Online Appendix For: Railroads and American Economic Growth: A 'Market Access' Approach

Dave Donaldson
Stanford University and NBER

Richard Hornbeck
University of Chicago and NBER

October 2015

I Data Appendix

I.A Data Sources

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

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

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

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

I.B Network Database

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

I.C Sample Definition

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

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

I.D Summary Statistics and Geographic Variation

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

II Results Appendix

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

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

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

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

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

[1] The standard errors for the estimates in columns 2 are calculated using the delta method, which transforms the standard errors in column 1.

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

[3] For column 2, because data are not available in 1890 for the total value of real estate and personal property, we use the estimated coefficient from column 1 and the counterfactual decline in market access from our baseline specification.

Our baseline measure of market access includes counties' access to all other counties, based on a model of trade among all US counties. Alternatively, agricultural counties may only benefit from selling/buying goods to/from cities. The estimated impact of market access is similar when only measuring counties' access to urban areas, cities, or only New York City (rows 8-10), which reflects the large influence of urban areas and cities in determining counties' overall market access.[4] The implied economic impact of railroads is somewhat smaller in these specifications, which could reflect railroads' comparative advantage in linking rural areas with each other (relative to the role of waterways).[5]

Calculating our measure of market access requires assuming a value of θ (the "trade elasticity"), for which our baseline measure assumes a value of 8.22. We obtained that value of θ by estimating which value of θ best fit the model to the data, but now we consider the results' robustness to alternative values of theta. Different values of θ will change the estimated impact of market access (column 1), though this largely reflects a mechanical re-scaling of "market access" and the relevant notion of robustness concerns the counterfactual impact on land value (column 2).[6] In row 11, we assume a value of 1 for θ, which makes our measure of market access more similar to an older notion of "market potential" (Harris, 1954). In rows 12 and 16, we assume values for θ of 3.60 and 12.86, which reflect the two extreme estimates in Eaton and Kortum (2002).[7] In rows 13 and 17, we assume values for θ of 3.73 and 26.83, which correspond to the 95% confidence interval around our estimated value of 8.22.[8] In row 14, we assume a value for θ of 3.80, which corresponds to the assumed value in our previous working paper.[9] In row 15, we assume a value for θ of 6.74, which reflects the mean estimate from a meta-survey by Head and Mayer (2014).[10] Overall, these alternative choices for θ mechanically change the estimated impact of market access (column 1), but have less effect on the implied impact of railroads (column 2) because there are also changes in the counterfactual impact on market access.

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

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

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

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

[8]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.

[9]Donaldson (2015) estimates an average θ of 3.80 (across 13 agricultural categories), and Simonovska and Waugh (2013) estimate θ =4.10.

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

II.B Robustness to Transportation Cost Parameters

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

Water transportation is always low cost, but the estimated impact of market access is not sensitive to further lowering the cost of sea routes or all water routes (rows 1 and 2).[11] The aggregate impact of railroads declines somewhat with lower water costs, as a counterfactual without railroads becomes more manageable. Our empirical estimates are less sensitive than social saving estimates, however, which change proportionally with the difference in point-to-point costs by rail and water.

One advantage to railroads was in reducing transshipment charges, incurred whenever transferring goods to/from a railroad car, river boat, canal barge, or ocean liner. The estimates are not sensitive, however, to eliminating transshipment charges within the waterway network (row 3). In addition, transportation through the railroad network was not entirely seamless: congestion, fragmented track ownership, or differences in gauges may have required periodic transshipment; and scheduled freight routes would be less direct than those calculated on the GIS network. We consider a higher railroad rate (0.735 cents) that reflects a need for two transshipment points within an average length railroad route, and an even higher railroad rate (0.878 cents) that makes railroad routes as indirect as wagon routes.[12] The estimated impact of market access is not sensitive to higher railroad rates (rows 4 and 5), whereas the aggregate impact of railroads declines slightly with a decrease in the relative advantage of railroads.

Wagon transportation costs are likely to be an important feature of the database. Fogel emphasized that the baseline wagon rate may be too high and his social saving estimates are substantially lower for decreased wagon costs of 14 cents per mile.[13] By contrast, this lower wagon rate increases the estimated impact of market access such that the aggregate impact of railroads is moderately higher (row 6).

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

[11] The lower waterway rate (0.198 cents per ton-mile) reflects Fogel's preliminary rate for waterway transportation, prior to his adjustments for supplemental costs associated with waterway transportation.

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

[13] To clarify, the cost of 14 cents per mile reflects Fogel's considered lower cost (10 cents per mile) and Fogel's adjustment factor for indirect wagon routes (1.4). It is more convenient for us to scale up the price of routes, rather than equivalently scale up the distance required.

region from the sample (row 7).[14] The Western regions are not central to the empirical analysis, which draws on within-state variation in changes in market access.

The estimated impact of market access is affected little by changes to the waterway route between the Pacific Ocean and the Atlantic Ocean (rows 8-10).[15] This waterway route has little impact on the calculation of market access because the transcontinental railroad was available in 1870 and 1890. By contrast, this waterway route has more influence on the estimated aggregate impact of railroads because it provides the only link between Western and Eastern markets in the absence of the railroads (aside from wagons). The costs of waterway shipping may have been driven down by competition with railway shipping, however, which we explore in later analysis.

II.C Robustness to Alternative Empirical Specifications

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

The baseline specification includes controls for changes by state and latitude/longitude, though we also consider additional controls for region-specific changes. Rows 1 and 2 also include year-specific controls for the fraction of each county in 20 "resource regions" or 145 "resource subregions" (Hornbeck, 2010).[16] Comparing counties within such local regions may exacerbate measurement error in market access, arising from simplifications in our network database and calculated county-to-county freight transportation costs, though the estimated magnitudes remain substantial and statistically significant. Rows 3 and 4 modify the baseline specification's controls for county longitude and latitude, replacing the year-specific third-order polynomials with fifth-order and first-order polynomials.

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

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

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

[16] Note that these regions were mapped in 1966, so they may be endogenous to the availability of railroads.

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

Some counties experience large increases in agricultural land value as new lands are settled between 1870 and 1890. In the theory and the empirical analysis, we consider the total supply of agricultural land fixed by counties' geographic borders and analyze increases in county land value on both the intensive and extensive margins. Our baseline estimates assume that newly settled farmland had zero agricultural value previously, as newly settled land came predominately from the public domain where land had not been desired and so was revealed to have had little economic value. In two robustness checks, we relax this assumption and adjust upward counties' value of land in 1870 by assuming new lands (settled between 1870 and 1890) had some value in 1870: either a small value ($0.0625 per acre) that reflects the cost of acquiring land through the homestead act (row 9), or a larger value ($1.25 per acre) that reflects the cost of acquiring land through preemption (row 10). The estimates are robust to these adjustments, though the estimated magnitude declines meaningfully when assuming unsettled land would have been worth the cost of acquiring land through preemption. We see that estimate as overly conservative, as the absence of settlement in 1870 indicates these lands were not worth the cost of settlement.[17]

In a related exercise, we consider whether the increases in agricultural land value are driven, in part, by increases in the intensity of land improvement. In the baseline analysis, we follow Fogel in deflating the Census-reported value of agricultural land, buildings, and improvements to obtain an estimate of the value of agricultural land only.[18] Fogel's deflator for 1890 may over-deflate land values in 1870, however, for counties with lower land improvement in 1870. We consider two robustness checks that decrease this deflator in 1870

[17] In addition, we only add to counties' 1870 value by truncating the quantity of "newly settled land" at zero, measured by the change in total farmland between 1870 and 1890.

[18] Deflating land values in this manner does not change the estimated impact of market access or the counterfactual decline in agricultural land value, as the state-level deflator is absorbed by state-by-year fixed effects. Deflating does affect the implied annual economic loss, however, which depends on the level of agricultural land value.

for counties that had less land improvement in 1870, which effectively dampens the increase in land value from 1870 to 1890 for counties that experience increases in land improvement. First, in row 11, we do not deflate land values for counties that had no land improvement in 1870, and partially deflate when land improvement was lower in 1870 than in 1890.[19] This is an overly aggressive correction because it assumes there is no value from buildings or other improvements when there is no improved acreage. Second, in row 12, we modify Fogel's state-level deflators in 1890 to derive the comparable county-level deflator in 1870 that depends on counties' acreage of improved land in 1870 and 1890, counties' reported value of land/buildings/improvements in 1870 and 1890, and state-level estimates of the cost of clearing land in 1870 and 1890 (Primack, 1977).[20] The regressions are robust to both of these approaches, and so accounting for land improvement costs does not appear to be quantitatively important for the results.[21]

Finally, we consider whether impacts on agricultural land values reflect impacts of market access on local non-agricultural sectors. In Section IV.D and the Theory Appendix, we discuss how the inclusion of non-agricultural sectors does not meaningfully change our analysis of how market access impacts agricultural land values. We do assume, however, that the supply of land to the agricultural sector is fixed by counties' geographic size. In practice, impacts of market access on non-agricultural sectors might impact agricultural land.[22] We can restrict the analysis to rural areas, however, as urban and manufacturing activity is highly spatially concentrated. In rows 13 and 14, we report the estimates' robustness to dropping from the regression sample those counties with any city population (i.e., population in areas above 25,000 people) and counties with any urban population (i.e., population in areas above 2,500 people).[23] The estimated magnitudes decline moderately, but remain similar.

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

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

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

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

[23]Note that the calculation of rural areas' market access still includes access to cities and other urban areas.

III Theory Appendix

III.A An Extended Model with Multiple Sectors

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

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

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

$$V(\mathbf{P}_o, Y_o) = \frac{Y_o}{(P_o^A)^{\mu}(P_o^M)^{1-\mu}}, \tag{1}$$

with P_o^M being a price index that is given by

$$(P_o^M)^{1-\sigma_M} = \int_0^{n_M} (p_o^M(j))^{1-\sigma_M} dj, \tag{2}$$

and n_M being the exogenous number of manufactured varieties. P_o^A follows the definition in Section IV of the main paper.

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

$$MC_o^M(j) = \frac{q_o^{\alpha_M} w_o^{\gamma_M} r_o^{1-\alpha_M-\beta_M-\gamma_M} (\Phi_o^A)^{\beta_M}}{z_o^M(j)} \quad \text{if } o \in \mathcal{U}, \tag{3}$$

where q_o is the land rental rate, w_o is the wage rate, and r_o is the capital rental rate in that region. The term Φ_o^A is a price index over the set of agricultural varieties that are used in the production of manufacturing goods; we assume that manufacturing production requires a bundle of agricultural varieties, with constant elasticity of transformation ψ_A across these varieties.[24] Similar to the agricultural sector (from Section IV of the main paper), $z_o^M(j)$ is a Hicks-neutral productivity shifter that is exogenous and local to region o and drawn from a Fréchet distribution with CDF given by: $F_o^M(z) = 1 - \exp(-B_o z^{-\theta_M})$, with $\theta_M > 1$. The manufacturing sector is assumed to be perfectly competitive.

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

$$MC_o^A(j) = \frac{q_o^{\alpha_A} w_o^{\gamma_A} r_o^{1-\alpha_A-\beta_A-\gamma_A} (\Phi_o^M)^{\beta_A}}{z_o^A(j)} \quad \text{if } o \in \mathcal{R}, \tag{4}$$

where $z_o^A(j)$ is drawn from the CDF $F_o^A(z) = 1 - \exp(-A_o z^{-\theta_A})$, with $\theta_A > 1$. Analogously to the case of the manufacturing sector, the term Φ_o^M is a CES price index for the set of manufacturing varieties used in agricultural production.[25] As before, the agricultural sector is assumed to be perfectly competitive.

Trade Costs: As in Section IV of the main paper, trading goods is costly. We denote the iceberg cost of shipping a good from region o to region d in the manufacturing sector as τ_{od}^M, and that in the agricultural sector as τ_{od}^A.

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

$$w_o = \bar{U}(P_o^A)^{\mu}(P_o^M)^{1-\mu}. \tag{5}$$

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

[24] This implies that $(\Phi_o^A)^{1-\psi_A} = \int_0^{n_A} (p_o^A(j))^{1-\psi_A} dj$.

[25] We denote by ψ_M the elasticity of substitution, such that $(\Phi_o^M)^{1-\psi_M} = \int_0^{n_M} (p_o^M(j))^{1-\psi_M} dj$.

<u>**Solving for Prices and Trade Flows:**</u> We follow similar derivations to those in Section IV of the main paper, this time separately for the manufacturing and agricultural sectors. We begin with the manufacturing sector. The consumer price index for manufactured goods in any region d is given by:

$$(6) \qquad (P_d^M)^{-\theta_M} = \chi_1 \sum_{o \in \mathcal{U}} B_o (q_o^{\alpha_M} w_o^{\gamma_M})^{-\theta_M} (IMA_o^A)^{\beta_M} (\tau_{od}^M)^{-\theta_M} \equiv CMA_o^M \quad \text{if } d \in \mathcal{U}, \mathcal{R},$$

where we denote by CMA_o^M the "consumer market access (manufacturing)" following the discussion in Section IV of the main paper.[26] We also refer to IMA_o^A — defined formally below — as "input market access (agriculture)" because it captures how cheaply manufacturing firms can access intermediate inputs (i.e., agricultural goods). The price index among agricultural firms, which captures the cost of their ideal bundle of manufacturing inputs, is similar:

$$(7) \qquad (\Phi_d^M)^{-\theta_M} = \chi_2 \sum_{o \in \mathcal{U}} B_o (q_o^{\alpha_M} w_o^{\gamma_M})^{-\theta_M} (IMA_o^A)^{\beta_M} (\tau_{od}^M)^{-\theta_M} \equiv IMA_o^M \quad \text{if } d \in \mathcal{R},$$

where IMA_o^M as "input market access (manufacturing)" because it captures how cheaply agricultural firms can access inputs made by manufacturing firms.[27] Note that CMA_d^M and IMA_d^M are very closely related but not identical: they are defined for different sets of regions d and in general $\chi_1 \neq \chi_2$.

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

$$(8) \quad X_{od}^M = \chi_1 B_o (q_o^{\alpha_M} w_o^{\gamma_M})^{-\theta_M} (IMA_o^A)^{\beta_M} (\tau_{od}^M)^{-\theta_M} (CMA_d^M)^{-1} X_d^M \quad \text{if } o \in \mathcal{U} \text{ and } d \in \mathcal{U}, \mathcal{R},$$

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

$$(9) \qquad \widetilde{X}_{od}^M = \chi_2 B_o (q_o^{\alpha_M} w_o^{\gamma_M})^{-\theta_M} (IMA_o^A)^{\beta_M} (\tau_{od}^M)^{-\theta_M} (IMA_d^M)^{-1} \widetilde{X}_d^M \quad \text{if } o \in \mathcal{U} \text{ and } d \in \mathcal{R},$$

where $\widetilde{X}_d^M$ denotes the expenditure on manufacturing goods among agricultural firms in

[26]Here, $\chi_1 \equiv [\Gamma(\frac{\theta_M + 1 - \sigma_M}{\theta_M})]^{-\theta_M/(1-\sigma_M)} r^{-(1-\alpha_M-\beta_M-\gamma_M)\theta_M}$, where $\Gamma(.)$ is the Γ function.

[27]Here, $\chi_2 \equiv [\Gamma(\frac{\theta_M + 1 - \psi_M}{\theta_M})]^{-\theta_M/(1-\psi_M)} r^{-(1-\alpha_M-\beta_M-\gamma_M)\theta_M}$.

region d.

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

$$(10) \qquad (P_d^A)^{-\theta_A} = \chi_3 \sum_{o\in\mathcal{R}} A_o (q_o^{\alpha_A} w_o^{\gamma_A})^{-\theta_A} (IMA_o^M)^{\beta_A} (\tau_{od}^A)^{-\theta_A} \equiv CMA_o^A \quad \text{if } d \in \mathcal{U}, \mathcal{R},$$

where we denote by CMA_o^A the "consumer market access (agriculture)" following our above definition of CMA_o^M.[28] The price index among manufacturing firms, which captures the cost of their ideal bundle of agricultural inputs, is similar:

$$(11) \qquad (\Phi_d^A)^{-\theta_A} = \chi_4 \sum_{o\in\mathcal{R}} A_o (q_o^{\alpha_A} w_o^{\gamma_A})^{-\theta_A} (IMA_o^M)^{\beta_A} (\tau_{od}^A)^{-\theta_A} \equiv IMA_o^A \quad \text{if } d \in \mathcal{U},$$

where we refer to IMA_o^A as "input market access (agriculture)" because it captures how cheaply manufacturing firms can access intermediate inputs (i.e., agricultural goods).[29]

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

$$(12) \quad X_{od}^A = \chi_3 A_o (q_o^{\alpha_A} w_o^{\gamma_A})^{-\theta_A} (IMA_o^M)^{\beta_A} (\tau_{od}^A)^{-\theta_A} (CMA_d^A)^{-1} X_d^A \quad \text{if } o \in \mathcal{R} \text{ and } d \in \mathcal{U}, \mathcal{R},$$

whereas agricultural trade to manufacturing firms follows

$$(13) \qquad \widetilde{X}_{od}^A = \chi_4 A_o (q_o^{\alpha_A} w_o^{\gamma_A})^{-\theta_A} (IMA_o^M)^{\beta_A} (\tau_{od}^A)^{-\theta_A} (IMA_d^A)^{-1} \widetilde{X}_d^A \quad \text{if } o \in \mathcal{R} \text{ and } d \in \mathcal{U},$$

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

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

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

[28] Here, $\chi_3 \equiv [\Gamma(\frac{\theta_A+1-\sigma_A}{\theta_A})]^{-\theta_A/(1-\sigma_A)} r^{-(1-\alpha_A-\beta_A-\gamma_A)\theta_A}$.

[29] Here, $\chi_4 \equiv [\Gamma(\frac{\theta_A+1-\psi_A}{\theta_A})]^{-\theta_A/(1-\psi_A)} r^{-(1-\alpha_A-\beta_A-\gamma_A)\theta_A}$.

where

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

and, analogously to Section IV of the main paper, FMA_o^A refers to "firm market access (agriculture)".[30] Equation (14) takes a similar form to equation (10) in Section IV of the main paper, which was the main log-linear equation relating agricultural land values to market access that drives our empirical analysis. But here, in equation (14), there are four different sources of market access that are potentially distinct from one another (even when trade costs are symmetric such that, in the one-sector model in Section IV of the main paper, CMA and FMA were proportional to one another and simplified to one term). That is, agricultural land values in region o are relatively high if consumers in region o have good access to agricultural goods (high CMA_o^A) or manufactured goods (high CMA_o^M), if agricultural firms in region o have good access to intermediate manufactured inputs (high IMA_o^M), or if agricultural firms in region o have good access to consumers and firms to whom they can sell agricultural output (high FMA_o^A).

The main insight from Section IV of the main paper — that the price of immobile land is higher in regions that have higher market access — continues to hold in even this substantially more complicated economic environment. There are now several notions of market access, which differ from each other in nuanced ways, but their basic functional relationships are very similar. In principle, it is possible to create different empirical measures for these separate market access terms, following steps analogous to those in Section IV of the main paper; in practice, however, these terms are so similar in our empirical setting that empirical proxies for these terms ($\ln CMA_o^A$, $\ln CMA_o^M$, $\ln IMA_o^M$, and $\ln FMA_o^A$) have joint correlations that exceed 0.999.[31]

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

[30] Here $\chi_5 \equiv \left(\frac{1}{1+\alpha_A\theta_A}\right)\ln\alpha - \left(\frac{\gamma_A\theta_A}{1+\alpha_A\theta_A}\right)\ln\bar{U}$. This derivation applies goods market clearing, noting that $Y_o^A = \sum_{d\in\mathcal{R},\mathcal{U}} X_{od}^A + \sum_{d\in\mathcal{U}} \widetilde{X}_{od}^A$.

[31] In particular, we created first-order empirical approximations to these terms that are analogous to our definition of market access in equation (12) of the main paper.

III.B Procedure for Counterfactual Simulations

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

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

$$P_o^{-\theta} = \sum_d \frac{\tau_{od}^{-\theta} P_d N_d}{\sum_i \tau_{di}^{-\theta} P_i^{1+\theta} N_i}. \tag{15}$$

Given data on trade costs τ_{od} from the factual 1890 scenario, population levels N_o from 1890 and a value for the parameter θ, we solve this system of equations for the unique (up to scale) price index P_o in each region in 1890.[32]

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

$$P_o^{-\theta} \bar{U}^{\theta(\alpha+\gamma)} = \xi_1 f(P)^{-\theta(1-\alpha-\gamma)} \sum_d C_d \tau_{od}^{-\theta} P_d^{-\theta(\alpha+\gamma)} N_d^{-\theta\alpha} \tag{16}$$

where $C_d \equiv L_d^{\alpha\theta} T_d$ and our assumption about the internationally mobility of capital amounts to assuming that $r/f(P) = \bar{r}$, where $\bar{r}$ is the exogenous foreign real interest rate and $f(P)$ is the price index of the goods that are implicitly traded for foreign capital.[33] Having solved for relative prices P_o in the factual 1890 equilibrium, from the solution to equation (15), we

[32] This is equivalent to the procedure described in footnote 57 of the main paper, which we use to construct a value of $MA_o(\theta)$, for any candidate value of θ, as an input into our NLS routine for the purposes of estimating θ. The results in Allen and Arkolakis (2014) imply that there is a unique (up to scale) solution of equation (15).

[33] Here, $\xi_1 \equiv \left(\left[\Gamma(\frac{\theta+1-\sigma}{\theta})\right]^{-\frac{\theta}{1-\sigma}} \bar{r}^{-\theta(1-\alpha-\gamma)} \right) \left(\frac{\alpha}{\theta}\right)^{-\alpha\theta}$ and $C_i = L_i^{\alpha\theta} T_i^{(g)}$. As discussed in footnote 35 of the main paper, our baseline case is that in which foreign capital is traded for goods in New York City (the major international trading exchange and financial center at the time).

use this solution in equation (16), along with data on τ_{od} and N_d from 1890, to solve for the vector of C_d terms (up to scale, which amounts to using the normalization that $\bar{U} = 1$ in the factual 1890 equilibrium).

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

References

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

Bureau of Statistics. 2003. *Statistical Abstract of the United States 1892.* Ross Publishing Incorporated. http://www2.census.gov/prod2/statcomp/documents/1892-02.pdf.

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

Carter, Susan B., Scott S. Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright. 2006. *Historical Statistics of the United States: Millennial Edition.* Cambridge UK:Cambridge University Press.

Costinot, Arnaud, Dave Donaldson, and Ivana Komunjer. 2012. "What Goods do Countries Trade? A Quantitative Exploration of Ricardo's Ideas." *Review of Economic Studies*, 79: 581–608.

Donaldson, Dave. 2015. "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure." *American Economic Review*, forthcoming.

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

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

Hacker, J. David. 2013. "New Estimates of Census Coverage in the United States, 1850-1930." *Social Science History*, 37(1): 71–101.

Haines, Michael R. 2005. *Historical, Demographic, Economic, and Social Data: US, 1790-2000.* ICPSR.

Harris, Chauncy D. 1954. "The Market as a Factor in the Localization of Industry in the United States." *Annals of the Association of American Geographers*, 64: 315–348.

Head, Keith, and Thierry Mayer. 2014. "Gravity Equations: Workhorse, Toolkit, Cookbook." *Handbook of International Economics Vol. 4*, , ed. Gita Gopinath, Elhanan Helpman and Kenneth Rogoff. New York: Elsevier.

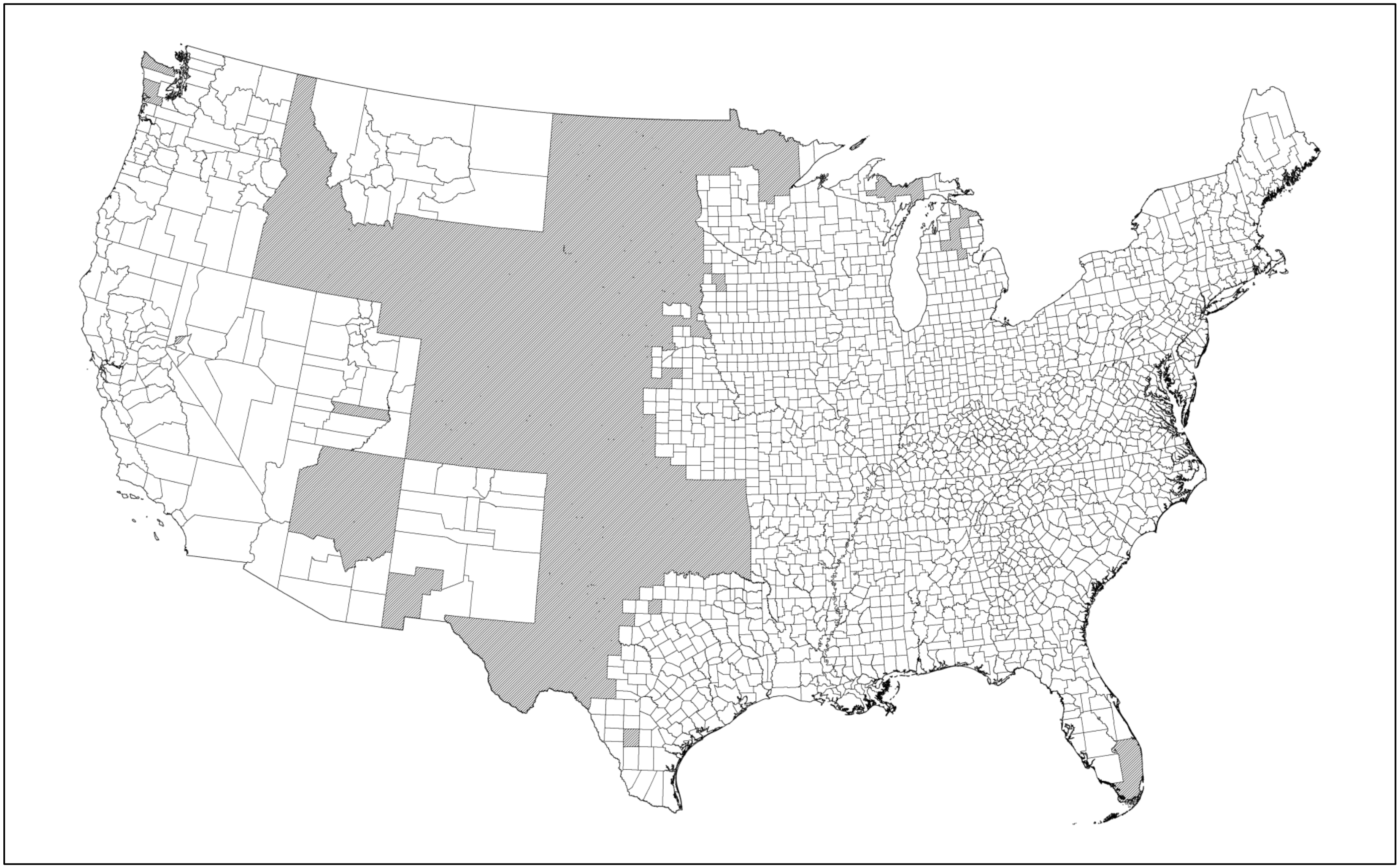
Hornbeck, Richard. 2010. "Barbed Wire: Property Rights and Agricultural Development." *Quarterly Journal of Economics*, 125: 767–810.

Minnesota Population Center. 2011. *National Historical Geographic Information System: Version 2.0.* Minneapolis, MN: University of Minnesota.

Primack, Martin L. 1977. *Farm Formed Capital in American Agriculture 1850 to 1910.* New York: Arno Press.

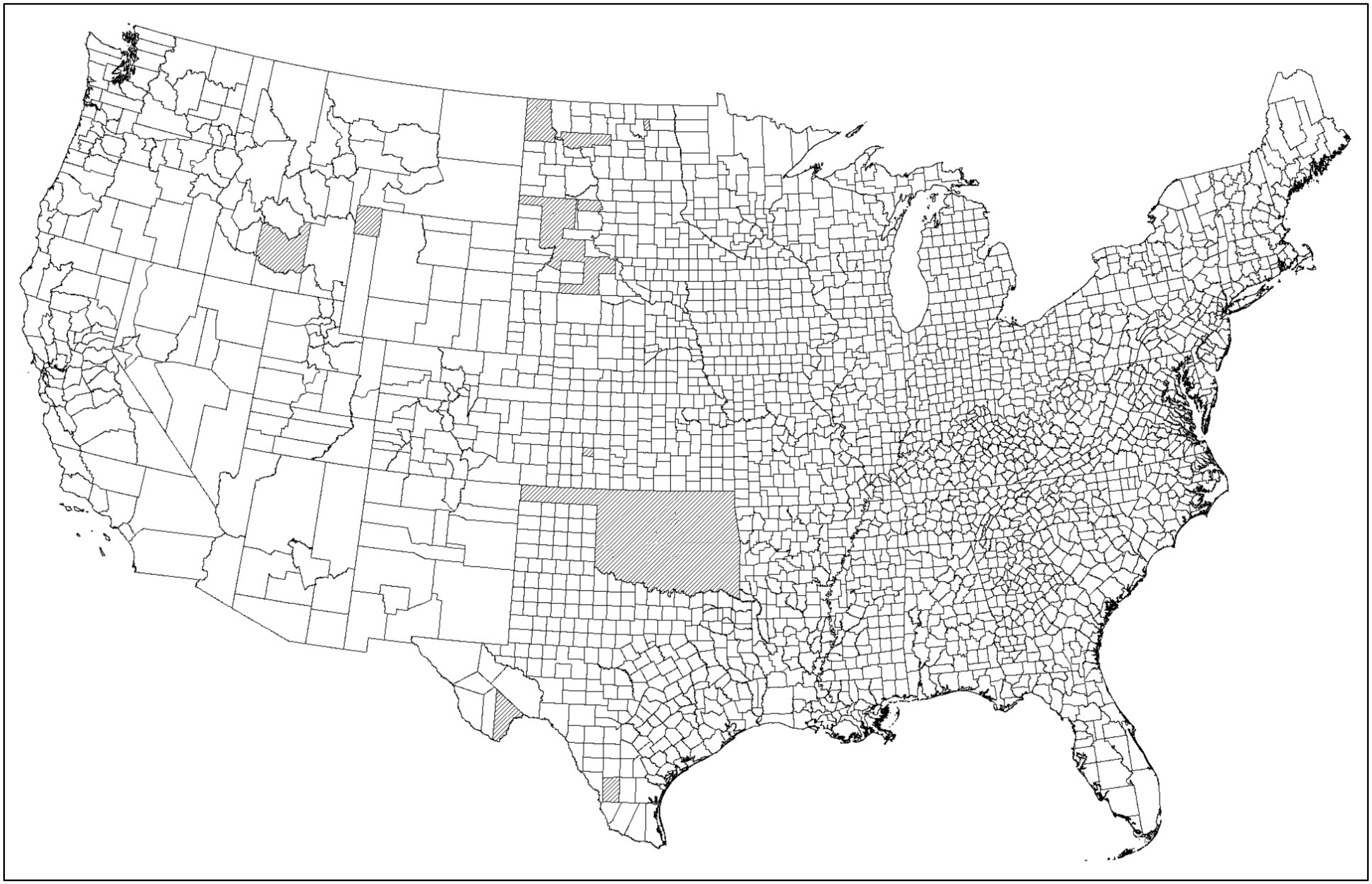
Simonovska, Ina, and Michael E. Waugh. 2013. "The Elasticity of Trade: Estimates and Evidence." *Journal of International Economics*, 92(1): 34–50.

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



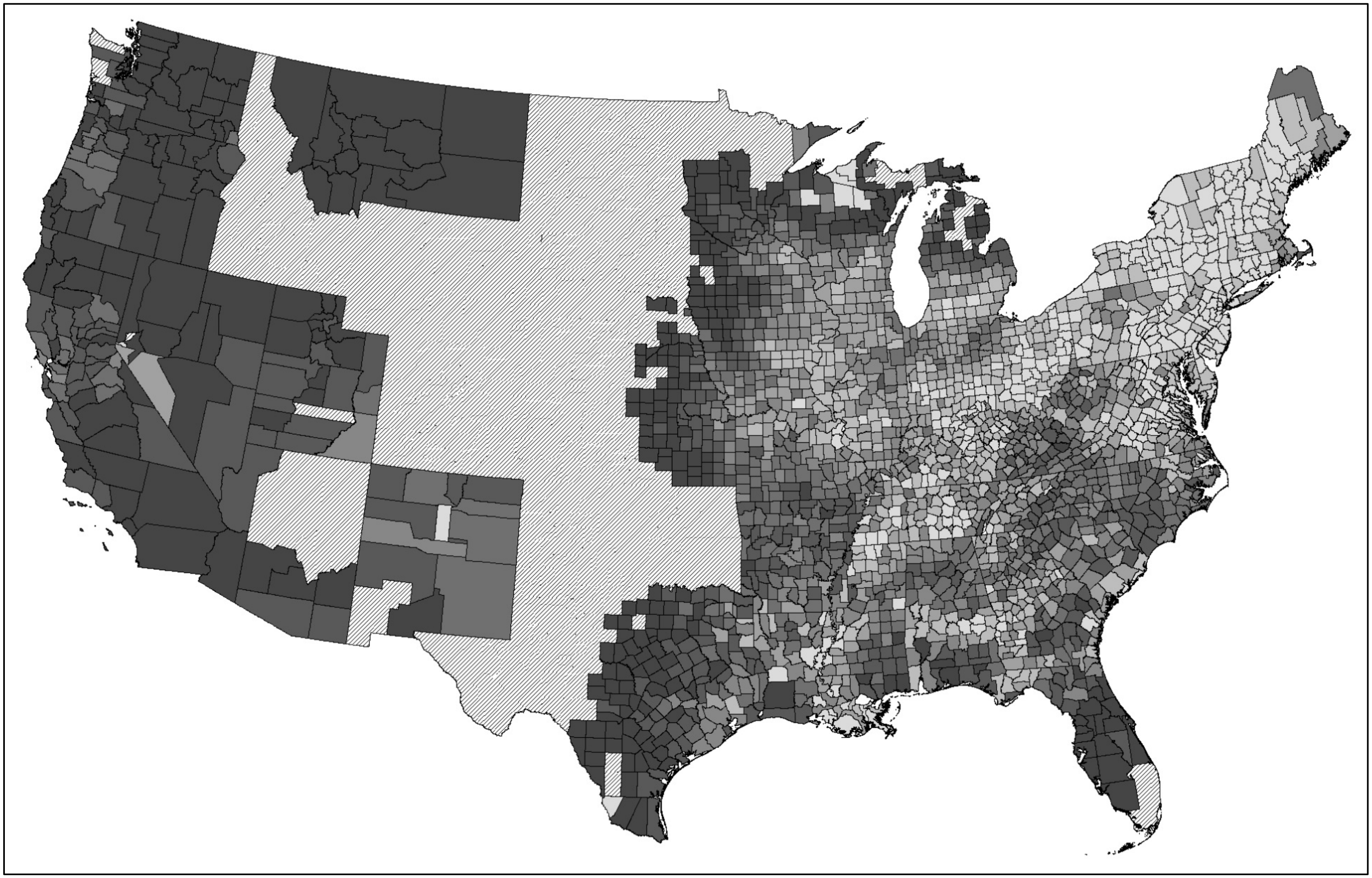
Notes: This map shows the 2,327 sample counties in the regression analysis, which are all counties that report non-zero land values in 1870 and 1890. The excluded geographic areas are hatched. County boundaries correspond to county boundaries in 1890.

Appendix Figure 2. Sample of 2,782 Counties in the Counterfactual Analysis



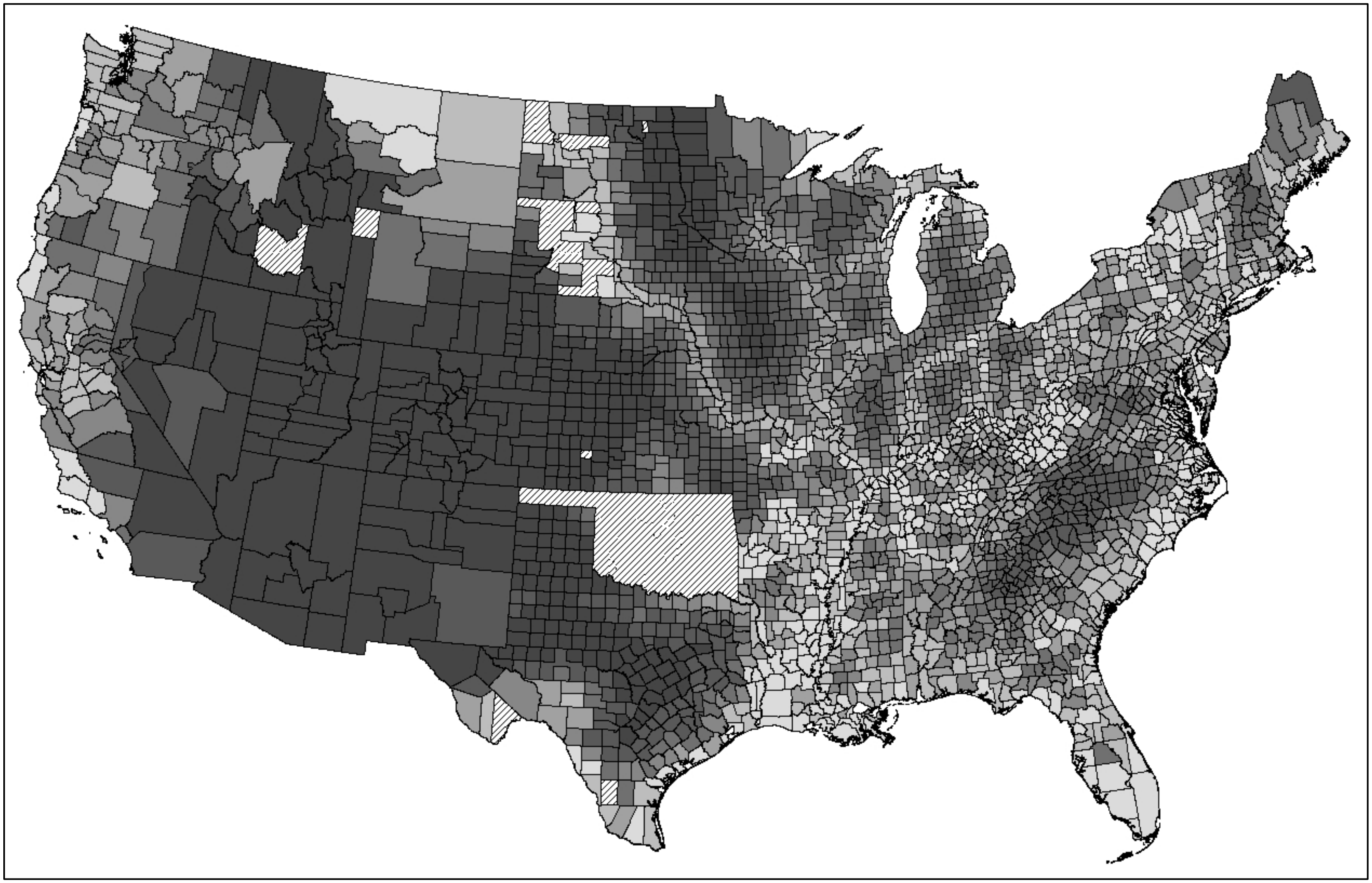
Notes: This map shows the 2,782 sample counties in the counterfactual analysis. Compared to Appendix Figure 1, the additional 455 counties are those that report land value data in 1890 only. The excluded geographic areas are hatched. County boundaries correspond to county boundaries in 1890.

Appendix Figure 3. Changes in Log Land Value from 1870 to 1890, by County



Notes: This map shows the 2,327 sample counties, shaded according to their change in log land value from 1870 to 1890. Counties are divided into seven groups (with an equal number of counties per group), and darker shades denote larger increases in land value. The seven groupings correspond to log changes of: greater than 2.12 (darkest), 2.12 to 1.20, 1.20 to 0.73, 0.73 to 0.42, 0.42 to 0.17, 0.17 to -0.09, less than -0.09 (lightest).

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



Notes: This map shows the 2,782 counterfactual sample counties, shaded according to their change in log market access from 1890 to the baseline counterfactual scenario. Counties are divided into seven groups (with an equal number of counties per group), and darker shades denote larger declines in market access. The seven groupings correspond to log changes of: less than -3.94 (darkest), -3.94 to -2.69, -2.63 to -1.97, -1.97 to -1.49, -1.49 to -1.18, -1.18 to -0.97, greater than -0.97 (lightest).

Appendix Table 1. Transportation Network Database Components

ID	Component Name	Component Definition	Construction Description	Baseline Cost (in Dollars)
0	Navigable Rivers	Fogel's definition of navigable rivers; time invariant component of network	Hand-traced from Fogel (1964)	0.0049 * [Length]
1	Constructed Canals	Fogel's definition of navigable canals; time invariant component of network	Hand-traced from Fogel (1964)	0.0049 * [Length]
2	Proposed Canals	Fogel's definition of proposed canals; time invariant component of network; only included in alternative counterfactual scenarios	Hand-traced from Fogel (1964)	0.0049 * [Length]
4	Sea/Lake Routes	Multiple point-to-point connections throughout the Great Lakes and Oceans; time invariant component of network	Created manually to effectively saturate area	0.0049 * [Length]
5	Railroad Harbor	Points where transshipment to Sea/Lake Routes is considered possible, created wherever the 1911 railroad network approaches the coastline; time invariant component of the network	Created manually as a short line from a Sea/Lake route	0.5
6	Railroads	Railroad lines as depicted on maps from 1870 (Colton 1870) and 1887 (Cram 1887); time variant component of the network.	Hand-traced from a railroad map in 1911 (Whitney and Smith 1911), which was most accurately geo-referenced to county borders. Maps for 1890 and 1870 were then created by manually deleting railroad lines that do not appear in the earlier periods.	0.0063 * [Length]
8	Wagon Routes (Centroid-to-Centroid)	Wagon Routes connecting any two centroids within a distance of 300km; time invariant component of network	Created automatically in ArcGIS	0.231 * [Length]
9	Sea Route Between Coasts	Direct sea route connecting the West Coast (near San Diego) to the East Coast (in the Gulf of Mexico); time invariant component of network	Created manually	8
130	In-county centroid-to-railroad connection	Represents the average wagon route from any point in the county to railroad lines that pass through the county; a transshipment cost is then incurred; time variant component of network	Created manually to connect a centroid to railroads within the county	0.5 + 0.231 * [Mean_Length]

140	Out-of-county centroid-to-railroad connection	Represents the average wagon route from any point in the county to relevant railroad lines outside the county border in various directions; a transshipment cost is then incurred; time invariant component of network	Created manually to connect a centroid to potentially-relevant railroads outside the county	0.5 + 0.231 * [Length]
15	Out-of-county centroid-to-harbor connection	Represents the average wagon route from any point in the county to Sea/Lake harbors (ID "5") outside the county in various directions; a transshipment cost is then incurred; time invariant component of network	Created manually to connect a centroid to potentially-relevant harbors outside the county	0.5 + 0.231 * [Length]
60	River Harbor	Points where transshipment to Sea/Lake Routes is considered possible, created wherever rivers, canals, or proposed canals meet the coastline; time invariant component of the network	Created manually as a short line from a Sea/Lake route	0.5
61	Canal Harbor			
62	Proposed Canal Harbor			
601	River-to-Canal Transshipment Point	Points where rivers and canals meet and transshipment is possible; time invariant component of the network	Created manually as a short line to connect the two modes of transportation	0.5
602	River-to-Proposed Canal Transshipment Point	Points where rivers and proposed canals meet and transshipment is possible; time invariant component of the network		
612	Canal-to-Proposed Canal Transshipment Point	Points where canals and proposed canals meet and transshipment is not necessary; time invariant component of the network		0.0049 * [Length]
700	Transshipment point between railroad and river, canal, or proposed canal	Points where inland waterways (rivers, canals, proposed canals) meet railroads and transshipment is possible; time invariant component of the network		0.5

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

Notes: For the above formulas, [Length] data measures each segment's length in meters and needs to be converted to miles by dividing by 1609.344. To calculate the respective costs, we create the field "length" for each element of the network, assign to it the length in miles (using "Calculate Geometry"), and then create a field "Cost" to which we assign a value based on the indicated formula (using "Field Calculator"). Along with ID "6", there is another railroad line with ID "7" that connects Sarnia and Buffalo based on Stratford's railway industry map.

In the above formulas, [Mean_Length] is again measured in miles and is our estimate of the average distance of a point in the county to the respective waterway or railroad line that passes through the county. In particular, we created 200 random points within each county (using the "Create Random Points" tool). We then calculate the minimum distance from each of those points to the respective mode of transportation (using the "Near" tool. We then collapse the 200 observations into 1 with the mean distance (using the "Dissolve" tool) and merge this data to each county's in-county connections to that mode of transportation (using as merge field the county's unique identifier in the "Join Field" tool). By contrast, when a waterway or railroad line passes outside a county, the measured distance from the centroid is a sufficient approximation of the average distance from points in the county.

Appendix Table 2. Summary Statistics For Main Regression Sample of 2,327 Counties

	Number of Counties	Mean in 1870	Mean in 1890	Log Change from 1870 to 1890
	(1)	(2)	(3)	(4)
Market Access	2,327	7,804,585	15,804,882	0.750
		[2,246,742]	[3,220,369]	[0.288]
Land Value	2,327	7,728,604	8,337,710	0.157
		[5,847,713]	[5,902,241]	[0.532]
Farmland Acres	2,327	288,202	321,174	0.121
		[134,718]	[141,237]	[0.297]
Population	2,327	41,296	60,096	0.278
		[61,841]	[117,292]	[0.313]

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

Appendix Table 3. Robustness to Changes in the Definition of Market Access

	Estimated Impact of Market Access	Percent Decline in Land Value Without Railroads
	(1)	(2)
Baseline Specification	0.511 (0.065)	60.2 (4.2)
Changes in Definition of Market Access		
1. Include access to international markets	0.511 (0.065)	59.7 (4.1)
2. Adjustment for Census undercounting	0.511 (0.065)	60.2 (4.2)
3. Measure access to county wealth	0.503 (0.065)	59.6 (4.2)
4. Include access to own market	0.523 (0.067)	60.4 (4.2)
5. Limit access to counties beyond 5 miles	0.503 (0.065)	60.1 (4.2)
6. Limit access to counties beyond 50 miles	0.491 (0.064)	59.7 (4.2)
7. Limit access to counties beyond 200 miles	0.491 (0.066)	60.4 (4.4)
8. Limit access to only urban areas	0.496 (0.063)	56.4 (4.0)
9. Limit access to only cities	0.484 (0.062)	53.8 (3.9)
10. Limit access to only New York City	0.500 (0.064)	49.1 (3.6)
11. Set parameter "theta" equal to 1	4.19 (0.54)	66.5 (4.3)
12. Set parameter "theta" equal to 3.60	1.17 (0.15)	63.7 (4.3)
13. Set parameter "theta" equal to 3.73	1.13 (0.15)	63.6 (4.2)
14. Set parameter "theta" equal to 3.80	1.11 (0.14)	63.5 (4.2)
15. Set parameter "theta" equal to 6.74	0.623 (0.080)	61.2 (4.2)
16. Set parameter "theta" equal to 12.86	0.331 (0.041)	57.3 (4.0)
17. Set parameter "theta" equal to 26.83	0.172 (0.021)	49.8 (3.9)

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

Appendix Table 4. Robustness to Changes in the Transportation Cost Parameters

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

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

Appendix Table 5. Robustness to Alternative Empirical Specifications

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

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