RAILROADS AND AMERICAN ECONOMIC GROWTH: A "MARKET ACCESS" APPROACH*

DAVE DONALDSON AND RICHARD HORNBECK

This article examines the historical impact of railroads on the U.S. 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 U.S. agricultural land by 60%, with limited potential for mitigating these losses through feasible extensions to the canal network or improvements to country roads. *JEL* Codes: N01, N51, N71, F1, O1, R1.

I. INTRODUCTION

Railroads spread throughout a growing United States in the nineteenth 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 U.S. economy or at least very influential for economic growth. Our understanding of the development of the U.S. 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"

*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, Julius Luettge, Meredith McPhail, Rui Wang, Sophie Wang, and Kevin Wu provided excellent research assistance. This material is based on work supported by the National Science Foundation under Grant No. 1156239.

The Quarterly Journal of Economics (2016), 799–858. doi:10.1093/qje/qjw002.
Advance Access publication on February 28, 2016.

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 railroads had only a small aggregate impact on the U.S. 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).[1]

There is an appeal to a methodology that directly estimates 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 article 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 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 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

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.

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 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 article 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 U.S. 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 and Fogel's estimates neglect many other channels through which railroads may have affected other economic sectors and/or technological growth.[2]

2. 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 then we 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 distribution of population across U.S. counties. Holding the total U.S. 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 U.S. population would need to fall substantially. An alternative scenario that we consider holds the total U.S. 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 Fogel proposed. 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 U.S. 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.[3] Beyond the substantial effects on agricultural land value, however, our analysis anticipates 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 our ability to measure and analyze impacts of "market access" will spur further research on the aggregate impacts of railroads throughout the U.S. economy.[4]

3. 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 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.

4. In 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 U.S. 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 article 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 evaluate the impact of counterfactual scenarios. This article's methodological approach is more suited to settings with high mobility of labor, which appears to reflect the historical U.S. economy more than the Indian economy. The concept of market access has been useful for empirical work (surveyed by Redding 2010), though this article 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.

More broadly, this article 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 article is organized as follows. Section II reviews and extends Fogel's 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.

5. In 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.

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. U.S. 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) developed 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, although 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 his calculations, using modern spatial analysis tools and digitized county-level data.

Fogel (1964) estimated that the social saving from railroads in the agricultural sector in 1890 was no more than 2.7% of GNP. He divided this impact into that coming from inter-regional trade (0.6%) and intraregional trade (2.1%). For inter-regional 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 calculated the annual inter-regional 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]

6. Indeed, 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 effect 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 referred 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 bounded 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 inter-regional analysis, Fogel calculated 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, although 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 emphasized 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 estimated 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

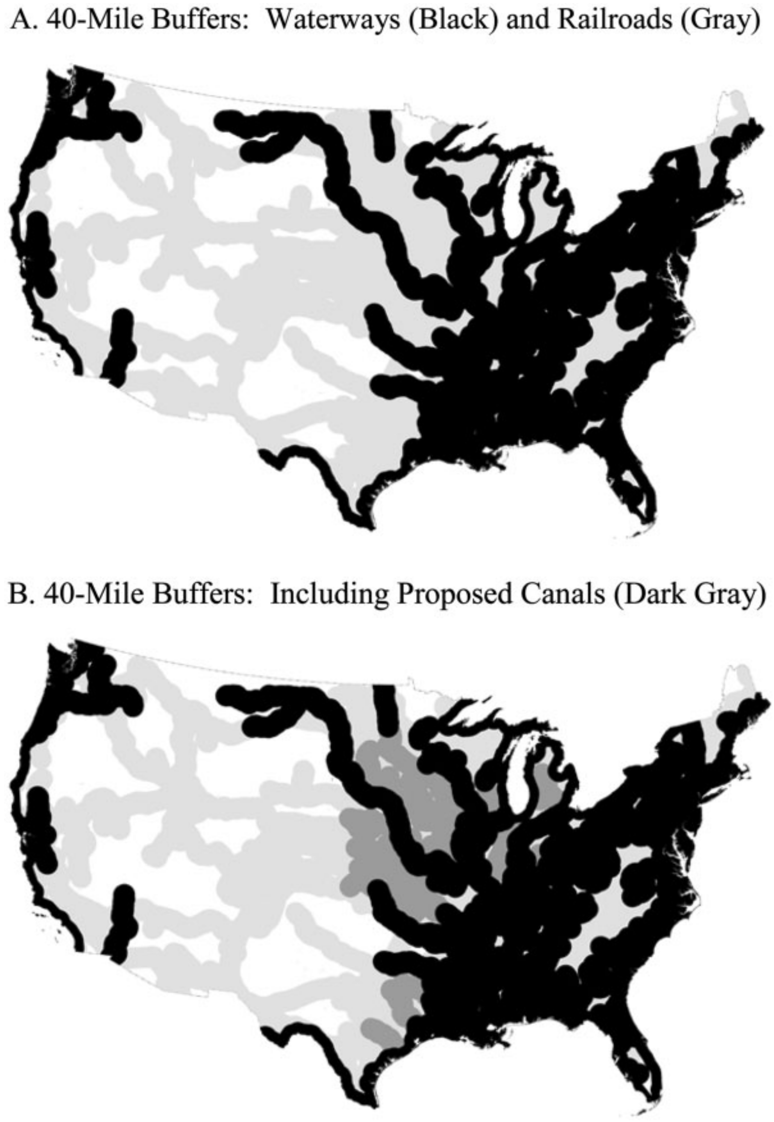

FIGURE I

Distance Buffers in 1890 around Waterways, Railroads, and Proposed Canals

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.

C. 10-Mile Buffers: Waterways (Black) and Railroads (Gray)

D. 10-Mile Buffers: Including Proposed Canals (Dark Gray)

FIGURE I

Continued

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 $186 million annual return on agricultural land in the infeasible region that is only moderately larger than Fogel's approximation of $154 million.[7] Consistent with 40 miles being a reasonable cutoff distance for the infeasible region, we calculate an annual return of only $5 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.[8]

7. 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 preferred measure of GNP in 1890 ($12 billion).

8. 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 zero 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 five-mile distance buffer of

Implementing this approach, we calculate the annual intraregional impact of removing railroads to be $325 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 analysis of distance buffers, we calculate an annual loss of $225 million or 1.9% 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 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.[9] We maintain Fogel's focus on the agricultural sector, as nonagricultural 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

a waterway and subtracting the county's land share within that buffer of a waterway or railroad.

9. There 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.

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 article uses a new data set 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: (i) transportation cost parameters that apply to a given unit length of each transportation mode (railroad, waterway, and wagon); (ii) a transportation network database that maps where freight could move along each transportation mode; and (iii) 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 ton-mile and waterway rates equal to 0.49 cents per ton-mile.[10] Transshipment costs 50 cents per ton, incurred whenever transferring

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

goods to/from a railroad car, river boat, canal barge, or ocean liner.[11] Wagon transportation costs 23.1 cents per ton-mile, defined as the straight-line distance between two points.[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 emphasized 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).[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.[14]

11. Fogel considers trans-shipment charges as a subcategory of water rates, but our modeling of trans-shipment points allows for a unified treatment of Fogel's inter-regional 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.

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.

13. First, year-specific maps of railroads are "georeferenced" to U.S. 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).

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

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 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] Trans-shipment 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 300 km.[19]

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.

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 trans-shipment costs. The need to include trans-shipment costs is the main reason 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.

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 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 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 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, Online 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, it is mainly restricted to transportation linkages within the United States.[21] The data only include U.S. counties'

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, because of technical limitations, Fogel approximated this average distance using one third of the distance from the farthest point in a county to a railroad.

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.

A. Natural Waterways

B. Natural Waterways and Canals

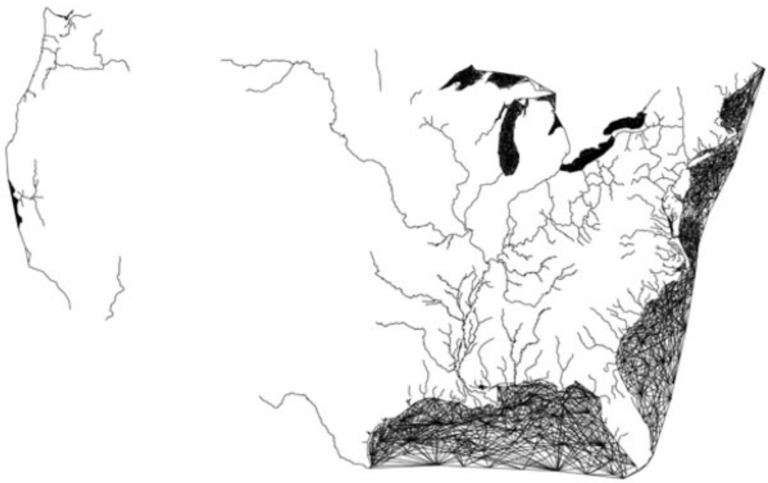

FIGURE II

Constructed Network Database (Partial)

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.

C. Natural Waterways, Canals, and 1870 Railroads

D. Natural Waterways, Canals, and 1890 Railroads

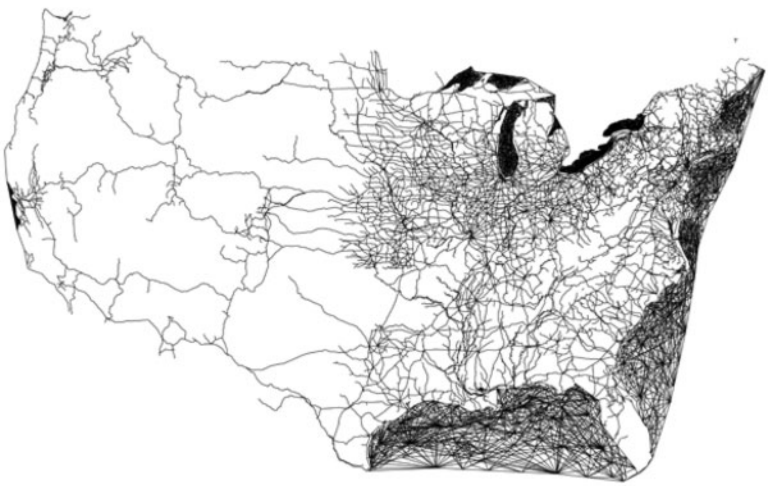

FIGURE II

Continued

access to other U.S. counties. As a robustness check, however, we incorporate international markets by assigning additional product demand and supply to U.S. 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.[22] Furthermore, 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 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 trans-shipment 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.

22. The Western regions are not central to the empirical analysis, which draws on within-state variation in changes in market access.

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, that is, 5 million calculations.[24] Initially, we calculate the lowest-cost routes under two scenarios: (i) the wagon, waterway, and railroad network in 1870; and (ii) the wagon, waterway, and railroad network in 1890. These transportation costs are used to calculate counties' market access in 1870 and 1890, so 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 U.S. 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).

Online Appendix Figure 1 maps the 2,327 counties included in the main regression analysis, which includes all counties with reported land value data in 1870 and 1890. Online Appendix Figure 2 maps a larger sample of 2,782 counties included in the

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).

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 U.S. 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 U.S. counties that specifies how each county is affected by

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.

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]

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 U.S. 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 σ).[27] Therefore, a consumer living in county o, who receives

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.

27. The elasticity of substitution is not restricted (beyond the discussion in note 9); that is, σ 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, σ indexes the inverse of the dispersion of the utility levels that agents enjoy from their ideal varieties).

income Y_o and faces a vector of prices $\mathbf{P_o}$, experiences indirect utility:

$$V(\mathbf{P}_o, Y_o) = \frac{Y_o}{P_o}, \tag{1}$$

where P_o is the ideal price index (a standard CES price index) over the continuum of varieties.[28]

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:

$$MC_o(j) = \frac{q_o^{\alpha} w_o^{\gamma} r_o^{1-\alpha-\gamma}}{z_o(j)}, \tag{2}$$

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 each county draws its productivity level, for any given variety j, from a Fréchet (or Type II extreme value) distribution with CDF given by: $F_o(z) = 1 - \exp(-A_o z^{-\theta})$, with $\theta > 1$.[29] This distribution captures how productivity differences across counties give incentives to specialize and trade, where these incentives are inversely related to θ.[30] We assume perfect competition among producers.[31]

28. That is, $P_o \equiv \left[\int_0^n (p_o(j))^{1-\sigma} dj\right]^{\frac{1}{1-\sigma}}$, 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.

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. 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.

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

31. An alternative (and observationally equivalent) formulation, following Melitz (2003), would assume that firms compete monopolistically with free entry

There are costs to trading varieties across counties. Remote locations pay high prices for imported varieties and receive low prices for varieties they produce, because 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 τ_{od} is applied to each unit of the variety shipped.[32] 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 $\tau_{od}^{k} \geq 1$.

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.[33] 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., $r_o = r$), although 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.[34] We assume that workers 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:

---

such that all firms' expected profits are zero and draw their productivity levels z, following Chaney (2008) and others, from a Pareto distribution $G_o(z) = 1 - (\frac{z}{A_o})^{-\theta}$, as typically seen in firm-level data sets (e.g., Axtell 2001).

32. 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.

33. 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.

34. 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 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 indexes.

$$w_o = \bar{U} P_o, \tag{3}$$

where $\bar{U}$ is the level of utility obtained by workers in each county. As we discuss further in Section IV.C, 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

1. 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:[35]

$$(P_d)^{-\theta} = \kappa_1 \sum_o A_o (q_o^{\alpha} w_o^{\gamma})^{-\theta} \tau_{od}^{-\theta} \equiv CMA_d. \tag{4}$$

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., τ_{od}).

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

$$X_{od} = \kappa_1 A_o (q_o^{\alpha} w_o^{\gamma})^{-\theta} \tau_{od}^{-\theta} CMA_d^{-1} Y_d. \tag{5}$$

35. Here, κ_1 is a constant defined by $\kappa_1 \equiv [\Gamma(\frac{\theta+1-\sigma}{\theta})]^{-\frac{\theta}{1-\sigma}} r^{-(1-\alpha-\gamma)\theta}$, where $\Gamma(\cdot)$ is the Γ 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).

2. Land Rental Rate. While trade flows between nineteenth-century U.S. 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). 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:[36]

$$(6)\quad (1+\alpha\theta)\ln q_o = \kappa_2 + \ln\left(\frac{A_o}{L_o}\right) - \gamma\theta\ln\bar{U} + \gamma\ln CMA_o + \ln FMA_o,$$

where FMA_o refers to "firm market access" for goods from origin o and is defined as:

$$(7)\qquad FMA_o \equiv \sum_d \tau_{od}^{-\theta} CMA_d^{-1} Y_d.$$

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 $\tau_{od}^{-\theta}$).

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

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:[37]

$$CMA_d = \sum_o \tau_{od}^{-\theta} FMA_o^{-1} Y_o. \tag{8}$$

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:[38]

$$MA_o = \kappa_3 \sum_d \tau_{od}^{-\theta} MA_d^{\frac{-(1+\theta)}{\theta}} N_d. \tag{9}$$

In words, a county's market access can be expressed as the sum over the cost of trading with each other county, that other county's population, and that other county's access to other markets.

Given the above simplifications, equation (6) becomes:[39]

$$\ln q_o = \kappa_4 + \left(\frac{1}{1+\alpha\theta}\right) \ln\left(\frac{A_o}{L_o}\right) + \left(\frac{1+\gamma}{1+\alpha\theta}\right) \ln(MA_o). \tag{10}$$

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 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

37. This result can be obtained by summing equation (5) over all destinations d and substituting $A_o(q_o^{\alpha} w_o^{\gamma})^{-\theta}$ into equation (4).

38. Here, $\kappa_3 \equiv \frac{\bar{U}\rho^{\frac{1+\theta}{\theta}}}{\gamma}$.

39. Here, $\kappa_4 \equiv \frac{1}{1+\alpha\theta}(\kappa_2 - \gamma \ln \rho - \gamma\theta \ln \bar{U})$.

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:[40]

$$\begin{aligned}\ln N_o =& \kappa_5 + \left(\frac{1}{1+\alpha\theta}\right)\ln(A_o) - \left(\frac{2+\alpha\theta}{1+\alpha\theta}\right)\ln(L_o)\\ &+ \left[\frac{1+\theta(1+\alpha+\gamma)}{\theta(1+\alpha\theta)}\right]\ln(MA_o).\end{aligned} \tag{11}$$

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

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.[41] Thus, both direct and indirect effects 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 τ_{od} into MA_o and then substituting the resulting

40. Here, $\kappa_5 \equiv \frac{\ln(\kappa_1\alpha\rho^{-\gamma})}{1+\alpha\theta} - \ln\left(\frac{\alpha}{\theta}\right) - \frac{1+\theta(\alpha+\gamma)}{1+\alpha\theta}\ln\bar{U}$.

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

counterfactual MAo 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 variables approaches that might purge the empirical estimates of endogeneity bias arising from local railroad placement decisions, all of which we pursue later.

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 $\overline{N} \equiv \sum_o N_o$ would change.[42] In the other extreme scenario, if there is a perfectly inelastic supply of international workers, the aggregate U.S. population would not change as a result of the counterfactual transportation costs but $\bar{U}$ would change.[43] As we discuss in Section VII, we can use the model here to solve for the resulting effects—that is, to solve for the change in either $\bar{U}$ or $\overline{N}$—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.

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

43. Furthermore, this would affect all counties' land rents because κ_3 in equation (10) depends on $\bar{U}$.

First, although 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 Online Appendix outlines a model with separate agricultural and manufacturing sectors, clarifying how railroads might affect 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 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).[44] We also present robustness checks that limit the sample to rural areas (where agricultural land would be minimally affected by local demand for land by manufacturing or housing).[45]

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 agriculture, we use the reported total value of land in farms and assume that

44. Furthermore, 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 U.S. city).

45. In the Online 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. Although manufacturing and housing use relatively little land, compared with 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.

land not in farms has zero agricultural value.[46] 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 nonagricultural purposes might also have agricultural value, which motivates robustness checks that restrict the sample to include only rural counties.

Third, although 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.[47] Land values reflect both contemporaneous rents and discounted future rents, so any correctly anticipated changes in market access would attenuate the estimated impact of changes in market access.[48] 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.[49] 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_o). 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

46. 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.

47. Formally, it is sufficient for us to assume that $V_o = \frac{q_o}{r}$, where V_o 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.

48. We suspect that residual changes in market access are difficult to anticipate. 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.

49. 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.

changes in county productivity are orthogonal to changes in market access from 1870 to 1890, after controlling for counties' geographic location (state, longitude, latitude).[50] 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 trade costs (τ_{od}). We approximate these trade costs using the calculated county-to-county lowest-cost freight transportation routes (described in Section III.D), 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), although we explore the sensitivity of our results to the particular freight rates that enter τ_{od} in MA_o.

Sixth, the market access term (MA_o) is not directly observed because some destination characteristics are unobserved.[51] 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.[52] 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:

50. Because the productivity term (A_o) 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.

51. From equation (4), the wage w_o and the technology term A_o are unobserved.

52. 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.

$$MA_o \approx \sum_d \tau_{od}^{-\theta} N_d. \tag{12}$$

The results are not sensitive to our use of the MA approximation in equation (12), as we document later, 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,[53] 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. 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 θ (a parameter known as the "trade elasticity").[54] Different values of θ 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 θ. While θ 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

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

54. As per equation (5), trade costs affect trade flows with this elasticity (in partial equilibrium, holding fixed exporter factor prices, importer income, and the importer's total price index).

8.28).[55] In Section VI.B and the Online Appendix, we verify the robustness of our results to choosing alternative values for θ within (and even outside of) this range. However, we focus on estimates when setting θ equal to 8.22, which is the value we obtain by drawing on the model's structure to estimate the value of θ that best fits the data in our empirical setting. We obtain this estimate from a nonlinear least squares (NLS) routine, noting that equation (10) is nonlinear in θ.[56] The estimated value of 8.22 has a 95% confidence interval between 3.73 and 26.83 (based on block-bootstrapping at the state-level, with 400 replications).[57]

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

55. In some other examples, Caliendo and Parro (forthcoming) estimate an average θ (across 20 industries) of 8.64 (with θ = 8.11 for agriculture); Costinot, Donaldson, and Komunjer (2012) estimate θ = 6.53; Donaldson (2015) estimates an average θ of 3.80 (across 13 agricultural categories); and Simonovska and Waugh (2013) estimate θ = 4.10. In the Head and Mayer (2014) meta-survey of estimates in the literature, based on those authors' preferred estimation strategy, the mean value of θ is 6.74 and the median value is 5.03.

56. In particular, writing equation (10) at time t we have $\ln q_{ot} = f(\theta)_{ot} + \varepsilon_{ot}$, where $f(\theta)_{ot} \equiv \left(\frac{1+\gamma}{1+\alpha\theta}\right) \ln MA(\theta)_{ot}$; $MA(\theta)_{ot}$ is defined implicitly as the solution to equation (9) at time t and for given θ, as well as given data on county-to-county transportation costs τ_{odt} and county populations N_{ot} in year t; and $\varepsilon_{ot} \equiv \left(\frac{1}{1+\alpha\theta}\right) \ln \left(\frac{A_{ot}}{L_{ot}}\right)$. This is an NLS problem because $f(\theta)_{ot}$ is nonlinear in the parameter θ. As with our main estimation strategy, described in Section IV.E, we assume that ε_{ot} is orthogonal to $f(\theta)_{ot}$ conditional on county fixed effects, state-by-year fixed effects, and cubic polynomials in latitude and longitude interacted with year fixed effects. In practical terms, we execute this NLS problem using a grid search over θ based on an evenly spaced grid with 1,000 points. This procedure requires a value for the land and labor shares in farm production, α and γ, respectively, to compute $f(\theta)_{ot}$; following Caselli and Coleman (2001), we use α = 0.19 and γ = 0.60.

57. Note that the bootstrapped confidence interval is highly skewed to the right, such that the 95% confidence interval extends to higher values not typically considered in the empirical literature.

of –θ rather than –1. Typical estimates of θ are substantially greater than one, including our own estimate of θ = 8.22, but we also report results from assuming a value of θ equal to 1.

IV.E. Main Empirical Specification

Summarizing the foregoing 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 (δ_o), state-by-year fixed effects (δ_{st}), and a cubic polynomial in county latitude and longitude interacted with year effects ($f(x_o, y_o)\delta_t$):

$$\ln V_{ot} = \beta \ln(MA_{ot}) + \delta_o + \delta_{st} + f(x_o, y_o)\delta_t + \epsilon_{ot}. \tag{13}$$

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

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 β 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 β 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 1,000 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.

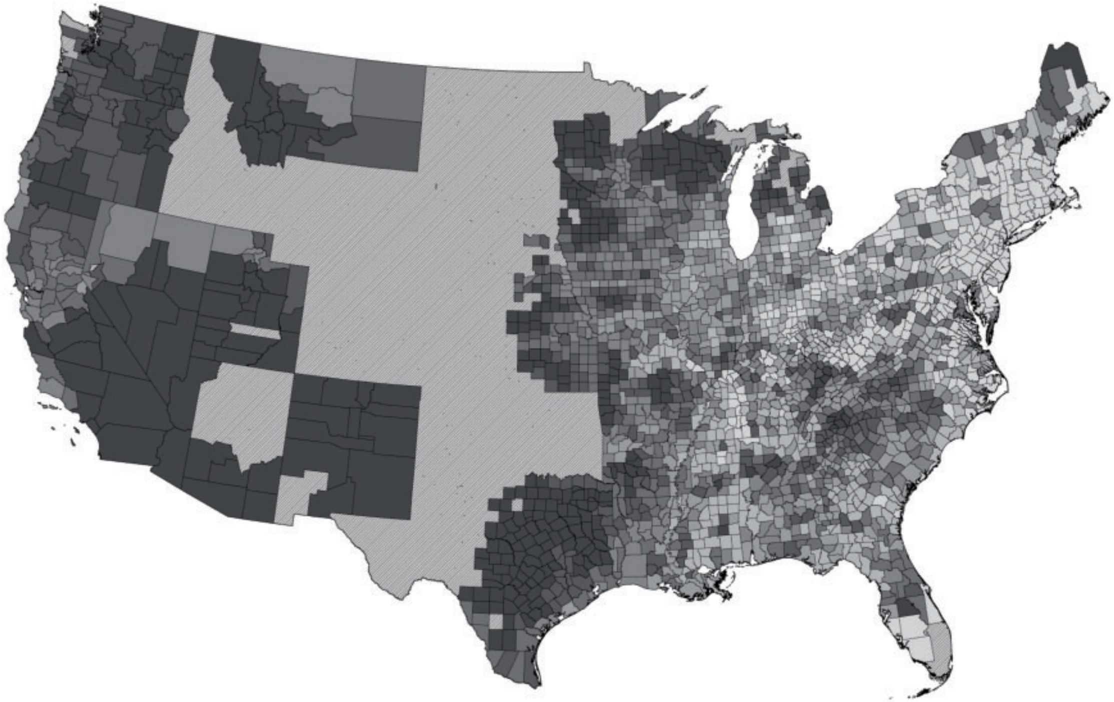

FIGURE III

Calculated Changes in Log Market Access from 1870 to 1890, by County

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. The seven groupings correspond to log changes of: greater than 1.60 (darkest), 1.60 to 1.06, 1.06 to 0.83, 0.83 to 0.73, 0.73 to 0.66, 0.66 to 0.61, and smaller than 0.61 (lightest). nonsample regions are shown hatched.

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 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 already, 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,

62. Counties are separated into seven equal-sized groups.

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.

TABLE I

Estimated Impact of Market Access on Agricultural Land Value

	Log Value of Agricultural Land				
	(1) Baseline Specification	(2) Model-Derived Market Access	(3) Fixed 1870 Population	(4) 100-Mile Buffer Market Access	(5) Unweighted
Log market access	0.511	0.587	0.510	0.487	0.506
	(0.065)	(0.073)	(0.065)	(0.064)	(0.124)
Number of counties	2,327	2,327	2,327	2,327	2,327
R-squared	0.625	0.627	0.625	0.621	0.606

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.

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). Furthermore, weighting by 1870 land value removes the arbitrary distinction between omitting

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.

counties with missing (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.[65] 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 amid 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.[66] 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

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 II

IMPACT OF MARKET ACCESS: ROBUSTNESS TO CONTROLS FOR LOCAL RAILROADS

	Log Value of Agricultural Land				
	(1)	(2)	(3)	(4)	(5)
Log market access	0.511	0.434	0.431	0.343	0.276
	(0.065)	(0.064)	(0.082)	(0.080)	(0.075)
Controls for:					
Any railroad	No	Yes	Yes	Yes	Yes
Railroad length	No	No	Yes	Yes	Yes
Railroads within nearby buffer	No	No	No	Yes	Yes
Railroads within further buffers	No	No	No	No	Yes
Number of counties	2,327	2,327	2,327	2,327	2,327
R-squared	0.625	0.627	0.632	0.640	0.653

Notes. Column (1) reports the estimated impact of market access from the baseline specification (Table I, 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.

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.

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. Furthermore, 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'

TABLE III

IMPACT OF MARKET ACCESS: INSTRUMENTING WITH WATERWAYS

	Change in Log Market Access (1870 to 1890)	Change in Log Value of Agricultural Land (1870 to 1890)		
	(1)	(2)	(3)	(4)
	OLS	OLS	2SLS	OLS
Change in log market access			1.142	0.511
			(0.290)	(0.065)
Log water market access in 1870	−0.096	−0.109		
	(0.030)	(0.024)		
Number of counties	2,327	2,327	2,327	2,327
R-squared	0.568	0.602	0.571	0.625

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.

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)).

Although we find this instrumental variables (IV) 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. Online 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 later.

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: (i) the original regressions are estimated using large changes in market access, and (ii) 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 1 standard deviation (plus or minus), Figure IV shows a

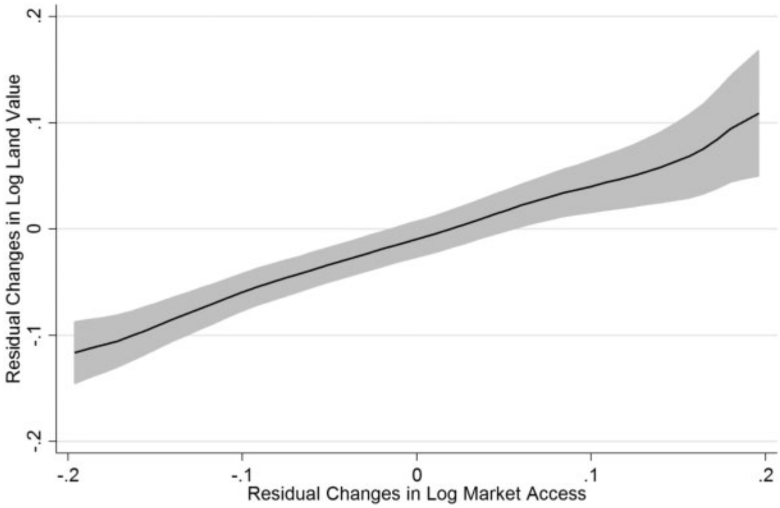

FIGURE IV

Local Polynomial Relationship between Changes in Log Land Value and Log Market Access, 1870 to 1890

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.

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 U.S. 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

67. The local polynomial represents the (default) Epanechnikov kernel with (default) bandwidth 0.06.

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 power to create greater measurement error in counties' market access.[70]

VI.B. Summary of Robustness Checks

We summarize 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.

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 nonsample counties make up 8% of the total counterfactual loss. In the absence of railroads, these nonsample counties experience larger average declines in market access than the regression sample counties (92% versus 78%), but their average land value is lower ($1.0 million versus $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.

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.

We also consider restrictions on which other locations provide counties with valued trading partners. For example, because 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 θ (the "trade elasticity"), which governs how much counties' market access depends on more distant locations. Furthermore, 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 nonagricultural 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 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. Furthermore, 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.[71]

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 draw further on 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.[72] We begin by assuming that

71. This procedure uses population shares observed when the railroad network was substantially less developed (in 1870 and 1850) or nonexistent (in 1830). Notably, many counties receive zero population when there were no overlaying counties in these earlier periods.

72. 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 note 9, 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 $\gamma = 0.60$.

TABLE IV

COUNTERFACTUAL IMPACTS ON LAND VALUE, ALLOWING FOR POPULATION REALLOCATION

	Percent Decline in Land Value without Railroads
Baseline counterfactual without railroads in 1890	60.2 (4.2)
Changes in the distribution of population (holding total population constant)	
1. Assuming the population distribution from 1870	59.1 (4.1)
2. Assuming the population distribution from 1850	59.3 (4.1)
3. Assuming the population distribution from 1830	60.1 (4.0)
4. Assigning the model-predicted counterfactual distribution of population	56.6 (4.0)
Changes in the distribution and total level of population (holding worker utility constant)	
5. Model-predicted estimate, allowing for changes in the level and distribution of population	58.4
Changes in the distribution of population and worker utility (holding total population constant)	
6. Model-predicted estimate, allowing for changes in worker utility and the distribution of population	19.0

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.

the total U.S. population level is held constant, which we denote by $\overline{N}$ ($\equiv \sum_o N_o$). 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 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

A. Counterfactual Changes in Log Population

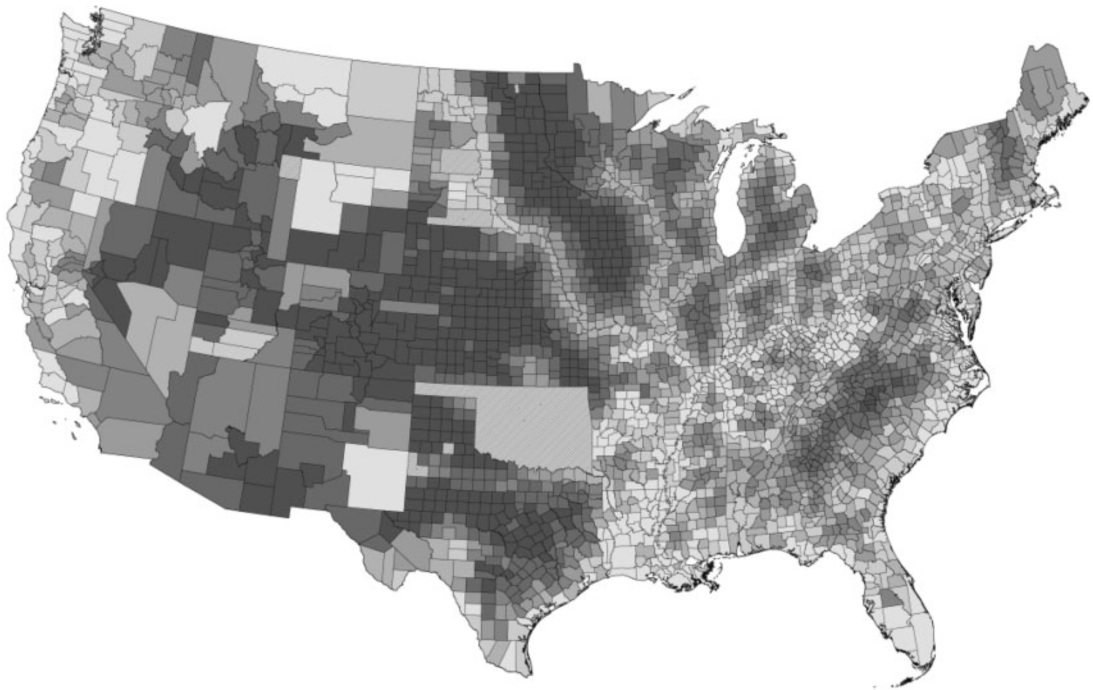

FIGURE V

Changes in Log Population, by County

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: less than −1.31 (darkest), −1.31 to

B. Changes in Log Population from 1870 to 1890

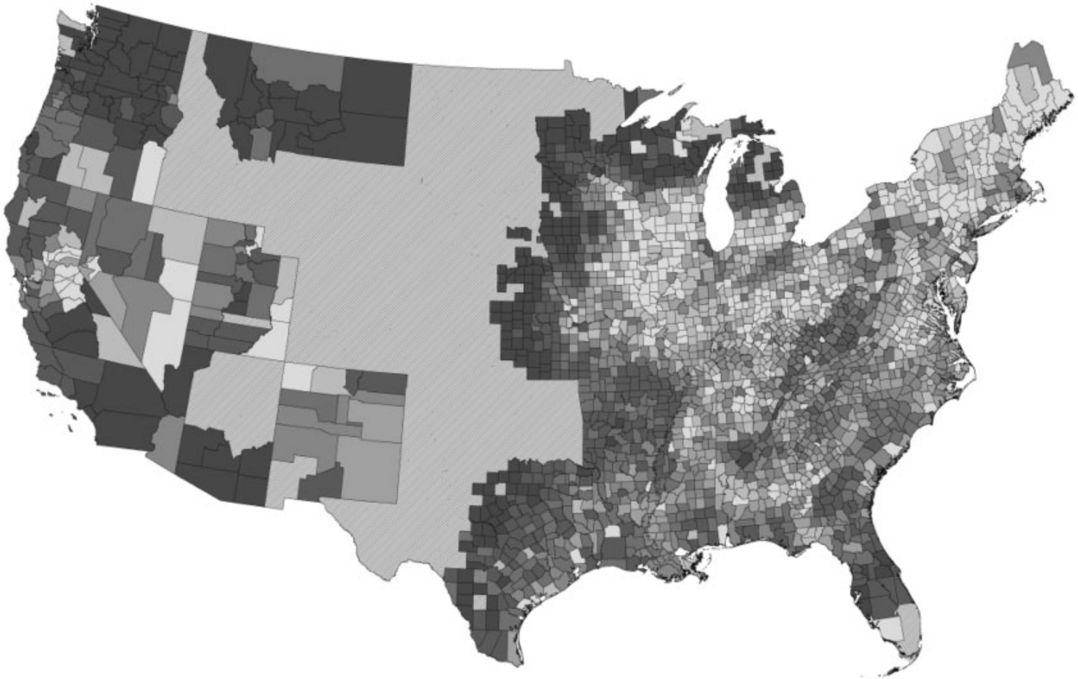

FIGURE V

Continued

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: greater than 1.46 (darkest), 1.46 to 0.81, 0.81 to 0.53, 0.53 to 0.37, 0.37 to 0.24, 0.24 to 0.11, and less than 0.11 (lightest). Nonsample regions are shown hatched.

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 myriad other factors.

Up to this point, our estimates have held fixed the total population $\overline{N}$ 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 $\overline{N}$ were to fall). Conversely, if workers did not leave the United States, so that $\overline{N}$ 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: (i) holding $\bar{U}$ fixed and allowing $\overline{N}$ to decline, and (ii) holding $\overline{N}$ 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 $\overline{N}$ 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%). Although this is surely a large adjustment of U.S. population, recall that our

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.

model contains only one sector for tractability. In a multisector U.S. economy, containing agricultural and manufacturing sectors with tradable goods (and service sectors with nontradable 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 U.S. 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 scenarios imply differing incidence of the economic loss from the removal of railroads. Although the exact change in the aggregate variables ($\bar{U}$ and $\overline{N}$) 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 affect 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 affected prices, wages, and worker utility in the rest of the world.

Against this backdrop, a natural question is which of the two extreme scenarios 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

74. In 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.

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

TABLE V
ESTIMATED IMPACT OF MARKET ACCESS ON POPULATION

	Log Population					
	(1) Baseline Specification	(2) Calibrated Market Access	(3) Fixed 1870 Population	(4) 100-Mile Buffer Market Access	(5) Unweighted	(6) Weighted by 1870 Population
Log market access	0.259	0.314	0.262	0.243	0.348	0.197
	(0.049)	(0.056)	(0.050)	(0.048)	(0.086)	(0.039)
Number of counties	2,327	2,327	2,327	2,327	2,327	2,327
R-squared	0.310	0.316	0.311	0.307	0.523	0.252

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.

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 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 U.S. 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, although 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

76. The 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.

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 U.S. 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 recalculate 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). The proposed 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

TABLE VI

COUNTERFACTUAL IMPACTS ON LAND VALUE, ALLOWING FOR TRANSPORTATION RESPONSES

	Percent Decline in Land Value
Baseline counterfactual without railroads in 1890	60.2 (4.2)
Allowing for transportation responses	
1. Extended canal network	52.4 (4.2)
2. Improved country roads, wagon cost of 14 cents	47.5 (3.9)
3. Extended canal network and improved country roads	40.0 (3.7)
4. Increased water shipping rates, doubled	72.5 (4.2)

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.

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 a 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%.[77]

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.[78] When replacing the railroads with a lower wagon cost of 5 cents or 2.5 cents, 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 cent.

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 trans-shipment 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%

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 cents, 3.5 cents, and 1.4 cents per ton-mile.

greater than the baseline estimated economic impact from removing railroads.

IX. Conclusion

This article develops a new approach for estimating the historical impact of railroads on the U.S. 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. Furthermore, declines in the agricultural sector may have also affected 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 with 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 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. Furthermore, 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 affected the development of the U.S. 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 U.S. economy.

As a broader methodological contribution, this article 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.

STANFORD UNIVERSITY AND NBER
UNIVERSITY OF CHICAGO AND NBER

SUPPLEMENTARY MATERIAL

An Online Appendix for this article can be found at QJE online (qje.oxfordjournals.org).

REFERENCES

Ahuja, Ravindra K., Thomas L. Magnanti, and James B. Orlin, *Network Flows: Theory, Algorithms, and Applications* (New York: Prentice Hall, 1993).

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

Anderson, James E., and Eric van Wincoop, "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, 93 (2003), 170–192.

———, "Trade Costs," *Journal of Economic Literature*, 42 (2004), 691–751.

Anderson, Simon P., André de Palma, and Jacques F. Thisse, *Discrete Choice Theory of Product Differentiation* (Cambridge, MA: MIT Press, 1992).

Anselin, Luc, *Spatial Econometrics: Methods and Models* (Dordrecht: Kluwer Academic, 1988).

Atack, Jeremy, "On the Use of Geographic Information Systems in Economic History: The American Transportation Revolution Revisited," *Journal of Economic History*, 73 (2013), 313–338.

Atack, Jeremy, Fred Bateman, Michael Haines, and Robert A. Margo, "Did Railroads Induce or Follow Economic Growth? Urbanization and Population Growth in the American Midwest, 1850–1860," *Social Science History*, 34 (2010), 171–197.

Atack, Jeremy, Michael Haines, and Robert A. Margo, "Railroads and the Rise of the Factory: Evidence for the United States, 1850–1870," in *Economic Evolution and Revolutions in Historical Time*, P. Rhode, J. Rosenbloom, and D. Weiman, eds. (Palo Alto, CA: Stanford University Press, 2011).

Atack, Jeremy, and Robert A. Margo, "The Impact of Access to Rail Transportation on Agricultural Improvement: The American Midwest as a Test Case, 1850–1860," *Journal of Transport and Land Use*, 4 (2011), 5–18.

Axtell, Robert L., "Zipf Distribution of U.S. Firm Sizes," *Science*, 293 (2001), 1818–1820.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian, "On the Road: Access to Transportation Infrastructure and Economic Growth in China," NBER Working Paper 17897, 2012.

Baum-Snow, Nathaniel, "Did Highways Cause Suburbanization?," *Quarterly Journal of Economics*, 122 (2007), 775–805.

Boyd, J. Hayden, and Gary M. Walton, "The Social Savings from Nineteenth-Century Rail Passenger Services," *Explorations in Economic History*, 9 (1971), 233–254.

Caliendo, Lorenzo, and Parro Fernando, "Estimates of the Trade and Welfare Effects of NAFTA," *Review of Economic Studies*, forthcoming.

Caselli, Francesco, and Wilbur John Coleman, "Cross-Country Technology Diffusion: The Case of Computers," *American Economic Review*, 91 (2001), 328–335.

Cervantes, Fernando Perez, *"Railroads and Economic Growth: A Trade Policy Approach,"* Mimeo, University of Chicago, 2013.
Chaney, Thomas, "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review*, 98 (2008), 1707–1721.
Colton, G. W., *Colton's Railroad Map (Intermediate Size) of the United States Reduced from "Colton's Railroad and Commercial Map of the United States"* (New York: Colton, 1870).
Combes, Pierre-Philippe, Thierry Mayer, and Jacques-François Thisse, *Economic Geography: The Integration of Regions and Nations* (Princeton, NJ: Princeton University Press, 2008).
Conley, Timothy G., "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, 92 (1999), 1–45.
Costinot, Arnaud, Dave Donaldson, and Ivana Komunjer, "What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas," *Review of Economic Studies*, 79 (2012), 581–608.
Cram, George F., *Standard American Atlas of the World* (Chicago: Cram, 1887).
David, Paul A., "Transport Innovation and Economic Growth: Professor Fogel on and off the Rails," *Economic History Review*, 22 (1969), 506–524.
Donaldson, Dave, "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*, forthcoming, 2015.
Eaton, Jonathan, and Samuel Kortum, "Technology, Geography and Trade," *Econometrica*, 70 (2002), 1741–1779.
Eaton, Jonathan, Samuel S. Kortum, and Sebastian Sotelo, "International Trade: Linking Micro and Macro," NBER Working Paper 17864, 2012.
Fishlow, Albert, *American Railroads and the Transformation of the Antebellum Economy* (Cambridge, MA: Harvard University Press, 1965).
Fogel, Robert W., *Railroads and American Economic Growth: Essays in Econometric History* (Baltimore, MD: Johns Hopkins University Press, 1964).
———, "Notes on the Social Saving Controversy," *Journal of Economic History*, 39 (1979), 1–54.
Hacker, J. David, "New Estimates of Census Coverage in the United States, 1850–1930," *Social Science History*, 37 (2013), 71–101.
Haines, Michael R., *Historical, Demographic, Economic, and Social Data: US, 1790–2000* (ICPSR, 2005).
Haines, Michael, and Roberet A. Margo, "Railroads and Local Development: The United States in the 1850s," in *Quantitative Economic History: The Good of Counting*, J. Rosenbloom , ed. (London: Routledge, 2008), 78–99.
Hanson, Gordon H., "Market Potential, Increasing Returns and Geographic Concentration," *Journal of International Economics*, 67 (2005), 1–24.
Harris, Chauncy D., "The Market as a Factor in the Localization of Industry in the United States," *Annals of the Association of American Geographers*, 44 (1954), 315–348.
Head, Keith, and Thierry Mayer, "Gravity, Market Potential and Economic Development," *Journal of Economic Geography*, 11 (2011), 281–294.
———, "Gravity Equations: Workhorse, Toolkit, Cookbook," *Handbook of International Economics*, vol. 4, Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, eds. (New York: Elsevier, 2014).
Herrendorf, Berthold, James A. Schmitz, and Arilton Teixeira, "Transportation and Development: Insights from the U.S., 1840–1860," Federal Reserve Bank of Minneapolis Research Department Staff Report 425, 2009.
Holmes, Thomas J., and James A. Schmitz, "Competition at Work: Railroads vs. Monopoly in the U.S. Shipping Industry," *Federal Reserve Bank of Minneapolis Quarterly Review*, 25 (2001), 3–29.
Hornbeck, Richard, "Barbed Wire: Property Rights and Agricultural Development," *Quarterly Journal of Economics*, 125 (2010), 767–810.
Lebergott, Stanley, "United States Transportation Advance and Externalities," *Journal of Economic History*, 26 (1966), 437–461.
Leunig, Timothy, "Time Is Money: A Re-Assessment of the Passenger Social Savings from Victorian British Railways," *Journal of Economic History*, 66 (2006), 635–673.
———, "Social Savings," *Journal of Economic Surveys*, 24 (2010), 775–800.

McClelland, Peter D., “Railroads, American Growth, and the New Economic History: A Critique,” *Journal of Economic History*, 28 (1968), 102–123.
Melitz, Marc J., “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity,” *Econometrica*, 71 (2003), 1695–1725.
Michaels, Guy, “The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System,” *Review of Economics and Statistics*, 90 (2008), 683–701.
Miguel, Edward, and Michael Kremer, “Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities,” *Econometrica*, 72 (2004), 159–217.
Nerlove, Marc, “Railroads and American Economic Growth,” *Journal of Economic History*, 26 (1966), 107–115.
Redding, Stephen, “The Empirics of New Economic Geography,” *Journal of Regional Science*, 50 (2010), 297–311.
Redding, Stephen, and Daniel M. Sturm, “The Cost of Remoteness, Evidence from German Division and Reunification,” *American Economic Review*, 98 (2007), 1766–1797.
Redding, Stephen, and Anthony Venables, “Economic Geography and International Inequality,” *Journal of International Economics*, 62 (2004), 53–82.
Simonovska, Ina, and Michael E. Waugh, “The Elasticity of Trade: Estimates and Evidence,” *Journal of International Economics*, 92 (2013), 34–50.
Swisher, Scott, “Reassessing Railroads: Accounting for Transport Network Endogeneity,” University of Cambridge, Working Paper, 2014.
White, Colin M., “The Concept of Social Savings in Theory and Practice,” *Economic History Review*, 24 (1976), 82–100.
Whitney, William D., and Benjamin E. Smith, *The Century Dictionary and Cyclopedia: With a New Atlas of the World: A Work of General Reference in all Departments of Knowledge* (Century, 1911).
Williamson, Jeffrey G., *Late Nineteenth-Century American Development: A General Equilibrium History* (New York: Cambridge University Press, 1974).