported "permanent improvements" of $5 million as total capital expenditure minus depreciation. To calculate the annual social return, we sum the annual private return ($285 million), our estimated annualized increase in agricultural land value ($425 million), and our estimated increase in annual productivity ($3.36 billion), divided by the cost of the railroads including equipment and land ($8.121 billion).

[96]For this calculation, using the available Census tabulations, we reduce the total US population by excluding both the foreign-born population in 1890 and the white native-born population in 1890 with foreign-born parents.

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 2.7% lower using only railroads that existed in 1880, 10.4% lower using railroads from 1870, 16.4% lower using railroads from 1860, or 23.4% lower using railroads from 1850 (Table 7, panel A, columns 2 – 5).

Additional canals might have been constructed to mitigate national productivity losses, in the absence of the railroad network, but we find that these canals would have been an ineffective substitute for the railroad network. We evaluate the system of feasible canals proposed by Fogel (1964), estimating that productivity would have been lower by 24.7% in 1890 when replacing the railroad network with these additional canals (Table 7, panel A, column 6). That is, the additional canals would have mitigated only 12% of the national aggregate productivity loss from removing the railroad network.[97]

By contrast, the railroads would have been "cheap at twice the price." We estimate that productivity would have been lower by 9.3% 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 67% of the national productivity decline.

Table 8 reports that our estimated counterfactual impacts on national aggregate productivity vary moderately with the estimated value of $\overline{P}$ (average price per ton of traded goods), which effectively re-scales railroad rates, and the estimated counterfactual impacts are not sensitive to the estimated value of θ (the trade elasticity).[98]

We would estimate zero dollar impact of the railroads on national aggregate productivity, mechanically, if we assumed zero gaps between inputs' value marginal product and marginal cost. Our baseline counterfactual assumes the measured frictions in the manufacturing sector also reflect frictions in the agricultural sector. Alternatively, we could assume zero gaps in the agricultural sector, and only apply our estimated manufacturing wedges to the county's manufacturing share of combined revenue across manufacturing and agriculture.[99] In this

[97]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.

[98]Our point estimate for θ is 2.788, with a 95% confidence interval between 1.815 and 3.556. For alternative values of θ, we estimate that removing the railroad network would have lowered productivity in 1890 by 28.9% ($\theta = 1.815$), 27.5% ($\theta = 3.556$), or 24.5% ($\theta = 8.22$). As in Donaldson and Hornbeck (2016), alternative values of θ effectively re-scale market access but have little effect on the estimated impacts. 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 θ. The estimated counterfactual impacts are more sensitive to the estimated value of the average price per ton of traded goods, $\overline{P} = 35.7$, because alternative values of $\overline{P}$ effectively re-scale the baseline transportation rates (whereby lower values of $\overline{P}$ magnify differences between the factual and counterfactual scenarios and higher values of $\overline{P}$ diminish differences between the factual and counterfactual scenarios). For alternative values of $\overline{P}$, we estimate counterfactual productivity declines of 36.4% ($\overline{P} = 20$) and 23.3% ($\overline{P} = 50$).

[99]We only measure revenue in manufacturing and agriculture, and so we are now effectively assuming that

case, we estimate that national aggregate productivity would have declined by 17.4% in 1890 without the railroad network (Table 8, column 7).

The estimated counterfactual impacts also become moderately smaller if we adjust counties' measured input expenditures, as in Section IV, using counties' measured materials wedges as a proxy for capital wedges or labor wedges (due to potential measurement error in county expenditures on capital or labor). We estimate that aggregate productivity would have declined by 25.6% when assigning capital expenditures to counties such that counties' capital wedge equals their materials' wedge (Table 8, column 9).[100] This aggregate productivity decline is 21.3% when also assigning labor expenditures to counties such that labor wedges are the same as materials wedges (Table 8, column 10).[101] We interpret these as conservative estimates, given the potential for input frictions in labor or capital markets to create greater distortions than for materials.

As a comparison for these estimates' magnitudes, we consider an alternative counterfactual that maintains the railroad network in 1890 but considers the national aggregate productivity gains from removing all input distortions. In this case, national aggregate productivity in 1890 would increase by 101.9%.[102] Holding total population fixed, national aggregate productivity would increase by 27.9% and worker utility would increase by 135% when removing all input distortions.

A back-of-the-envelope calculation helps to understand the magnitude of our baseline 28% counterfactual impact on national aggregate productivity. This estimated impact on national aggregate productivity is driven by county-level gaps between the marginal value of inputs and their marginal cost, multiplied by counterfactual changes in county-level production inputs. The national revenue-weighted average gaps are 0.073 (for materials), 0.046 (for labor), and 0.012 (for capital), such that the sum of these gaps is 0.131. If we assume that the counterfactual declines in production inputs were the same in each county (67% or 1.11 log points), this corresponds to a national aggregate productivity decline of 21% or 0.233 log points (0.131 times 1.11 times 1.6, where 1.6 is the sum of the county-level Domar weights

wedges in other sectors are equal to the weighted average wedges across manufacturing (from Section IV) and agriculture ($\psi = 0$ for this alternative case only).

[100]The measurement of capital expenditures is particularly subject to measurement error, but capital expenditures are a small share of total input expenditures and assuming zero misallocation in capital (i.e., fixing the capital wedge equal to 1 and adjusting capital expenditures in each county such that its revenue share is equal to its measured cost share) reduces the estimated productivity loss from 28.0% to 24.6% (Table 8, column 8).

[101]If we decrease the dispersion in capital wedges (or all input wedges) wedges by 5, 10, or 25 percent, as in Appendix Table 2, the counterfactual estimates are within one percentage point of our baseline estimate.

[102]In this counterfactual, further increases in inputs are no longer marginally productive, but the inframarginal increases in input-use generate allocative efficiency growth (multiplying the change in input-use by the Domar-weighted average of factual 1890 gaps and zero).

in equation 20). Our calculated counterfactual decline in national aggregate productivity is larger (28%), which reflects cross-county heterogeneity in the gaps and input changes. Indeed, the population-constant counterfactual decline in national aggregate productivity (5.8%) is roughly the difference between the actual counterfactual (28%) and that back-of-the-envelope calculation (21%).

National aggregate productivity falls in the counterfactual scenarios because of gaps between the value marginal product of inputs and their marginal cost, but the average gap in 1890 (0.13) is not large in comparison to other eras. For the United States, in 1997, the manufacturing gap is around 0.3.[103] Thus, the substantial impacts of the railroads on national aggregate productivity are not driven by especially large measured gaps in the historical data; rather, the effects are driven by substantial impacts of the railroads on both the relative allocation of inputs across counties and aggregate inputs in the United States.

VI Interpretation

We estimate substantially larger economic gains from the railroads, as a share of GDP, than previous estimates of 3.2% (Donaldson and Hornbeck, 2016) or 2.7% (Fogel, 1964). Our estimated impact on national aggregate productivity (28%) also understates the total economic gains because it supplements those previous estimates: our counterfactual estimates would indicate no impact on national aggregate productivity from the railroads if there were no differences between counties' value marginal product of inputs and their marginal cost (as assumed by Fogel (1964) and Donaldson and Hornbeck (2016)), whereas the economy would still benefit from the railroads decreasing resources spent on transportation (as in Fogel 1964 and Fishlow 1965 or economic gains capitalized in land values (as in Donaldson and Hornbeck, 2016).

Our analysis starts with the manufacturing sector and extends this analysis to the broader economy, whereas Fogel (1964) and Donaldson and Hornbeck (2016) start with the agricultural sector and extend their analyses to the broader economy. In considering impacts on the broader economy, the key difference in our approaches is where those economic gains will appear: for Fogel (1964), the benefits from railroads are confined to the transportation sector through savings in transportation costs; for Donaldson and Hornbeck (2016), their aggregate counterfactual analysis uses a one-sector model of the entire economy (not just the agricultural sector) and the aggregate impacts are capitalized in land values.[104]

[103] The modern calculations use the NBER-CES database (Becker, Gray and Marvakov, 2013) and assume that the cost of capital is 8%.

[104] Donaldson and Hornbeck (2016) suggest that further research might explore impacts on the manufacturing sector, through increases in technical efficiency, but the assumption of no market distortions (and perfectly elastic supply of capital and labor) focuses the economic incidence of the railroads on the value of the fixed factor (land). There are no county-level data on the value of non-agricultural land or other

Donaldson and Hornbeck (2016) estimate that land values would have fallen by 58.4% without the railroad network, which implies economic losses equal to 3.2% of GDP on an annual basis.[105] The largest annual economic loss that Donaldson and Hornbeck (2016) could have estimated is 5.35%, from a 100% decline in land value, though land value could not go to zero in the model without the whole economy going to zero.[106] In our model that allows for market distortions, the difference between output value and input costs is not capitalized in land values, so there can be national aggregate productivity gains (or losses) that are not captured by changes in total land value.

One general implication for measuring the economic incidence of new infrastructure or new technologies is that increased payments to land (or labor or capital) do not include all economic gains when there are market distortions (e.g., due to input frictions or market power).[107] We show that these additional economic gains can be substantively large, particularly when new infrastructure or new technologies are broadly used and encourage substantial expansion of economic activity. As in Baqaee and Farhi (2020), TFP growth in one sector (transportation) can increase production in other sectors that were inefficiently small and thereby generate larger aggregate productivity gains than implied by the Domar-weighted increase in transportation sector TFP.[108]

---

fixed factors in the non-agricultural sector, particularly without the cost of sunk investments, and so the calculated dollar losses come from the value of agricultural land only.

[105]While allowing for distortions does not matter for counterfactual effects of the railroads on land values, allowing for intermediate goods and county-specific production functions does change the counterfactual impacts on land values (and we estimate an aggregate 67.0% counterfactual decline in land value).

[106]Similarly, Fogel's calculations also use the total value of agricultural land in areas to bound the lost economic value of those areas (more than 40 miles from a waterway), which would "at worst" be abandoned in the absence of the railroads, and then considers the increases in transportation costs without the railroads for areas within 40 miles of a waterway.

[107]There are aggregate productivity gains when production is inefficiently low due to input frictions, but also when those distortions are due to market power (i.e., markups): firms with market power keep their price high by producing too little output and using too few inputs, and so policies that encourage a monopolist to grow can increase aggregate productivity (Boar and Midrigan, 2020). Market power leads to inefficiencies, but the inefficiency is that firms with market power are producing too little and so aggregate productivity increases as inputs are reallocated to firms with more market power (even if that increases the average markup in the economy). Even if high-markup firms are not especially productive, in a physical sense, their high price means that consumers on the margin would value more of that good than the goods being produced with those inputs by lower-markup firms. Increases in aggregate productivity do not depend on the average productivity of firms; rather, it depends on the marginal productivity of firms, and aggregate productivity increases when inputs move to higher marginal productivity firms even if those firms are less productive on average.

[108]Quantifying these direct gains from transportation sector TFP growth is effectively Fogel's social savings approach. The economic gains from decreasing resources spent on transportation can be measured directly by calculating the decreases in transportation costs using the railroads instead of the waterways: this is the social savings calculation in Fogel (1964), which implies that our estimated impact on productivity is in addition to Fogel's estimate and that the social savings approach misses the broader impacts on national aggregate productivity that we consider. Allen and Arkolakis (2020) derive Fogel (1964)'s social savings calculation in general equilibrium and show how it can break down with departures from the neoclassical

Our estimates highlight the importance of considering market distortions when quantifying economic impacts of new technologies or infrastructure. One new technology may appear to have small aggregate effects, when considering only productivity gains within that particular sector, but may have much larger effects through induced changes in other sectors that are producing too little (or too much) from society's perspective. The railroads decreased transportation costs, effectively subsidizing the expansion of economic activities throughout the economy that had a positive social return (i.e., activities whose value marginal product exceeded their marginal cost). The more that economic activity expands in response to decreased transportation costs, the greater the aggregate economic gains, which is opposite to the intuition of Fogel (1964, 1979) in which the railroads' impacts were supposed to be bounded above by assuming an inelastic demand for transportation.[109] When there are market distortions and other activities are under-provided (i.e., their social marginal return is greater than their private marginal cost), then new technologies or infrastructure can generate much larger economic gains by encouraging the expansion of those other activities.

We estimate that the railroads substantially increased the scale of the US 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 national aggregate productivity gains by encouraging the expansion of an economy with market distortions. Market integration need not decrease market distortions; indeed, estimated distortions in modern US data are generally larger than in this historical era, such that there continue to be large potential aggregate productivity gains from further increases in input-use (e.g., through immigration,

model (in their case, in the presence of agglomeration economies). This echoes the critique of (Fogel, 1964) by David (1969), who emphasizes increasing returns to scale as the reason for market inefficiencies. In response, Fogel (1979) makes this assumption more explicit: that in non-transportation sectors, firms' value marginal product is equal to their marginal cost. Donaldson and Hornbeck (2016) maintain this assumption from Fogel (1964, 1979). Our main departure from Fogel (1964, 1979) and Donaldson and Hornbeck (2016) is to allow for various distortions that can drive a wedge between the social benefit and private cost of firms expanding production (e.g., firm markups or input market frictions).

[109] This issue is related to the seminal work of Harberger (1964), who simplifies the economic impacts of a marginal tax change by assuming there are no other distortions in the economy, but Harberger (1964) notes the true impacts of the tax change are probably understated because of resulting changes in other activities that are also distorted. In a similar spirit, Hulten (1978) presents a general motivation for focusing only on within-sector productivity improvements while a recent literature on production networks (Baqaee and Farhi, 2019*b*; Liu, 2019) has used general input-output models to move beyond this assumption. 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 important omission comes from not considering impacts from resulting changes in other activities.

increased infrastructure spending, or other capital investment).

We do not find that the railroads increased county TFPR, and we hold counties' technical efficiency fixed in our counterfactual estimates. Our estimated impacts on county TFPR may understate impacts on physical productivity and technological innovation, as the model predicts output prices decline with county market access, and future research can revisit these potential effects with data on physical productivity and technological innovation.

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

VII Conclusion

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

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

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

Our counterfactual analysis does not include impacts of the railroads on physical productivity (technical efficiency), or ways in which technological innovation might respond to increases in market access. We estimate little effect of county market access on county TFPR (revenue total factor productivity), which is often correlated with physical productivity, but decreases in output prices would cause our estimated impacts on county TFPR to understate increases in physical productivity.

We also do not consider a variety of other mechanisms through which railroads may have impacted the US economy. Our analysis does not consider how the construction and operation of the railroads may have directly affected the economy, such as through the development of improved management practices (Chandler, 1965). We also do not consider how the railroads may have impacted worker mobility, both across counties and within urban areas. The railroads likely encouraged certain economic activities to agglomerate in major urban centers, particularly wholesale trade, with effects highlighted by qualitative analyses of individual major cities (e.g., Cronon, 1991). Our empirical analysis complements these city histories, examining how a broad range of counties were induced to grow by the railroads and increases in market access. The railroads shifted economic activity from some counties to others, along with increasing aggregate economic activity in the United States, and these effects combined to generate substantial national aggregate productivity gains that were worth roughly 50 years of technological innovation in this era.

References

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

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

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

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

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

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

Asker, John, Allan Collard-Wexler, and Jan De Loecker. 2014. "Dynamic inputs and resource (mis) allocation." *Journal of Political Economy*, 122(5): 1013–1063.

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

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

Atack, Jeremy, and Fred Bateman. 1999. "Nineteenth-century US industrial development through the eyes of the Census of Manufactures." *Historical Methods*, 32(4): 177–188.

Atack, Jeremy, and Robert A Margo. 2019. "Gallman revisited: blacksmithing and American manufacturing, 1850–1870." *Cliometrica*, 13(1): 1–23.

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

Balboni, Clare. 2019. "In Harm's Way? Infrastructure Investments and the Persistence of Coastal Cities." *Working Paper*.

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

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

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

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

Baylis, Patrick. 2020. "Temperature and temperament: Evidence from Twitter." *Journal* *of Public Economics*, 184: 104161.

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

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

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

Blackwood, G Jacob, Lucia S Foster, Cheryl A Grim, John Haltiwanger, and Zoltan Wolf. Forthcoming. “Macro and Micro Dynamics of Productivity: From Devilish Details to Insights.” *AEJ: Macroeconomics.*

Boar, Corina, and Virgiliu Midrigan. 2020. “Markups and Inequality.” *Working Paper.*

Boberg-Fazlic, Nina, and Paul Richard Sharp. 2020. “Immigrant Communities and Knowledge Spillovers: Danish-Americans and the Development of the Dairy Industry in the United States.” *CEPR Discussion Paper No. DP13757.*

Bodenhorn, Howard, and Hugh Rockoff. 1992. “Regional interest rates in antebellum America.” In *Strategic factors in nineteenth century American economic history: A volume to honor Robert W. Fogel.* 159–187. University of Chicago Press.

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

Caplin, Andrew, and John Leahy. 2010. “Economic Theory and the World of Practice: A Celebration of the (S, s) Model.” *Journal of Economic perspectives*, 24(1): 183–202.

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.

Conley, Timothy G. 1999. “GMM estimation with cross sectional dependence.” *Journal of econometrics*, 92(1): 1–45.

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.” *Working Paper.*

Cronon, William. 1991. *Nature’s metropolis: Chicago and the Great West.*

Cunningham, Cindy, Lucia Foster, Cheryl Grim, John Haltiwanger, Sabrina Wulff Pabilonia, Jay Stewart, and Zoltan Wolf. 2021. “Dispersion in dispersion: Measuring establishment-level differences in productivity.”

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

Davis, Lance E, and Robert E Gallman. 2019. "The Share of Savings and Investment in Gross National Product During the 19th Century in the USA." 437–466, De Gruyter Mouton.

De Chaisemartin, Clément, and Xavier d'Haultfoeuille. 2020. "Two-way fixed effects estimators with heterogeneous treatment effects." *American Economic Review*, 110(9): 2964–96.

De Loecker, Jan, and Frederic Warzynski. 2012. "Markups and firm-level export status." *American Economic Review*, 102(6): 2437–71.

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

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

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.

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

Fisher, Franklin M. 1987. "On the misuse of the profits-sales ratio to infer monopoly power." *The Rand Journal of Economics*, 384–396.

Fisher, Franklin M, and John J McGowan. 1983. "On the misuse of accounting rates of return to infer monopoly profits." *American Economic Review*, 73(1): 82–97.

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

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

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

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

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

Gavazza, Alessandro, and Alessandro Lizzeri. 2021. "Frictions in Product Markets." *Working Paper.*

Gross, Daniel P. 2020. "Collusive Investments in Technological Compatibility: Lessons from US Railroads in the Late 19th Century." *Management Science*, 66(12): 5683–5700.

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 Solow's productivity residual." *Growth/Productivity/Unemployment: Essays to Celebrate Bob Solow's 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." *Quarterly Journal of Economics*, 125(2): 767–810.

Hsiang, Solomon M. 2010. "Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America." *PNAS*, 107(35): 15367–15372.

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

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

Hulten, Charles R. 1991. "The measurement of capital." In *Fifty years of economic measurement: The jubilee of the conference on research in income and wealth.* 119–158.

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

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

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

Lafortune, Jeanne, Ethan G Lewis, José Pablo Martínez, and José Tessada. 2021.

"Changing Returns to Scale in Manufacturing 1880-1930: The Rise of (Skilled) Labor?" *Working Paper.*

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

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

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

Olea, José Luis Montiel, and Carolin Pflueger. 2013. "A robust test for weak instruments." *Journal of Business & Economic Statistics*, 31(3): 358–369.

Petrin, Amil, and James Levinsohn. 2012. "Measuring aggregate productivity growth using plant-level data." *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, and T Kirk White. 2020. "Measuring Cross-Country Differences in Misallocation." *US Census Bureau Center for Economic Studies Paper*, 16-50.

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

Rusanov, Vasily. 2021. "Internal Migration and the Diffusion of Schooling in the US." *Working Paper.*

Sequeira, Sandra, Nathan Nunn, and Nancy Qian. 2020. "Immigrants and the Making of America." *The Review of Economic Studies*, 87(1): 382–419.

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

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

Solomon, Ezra. 1970. "Alternative rate or return concepts and their implications for utility regulation." *The Bell Journal of Economics and Management Science*, 65–81.

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

Sraer, David, and David Thesmar. 2019. "A sufficient statistics approach for aggregating firm-level experiments."

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

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

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

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

United States Census Bureau. 1860*a*. "Instructions to U.S. Marshals. Instructions to Assistants."

United States Census Bureau. 1860*b*. "Manufacturers of the United States in 1860."

United States Census Bureau. 1870. "Volume 3. The Statistics of Wealth and Industry of the United States."

United States Census Bureau. 1880. "Report on the Manufactures of the United States at the Tenth Census."

United States Census Bureau. 1975. *Historical Statistics of the United States, Colonial Times to 1970.* US Department of Commerce, Bureau of the 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." *IMF Working Paper*, 18(276).

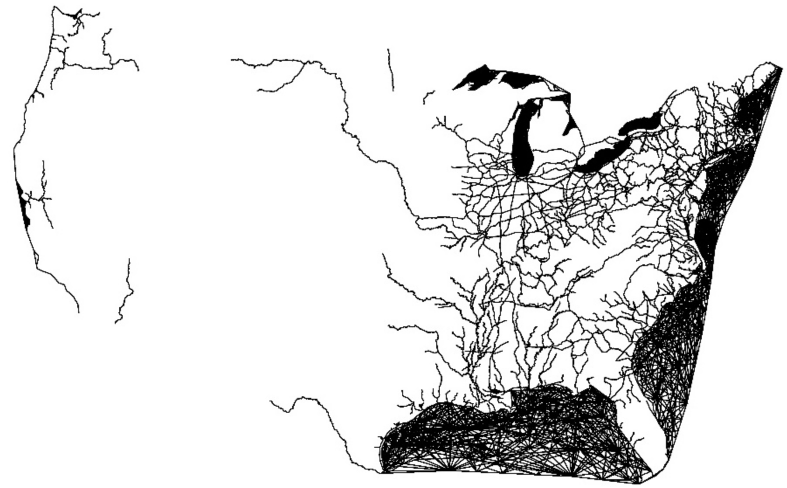
Zárate, Román David. 2021. "Factor Allocation, Informality and Transit Improvements: Evidence from Mexico City."

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

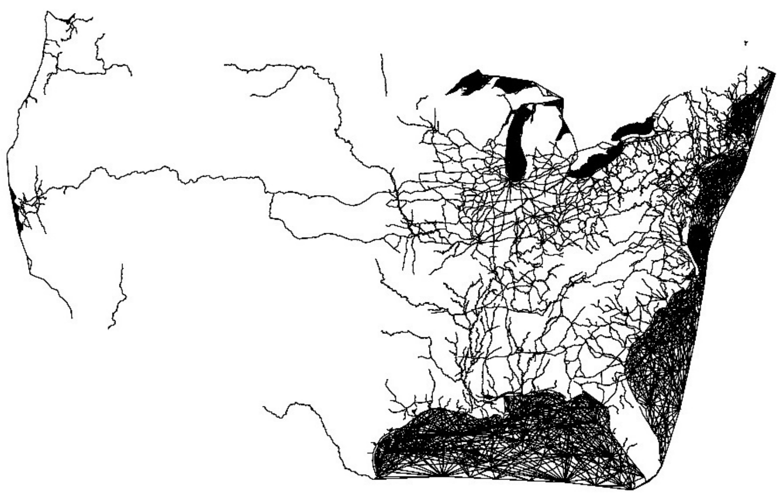
Figure 1. Waterways and Railroads, by Decade

A. Waterways

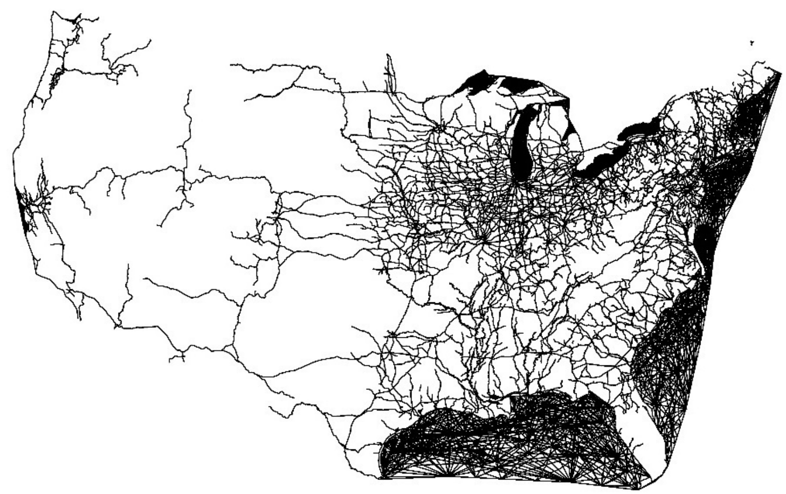
B. Waterways and 1860 Railroads



C. Waterways and 1870 Railroads



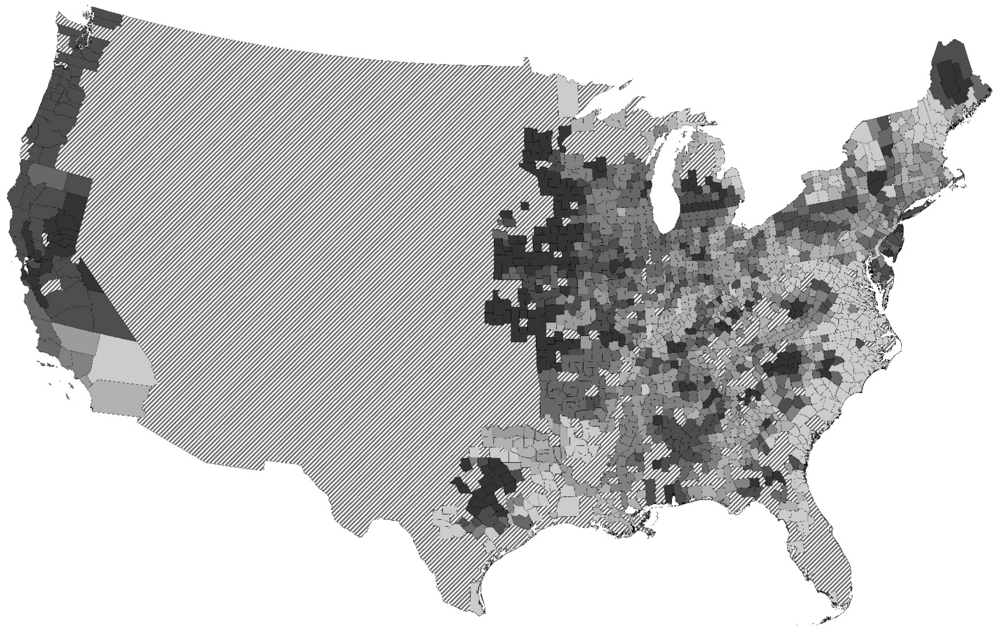
D. Waterways and 1880 Railroads



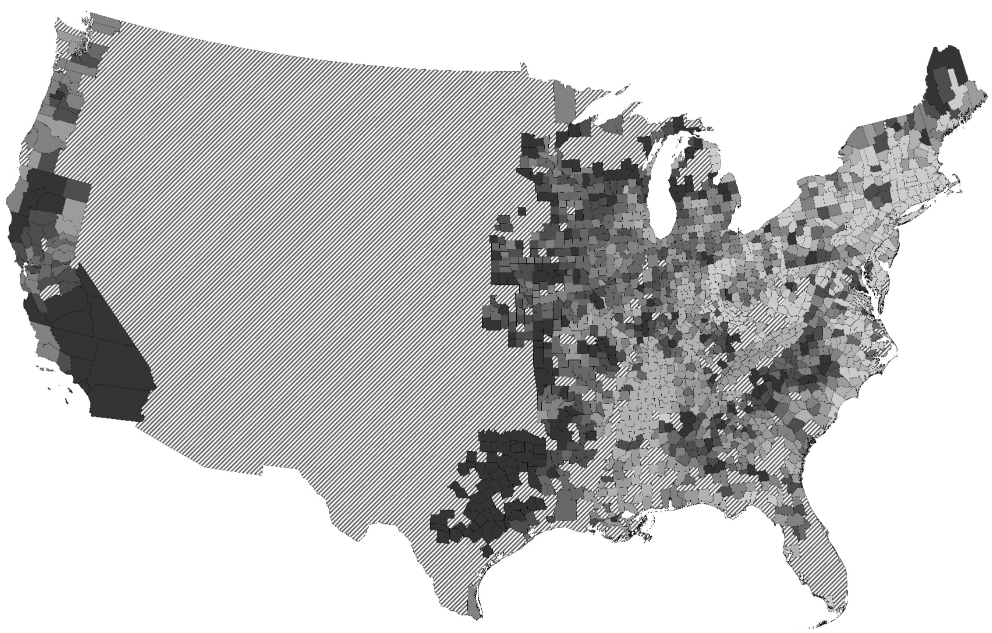
Notes: Panel A shows the waterway network: natural waterways (including navigable rivers, lakes, and oceans) and constructed canals. Panel B adds railroads constructed by 1860, Panel C adds railroads constructed between 1860 and 1870, and Panel D adds railroads constructed between 1870 and 1880.

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

A. From 1860 to 1870



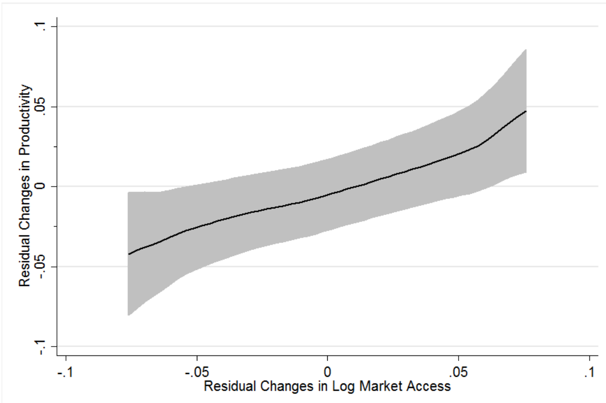
B. From 1870 to 1880



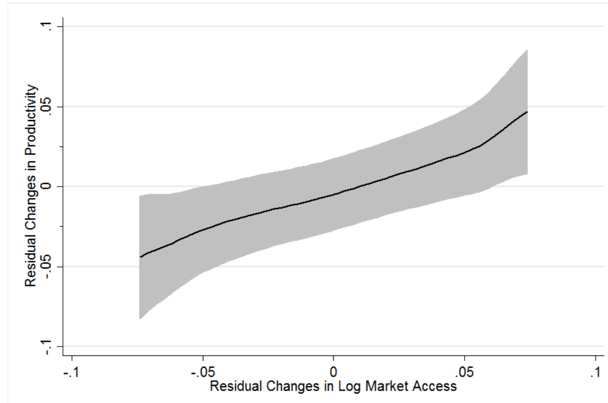
Notes: In each Panel, counties are shaded according to their calculated change in market access from 1860 to 1870 (Panel A) and from 1870 to 1880 (Panel B). Counties are divided into seven groups (with an equal number of counties per group), and darker shades denote larger increases in market access.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

Figure 3. Local Polynomial Relationships Between County Productivity and Market Access, Using Approximated Market Access and Model-Defined Market Access

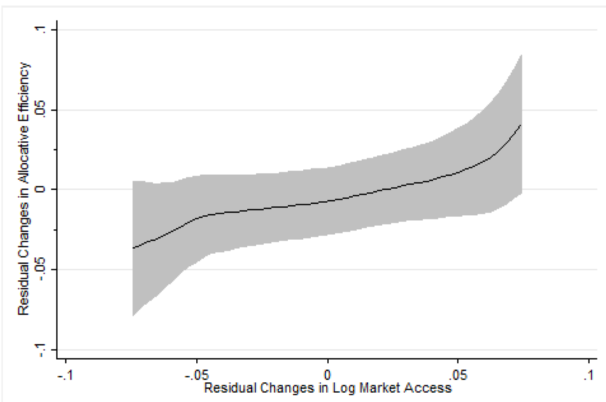
A. County Productivity and Approximated MA



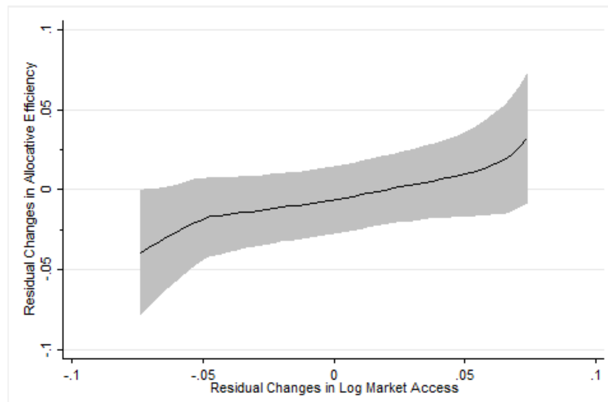
B. County Productivity and Model-Defined MA



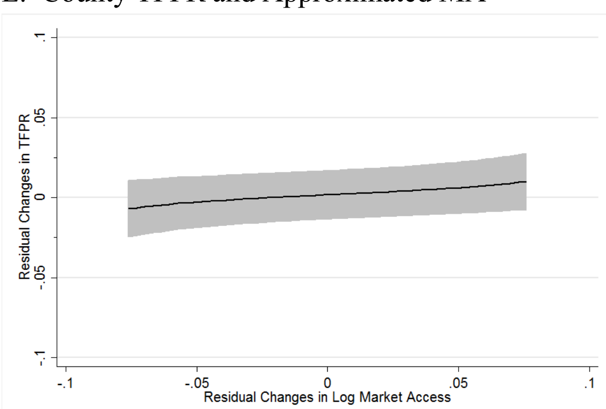
C. County AE and Approximated MA



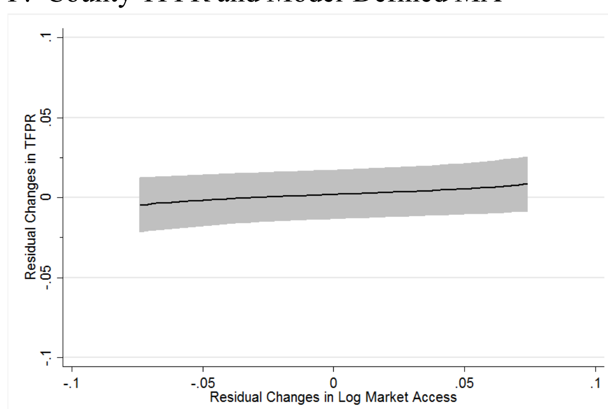
D. County AE and Model-Defined MA



E. County TFPR and Approximated MA

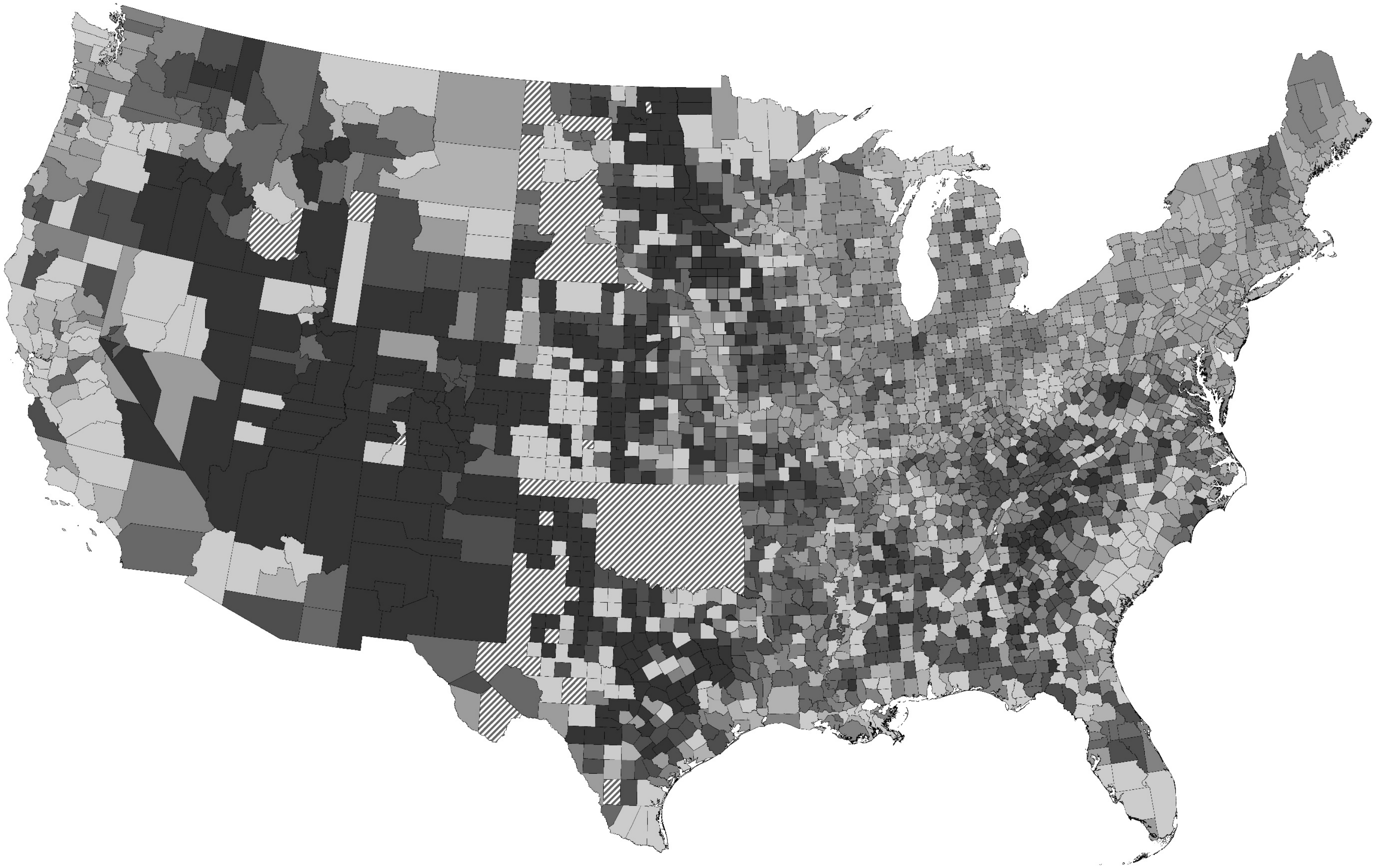


F. County TFPR and Model-Defined MA



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 0.03. The shaded region reflects the 95% confidence interval.

Figure 4. Counterfactual Changes in Productivity, by County



Notes: This map shows counties shaded according to their change in productivity from 1890 to the baseline counterfactual scenario: darker shades denote larger declines in productivity, and counties are divided into seven equal groups. This counterfactual sample includes all 2,722 counties that in 1890 report positive population and positive revenue (agriculture and/or manufacturing). The excluded geographic areas are cross-hashed. County boundaries correspond to county boundaries in 1890.

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

	Baseline Specification	Fixed 1860 Population	100-Mile Buffer Market Access	County-level Data Only: 1860 to 1900	1850 to 1900
	(1)	(2)	(3)	(4)	(5)
Panel A. Log Revenue					
Log Market Access	0.192	0.185	0.185	0.258	0.236
	(0.049)	(0.047)	(0.048)	(0.060)	(0.055)
Panel B. Log Capital Expenditure					
Log Market Access	0.158	0.151	0.153	0.225	0.209
	(0.053)	(0.051)	(0.052)	(0.060)	(0.054)
Panel C. Log Labor Expenditure					
Log Market Access	0.196	0.188	0.188	0.293	
	(0.061)	(0.059)	(0.060)	(0.068)	
Panel D. Log Materials Expenditure					
Log Market Access	0.183	0.176	0.176	0.243	
	(0.050)	(0.048)	(0.049)	(0.061)	
Panel E. Log Productivity					
Log Market Access	0.204	0.196	0.196	0.280	
	(0.051)	(0.049)	(0.050)	(0.057)	
Number of Counties	1,802	1,802	1,802	1,802	1,437
County-Year Obs.	5,406	5,406	5,406	9,010	8,622

Notes: Column 1 reports estimates from equation 8: for the indicated outcome variable in each panel, we report the estimated impacts of log market access (as defined in equation 1), controlling for county fixed effects, state-by-year fixed effects, and year-interacted cubic polymials in county latitude and longitude. The sample is a balanced panel of 1,802 counties (1860, 1870, 1880).

The outcome variables are: the log of total county manufacturing annual revenue (Panel A); the log of total county manufacturing annual expenditures on capital, labor, and materials (Panels B, C, D); and the log of total county manufacturing revenue minus the weighted logs of total county manufacturing expenditures on capital, labor, and materials (where those weights are the county's average revenue share for that input, and the variable is scaled by the ratio of average county revenue to average county productivity, as defined in equation 5).

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

Column 2 reports estimates using a measure of counties' market access in each decade that holds counties' population levels fixed at 1860 levels. Column 3 uses a measure of counties' market access only to counties beyond 100 miles of a county. Columns 4 and 5 use county-level data only, rather than county-by-industry data, which only affects the definition of Log Productivity in Panel E. Column 4 reports estimates for the 1860 to 1900 period, and Column 5 reports estimates for the 1850 to 1900 period using available data on county revenue and county capital expenditures in 1850.

Robust standard errors clustered by state are reported in parentheses.

Table 2. Impacts on County Productivity, Decomposed into TFPR and AE (Allocative Efficiency)

	Baseline Specification	Detailed Industry Groups	County-level Data Only 1860 to 1880	County-level Data Only 1860 to 1900	Adjusted Capital Misallocation: Zero	Adjusted Capital Misallocation: Materials
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Log County Productivity						
Log Market Access	0.204	0.204	0.204	0.280	0.207	0.206
	(0.051)	(0.051)	(0.051)	(0.057)	(0.049)	(0.049)
Panel B. County TFPR (Revenue Total Factor Productivity)						
Log Market Access	0.036	0.038	0.038	0.020	0.028	0.030
	(0.025)	(0.025)	(0.026)	(0.026)	(0.028)	(0.029)
Panel C. County AE (Allocative Efficiency)						
Log Market Access	0.168	0.166	0.166	0.259	0.179	0.176
	(0.051)	(0.052)	(0.054)	(0.066)	(0.052)	(0.053)
Number of Counties	1,802	1,802	1,802	1,802	1,802	1,802
County-Year Obs.	5,406	5,406	5,406	9,010	5,406	5,406

Notes: Column 1, panel A, corresponds to the estimate reported in Panel E of Column 1 in Table 1. Column 1 reports estimates from equation 8: for the indicated outcome variable in each panel, we report the estimated impacts of log market access (as defined in equation 1), controlling for county fixed effects, state-by-year fixed effects, and year-interacted cubic polymials in county latitude and longitude. The sample is a balanced panel of 1,802 counties (1860, 1870, 1880).

Panel A reports the estimated impacts on log county productivity (as in Panel E of Table 1), and Panels B and C report the impacts on productivity through changes in county TFPR or revenue total factor productivity (as defined in equation 6) and through changes in county AE or allocative efficiency (as defined in equation 7).

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

Column 2 calculates the outcome variables using county-by-industry data based on 193 industry categories, rather than the 31 industry categories used in column 1. Columns 3 and 4 calculate the outcome variables using county-level data, rather than county-by-industry data, for the same period from 1860 to 1880 (in column 3) and an extended period from 1860 through 1900 (in column 4). Column 5 calculates county productivity and its components when assuming zero misallocation in capital (such that the county's capital revenue share is equal to its cost share). Column 6 calculates county productivity and its components when assuming that capital misallocation is equal to materials misallocation (such that the county's capital revenue share is adjusted so the ratio of its cost share to revenue share is equal to the ratio of that county's materials cost share to materials revenue share).

Robust standard errors clustered by state are reported in parentheses.

Table 3. Impacts of Market Access, Controlling Flexibly for Local Railroad Construction

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Log County Productivity						
Log Market Access	0.204	0.218	0.199	0.197	0.175	0.142
	(0.051)	(0.056)	(0.057)	(0.058)	(0.058)	(0.056)
Panel B. County TFPR (Revenue Total Factor Productivity)						
Log Market Access	0.036	0.049	0.047	0.041	0.033	0.013
	(0.025)	(0.030)	(0.031)	(0.031)	(0.032)	(0.035)
Panel C. County AE (Allocative Efficiency)						
Log Market Access	0.168	0.169	0.152	0.156	0.143	0.129
	(0.051)	(0.055)	(0.056)	(0.055)	(0.055)	(0.054)
Additional Controls for:						
Any Railroad	No	Yes	Yes	Yes	Yes	Yes
Railroad Length	No	No	Yes	Yes	Yes	Yes
Railroad Length Polynomial	No	No	No	Yes	Yes	Yes
Railroads in Nearby Buffer	No	No	No	No	Yes	Yes
Railroads in Further Buffers	No	No	No	No	No	Yes
Number of Counties	1,802	1,802	1,802	1,802	1,802	1,802
County-Year Obs.	5,406	5,406	5,406	5,406	5,406	5,406

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

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

Table 4. Impacts of Market Access, Instrumenting with Baseline Access through Waterways

	Log Market Access	County Productivity		County TFPR Revenue Total Factor Productivity		County AE Allocative Efficiency	
	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log Market Access		0.222	0.204	0.041	0.036	0.192	0.168
		[0.030,0.437]	(0.051)	[-0.110,0.317]	(0.025)	[-0.0004,0.367]	(0.051)
Instruments:							
Log Water Market Access	-0.242						
in 1860 X year=1870	(0.052)						
Log Water Market Access	-0.430						
in 1860 X year=1880	(0.059)						
Kleibergen-Paap F statistic		26.90		26.90		26.90	
Number of Counties	1,802	1,802	1,802	1,802	1,802	1,802	1,802
County-Year Obs.	5,406	5,406	5,406	5,406	5,406	5,406	5,406

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

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

Table 5. Sources of Growth in County Allocative Efficiency (AE)

	Allocative Efficiency by Input	County Input Gap	County Input Wedge	County Input Cost Share	County Std. Dev. of Wedges
	(1)	(2)	(3)	(4)	(5)
Panel A. Capital					
Log Market Access	-0.004	0.001	0.022	-0.0005	-0.015
	(0.009)	(0.002)	(0.036)	(0.0006)	(0.044)
Panel B. Labor					
Log Market Access	0.066	-0.001	-0.050	-0.0013	-0.022
	(0.015)	(0.005)	(0.067)	(0.0033)	(0.044)
Panel C. Materials					
Log Market Access	0.107	0.012	-0.027	0.0017	0.032
	(0.049)	(0.006)	(0.039)	(0.0037)	(0.054)
Number of Counties	1,802	1,802	1,802	1,802	1,802
County/Year Obs.	5,406	5,406	5,406	5,406	5,406

Notes: For the indicated outcome variable, each column and panel reports the estimated impact of log market access from our baseline specification (as in column 1 of Table 1). Column 1 reports impacts on county productivity through changes in county allocative efficiency (as in Table 2, panel C, column 1) through changes in capital (panel A), labor (panel B), and materials (panel C). Column 2 reports impacts on county-level input gaps (defined as the input's cost share minus its revenue share in that decade), Column 3 reports impacts on county-level input wedges (defined as the input's cost share divided by its revenue share in that decade), and Column 4 reports impacts on county-level cost shares (defined as the national industry-level cost shares in each decade multiplied by the share of county revenue in each industry in that decade). Note that the county-level outcomes in columns 2 - 4 are defined in each decade, whereas our baseline definition of county productivity (and county TFPR and county AE) uses counties' average values over the sample period (or, in alternative specifications, counties' value in 1860 only). Column 5 reports impacts on counties' standard deviation of input wedges across industries in that county and decade.

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

Table 6. Impacts of Market Access on County Industries, Establishments, and Sector Shares

	Log Number of	Log Average Estab. Size:		Log Number of	County Manufacturing Share of:			
	Industries	Revenue / Estab.	Workers / Estab.	Establishments	Revenue	Value-Added	Surplus	Employment
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Market Access	0.029	0.021	0.025	0.171	0.0077	0.0002	-0.0017	0.0045
	(0.021)	(0.030)	(0.043)	(0.034)	(0.0078)	(0.0059)	(0.0090)	(0.0048)
Number of Counties	1,802	1,802	1,802	1,802	1,774	1,774	1,713	1,687
County/Year Obs.	5,406	5,406	5,406	5,406	5,322	5,322	5,139	5,061

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

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

Table 7. Counterfactual Impacts on National Aggregate Productivity

	Baseline: No Railroads	Restricted Railroad Networks: Only 1850 RRs	Only 1860 RRs	Only 1870 RRs	Only 1880 RRs	No Railroads, Extended Canals	All Railroads, Twice the Cost
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Counterfactual scenario, holding worker utility constant							
Change in Aggregate Productivity	-28.0%	-23.4%	-16.4%	-10.4%	-2.7%	-24.7%	-9.3%
Panel B. Counterfactual scenario, holding total population constant							
Change in Aggregate Productivity	-5.8%	-5.3%	-4.3%	-4.3%	-0.7%	-4.6%	-4.6%
Change in Utility	-34.4%	-28.6%	-19.3%	-19.3%	-3.1%	-30.5%	-30.5%

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

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

Table 8. Counterfactual Impacts on National Aggregate Productivity, Robustness

		Alternative Trade Elasticities:			Alternative Average Prices:				Use Materials Wedge for:	
	Baseline	Θ = 1.8	Θ = 3.6	Θ = 8.2	$\overline{P}$ = 20	$\overline{P}$ = 50	Efficient Agriculture	Efficient Capital	Capital	Capital and Labor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A. Fixed Worker Utility										
Change in Aggregate Productivity	-28.0%	-28.9%	-27.5%	-24.5%	-36.4%	-23.3%	-17.4%	-24.6%	-25.6%	-21.3%
Panel B. Fixed Total Population										
Change in Aggregate Productivity	-5.8%	-6.0%	-5.7%	-4.1%	-7.7%	-4.7%	-4.9%	-4.3%	-4.8%	-6.5%
Change in Utility	-34.4%	-35.8%	-33.5%	-30.0%	-43.9%	-28.9%	-34.6%	-34.4%	-34.4%	-34.6%

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 allows to relocate endogenously within the country. The sample includes all 2,722 counties that in 1890 report positive population and positive revenue (agriculture and/or manufacturing).

Columns 2 to 10 report robustness of our baseline estimates (in column 1) under alternative parameters. Our baseline estimates use our estimated value for Θ, the trade elasticity, of 2.788 with a 95% confidence interval between 1.815 and 3.556. In columns 2 and 3, we alternatively impose values for Θ of 1.815 or 3.556; in column 4, we impose a value of 8.22 from Donaldson and Hornbeck (2016). Our baseline estimates also use our estimated value for $\overline{P}$ of 35.7, 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 $\overline{P}$ of 20 or 50. In column 7, we reduce the estimated degree of input distortions in each county by assuming that the agricultural sector is efficient, and only apply our estimated manufacturing wedges to the county's manufacturing share of combined output across manufacturing and agriculture. Column 8 assumes that capital-use is efficient, such that there is zero gap for capital (or a wedge of 1). Columns 9 and 10 use the estimated materials wedge in each county to assign counties' capital wedge (Column 9) or capital wedge and labor wedge (Column 10).

Appendices

A Data Appendix

A.1 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 (United States Census Bureau, 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.[110] 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 193 categories that were more consistent across decades, and further grouped these industries into 31 categories. Our estimates are not sensitive to these industry groupings (Table 2), 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 Revenue (PQ). 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, United States Census Bureau, 1860*a*).

Manufacturing Materials Expenditure ($W^M X^M$). 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, United States Census Bureau, 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 Labor Expenditure ($W^L X^L$). 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

[110] In 1860 only, the Census of Manufactures also collected information for sectors outside of manufacturing (fisheries and mining) that we drop from our analysis for consistency.

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" (United States Census Bureau, 1860*a*).

Manufacturing Capital Expenditure ($W^K X^K$). 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 establishment's capital value was directed to include "capital invested in real and personal estate in the business" (United States Census Bureau, 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" (United States Census Bureau, 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" (United States Census Bureau, 1860*a*).

Civil War Related Industries. We coded two sets of industries as being "Civil War related." Our strict classification includes: artificial limbs and surgical appliances; awnings and tents; coffins; cutlery, edge tools, and axes; drugs; chemicals and medicines; explosives and fireworks; flags and banners; gun- and lock-smithing; gunpowder; lead; military goods; and ship and boat building. Our broad classification adds: bronze; canning and preserving; carriage and wagon materials; carriages and wagons; clothing (general); cooperage; gloves and mittens; and hats and caps.

A.2 Main Outcomes

The table below is a reference for the formulas used in calculating county productivity and its components. We use an upper bar to denote averages over the sample period. County-level values of revenue and input expenditures in each year reflect a sum of county-industry values in that year.

Component	Formula	Notes
Output	$P_{ct}Q_{ct}$	Gate value of revenue in the Census.
Capital	$W^K_{ct}X^K_{ct}$	Book value of capital in the Census, multiplied by interest rate.
Labor	$W^L_{ct}X^L_{ct}$	Wage bill in the Census.
Materials	$W^M_{ct}X^M_{ct}$	Gate value of materials in the Census.
s^k_{ct}	$\frac{W^k_{ct}X^k_{ct}}{P_{ct}Q_{ct}}$	Revenue share of input k in county c in year t, with $\overline{s^k_c}$ representing the average across years.
α^k_{ct}	$\sum_i \frac{P_{cit}Q_{cit}}{\sum_j P_{cjt}Q_{cjt}} \frac{\sum_c W^k_{cit}X^k_{cit}}{\sum_c \sum_\ell W^\ell_{cit}X^\ell_{cit}}$	County-level revenue share weighted sum of input's national industry cost share
ξ	$\frac{1}{1-\left(\frac{1}{C}\sum_c \sum_k \overline{s^k_c}\right)}$	Used for re-scaling percent growth in county output into percent growth in county productivity.
Productivity	$\xi\left[P_{ct}Q_{ct} - \sum_k \overline{s^k_c} \ln W^k_{ct}X^k_{ct}\right]$	
TFPR	$\xi\left[P_{ct}Q_{ct} - \sum_k \overline{\alpha^k_c} \ln W^k_{ct}X^k_{ct}\right]$	
Allocative Efficiency (AE)	$\xi\left[\left(\overline{\alpha^k_c} - \overline{s^k_c}\right) \ln W^k_{ct}X^k_{ct}\right]$	
Productivity Robustness: County Scalar	$\xi_c\left[P_{ct}Q_{ct} - \sum_k \overline{s^k_c} \ln W^k_{ct}X^k_{ct}\right]$	$\xi_c = \frac{1}{1-\left(\sum_k \overline{s^k_c}\right)}$ Drop counties with negative scalar values and top 1% of values
Productivity Robustness: Median Scalar	$\xi\left[P_{ct}Q_{ct} - \sum_k \widetilde{s^k_c} \ln W^k_{ct}X^k_{ct}\right]$	$\xi = \frac{1}{1-\left(\frac{1}{C}\sum_c \sum_k \widetilde{s^k}\right)}$ $\widetilde{s^k_c}$ is the median revenue share for k in county c, $\widetilde{s^k}$ is its national median
Productivity Robustness: 1860 Scalar	$\xi\left[P_{ct}Q_{ct} - \sum_k s^k_{c1860} \ln W^k_{ct}X^k_{ct}\right]$	$\xi = \frac{1}{1-\left(\frac{1}{C}\sum_c \sum_k s^k_{c1860}\right)}$ We also use α^k_{c1860} for decomposition into TFPR and AE

A.3 Other County-Level Data

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

Other county-level data are from the United States 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 is estimated to vary 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 (Hacker, 2013). Agricultural land value is defined as the total value of land in farms, including the value of farm buildings and improvements.[111]

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 United States 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 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.

The regression sample is 1,802 counties that report county-industry manufacturing data in 1860, 1870, and 1880 (see Figure 2). The counterfactual sample is 2,722 counties with positive population and positive agricultural or manufacturing revenue in 1890 (see Figure 4).

[111]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).

B Results Appendix

B.1 Robustness: Measurement of Productivity

There may be various sources of measurement error in our calculation of county productivity and its decomposition into county TFPR (revenue total factor productivity) and county AE (allocative efficiency). Appendix Table 2 shows how our regression estimates differ under a variety of alternative approaches for calculating county productivity and its decomposition, and in this section we describe the specifications in that table.

Table 2 showed that our estimates were similar when replacing the observed capital wedge in each county with the materials wedge, motivated by De Loecker and Warzynski (2012) who use the materials wedge to measure the common component of distortions across inputs. In Row 5 of Appendix Table 2, we additionally replace the labor wedge with the materials wedge and estimate similar percent effects of market access.

As described in Section IV, in rows 2, 3, and 4 we change how we measure the flow of capital services to measure sensitivity to measurement error in capital. Row 2 doubles the baseline capital interest rate and Row 3 triples it. One additional motivation for this adjustment could be that the nominal mortgage rate reflects a different depreciation rate for housing than the relevant one for manufacturing equipment. Estimated depreciation rates for equipment in this historical era are around 6% (Davis and Gallman, 2019), which is in the range of modern values estimated by the BEA (although at the lower end, perhaps because various high-depreciation capital inputs for modern factories did not exist at the time, such as internal combustion engines and computers). Row 4 shows results using the average national interest rate instead of state-specific rates.

Given that the measurement of reallocation gains can be sensitive to the upper and lower tails (Rotemberg and White, 2020), in rows 6–11 we test the sensitivity of our results to lowering dispersion in input distortions. In rows 6–8, we shrink dispersion in the capital wedge by 5 percent, 10 percent, and 25 percent.[112] While this adjustment mechanically lowers the potential gains from reallocation in exercises like those in Hsieh and Klenow (2009), our regression estimates are stable. In rows 9–11, we shrink dispersion in all of the input wedges, which also has little effect on our regression estimates.[113]

Our main estimates hold fixed counties' revenue shares and production function elasticities (and therefore the wedges), using the observed average from 1860 to 1880 (as in Petrin

[112] In particular, we replace the observed wedge for capital with the weighted average of the observed capital wedge and its national median, where the weights on the county's observed values are respectively 95 percent, 90 percent, and 75 percent. We then impute consistent values for capital (so the relationship between imputed capital and observed output is consistent with the imputed wedge).

[113] For the counterfactuals, replacing the observed wedges with their shrunk counterparts never affects the welfare losses from removing the railroads by more than 1 percentage point.

and Levinsohn 2012's Törnqvist-Divisia approximation). In row 12, we instead hold these values fixed at their 1860 levels (and correspondingly use those values to calculate the scaling factor). In row 13, in addition to using these 1860 values in the measurement of county productivity in each decade, based on county revenue and county input expenditures in each decade, we hold county populations fixed at their 1860 values for calculating market access (as in Table 1, column 2).

In rows 14 and 15, we use alternative scaling factors in the definition of county productivity. The percent impact of market access on county productivity reflects a scaled percent impact of market access on revenue and input expenditures, as discussed in Section V.C, and our baseline estimates define this scaling factor as the average ratio of county output to county productivity over the 1860-1880 period. Alternatively, in row 14, we define this scaling factor based on the median ratio of county revenue to county productivity. In row 15, we use county-specific scaling factors instead of the national average (dropping counties with negative values and the top 1% of values, as these scaling factors become undefined as productivity approaches zero).

As an alternative approach to dealing with extreme values, we show similar impacts of market access when excluding sample counties with the largest and smallest changes in productivity from 1860 to 1880: row 16 excludes the top and bottom 1% of counties, and row 17 excludes the top and bottom 5% of counties.

Establishment inputs may be under-reported in the Census of Manufactures, which would over-state establishments' productivity. For example, Census enumerators ask establishments for their labor costs and these were intended to include in-kind boarding costs and labor supplied by establishment owners working on their own account, but there has been debate about whether establishment owner labor is fully reflected in these costs (Weeks, 1886; Atack, 1977; Sokoloff, 1984; Margo, 2014). Row 18 reports similar impacts of county market access, though, when inflating labor costs to reflect potentially omitted labor costs.[114]

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 decomposition of county productivity into TFPR and AE. In row 19 we assume decreasing returns to scale and re-scale the production function elasticities to add up to 0.95, following the low estimates of Lafortune et al. (2021); in row 20 we correspondingly assume the returns to scale are 1.05. Assuming decreasing returns to scale increases the impact on county productivity through TFPR growth (column 2) and reduces the impact through AE growth (column 3), while mechanically leaving unchanged the impact on county productivity

[114]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.

(column 1). The general view in the literature is that historical manufacturing firm returns to scale were roughly constant, with evidence for decreasing returns or increasing returns depending on adjustments to measured inputs (see, e.g., Atack, 1977; Sokoloff, 1984; Margo, 2014), which is also the view for modern manufacturing (Blackwood et al., Forthcoming).

We also show the sensitivity of our results to alternative methods for calculating the production function elasticities. These adjustments mechanically have no effect on the measurement of county productivity (in column 1) and, in practice, have little substantive effect on the estimated impacts through county TFPR (column 2) and county AE (column 3) because market access had similar percent effects on each input. In row 21, we calculate county-level elasticities averaging industry-level cost shares with weights equal to an industry's share of total expenditure in that county, rather than an industry's share of total revenue in that county (which could over-weight the influence of high-markup sectors). In row 22, we calculate production function elasticities only using the "most efficient" counties for calculating each industry's cost shares (specifically, those counties whose total input expenditure gap is within one standard deviation of zero). In row 23, instead of using national values for calculating industry cost shares, we leave out each specific county-industry when calculating its production function elasticity. In row 24, we instead use only other counties located in the same state to calculate industry-level cost shares, and in row 25 we leave out that county in the calculation of state-level industry cost shares. In row 26, we use each observed county-industry cost share as our measure of production function elasticities, which imposes a constant wedge across inputs within the county.

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, row 27 reports estimates when increasing the cost share for labor by 5 percentage points and proportionally decreasing the cost shares for materials and capital. Similarly, we increase the cost share for materials by 5 percentage points (in row 28) or for capital by 5 percentage points (in row 29).

We measure manufacturing productivity using data from the Census of Manufactures, though there may be additional manufacturing activity not included in the Census of Manufactures. The Census of Manufactures 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.[115] If we exclude the butter and cheese industry in each decade, though, we estimate similar impacts of market access (row 30).

[115] In 1880, the Census of Manufactures 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, 2020). The Census of Agriculture reports quantities of butter and cheese produced, but does not report data on their values or associated inputs.

The Census of Manufactures would have missed some manufacturing establishments, particularly smaller establishments (United States Census Bureau, 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 in Table 6 that changes in market access are not associated with changes in average establishment 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 Manufactures.[116] The national value of home manufactures is only 1.2% of total manufacturing revenue in 1860 and 0.4% of total manufacturing revenue in 1870, summing output values from the Census of Agriculture and the Census of Manufacturing for our sample counties, though home manufactures are substantively important in some counties. Among our sample counties, the median revenue shares of home manufacturing are 3.6% in 1860 and 1.0% in 1870 and the average revenue shares of home manufacturing are 11.8% in 1860 and 6.5% in 1870.[117] The estimated impact of county market access on the value of manufacturing output is 0.192 (0.049), using only data from the Census of Manufactures, and the effect is 0.160 (0.043) when adding the value of home manufactures to data from the Census of Manufactures. Data on home manufactures also allow us to expand the balanced sample of counties to include 149 additional counties that do not report manufacturing revenue in at least one decade of the Census of Manufactures, but do report home manufactures in that decade. The estimated impact of county market access on county manufacturing revenue increases to 0.269 (0.041) when including these 149 additional counties and adding the value of home manufactures to data from the Census of Manufactures.

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 includes 91% of population and 99% of manufacturing revenue 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 4 percentage point in-

[116]For this analysis, we assume that home manufactures are not already included in the Census of Manufactures. Spot checks of the Census of Agriculture manuscripts show the values of home manufactures tend to be substantially less than the $500 threshold used by the Census of Manufactures.

[117]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.

crease in the probability that a county reports any manufacturing activity in the Census of Manufactures. 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.

B.2 Robustness: Regional Shocks

Our estimated effects of county market access on county productivity may be influenced by other omitted sources of growth in county productivity that are correlated with relative changes in county market access. In Appendix Table 3, we show how our regression estimates differ under a variety of alternative specifications that look to control for these omitted variables or adjust our sample definition.

An expanding national railroad network affected different counties' market access from 1860 to 1870 and from 1870 to 1880, and over each decade there were similar effects of market access on county productivity. Splitting our baseline analysis by decade pair (1860 and 1870, 1870 and 1880, 1860 and 1880), increases in county market access lead to substantial increases in county productivity that are more often driven by increases in county AE than by increases in county TFPR (Rows 2, 3, 4).[118] These separate estimates also avoid potential issues with interpreting two-way fixed effects models with multiple time periods (De Chaisemartin and d'Haultfoeuille, 2020).

In row 5, we extend the specification from column 3 of Table 3 (controlling for any railroad and railroad length) to include future values of market access and railroads. These additional controls do not substantively change the measured effect of contemporaneous market access, and there is no systematic effect of future market access on current outcomes (i.e., growth in county market access was not associated with differential trends in county manufacturing activity over the prior decade).[119]

Rows 6–17 control for county characteristics in 1860, interacted with year, to flexibly allow for those characteristics to have time-varying influences on county productivity. Rows 6 and 7 control for 1860 market access without railroads and actual 1860 market access. Rows 8 and 9 control for the share of counties' 1860 revenue in each industry, interacted with year, given the potential for relative changes in industry output prices and input prices or other industry-

[118] While the decomposition puts more weight on TFPR growth than AE growth in the 1870-1880 period, we estimate that AE growth is the dominant component in the 1870-1900 period when using county-level data and extending the sample, so we hesitate to over-interpret the statistically insignificant differences between these decade pairs.

[119] From estimating this specification on an extended sample back to 1850, using the available county-level data on revenue from 1850, we also estimate no significant effect of future market access on manufacturing revenue growth from 1850 to 1860. Similarly, we estimate no significant effect on revenue growth (or productivity growth) from 1860 to 1870 from market access growth from 1870 to 1880.

specific shocks to differentially impact counties' growth. Rows 10 to 17 control for other county characteristics in 1860: counties' input-specific gaps in 1860 (the difference between the output elasticities and the revenue shares), the 1860 production function elasticities; 1860 input revenue shares; 1860 input wedges (the ratio of the output elasticities to the revenue shares); the 1860 HHIs of manufacturing industry revenue and employment; whether a county was on the "frontier" in 1860 (following the definition from Bazzi, Fiszbein and Gebresilasse, 2020); and jointly controlling for the gaps, elasticities, wedges, HHIs, and frontier status.[120]

As the Civil War occurred within our sample period, and had substantially different implications for different areas of the country, we explore the sensitivity of our results to adjusting for differential impacts of the Civil War and the abolition of slavery. First, counties that were initially concentrated in industries that produced more war-related goods may have changed differently over this period even in the absence of changes in the railroad network. For row 18, before calculating county productivity, we drop all industries strictly related to war production and in row 19 we drop industries more broadly related to war production (as defined in the Section A). Alternatively, in rows 20 and 21, we instead control for the 1860 share of revenue in war-related sectors under each definition. These adjustments do not have large effects on the estimated coefficients. The Civil War itself may have had a direct effect on outcomes, and row 22 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.[121] Row 23 instead drops all counties with battles with over 500 casualties, while row 24 drops all counties with any noted battle. Row 25 drops counties on the border of the Union and the Confederacy. Row 26 drops Confederate states, row 27 drops slave states, and row 28 drops the Southern region.[122] The estimates are stable across these sample changes.

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 rows 29 and 30, 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

[120]The share of counties' 1860 revenue in each industry, the HHIs, and frontier status are moderately predictive of 1860 county gaps (a within R-squared of 0.15, after conditioning on state fixed effects and latitude/longitude, which mostly reflects the influence of the 31 industry shares).

[121]Our battle data includes all battlefields identified as significant by the Civil War Sites Advisory Commission's Report on the NationâĂŹs Civil War Battlefields.

[122]When excluding areas from the regression sample, we continue to include them in the measurement of other counties' market access.

boundaries.[123]

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, we find similar estimates when controlling for fifth-order polynomials (row 31) or linear functions of counties' latitude and longitude (row 32). We also estimate similar impacts of market access when excluding from our sample the Plains and West Coast regions of the United States in row 33.[124]

B.3 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. In Appendix Table 4, we show how the estimated effects of county market access on county productivity (and TFPR and AE) differ under a variety of alternative approaches for calculating market access. In this section, we describe the specifications in that table.

In rows 2 and 3, we show that the estimated impacts on county productivity are not sensitive to omitting counties with the largest and smallest changes in market access. The estimated impact on county productivity through TFPR growth becomes moderately larger when omitting more counties (in row 3), but the estimates are not statistically different.

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, waterways, and wagons. 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 (see, e.g., Gross (2020) on North-South gauge differences).

In calculating counties' market access, the "iceberg trade costs" (τ_{cd}) reflect the measured county-to-county transportation costs (t_{cd}) scaled by the average price per ton of transported goods ($\tau_{cd} = 1 + t_{cd}/\overline{P}$). Our baseline estimates use an average price of 35.7 that best fits the data, but rows 10 and 11 report similar estimated effects of market access if we instead use 20 or 50 dollars per ton.[125]

[123] We assign each county the share of its area in each region or subregion, and control for those shares interacted with year fixed effects.

[124] The West Coast sample states are California, Oregon, and Washington. The Plains sample states are Kansas, Nebraska, and Texas.

[125] We continue to focus on the estimated impact of a one standard deviation greater increase in county market access, as market access itself is re-scaled when changing the transportation cost parameters.

The impact of market access is also similar when using alternative values of θ, the trade elasticity. Smaller values of θ compress the distribution of changes in market access, just as for larger average prices, but increase the effect of gaining access such that there remains a similar effect from a one standard deviation greater increase in market access. Rows 12 to 14 show similar results using the extremes of the 95% confidence interval around our baseline estimate of θ (1.815 to 3.556) and a larger value of θ (8.22) from Donaldson and Hornbeck (2016).

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 (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 by Hacker (2013). Our baseline measure of market access considers counties' access to population, 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).

C 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.

C.1 Market Access and Productivity

As described in equation 11, trade flows follow a gravity equation:

$$E_{od} = \kappa_1 A_o \left(\Pi_k \left(\left(1+\psi_o^k\right) W_o^k\right)^{\alpha_o^k}\right)^{-\theta} \tau_{od}^{-\theta} Y_d P_d^{\theta}, \tag{21}$$

where the price index in equation 12 is:

$$CMA_d = P_d^{-\theta} = \kappa_1 \sum_o \tau_{od}^{-\theta} A_o \left(\Pi_k \left(\left(1+\psi_o^k\right) W_o^k\right)^{\alpha_o^k}\right)^{-\theta}. \tag{22}$$

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 21 over all counties and taking logs gives:[126]

$$\ln\left(Y_o\right) = \kappa_1 + \varkappa_{1o} + \left(\alpha_o^M + \alpha_o^L\right) \ln\left(P_o^{-\theta}\right) - \theta\alpha_o^T \ln q_o + \ln\left(\sum_d \left(\frac{P_d}{\tau_{od}}\right)^{\theta} Y_d\right). \tag{23}$$

Plugging $P_d^{-\theta}$ from equation 22 into equation 23, and combining terms, gives:[127]

$$\ln Y_o = \kappa_1 + \varkappa_{2o} + \left(\frac{\alpha_o^M + \alpha_o^L + 1}{1+\theta\alpha_o^T}\right) \ln\left(P_o^{-\theta}\right). \tag{24}$$

Once we have solved for output in each county, input quantities and input expenditures 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 and other distortions 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

[126] $\kappa_1 = \left(-\frac{\theta}{1-\sigma}\right) \ln\left(\Gamma\left(\frac{\theta+1-\sigma}{\theta}\right)\right)$ and $\varkappa_{1o} = \ln\left(A_o\right) - \theta\alpha_o^L \ln\left(\left(1+\psi_o^L\right)\bar{U}\right) - \theta\alpha_o^K \ln\left(\left(1+\psi_o^K\right) r\right) - \theta\alpha_o^M \ln\left(1+\psi_o^M\right)$

[127] $\varkappa_{2o} = \frac{\varkappa_{1o} + \ln\xi_o - \theta\alpha_o^T \ln\frac{\alpha_o^T}{T_o}}{1+\theta\alpha_o^T}$

access, as described in equation 18:

$$\frac{d \ln q_o}{d \ln MA_o} = \frac{\alpha_o^M + \alpha_o^L + 1}{1 + \theta \alpha_o^T}. \tag{25}$$

The local prices for labor (w_o) and materials (P_o) are log-linear in market access, as described by equation 22:

$$\frac{d \ln w_o}{d \ln MA_o} = \frac{d \ln P_o}{d \ln MA_o} = -\frac{1}{\theta}. \tag{26}$$

The impact of market access on productivity, equation 17, is then given by substituting into equation 16 the impacts of log market access on log real inputs, and converting from the gap to the wedge:

$$\begin{aligned} \frac{d \ln PR_o}{d \ln MA_o} = \quad & \frac{P_o Q_o}{Pr_o} \quad \Bigg[\left(\alpha_o^L - s_o^L\right) \quad \left(\frac{1}{\theta} + \frac{\alpha_o^M + \alpha_o^L + 1}{1 + \theta \alpha_o^T}\right) \\ & + \quad \left(\alpha_o^M - s_o^M\right) \quad \left(\frac{1}{\theta} + \frac{\alpha_o^M + \alpha_o^L + 1}{1 + \theta \alpha_o^T}\right) \\ & + \quad \left(\alpha_o^K - s_o^K\right) \quad \left(\frac{\alpha_o^M + \alpha_o^L + 1}{1 + \theta \alpha_o^T}\right) \Bigg]. \end{aligned} \tag{27}$$

C.2 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 (θ) and the average price of traded goods ($\overline{P}$), we estimate using supplemental data.

Our general approach to estimating counterfactuals is to rationalize the observed distribution of population by estimating each county's technical efficiency (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:

$$P_d^{-\theta} = \sum_o \frac{\tau_{od}^{-\theta} P_o \frac{\left(1+\psi_o^L\right) L_o}{\alpha_o^L}}{\sum_i \tau_{oi}^{-\theta} P_i^{1+\theta} \frac{\left(1+\psi_i^L\right) L_i}{\alpha_i^L}}. \tag{28}$$

This equation matches equation 15 in the appendix to Donaldson and Hornbeck (2016), with two changes: we allow for market distortions $\left(\psi_o^L\right)$, and we allow variation across counties in the production function elasticity for labor $\left(\alpha_o^L\right)$. 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:[128]

(29)
$$C_i \equiv A_i T_i^{\theta\alpha_i^T} r_i^{-\theta\alpha_i^K} \alpha_i^{T-\theta\alpha_i^T} \left[\Gamma\left(\frac{\theta+1-\sigma}{\theta}\right)\right]^{-\frac{\theta}{1-\sigma}} \left(1+\psi_i^L\right)^{-\theta\alpha_i^L} \left(1+\psi_{ki}\right)^{-\theta\alpha_i^K} \left(1+\psi_{mi}\right)^{-\theta\alpha_i^M}.$$

We can then rewrite equation 11 as

$$C_o \sum_i \tau_{oi}^{-\theta} P_i^{1+\theta} \frac{\left(1+\psi_i^L\right) L_i}{\alpha_i^L} = P_o^{1+\theta\left(\alpha_o^L+\alpha_o^T+\alpha_o^M\right)} \left(\frac{\left(1+\psi_o^L\right) L_o}{\alpha_o^L}\right)^{1+\theta\alpha_o^T} \bar{U}^{\theta\left(\alpha_o^L+\alpha_o^T\right)}, \tag{30}$$

and re-write equation 12 as

$$P_o^{-\theta} = \sum_i C_i \bar{U}^{-\theta\left(\alpha_i^L+\alpha_i^T\right)} P_i^{-\theta\left(\alpha_i^L+\alpha_i^T+\alpha_i^M\right)} \left(\frac{\left(1+\psi_i^L\right) L_i}{\alpha_i^L}\right)^{-\theta\alpha_i^T} \tau_{oi}^{-\theta}. \tag{31}$$

To solve this system above, we define:

$$\phi_o \equiv \frac{\left(P_o^{-\theta} C_o\right)}{\left(\frac{(1+\psi_o^L)BL_o}{\alpha_o^L}\right)^{1+\theta\alpha_o^T} \bar{U}^{\theta(\alpha_o^L+\alpha_o^T)} P_o^{1+\theta(\alpha_o^L+\alpha_o^T+\alpha_o^M)}}. \tag{32}$$

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

We now have the equations:

$$1 = \sum_i \frac{C_o}{P_o^{1+\theta(\alpha_o^L+\alpha_o^T+\alpha_o^M)} \left(\frac{(1+\psi_o^L)BL_o}{\alpha_o^L}\right)^{1+\theta\alpha_o^T} \bar{U}^{\theta(\alpha_o^L+\alpha_o^T)}} \tau_{oi}^{-\theta} P_i^{1+\theta} \left(\frac{1+\psi_i^L}{\alpha_i^L}\right) BL_i \tag{33}$$

and

$$\phi_o = \sum_i \frac{C_o}{P_o^{1+\theta(\alpha_o^L+\alpha_o^T+\alpha_o^M)} \left(\frac{(1+\psi_o^L)BL_o}{\alpha_o^L}\right)^{1+\theta\alpha_o^T} \bar{U}^{\theta(\alpha_o^L+\alpha_o^T)}} \tau_{oi}^{-\theta} P_i^{1+\theta} \left(\frac{1+\psi_i^L}{\alpha_i^L}\right) BL_i \phi_i. \tag{34}$$

[128] We will be able to solve for C_i using the factual data, our assumption on interest rates, and our estimated value of θ.

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 λ and a unique (up to proportionality) eigenvector $\hat{\phi}$ with all its elements positive. We define the matrix A with elements $[A_{od}] = \left[\frac{C_o}{P_o^{1+\theta\left(\alpha_o^L+\alpha_o^T+\alpha_o^M\right)} L_o^{1+\theta\alpha_o^T}} \tau_{od}^{-\theta} P_d^{1+\theta} \left(\frac{1+\psi_d^L}{\alpha_d^L}\right) L_d\right]$. For any possible values of C_o, P_o, N_o, and τ_{od}, the matrix A has all of its elements positive. Perron-Frobenius therefore implies that, for any values of P_o, N_o, τ_{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 33 and 34 holds, we have two eigenvectors. The two eigenvectors are ϕ_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 ϕ^{-1}. Thus, we know that

$$P_o^{1+\theta\left(1+\alpha_o^T+\alpha_o^L+\alpha_o^M\right)} = \phi L_o^{-\left(1+\theta\alpha_o^T\right)} C_o. \tag{35}$$

We then solve for P_o in equation 35, and plug into equation 30 to get:

$$\begin{aligned} C_o \sum_i \tau_{oi}^{-\theta} \left(\phi \bar{U}^{-\theta\left(\alpha_i^L+\alpha_i^T\right)} \left(\frac{\left(1+\psi_i^L\right) BL_i}{\alpha_i^L}\right)^{-\left(1+\theta\alpha_i^T\right)} C_i\right)^{\frac{1+\theta}{1+\theta\left(1+\alpha_i^T+\alpha_i^L+\alpha_i^M\right)}} \frac{\left(1+\psi_i^L\right) BL_i}{\alpha_i^L} = \\ \left(\phi \bar{U}^{-\theta\left(\alpha_o^L+\alpha_o^T\right)} \left(\frac{\left(1+\psi_o^L\right) BL_o}{\alpha_o^L}\right)^{-\left(1+\theta\alpha_o^T\right)} C_o\right)^{\frac{1+\theta\left(\alpha_o^L+\alpha_o^T+\alpha_o^M\right)}{1+\theta\left(1+\alpha_o^L+\alpha_o^T+\alpha_o^M\right)}} \\ \times \left(\frac{\left(1+\psi_o^L\right) BL_o}{\alpha_o^L}\right)^{1+\theta\alpha_o^T} \bar{U}^{\theta\left(\alpha_o^L+\alpha_o^T\right)} \end{aligned} \tag{36}$$

Rearranging and combining like terms, we get:

$$\begin{aligned} \left(\phi \bar{U}^{-\theta\left(\alpha_o^L+\alpha_o^T\right)}\right)^{-\frac{1+\theta\left(\alpha_o^L+\alpha_o^T+\alpha_o^M\right)}{1+\theta\left(1+\alpha_o^L+\alpha_o^T+\alpha_o^M\right)}} \bar{U}^{-\theta\left(\alpha_o^L+\alpha_o^T\right)} C_o^{\frac{\theta}{1+\theta\left(1+\alpha_o^L+\alpha_o^T+\alpha_o^M\right)}} * \\ \sum_i \tau_{oi}^{-\theta} \left(\phi \bar{U}^{-\theta\left(\alpha_i^L+\alpha_i^T\right)} C_i\right)^{\frac{1+\theta}{1+\theta\left(1+\alpha_i^T+\alpha_i^L+\alpha_i^M\right)}} \left(\frac{\left(1+\psi_i^L\right) BL_i}{\alpha_i^L}\right)^{1-\frac{\left(1+\theta\alpha_i^T\right)(1+\theta)}{1+\theta\left(1+\alpha_i^T+\alpha_i^L+\alpha_i^M\right)}} \\ = \left(\frac{\left(1+\psi_o^L\right) BL_o}{\alpha_o^L}\right)^{\left(1+\theta\alpha_o^T\right)\left(1-\frac{1+\theta\left(\alpha_o^L+\alpha_o^T+\alpha_o^M\right)}{1+\theta\left(1+\alpha_o^L+\alpha_o^T+\alpha_o^M\right)}\right)}. \end{aligned} \tag{37}$$

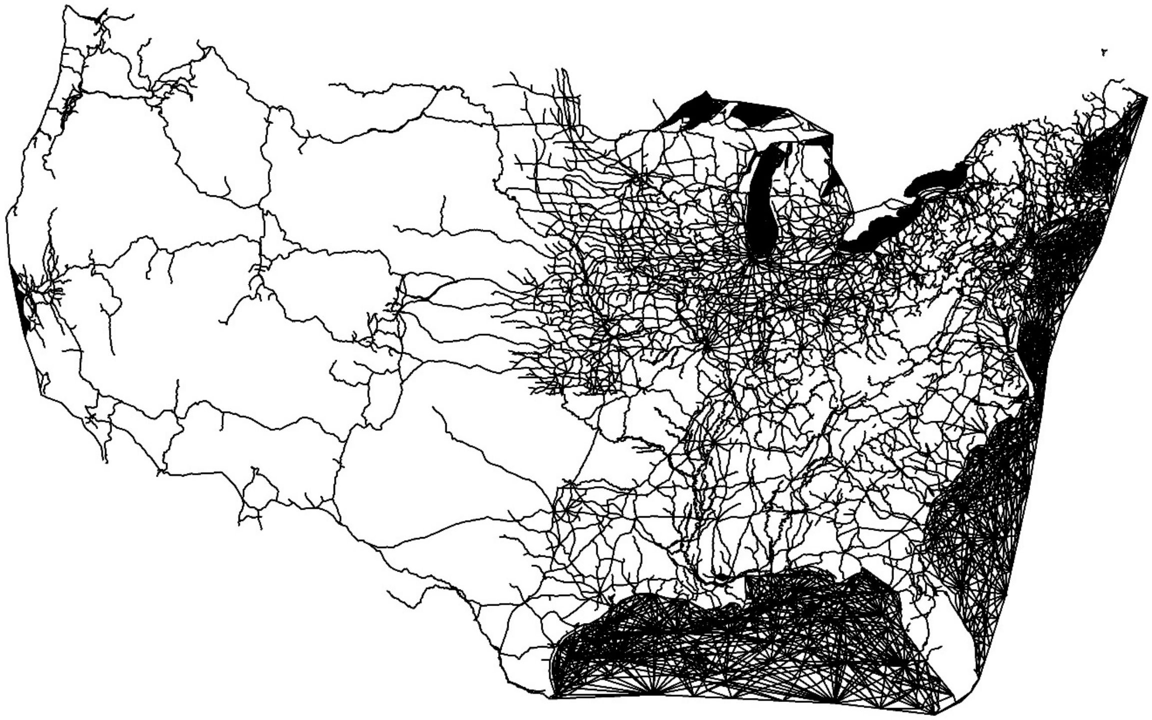
We know that there is a unique combination of ϕ and B that solves equation 37. To find that solution, we grid-search over the parameter space. For each initial guess, we use the Fujimoto-Krause algorithm to solve for the distribution of population (L_i) and we pick the parameters for which equation 37 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 θ and $\overline{P}$, we estimate how much production inputs would have changed in each county given a different vector of costs τ_{od} (e.g., without the railroads) and 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).[129] The counterfactual impact on national aggregate productivity is then given by the Domar-weighted sum of these counterfactual changes in county production inputs multiplied by the county-level gap for that input (equation 20).

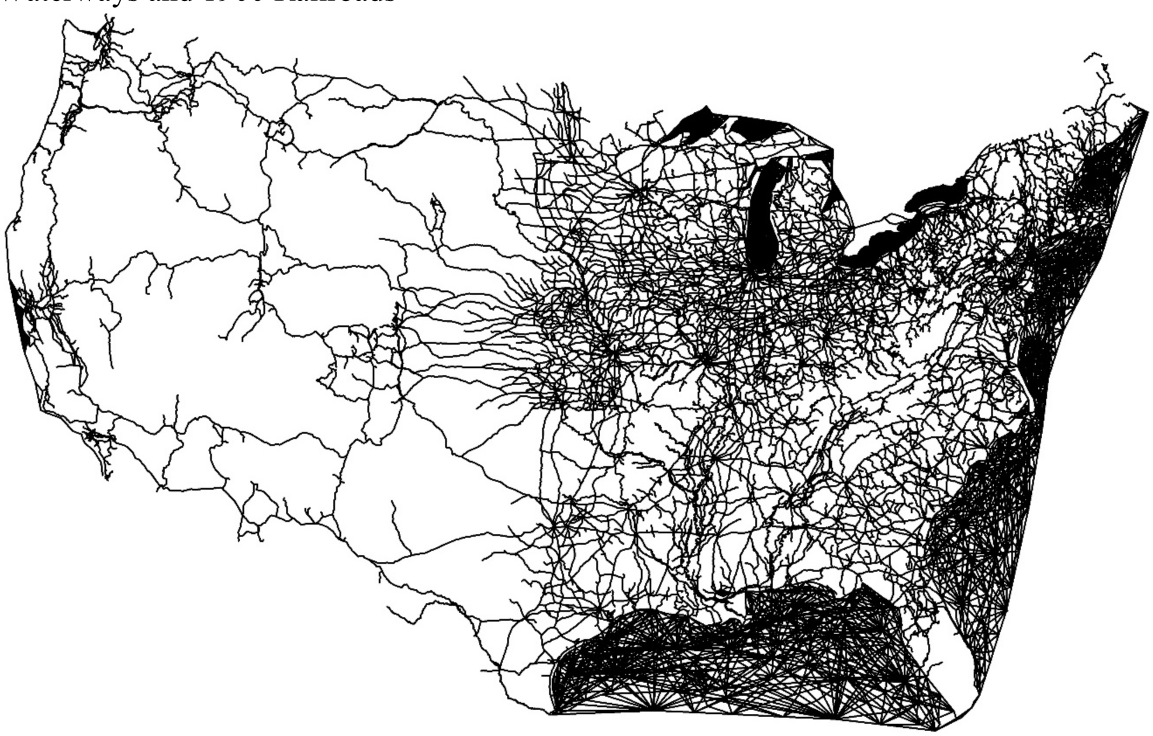
[129] Note that relative population levels are not independent of total population levels, unlike in Donaldson and Hornbeck (2016), because the production function elasticities vary over space but quantitatively this effect is relatively small.

Appendix Figure 1. Waterways and Railroads, 1890 and 1900

A. Waterways and 1890 Railroads



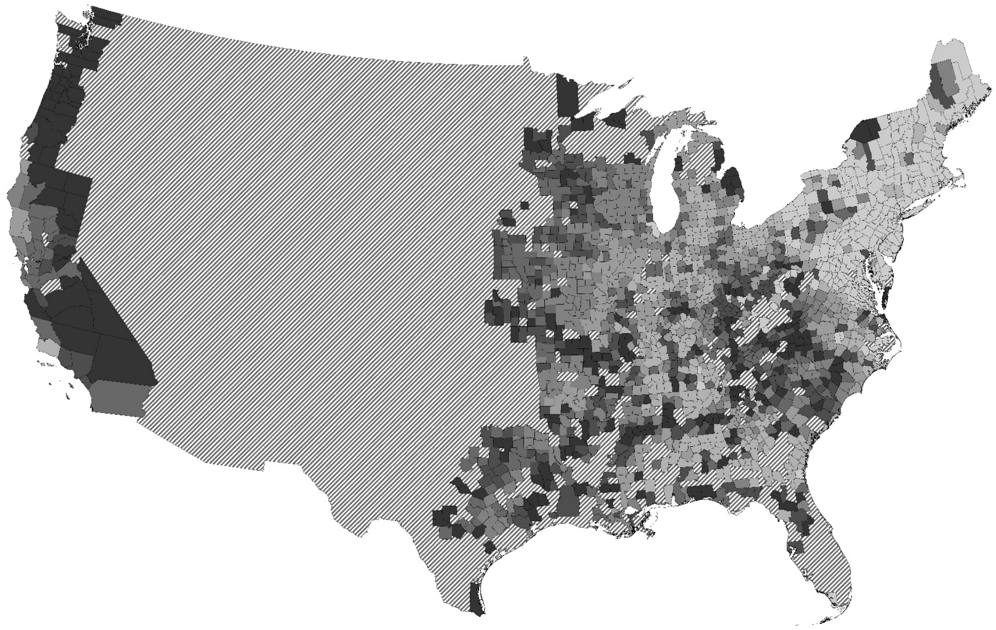
B. Waterways and 1900 Railroads



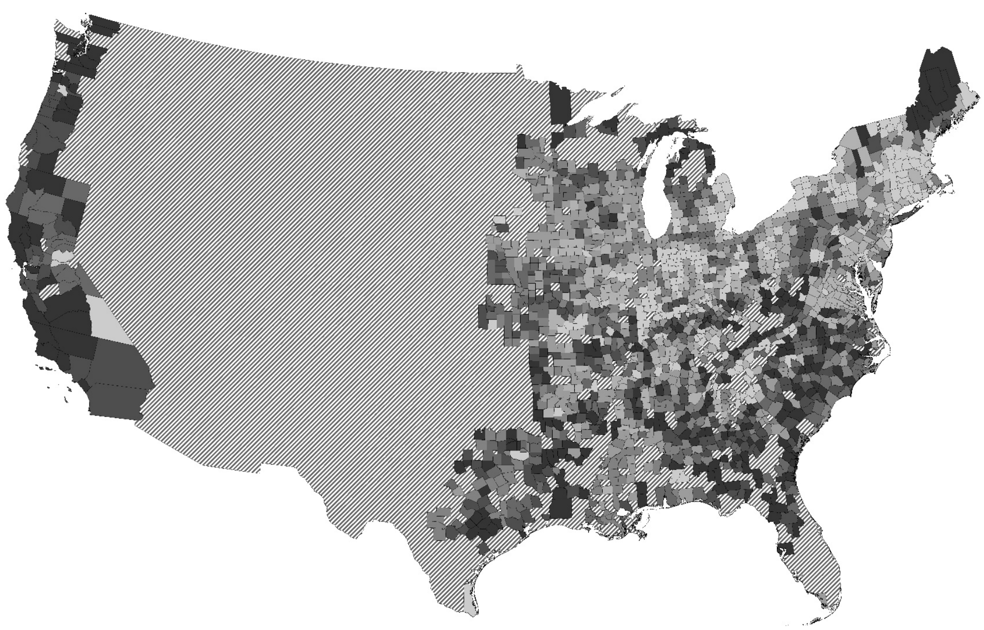
Notes: Similar to Figure 1, Panel A shows the railroads constructed by 1890, as well as the natural waterways (including navigable rivers, lakes, and oceans) and constructed canals. Panel B adds railroads constructed between 1890 and 1900.

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

A. From 1880 to 1890



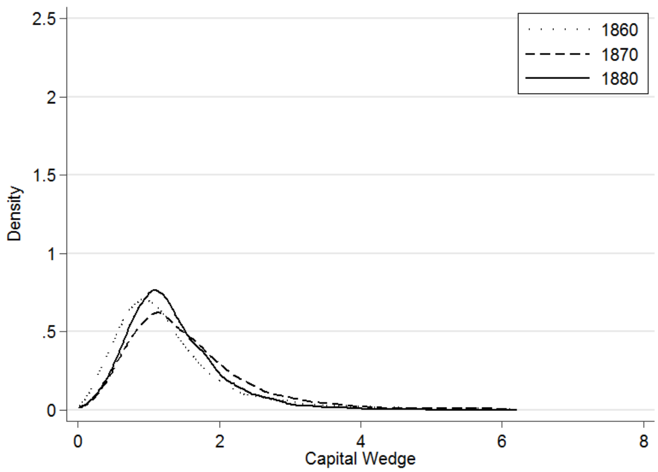
B. From 1890 to 1900



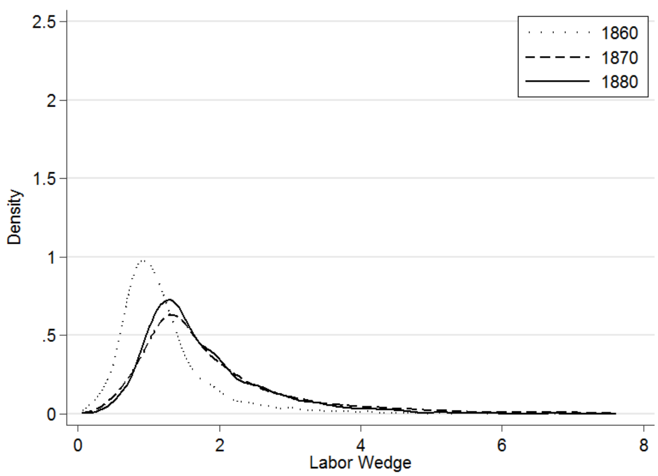
Notes: In each panel, counties are shaded according to their calculated change in market access from 1880 to 1890 (Panel A) and from 1890 to 1900 (Panel B). Counties are divided into seven groups (with an equal number of counties per group), and darker shades denote larger increases in market access. These maps include the 1,802 sample counties in the regression analysis, which are all counties that report non-zero manufacturing activity from 1860, 1870, and 1880. The excluded geographic areas are cross-hashed. County boundaries correspond to county boundaries in 1890.

Appendix Figure 3. Cross-County Dispersion in Input Wedges, by Decade

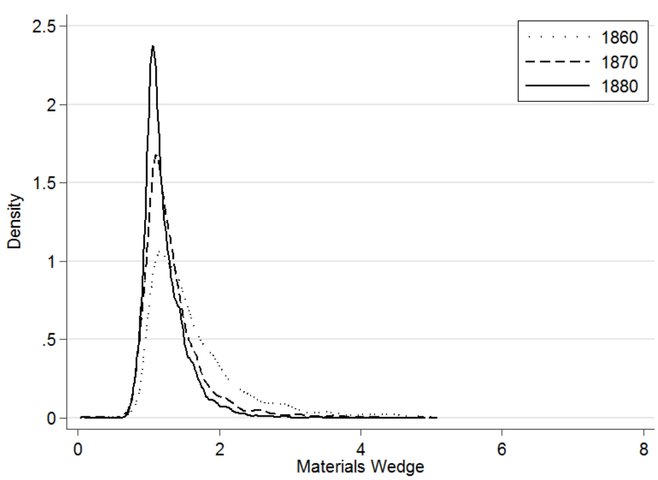
A. Cross-County Dispersion in Capital Wedges, by Decade



B. Cross-County Dispersion in Labor Wedges, by Decade

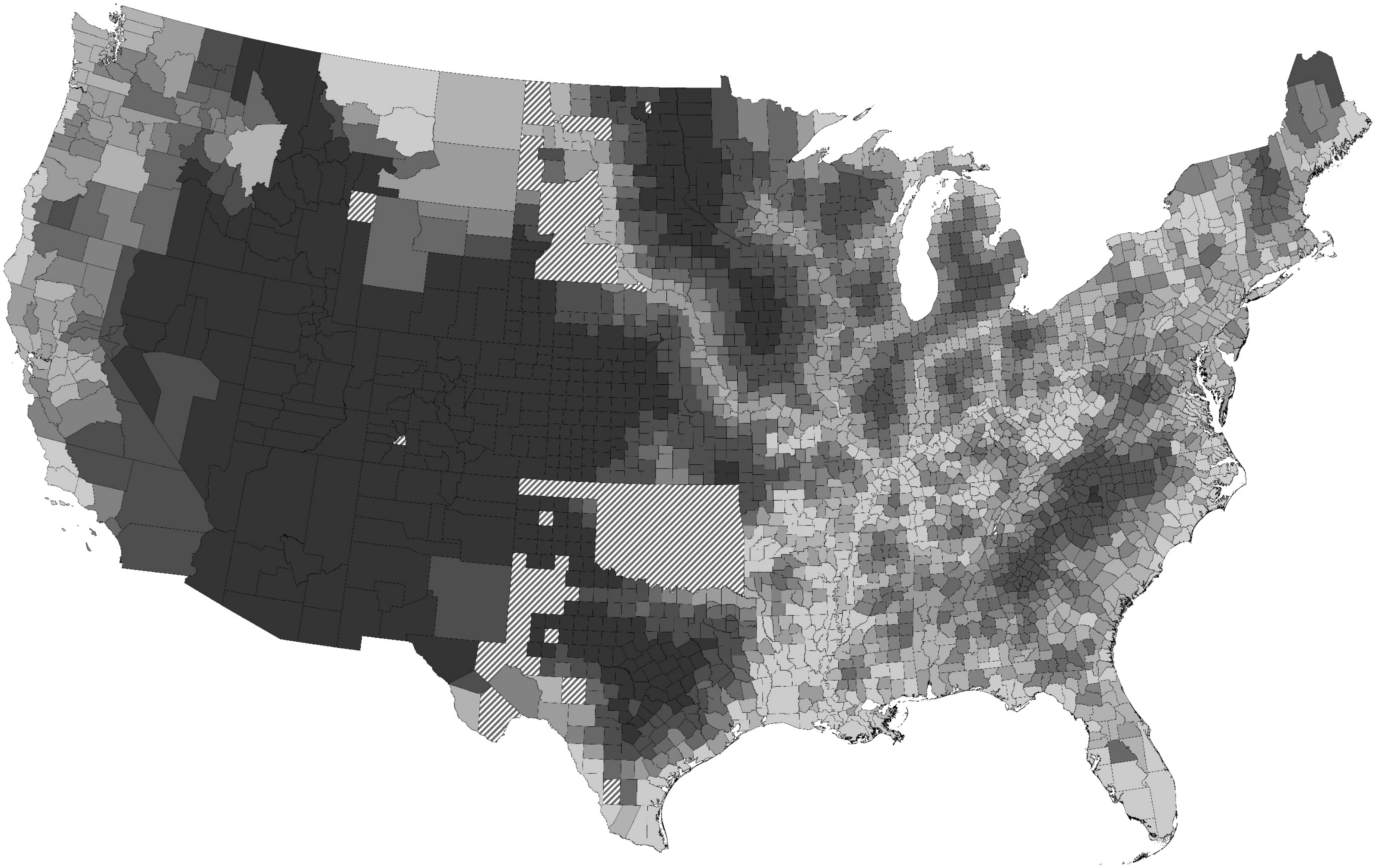


C. Cross-County Dispersion in Materials Wedges, by Decade



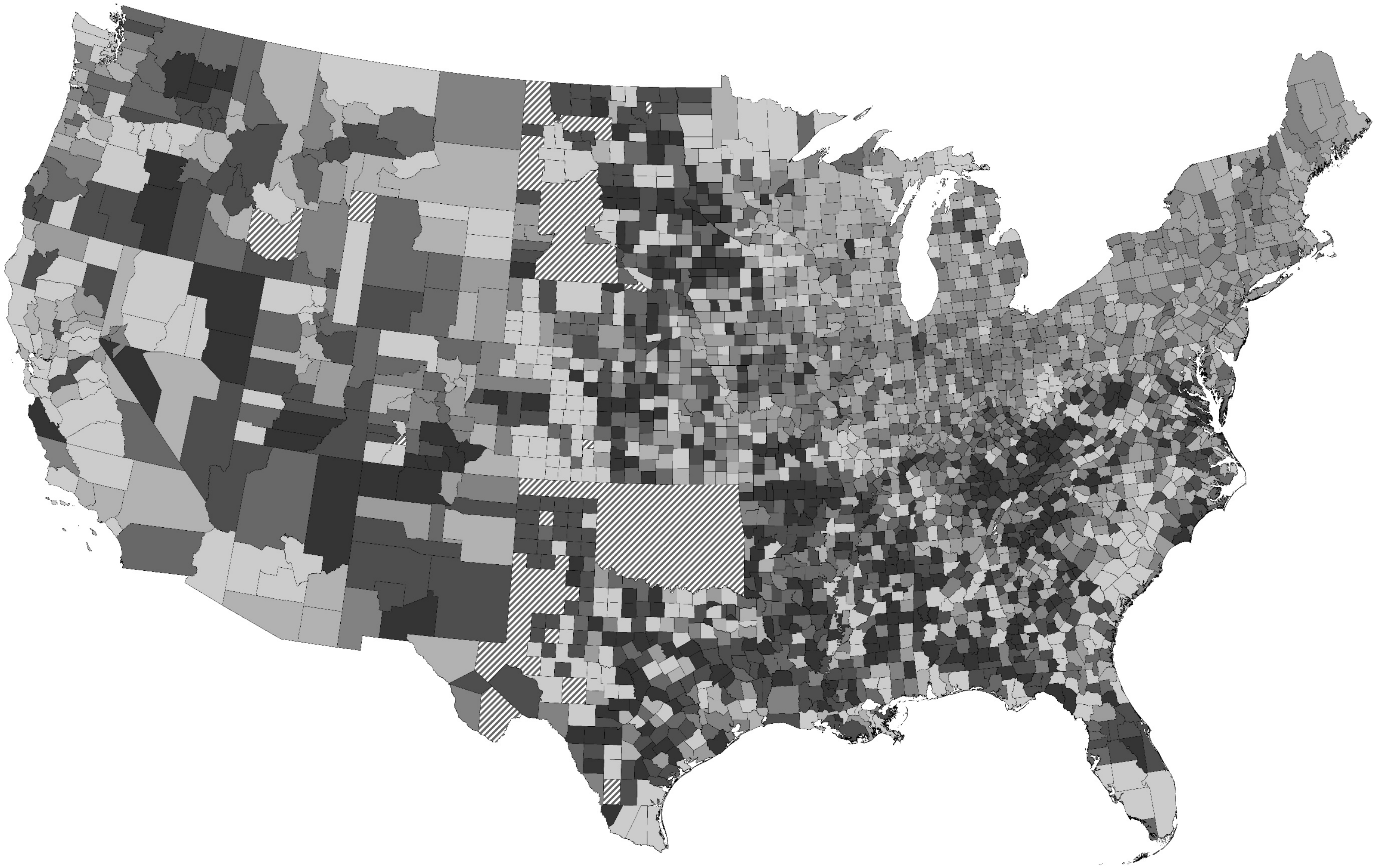
Notes: Each panel plots the cross-county dispersion in input wedges (output elasticity / revenue share), by decade.

Appendix Figure 4. Counterfactual Changes in Market Access, by County



Notes: This map shows counties shaded according to their change in market access from 1890 to the baseline counterfactual scenario: darker shades denote larger declines in market access, and counties are divided into seven equal groups. This counterfactual sample includes all 2,722 counties that in 1890 report positive population and positive revenue (agriculture and/or manufacturing). The excluded geographic areas are cross-hashed. County boundaries correspond to county boundaries in 1890.

Appendix Figure 5. County-level Gaps in the Counterfactual Analysis



Notes: This map shows counties shaded according to their estimated sum of gaps between the output elasticity for each input (materials, labor, capital) and the revenue share for that input: darker shades denote a larger sum of input gaps, and counties are divided into seven equal groups. This counterfactual sample includes all 2,722 counties that in 1890 report positive population and positive revenue (agriculture and/or manufacturing). The excluded geographic areas are cross-hashed. County boundaries correspond to county boundaries in 1890.

Appendix Table 1. Aggregated Industry Group Revenue Shares and Cost Shares

	Revenue	Cost Shares:		
	Share	Materials	Labor	Capital
	(1)	(2)	(3)	(4)
Panel A. All Industries				
Total in 1860	1	0.72	0.24	0.04
Total in 1870	1	0.73	0.23	0.04
Total in 1880	1	0.76	0.20	0.04
Panel B. Aggregated Industry Groups				
Clothing, Textiles, Leather in 1860	0.32	0.69	0.27	0.04
Clothing, Textiles, Leather in 1870	0.23	0.76	0.21	0.03
Clothing, Textiles, Leather in 1880	0.22	0.73	0.23	0.04
Food and Beverage in 1860	0.25	0.90	0.07	0.03
Food and Beverage in 1870	0.21	0.89	0.08	0.03
Food and Beverage in 1880	0.29	0.91	0.06	0.02
Lumber and Wood Products in 1860	0.17	0.59	0.36	0.05
Lumber and Wood Products in 1870	0.16	0.64	0.31	0.05
Lumber and Wood Products in 1880	0.14	0.68	0.27	0.05
Metals and Metal Products in 1860	0.15	0.63	0.32	0.05
Metals and Metal Products in 1870	0.18	0.66	0.29	0.05
Metals and Metal Products in 1880	0.13	0.68	0.27	0.05
Other Industries in 1860	0.12	0.66	0.29	0.05
Other Industries in 1870	0.22	0.66	0.30	0.04
Other Industries in 1880	0.22	0.64	0.29	0.07

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. Panel B reports these statistics for aggregated industry groups, along with that industry group's share of total revenue (column 1). The "Clothing, Textiles, Leather" industry group contains: clothing; yarn, cloth, and other textiles; leather; leather products; boots and shoes. The "Food and Beverage" industry group contains: flour and grist mills; bread and bakery products; butter and cheese; tobacco; liquors and beverages. The "Lumber and Wood Products" industry group contains: lumber; wood products; cooperage; carriages and wagons; furniture; paper; printing and publishing; ship and boat building. The "Metals and Metal Products" industry group contains: iron and steel; iron and steel products; brass and other metal products; tin, copper, and sheet-iron ware; jewelry, pottery, and decorative work.

Appendix Table 2. Impacts of Market Access, Robustness (Measurement of Productivity)

	Estimated Impact of Market Access on:		
	Productivity	TFPR	AE
	(1)	(2)	(3)
1. Baseline Specification	0.204	0.036	0.168
	(0.051)	(0.025)	(0.051)
2. Doubling firm capital costs	0.203	0.081	0.122
	(0.061)	(0.038)	(0.058)
3. Tripling firm capital costs	0.184	0.039	0.144
	(0.053)	(0.031)	(0.049)
4. Using National instead of State Interest Rates	0.209	0.036	0.173
	(0.050)	(0.025)	(0.051)
5. Use materials wedge for capital and labor wedges	0.202	0.052	0.149
	(0.073)	(0.064)	(0.065)
6. Decrease dispersion of capital wedges by 5%	0.206	0.030	0.177
	(0.049)	(0.024)	(0.050)
7. Decrease dispersion of capital wedges by 10%	0.205	0.026	0.179
	(0.049)	(0.023)	(0.050)
8. Decrease dispersion of capital wedges by 25%	0.199	0.016	0.183
	(0.049)	(0.020)	(0.051)
9. Decrease dispersion of all wedges by 5%	0.208	0.029	0.179
	(0.049)	(0.024)	(0.050)
10. Decrease dispersion of all wedges by 10%	0.208	0.025	0.183
	(0.049)	(0.023)	(0.051)
11. Decrease dispersion of all wedges by 25%	0.206	0.015	0.191
	(0.050)	(0.019)	(0.051)
12. Using 1860 values for Wedges and Scaling Factor	0.204	0.045	0.159
	(0.066)	(0.022)	(0.067)
13. Using 1860 values for Wedges and Scaling Factor,	0.196	0.043	0.153
and 1860 population for calculating market access	(0.064)	(0.021)	(0.064)
14. Using Median Scaling Factor	0.225	0.028	0.197
	(0.055)	(0.023)	(0.051)
15. Using County-Specific Scaling Factors	0.223	0.040	0.183
	(0.055)	(0.024)	(0.052)
16. Dropping top/bottom centile, change in Productivity	0.190	0.043	0.147
	(0.060)	(0.030)	(0.064)
17. Dropping top/bottom 5 centiles, change in Productivity	0.225	0.100	0.126
	(0.081)	(0.041)	(0.087)
18. Inflating firm labor costs	0.184	0.039	0.144
	(0.053)	(0.031)	(0.049)

19	Decreasing returns to scale (0.95)	0.204 (0.051)	0.081 (0.027)	0.123 (0.042)
20.	Increasing returns to scale (1.05)	0.204 (0.051)	-0.014 (0.030)	0.218 (0.064)
21.	Using elasticities weighted by costs instead of revenues	0.204 (0.051)	0.031 (0.027)	0.173 (0.053)
22.	Using elasticities from most-efficient counties	0.204 (0.051)	0.035 (0.028)	0.169 (0.052)
23.	Using national industry cost shares, omitting own county	0.204 (0.051)	0.034 (0.025)	0.171 (0.052)
24.	Using state industry cost shares	0.204 (0.051)	0.042 (0.026)	0.162 (0.052)
25.	Using state industry elasticities, omitting own county	0.204 (0.051)	0.038 (0.029)	0.167 (0.054)
26.	Using local elasticities	0.204 (0.051)	0.045 (0.021)	0.159 (0.049)
27.	Inflate labor cost share by 5 percentage points	0.204 (0.051)	0.030 (0.023)	0.174 (0.056)
28.	Inflate materials cost share by 5 percentage points	0.204 (0.051)	0.035 (0.032)	0.169 (0.052)
29.	Inflate capital cost share by 5 percentage points	0.204 (0.051)	0.041 (0.022)	0.163 (0.052)
30.	Exclude butter and cheese industry	0.206 (0.049)	0.034 (0.025)	0.172 (0.050)

Notes: Row 1 reports our baseline estimates, from Table 2, for the impacts of market accesss on county productivity (column 1) and the impacts through changes in county TFPR or revenue total factor productivity (column 2) and county AE or allocative efficiency (column 3). All regressions include county fixed effects, state-by-year fixed effects, and year-interacted cubic polymials in county latitude and longitude. We continue to report the estimated impact of a one standard deviation greater change in market access from 1860 to 1880. The sample is a balanced panel of 1,802 counties (1860, 1870, 1880). Robust standard errors clustered by state are reported in parentheses.

Rows 2 to 30 report alternative estimates, which generally relate to adjusting our measurement of productivity and its decomposition into TFPR and AE. Rows 2 and 3 inflate firm capital costs, doubling or tripling the assumed interest rate to calculate annual capital expenditures. Row 4. Row 5 reports estimates when using counties' materials wedge to proxy for their capital and labor wedges, adjusting county revenue shares for capital and labor to equal their output elasticities divided by the materials wedge. Rows 6 to 8 report estimates when lowering the difference between measured county capital wedges and the median county capital wedges by 5%, 10%, or 25%, and Rows 9 to 11 report estimates when lowering the cross-county dispersion in all input wedges. Row 12 fixes county input wedges at their 1860 levels, rather than counties' average wedges over the 1860-1880 sample period, and fixes the scaling factor at its 1860 level of 4.9 (the average of county revenue divided by county productivity) and Row 13 does the same while also calculating counties' market access holding counties' population fixed at 1860 levels. Row 14 uses the median county scaling factor (4.9), rather than the average county scaling factor (5.1). Row 15 uses county-specific scaling factors, dropping counties with negative values and the top 1% of values. Rows 16 and 17 drop counties with the largest and smallest changes in productivity from 1860 to 1880: row 16 excludes the top and bottom 1% of counties, and row 17 excludes the top and bottom 5% of counties. Row 18 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. Rows 19 and 20 modify our baseline assumption of constant returns to scale, and re-scale the cost shares to add up to 0.95 (row 19) or 1.05 (row 20). Rows 21 to 26 adjust the measurement of county output elasticities: averaging industry-level cost shares with weights equal to an industry's share of total expenditure in that county (rather than revenue); averaging over industry-level cost shares from only the most efficient counties (those with gaps within one standard deviation of zero); calculating leave-out elasticities based on industry-level cost shares in other counties (omitting own industries); calculating state-specific industry-level cost shares; and calculating county-industry cost shares (which imposes a constant wedge across inputs). Rows 27 to 29 modify the relative cost shares for each factor, inflating by 5 percentage points the cost shares of labor (row 27), materials (row 28), and capital (row 29), and proportionally reducing the cost shares of the other factors. Row 30 excludes the butter and cheese industry from the analysis, for which coverage in the Census of Manufactures changes from 1860 to 1870.

Appendix Table 3. Impacts of Market Access, Robustness (Regional Shocks)

		Estimated Impact of Market Access on:		
		Productivity	TFPR	AE
		(1)	(2)	(3)
1.	Baseline Specification	0.204	0.036	0.168
		(0.051)	(0.025)	(0.051)
2.	Only 1860 and 1870	0.188	0.020	0.168
		(0.078)	(0.038)	(0.079)
3.	Only 1870 and 1880	0.199	0.121	0.078
		(0.102)	(0.059)	(0.113)
4.	Only 1860 and 1880	0.216	0.006	0.209
		(0.067)	(0.034)	(0.057)
5.	Current Market Access, Controlling for Market Access	0.177	0.043	0.134
	10 Years in the Future	(0.045)	(0.025)	(0.048)
	Market Access 10 Years in the Future, Controlling for	0.037	0.013	0.024
	Current Market Access	(0.085)	(0.066)	(0.070)
6.	Controlling for 1860 Waterway Access	0.202	0.032	0.170
		(0.060)	(0.047)	(0.073)
7.	Controlling for 1860 Market Access	0.253	0.058	0.195
		(0.089)	(0.050)	(0.082)
8.	Controls for Industry Shares	0.198	0.025	0.174
		(0.056)	(0.026)	(0.058)
9.	Controls for Detailed Industry Shares	0.162	0.037	0.125
		(0.066)	(0.023)	(0.074)
10.	Controls for County Gaps in 1860	0.182	0.003	0.179
		(0.045)	(0.025)	(0.059)
11.	Controls for County Elasticities in 1860	0.217	0.039	0.178
		(0.052)	(0.025)	(0.055)
12.	Controls for County Elasticities and Gaps in 1860	0.194	0.004	0.190
		(0.049)	(0.024)	(0.063)
13.	Controls for County Revenue Shares in 1860	0.195	0.002	0.194
		(0.047)	(0.023)	(0.060)
14.	Controls for County Wedges in 1860	0.207	0.065	0.142
		(0.055)	(0.023)	(0.056)
15.	Controls for County HHIs in 1860	0.190	0.046	0.144
		(0.053)	(0.026)	(0.054)
16.	Controls for Frontier in 1860	0.166	0.040	0.126
		(0.049)	(0.026)	(0.046)
17.	Controls for Gaps, Elasticities, Wedges, HHIs, and	0.163	0.035	0.127
	Frontier in 1860	(0.052)	(0.020)	(0.062)

18. Excludes Civil War related industries (strict)	0.209	0.037	0.172
	(0.050)	(0.026)	(0.052)
19. Excludes Civil War related industries (broad)	0.221	0.047	0.174
	(0.050)	(0.027)	(0.052)
20. Controls for share of War Related Industries (strict)	0.205	0.035	0.170
	(0.051)	(0.026)	(0.052)
21. Controls for share of War Related Industries (broad)	0.223	0.035	0.188
	(0.048)	(0.025)	(0.050)
22. Controls for Civil War battles and casualties	0.204	0.037	0.166
	(0.052)	(0.026)	(0.051)
23. Drop Counties with battles, > 500 casualties	0.216	0.038	0.178
	(0.054)	(0.025)	(0.053)
24. Drop Counties with Civil War battles	0.199	0.033	0.166
	(0.050)	(0.027)	(0.054)
25. Drop Counties on Civil War Border	0.235	0.040	0.194
	(0.056)	(0.027)	(0.054)
26. Drop Confederate states	0.210	0.010	0.201
	(0.066)	(0.037)	(0.063)
27. Drop Slave states	0.172	-0.001	0.173
	(0.077)	(0.044)	(0.077)
28. Drop Southern region	0.198	0.030	0.168
	(0.064)	(0.028)	(0.061)
29. Fixed Effects for 20 "resource regions"	0.229	0.046	0.183
	(0.054)	(0.029)	(0.052)
30. Fixed Effects for 106 "resource subregions"	0.185	0.038	0.147
	(0.067)	(0.049)	(0.056)
31. Fifth Order Polynomial	0.207	0.027	0.180
	(0.055)	(0.029)	(0.059)
32. First Order Polynomial	0.176	0.023	0.153
	(0.047)	(0.026)	(0.042)
33. Drop Western region (Plains and West Coast)	0.204	0.016	0.188
	(0.079)	(0.035)	(0.077)

Notes: Row 1 reports our baseline estimates, from Table 2, for the impacts of market accesss on county productivity (column 1) and the impacts through changes in county TFPR or revenue total factor productivity (column 2) and county AE or allocative efficiency (column 3). All regressions include county fixed effects, state-by-year fixed effects, and year-interacted cubic polymials in county latitude and longitude. We continue to report the estimated impact of a one standard deviation greater change in market access from 1860 to 1880. The sample is a balanced panel of 1,802 counties (1860, 1870, 1880). Robust standard errors clustered by state are reported in parentheses.

Rows 2 to 33 report alternative estimates, which generally relate to controlling for other regional shocks. Rows 2, 3, and 4 report estimates when restricting the sample period to two decades only, focusing on changes over that period only. Row 5 reports the impacts of current market access along with the impacts of future market access (i.e., the 10-year pre-trend), controlling for contemporaneous and future values for whether a county has any railroad and the length of its railroads. Row 6 controls for counties' market access through waterways in 1860, interacted with decade. Row 7 controls for counties' market access in 1860, interacted with decade. Row 8 controls for counties' 1860 share of revenue in each of 31 industries, interacted with decade. Row 9 controls for counties 1860 share of revenue in each of 193 industries, interacted with decade. Rows 10 to 16 control for other county characteristics in 1860, interacted with decade: county gaps; county elasticities; county elasticities and gaps; county revenue shares; county wedges; county employment concentration within manufacturing and across sectors; and whether a county was on the "frontier." Row 17 controls for all of those variables, interacted with decade. Rows 18 and 19 exclude from the data those industries most related to the Civil War or more broadly related to the Civil War (see Data Appendix for the list of industries). Rows 20 and 21 instead control for counties' revenue share in Civil War related production, interacted with decade. Row 22 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 decade fixed effects. Row 23 excludes 99 counties with recorded Civil War battles that had more than 500 recorded casualties, and Row 24 excludes 177 counties with recorded Civil War battles. Row 25 drops 93 counties on the North-South border, Row 26 drops 745 counties in Confederate states, Row 27 drops 980 counties in slave states, and Row 28 drops 765 counties in the Southern region. Row 29 controls for region-by year fixed effects (20 regions), and row 30 controls for subregion-by-year fixed effects (106 subregions). Rows 31 and 32 modify the controls for county latitude and longitude to be a fifth-order polynomial or first-order polynomial, respectively. Row 33 excludes 201 counties in the Plains region and West Coast region of the sample.

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

		Estimated Impact of Market Access on:		
		Productivity	TFPR	AE
		(1)	(3)	(2)
1.	Baseline Specification	0.204	0.036	0.168
		(0.051)	(0.025)	(0.051)
2.	Dropping top/bottom centile, change in market access	0.190	0.043	0.147
		(0.060)	(0.030)	(0.064)
3.	Dropping top/bottom 5 centiles, change in market access	0.225	0.100	0.126
		(0.081)	(0.041)	(0.087)
4.	Reduces the cost of water to 0.198 cents per ton mile	0.173	0.028	0.145
		(0.050)	(0.025)	(0.045)
5.	Reduces the cost of wagons to 14 cents per ton mile	0.220	0.037	0.183
		(0.055)	(0.028)	(0.057)
6.	No transshipment costs between waterways	0.202	0.033	0.169
		(0.051)	(0.025)	(0.051)
7.	Include transshipment between Northern and Southern RRs	0.205	0.036	0.168
		(0.052)	(0.026)	(0.051)
8.	Raise railroad cost to 0.735 cents per ton mile	0.201	0.034	0.167
		(0.051)	(0.025)	(0.050)
9.	Raise railroad cost to 0.878 cents per ton mile	0.193	0.032	0.162
		(0.052)	(0.025)	(0.049)
10.	Average price of goods, $\overline{P}$, set to 20	0.205	0.037	0.168
		(0.051)	(0.026)	(0.051)
11.	Average price of goods, $\overline{P}$, set to 50	0.203	0.035	0.168
		(0.052)	(0.025)	(0.051)
12.	Trade elasticity, Θ, set to 1.815	0.204	0.036	0.168
		(0.051)	(0.025)	(0.051)
13.	Trade elasticity, Θ, set to 3.556	0.205	0.036	0.169
		(0.051)	(0.025)	(0.051)
14.	Trade elasticity, Θ, set to 8.22	0.208	0.037	0.171
		(0.051)	(0.026)	(0.051)
15.	Include access to international markets	0.203	0.036	0.167
		(0.051)	(0.025)	(0.051)
16.	Adjustment for Census undercounting	0.204	0.036	0.168
		(0.051)	(0.025)	(0.051)

17. Measure access to county wealth	0.197	0.033	0.164
	(0.049)	(0.025)	(0.050)
18. Include access to own market	0.205	0.036	0.170
	(0.051)	(0.025)	(0.051)
19. Limit access to counties beyond 5 miles	0.203	0.036	0.166
	(0.051)	(0.025)	(0.051)
20. Limit access to counties beyond 50 miles	0.200	0.036	0.164
	(0.051)	(0.025)	(0.051)
21. Limit access to counties beyond 200 miles	0.185	0.032	0.153
	(0.048)	(0.024)	(0.049)

Notes: Row 1 reports our baseline estimates, from Table 2, for the impacts of market accesss on county productivity (column 1) and the impacts through changes in county TFPR or revenue total factor productivity (column 2) and county AE or allocative efficiency (column 3). All regressions include county fixed effects, state-by-year fixed effects, and year-interacted cubic polymials in county latitude and longitude. We continue to report the estimated impact of a one standard deviation greater change in market access from 1860 to 1880. The sample is a balanced panel of 1,802 counties (1860, 1870, 1880). Robust standard errors clustered by state are reported in parentheses.

Rows 2 to 21 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 transfering 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 (35.7) with alternative assumed values of 20 or 50. Rows 12 and 13 replace our baseline estimated trade elasticity (2.788) with alternative assumed values that reflect its estimated 95% confidence internal (1.815 to 3.556), 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-enumeration 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)
Panel A. Sum of Average Input Gaps						
1860	0.22	0.33	0.23	0.21	0.22	0.21
	[0.06]	[0.18]	[0.09]	[0.06]	[0.04]	[0.09]
1870	0.19	0.24	0.22	0.21	0.19	0.22
	[0.07]	[0.09]	[0.10]	[0.08]	[0.06]	[0.08]
1880	0.15	0.16	0.15	0.15	0.15	0.17
	[0.06]	[0.06]	[0.03]	[0.04]	[0.07]	[0.06]
Panel B. Average Materials Gap						
1860	0.15	0.27	0.26	0.16	0.14	0.16
	[0.05]	[0.11]	[0.09]	[0.05]	[0.04]	[0.07]
1870	0.14	0.20	0.21	0.15	0.13	0.16
	[0.07]	[0.09]	[0.08]	[0.08]	[0.06]	[0.10]
1880	0.11	0.13	0.10	0.10	0.11	0.13
	[0.06]	[0.06]	[0.04]	[0.05]	[0.07]	[0.07]
Panel C. Average Labor Gap						
1860	0.06	0.05	-0.03	0.04	0.07	0.05
	[0.04]	[0.10]	[0.05]	[0.04]	[0.03]	[0.05]
1870	0.05	0.03	0.01	0.05	0.05	0.06
	[0.03]	[0.07]	[0.04]	[0.03]	[0.03]	[0.06]
1880	0.04	0.02	0.04	0.04	0.03	0.04
	[0.03]	[0.04]	[0.03]	[0.03]	[0.02]	[0.06]
Panel D. Average Capital Gap						
1860	0.01	0.01	0.00	0.01	0.01	0.00
	[0.01]	[0.03]	[0.06]	[0.01]	[0.01]	[0.03]
1870	0.01	0.00	-0.01	0.00	0.01	0.01
	[0.01]	[0.02]	[0.02]	[0.01]	[0.01]	[0.02]
1880	0.01	0.01	0.01	0.01	0.00	0.01
	[0.02]	[0.04]	[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 output elasticity 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 revenue 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

		By Region:				
	National	Plains	West Coast	Midwest	Northeast	South
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Average of Input Wedges						
1860	1.36	1.80	1.30	1.30	1.37	1.33
	[0.21]	[1.02]	[0.20]	[0.27]	[0.14]	[0.28]
1870	1.31	1.38	1.14	1.29	1.30	1.49
	[0.20]	[0.40]	[0.11]	[0.18]	[0.12]	[0.59]
1880	1.24	1.27	1.24	1.31	1.21	1.31
	[0.18]	[0.28]	[0.12]	[0.18]	[0.14]	[0.33]
Panel B. Average Materials Wedge						
1860	1.28	1.91	1.63	1.28	1.27	1.34
	[0.23]	[1.30]	[0.45]	[0.19]	[0.10]	[0.47]
1870	1.25	1.44	1.42	1.29	1.23	1.31
	[0.21]	[0.27]	[0.21]	[0.34]	[0.11]	[0.30]
1880	1.17	1.21	1.17	1.15	1.18	1.22
	[0.09]	[0.11]	[0.08]	[0.08]	[0.08]	[0.15]
Panel C. Average Labor Wedge						
1860	1.35	1.62	0.91	1.32	1.37	1.35
	[0.27]	[1.07]	[0.17]	[0.38]	[0.17]	[0.38]
1870	1.32	1.43	1.10	1.39	1.27	1.73
	[0.45]	[1.04]	[0.19]	[0.33]	[0.17]	[1.58]
1880	1.23	1.19	1.21	1.29	1.20	1.35
	[0.23]	[0.53]	[0.15]	[0.22]	[0.14]	[0.66]
Panel D. Average Capital Wedge						
1860	1.43	1.87	1.37	1.31	1.48	1.31
	[0.40]	[1.54]	[0.47]	[0.47]	[0.31]	[0.56]
1870	1.35	1.28	0.90	1.20	1.41	1.43
	[0.34]	[0.51]	[0.24]	[0.27]	[0.29]	[0.64]
1880	1.32	1.42	1.35	1.49	1.24	1.36
	[0.44]	[0.53]	[0.30]	[0.45]	[0.41]	[0.57]

Notes: This table reports measured wedges in the manufacturing sector, by decade, where the input wedges are equal to that input's output elasticity 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 revenue 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. Measured Manufacturing Revenue Shares by Decade and Region

		By Region:				
	National	Plains	West Coast	Midwest	Northeast	South
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Average Materials Revenue Share						
1860	0.56	0.46	0.48	0.60	0.55	0.57
	[0.07]	[0.18]	[0.11]	[0.08]	[0.05]	[0.12]
1870	0.59	0.52	0.54	0.59	0.59	0.58
	[0.07]	[0.11]	[0.09]	[0.08]	[0.07]	[0.11]
1880	0.64	0.66	0.61	0.68	0.63	0.63
	[0.09]	[0.08]	[0.05]	[0.07]	[0.09]	[0.07]
Panel B. Average Labor Revenue Share						
1860	0.19	0.17	0.24	0.16	0.20	0.17
	[0.05]	[0.09]	[0.07]	[0.05]	[0.04]	[0.06]
1870	0.19	0.21	0.19	0.17	0.20	0.16
	[0.05]	[0.11]	[0.05]	[0.05]	[0.04]	[0.07]
1880	0.17	0.15	0.19	0.14	0.19	0.16
	[0.05]	[0.06]	[0.03]	[0.04]	[0.04]	[0.06]
Panel C. Average Capital Revenue Share						
1860	0.03	0.04	0.05	0.03	0.03	0.04
	[0.02]	[0.03]	[0.06]	[0.02]	[0.01]	[0.03]
1870	0.03	0.03	0.05	0.03	0.03	0.03
	[0.01]	[0.02]	[0.02]	[0.01]	[0.01]	[0.02]
1880	0.04	0.04	0.04	0.03	0.04	0.04
	[0.01]	[0.03]	[0.02]	[0.02]	[0.01]	[0.02]

Notes: This table reports measured revenue shares in the manufacturing sector, by decade, where the input revenue shares are equal to county expenditure on that input divided by county revenue. Column 1 reports these revenue shares at the national level, and columns 2 to 6 report these revenue shares by region, which weight county-level revenue shares by county revenue in that decade. Panels A to C report revenue shares for materials, labor, and capital. Standard deviations are reported in brackets.

Appendix Table 8. Measured Manufacturing Output Elasticities by Decade and Region

		By Region:				
	National	Plains	West Coast	Midwest	Northeast	South
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Average Materials Elasticity						
1860	0.71	0.73	0.74	0.76	0.69	0.73
	[0.05]	[0.10]	[0.06]	[0.06]	[0.04]	[0.07]
1870	0.72	0.72	0.75	0.74	0.71	0.74
	[0.04]	[0.07]	[0.03]	[0.04]	[0.03]	[0.06]
1880	0.75	0.79	0.72	0.77	0.74	0.76
	[0.05]	[0.07]	[0.05]	[0.05]	[0.05]	[0.05]
Panel B. Average Labor Elasticity						
1860	0.25	0.22	0.22	0.20	0.26	0.23
	[0.05]	[0.09]	[0.05]	[0.06]	[0.04]	[0.07]
1870	0.24	0.24	0.21	0.22	0.25	0.22
	[0.04]	[0.06]	[0.03]	[0.04]	[0.03]	[0.05]
1880	0.21	0.16	0.23	0.18	0.22	0.19
	[0.05]	[0.06]	[0.04]	[0.05]	[0.04]	[0.05]
Panel C. Average Capital Elasticity						
1860	0.04	0.05	0.04	0.04	0.04	0.04
	[0.01]	[0.01]	[0.01]	[0.01]	[0.00]	[0.01]
1870	0.04	0.04	0.04	0.04	0.04	0.04
	[0.00]	[0.01]	[0.01]	[0.00]	[0.00]	[0.01]
1880	0.04	0.05	0.06	0.04	0.04	0.05
	[0.01]	[0.02]	[0.01]	[0.01]	[0.01]	[0.01]

Notes: This table reports measured output elasticities in the manufacturing sector, by decade. County output elasticities for each input in each decade are equal to: each national industry's expenditure on that input divided by total industry expenditure, multiplied by the share of county revenue in that industry. Column 1 reports these elasticities at the national level, and columns 2 to 6 report these elasticities by region, which weight county elasticities by county revenue in that decade. Panels A to C report elasticities for materials, labor, and capital. Standard deviations are reported in brackets.

Appendix Table 9. Impacts of Market Access on Input Gaps and Wedges, by Region

		By Region:					Frontier
	National	Plains	West Coast	Midwest	Northeast	South	Counties
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Input Gaps							
Materials	0.012	0.020	0.095	0.019	0.027	0.021	0.019
	(0.006)	(0.008)	(0.009)	(0.009)	(0.025)	(0.017)	(0.013)
Labor	-0.001	0.000	0.047	-0.015	0.004	-0.014	-0.003
	(0.005)	(0.009)	(0.011)	(0.003)	(0.016)	(0.009)	(0.007)
Capital	0.001	0.000	0.006	0.004	0.000	0.002	0.001
	(0.002)	(0.002)	(0.007)	(0.004)	(0.002)	(0.002)	(0.004)
Panel B. Input Wedges							
Materials	-0.027	0.021	-0.023	0.031	0.022	0.132	-0.097
	(0.039)	(0.021)	(0.031)	(0.039)	(0.027)	(0.089)	(0.127)
Labor	-0.050	0.062	0.101	-0.165	-0.038	-0.478	0.002
	(0.067)	(0.045)	(0.074)	(0.053)	(0.147)	(0.109)	(0.065)
Capital	0.022	0.173	-0.097	-0.005	0.060	-0.006	0.145
	(0.036)	(0.115)	(0.191)	(0.064)	(0.046)	(0.112)	(0.081)

Notes: Panels A and B report the estimated impacts of market access on county-level input gaps and county-level input wedges in manufacturing. Column 1 reports estimates at the national level, as in Columns 2 and 3 of Table 5. Columns 2 to 6 report estimates from separate regressions for each region, and Column 7 reports estimates in the sample of "Frontier Counties," defined following Bazzi et al. (2020) as counties with between two and six people per square mile in 1860 and that are within 100km of the boundary where population density fell below two people per square mile in 1860.

All regressions include county fixed effects, state-by-year fixed effects, and year-interacted cubic polymials in county latitude and longitude. The sample is our main balanced panel of 1,802 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.

Appendix Table 10. Impacts of Market Access, County-by-Industry Level Regressions

			By Industry Group:			
	Pooled Specification	Weighted by 1860 Revenue Share	Clothing, Textiles, Leather	Food and Beverage	Lumber and Wood Products	Metals and Metal Products
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Log County-Industry Productivity						
Log Market Access	0.189	0.160	0.187	0.008	0.220	0.302
	(0.058)	(0.052)	(0.096)	(0.056)	(0.148)	(0.158)
Panel B. County-Industry TFPR (Revenue Total Factor Productivity)						
Log Market Access	0.056	0.044	0.060	-0.012	0.075	0.054
	(0.023)	(0.025)	(0.073)	(0.023)	(0.038)	(0.123)
Panel C. County-Industry AE (Allocative Efficiency)						
Log Market Access	0.132	0.116	0.127	0.020	0.145	0.248
	(0.058)	(0.043)	(0.089)	(0.063)	(0.139)	(0.119)
Number of Counties	1,800	1,800	994	1,338	1,480	709
County-Year Obs.	5,400	5,400	2,640	3,665	3,984	1,860

Notes: this table reports estimates from regressions at the county-by-industry level, after aggregating the more-detailed industries to five industry groups: clothing, textiles, leather; food and beverage; lumber and wood products; metals and metal products; and other industries. We extend our baseline estimating equation 8, interacting the control variables with industry group. The specification then includes: county-industry fixed effects, state-year-industry fixed effects, and year-specific cubic polynomials in county latitude and longitude. The sample is drawn from our main balanced panel of 1,802 counties in 1860, 1870, and 1880, though each industry group is not reported in each county and decade. We omit county-industries that appear only once, but do not restrict the sample to county-industries that appear all three years. We continue to report the estimated impact of a one standard deviation greater change in market access from 1860 to 1880. Robust standard errors clustered by state are reported in parentheses.

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

Appendix Table 11. Measured Manufacturing Gaps, by Decade and Industry

	All Industries	Clothing, Textiles, Leather	Food and Beverage	Lumber and Wood Products	Metals and Metal Products
	(1)	(2)	(3)	(4)	(5)
Panel A. Sum of Average Input Gaps					
1860	0.22	0.22	0.26	0.26	0.23
	[0.09]	[0.06]	[0.09]	[0.09]	[0.09]
1870	0.19	0.16	0.25	0.25	0.19
	[0.16]	[0.06]	[0.09]	[0.09]	[0.08]
1880	0.15	0.16	0.18	0.18	0.17
	[0.12]	[0.04]	[0.08]	[0.08]	[0.08]
Panel B. Average Materials Gap					
1860	0.15	0.18	0.17	0.17	0.17
	[0.09]	[0.06]	[0.09]	[0.09]	[0.10]
1870	0.11	0.08	0.12	0.12	0.12
	[0.14]	[0.07]	[0.10]	[0.10]	[0.10]
1880	0.07	0.10	0.05	0.05	0.09
	[0.12]	[0.05]	[0.11]	[0.11]	[0.10]
Panel C. Average Labor Gap					
1860	0.06	0.04	0.08	0.08	0.05
	[0.06]	[0.05]	[0.07]	[0.07]	[0.06]
1870	0.07	0.07	0.11	0.11	0.06
	[0.07]	[0.04]	[0.06]	[0.06]	[0.06]
1880	0.07	0.06	0.13	0.13	0.07
	[0.06]	[0.04]	[0.06]	[0.06]	[0.07]
Panel D. Average Capital Gap					
1860	0.01	0.00	0.01	0.01	0.01
	[0.02]	[0.01]	[0.02]	[0.02]	[0.04]
1870	0.01	0.00	0.01	0.01	0.01
	[0.02]	[0.01]	[0.01]	[0.01]	[0.02]
1880	0.00	0.00	0.01	0.01	0.01
	[0.03]	[0.01]	[0.02]	[0.02]	[0.02]

Notes: This table reports measured gaps in the manufacturing sector, by decade, where the input gaps are equal to that input's output elasticity minus its revenue share. Column 1 reports these gaps at the national level, and columns 2 to 5 report these gaps for consistent aggregated industry groups, which weight county-industry gaps by county-industry revenue 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 12. Measured Manufacturing Wedges, by Decade and Industry

	All Industries	Clothing, Textiles, Leather	Food and Beverage	Lumber and Wood Products	Metals and Metal Products
	(1)	(2)	(3)	(4)	(5)
Panel A. Average Input Wedge					
1860	1.47	1.29	1.46	1.46	1.36
	[0.46]	[0.23]	[0.43]	[0.43]	[0.26]
1870	1.49	1.33	1.46	1.46	1.38
	[0.58]	[0.33]	[0.40]	[0.40]	[0.88]
1880	1.47	1.25	1.38	1.38	1.34
	[0.51]	[0.22]	[0.23]	[0.23]	[0.32]
Panel B. Average Materials Wedge					
1860	1.34	1.35	1.48	1.48	1.42
	[0.67]	[0.27]	[0.55]	[0.55]	[0.41]
1870	1.25	1.14	1.37	1.37	1.27
	[0.43]	[0.14]	[0.80]	[0.80]	[0.32]
1880	1.14	1.18	1.12	1.12	1.21
	[0.20]	[0.10]	[0.18]	[0.18]	[0.35]
Panel C. Average Labor Wedge					
1860	1.62	1.26	1.43	1.43	1.29
	[0.89]	[0.42]	[0.60]	[0.60]	[0.42]
1870	1.64	1.49	1.60	1.60	1.32
	[0.91]	[0.62]	[0.59]	[0.59]	[0.61]
1880	1.76	1.38	1.76	1.76	1.46
	[0.93]	[0.54]	[0.68]	[0.68]	[0.57]
Panel D. Average Capital Wedge					
1860	1.45	1.26	1.48	1.48	1.37
	[0.67]	[0.51]	[0.70]	[0.70]	[0.51]
1870	1.57	1.35	1.42	1.42	1.53
	[1.17]	[0.61]	[0.59]	[0.59]	[2.18]
1880	1.52	1.18	1.28	1.28	1.35
	[0.87]	[0.41]	[0.43]	[0.43]	[0.65]

Notes: This table reports measured wedges in the manufacturing sector, by decade, where the input wedges are equal to that input's output elasticity divided by its revenue share. Column 1 reports these wedges at the national level, and columns 2 to 5 report these wedges for consistent aggregated industry groups, which weight county-industry wedges by county-industry revenue 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 13. Impacts of Market Access on Input Gaps and Wedges, by Industry Group

		By Industry Group:			
	All Industries	Clothing, Textiles, Leather	Food and Beverage	Lumber and Wood Products	Metals and Metal Products
	(1)	(2)	(3)	(4)	(5)
Panel A. Input Gaps					
Materials	0.018	-0.001	-0.004	0.010	-0.021
	(0.014)	(0.008)	(0.004)	(0.004)	(0.015)
Labor	0.010	0.015	0.002	0.000	0.026
	(0.004)	(0.005)	(0.003)	(0.006)	(0.015)
Capital	-0.001	0.001	0.000	0.005	0.004
	(0.005)	(0.003)	(0.002)	(0.003)	(0.006)
Panel B. Input Wedges					
Materials	-0.004	-0.022	0.010	0.027	-0.071
	(0.019)	(0.035)	(0.016)	(0.040)	(0.096)
Labor	0.050	0.092	-0.009	0.005	0.420
	(0.029)	(0.078)	(0.078)	(0.034)	(0.259)
Capital	0.001	0.062	-0.068	0.071	-0.039
	(0.038)	(0.076)	(0.049)	(0.070)	(0.105)

Notes: Panels A and B report the estimated impacts of market access on county-industry input gaps and county-industry input wedges, where industry is defined using these four aggregated industry groups and other industries. Column 1 reports pooled estimates, from an unweighted regression, and Columns 2 to 5 allow the effect of market access to vary in each consistent aggregated industry group.

These are county-industry regressions, as in Appendix Table 10. All regressions include county-industry fixed effects, state-year-industry fixed effects, and year-specific cubic polymials in county latitude and longitude. The sample is drawn from our main balanced panel of 1,802 counties in 1860, 1870, and 1880, though each industry group is not reported in each county and decade. We omit county-industries that appear only once, but do not restrict the sample to county-industries that appear all three years. We continue to report the estimated impact of a one standard deviation greater change in market access from 1860 to 1880. Robust standard errors clustered by state are reported in parentheses.

Appendix Table 14. Impacts of Market Access on County Specialization

	Revenue Shares	Value-Added Shares	Surplus Shares	Employment Shares
	(1)	(2)	(3)	(4)
Panel A. Cross-Sector Specialization Index (Manufacturing vs. Agriculture)				
Log Market Access	-0.0103	0.0007	-0.0019	0.0016
	(0.0111)	(0.0062)	(0.0119)	(0.0053)
Number of Counties	1,774	1,774	1,713	1,687
County/Year Obs.	5,322	5,322	5,139	5,061
Panel B. Within-Manufacturing Specialization Index (Across Industries)				
Log Market Access	-0.0241	-0.0369	-0.0278	-0.0099
	(0.0100)	(0.0677)	(0.0403)	(0.0084)
Number of Counties	1,802	1,802	1,802	1,802
County/Year Obs.	5,406	5,406	5,406	5,406

Notes: For the indicated outcome variable, each column and panel reports the estimated impact of log market access from our baseline specification (as in column 1 of Table 1). 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 revenue (column 1), value-added (column 2), surplus (column 3), and employment (column 4) as defined in Table 6. 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 polymials in county latitude and longitude. The samples are drawn from our main balanced panel of 1,802 counties in 1860, 1870, and 1880, which 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 1,802 counties. Robust standard errors clustered by state are reported in parentheses.