Railroads and American Economic Growth:
A “Market Access” Approach*

Dave Donaldson
Stanford University and NBER

Richard Hornbeck
University of Chicago and NBER

July 2015

Abstract

This paper examines the historical impact of railroads on the American economy, with a focus on quantifying the aggregate impact on the agricultural sector in 1890. Expansion of the railroad network may have affected all counties directly or indirectly — an econometric challenge that arises in many empirical settings. However, the total impact on each county is captured by changes in that county’s “market access,” a reduced-form expression derived from general equilibrium trade theory. We measure counties’ market access by constructing a network database of railroads and waterways and calculating lowest-cost county-to-county freight routes. We estimate that county agricultural land values increased substantially with increases in county market access, as the railroad network expanded from 1870 to 1890. Removing all railroads in 1890 is estimated to decrease the total value of US agricultural land by 60%, with limited potential for mitigating these losses through feasible extensions to the canal network or improvements to country roads.

*For helpful comments and suggestions, we thank Larry Katz, anonymous referees, and many colleagues and seminar participants at Boston University, Brown, Chicago, Colorado, Dartmouth, EIEF, George Mason, George Washington, Harvard, LSE, NBER, Northwestern, Santa Clara, Simon Fraser, Stanford, Stanford GSB, Toronto, Toulouse, UBC, UCL, UC-Berkeley, UC-Davis, UC-Irvine, UC-Merced, UC-San Diego, Warwick, and the ASSA and EHA conferences (including our discussants, Gilles Duranton and Jeremy Atack). We are grateful to Jeremy Atack and coauthors for sharing their data and our conversations. Georgios Angelis, Irene Chen, Andrew Das Sarma, Manning Ding, Jan Kozak, Meredith McPhail, Rui Wang, Sophie Wang, and Kevin Wu provided excellent research assistance. This material is based upon work supported by the National Science Foundation under Grant No. 1156239.
I Introduction

Railroads spread throughout a growing United States in the 19th century as the economy rose to global prominence. Railroads became the dominant form of freight transportation and areas around railroad lines prospered. The early historical literature often presumed that railroads were indispensable to the United States’ economy or, at least, very influential for economic growth. Our understanding of the development of the American economy is shaped by an understanding of the impact of railroads and, more generally, the impact of market integration.

In *Railroads and American Economic Growth*, Fogel (1964) transformed the academic literature by using a “social saving” methodology to focus attention on counterfactuals: in the absence of railroads, agricultural freight transportation by rivers and canals would have been only moderately more expensive along most common routes. Fogel argued that small differences in freight rates caused some areas to thrive relative to others, but that railroads had only a small aggregate impact on the American agricultural sector. This social saving methodology has been widely applied to transportation improvements and other technological innovations, though many scholars have discussed both practical and theoretical limitations of the approach (see, e.g., Lebergott, 1966; Nerlove, 1966; McClelland, 1968; David, 1969; White, 1976; Fogel, 1979; Leunig, 2010).  

There is an appeal to a methodology that estimates directly the impacts of railroads, using increasingly available county-level data and digitized railroad maps. Recent work has compared counties that received railroads to counties that did not (Haines and Margo, 2008; Atack and Margo, 2011; Atack et al., 2010; Atack, Haines and Margo, 2011), and similar methods have been used to estimate impacts of railroads in modern China (Banerjee, Duflo and Qian, 2012) or highways in the United States (Baum-Snow, 2007; Michaels, 2008). These studies estimate relative impacts of transportation improvements; for example, due to displacement and complementarities, areas without railroads and areas with previous railroads are also affected when railroads are extended to new areas.

This paper develops a methodology for estimating aggregate impacts of railroads. We argue that it is natural to measure how expansion of the railroad network affects each county’s “market access,” a reduced-form expression derived from general equilibrium trade theory, and then to estimate how enhanced market access is capitalized into each county’s value of agricultural land. A county’s market access increases when it becomes cheaper to trade

---

1 One alternative approach is to create a computational general equilibrium model, with the explicit inclusion of multiple regions separated by a transportation technology (e.g., Williamson, 1974; Herrendorf, Schmitz and Teixeira, 2009). Cervantes (2013) presents estimates from a calibrated trade model. Swisher (2014) calibrates a simpler economic model but models the strategic interaction between railroad and canal companies in building networks.
with another county, particularly when that other county has a larger population and higher trade costs with other counties. In a wide class of multiple-region models, changes in market access summarize the total direct and indirect impacts on each county from changes in the national railroad network.

We measure counties’ market access by constructing a network database of railroads and waterways and calculating lowest-cost county-to-county freight routes. As the national railroad network expanded from 1870 to 1890, we estimate that county-level increases in market access were capitalized into substantially higher agricultural land values.

Another empirical advantage to estimating the impact of market access, rather than estimating the impact of local railroad density, is that counties’ market access is influenced by changes elsewhere in the railroad network. The estimated impact of market access on agricultural land values is largely robust to using only variation in access to more-distant markets or to controlling for changes in counties’ own railroad track, despite concerns about exacerbating attenuation bias from measurement error. Another identification approach uses the fact that counties close to navigable waterways are naturally less dependent on expansion of the railroad network to obtain access to markets. The estimated impact of market access is larger, but much less precise, when instrumenting for changes in market access with counties’ initial market access through waterways only.

The paper then estimates the aggregate impact of railroads on the agricultural sector in 1890, based on the calculated decline in counties’ market access without railroads and the estimated impact of market access on agricultural land values. Removing all railroads in 1890 is estimated to lower the total value of US agricultural land by 60.2%. This reduction in agricultural land value generates annual economic losses equal to 3.22% of GNP, which is moderately larger than comparable social saving estimates by Fogel (1964). Railroads were critical to the agricultural sector, though the total loss of all agricultural land value would only generate annual economic losses equal to 5.35% of GNP. Notably, these estimates and Fogel’s estimates neglect many other channels through which railroads may have impacted other economic sectors and/or technological growth.\footnote{For example, railroads may have had substantial economic impacts through: enabling the transportation of perishable or time-sensitive products, spreading access to natural resources, generally benefiting manufacturing through increased scale and coordination, encouraging technological growth, and increasing labor mobility.}

The initial counterfactual analysis assumes that the population distribution is held fixed in the counterfactual, but we then seek to relax that assumption. First, and most simply, we report similar impacts on agricultural land values when setting the counterfactual distribution of population equal to the historical distribution of population in 1870, 1850, or 1830. Second, drawing on the full structure of the model, we solve for the counterfactual distribu-
tion of population across US counties. Holding the total US population fixed, the estimated impacts on agricultural land are insensitive to the substantial reallocation of workers across the country.

The initial counterfactual analysis also assumes that worker utility is held fixed in the counterfactual, such that all welfare impacts of railroads are capitalized into land values. For worker utility to be held fixed in the counterfactual, however, the model’s structure predicts that total US population would need to fall substantially. An alternative scenario that we consider holds the total US population fixed, but where worker utility is then determined endogenously (and would need to fall substantially in the counterfactual without railroads). In this case land values decline substantially less than in the fixed worker utility case because much of the economic loss is shifted between production factors (i.e., from land to labor). In either case, the counterfactual impacts on population and welfare reflect additional aggregate losses from the removal of railroads, which were not reflected in our baseline estimates or in Fogel’s estimates that are based on losses in agricultural land value only.

Finally, we consider whether alternative transportation improvements had the potential to substitute for the absence of railroads. First, in the absence of railroads, additional canals might have been constructed to bring many areas closer to low-cost waterways (Fogel, 1964). However, we measure substantial declines in counties’ market access when replacing railroads with the extended canal network proposed by Fogel. The proposed canals mitigate only 13% of the losses from removing the railroad network, though the implied annual economic benefits of these hypothetical canals would have exceeded their estimated annual capital costs. Second, in the absence of railroads, country roads might have been improved to reduce the costs of long-distance wagon transportation (Fogel, 1964). Replacing railroads with lower wagon transportation costs would have mitigated 21% of the losses from removing the railroad network. Most of this benefit to improved country roads would have continued in the presence of railroads, however, which suggests that railroads did not substantially discourage improvements in country roads. The absence of railroads might also have increased waterway shipping rates (Holmes and Schmitz, 2001), which is estimated to exacerbate by 20% the economic losses from removing railroads.

In summary, revisiting the historical impact of railroads on the American economy suggests a larger aggregate economic impact from railroads and market integration. Fogel (1964) calculates the impact of railroads based on willingness to pay for the transportation of agricultural goods, and our methodology is based on a similar willingness to pay for agricultural land.\footnote{We see our methodology as a natural extension of Fogel’s intuition, drawing on recent advances in trade theory, county-level data, and spatial computational tools. Whereas Fogel adds up the impact of railroads...} Beyond the substantial impacts on agricultural land value, however, our analysis an-
ticipates substantial declines in consumer welfare and total population in the absence of the railroads. Our estimates neglect further potential impacts on other sectors and technological growth, yet we hope that our ability to measure and analyze impacts of “market access” will spur further research on the aggregate impacts of railroads throughout the American economy.4

More broadly, this paper takes on the general methodological challenge of estimating aggregate treatment effects in empirical settings with substantial treatment spillover effects. Local railroad construction affects agricultural land values in all counties, to some degree, through interlinked trade networks. If railroads’ spillover impacts were confined to nearby areas, then the unit of analysis might be aggregated (e.g., Miguel and Kremer, 2004). As in many empirical settings, however, sufficient aggregation is empirically intractable. Our proposed solution uses economic theory to characterize how much railroads change each area’s market access; once the intensity of treatment is defined to reflect both direct and indirect impacts, relative empirical comparisons estimate the aggregate treatment effect of railroads on land values.5 Using economic theory as a guide, it is possible to estimate aggregate treatment effects in a reduced-form manner using relative variation. Extended results may then draw further on the model’s structure. Empirical research is increasingly estimating relative magnitudes by comparing areas more affected or less affected by some plausibly exogenous variation in treatment; we hope to encourage an extension of this research agenda to address the many important questions that are more aggregate in nature.

The rest of the paper is organized as follows. Section II reviews and extends Fogel’s partly by assuming the complete loss of agricultural land more than 40 miles from a natural waterway, we directly estimate the impact of railroads on all counties’ agricultural land values.

4In related work using a similar model, Redding and Sturm (2007) estimate the impact on population from changes in market access following the division and reunification of Germany, Hanson (2005) studies the correlation between US county-level wages and county-level market access from 1970 to 1990, and Redding and Venables (2004) and Head and Mayer (2011) study the relationship between national GDP and country market access. Donaldson (2015) estimates the income benefits from India’s railroads and shows that these are consistent with an Eaton and Kortum (2002) model similar to that used here. In contrast to Donaldson (2015), this paper measures the impact of railroads on market access (as derived from an Eaton-Kortum model extended to allow for labor mobility) to estimate the aggregate impact of railroads and to evaluate the impact of counterfactual scenarios. This paper’s methodological approach is more suited to settings with high mobility of labor, which appears to reflect the historical US economy more than the Indian economy. The concept of “market access” has been useful for empirical work (surveyed by Redding (2010)), though this paper is the first to leverage the concept of “market access” to estimate aggregate effects of place-based treatments (such as transportation infrastructure) from spatial comparisons using micro-geographical data. Redding (2010) highlights the surprising absence of research in this field that uses the price of an immobile factor, such as our use of land values, to estimate the benefits to each location in the presence of mobile factors.

5In the absence of an economic model, the spatial econometrics literature provides estimators for when treatment spillovers are a known function of geographic or economic distance Anselin (1988). Estimation of aggregate treatment effects requires a cardinal ranking of how much areas (or people) are exposed to the treatment, whereas an ordinal ranking is insufficient.
analysis of the railroads’ impact on the agricultural sector. Section III discusses our data collection and, in particular, our construction of a network database for calculating county-to-county transportation costs. Section IV derives our theoretical notion of “market access” and the resulting main empirical specification. Section V presents empirical estimates of the impact of changes in market access on changes in agricultural land value from 1870 to 1890. Section VI presents the baseline counterfactual impacts in 1890 from removing the railroad network, and summarizes the results’ robustness. Section VII analyzes counterfactual impacts on population and worker utility. Section VIII analyzes counterfactual impacts from replacing the railroad network with alternative transportation improvements. Section IX concludes. An online appendix contains accompanying material: additional details on the data construction and summary statistics; additional details on the robustness checks and the accompanying tables; and supplementary theoretical results from an extended version of our baseline model.

II American Railroads and “Social Saving” Estimates

By 1890, expansion of the railroad network had enabled a dramatic shift westward in the geographic pattern of agricultural production. Large regional trade surpluses and deficits in agricultural goods reflected the exploitation of comparative advantage. Fogel (1964) develops a “social saving” methodology for calculating the aggregate impact of railroads on the agricultural sector. We develop a different “market access” methodology for estimating the aggregate impact of railroads on the agricultural sector, though some aspects of our approach draw on Fogel’s intuition. It is therefore useful to begin with a summary of Fogel’s social saving analysis. We also take the opportunity to extend some of Fogel’s calculations, using modern spatial analysis tools and digitized county-level data.

Fogel (1964) estimates that the social saving from railroads in the agricultural sector in 1890 was no more than 2.7% of GNP. He divides this impact into that coming from interregional trade (0.6%) and intraregional trade (2.1%). For interregional trade, defined as occurring from 9 primary markets in the Midwest to 90 secondary markets in the East and South, freight costs were only moderately cheaper with the availability of railroads than when using only natural waterways and canals. Multiplying the difference in freight costs (with and without railroads) by the quantity of transported agricultural goods (in 1890), Fogel calculates the annual interregional social saving from railroads to be no more than $73 million or 0.6% of GNP. This number is proposed as an upper bound estimate because the approach assumes perfectly inelastic demand for transport, whereas the quantity of transported goods should be expected to decline with increased transportation costs.6

6Indeed, the total cost of agricultural interregional shipments would have nearly doubled in the absence of railroads.
For intraregional trade, defined as the trade from farms to primary markets, the impact of railroads was mainly to reduce distances of expensive wagon transportation. In the absence of railroads, farms would have incurred substantially higher costs in transporting goods by wagon to the nearest waterway to be shipped to the nearest primary market. In areas more than 40 miles from a waterway, wagon transportation may have become prohibitively expensive; indeed, Fogel refers to all land more than 40 miles from a navigable waterway as the “infeasible region” because it may have become infeasible for agricultural production if railroads were removed. Figure I, panel A, largely reproduces Fogel’s map of areas within 40 miles of a navigable waterway (shaded black), with the addition of areas within 40 miles of a railroad in 1890 (shaded light gray). Fogel bounds the economic loss in the “infeasible region” by the value of agricultural land in areas more than 40 miles from a waterway, which he calculates to generate approximately $154 million in annual rent. Adding the additional increase in transportation costs within the feasible region, which is bounded by $94 million using a similar approach to the interregional analysis, Fogel calculates the total annual intraregional impact to be no more than $248 million or 2.1% of GNP.

Fogel’s total social saving estimate of $321 million, or 2.7% of GNP, is generally interpreted as indicating a limited impact of the railroads, though the total loss of all agricultural land could only generate annual losses of $642 million or 5.35% of GNP. Fogel’s methodology is typically associated with the interregional social saving calculation and the analogous approach for the intraregional impact in the feasible region, though the annual rents from land value in the infeasible region is the largest component of the total estimate. Fogel emphasizes that losses in the infeasible region may well be overstated, as the railroad network could have been replaced with an extended canal network to bring most of the infeasible region (by value) within 40 miles of a waterway. Figure I, panel B, shows that much of the area beyond 40 miles from a navigable waterway would be within 40 miles of canals that might plausibly have been built if railroads did not exist (shaded dark gray). Fogel estimates that these canals would mitigate 30% of the intraregional impact from removing railroads.

Fogel faced a number of challenges in calculating the intraregional impact of railroads, some of which can be partly overcome by using modern computer software and digitized county-level data. One challenge was in measuring the area of the infeasible region, which is more accurate with the benefit of modern computer software. Using digitized maps of Fogel’s waterways and county-level data on agricultural land values (as opposed to state-level averages), we calculate a $181 million annual return on agricultural land in the infeasible region that is only moderately larger than Fogel’s approximation of $154 million. Consistent

\footnote{Unless otherwise noted, we use Fogel’s preferred mortgage interest rate (7.91%) to convert agricultural land values to an annual economic value. We also express annual impacts as a percent of GNP using Fogel’s...}
with 40 miles being a reasonable cutoff distance for the infeasible region, we calculate an annual return of only $4 million on agricultural land more than 40 miles from a waterway or railroad in 1890 (an infeasible calculation in Fogel’s era).

Fogel faced another challenge in calculating the intraregional social saving in the feasible region. Data limitations required a number of practical approximations and there are theoretical concerns about whether an upper bound estimate is meaningful given the potentially large declines in transported goods without railroads. An alternative approach, extending Fogel’s treatment of the infeasible region, is to assume that agricultural land declines in value the further it is from the nearest waterway or railroad. A simple implementation of this idea, though computationally infeasible in Fogel’s era, is to assume that land value decays linearly as it lies between 0 miles and 40 miles from the nearest waterway or railroad. Using modern computer software, we can calculate the fraction of each county within arbitrarily small distance buffers of waterways and/or railroads. Implementing this approach, we calculate the annual intraregional impact of removing railroads to be $319 million or 2.7% of GNP.

Figure I, panel C, shows smaller geographic buffers around waterways and railroads. In contrast to the 40-mile buffers in panel A, panel C shows 10-mile buffers that reflect the average wagon haul from a farm to a rail shipping point in 1890. The comparative advantage of railroads’ high density is more apparent at smaller distance buffers. Panel D adds 10-mile buffers around the proposed canals, which mainly run through sections of the Midwest and Eastern Plains. Replicating the above analysis of distance buffers, we calculate an annual loss of $221 million or 1.8% of GNP when replacing railroads with the proposed canals. This preliminary exercise finds that the proposed canals mitigate 31% of the intraregional impact from removing railroads, which is very close to Fogel’s original estimate of 30%.

The waterway network, particularly with extended canals, is moderately effective in bringing areas near some form of low-cost transportation. Construction of railroads was hardly limited to providing a similarly sparse network, however, and our later empirical estimates will show that high density railroad construction was particularly effective in providing preferred measure of GNP in 1890 ($12 billion).

\[\text{In practice, we take a discrete approximation to this linear decay function and assume that agricultural land loses 100\% of its value beyond 40 miles, 93.75\% of its value between 40 and 35 miles, 81.25\% of its value between 35 and 30 miles, and so forth until losing 6.25\% of its value between 5 and 0 miles. We calculate the share of each county that lies within each of these buffer zones (e.g., between 40 miles and 35 miles from a waterway or railroad). In addition, to avoid overstating the impact of railroads, we modify Fogel's calculation of the infeasible region to also reflect counties' imperfect access to railroads: since no county has all of its land within 0 miles of a waterway or railroad, all 1890 county land values already capitalize some degree of imperfect access. To calculate percent declines off the correct base, we adjust observed county agricultural land values to reflect their implied value if not for distance to a waterway or railroad. In the end, we calculate the implied decline in land value based on each county's land share within each 5-mile distance buffer of a waterway and subtracting the county's land share within that buffer of a waterway or railroad.}\]
nearby low-cost routes to markets.

Our empirical analysis will extend much of Fogel’s intuition for evaluating railroads’ aggregate impact on the agricultural sector in 1890. We maintain Fogel’s focus on the agricultural sector, as non-agricultural freight was geographically concentrated in areas with low transportation costs along waterways. We build on Fogel’s intuition that the value of agricultural land, as an immobile factor, should reflect the cost of getting agricultural goods to market. We choose transportation cost parameters to be comparable to Fogel’s chosen values (discussed in Section III.A), but explore robustness to these parameter choices in Section VI.B and the online appendix. Crucially, rather than follow Fogel in assuming a relationship between agricultural land values and the transportation network, we estimate this relationship. Rather than follow Fogel in assuming where goods are transported, we use insights from a general equilibrium trade model to help measure how counties value the transportation network. In particular, we measure how expansion of the railroad network affects counties’ market access and then estimate the impact of market access on agricultural land values. We then calculate the implied impact on land values from decreases in market access if railroads were eliminated, if railroads were replaced with the proposed canals, or under other counterfactual scenarios.

III Data Construction

This paper uses a new dataset on predicted county-to-county freight transportation costs, calculated using a newly-constructed geographic information system (GIS) network database. This network database shares some similarities to a hypothetical historical version of Google Maps, as a digital depiction of all journeys that were possible in 1870 and 1890 using available railroads, canals, natural waterways, and wagons.

Our measurement of market access relies on three components: (1) transportation cost parameters that apply to a given unit length of each transportation mode (railroad, waterway, and wagon); (2) a transportation network database that maps where freight could move along each transportation mode; and (3) the computation of lowest-cost freight routes along the network for given cost parameters. In this section, we describe the construction of these components and some data limitations.

III.A Transportation Cost Parameters

Our guiding principle in choosing transportation cost parameters has been to follow Fogel’s choice of these same parameters. We therefore set railroad rates equal to 0.63 cents per

---

9There has been extensive debate — surveyed by Fogel (1979) — regarding the social saving methodology and its application to evaluating the aggregate impact of railroads. We do not relitigate these issues, as most do not relate directly to our alternative methodological approach. Where relevant, we address some of the associated issues.
ton-mile and waterway rates equal to 0.49 cents per ton-mile.\textsuperscript{10} Transshipment costs 50 cents per ton, incurred whenever transferring goods to/from a railroad car, river boat, canal barge, or ocean liner.\textsuperscript{11} Wagon transportation costs 23.1 cents per ton-mile, defined as the straight line distance between two points.\textsuperscript{12} We later highlight some potentially important simplifications embedded in these cost parameter choices, and explore the results’ robustness to alternative transportation cost parameters.

Because wagon transportation is much more expensive than railroad or waterway transportation, the most important aspects of network database construction concern the required distances of wagon transportation. Indeed, Fogel (1964) and Fishlow (1965) both emphasize that railroads mainly lowered transportation costs by decreasing expensive wagon transportation through the interior of the United States.

### III.B Transportation Network Database

Creation of the network database begins with digitized maps of constructed railroads around 1870 and 1890. We are grateful to Jeremy Atack and co-authors for providing these initial GIS railroad files (Atack, 2013).\textsuperscript{13} These railroad files were originally created to define mileage of railroad track by county and year; by contrast, for our purposes, railroad lines are modified to ensure that GIS software recognizes that travel is possible through the railroad network.\textsuperscript{14}

The second step adds the time-invariant locations of canals, navigable rivers, and other natural waterways. We use Fogel’s definition of navigable rivers, which are enhanced to

\textsuperscript{10}Rates reflect an output-weighted average of rates for transporting grain and meat. Waterway rates include insurance charges for lost cargo (0.025 cents), inventory and storage costs for slower transport and non-navigable winter months (0.194 cents), and the social cost of public waterway investment (0.073 cents).

\textsuperscript{11}Fogel considers transshipment charges as a sub-category of water rates, but our modeling of transshipment points allows for a unified treatment of Fogel’s interregional and intraregional scenarios. Fogel’s sources record higher railroad freight costs per ton-mile for shorter routes, but we approximate these higher costs with a 100 cent fixed fee and a 0.63 cent fee per mile.

\textsuperscript{12}This rate reflects a cost of 16.5 cents per mile traveled and Fogel’s adjustment factor of 1.4 between the shortest straight line distance and miles traveled.

\textsuperscript{13}First, year-specific maps of railroads are “georeferenced” to US county borders. Second, railroad lines are hand-traced in GIS software to create a digital map of railroad line locations. The best practical approach has been to trace railroad lines from excellent maps in 1911 (Whitney and Smith, 1911), and then remove lines that do not appear in maps from 1887 (Cram, 1887) and 1870 (Colton, 1870).

\textsuperscript{14}We use GIS topology tools to ensure exact connections between all railroad line segments. Hand-traced railroad lines often contain small internal gaps that we have “snapped” together, though we have tried to maintain these gaps when appropriate (e.g., across the Mississippi river in the absence of a railroad bridge). The default option in GIS is for intersecting lines to reflect an overpass without a connection, but we have broken the network into segments that permit turns at each intersection. These modifications to the railroad network have little effect on total railroad track mileage by county and year. To minimize measurement error in changes, we created a final 1890 railroad file and modified that file to create a version for 1870 that omits lines constructed between 1870 and 1890.
follow natural river bends.¹⁵ For lakes and oceans, we saturate their area with “rivers” that allow for a large number of possible routes.¹⁶ Transshipment costs are incurred whenever freight is transferred to/from one of the four transportation methods: railroad, canal, river, and lake or ocean.¹⁷

The third step connects individual counties to the network of railroads and waterways. We measure average travel costs between counties by calculating the travel cost between the geographical center (or centroid) of each pair of counties. County centroids must be connected to the network of railroads and waterways; otherwise, lowest-cost travel calculations assume that freight travels freely to the closest railroad or waterway. We create wagon routes from each county centroid to each nearby type of transportation route in each relevant direction.¹⁸ Because the network database only recognizes lines, we also create direct wagon routes from every county centroid to every other county centroid within 300km.¹⁹

The fourth step refines centroid-to-network connections due to the importance of wagon distances to overall freight costs. For example, when a railroad runs through a county, the centroid’s nearest distance to a railroad does not reflect the average distance from county points to a railroad.²⁰ We create 200 random points within each county, calculate the distance from each point to the nearest railroad, and take the average of these nearest distances. We then adjust the cost of travel along each centroid connection to within-county railroads to reflect that county’s average travel cost to a railroad. We then repeat this procedure for centroid connections to navigable rivers and canals. This refinement to the network database allows the empirical analysis to exploit precise variation on the intensive margin of county access to railroads and waterways as the density of the railroad network increases from 1870.

¹⁵ Fogel’s classification of “navigable” rivers may be overly generous in some cases (Atack, 2013).
¹⁶ We do not permit direct access to lakes and oceans at all points along the coast; rather, we restrict access to “harbors” where the coast intersects interior waterways. We create additional “harbors” where the railroad network in 1911 approaches the coastline, which also permits direct “wagon” access to the coast at these points.
¹⁷ Overlapping railroads and waterways do not connect by default; instead, we create connections among railroads and waterways to allow for fixed transshipment costs. The need to include transshipment costs is the main reason why it is not possible to model the network using a raster, assigning travel costs to each map pixel.
¹⁸ Many such connections were created by hand, which raises the potential for errors, but we have used GIS topology tools to ensure that these connections are exactly “snapped” and classified correctly by type (centroid-to-railroad, centroid-to-river, etc.).
¹⁹ The direct wagon routes are restricted to be over land, but there is no adjustment for mountains or other terrain; in practice, the long-distance wagon routes are already very costly. The cost of wagon transportation also already includes an adjustment for the general inability to travel in straight lines along the most direct route.
²⁰ Fogel recognized the importance of measuring this within-county distance and his ideal solution was to break each county into small grids and take the average of nearest distances from each grid to a railroad. However, due to technical limitations, Fogel approximated this average distance using one-third of the distance from the farthest point in a county to a railroad.
to 1890.

Figure II shows part of the created network database. Panel A shows natural waterways, including the navigable rivers and routes within lakes and oceans. Panel B adds the canal network, which is highly complementary with natural waterways. Panel C adds railroads constructed in 1870, and then Panel D adds railroads constructed between 1870 and 1890. Early railroads were complementary with the waterway network; by 1870 and especially by 1890, however, the railroad network is more of a substitute for the waterway network.

As a summary, in the accompanying online appendix, Appendix Table 1 lists each segment of the transportation network database, a brief description, and its assigned cost.

III.C Limitations of the Network Database

There are several limitations of the constructed network database. First, the constructed network database is mainly restricted to transportation linkages within the United States. The data only include US counties’ access to other US counties. As a robustness check, however, we incorporate international markets by assigning additional product demand and supply to US counties with major international ports.

Second, freight rates are held constant throughout the network database. Freight rates may vary with local demand and market power in the transportation sector, though local variation in freight rates would then be partly endogenous to local economic outcomes. Thus, there are econometric advantages to using Fogel’s average national rates. We hold rates fixed in 1890 and 1870, such that measured changes in trade costs and market access are determined by changes in the railroad network.

Third, freight rates are not allowed to vary by direction. Some freight rates might vary by direction due to back-haul trade relationships or waterway currents, so we explore the results’ robustness to different waterway and railroad rates. Wagon rates may dramatically overstate some transportation costs in Western states where cattle were driven to market, though we also report estimates when excluding the Western states. Further, we examine how particular regions are influencing the results by allowing the impact of market access to vary by region.

Fourth, there are no congestion effects or economies of scale in transporting goods. We do not restrict locations where trains can turn or switch tracks, so actual railroad transportation routes may be less direct. We also do not measure differences in railroad gauges, which

---

21 There are two exceptions. First, the network database includes a Canadian railroad line between New York and Michigan. Second, the database includes a waterway route from the Pacific Ocean to the Atlantic Ocean (i.e., around Cape Horn), and the empirical analysis explores the results’ robustness to varying the cost of this waterway connection.

22 The Western regions are not central to the empirical analysis, which draws on within-state variation in changes in market access.
required some additional costs in modifying railroad cars and tracks. In robustness checks, we allow for higher railroad costs that reflect less-direct routes or periodic transshipment within the railroad network.

Fifth, we do not directly consider the speed or bulk of freight transportation. While railroads might transform the long distance trade of time-sensitive goods, we maintain Fogel’s focus on railroads’ aggregate importance in the bulk transportation of storable agricultural commodities. The assumed waterway rate, taken from Fogel, includes an adjustment for higher storage and inventory costs associated with slower water transportation, which makes up 40% of the total waterway rate.

Overall, we should expect that measurement of transportation costs will be robust to even large percent differences in the chosen railroad and waterway rates. Recall that 10 miles of wagon transportation are roughly equivalent to 375 – 475 miles of railroad or waterway transportation. Thus, the estimated transportation costs are dominated by the order-of-magnitude difference between the cost of wagons and the cost of railroads or waterways.\(^{23}\)

### III.D Transportation Route Cost Calculations

We use the complete network database to calculate the lowest-cost route between each pair of counties, i.e., 5 million calculations.\(^{24}\) Initially, we calculate the lowest-cost routes under two scenarios: (1) the wagon, waterway, and railroad network in 1870, and (2) the wagon, waterway, and railroad network in 1890. These transportation costs are used to calculate counties’ market access in 1870 and 1890, so that we can estimate the impact of changing market access on changes in land values. For the later analysis, we calculate the lowest-cost routes under counterfactual scenarios: removing the 1890 railroad network; replacing the 1890 railroad network with an extended canal network; replacing the 1890 railroad network with improved country roads (decreased wagon freight rates); and removing the 1890 railroad network and increasing water freight rates (due to decreased competition).

### III.E County-level Census Data

County-level data are drawn from the US Censuses of Agriculture and Population (Haines, 2005). The two main variables of interest are: the total value of agricultural land, and the total population. We adjust data from 1870 to reflect 1890 county boundaries (Hornbeck, 2010).

In the accompanying online appendix, Appendix Figure 1 maps the 2,327 counties included in the main regression analysis, which includes all counties with reported land value

---

\(^{23}\)In robustness checks, we also allow for lower transportation costs by wagon.

\(^{24}\)In principle, it is a daunting task to find the optimal route between two points on such a dense network; in practice, the computation is improved dramatically by applying Dijkstra’s algorithm (see, e.g., Ahuja, Magnanti and Orlin, 1993, for a textbook treatment).
data in 1870 and 1890. Appendix Figure 2 maps a larger sample of 2,782 counties included in the counterfactual analysis, which includes 455 additional counties that report land value data in 1890 (but not in 1870). In our calculation of counties’ market access, we calculate counties’ access to all other counties that existed in that period, regardless of whether those other counties are included in the regression sample.

For the data on agricultural land value, the reported data include the combined value of agricultural land, buildings, and improvements. We follow Fogel in deflating the reported Census data to reflect the “pure” value of agricultural land (Fogel, 1964, pp. 82-83), such that gains in land value reflect true economic gains and not the cost of fixed investments.\(^{25}\) In robustness checks, we explore further adjustments to the land value data that reflect county-level changes in land settlement or land improvement.

For the data on population, there are some known challenges with undercounting in the Census. In robustness checks, we adjust population data to reflect undercounting that is systematically more severe in 1870 and in the South (Hacker, 2013). We also explore adjusting population data to reflect the presence of trade with international markets, inflating the population in major US ports to reflect the value of imports and exports (divided by income per capita).

The online appendix provides some additional details on these county-level data. Appendix Table 2 provides summary statistics on county characteristics in 1870, in 1890, and changes between 1870 and 1890. Appendix Figure 3 maps counties’ change in land value between 1870 and 1890, with darker shades representing greater increases in land value.

IV A “Market Access” Approach to Valuing Railroads

The empirical analysis is guided by a model of trade among US counties that specifies how each county is affected by changes in the national matrix of county-to-county trade costs. The model contains thousands of counties, each with interacting goods markets and factor markets, that generate positive and negative spillovers on other counties. Nevertheless, under a set of assumptions that are standard among modern trade models, all direct and indirect impacts of changing trade costs are reflected, in equilibrium, in changes to a county’s “market access.”\(^{26}\)

\(^{25}\)Fogel reports the “pure” value of agricultural land by state, after subtracting estimates for the value of buildings and land improvements. We multiply counties’ reported Census data by Fogel’s estimated “pure” value of agricultural land (in their state) divided by the reported Census value of agricultural land (in their state), which reduces the total value of agricultural land in our sample by 39%. This adjustment to land value data affects the magnitude of the counterfactual estimates, but does not affect the regression estimates that are conditional on state-by-year fixed effects.

\(^{26}\)These modeling assumptions are used extensively in the fields of international trade and economic geography, and reflect recent best practice to gain traction in general equilibrium spatial settings with many regions that trade subject to trade costs.
The model implies a simple log-linear relationship between county agricultural land value and county market access, appropriately defined. While the model requires particular assumptions to arrive at this parsimonious solution to the challenges posed by general equilibrium spatial spillovers, the predicted relationship also has an atheoretical appeal in capturing the impact of railroads. County market access increases when it becomes cheaper to trade with another county, particularly when that other county has a larger population. Guided by the model, we present our main empirical specification that regresses county agricultural land value on county market access and a set of control variables.

IV.A A Model of Trade Among US Counties

The economy consists of many trading counties, each indexed by $o$ if the origin of a trade and by $d$ if the destination. Our baseline model contains just one sector, though the online appendix includes an extended model with an additional sector (and where the two sectors interact through input-output linkages as well as factor and product markets). Agents in the model consume a continuum of differentiated goods varieties (indexed by $j$), and tastes over these varieties take a CES form (with elasticity $\sigma$).

Therefore, a consumer living in county $o$, who receives income $Y_o$ and faces a vector of prices $P_o$, experiences indirect utility:

\begin{equation}
V(P_o, Y_o) = \frac{Y_o}{P_o},
\end{equation}

where $P_o$ is the ideal price index (a standard CES price index) over the continuum of varieties.

Producers in each county use a Cobb-Douglas technology to produce varieties from labor, capital, and land. The marginal cost of producing variety $j$ in county $o$ is:

\begin{equation}
MC_o(j) = q_o^\alpha w_o^\gamma r_o^{1-\alpha-\gamma} z_o(j),
\end{equation}

where $q_o$ is the agricultural land rental rate, $w_o$ is the wage rate, $r_o$ is the capital rental rate, and $z_o(j)$ is a Hicks-neutral productivity shifter that is exogenous and local to county $o$. We follow Eaton and Kortum (2002) in modeling these productivity shifters by assuming that

\footnote{The elasticity of substitution is not restricted (beyond the discussion in footnote 30 below); that is, $\sigma$ could be high if varieties are similar. Anderson, de Palma and Thisse (1992) provide an attractive microfoundation for aggregate-level CES preferences: if individual agents desire only one variety of the good (their “ideal variety”) and agents’ utilities from their ideal varieties are distributed in an extreme value (or “logit”) fashion, then aggregate consumption data from a population of many such agents behaves as though all agents have CES preferences over all varieties (where, in such an interpretation, $\sigma$ indexes the inverse of the dispersion of the utility levels that agents enjoy from their ideal varieties).}

\footnote{That is, $P_o \equiv \left[ \int_0^\infty (p_o(j))^{1-\sigma} dj \right]^{1/(1-\sigma)}$, where $n$ denotes the (exogenous) measure of varieties available to consumers and $p_o(j)$ is the price for which variety $j$ sells in county $o$.}

14
each county draws its productivity level, for any given variety \( j \), from a Fréchet (or Type II extreme value) distribution with CDF given by: 
\[
F_o(z) = 1 - \exp(-A_o z^{-\theta}), \quad \text{with } \theta > 1. \tag{29,30}
\]

This distribution captures how productivity differences across counties give incentives to specialize and trade, where these incentives are inversely related to \( \theta \).\(^{31}\) We assume perfect competition among producers.\(^{32}\)

There are costs to trading varieties across counties. Remote locations pay high prices for imported varieties and receive low prices for varieties they produce, as this is the only way that locations can be competitive in distant markets. We model trade costs using a simple and standard “iceberg” formulation: a proportional trade cost \( \tau_{od} \) is applied to each unit of the variety shipped.\(^{33}\) When a variety is made in county \( o \) and sold locally in county \( o \), its price is \( p_{oo}(j) \); but when this same variety is made in county \( o \) and shipped to county \( d \), it will sell for \( p_{od}(j) = \tau_{od} p_{oo}(j) \). Trade is potentially costly, so \( \tau_{od} \geq 1 \).

The physical supply of land is fixed by county geographic borders, with \( L_o \) units available in county \( o \), and we consider impacts on the total value of agricultural land in each county.\(^{34}\) Given that much land is unsettled prior to the railroads, our empirical analysis also estimates a decomposition of the total impact on agricultural land value per county acre into impacts on the intensive margin (land value per farm acre) and the extensive margin (farm acres per county acre). We assume that capital is perfectly mobile, such that the return to capital is equalized across counties (i.e., \( r_o = r \)), though the empirical analysis will include geographic controls that absorb regional variation in the interest rate. We further assume that the United States faces a perfectly elastic supply of capital.\(^{35}\) We also assume that workers

\(^{29}\)Following Eaton and Kortum (2002), an intuitive rationale for this particular functional form for the distribution of productivities is that it reflects the limiting distribution when producers receive technologies from any distribution and discard all but the best.

\(^{30}\)An additional parameter restriction, \( \theta > \sigma - 1 \), is required for the integral in \( P_o \) to be finite. However, Eaton, Kortum and Sotelo (2012) demonstrate this restriction is no longer required when there are a finite number of varieties, as in reality. Our continuum of varieties assumption can be thought of as an analytically convenient approximation to the true, finite number of varieties.

\(^{31}\)More specifically, the parameter \( A_o \) captures county-specific (log) mean productivity, which corresponds to each county’s level of absolute advantage. The parameter \( \theta \) captures, inversely, the (log) standard deviation of productivity, which corresponds to the scope for comparative advantage. A low \( \theta \) means county productivity draws are dispersed, creating large incentives to trade on the basis of productivity differences.

\(^{32}\)An alternative (and observationally equivalent) formulation, following Melitz (2003), would assume that firms compete monopolistically with free entry such that all firms’ expected profits are zero and draw their productivity levels \( z \), following Chaney (2008) and others, from a Pareto distribution \( G_o(z) = 1 - (z/A_o)^{-\theta} \), as typically seen in firm-level datasets (e.g., Axtell, 2001).

\(^{33}\)While we measure the absolute cost of trade between counties, we express this cost in proportional terms using Fogel’s average value of transported agricultural goods.

\(^{34}\)Landowners are not restricted to own land in their county of residence but, because we do not observe land ownership by county, we assume that land is owned (and hence the rents earned by landowners are spent) in proportion to county populations.

\(^{35}\)Specifically, our baseline assumption — which is not needed until we solve for general equilibrium counterfactuals in Section VII — is that the nominal price of capital relative to the price index in New
are perfectly mobile across counties, at least over a period of many years. As a result of workers’ endogenous option to work in other counties, workers’ utility levels are equalized across counties in equilibrium and hence nominal wages satisfy:

\( w_o = \bar{U} P_o, \)

where \( \bar{U} \) is the level of utility obtained by workers in each county. As we discuss further in Section IV.C below, the level of \( \bar{U} \) does not affect any of the regressions that we estimate, as any changes in \( \bar{U} \) are not predicted to be proportionally differential by county, and are therefore absorbed by the regression constant in our log-linear regressions. However, the endogenous determination of \( \bar{U} \) is potentially important for our counterfactual exercises and we discuss this further in Sections IV.C and VII.

IV.B Solving the Model

Prices and trade flows: First, we solve for the trade in goods from each origin county \( o \) to each other destination county \( d \). Due to perfect competition, the marginal cost of producing each variety is equal to its price. Substituting marginal costs from each supply location \( o \) (equation 2) into the demand for varieties in county \( d \), and allowing consumers to buy goods from their cheapest source of supply in equilibrium, Eaton and Kortum (2002) derive two important results for our application. The first is that the consumer price in destination location \( d \) is given by:

\[
(P_d)^{-\theta} = \kappa_1 \sum_o A_o (q_o^\gamma w_o^\gamma)^{-\theta} \tau_{od}^{-\theta} \equiv CMA_d.
\]

We follow Redding and Venables (2004) in referring to this (inverse transformation of the) price index as \( CMA_d \) or “consumer market access.” Consumer market access in county \( d \) represents its access to cheap products: it is a weighted sum of productivity-adjusted costs of production in each origin market \( o \) that could supply market \( d \), with weights declining in the cost of trading from \( o \) to \( d \) (i.e., \( \tau_{od} \)).

A second important result from Eaton and Kortum (2002) describes \( X_{od} \), the value of total exports from \( o \) to \( d \), as:

\[
X_{od} = \kappa_1 A_o (q_o^\gamma w_o^\gamma)^{-\theta} \tau_{od}^{-\theta} CMA_d^{-1} Y_d.
\]

York City (i.e., the largest point of entry and exit for internationally-traded goods and a financial center) is fixed. We obtain very similar results if the nominal price of capital is instead constant relative to a population-weighted average of all counties’ price indices.

\[ \text{Here, } \kappa_1 \text{ is a constant defined by } \kappa_1 = \left[ \Gamma \left( \frac{\theta + 1 - \sigma}{\theta} \right) \right]^{-\theta/(1-\sigma)} \Gamma^{-(1-\alpha-\gamma)\theta}, \text{ where } \Gamma(\cdot) \text{ is the } \Gamma \text{ function defined by } \Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dx. \]
From equation (5), county \( o \) sends more goods to county \( d \) if county \( o \) is relatively productive (high \( A_o \)) or relatively low cost (low \( w_o \) or low \( q_o \)). County \( o \) also sends more goods to county \( d \) if county \( d \) has high total income (high \( Y_d \)) or low overall consumer market access (low \( CMA_d \)), meaning that county \( o \) faces less competition when selling to market \( d \).

Equation (5) is known as a gravity equation, which governs trade flows in this model. The gravity equation is appealing because it dramatically simplifies a complex general equilibrium problem of spatial competition. In addition, an empirical appeal of the gravity equation is that it appears to provide a strong fit for trade-flow data in many contexts (e.g., Anderson and van Wincoop, 2003, 2004; Combes, Mayer and Thisse, 2008; Head and Mayer, 2014).

Land rental rate: While trade flows between 19th century US counties are unobserved, the gravity equation implies tractable and empirically useful expressions for the land rental rate (\( q_o \)), a proxy for which is observed (as discussed in Section III.E above). Under the assumption of Cobb-Douglas technology, land is paid a fixed share of total output \( Y_o \), so

\[
q_o L_o = \alpha Y_o.
\]

Using equations (3) and (4) and taking logs, equation (5) implies:\(^\text{37}\)

\[
(1 + \alpha \theta) \ln q_o = \kappa_2 + \ln \left( \frac{A_o}{L_o} \right) - \gamma \theta \ln U + \gamma \ln CMA_o + \ln FMA_o,
\]

where \( FMA_o \) refers to “firm market access” for goods from origin \( o \) and is defined as:

\[
FMA_o \equiv \sum_d \tau_{od}^{-\theta} CMA_d^{-1} Y_d.
\]

Firm market access (\( FMA_o \)) is a sum of terms over all destination counties \( d \) to which county \( o \) sells its goods. These terms include the size of the destination market (given by total income, \( Y_d \)) and the competitiveness of the destination market (given by its \( CMA_d \) term). All terms are inversely weighted by the cost of trading with each distant market (i.e., by \( \tau_{od}^{-\theta} \)).

Firm market access is conceptually similar to consumer market access, as both are increasing in cheap access to large markets with few trade partners. To see this similarity explicitly, note that it is possible to write \( CMA_d \) as:\(^\text{38}\)

\[
CMA_d = \sum_o \tau_{od}^{-\theta} FMA_o^{-1} Y_o.
\]

\(^\text{37}\) Here, \( \kappa_2 \equiv \ln(\kappa_1 \alpha) \). Goods markets clear, so all produced goods are bought (\( Y_o = \sum_d X_{od} \)).

\(^\text{38}\) This result can be obtained by summing equation (5) over all destinations \( d \) and substituting \( A_o (q_o^\theta w_o^\gamma)^{-\theta} \) into equation (4).
Under the additional restriction that trade costs are symmetric (i.e., $\tau_{od} = \tau_{do}$ for all counties $d$ and $o$), which is satisfied by the freight costs data we have constructed in Section III.D, any solution to equations (7) and (8) must satisfy $FMA_o = \rho CMA_o$ for some scalar $\rho > 0$. That is, $FMA$ and $CMA$ are equal to one another up to a proportionality whose value does not affect our analysis. Therefore, we simply refer to “market access” ($MA$) to reflect both concepts of market access. Formally, we let $MA_o \equiv FMA_o = \rho CMA_o$ for all counties $o$. Using the fact that $Y_d = \frac{w_d N_d}{\gamma}$, where $N_d$ refers to the (endogenous) number of workers living in county $d$, as well as equation (3), equation (7) implies that:

$$\ln q_o = \kappa_3 + \left(1 - \frac{\theta}{1 + \alpha \theta}\right) \ln (A_o) - \frac{2 + \alpha \theta}{1 + \alpha \theta} \ln (L_o) + \left[1 + \theta (1 + \alpha + \gamma) \theta \ln \bar{U}\right] \ln (MA_o).$$

Equation (10) provides a useful guide for the empirical analysis. Equilibrium land rental rates ($q_o$) are log-linear in just one endogenous county-specific economic variable: market access ($MA_o$). This notion of market access both captures firms’ desire to sell goods elsewhere for a high price and captures consumers’ desire to buy goods from elsewhere at a low price. Immobile land in county $o$ will be more valuable if county $o$ has cheaper access to large uncompetitive markets and/or cheaper access to labor (by offering mobile workers a location in which they can enjoy cheaper access to goods).

Finally, similar derivations imply that the equilibrium population $N_o$ in any location $o$ obeys a similar relationship:

$$\ln N_o = \kappa_5 + \left(1 - \frac{\theta}{1 + \alpha \theta}\right) \ln (A_o) - \frac{2 + \alpha \theta}{1 + \alpha \theta} \ln (L_o) + \left[1 + \theta (1 + \alpha + \gamma) \theta \ln \bar{U}\right] \ln (MA_o).$$

That is, county population also responds log-linearly to differences in market access, in this setting with free labor mobility. We also estimate this relationship empirically in Section VII.

---

39 Here, $\kappa_3 \equiv \frac{\bar{U} \rho^\frac{1+\theta}{\gamma}}{1+\alpha \theta}$.  
40 Here, $\kappa_4 \equiv \frac{1}{1+\alpha \theta} (\kappa_2 - \gamma \ln \rho - \gamma \theta \ln \bar{U})$.  
41 Here, $\kappa_5 \equiv \frac{\ln (\kappa_1 \alpha \theta^{-\gamma})}{1+\alpha \theta} - \ln \left(\frac{\alpha}{\theta}\right) - \frac{1+\theta (\alpha + \gamma)}{1+\alpha \theta} \ln \bar{U}$.
IV.C Using the Model to Inform Empirical Work

Equation (10) has three key implications for the empirical analysis. First, all economic forces that make goods markets and factor markets interdependent across counties are represented by “market access.” Thus, both direct and indirect impacts of railroads are captured by analyzing changes in market access. For example, county A receiving a railroad line would affect other counties: those that can now trade with county A, those that had been trading with county A, those that had traded with county A’s previous trade partners, those that had traded with county A’s new trade partners, and so on. Even if access to railroads were randomly assigned to a “treatment” county, “control” counties would be affected and a regression of land rents on railroad access would produce biased estimates of railroads’ aggregate impact. However, a regression of land rents on market access would be free of this bias, in the context of our model, because all counties’ market access will adjust to changes in the railroad network. In addition, the aggregate effect of counterfactual changes to the transportation network (such as the removal of railroad lines or their replacement with a proposed canal network) can be calculated by substituting the counterfactual values of $\tau_{ad}$ into $MA_o$ and then substituting the resulting counterfactual $MA_o$ into equation (10). We perform such calculations in Section VI.

The second key implication of equation (10) is that a county’s market access can increase or decrease due to changes in the railroad network far beyond that county’s borders. Thus, the empirical estimation is not identified only from particular counties gaining railroad access, which might otherwise be correlated with land rental rates. This prediction of the model suggests some robustness checks, control variables, and instrumental variable approaches that might purge the empirical estimates of endogeneity bias arising from local railroad placement decisions, all of which we pursue below.

Finally, a counterfactual change in the transportation network might affect aggregate worker utility ($\bar{U}$). Two extreme scenarios are possible, with reality surely in-between the two cases. In one extreme scenario, if there is a perfectly elastic supply of international migrants, $\bar{U}$ would be pinned down by workers’ utility levels abroad but the aggregate number of workers $\bar{N} \equiv \sum_o N_o$ would change. In the other extreme scenario, if there is a perfectly inelastic supply of international workers, the aggregate US population would not change as a result of the counterfactual transportation costs but $\bar{U}$ would change. As we

---

42 This statement is true holding constant aggregate worker utility ($\bar{U}$), which is an assumption to which we return shortly.

43 This scenario assumes that the US labor market is vanishingly small relative to the world labor market, such that technological changes affecting labor demand in the US have no appreciable effect on the level of world worker utility $\bar{U}$.

44 Further, this would affect all counties’ land rents because $\kappa_3$ in equation (10) depends on $\bar{U}$. 

19
discuss in Section VII below, we can use the model here to solve for the resulting effects — that is, to solve for the change in either $\bar{U}$ or $\bar{N}$ — in these two extreme scenarios and calculate the associated impacts on land value. We do so to explore the relevance of these aggregate phenomena, and we gauge the potential for intermediate cases by estimating the domestic responsiveness of counties’ population to changes in counties’ market access, as guided by equation (11).

IV.D From Theory to an Empirical Specification

While equation (10) provides a useful guide for the empirical analysis, there are several issues involved with its direct empirical implementation.

First, while the model describes economic impacts in a one-sector model, we now turn to estimating the impact in the agricultural sector using data on agricultural land values only. Our main model refers to the price of land generally, but the online appendix derives predictions for the price of agricultural land specifically. In particular, the appendix outlines a model with separate agricultural and manufacturing sectors, clarifying how railroads might impact agriculture through impacts on: consumption of manufactured goods, agricultural firms selling inputs to manufacturing firms and consumers, and agricultural firms buying inputs from manufacturing firms. In all cases, the value of agricultural land remains log-linear in a series of different “market access” terms that take a similar functional form to the single notion of market access in equations (9) and (10). These “market access” terms all reflect a trade-cost-weighted sum over population in areas producing or consuming particular goods (e.g., urban areas, rural areas, all areas), such that, at least in our setting, empirical approximations of these terms are extremely highly correlated with each other. There is little hope of distinguishing their impacts, so we simply note that market access might reflect any number of these different mechanisms. We later verify the robustness of our estimates to restricting the definition of market access to include counties’ access to urban areas only (which might be the particular markets that rural areas value gaining access to).\footnote{Further, for an extreme case in which prices are pinned down by the cost of reaching international markets and counties only value access to the “world economy,” we consider measuring counties’ access to only New York City (i.e., the largest hub for international trade and the most populated US city).}

Second, a related challenge is that the Census of Agriculture does not report on the agricultural value of all land in each county. To measure counties’ total value of land for agricultural use, we consider restricting the sample to rural areas (where agricultural land would be minimally affected by local demand for land by manufacturing or housing).\footnote{In the appendix, we continue to assume that the supply of agricultural land is fixed. As a consequence, impacts of market access on local manufacturing do not directly change the supply of land to the agricultural sector. While manufacturing and housing use relatively little land, compared to the agricultural sector, we consider restricting the sample to rural areas to reduce the potential for changes in manufacturing and housing to directly impact the supply of agricultural land.}
agriculture, we use the reported total value of land in farms and assume that land not in farms has zero agricultural value. Unsettled land in the public domain could have been obtained at very low cost, and we explore the results’ robustness to assuming that lands settled between 1870 and 1890 had some unmeasured value even in 1870. Land used for non-agricultural purposes might also have agricultural value, which motivates robustness checks that restrict the sample to include only rural counties.

Third, while we can obtain data on the value of agricultural land, the model relates counties’ market access to the rental cost of land. Of course, land values are closely related to land rents and the interest rate. Land values reflect both contemporaneous rents and discounted future rents, and so any correctly anticipated changes in market access would attenuate the estimated impact of changes in market access. The model refers to the per acre price of land, where the quantity of land is fixed by counties’ geographic borders, and so we analyze the total value of land per county acre. We then consider how the estimated total impact decomposes into impacts on the intensive margin (land value per farm acre) and the extensive margin (farm acres per county acre).

Fourth, a potential challenge is that we do not directly observe county productivity \( A_0 \). Our analysis relates changes in land value to changes in market access, however, which absorbs any fixed component of county productivity. We then assume that changes in county productivity are orthogonal to changes in market access from 1870 to 1890, after controlling for counties’ geographic location (state, longitude, latitude). Additional robustness checks include controls for region-specific or subregion-specific changes in productivity.

Fifth, the calculation of market access (via equation 9) requires the measurement of all

---

47 Section III.E discusses modifying the reported Census data to obtain a measure of the value of agricultural land only, which does not include the value of agricultural buildings and improvements.

48 Formally, it is sufficient for us to assume that \( V_0 = q_0/r \), where \( V_0 \) is the land value and \( r \) is a fixed interest rate. In practice, the empirical results would be unaffected if the interest rate varies by county, state-year, or with any of the control variables in the empirical specifications.

49 We suspect that residual changes in market access are difficult to anticipate. But because some may have been able to anticipate local railroad construction, we also report estimates that control directly for changes in local railroad density. Land values may diverge from land rents during periods of systematic optimism or pessimism, though we control for regional shocks to land values.

50 In practice, we analyze the total value of land in the county, but this is numerically equivalent because the number of acres in the county is absorbed by county fixed effects or by differencing the regression (in logs). We assume that land not in farms has zero agricultural value, though we later relax that assumption. By contrast, it would be inappropriate to analyze only the value of land per acre in farms: this would neglect the central effect from increased economic value of previously “infeasible” land, and there would be bias from changes in the composition of farmland.

51 Because the productivity term \( A_0 \) enters log-linearly in equation (10), we control for this term using county fixed effects, state-by-year fixed effects, and year-interacted cubic polynomials in the latitude and longitude of the country centroid. We include cubic polynomials in counties’ longitude and latitude to control flexibly for geographic differences, though we also explore robustness to lower-order and higher-order polynomials.
trade costs ($\tau_{od}$). We approximate these trade costs using the calculated county-to-county lowest-cost freight transportation routes (described in Section III.D above), expressed in proportional terms based on the average value of transported agricultural goods. We treat each county as a point with common prices and wages throughout, though the calculated centroid-to-network distances were adjusted to reflect average distances from many points in each county. The baseline results use trade costs calculated using freight rates drawn from Fogel (1964), though we explore the sensitivity of our results to the particular freight rates that enter $\tau_{od}$ in $MA_o$.

Sixth, the market access term ($MA_o$) is not directly observed because some destination characteristics are unobserved.\textsuperscript{52} Based on equation (9), however, it is possible to use data on each county’s population ($N_o$) to express each county’s market access $MA_o$ as an implicit function of the market access of all other counties. We can solve this implicit function numerically and report empirical estimates that use counties’ derived market access in 1870 and 1890.\textsuperscript{53} This approach accords exactly with equation (9), but the calculation of these terms depends on running the data through the particular structure of the model. A simpler approach, which is also less model dependent, uses the following expression that provides a first-order approximation to counties’ market access:

\begin{equation}
MA_o \approx \sum_d \tau_{od}^{-\theta} N_d.
\end{equation}

The results are not sensitive to our use of the $MA$ approximation in equation (12), as we document below, because (the log of) this approximated term is highly correlated with (the log of) the $MA$ term derived from solving equation (9). We also explore robustness to proxying for market demand using the Census-reported value of real and personal property, rather than using population $N_d$ as in equation (12).

Seventh, the population $N_d$ in each county $d$ is endogenously co-determined with the land rental rate $q_o$ in county $o$, which would generate endogeneity bias in a regression based on equation (10). A particular instance of this concern arises because $N_o$ is included in the definition of $MA_o$ in equation (12). For this reason, we exclude each county’s own population from its measure of market access,\textsuperscript{54} though our results are insensitive to this decision because the contribution of $N_o$ to $MA_o$ is small for most counties. More generally, a county’s land value may be affected by local shocks that affect nearby counties’ population.

\textsuperscript{52}From equation (4), the wage $w_o$ and the technology term $A_o$ are unobserved.

\textsuperscript{53}With $C$ counties, equation (9) becomes a system of $C$ equations in $C$ unknowns. Following the results cited in Allen and Arkolakis (2014) this system has a unique solution up to a scalar multiple that affects all counties’ $MA_o$ values equally.

\textsuperscript{54}Throughout the empirical analysis, we work with the variable $MA_o \approx \sum_{d \neq o} \tau_{od}^{-\theta} N_d$. 

22
In robustness checks, we calculate each county’s market access when omitting other counties within particular distance buffers around that county. In further robustness checks, we calculate each county’s market access in 1870 and 1890 when holding all counties’ population fixed at 1870 levels.

Eighth, and finally, the expression for market access in equation (12) requires an estimate of $\theta$ (a parameter known as the “trade elasticity”). Different values of $\theta$ will have a mechanical influence on the estimated impact of market access, by changing the definition of market access, but it is more relevant to consider whether the estimated counterfactual impacts are sensitive to the choice of $\theta$. While $\theta$ depends on the empirical context, values estimated and used in the literature have typically straddled the two extreme estimates in Eaton and Kortum (2002) of 3.60 and 12.86 (though Eaton and Kortum’s (2002) preferred estimate is 8.28). In Section VI.B and the online appendix, we verify the robustness of our results to choosing alternative values for $\theta$ within (and even outside of) this range. However, we focus on estimates when setting $\theta$ equal to 8.22, which is the value we obtain by drawing on the model’s structure to estimate the value of $\theta$ that best fits the data in our empirical setting. We obtain this estimate from a non-linear least squares (NLS) routine, noting that equation (10) is non-linear in $\theta$. The estimated value of 8.22 has a 95% confidence interval between 3.37 and 13.18 (based on block-boosters at the state-level, with 200 replications).

Our main empirical approximation of market access recalls an older concept of “market potential,” based on the number and size of markets available at low trade costs (Harris 1954). Harris’s market potential term effectively equals $\sum_{d\neq o} (\tau_{od})^{-1} N_d$, though Harris used distance as a proxy for trade costs. Our focus on trade costs, rather than distance, allows us to consider how changes in the national railroad network affect each county even though...
geographic distances remain fixed. The remaining practical difference for our initial empirical estimation is that we allow trade costs to affect the importance of distant market sizes with a power of \(-\theta\) rather than minus one. Typical estimates of \(\theta\) are substantially greater than one, including our own estimate of \(\theta = 8.22\), but we also report results from assuming a value of \(\theta\) equal to one.

IV.E Main Empirical Specification

Summarizing the above discussion, we begin by regressing the log value of agricultural land in county \(o\) and year \(t\) on log market access \((MA_{ot})\), a county fixed effect \((\delta_o)\), state-by-year fixed effects \((\delta_{st})\), and a cubic polynomial in county latitude and longitude interacted with year effects \((f(x_o, y_o)\delta_t)\):

\[
(13) \ln V_{ot} = \beta \ln(MA_{ot}) + \delta_o + \delta_{st} + f(x_o, y_o)\delta_t + \varepsilon_{ot}.
\]

In practice, and equivalently in the case of two time periods, we estimate equation (13) in differences and often find it convenient to discuss relating changes in log land value to changes in log market access.\(^{58}\) The regressions are weighted by counties’ land value in 1870, both to minimize the influence of outliers and to estimate the appropriate average effect for the counterfactual analysis.\(^{59}\) Standard errors are clustered at the state level to adjust for heteroskedasticity and within-state correlation over time.\(^{60}\)

The regression sample is a balanced panel of 2,327 counties with land value data in 1870 and 1890.\(^{61}\) Figure III shows the sample counties, which are shaded to reflect their change in market access from 1870 to 1890.\(^{62}\) Darker shades correspond to larger increases in market access. There is substantial variation within geographic regions, though Figure III is unable to illustrate the full degree of within-region variation due to the coarseness of the shaded bins. Subsequent robustness checks further exploit this detailed variation within geographic

\(^{58}\)Analogously, the baseline specification (in differences) controls for state fixed effects and flexible polynomials in a county’s latitude and longitude.

\(^{59}\)We use the estimated \(\beta\) to calculate the percent decline in each county’s land value associated with the counterfactual decline in each county’s market access, and multiply this percent decline in land value by each county’s land value in 1890. The aggregate counterfactual loss gives greater weight to counties with greater land value so, if the impact of market access varies across sample counties, it is natural to estimate \(\beta\) weighting by county land value.

\(^{60}\)The estimated standard errors are similar when allowing for spatial correlation among sample counties (Conley, 1999), assuming that spatial correlation declines linearly up to a distance of 700 miles and is zero thereafter. Compared to unweighted standard errors clustered by state, the spatial standard error on the baseline estimate is similar with distance cutoffs of 600 – 800 miles, lower by 10–25% for distance cutoffs between 500 and 200 miles, and higher by 7–9% for distance cutoffs between 900 and 1000 miles.

\(^{61}\)Our measure of county market access includes the cost of trading with each other county that has population data, even if that county is not in the regression sample.

\(^{62}\)Counties are separated into seven equal-sized groups.
areas, through the inclusion of additional control variables.

V  Estimated Impact of Market Access on Agricultural Land Value

V.A Baseline Estimate

Table I reports our baseline result from estimating equation (13). Market access is estimated to have a large and statistically significant impact on land values: a 1% increase in market access increases land values by approximately 0.51% (column 1). The total impact on land value decomposes into roughly equal impacts on the intensive margin (value per farm acre) and the extensive margin (farm acres per county acre).\(^{63}\)

Column 2 reports a similar elasticity for our model-derived measure of market access, discussed above, which reflects a close correlation between log changes in the two measures of market access. We now focus on using the approximated measure of market access, given that its definition is more transparent and simpler to compute under many alternative assumptions, and we consider again the model-derived measure when we present model-derived counterfactual estimates in Section VII.

In our baseline specification, county market access increases due to expansion of the railroad network and growth in other counties’ population. To better understand the main identifying variation in market access, we calculate counties’ market access in 1870 and 1890 when holding all counties’ population fixed at 1870 levels. The estimated impact of market access remains very similar (column 3), as relative changes in market access are primarily determined by the changing transportation network rather than by differential population growth among counties’ trade partners.

In a related exercise, we calculate county \(o\)’s market access based only on those counties \(d\) that are located beyond some distance buffer from county \(o\).\(^{64}\) For a distance buffer of 100 miles, column 4 reports a similar impact of market access on land values. A county’s market access mainly reflects trade with more distant counties, which reduces the potential for bias from local shocks increasing both land values and access to local markets.

Column 5 reports an unweighted estimate, which is of the same magnitude but lower statistical precision than the baseline estimate. Our preferred estimates weight by 1870 land value to reduce the influence of outliers (e.g., counties with very low land value in 1870 that experience large percent increases in land value from 1870 to 1890). Further, weighting by 1870 land value removes the arbitrary distinction between omitting counties with missing

\(^{63}\)In particular, estimating equation (13) produces an estimated impact of 0.245% on the log value of agricultural land per acre of farmland, and an estimated impact of 0.266% on log acres of farmland per county acre.

\(^{64}\)We measure which counties’ borders fall within a distance buffer of each county, and calculate that county’s market access when setting nearby counties’ market size to zero.
(or zero) land value in 1870 and including counties with nearly zero land value in 1870.

V.B Endogeneity of Railroad Construction

Perhaps the main empirical concern is that expansion of the railroad network is endogenous, which may create spurious correlation between increases in county market access and agricultural land value. In particular, railroad construction may occur in counties that would otherwise have experienced relative increases in agricultural land values. Some variation in local railroad construction may be exogenous, perhaps affected by politics, terrain, or incentives to connect particular large cities, but it is difficult to isolate this variation amidst the high-density railroad network in the historical United States.

A useful feature of our definition of market access is that much variation in a county’s market access is not determined solely by that county’s own railroad track or even nearby railroad track. Thus, we can examine changes in counties’ market access that are orthogonal to changes in counties’ own railroads or nearby railroads. We report these estimates in Table II, where column 1 reports the baseline result as a basis for comparison.

Column 2 of Table II reports estimates from a modified version of equation (13), which now controls for whether a county has any railroad track. Column 3 also controls for a flexible function of the county’s mileage of railroad track. Column 4 adds controls for railroad track within a 10-mile buffer of the county, including whether there is any track and the mileage of track. Column 5 also adds the same controls for railroad track within distance buffers of 20 miles, 30 miles, and 40 miles. The estimated impact of market access declines across these specifications, as local railroad construction increases county market access and including these control variables may, in part, exacerbate attenuation bias from measurement error in market access. The estimated impact of market access remains substantial and statistically significant, however, when exploiting variation in market access that is independent of local railroad construction.

In another empirical exercise, we exploit the inherent substitutability between railroads and waterways. Counties close to navigable waterways are naturally less dependent on the railroad network to obtain access to markets. Thus, setting aside how the railroad network actually changed from 1870 to 1890, we might expect market access to increase by less in counties with better access to waterways. To proxy for counties’ access to markets through natural waterways (i.e., county “water market access”), we calculate counties’ market access in 1870 based on county-to-county trade costs in the absence of railroads.

65 In practice, this concern may remain after controlling for changes by state and counties’ longitude and latitude.

66 For this specification, and the following specifications, we control flexibly for railroad track mileage by including a cubic polynomial function of railroad track mileage.
Table III, column 1, reports that counties with greater “water market access” in 1870 experienced less of an increase in market access from 1870 to 1890. Further, counties with greater “water market access” in 1870 experienced a relative decline in agricultural land values (column 2). If we assume that counties with greater “water market access” would have changed similarly to counties with lower “water market access,” aside from their differential changes in market access, then we can instrument for the change in market access with counties’ initial “water market access.” Column 3 reports the implied instrumental variables estimate, which corresponds to the ratio of the coefficient in column 2 to the coefficient in column 1. The implied impact of market access on agricultural land value is larger, but much less precise, than the baseline estimate (column 4).

While we find this instrumental variable approach reasonable *ex ante*, the substantially larger magnitude may well reflect a violation of the IV identification assumptions. For example, counties further from natural waterways may have experienced greater relative increases in agricultural land value for a variety of other reasons. The IV estimate certainly does not reject a substantial impact of market access on agricultural land value, but we emphasize our empirical approaches that focus instead on directly controlling for local shocks to agricultural land value (including additional robustness checks summarized below).

VI Economic Impact of Removing Railroads in 1890

Drawing on the estimated impact of market access, we now turn to estimating the economic impact of removing all railroads in 1890 and the robustness of both estimates to various choices in the empirical analysis.

VI.A Baseline Estimate

Using our transportation network database, we calculate county-to-county lowest-cost freight routes in the absence of any railroads. Given these counterfactual trade costs, and the population of each county, we calculate counties’ counterfactual market access. Appendix Figure 4 maps counties’ change in market access from 1890 to our baseline counterfactual scenario without railroads, with darker shades corresponding to greater declines in market access. We begin by assuming that counties’ populations are held fixed at 1890 levels, but relax this assumption in a later section.

Counties’ market access in 1890 declines by 80%, on average, when all railroads are eliminated. The standard deviation of this decline is 15%, while the 5th and 95th percentiles are 55% and 99.6% declines. Projecting the impact of large counterfactual changes in market access is more credible when two conditions hold: (1) the original regressions are estimated using large changes in market access, and (2) the impact of market access is (log-)linear.

In support of the first condition, the measured changes in market access between 1870
and 1890 have a similar range as that in our counterfactual scenarios. The average percent decline in market access from 1890 to 1870 is 60%, with a standard deviation of 16%, a 5th percentile decline of 43%, and a 95th percentile decline of 94.7%. Log changes in market access from 1870 to 1890 remain large when controlling for state fixed effects and counties’ longitude and latitude: the residual standard deviation is 0.20 logs (weighted) and 0.54 logs (unweighted), whereas the unconditional standard deviation is 0.75 logs (weighted) and 0.89 logs (unweighted).

In support of the second condition, the estimated impact of market access on land value does appear to be (log-)linear. We calculate residual changes in log land value and log market access, after conditioning on the control variables in equation (13). Limiting the sample to residual changes in market access within one standard deviation (plus or minus), Figure IV shows a kernel-weighted local polynomial and its 95% confidence interval.67 There does appear to be a roughly linear functional relationship between changes in log land value and changes in log market access. The theoretical model also predicts that this relationship is log-linear, which gives some additional confidence in predicting counterfactual impacts based on this functional form.

Removing all railroads in 1890 is predicted to decrease the total value of US agricultural land by 60.2% (with a standard error of 4.2%), based on the calculated decline in market access and the estimated impact of market access on agricultural land value.68 In 1890 dollars, this loss corresponds to $4.9 billion. Using Fogel’s preferred mortgage rate of interest, the implied annual economic loss (and standard error) is $386 million ($27 million) or 3.22% (0.22%) of GNP in 1890. The largest possible annual economic impact of railroads is only 5.35% of GNP, which would reflect the complete loss of all agricultural land value in the sample region.69

Decreases in agricultural land value are largest in the Midwest, but are substantial in all regions of the United States. The decline in agricultural land value by region is: $2.5 billion in the Midwest, $0.9 billion in the Plains, and $0.5 billion in the Northeast, South, and Far West. When allowing the impact of market access to vary by region, the impact of railroads declines somewhat in the Northeast and Far West, where we expect congestion and market

---

67 The local polynomial represents the (default) Epanechnikov kernel with (default) bandwidth 0.06.
68 County agricultural land value falls by 0.511 logs for every 1 log decline in market access (Table I, column 1), and the implied percent decline in each county’s land value is multiplied by each county’s land value in 1890. We include all counties from 1890 in these counterfactual estimates, though 455 counties are omitted from the regression sample due to missing data in 1870 (e.g., the counties did not exist in 1870). Losses in these non-sample counties make up 8% of the total counterfactual loss. In the absence of railroads, these non-sample counties experience larger average declines in market access than the regression sample counties (92% vs. 78%), but their average land value is lower ($1.0 million vs. $3.3 million).
69 Fogel also reports state-level mortgage interest rates: using these rates, the implied annual economic loss is 3.01% of GNP and the largest possible loss is 4.92% of GNP.
power to create greater measurement error in counties’ market access.\textsuperscript{70}

VI.B Summary of Robustness Checks

We summarize here various robustness checks, which the online appendix discusses in detail.

For our main measure of counties’ market access, we calculate counties’ access to population in all other counties. We consider several modifications to this use of population, including: inflating the population in counties with major ports to reflect access to international exports and imports; inflating counties’ population by decade and region to adjust for estimates of Census undercounting; and replacing counties’ population with counties’ wealth as an alternative proxy for counties’ market size.

We also consider restrictions on which other locations provide counties with valued trading partners. For example, as agricultural counties might only benefit from trade with urban locations, we define counties’ market access over only urban areas, only major cities, or only New York City. The latter specification, in particular, approximates a model in which counties only value decreased transportation costs to world markets with fixed prices (pinned down in New York City). We also consider defining market access using alternative values of $\theta$ (the “trade elasticity”), which governs how much counties’ market access depends on more-distant locations. Further, we report the results’ robustness to choosing alternative transportation cost parameters, such as lower waterway costs, higher railroad costs, or lower wagon costs.

We also consider modifications to our main empirical specification, such as including additional controls for regional and subregional changes in agricultural land value. Because some frontier regions might experience extreme changes, we also consider excluding outlier values for changes in market access and land value. Much of the increase in counties’ agricultural land value was associated with increased settlement of farmland, so we also consider whether previously unsettled land had some unobserved positive value and how much increased land values might reflect greater fixed investments in improving farmland. Finally, we consider whether impacts on agricultural land values might be driven by impacts of market access on local non-agricultural sectors, restricting the analysis to rural areas with less urban land-use or manufacturing.

Overall, the empirical results are similar across these alternative modeling assumptions. The online appendix reports the robustness of both the estimated impact of market access and the estimated percent decline in land value without railroads, although the latter estimate is the most relevant notion of robustness when changes in the definition of market

\textsuperscript{70}When allowing the impact of market access to vary by region, the estimated impacts of market access (and standard error) are: 0.587 (0.245) in the Plains, 0.519 (0.076) in the Midwest, 0.546 (0.125) in the South, 0.476 (0.153) in the Far West, and 0.306 (0.063) in the Northeast.
access have a mechanical effect on its estimated impact.

VII Counterfactual Impacts on Population and Worker Utility

The previous counterfactual estimates reflect the predicted change in agricultural land value in the absence of railroads, but under the assumption that removal of the railroads would not cause reallocation of population across the United States (or between the United States and the rest of the world). This section explores how the counterfactual estimates change when allowing for this reallocation of population. Further, we consider the associated impacts on worker utility from declines in the agricultural sector.

We begin by using simple proxies for the potential distribution of population within the United States in a counterfactual world without railroads: the distribution of population actually observed in earlier decades (1870, 1850, and 1830). In particular, when calculating each county’s counterfactual market access in the no-railroad scenario, we assign each county a population share that is equal to its share of the national population in the earlier decade (1870, 1850, or 1830), but where total national population is held constant at its 1890 level.\footnote{This procedure uses population shares observed when the railroad network was substantially less-developed (in 1870 and 1850) or non-existent (in 1830). Notably, many counties receive zero population when there were no overlaying counties in these earlier periods.}

Table IV reports the counterfactual impacts on land value under these alternative scenarios, removing railroads and allowing for changes in the distribution of population. Rows 1, 2, and 3 report the results from reallocating population to reflect population shares from 1870, 1850, and 1830, respectively. These three alternative estimates are similar to each other, with counterfactual losses in agricultural land value between 59.1% and 60.1%. These estimates are also similar to our baseline estimate of 60.2%, reported at the top of Table IV, which suggests that the estimated impacts on agricultural land value are insensitive to the domestic reallocation of population shares.

However, the counterfactual distribution of population — after allowing workers to relocate optimally in a no-railroads scenario — may differ in important ways from earlier historical distributions of population. To explore this phenomenon, we now draw further upon the model. The model’s structure is sufficient to simulate the distribution of population in 1890 in the absence of railroads, but holding all else constant.\footnote{In particular, we assume that each county’s productivity ($A_o$) is held constant at its 1890 level and solve for the new equilibrium distribution of population when trade costs change in the no-railroads counterfactual. As described in footnote (57), we develop a procedure that allows us to back out the productivity parameters ($A_o$) for each county in 1890 by using data on the population share in each county in 1890 and trade costs in 1890. This procedure draws on our estimate of $\theta = 8.22$ and the assumed parameter values $\alpha = 0.19$ and $\gamma = 0.60$.}

We begin by assuming that the total United States population level is held constant, which we denote by $\bar{N}$ ($\equiv \sum_o N_o$). We then use the new county populations to calculate each county’s market
access in the no-railroad counterfactual, following equation (12), as before. Table IV, Row 4, reports an estimated counterfactual decline in land value of 56.6%, which suggests that the endogenous reallocation of population in response to the removal of railroads has only a small effect on the loss in land value attributable to the removal of railroads.

The above calculations provide a predicted counterfactual population for each county in the no-railroad scenario, which is itself of interest. Figure V, panel A, maps the substantial counterfactual changes in population, in which darker shades correspond to greater declines in population. In the absence of railroads, there is a clear shift of population from interior regions further from navigable waterways toward interior regions of the country close to navigable waterways. We suspect that the land value results are not sensitive to this counterfactual reallocation of population because the major cities continue to be major cities, and counties’ market access is dependent heavily on their access to these major cities. As a comparison, panel B maps the actual change in population from 1870 to 1890, when there was a more general movement of population toward more Western regions due to railroad network expansion and a myriad of other factors.

Up to this point, our estimates have held fixed the total population \( \bar{N} \) in the United States at its 1890 level even as railroads are removed. As discussed in Section IV.C, however, we have also assumed that workers’ utility level \( \bar{U} \) has been held fixed as well. These assumptions are inconsistent with each other, as \( \bar{U} \) could only remain constant in the counterfactual if workers left the United States (i.e., if \( \bar{N} \) were to fall). Conversely, if workers did not leave the United States, so that \( \bar{N} \) were to remain constant, then \( \bar{U} \) would decline. Drawing on the full general equilibrium structure of the model we now consider these two extreme and opposite scenarios: (1) holding \( \bar{U} \) fixed and allowing \( \bar{N} \) to decline, and (2) holding \( \bar{N} \) fixed and allowing \( \bar{U} \) to decline.

In the first extreme scenario, we continue to hold worker utility \( \bar{U} \) fixed and allow for international migration such that the total population of the United States \( \bar{N} \) falls in the no-railroad scenario. The model parameters imply that land value would fall by 58.4% (row 5).\(^{73}\) The rigid structure of the model, in particular the assumption regarding Cobb-Douglas production, coupled with our choice of numeraire, implies that total population would also fall by the same proportion (58.4%). While this is surely a large adjustment of US population,

\(^{73}\) These magnitudes now reflect our model-derived measure of market access, rather than our empirical approximation of market access. There is a conceptual issue in considering these counterfactual changes in land value from changes in the model-derived measure of market access, which concerns what a “dollar” means in the counterfactual. The absolute level of any county’s market access is, like the price level in any general equilibrium economy, indeterminate without the choice of a numeraire good. The estimates reported here take the population-weighted average of all counties’ good price indices as the numeraire, as might approximate a conventional consumer price index. We obtain similar results when using other plausible choices as the numeraire, such as the price index in New York City.
recall that our model contains only one sector for tractability. In a multi-sector US economy, containing agricultural and manufacturing sectors with tradable goods (and service sectors with non-tradable goods), part of this aggregate population adjustment could occur across sectors within the United States.

In the second extreme scenario, we allow worker utility to decline and restrict international migration such that the total US population is held fixed. The model parameters imply that land value declines by 19%, as reported in row 6, which is substantially smaller than the previous land value estimates because much of the economic incidence has been shifted onto labor rather than land.⁷⁴ Worker utility also declines by 19%, such that the aggregate economic impact becomes an even larger share of the total economy.⁷⁵

These two extreme scenarios imply differing incidence of the economic loss from the removal of railroads. While the exact change in the aggregate variables (\( \bar{U} \) and \( \bar{N} \)) is necessarily hard to pin down empirically, the simple general equilibrium model proposed here suggests that there were large economic impacts from the railroads in the agricultural sector that go beyond the impact on agricultural land value. While the complete loss of the agricultural sector could only generate annual direct losses equal to 5.35% of GNP (i.e., multiplying the value of all agricultural land by Fogel’s preferred mortgage interest rate), the results here highlight that the loss of agricultural production can also impact total population and/or worker utility. In addition, because product and labor markets in the United States interacted with those in other countries via trade and migration, railroads would have impacted prices, wages, and worker utility in the rest of the world.

Against this backdrop, a natural question is which of the two extreme scenarios discussed above might be closer to the truth. To provide guidance on this we now turn to estimating the mobility of population within the United States in this time period in response to changes in counties’ market access. Table V reports estimates analogous to our previous empirical estimates from Table I, but with log population as the outcome variable; this log-linear specification is motivated by our model in equation (11). Our baseline estimate implies that county population increases by 0.26% from a 1% increase in market access. This reflects a substantial population response, but one that is about one third as large as that predicted by our model (in which domestic population mobility is assumed to be perfectly elastic).⁷⁶ Given an intermediate responsiveness of domestic population, and the prospect of intermediate responsiveness of international migration, we suspect that counterfactual impacts might

---

⁷⁴In the previous case, with a perfectly elastic supply of labor (i.e., row 5), land is the only fixed factor and bears the entire cost from removal of railroads.

⁷⁵The Cobb-Douglas production function and numeraire choice imply that worker utility falls by the same percentage as land value.

⁷⁶The coefficient predicted by our model is \( \frac{1+\theta(1+\alpha+\gamma)}{\theta(1+\alpha\theta)} \), equal to 0.75 at our preferred parameter values.
include meaningful declines in both total population and worker utility.

Overall, we draw four main conclusions from this section. First, our estimated counterfactual impacts on agricultural land value are not sensitive to population reallocation within the United States in the absence of railroads. Second, the absence of railroads is likely to have caused substantial declines in total United States population and worker utility for intermediate levels of international labor mobility. Third, agricultural land values would have continued to decline substantially, were railroads to be removed, along with these declines in population and worker utility. Finally, while the loss of agricultural land value can generate only small direct economic losses as a share of the national economy, the accompanying declines in population or worker utility would generate substantial losses in GNP or aggregate welfare.

VIII Counterfactual Transportation Responses to the Absence of Railroads

In evaluating the importance of railroads to the United States economy, and whether railroads were “indispensable” in some sense, it is interesting to consider whether other transportation infrastructure investments might have compensated for the absence of the railroads. Extending our baseline counterfactual analysis, and setting aside impacts on population and worker utility, we report impacts on agricultural land value under alternative counterfactual scenarios. These estimates effectively quantify how much less (or more) market access might have declined in the absence of the railroads due to particular endogenous responses in the transportation network.

Prior to the railroads, many resources were devoted to building a canal network in the Eastern United States; in the absence of the railroads, a system of canals might have been built through portions of the Midwest and Eastern Plains. As described in Section II, Fogel (1964) proposes a feasible system of canals that would have brought 70% of the “infeasible region” within 40 miles of a navigable waterway. In Fogel’s estimates, and in our preliminary extension of Fogel’s analysis in Section II, this system of canals mitigates 30% of the intraregional losses from removing railroads. These estimates require assuming how much land values are affected by distance to a waterway, however, and counties’ distance to a waterway is an imperfect proxy for what counties actually value: access to markets.

To measure the impact of Fogel’s proposed canals, we calculate county-to-county lowest cost transportation routes for a counterfactual network database that replaces all railroads with Fogel’s proposed extension to the canal network. Using these costs to re-calculate counties’ reduction in market access in 1890 without railroads, and multiplying this decline by the estimated impact of market access, we estimate that agricultural land values would still decline by 52.4% with a standard error of 4.2% (Table VI, row 1).
canals are a limited substitute for the railroad network, mitigating only 13% of the losses from removing railroads. While canals reach within 40 miles of many Midwestern areas, the railroad network provides substantially better access to markets. This result is foreshadowed by the remarkably dense railroad network in 1890 seen in Figure II.

Fogel’s proposed canals would have generated annual gains of $50 million in the absence of the railroads, which does exceed their estimated annual capital cost of $34 million. Fogel’s proposed canals were not actually built, presumably because they were made unnecessary by the presence of the railroads. Indeed, using a network database that includes both railroads and the canal extensions, we estimate that the proposed canals generate an annual economic benefit of just $0.20 million.

As an alternative technological solution, in the absence of railroads, there may have been substantial improvements in road-based transportation. Fogel speculates that motor trucks might have been introduced earlier, but a more immediate response could have been the improvement of country roads. For a counterfactual network database that excludes railroads and reduces the cost of wagon transportation to the cost along improved roads (10 cents per mile traveled, 14 cents for a straight route; down from 16.5 and 23.1 respectively in our baseline network database), agricultural land values still decline by 47.5% (Table VI, row 2). Adaptation through improved country roads therefore mitigates only 21% of the loss from removing railroads. We do not find that improving country roads is particularly complementary with extending the canal network: doing both together mitigates 33.6% of the loss from removing railroads (Table VI, row 3), compared to their summed impact of mitigating 34.1%.

This alternative technological solution is predicated on the notion that the absence of railroads would heighten incentives to improve country roads. In a world without railroads, we estimate a $81 million annual benefit from decreased wagon costs; in a world with railroads, we estimate a $50 million annual benefit from decreased wagon costs. Railroads indeed reduce the gains from decreasing wagon transportation costs, but there remain large gains from improving country roads in a world with railroads.

It is difficult to quantify whether the absence of railroads might have encouraged the earlier introduction of motorized trucking, but we can measure how much wagon costs would need to decline to compensate for the absence of railroads. We calculate counterfactual scenarios without railroads, decreasing the wagon cost to 5 cents, 2.5 cents, and 1 cent per ton mile. When replacing the railroads with a lower wagon cost of 5 cents or 2.5 cents,

---

77 For this exercise, we calculate market access for a counterfactual scenario that both includes proposed canals and reduces the cost of wagon transportation.

78 After adjusting for straight routes, the assumed wagon costs are 7.5 cents, 3.5 cents, and 1.4 cents per ton mile.
agricultural land values fall by 30.6% (3.0%) and 15.2% (1.7%), respectively. Agricultural land values increase by 7.5% (1.1%) when replacing railroads with a lower wagon cost of 1 cent, at which point the wagon rate is nearing the railroad rate of 0.63 cents.

Notably, other counterfactual changes might exacerbate the absence of railroads, whereas Fogel focuses on compensatory responses that mitigate the impact on transportation costs from removing railroads. In particular, competition from railroads may have dramatically lowered the costs of shipping by waterway. Holmes and Schmitz (2001) discuss how waterway shipping rates may have roughly doubled in the absence of the railroads, due to increased holdup at transshipment points and adoption of higher-cost shipping technologies. For a counterfactual network database that excludes railroads and doubles the cost of water transportation, we estimate that agricultural land values would decline by 72.5% (Table VI, row 4). This economic loss is 20% greater than the baseline estimated economic impact from removing railroads.

IX Conclusion

This paper develops a new approach for estimating the historical impact of railroads on the American economy. Our analysis uses a new database of county-to-county transport costs to characterize counties’ access to markets. Drawing on recent trade research, changes in counties’ market access summarize the direct and indirect channels through which expansion of the railroad network impacts each county in general equilibrium. We directly estimate the impact of market access on counties’ agricultural land value, which has an intuitive appeal even in the absence of the model: locations benefit from increased access to markets through railroad network expansion, rather than access to railroads per se, and this economic gain to each location is capitalized into the value of agricultural land (i.e., the fixed factor). The empirical analysis exploits county-level variation in market access, controlling for regional changes in agricultural land value and even local changes in the railroad network.

Our estimates imply that railroads were critical to the agricultural sector in 1890: the absence of railroads would have decreased agricultural land values by 60%. Railroads’ contributions to the agricultural sector were largely irreplaceable, either through extensions to the canal network or improvements in country roads. Further, declines in the agricultural sector may have also impacted total population in the United States and workers’ utility.

Whereas Fogel’s estimates depend in large part on assuming the impact on land values from changes in transportation distances, our empirical analysis ultimately lets the data estimate how new railroads improve market access and how market access raises land values. The data indicate that county land values are affected strongly by market access, and that railroads had a critical and irreplaceable role in increasing counties’ market access. In
contrast to Fogel’s analysis, alternative transportation improvements had a limited ability to substitute for the loss of railroads. Our analysis draws on county-level data and a GIS network database that was infeasible in Fogel’s era, along with recent advances in general equilibrium trade theory, but our model maintains the neoclassical framework underlying Fogel’s social saving approach.

Compared to Fogel’s social saving approach, our market access approach finds moderately larger economic impacts from the railroads, and substantially larger impacts on GNP and aggregate welfare once we consider potential impacts on population and worker utility. Fogel’s approach and our approach both focus on railroads’ impacts through the transportation of agricultural goods, but Fogel’s estimates neglect ways in which agricultural land value fails to bound the economic losses from impacts on the agricultural sector.

Our analysis neglects many other potential benefits from the railroads, following Fogel in focusing on gains within the agricultural sector. We view this neglect as opportunities for further research, rather than a presumption that railroads had minimal impacts through other channels. For example, we have neglected impacts on the manufacturing sector, for which railroads may increase access to inputs and consumers. Further, railroads would generate direct gains to workers in the form of decreased passenger rates (e.g., Fishlow, 1965; Boyd and Walton, 1971; Leunig, 2006). Whereas our analysis only measures static gains from specialization and the exploitation of comparative advantage, we suspect larger dynamic gains from increases in technological innovation. We hope that future research might quantify additional channels through which railroads impacted the development of the American economy, perhaps drawing on our measurement of market access to quantify aggregate impacts in addition to relative impacts. We also hope that our data on market access will be useful when examining other changes in the American economy.

As a broader methodological contribution, this paper demonstrates a tractable approach to estimating aggregate treatment effects in the presence of spillover effects. For general settings in which spillover effects are at a national or global scale, some amount of theoretical structure is needed to move beyond estimating relative impacts in more-affected areas. While dealing with this empirical challenge requires some theoretical guidance, the empirical analysis can then proceed in a fairly reduced-form manner. In our case, drawing on a wide class of trade models, the general equilibrium impacts of railroads are captured by measuring the changes induced in counties’ “market access.” Empirical research in all fields of economics is increasingly estimating relative magnitudes by comparing areas that are relatively more or less affected by some plausibly exogenous treatment, but we hope our efforts might encourage similar attempts to exploit relative variation in addressing questions that are more aggregate in nature.
References


Figure I. Distance Buffers in 1890 around Waterways, Railroads, and Proposed Canals

A. 40-Mile Buffers: Waterways (Black) and Railroads (Gray)
B. 40-Mile Buffers: Including Proposed Canals (Dark Gray)

C. 10-Mile Buffers: Waterways (Black) and Railroads (Gray)
D. 10-Mile Buffers: Including Proposed Canals (Dark Gray)

Notes: In Panel A, areas shaded light gray are within 40 miles of a railroad in 1890 but not within 40 miles of a waterway (shaded black). In Panel B, areas shaded dark gray are further than 40 miles from a waterway but within 40 miles of Fogel's proposed canals. Panels C and D are equivalent for 10-mile buffers.
Figure II. Constructed Network Database (Partial)

A. Natural Waterways

B. Natural Waterways and Canals

C. Natural Waterways, Canals, and 1870 Railroads

D. Natural Waterways, Canals, and 1890 Railroads

Notes: Panel A shows all natural waterways, including navigable rivers and routes across lakes and oceans. Panel B adds the canal network (as actually constructed in 1870 and 1890). Panel C adds railroads constructed in 1870, and then Panel D adds railroads constructed between 1870 and 1890.
Figure III. Calculated Changes in Log Market Access from 1870 to 1890, by County

Notes: This map shows the 2,327 sample counties, shaded according to their calculated change in market access from 1870 to 1890. Counties are divided into seven groups (with an equal number of counties per group), and darker shades denote larger increases in market access.
Figure IV. Local Polynomial Relationship Between Changes in Log Land Value and Log Market Access, 1870 to 1890

Notes: Residual changes in sample counties are calculated by regressing changes in the indicated variable on state fixed effects and county longitude and latitude, as in equation (13). This figure then plots the local polynomial relationship between residual changes in log land value and residual changes in log market access, based on an Epanechnikov kernel function with default bandwidth of 0.06. The shaded region reflects the 95% confidence interval.
Figure V. Changes in Log Population, by County

A. Counterfactual Changes in Log Population

Notes: Panel A shows the 2,782 counterfactual sample counties, shaded according to their change in log population from 1890 to the counterfactual scenario. Counties are divided into seven equal-sized groups: darker shades denote larger declines in population, and lighter shades denote larger increases in population. The seven groupings correspond to log changes of: -2.96 to -1.31 (darkest), -1.31 to -0.76, -0.76 to -0.34, -0.34 to -0.02, -0.02 to 0.19, 0.19 to 0.34, and 0.34 to 0.71 (lightest).

B. Changes in Log Population from 1870 to 1890

Notes: Panel B shows the 2,327 sample counties, shaded according to their change in log population from 1870 to 1890. Counties are divided into seven equal-sized groups: darker shades denote larger increases in population, and lighter shades smaller increases. The seven groupings correspond to log changes of: 8.19 to 1.17 (darkest), 1.17 to 0.74, 0.74 to 0.50, 0.50 to 0.35, 0.35 to 0.23, 0.23 to 0.10, and 0.10 to -2.17 (lightest).
Table I. Estimated Impact of Market Access on Agricultural Land Value

<table>
<thead>
<tr>
<th>Log Value of Agricultural Land</th>
<th>Baseline Specification</th>
<th>Model-Derived Market Access</th>
<th>Fixed 1870 Population</th>
<th>100-mile Buffer Market Access</th>
<th>Unweighted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Market Access</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td></td>
<td>0.511</td>
<td>0.587</td>
<td>0.510</td>
<td>0.487</td>
<td>0.506</td>
</tr>
<tr>
<td></td>
<td>(0.065)</td>
<td>(0.073)</td>
<td>(0.065)</td>
<td>(0.064)</td>
<td>(0.124)</td>
</tr>
<tr>
<td>Number of Counties</td>
<td>2,327</td>
<td>2,327</td>
<td>2,327</td>
<td>2,327</td>
<td>2,327</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.625</td>
<td>0.627</td>
<td>0.625</td>
<td>0.621</td>
<td>0.606</td>
</tr>
</tbody>
</table>

Notes: Column 1 reports estimates from equation (13) in the text: for a balanced panel of 2,327 counties in 1870 and 1890, the Log Value of Agricultural Land is regressed on Log Market Access (as defined in equation 12), county fixed effects, state-by-year fixed effects, and year-specific cubic polynomials in county latitude and longitude. The regression is weighted by counties' 1870 value of agricultural land. Columns 2 through 5 report robustness checks, as discussed in the text: column 2 uses a model-derived measure of market access (equation 9 in the text); column 3 uses a measure of market access for 1890 that holds counties' population levels fixed at 1870 levels; column 4 uses a measure of market access only to counties beyond 100 miles of a county; and column 5 reports estimates from the baseline specification when not weighting by counties' 1870 land value. Robust standard errors clustered by state are reported in parentheses.
<table>
<thead>
<tr>
<th>Controls for:</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any Railroad</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Railroad Length</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Railroads within nearby buffer</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Railroads within further buffers</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Number of Counties</td>
<td>2,327</td>
<td>2,327</td>
<td>2,327</td>
<td>2,327</td>
<td>2,327</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.625</td>
<td>0.627</td>
<td>0.632</td>
<td>0.640</td>
<td>0.653</td>
</tr>
</tbody>
</table>

Notes: Column 1 reports the estimated impact of market access from the baseline specification (Table 1, column 1). Column 2 includes an additional control for whether a county contains any railroad track. Column 3 also controls for a cubic polynomial function of the railroad track mileage in a county. Column 4 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 5 includes additional controls for any railroad track and mileage of railroad track within 20 miles, 30 miles, and 40 miles of the county (as in Column 4). All regressions include county fixed effects, state-by-year fixed effects, and year-specific cubic polynomials in county latitude and longitude. All regressions are weighted by counties' 1870 value of agricultural land. Robust standard errors clustered by state are reported in parentheses.
Table III. Impact of Market Access: Instrumenting with Waterways

<table>
<thead>
<tr>
<th></th>
<th>Change in Log Market Access (1870 to 1890)</th>
<th>Change in Log Value of Agricultural Land (1870 to 1890)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS (1)</td>
<td>OLS (2)                                           2SLS (3)</td>
</tr>
<tr>
<td>Change in Log Market Access</td>
<td>1.14</td>
<td>0.511</td>
</tr>
<tr>
<td></td>
<td>(0.290)</td>
<td>(0.065)</td>
</tr>
<tr>
<td>Log Water Market Access in 1870</td>
<td>-0.096</td>
<td>-0.109</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>Number of Counties</td>
<td>2,327</td>
<td>2,327</td>
</tr>
<tr>
<td></td>
<td>2,327</td>
<td>2,327</td>
</tr>
<tr>
<td></td>
<td>2,327</td>
<td>2,327</td>
</tr>
<tr>
<td></td>
<td>2,327</td>
<td>2,327</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.568</td>
<td>0.602</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Columns 1 and 2 report the impact of Log Water Market Access in 1870 on changes in the indicated outcome variable between 1870 and 1890. Column 3 reports the estimated impact of a change in Log Market Access on the change in Log Value of Agricultural Land, instrumenting for the change in Log Market Access with Log Water Market Access in 1870. Column 4 reports the baseline estimate for comparison (from column 1 of Table I). All regressions include state fixed effects and cubic polynomials in county latitude and longitude, and are weighted by counties' 1870 value of agricultural land. Robust standard errors clustered by state are reported in parentheses.
### Table IV. Counterfactual Impacts on Land Value, Allowing for Population Reallocation

<table>
<thead>
<tr>
<th>Changes in the Distribution of Population (Holding Total Population Constant)</th>
<th>Percent Decline in Land Value Without Railroads</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Counterfactual Without Railroads in 1890</td>
<td>60.2 (4.2)</td>
</tr>
<tr>
<td>1. Assuming the population distribution from 1870</td>
<td>59.1 (4.1)</td>
</tr>
<tr>
<td>2. Assuming the population distribution from 1850</td>
<td>59.3 (4.1)</td>
</tr>
<tr>
<td>3. Assuming the population distribution from 1830</td>
<td>60.1 (4.0)</td>
</tr>
<tr>
<td>4. Assigning the model-predicted counterfactual distribution of population</td>
<td>56.6 (4.0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Changes in the Distribution and Total Level of Population (Holding Worker Utility Constant)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5. Model-predicted estimate, allowing for changes in the level and distribution of population</td>
<td>58.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Changes in the Distribution of Population and Worker Utility (Holding Total Population Constant)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6. Model-predicted estimate, allowing for changes in worker utility and the distribution of population</td>
<td>19.0</td>
</tr>
</tbody>
</table>

**Notes:** Each row reports the counterfactual impact on land value from the removal of railroads, given some response in county populations (as described in Section VII). In row 1, the railroad network is removed and county population shares are shifted to their population share in 1870 (holding fixed the total population in the country). Similarly, in rows 2 and 3, the railroad network is removed and county population shares are shifted to their population shares in 1850 and 1830, respectively. In row 4, the railroad network is removed and county population shares are shifted to those predicted by the model for the counterfactual scenario. In rows 1 - 4, the counterfactual calculations follow the same procedure as in the baseline counterfactual analysis (reported at the top of the Table for comparison). In row 5, we report the impact on land value implied by the model parameters for counterfactual transportation costs without railroads, holding worker utility constant and allowing total population to decline along with the reallocation of population within the country. In row 6, we report the impact on land value implied by the model parameters for counterfactual transportation costs without railroads, holding total population constant and allowing worker utility to decline along with the reallocation of population within the country. Robust standard errors clustered by state are reported in parentheses, when available.
Table V. Estimated Impact of Market Access on Population

<table>
<thead>
<tr>
<th></th>
<th>Baseline Specification</th>
<th>Calibrated Market Access</th>
<th>Fixed 1870 Population</th>
<th>100-mile Buffer Market Access</th>
<th>Unweighted</th>
<th>Weighted by 1870 Population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td>Log Market Access</td>
<td>0.259</td>
<td>0.314</td>
<td>0.262</td>
<td>0.243</td>
<td>0.348</td>
<td>0.197</td>
</tr>
<tr>
<td></td>
<td>(0.049)</td>
<td>(0.056)</td>
<td>(0.050)</td>
<td>(0.048)</td>
<td>(0.086)</td>
<td>(0.039)</td>
</tr>
<tr>
<td>Number of Counties</td>
<td>2,327</td>
<td>2,327</td>
<td>2,327</td>
<td>2,327</td>
<td>2,327</td>
<td>2,327</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.311</td>
<td>0.316</td>
<td>0.311</td>
<td>0.307</td>
<td>0.523</td>
<td>0.252</td>
</tr>
</tbody>
</table>

Notes: Columns 1 through 5 report estimates from empirical specifications analogous to those reported in columns 1 through 5 of Table I, but for the outcome variable Log Population. In column 6, the regression from column 1 is weighted by counties' population in 1870 (rather than counties' land value in 1870). Robust standard errors clustered by state are reported in parentheses.
<table>
<thead>
<tr>
<th>Table VI. Counterfactual Impacts on Land Value, Allowing for Transportation Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent Decline in Land Value</td>
</tr>
<tr>
<td>(1)</td>
</tr>
<tr>
<td>Baseline Counterfactual Without Railroads in 1890</td>
</tr>
<tr>
<td>Allowing for Transportation Responses</td>
</tr>
<tr>
<td>1. Extended canal network</td>
</tr>
<tr>
<td>2. Improved country roads, wagon cost of 14 cents</td>
</tr>
<tr>
<td>3. Extended canal network and improved country roads</td>
</tr>
<tr>
<td>4. Increased water shipping rates, doubled</td>
</tr>
</tbody>
</table>

Notes: Each row reports the counterfactual impact on land value from the removal of railroads, given some potential response in the transportation network (as described in Section VIII). In row 1, the railroad network is removed and the canal network is extended. In row 2, the railroad network is removed and the wagon freight rate is lowered to reflect improvements in country roads. In row 3, the railroad network is removed and both adjustments are made from rows 1 and 2. In row 4, the railroad network is removed and waterway freight rates are doubled. Robust standard errors clustered by state are reported in parentheses.
I Data Appendix

I.A Data Sources

County-level data are from the US Censuses of Agriculture and Population (Haines, 2005).

Land value is defined as the total value of land in farms, including the value of farm buildings and improvements. We deflate these reported data, however, using Fogel’s state-level estimates of the value of agricultural land only (Fogel, 1964, pp. 82-83). We multiply counties’ reported Census data by Fogel’s estimated value of agricultural land only (in their state) divided by the reported Census value of agricultural land (in their state). We analyze the total value of agricultural land in the county, including land in farms (valued above) and land not in farms (valued at zero), though in robustness checks we relax the assumption that non-farmland has zero agricultural value. Farmland is defined to be the total number of acres of land in farms, including all improved land and all unimproved land.

Population is defined as the reported total population in each county. For one robustness check, we inflate these population data due to potential undercounting in the Census that varies by region and year: 8.8% undercounting in the South in 1870, 6.0% undercounting in the North in 1870, 4.95% undercounting in the South in 1890, and 4.05% undercounting in the North in 1890 (Hacker, 2013, pp. 92-93). Regional estimates of undercounting were unavailable for 1890, so we averaged the undercounting rates by region for 1880 and 1900. We sometimes use data on city population, which is defined as the county’s population living in cities of over 25,000 people. We also sometimes use data on urban population, which is defined as the county’s population living in urban areas of over 2,500 people. We also sometimes use data on county wealth in 1870, which is defined as the true value (rather than assessed value) of real estate and personal estate in the county.

We adjust county-level data to maintain consistent county definitions in 1870 and 1890. We adjust data from 1870 to reflect 1890 county boundaries (Hornbeck, 2010). Using historical US county boundary files (Minnesota Population Center, 2011), county borders in 1890 are intersected with county borders in 1870. When counties in 1890 fall within more than one 1870 county, data for each piece are calculated by multiplying the later county data by the share of its area in the 1870 county. For later periods, each 1870 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 1890 there is little overlap with a second 1870 county. In three instances we combine separate cities into a neighboring county: Baltimore City is combined into Baltimore County for 1890 (where it is contained in 1870); St. Louis City is combined into St. Louis County for 1890 (where it is contained in 1870); and Washington DC is combined into Montgomery County for both periods.
I.B  Network Database

We have created a GIS network database for the purpose of calculating county-to-county lowest-cost routes. We describe this network in the main text, and the tradeoffs involved in its construction. Appendix Table 1 lists each element of the network, a brief description, a brief description of its construction, and its assigned cost for our baseline estimates. The text discusses a number of robustness checks that adjust these assigned costs.

I.C  Sample Definition

For the regression analysis, the sample includes all 2,327 counties that report land value data for both 1870 and 1890. These 2,327 counties are defined by their boundaries in 1890 (see above), and Appendix Figure 1 maps this sample of counties included in the regression analysis. In some robustness checks, we modify this sample: excluding counties with any city population, excluding counties with any urban population, and excluding counties with outlier changes in market access or land value.

For the counterfactual analysis, note that we expand the sample to include all 2,782 counties that report land value data in 1890. That is, we use the regression estimate for how market access influences county land value and then calculate the implied decline in agricultural land value for all counties that existed in 1890 according to their counterfactual decline in market access. Appendix Figure 2 maps all counties included in the counterfactual analysis. Even for robustness checks, when the regression sample is further restricted, the counterfactual sample includes all counties for comparability.

I.D  Summary Statistics and Geographic Variation

Appendix Table 2 provides summary statistics for the sample of counties in the regression analysis. Appendix Figure 3 maps the geographic distribution of counties’ change in Log Land Value from 1870 to 1890, grouping the regression sample counties into seven equal-sized bins (with larger increases shaded darker). For the sample of counties included in the counterfactual analysis, Appendix Figure 4 maps counties’ loss in market access in going from the existing 1890 network to our baseline counterfactual scenario (with greater losses shaded darker).
II Results Appendix

II.A Robustness to Changes in the Definition of Market Access

Appendix Table 3 reports the baseline results’ robustness to changes in the definition of market access. Column 1 reports the estimated impact of market access, and column 2 reports the estimated percent decline in national agricultural land value without the railroads.\(^1\)

The results are similar when we adjust for the potential influence of international exports and imports (row 1). While we only directly measure US counties’ access to other US counties, we can effectively increase market sizes in counties with major international ports. In particular, we assign additional population to 11 counties, based on the value of merchandise traded and nominal GDP per capita. These 11 ports cover 90% to 93% of international trade in 1870 and 1890, yet this adjustment has little impact on the empirical results.\(^2\)

We also consider two further adjustments to the population data, for the purpose of defining counties’ market access in 1870 and 1890. The Census of Population is known to undercount population, with particularly large undercounts in 1870 and in the South. We inflate counties’ population by decade and region, based on an estimated degree of undercounting (Hacker, 2013), though this has little impact on the empirical estimates (row 2). As an alternative to using population data to proxy for county market size, we also consider using the Census-reported total value of real estate and personal property. While these data are only available in 1870, we calculate market access in 1870 and 1890 using these fixed market sizes (as in the main paper’s Table I, column 3). The empirical results remain similar (row 3), as county population is closely correlated with county value of real estate and personal property.\(^3\)

One technical issue is that a county’s market access should theoretically include access to its own population, though our baseline measure omits this term due to simultaneity concerns. A county’s own population forms a small share of its total market access, however, so including its own market has little impact on the results (row 4). In a similar trade-off between simultaneity concerns and the model’s suggested measure, we omit counties’ access to nearby counties and the results are similar (rows 5-7).

---

\(^1\)The standard errors for the estimates in columns 2 are calculated using the delta method, which transforms the standard errors in column 1.

\(^2\)For 11 major international ports, we assign additional county population in 1870 and 1890 based on the ports’ average value of exports and imports divided by nominal GDP per capita in 1870 and 1890 (Bureau of Statistics, 2003; Carter et al., 2006). This adjustment mainly increases the “effective population” in New York City, New Orleans, Boston, Baltimore, Philadelphia, and San Francisco. There are also large percent increases in Galveston, Savannah, and Charleston, and smaller percent increases in Norfolk and Portland.

\(^3\)For column 2, because data are not available in 1890 for the total value of real estate and personal property, we use the estimated coefficient from column 1 and the counterfactual decline in market access from our baseline specification.
Our baseline measure of market access includes counties’ access to all other counties, based on a model of trade among all US counties. Alternatively, agricultural counties may only benefit from selling/buying goods to/from cities. The estimated impact of market access is similar when only measuring counties’ access to urban areas, cities, or only New York City (rows 8-10), which reflects the large influence of urban areas and cities in determining counties’ overall market access. The implied economic impact of railroads is somewhat smaller in these specifications, which could reflect railroads’ comparative advantage in linking rural areas with each other (relative to the role of waterways).

Calculating our measure of market access requires assuming a value of $\theta$ (the “trade elasticity”), for which our baseline measure assumes a value of 8.22. We obtained that value of $\theta$ by estimating which value of $\theta$ best fit the model to the data, but now we consider the results’ robustness to alternative values of theta. Different values of $\theta$ will change the estimated impact of market access (column 1), though this largely reflects a mechanical re-scaling of “market access” and the relevant notion of robustness concerns the counterfactual impact on land value (column 2). In row 11, we assume a value of 1 for $\theta$, which makes our measure of market access more similar to an older notion of “market potential” (Harris, 1954). In rows 12 and 17, we assume values for $\theta$ of 3.37 and 13.18, which correspond to the 95% confidence interval around our estimated value of 8.22. In rows 13 and 16, we assume values for $\theta$ of 3.60 and 12.86, which reflect the two extreme estimates in Eaton and Kortum (2002). In row 15, we assume a value for $\theta$ of 6.74, which reflects the mean estimate from a meta-survey by Head and Mayer (2014). In row 14, we assume a value for $\theta$ of 3.80, which corresponds to the assumed value in our previous working paper. Overall, these alternative choices for $\theta$ mechanically change the estimated impact of market access (column 1), but have little effect on the implied impact of railroads (column 2) because there are also changes in the counterfactual impact on market access.

---

4 The Census defines urban areas as places with population greater than 2,500, and defines cities as areas with population greater than 25,000.

5 Note that the “distance buffer” (rows 5-7) and “city access” (rows 8-10) estimates would suffer from omitted variables bias if counties value access to non-included areas, so we emphasize these results mainly as a sensitivity check on our baseline estimates. These estimates do not strictly decompose the benefits of market access to cities, to urban areas, and to rural areas.

6 Note that this is a non-linear re-scaling of market access; otherwise, there would be no impact on estimating the impact of market access in logs.

7 Eaton and Kortum’s (2002) preferred estimate of $\theta$ is 8.28, which is remarkably similar to our estimate of 8.22. Similarly, Caliendo and Parro (forthcoming) estimate an average $\theta$ (across 20 industries) of 8.64 (with $\theta = 8.11$ for agriculture).

8 Similarly, Costinot, Donaldson and Komunjer (2012) estimate $\theta = 6.53$.

9 Donaldson (2015) estimates an average $\theta$ of 3.80 (across 13 agricultural categories), and Simonovska and Waugh (2013) estimate $\theta = 4.10$. 
II.B Robustness to Transportation Cost Parameters

Appendix Table 4 reports the robustness of the empirical results to alternative parameter choices for freight transportation costs in the network database.

Water transportation is always low cost, but the estimated impact of market access is not sensitive to further lowering the cost of sea routes or all water routes (rows 1 and 2).\(^{10}\) The aggregate impact of railroads declines somewhat with lower water costs, as a counterfactual without railroads becomes more manageable. Our empirical estimates are less sensitive than social saving estimates, however, which change proportionally with the difference in point-to-point costs by rail and water.

One advantage to railroads was in reducing transshipment charges, incurred whenever transferring goods to/from a railroad car, river boat, canal barge, or ocean liner. The estimates are not sensitive, however, to eliminating transshipment charges within the waterway network (row 3). In addition, transportation through the railroad network was not entirely seamless: congestion, fragmented track ownership, or differences in gauges may have required periodic transshipment; and scheduled freight routes would be less direct than those calculated on the GIS network. We consider a higher railroad rate (0.735 cents) that reflects a need for two transshipment points within an average length railroad route, and an even higher railroad rate (0.878 cents) that makes railroad routes as indirect as wagon routes.\(^{11}\) The estimated impact of market access is not sensitive to higher railroad rates (rows 4 and 5), whereas the aggregate impact of railroads declines slightly with a decrease in the relative advantage of railroads.

Wagon transportation costs are likely to be an important feature of the database. Fogel emphasized that the baseline wagon rate may be too high and his social saving estimates are substantially lower for decreased wagon costs of 14 cents per mile.\(^{12}\) By contrast, this lower wagon rate increases the estimated impact of market access such that the aggregate impact of railroads is moderately higher (row 6).

In some cases, wagon costs may not reflect the true cost of transporting goods to market. For example, cattle from Western ranches could be driven to market at much lower costs. While this is a one-way trade, it highlights adaptations in areas where long-distance wagon transportation was very expensive. The empirical results are robust to excluding the Western

\(^{10}\) The lower waterway rate (0.198 cents per ton-mile) reflects Fogel’s preliminary rate for waterway transportation, prior to his adjustments for supplemental costs associated with waterway transportation.

\(^{11}\) In the first case, we consider an average length railroad route of 926 miles (Fogel, 1964) and assign an additional dollar over this distance (which becomes 0.108 cents per mile). In the second case, we increase the baseline railroad rate by the same factor (1.4) used to adjust for indirect wagon routes (Fogel, 1964).

\(^{12}\) To clarify, the cost of 14 cents per mile reflects Fogel’s considered lower cost (10 cents per mile) and Fogel’s adjustment factor for indirect wagon routes (1.4). It is more convenient for us to scale up the price of routes, rather than equivalently scale up the distance required.
region from the sample (row 7). The Western regions are not central to the empirical analysis, which draws on within-state variation in changes in market access.

The estimated impact of market access is affected little by changes to the waterway route between the Pacific Ocean and the Atlantic Ocean (rows 8-10). This waterway route has little impact on the calculation of market access because the transcontinental railroad was available in 1870 and 1890. By contrast, this waterway route has more influence on the estimated aggregate impact of railroads because it provides the only link between Western and Eastern markets in the absence of the railroads (aside from wagons). The costs of waterway shipping may have been driven down by competition with railway shipping, however, which we explore in later analysis.

II.C Robustness to Alternative Empirical Specifications

Appendix Table 5 reports the baseline results’ robustness to alternative empirical specifications. These alternative specifications may change the estimated impact of market access, but do not also change the counterfactual decline in market access without railroads. Thus, in these cases, changes in the estimated coefficient (column 1) directly map into changes in the implied impact of railroads (column 2), though by less in percent terms due to the log functional form.

The baseline specification includes controls for changes by state and latitude/longitude, though we also consider additional controls for region-specific changes. Rows 1 and 2 also include year-specific controls for the fraction of each county in 20 “resource regions” or 145 “resource subregions” (Hornbeck, 2010). Comparing counties within such local regions may exacerbate measurement error in market access, arising from simplifications in our network database and calculated county-to-county freight transportation costs, though the estimated magnitudes remain substantial and statistically significant. Rows 3 and 4 modify the baseline specification’s controls for county longitude and latitude, replacing the year-specific third-order polynomials with fifth-order and first-order polynomials.

Due to concerns that outlier values might unduly influence the empirical estimates, we consider excluding outlier values for changes in market access and changes in land value.

---

13 For this specification, we exclude counties in Texas, Oklahoma, Kansas, Nebraska, South Dakota, North Dakota, and all states further West.

14 The network database includes a waterway route from the Pacific (near San Diego) to the Atlantic (near Florida), which reflects the potential to transport goods around Cape Horn (or through Panama by railroad). The transcontinental railroad was generally preferred after its construction, but waterway routes were used for some goods and the baseline estimates assume a Pacific-to-Atlantic waterway cost of $8 such that the overall GIS-calculated cost of transporting goods from San Francisco to New York City was similar by rail and water. Robustness checks assign this waterway cost such that the GIS-calculated costs by rail and water were similar between Seattle and New York City (a $5 connection) or between San Diego and New York City (a $11 connection).

15 Note that these regions were mapped in 1966, so they may be endogenous to the availability of railroads.
When excluding the largest and smallest 1% or 5% of changes in market access (rows 5 and 6), the estimated impact of market access is similar or slightly larger. These coefficients imply a similar or slightly larger impact of railroads, as we include all counties when calculating the implied decline in agricultural land value. When excluding the largest and smallest 1% or 5% of changes in land value (rows 7 and 8), the estimated impact of market access (and railroads) is similar or slightly smaller. Increasingly excluding outlier values in the outcome variable will mechanically attenuate estimates, but it is useful to verify that the estimates are not driven by large percent changes in a small number of sample counties. Indeed, part of the rationale for weighting the regressions is to reduce the influence of sparsely settled counties experiencing large increases in land value.

Some counties experience large increases in agricultural land value as new lands are settled between 1870 and 1890. In the theory and the empirical analysis, we consider the total supply of agricultural land fixed by counties’ geographic borders and analyze increases in county land value on both the intensive and extensive margins. Our baseline estimates assume that newly settled farmland had zero agricultural value previously, as newly settled land came predominately from the public domain where land had not been desired and so was revealed to have had little economic value. In two robustness checks, we relax this assumption and adjust upward counties’ value of land in 1870 by assuming new lands (settled between 1870 and 1890) had some value in 1870: either a small value ($0.0625 per acre) that reflects the cost of acquiring land through the homestead act (row 9), or a larger value ($1.25 per acre) that reflects the cost of acquiring land through preemption (row 10). The estimates are robust to these adjustments, though the estimated magnitude declines meaningfully when assuming unsettled land would have been worth the cost of acquiring land through preemption. We see that estimate as overly conservative, as the absence of settlement in 1870 indicates these lands were not worth the cost of settlement.\footnote{In addition, we only add to counties’ 1870 value by truncating the quantity of “newly settled land” at zero, measured by the change in total farmland between 1870 and 1890.}

In a related exercise, we consider whether the increases in agricultural land value are driven, in part, by increases in the intensity of land improvement. In the baseline analysis, we follow Fogel in deflating the Census-reported value of agricultural land, buildings, and improvements to obtain an estimate of the value of agricultural land only.\footnote{Deflating land values in this manner does not change the estimated impact of market access or the counterfactual decline in agricultural land value, as the state-level deflator is absorbed by state-by-year fixed effects. Deflating does affect the implied annual economic loss, however, which depends on the level of agricultural land value.} Fogel’s deflator for 1890 may over-deflate land values in 1870, however, for counties with lower land improvement in 1870. We consider two robustness checks that decrease this deflator in 1870.
for counties that had less land improvement in 1870, which effectively dampens the increase in land value from 1870 to 1890 for counties that experience increases in land improvement. First, in row 11, we do not deflate land values for counties that had no land improvement in 1870, and partially deflate when land improvement was lower in 1870 than in 1890. This is an overly aggressive correction because it assumes there is no value from buildings or other improvements when there is no improved acreage. Second, in row 12, we modify Fogel’s state-level deflators in 1890 to derive the comparable county-level deflator in 1870 that depends on counties’ acreage of improved land in 1870 and 1890, counties’ reported value of land/buildings/improvements in 1870 and 1890, and state-level estimates of the cost of clearing land in 1870 and 1890 (Primack, 1977). The regressions are robust to both of these approaches, and so accounting for land improvement costs does not appear to be quantitatively important for the results.

Finally, we consider whether impacts on agricultural land values reflect impacts of market access on local non-agricultural sectors. In Section IV.D and the Theory Appendix, we discuss how the inclusion of non-agricultural sectors does not meaningfully change our analysis of how market access impacts agricultural land values. We do assume, however, that the supply of land to the agricultural sector is fixed by counties’ geographic size. In practice, impacts of market access on non-agricultural sectors might impact agricultural land. We can restrict the analysis to rural areas, however, as urban and manufacturing activity is highly spatially concentrated. In rows 13 and 14, we report the estimates’ robustness to dropping from the regression sample those counties with any city population (i.e., population in areas above 25,000 people) and counties with any urban population (i.e., population in areas above 2,500 people). The estimated magnitudes decline moderately, but remain similar.

---

18 For 1870, we apply a deflator to county-level data that reflects the weighted average of Fogel’s deflator and no deflator, where the weight on Fogel’s deflator is the ratio of improved acreage in 1870 to improved acreage in 1890 (bounded above by one).

19 The derivation assumes that the value of buildings (and other unobserved fixed investments) is proportional to the reported total value of land/buildings/improvements, which is a necessary assumption in the absence of early data on their separate values (and Fogel also made this necessary assumption). The implied value of improvements is sometimes greater than the reported value of land, buildings, and improvements, and so we restrict the regression to counties for which the imputed value of agricultural land is positive in 1870.

20 We continue to weight the regressions by counties’ unadjusted land value in 1870, for comparability, but the estimates are similar when the regressions are also weighted by these adjusted 1870 land values.

21 For example, total land value could decrease as land is moved into housing or non-agricultural production; or, alternatively, agricultural land value could increase in anticipation of being sold for non-agricultural purposes.

22 Note that the calculation of rural areas’ market access still includes access to cities and other urban areas.
III Theory Appendix

III.A An Extended Model with Multiple Sectors

We now describe an extended version of our model that features two sectors — an agricultural sector (as in Section IV of the main paper) and a “manufacturing” sector — with multiple sources of interactions across them. Our goal is to demonstrate how the logic of our main result (equation 10 in the main paper), which links agricultural land values to market access, can be interpreted in this enriched economic environment.

We continue to index regions with the index $o$, and occasionally also by $d$ when a region is the destination of a trade flow. We now imagine that each county in the US is comprised of both a rural and an urban region, though in practice one or the other could be unpopulated and non-existent. To distinguish these two types of regions we denote rural regions by $o \in R$ and urban regions by $o \in U$.

**Tastes:** As in Section IV of the main paper, agricultural goods are available in a continuum of varieties. Manufactured goods are also available in a continuum of varieties. The representative consumer has CES preferences over these sets of varieties with elasticities of substitution $\sigma_A$ and $\sigma_M$, respectively. That is, an agent living in region $o$ and receiving income $Y_o$ has indirect utility:

$$V(P_o, Y_o) = \frac{Y_o}{(P^A_o)^{\mu}(P^M_o)^{1-\mu}},$$

with $P^M_o$ being a price index that is given by

$$\left(P^M_o\right)^{1-\sigma_M} = \int_0^{n_M} (p^M_o(j))^{1-\sigma_M} dj,$$

and $n_M$ being the exogenous number of manufactured varieties. $P^A_o$ follows the definition in Section IV of the main paper.

**Technology for Producing Manufactured Goods:** Manufactured goods are assumed to be made only by urban regions, reflecting the high degree of spatial concentration in manufacturing. Each urban region has access to a constant-returns Cobb-Douglas technology for turning land, labor, capital, and raw agricultural goods into manufacturing goods. The marginal cost of producing variety $j$ of manufactured goods in region $o$ is given by:

$$MC^M_o(j) = g^{\alpha_M}w^\gamma_Mr^\beta_M\Phi^{\beta_M}$$

$$\frac{1-\alpha_M-\beta_M-\gamma_M}{z^\beta_M(j)}$$

if $o \in U$. 

9
where \( q_o \) is the land rental rate, \( w_o \) is the wage rate, and \( r_o \) is the capital rental rate in that region. The term \( \Phi^A_o \) is a price index over the set of agricultural varieties that are used in the production of manufacturing goods; we assume that manufacturing production requires a bundle of agricultural varieties, with constant elasticity of transformation \( \psi_A \) across these varieties.\(^{23}\) Similar to the agricultural sector (from Section IV of the main paper), \( z^M_o(j) \) is a Hicks-neutral productivity shifter that is exogenous and local to region \( o \) and drawn from a Fréchet distribution with CDF given by: \( F^M_o(z) = 1 - \exp(-B_o z^{-\theta_M}) \), with \( \theta_M > 1 \). The manufacturing sector is assumed to be perfectly competitive.

**Technology for Producing Agricultural Goods:** Agricultural goods are assumed to be made only by rural regions, using the same production technology as in Section IV of the main paper but augmented to include the use of intermediate inputs produced in the manufacturing sector. Generalizing the notation slightly, this technology is given by:

\[
MC^A_o(j) = \frac{q^A_o w^A_o r^A_o (\Phi^A_o)^{\beta_A}}{z^A_o(j)} \quad \text{if} \quad o \in \mathcal{R},
\]

where \( z^A_o(j) \) is drawn from the CDF \( F^A_o(z) = 1 - \exp(-A_o z^{-\theta_A}) \), with \( \theta_A > 1 \). Analogously to the case of the manufacturing sector, the term \( \Phi^M_o \) is a CES price index for the set of manufacturing varieties used in agricultural production.\(^{24}\) As before, the agricultural sector is assumed to be perfectly competitive.

**Trade Costs:** As in Section IV of the main paper, trading goods is costly. We denote the iceberg cost of shipping a good from region \( o \) to region \( d \) in the manufacturing sector as \( \tau^M_{od} \), and that in the agricultural sector as \( \tau^A_{od} \).

**Factor Supply:** As in Section IV of the main paper, we continue to assume that both capital and labor are freely mobile within the United States and internationally. This implies that the capital rental rate is pinned down to the international level. It also implies that worker utility, from equation (1), is pinned down to the level of utility that US workers could earn elsewhere, denoted \( \bar{U} \). Therefore, equilibrium nominal wages must satisfy:

\[
w_o = \bar{U} (P^A_o)^\mu (P^M_o)^{1-\mu}.
\]

Finally, we assume that land is immobile and in fixed supply \( L_o \) in each region.

\(^{23}\)This implies that \( (\Phi^A_o)^{1-\psi_A} = \int_0^n A_o (p^A_o(j))^{1-\psi_A} dj \).

\(^{24}\)We denote by \( \psi_M \) the elasticity of substitution, such that \( (\Phi^M_o)^{1-\psi_M} = \int_0^n M_o (p^M_o(j))^{1-\psi_M} dj \).
Solving for Prices and Trade Flows: We follow similar derivations to those in Section IV of the main paper, this time separately for the manufacturing and agricultural sectors. We begin with the manufacturing sector. The consumer price index for manufactured goods in any region \( d \) is given by:

\[
(P_d^M)^{-\theta_M} = \chi_1 \sum_{o \in U} B_o(q_o^{\alpha_M} w_o^{\gamma_M})^{-\theta_M} (IMA_o^A)^{\beta_M} (\tau_{od}^M)^{-\theta_M} \equiv CMA_o^M \quad \text{if } d \in U, R,
\]

where we denote by \( CMA_o^M \) the “consumer market access (manufacturing)” following the discussion in Section IV of the main paper.\(^{25}\) We also refer to \( IMA_o^A \) — defined formally below — as “input market access (agriculture)” because it captures how cheaply manufacturing firms can access intermediate inputs (i.e., agricultural goods). The price index among agricultural firms, which captures the cost of their ideal bundle of manufacturing inputs, is similar:

\[
(\Phi_d^M)^{-\theta_M} = \chi_2 \sum_{o \in U} B_o(q_o^{\alpha_M} w_o^{\gamma_M})^{-\theta_M} (IMA_o^A)^{\beta_M} (\tau_{od}^M)^{-\theta_M} \equiv IMA_o^M \quad \text{if } d \in R,
\]

where \( IMA_o^M \) as “input market access (manufacturing)” because it captures how cheaply agricultural firms can access inputs made by manufacturing firms.\(^{26}\) Note that \( CMA_d^M \) and \( IMA_d^M \) are very closely related but not identical: they are defined for different sets of regions \( d \) and in general \( \chi_1 \neq \chi_2 \).

Following the logic in Section IV of the main paper, the trade flows in manufactured goods follows a gravity equation, though we must keep track of demand for these goods from both consumers and agricultural firms. First, trade from manufacturing regions to consumers satisfies:

\[
X_{od}^M = \chi_1 B_o(q_o^{\alpha_M} w_o^{\gamma_M})^{-\theta_M} (IMA_o^A)^{\beta_M} (\tau_{od}^M)^{-\theta_M} (CMA_d^M)^{-1} X_d^M \quad \text{if } o \in U \text{ and } d \in U, R,
\]

where \( X_d^M \) denotes the consumer expenditure on manufacturing goods in region \( d \). Similarly, trade flows from manufacturing regions to agricultural firms follows

\[
\tilde{X}_{od}^M = \chi_2 B_o(q_o^{\alpha_M} w_o^{\gamma_M})^{-\theta_M} (IMA_o^A)^{\beta_M} (\tau_{od}^M)^{-\theta_M} (IMA_d^M)^{-1} \tilde{X}_d^M \quad \text{if } o \in U \text{ and } d \in R,
\]

where \( \tilde{X}_d^M \) denotes the expenditure on manufacturing goods among agricultural firms in

\(^{25}\) Here, \( \chi_1 \equiv [\Gamma(\theta_M/(1-\alpha_M-\beta_M-\gamma_M))^{\theta_M/\theta_M} (1-\alpha_M-\beta_M-\gamma_M)^{-\theta_M}]_{\theta_M} \), where \( \Gamma(.) \) is the \( \Gamma \) function.

\(^{26}\) Here, \( \chi_2 \equiv [\Gamma(\theta_M/(1-\psi_M))^{\theta_M/\theta_M} (1-\alpha_M-\beta_M-\gamma_M)^{-\theta_M}]_{\theta_M} \).
region $d$.

Derivations are similar in the agricultural sector. The consumer price index for agricultural goods in any region $d$ is given by:

$$
(P_d^A)^{-\theta_A} = \chi_3 \sum_{o \in R} A_o (q_o^{oA} w_o^\gamma_A)^{-\theta_A} (IMA_o^M)^{\beta_A} (\tau_{od}^A)^{-\theta_A} \equiv CMA_o^A \quad \text{if } d \in U, R,
$$

where we denote by $CMA_o^A$ the “consumer market access (agriculture)” following our above definition of $CMA_o^M$.\textsuperscript{27} The price index among manufacturing firms, which captures the cost of their ideal bundle of agricultural inputs, is similar:

$$
(\Phi_d^A)^{-\theta_A} = \chi_4 \sum_{o \in R} A_o (q_o^{oA} w_o^\gamma_A)^{-\theta_A} (IMA_o^M)^{\beta_A} (\tau_{od}^A)^{-\theta_A} \equiv IMA_o^A \quad \text{if } d \in U,
$$

where we refer to $IMA_o^A$ as “input market access (agriculture)” because it captures how cheaply manufacturing firms can access intermediate inputs (i.e., agricultural goods).\textsuperscript{28}

Turning to predictions about agricultural trade, all results are analogous to those in the manufacturing sector. Trade from agricultural regions to consumers in all regions $d$ satisfies

$$
X_{od}^A = \chi_3 A_o (q_o^{oA} w_o^\gamma_A)^{-\theta_A} (IMA_o^M)^{\beta_A} (\tau_{od}^A)^{-\theta_A} (CMA_d^A)^{-1} X_d^A \quad \text{if } o \in R \text{ and } d \in U, R,
$$

whereas agricultural trade to manufacturing firms follows

$$
\tilde{X}_{od}^A = \chi_4 A_o (q_o^{oA} w_o^\gamma_A)^{-\theta_A} (IMA_o^M)^{\beta_A} (\tau_{od}^A)^{-\theta_A} (IMA_d^A)^{-1} \tilde{X}_d^A \quad \text{if } o \in R \text{ and } d \in U,
$$

where $X_d^A$ is consumer demand and $\tilde{X}_d^A$ is manufacturing firm demand for agricultural goods.

**Solving for Agricultural Land Values:** Following a similar derivation to that in Section IV of the main paper, the rental rate of land ($q_o$) in agricultural region $o \in R$ is given by:

$$
\ln q_o = \chi_5 + \left( \frac{1}{1 + \alpha_A \theta_A} \right) \ln \left( \frac{A_o}{L_o} \right) + \left( \frac{\gamma_A \mu}{1 + \alpha_A \theta_A} \right) \ln CMA_o^A + \left[ \frac{\gamma_A (1 - \mu) \theta_A}{\theta_M (1 + \alpha_A \theta_A)} \right] \ln CMA_o^M + \left( \frac{\beta_A}{1 + \alpha_A \theta_A} \right) \ln IMA_o^M + \left( \frac{1}{1 + \alpha_A \theta_A} \right) \ln FMA_o^A \quad \text{if } o \in R,
$$

\textsuperscript{27}Here, $\chi_3 \equiv [\Gamma(\theta_A/(1 - \sigma_A))]^{-\theta_A/(1 - \sigma_A)} r^{-\sigma_A - \beta_A - \gamma_A} \theta_A$.

\textsuperscript{28}Here, $\chi_4 \equiv [\Gamma(\theta_A/(1 - \psi_A))]^{-\theta_A/(1 - \psi_A)} r^{-\psi_A - \beta_A - \gamma_A} \theta_A$. 

12
where

\[(14) \quad FMA_o^A \equiv \chi_3 \sum_{d \in \mathcal{R}} (\tau_{od}^A - \theta_A(CMA_o^A)^{-1}X_d^A) + \chi_4 \sum_{d \in \mathcal{U}} (\tau_{od}^A - \theta_A(IMA_o^M)^{-1}\tilde{X}_d^A) \quad \text{if } o \in \mathcal{R},\]

and, analogously to Section IV of the main paper, \(FMA_o^A\) refers to “firm market access (agriculture)”.\(^{29}\) Equation (14) takes a similar form to equation (10) in Section IV of the main paper, which was the main log-linear equation relating agricultural land values to market access that drives our empirical analysis. But here, in equation (14), there are four different sources of market access that are potentially distinct from one another (even when trade costs are symmetric such that, in the one-sector model in Section IV of the main paper, \(CMA\) and \(FMA\) were proportional to one another and simplified to one term). That is, agricultural land values in region \(o\) are relatively high if consumers in region \(o\) have good access to agricultural goods (high \(CMA_o^A\)) or manufactured goods (high \(CMA_o^M\)), if agricultural firms in region \(o\) have good access to intermediate manufactured inputs (high \(IMA_o^M\)), or if agricultural firms in region \(o\) have good access to consumers and firms to whom they can sell agricultural output (high \(FMA_o^A\)).

The main insight from Section IV of the main paper — that the price of immobile land is higher in regions that have higher market access — continues to hold in even this substantially more complicated economic environment. There are now several notions of market access, which differ from each other in nuanced ways, but their basic functional relationships are very similar. In principle, it is possible to create different empirical measures for these separate market access terms, following steps analogous to those in Section IV of the main paper; in practice, however, these terms are so similar in our empirical setting that empirical proxies for these terms (\(\ln CMA_o^A\), \(\ln CMA_o^M\), \(\ln IMA_o^M\), and \(\ln FMA_o^A\)) have joint correlations that exceed 0.999.\(^{30}\)

Our main conclusion is that, while there is little hope of estimating such a model in our setting, our notion of “market access” effectively proxies for many types of economic interactions along these lines. For interpreting the empirical results, and why market access matters for agricultural land values, it could be any of these channels that reflect economic gains from trade with other markets: trade in consumer goods or intermediate goods, and from a worker’s perspective or a firm’s perspective.

\(^{29}\)Here \(\chi_5 \equiv \left(1 - \frac{1}{1 + \alpha A \theta_A}\right) \ln \alpha - \left(\frac{\gamma A \theta A}{1 + \alpha A \theta_A}\right) \ln \bar{U}.\) This derivation applies goods market clearing, noting that \(Y_o^A = \sum_{d \in \mathcal{R}, \mathcal{U}} X_{od}^A + \sum_{d \in \mathcal{U}} \tilde{X}_{od}^A.\)

\(^{30}\)In particular, we created first-order empirical approximations to these terms that are analogous to our definition of market access in equation (12) of the main paper.
III.B Procedure for Counterfactual Simulations

We provide here more details concerning the counterfactual simulations that appear in Section VII of the main paper; we therefore revert to the one-sector model of Section IV of the main paper. Recall that our goal is to solve for the new equilibrium allocation — new population distribution and then either the new level of worker utility, $\bar{U}$, or the new level of aggregate population, $\bar{N}$, depending on the assumption made about international worker mobility — that would arise were railroads to be removed from the 1890 economy. This counterfactual exercise therefore amounts to changing the matrix of trade costs $\tau_{od}$ in the model (from those that relate to the 1890 transportation network to the transportation network that would arise were railroads to be removed) while holding the productivity terms $A_o$ and exogenous land areas $L_o$ for each county fixed at their factual 1890 levels. The challenge in performing this simulation is that the terms $A_o$ and $L_o$ are unknown. We therefore describe here a procedure for first estimating these terms, through the structure of our model, from the available 1890 data on the population distribution.

We begin by noting that equation (9) in the main paper can be written as

$$P_o^{-\theta} = \sum_d \frac{\tau_{od}^{-\theta} P_d N_d}{\sum_i \tau_{di}^{-\theta} P_i^{1+\theta} N_i^\theta},$$

Given data on trade costs $\tau_{od}$ from the factual 1890 scenario, population levels $N_o$ from 1890 and a value for the parameter $\theta$, we solve this system of equations for the unique (up to scale) price index $P_o$ in each region in 1890.\footnote{This is equivalent to the procedure described in footnote 57 of the main paper, which we use to construct a value of $MA_o(\theta)$, for any candidate value of $\theta$, as an input into our NLS routine for the purposes of estimating $\theta$. The results in Allen and Arkolakis (2014) imply that there is a unique (up to scale) solution of equation (15).}

Substituting equation (3) from the main paper into equation (4) from the main paper, we obtain

$$P_o^{-\theta} \bar{U}^{(\alpha+\gamma)} = \xi_1 f(P)^{\theta(1-\alpha-\gamma)} \sum_d C_d^{\tau_{od}^{-\theta}} P_d^{\theta(\alpha+\gamma)} N_d^{-\theta\alpha},$$

where $C_d \equiv L_d^{\theta} T_d$ and our assumption about the internationally mobility of capital amounts to assuming that $r/f(P) = \bar{r}$, where $\bar{r}$ is the exogenous foreign real interest rate and $f(P)$ is the price index of the goods that are implicitly traded for foreign capital.\footnote{Here, $\xi_1 \equiv \left( \left[ \Gamma \left( \frac{\theta+1-\sigma}{\theta} \right) \right]^{-\theta} \bar{r}^{-\theta(1-\alpha-\gamma)} \left( \frac{\alpha}{\theta} \right)^{-\alpha\theta} \right)$ and $C_i = L_i^{\theta} T_i^{(\theta)}$. As discussed in footnote 35 of the main paper, our baseline case is that in which foreign capital is traded for goods in New York City (the major international trading exchange and financial center at the time).} Having solved for relative prices $P_o$ in the factual 1890 equilibrium, from the solution to equation (15), we
use this solution in equation (16), along with data on $\tau_{od}$ and $N_d$ from 1890, to solve for the vector of $C_d$ terms (up to scale, which amounts to using the normalization that $\bar{U} = 1$ in the factual 1890 equilibrium).

At this stage we have therefore identified the unknown terms $C_d \equiv L_d^{a_d}T_d$ for each county in 1890. While the variable $C_d$ is, for any county, an unknown mixture of productivity ($T_d$) and land endowment ($L_d$) it turns out to be exactly this mixture that is necessary for the calculations below. We now go on to solve for a new counterfactual equilibrium in which $C_d$ is held constant (at its 1890 level) but new trade costs $\tau_{od}$ prevail, corresponding to the removal of railroads. Following the logic in Allen and Arkolakis (2014), we know that there exists a unique equilibrium at our parameter values and that this equilibrium can be found by a straightforward iterative algorithm. This procedure solves for all relative population levels and land values (relative to a numeraire good) in the new counterfactual equilibrium. The overall level can then be solved for either by assuming that the total population level $\bar{N}$ does not change (as in Allen and Arkolakis (2014)) or by assuming that worker utility $\bar{U}$ does not change (and hence $\bar{N}$ does change).
References


Appendix Figure 1. Sample of 2,327 Counties in the Regression Analysis

Notes: This map shows the 2,327 sample counties in the regression analysis, which are all counties that report non-zero land values in 1870 and 1890. The excluded geographic areas are cross-hashed. County boundaries correspond to county boundaries in 1890.
Appendix Figure 2. Sample of 2,782 Counties in the Counterfactual Analysis

Notes: This map shows the 2,782 sample counties in the counterfactual analysis. Compared to Appendix Figure 1, the additional 455 counties are those that report land value data in 1890 only. The excluded geographic areas are cross-hashed. County boundaries correspond to county boundaries in 1890.
Appendix Figure 3. Changes in Log Land Value from 1870 to 1890, by County

Notes: This map shows the 2,327 sample counties, shaded according to their change in log land value from 1870 to 1890. Counties are divided into seven groups (with an equal number of counties per group), and darker shades denote larger increases in land value.
Appendix Figure 4. Counterfactual Changes in Log Market Access, by County

Notes: This map shows the 2,782 counterfactual sample counties, shaded according to their change in log market access from 1890 to the baseline counterfactual scenario. Counties are divided into seven groups (with an equal number of counties per group), and darker shades denote larger declines in market access.
### Appendix Table 1. Transportation Network Database Components

<table>
<thead>
<tr>
<th>ID</th>
<th>Component Name</th>
<th>Component Definition</th>
<th>Construction Description</th>
<th>Baseline Cost (in Dollars)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Navigable Rivers</td>
<td>Fogel's definition of navigable rivers; time invariant component of network</td>
<td>Hand-traced from Fogel (1964)</td>
<td>$0.0049 \times [\text{Length}]</td>
</tr>
<tr>
<td>1</td>
<td>Constructed Canals</td>
<td>Fogel's definition of navigable canals; time invariant component of network</td>
<td>Hand-traced from Fogel (1964)</td>
<td>$0.0049 \times [\text{Length}]</td>
</tr>
<tr>
<td>2</td>
<td>Proposed Canals</td>
<td>Fogel's definition of proposed canals; time invariant component of network; only included in alternative counterfactual scenarios</td>
<td>Hand-traced from Fogel (1964)</td>
<td>$0.0049 \times [\text{Length}]</td>
</tr>
<tr>
<td>3</td>
<td>Sea/Lake Routes</td>
<td>Multiple point-to-point connections throughout the Great Lakes and Oceans; time invariant component of network</td>
<td>Created manually to effectively saturate area</td>
<td>$0.0049 \times [\text{Length}]</td>
</tr>
<tr>
<td>5</td>
<td>Railroad Harbor</td>
<td>Points where transshipment to Sea/Lake Routes is considered possible, created wherever the 1911 railroad network approaches the coastline; time invariant component of the network</td>
<td>Created manually as a short line from a Sea/Lake route</td>
<td>$0.5</td>
</tr>
<tr>
<td>6</td>
<td>Railroads</td>
<td>Railroad lines as depicted on maps from 1870 (Colton 1870) and 1887 (Cram 1887); time variant component of the network.</td>
<td>Hand-traced from a railroad map in 1911 (Whitney and Smith 1911), which was most accurately geo-referenced to county borders. Maps for 1890 and 1870 were then created by manually deleting railroad lines that do not appear in the earlier periods.</td>
<td>$0.0063 \times [\text{Length}]</td>
</tr>
<tr>
<td>8</td>
<td>Wagon Routes (Centroid-to-Centroid)</td>
<td>Wagon Routes connecting any two centroids within a distance of 300km; time invariant component of network</td>
<td>Created automatically in ArcGIS</td>
<td>$0.231 \times [\text{Length}]</td>
</tr>
<tr>
<td>9</td>
<td>Sea Route Between Coasts</td>
<td>Direct sea route connecting the West Coast (near San Diego) to the East Coast (in the Gulf of Mexico); time invariant component of network</td>
<td>Created manually</td>
<td>$8</td>
</tr>
<tr>
<td>13</td>
<td>In-county centroid-to-railroad connection</td>
<td>Represents the average wagon route from any point in the county to railroad lines that pass through the county; a transshipment cost is then incurred; time variant component of network</td>
<td>Created manually to connect a centroid to railroads within the county</td>
<td>$0.5 + 0.231 \times [\text{Mean}_\text{Length}]</td>
</tr>
<tr>
<td></td>
<td>Description</td>
<td>Network Representation</td>
<td>Length Component</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------</td>
<td>------------------</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Out-of-county centroid-to-railroad connection</td>
<td>Represents the average wagon route from any point in the county to relevant railroad lines outside the county in various directions; a transshipment cost is then incurred; time invariant component of network</td>
<td>$0.5 + 0.231 \times \text{Length}$.</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Out-of-county centroid-to-harbor connection</td>
<td>Represents the average wagon route from any point in the county to relevant harbors (see below) outside the county in various directions; a transshipment cost is then incurred; time invariant component of network</td>
<td>$0.5 + 0.231 \times \text{Length}$.</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>River Harbor</td>
<td>Points where transshipment to Sea/Lake Routes is considered possible, created wherever rivers, canals, or proposed canals meet the coastline; time invariant component of the network</td>
<td></td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>Canal Harbor</td>
<td>Points where rivers and canals meet and transshipment is possible; time invariant component of the network</td>
<td>$0.5$.</td>
<td></td>
</tr>
<tr>
<td>62</td>
<td>Proposed Canal Harbor</td>
<td>Points where canals and proposed canals meet and transshipment is not necessary; time invariant component of the network</td>
<td></td>
<td></td>
</tr>
<tr>
<td>601</td>
<td>River-to-Canal Transshipment Point</td>
<td>Points where rivers and canals meet and transshipment is possible; time invariant component of the network</td>
<td>$0.5$.</td>
<td></td>
</tr>
<tr>
<td>602</td>
<td>River-to-Proposed Canal Transshipment Point</td>
<td>Points where rivers and proposed canals meet and transshipment is possible; time invariant component of the network</td>
<td></td>
<td></td>
</tr>
<tr>
<td>612</td>
<td>Canal-to-Proposed Canal Transshipment Point</td>
<td>Points where canals and proposed canals meet and transshipment is not necessary; time invariant component of the network</td>
<td>$0.0049 \times \text{Length}$.</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>River-to-Railroad Transshipment Point</td>
<td>Points where rivers and railroads meet and transshipment is possible; time invariant component of the network</td>
<td></td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>Canal-to-Railroad Transshipment Point</td>
<td>Points where canals and railroads meet and transshipment is possible; time invariant component of the network</td>
<td>$0.5$.</td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>Proposed Canal-to-Railroad Transshipment Point</td>
<td>Points where proposed canals and railroads meet and transshipment is possible; time invariant component of the network</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

23
<table>
<thead>
<tr>
<th></th>
<th>In-county centroid-to-river connection</th>
<th>Represents the average wagon route from any point in the county to waterway lines that pass through the county (river, canal, or proposed canal); a transshipment cost is then incurred; time variant component of network</th>
<th>Created manually to connect a centroid to waterways within the county</th>
<th>0.5 + 0.231 * [Mean_Length]</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>In-county centroid-to-canal connection</td>
<td></td>
<td></td>
<td>0.5 + 0.231 * [Length]</td>
</tr>
<tr>
<td>81</td>
<td>In-county centroid-to-proposed canal connection</td>
<td></td>
<td></td>
<td>0.5 + 0.231 * [Length]</td>
</tr>
<tr>
<td>82</td>
<td>Out-of-county centroid-to-river connection</td>
<td>Represents the average wagon route from any point in the county to relevant waterway lines outside the county border in various directions (river, canal, or proposed canal); a transshipment cost is then incurred; time invariant component of network</td>
<td>Created manually to connect a centroid to potentially-relevant waterways outside the county</td>
<td>0.5 + 0.231 * [Length]</td>
</tr>
<tr>
<td>90</td>
<td>Out-of-county centroid-to-canal connection</td>
<td></td>
<td></td>
<td>0.5 + 0.231 * [Length]</td>
</tr>
<tr>
<td>91</td>
<td>Out-of-county centroid-to-proposed canal connection</td>
<td></td>
<td></td>
<td>0.5 + 0.231 * [Length]</td>
</tr>
<tr>
<td>N/A</td>
<td>County Borders</td>
<td>1890 County Borders</td>
<td>Downloaded from nhgis.org</td>
<td>N/A</td>
</tr>
<tr>
<td>N/A</td>
<td>County Centroids</td>
<td>The geographic center (centroid) of each 1890 county</td>
<td>Created using ArcGIS &quot;Feature to point&quot; tool</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Notes: In the above formulas, [Length] refers to each segment's length in miles. To calculate the respective costs, we create the field "length" for each element of the network, assign to it the length in miles (using "Calculate Geometry"), and then create a field "Cost" to which we assign a value based on the indicated formula (using "Field Calculator").

In the above formulas, [Mean_Length] is again measured in miles and is our estimate of the average distance of a point in the county to the respective waterway or railroad line that passes through the county. In particular, we created 200 random points within each county (using the "Create Random Points" tool). We then calculate the minimum distance from each of those points to the respective mode of transportation (using the "Near" tool and converting these distances to miles. We then collapse the 200 observations into 1 with the mean distance (using the "Dissolve" tool) and merge this data to each county's in-county connections to that mode of transportation (using as merge field the county's unique identifier in the "Join Field" tool). By contrast, when a waterway or railroad line passes outside a county, the measured distance from the centroid is a sufficient approximation of the average distance from points in the county.
<table>
<thead>
<tr>
<th>Number of Counties</th>
<th>Mean in 1870</th>
<th>Mean in 1890</th>
<th>Log Change from 1870 to 1890</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Market Access</td>
<td>2,327</td>
<td>7,804,585</td>
<td>15,804,880</td>
</tr>
<tr>
<td></td>
<td>[2,246,742]</td>
<td>[3,220,369]</td>
<td>[0.288]</td>
</tr>
<tr>
<td>Land Value</td>
<td>2,327</td>
<td>7,728,604</td>
<td>8,337,710</td>
</tr>
<tr>
<td></td>
<td>[5,847,713]</td>
<td>[5,902,241]</td>
<td>[0.532]</td>
</tr>
<tr>
<td>Farmland Acres</td>
<td>2,327</td>
<td>288,202</td>
<td>321,174</td>
</tr>
<tr>
<td></td>
<td>[134,718]</td>
<td>[141,237]</td>
<td>[0.297]</td>
</tr>
<tr>
<td>Population</td>
<td>2,327</td>
<td>41,296</td>
<td>60,096</td>
</tr>
<tr>
<td></td>
<td>[61,841]</td>
<td>[117,292]</td>
<td>[0.313]</td>
</tr>
</tbody>
</table>

Notes: Each cell reports average county characteristics for the main regression sample of 2,327 counties (e.g., Appendix Figure 1), where counties are weighted by their value of land in 1870 (i.e., corresponding to the main regression specification). For each indicated county characteristic (by row): column 2 reports counties' weighted average in 1870, column 3 reports counties' weighted average in 1890, and column 4 reports counties' weighted average change from 1870 to 1890 (in logs). Standard deviations are reported in brackets.
## Appendix Table 3. Robustness to Changes in the Definition of Market Access

<table>
<thead>
<tr>
<th>Changes in Definition of Market Access</th>
<th>Estimated Impact of Market Access</th>
<th>Percent Decline in Land Value Without Railroads</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Specification</td>
<td>0.511 (0.065)</td>
<td>60.2 (4.2)</td>
</tr>
<tr>
<td>1. Include access to international markets</td>
<td>0.511 (0.065)</td>
<td>59.7 (4.1)</td>
</tr>
<tr>
<td>2. Adjustment for Census undercounting</td>
<td>0.511 (0.065)</td>
<td>60.2 (4.2)</td>
</tr>
<tr>
<td>3. Measure access to county wealth</td>
<td>0.503 (0.065)</td>
<td>59.6 (4.2)</td>
</tr>
<tr>
<td>4. Include access to own market</td>
<td>0.523 (0.067)</td>
<td>60.4 (4.2)</td>
</tr>
<tr>
<td>5. Limit access to counties beyond 5 miles</td>
<td>0.503 (0.065)</td>
<td>60.1 (4.2)</td>
</tr>
<tr>
<td>6. Limit access to counties beyond 50 miles</td>
<td>0.491 (0.064)</td>
<td>59.7 (4.2)</td>
</tr>
<tr>
<td>7. Limit access to counties beyond 200 miles</td>
<td>0.491 (0.066)</td>
<td>60.4 (4.4)</td>
</tr>
<tr>
<td>8. Limit access to only urban areas</td>
<td>0.496 (0.063)</td>
<td>56.4 (4.0)</td>
</tr>
<tr>
<td>9. Limit access to only cities</td>
<td>0.484 (0.062)</td>
<td>53.8 (3.9)</td>
</tr>
<tr>
<td>10. Limit access to only New York City</td>
<td>0.500 (0.064)</td>
<td>49.1 (3.6)</td>
</tr>
<tr>
<td>11. Set parameter &quot;theta&quot; equal to 1</td>
<td>4.19 (0.54)</td>
<td>66.5 (4.3)</td>
</tr>
<tr>
<td>12. Set parameter &quot;theta&quot; equal to 3.37</td>
<td>1.25 (0.16)</td>
<td>63.9 (4.3)</td>
</tr>
<tr>
<td>13. Set parameter &quot;theta&quot; equal to 3.60</td>
<td>1.17 (0.15)</td>
<td>63.7 (4.3)</td>
</tr>
<tr>
<td>14. Set parameter &quot;theta&quot; equal to 3.80</td>
<td>1.11 (0.14)</td>
<td>63.5 (4.2)</td>
</tr>
<tr>
<td>15. Set parameter &quot;theta&quot; equal to 6.74</td>
<td>0.623 (0.080)</td>
<td>61.2 (4.2)</td>
</tr>
<tr>
<td>16. Set parameter &quot;theta&quot; equal to 12.86</td>
<td>0.331 (0.041)</td>
<td>57.3 (4.0)</td>
</tr>
<tr>
<td>17. Set parameter &quot;theta&quot; equal to 13.18</td>
<td>0.324 (0.040)</td>
<td>57.1 (4.0)</td>
</tr>
</tbody>
</table>

Notes: Each row reports estimates from the indicated specification, as discussed in the appendix text (section II.A). Column 1 reports the estimated impact of Log Market Access on Log Value of Agricultural Land, and column 2 reports the estimated percent decline in agricultural land value for an 1890 counterfactual scenario with no railroads. Robust standard errors clustered by state are reported in parentheses.
# Appendix Table 4. Robustness to Changes in the Transportation Cost Parameters

<table>
<thead>
<tr>
<th></th>
<th>Estimated Impact of Market Access</th>
<th>Percent Decline in Land Value Without Railroads</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Baseline Specification</td>
<td>0.511 (0.065)</td>
<td>60.2 (4.2)</td>
</tr>
<tr>
<td><strong>Alternative Transportation Cost Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Reduce sea routes to 0.198 cents</td>
<td>0.509 (0.066)</td>
<td>57.1 (4.1)</td>
</tr>
<tr>
<td>2. Reduce water costs to 0.198 cents</td>
<td>0.504 (0.068)</td>
<td>51.2 (4.1)</td>
</tr>
<tr>
<td>3. Remove transshipment within waterways</td>
<td>0.512 (0.065)</td>
<td>59.0 (4.1)</td>
</tr>
<tr>
<td>4. Raise railroad cost to 0.735 cents</td>
<td>0.523 (0.066)</td>
<td>58.4 (4.1)</td>
</tr>
<tr>
<td>5. Raise railroad cost to 0.878 cents</td>
<td>0.534 (0.068)</td>
<td>56.1 (4.0)</td>
</tr>
<tr>
<td>6. Reduce wagon cost to 14 cents</td>
<td>0.784 (0.100)</td>
<td>65.3 (4.2)</td>
</tr>
<tr>
<td>7. Exclude Western region</td>
<td>0.464 (0.061)</td>
<td>57.0 (4.3)</td>
</tr>
<tr>
<td>8. Reduce Pacific-to-Atlantic cost to $5</td>
<td>0.510 (0.065)</td>
<td>59.7 (4.1)</td>
</tr>
<tr>
<td>9. Increase Pacific-to-Atlantic cost to $11</td>
<td>0.513 (0.065)</td>
<td>60.8 (4.2)</td>
</tr>
<tr>
<td>10. Exclude Pacific-to-Atlantic connection</td>
<td>0.514 (0.066)</td>
<td>62.3 (4.2)</td>
</tr>
</tbody>
</table>

Notes: Each row reports estimates from the indicated specification, as discussed in the appendix text (section II.B). Column 1 reports the estimated impact of Log Market Access on Log Value of Agricultural Land, and column 2 reports the estimated percent decline in agricultural land value for an 1890 counterfactual scenario with no railroads. Robust standard errors clustered by state are reported in parentheses.
### Appendix Table 5. Robustness to Alternative Empirical Specifications

<table>
<thead>
<tr>
<th></th>
<th>Estimated Impact of Market Access</th>
<th>Percent Decline in Land Value Without Railroads</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline Specification</strong></td>
<td>0.511 (0.065)</td>
<td>60.2 (4.2)</td>
</tr>
<tr>
<td><strong>Alternative Empirical Specifications</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Fixed Effects for 20 &quot;resource regions,&quot; by year</td>
<td>0.444 (0.063)</td>
<td>55.6 (4.6)</td>
</tr>
<tr>
<td>2. Fixed Effects for 145 &quot;resource subregions,&quot; by year</td>
<td>0.403 (0.059)</td>
<td>52.4 (4.7)</td>
</tr>
<tr>
<td>3. Fifth-order polynomial in latitude and longitude</td>
<td>0.438 (0.074)</td>
<td>55.1 (5.5)</td>
</tr>
<tr>
<td>4. First-order polynomial in latitude and longitude</td>
<td>0.563 (0.087)</td>
<td>63.3 (5.0)</td>
</tr>
<tr>
<td>5. Drop top/bottom centile, change in market access</td>
<td>0.517 (0.066)</td>
<td>60.6 (4.2)</td>
</tr>
<tr>
<td>6. Drop top/bottom 5 centiles, change in market access</td>
<td>0.548 (0.073)</td>
<td>62.4 (4.4)</td>
</tr>
<tr>
<td>7. Drop top/bottom centile, change in land value</td>
<td>0.519 (0.067)</td>
<td>60.7 (4.2)</td>
</tr>
<tr>
<td>8. Drop top/bottom 5 centiles, change in land value</td>
<td>0.472 (0.061)</td>
<td>57.6 (4.2)</td>
</tr>
<tr>
<td>9. Adjustment to land value, homestead fees</td>
<td>0.499 (0.064)</td>
<td>59.4 (4.2)</td>
</tr>
<tr>
<td>10. Adjustment to land value, preemption fees</td>
<td>0.378 (0.061)</td>
<td>50.4 (5.1)</td>
</tr>
<tr>
<td>11. Adjustment for cost of land improvements, Fogel</td>
<td>0.446 (0.057)</td>
<td>55.7 (4.1)</td>
</tr>
<tr>
<td>12. Adjustment for cost of land improvements, Primack</td>
<td>0.596 (0.077)</td>
<td>65.2 (4.2)</td>
</tr>
<tr>
<td>13. Drop counties with any city population</td>
<td>0.494 (0.070)</td>
<td>59.1 (4.6)</td>
</tr>
<tr>
<td>14. Drop counties with any urban population</td>
<td>0.429 (0.079)</td>
<td>54.5 (6.0)</td>
</tr>
</tbody>
</table>

Notes: Each row reports estimates from the indicated specification, as discussed in the appendix text (section II.C). Column 1 reports the estimated impact of Log Market Access on Log Value of Agricultural Land, and column 2 reports the estimated percent decline in agricultural land value for an 1890 counterfactual scenario with no railroads. Robust standard errors clustered by state are reported in parentheses.