

NBER WORKING PAPER SERIES

AGRICULTURAL IMPROVEMENTS AND ACCESS TO RAIL TRANSPORTATION:
THE AMERICAN MIDWEST AS A TEST CASE, 1850-1860

Jeremy Atack
Robert A. Margo

Working Paper 15520
<http://www.nber.org/papers/w15520>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2009

We are grateful to David Levinson and workshop participants at Rutgers and Oberlin for helpful comments. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2009 by Jeremy Atack and Robert A. Margo. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Agricultural Improvements and Access to Rail Transportation: The American Midwest as a Test Case, 1850-1860

Jeremy Atack and Robert A. Margo

NBER Working Paper No. 15520

November 2009

JEL No. N51,N71,N91

ABSTRACT

During the 1850s, land in U.S. farms surged by more than 100 million acres while almost 50 million acres of land were transformed from their raw, natural state into productive farmland. The time and expense of transforming this land into a productive resource represented a significant fraction of domestic capital formation at the time and was an important contributor to American economic growth. Even more impressive, however, was the fact that almost half of these total net additions to cropland occurred in just seven Midwestern states which comprised barely less than one-eighth of the land area of the country at that time. Using a new GIS-based transportation database linked to county-level census, we estimate that at least a quarter (and possibly two-thirds or more) of this increase can be linked directly to the coming of the railroad to the region. Farmers responded to the shrinking transportation wedge and rising revenue productivity by rapidly expanding the area under cultivation and these changes, in turn, drove rising farm and land values.

Jeremy Atack

Department of Economics

Vanderbilt University

VU Station B #351819

2301 Vanderbilt Place

Nashville, TN 37235-1819

and NBER

jeremy.atack@vanderbilt.edu

Robert A. Margo

Department of Economics

Boston University

270 Bay State Road

Boston, MA 02215

and NBER

margora@bu.edu

Agricultural Improvements and Access to Rail Transportation: The American Midwest as a Test Case, 1850-1860

Jeremy Atack and Robert A. Margo

Introduction

Capital formation played an especially prominent role in nineteenth century American economic growth where it accounted for approximately one third of the growth of per capita GNP from 1800 to 1840 and 46 percent from 1840 to 1900 (Gallman 2000). Although some of this growth in the capital stock came from foreign sources, much was created through domestic savings and investment. For example, the best estimates put the domestic savings rate in the 1850s at between 17 and 20 percent depending on one's assumptions about prices and home manufacturing (Gallman 2000). A considerable portion of this domestic investment took the form of land clearing and associated improvements to make the land suitable for agriculture. For example, according to contemporary estimates by Ezra Seaman, the cost of clearing and fencing land averaged about \$12 per acre around 1840 declining to about \$11 per acre by 1850 as forested land in the east gave way to open prairie in the (mid)west (Seaman 1852). Unfortunately, we do not know about land in farms or improved acreage earlier on since these data were not collected as a part of the first (partial) census of agriculture in 1840. However, we do know that during the 1840s, despite the bursting of the 1830s speculative land bubble, the federal government sold off over 19 million acres—about 30,000 square miles—of public land (Carter 2006: Table Cf79). Much of this land went (eventually) to farmers and, by 1850, the Census reported over 293 million acres of land in farms, about 38.5% of which had been improved (United States. Census Office and De Bow 1854). Public land sales surged during the 1850s to more than 52 million acres—an area 80 percent larger than the state of Ohio or 40 percent larger than Illinois (Carter, Gartner et al. 2006: Table Cf79). Although it took a while before this new land was cleared and entered agricultural production, improved acreage grew

by almost 50 million acres during the decade and land in farms expanded by well over 100 million acres so that, by 1860, a little over 40 percent of all land in farms had been improved (United States. Census Office. and Kennedy 1864). According to Gallman (2000), such additions to the stock of improved acreage may have accounted for more than half of all capital formation before 1820 but even as late as 1860, fully 22 percent of annual capital formation comprised additions to the stock of improved acreage.

The quantitative significance of these improvements to farm land has been widely recognized (Primack 1962; Primack 1977; Atack, Bateman et al. 2000) but the relative importance of the causal factors that lay behind such investments are less clear. In this paper we use newly collected archival data to assess the role of one specific factor – falling costs of internal transportation as captured by the diffusion of the railroad – in explaining increases in the growth in the share of improved acreage in the Midwest. We focus on the decade of the 1850s since this was when most of the diffusion of the railroad occurred in the region. It is also a period of substantial growth in farm acreage in the region in both relative and absolute terms.

The new data in this paper consist of county-level information on the spread of the railroad network in seven Midwestern states (Illinois, Indiana, Iowa, Michigan, Missouri, Ohio and Wisconsin) during the 1850s which have been linked to county level data on farm outcomes at the benchmark census dates. The transportation data, collected as part of a larger project, derive from GIS (geographic information systems) processing of digitized nineteenth century railroad maps and related sources as described below.

We measure the causal impact of the railroad on improved acreage through a difference-in-differences (DID) analysis, comparing the change in farm outcomes between 1850 and 1860 in a treatment group of counties – those which gained access to the railroad during the period– versus a control group which did not. We have excluded those counties which already had rail links in 1850 so that our treatment group consists of counties that gained rail access during the 1850s versus those counties

which did not have access prior to the Civil War. This analysis is complicated by the fact that some county boundaries were redrawn during the period while others were not even created until the 1840s. Consequently, we have restricted our analysis to the 278 counties in the Midwest that were in existence as of 1840 and did not change their boundaries during the 1850s.

The principal finding from the DID analysis is that the coming of the railroad had a positive and statistically significant effect on the share of improved acreage in agriculture. Our preferred DID estimate of the treatment effect is 0.048, implying that gaining access to the railroad increased the percentage of farm acreage that was improved in a county by 4.8 percentage points which is about 12 percent of the mean value in the sample counties in 1850—a sizeable impact.

We supplement this analysis with two robustness checks. First, because improved land was more valuable than unimproved land, we should also find increases in the value of farm acreage associated with the coming of the railroad. We confirm that the railroad, indeed, had a treatment effect on farm values, again using a DID analysis. When we repeat the DID analysis of farm values but control for the percent of land that was improved we are able to explain some but not all of the increase in farm values, suggesting that the railroad's effects on land value were general – that is a pure location effect.

A key assumption in our DID analysis is that, conditional on various observable determinants, the coming of the railroad to a county was a random event. If this assumption is false then the estimated treatment effect of the railroad on farm improvements would be biased. In particular, if county residents anticipated getting a railroad it would make economic sense to “build-in-advance” – that is, improve acreage in advance of the railroad. If this were the case, our DID treatment effect would be biased downwards.

To address this possible bias we compare the results of the DID analysis with those using an instrument variable. This instrument is derived from the federal government transportation surveys that were conducted in the 1820s and early 1830s—that is to say, right at the start of the railroad age. Our IV

estimate of the treatment effect of rail access on improved is positive and significantly larger in magnitude than the DID estimate. Indeed, the IV results suggest that all of the growth in the fraction of farm acreage that was improved in Midwestern agriculture in the 1850s can be attributed to the diffusion of the railroad.

Land Clearing and Transportation Improvements in Nineteenth Century America

While the first step to becoming a farmer was to acquire title to the land either through private purchase or directly from the federal government under the terms of the prevailing land legislation, its purchase was generally a minor cost compared with the cost of clearing the virgin land. Where the land was forested, trees had to be cut down and burned or hauled away. Even more laborious and labor intensive was the removal of the stumps which were left. Where the land was already clear of trees, as on the Midwestern prairies, heavy, specialized equipment and crews were needed to break the prairie sod (Stellingwerff and Swierenga 2004) but much of the other work could be done by the farmer with the assistance of family members (or neighbors) using only the tools and implements that came readily to hand on a nineteenth century farm. Even when the land had been cleared of its native vegetation, it still had to be fenced to keep the livestock in and the predators out. These activities were often arduous and certainly time-consuming and expensive. Farmers rarely cleared all the land they needed or wanted at once. A general rule of thumb was that a farmer with a typical size household could clear, perhaps, 5-10 acres a year and so, clearing the approximately 1.3 million acres of land (much of it forested), that was cleared annually in the Midwest in the 1850s must have occupied the time of about one-sixth of the Midwestern labor force at the time (Primack 1962; Primack 1977). Needless to say, with this kind of investment in cash and/or time, cleared land was highly valuable compared with virgin stock, accounting for much of the resale price of farm land. These out-of-pocket and opportunity costs doubtless account

for much of the \$11-12 per acre land clearing costs in Seaman's estimates—or about ten times the cost of land purchased directly from the federal government prior to passage of the Graduation Act in 1854.¹

The economic logic linking transportation costs to land improvements can be illuminated using the Von-Thünen model of agricultural land rents that was first proposed in 1826 (Thünen and Hall 1966). To fix ideas, suppose that each farmer has a farm of size one acre (of improved land), located at distance t from a “central place” where trade with the hinterland (i.e., the farmer) takes place. The farm supplies labor z per acre at an opportunity cost of w . The costs of transporting output per unit are c and the price of output (at the central place) is p .

In a competitive market these assumptions imply that the rental price of a farm at distance t , $R(t)$, from the central place is

$$R(t) = q(p - w \cdot z - c \cdot t)$$

Note that, if $t = 0$ – the farm is located on the boundary of the central place and $R(t)$ is at a maximum as determined by the yield of the land (q) given the labor input (z), output price, and the opportunity cost of the farmer's time. Conversely, there will exist a t^* that defines the distance from the central place beyond which it is not profitable to farm, $t^* = (p - w \cdot z)/c$. Economists sometimes refer to this as the extensive margin for market production.

It is straightforward to show that $dR(t)/dc = -q \cdot c < 0$ and $dt^*/dc = (p - w \cdot z)/c^2 > 0$; that is to say, a reduction in transport costs “flattens” the relationship between land rents and distance to the central place and, therefore, increases the distance t^* at which farming is still just profitable. A fall in transport costs therefore makes it economical to bring more land into production. However, before this land can be brought into production it has to be improved – that is, cleared and fenced– hence, the link between transport costs and land improvements.

¹ And substantially more than the average price paid for public lands bought using land warrants distributed to veterans. See Oberly, J. W. (1990). Sixty million acres : American veterans and the public lands before the Civil War. Kent, Ohio, Kent State University Press.

Of course, other factors will also determine t^* and therefore affect the amount of land in production. A rise in p , for example, or a reduction in w , will also increase t^* . Moreover, in an explicitly dynamic setup the permanent versus transitory nature of changes in these parameters will influence the likelihood of improving land and bringing it into production, as will expectations about future values. In particular, railroad building was (in general) a permanent change in the environment and thus could be expected to generate a greater response in land improvements than, say, what might be only a transitory shift in output prices. One reason for this was that much of the railroad investment was not only fixed but also sunk in that it was embedded in location-specific improvements. These could not be relocated and their benefits could only be fully captured by a transportation medium—investments in grading, filling and cuttings, for example. According to the 1880 Census (United States. Census Office. 10th census 1880., Walker et al. 1884), over 80 percent of railroad investment went into construction costs, of which only one or two percent represented the cost of the land itself; the rest was spent on surveying, grading, removing or bridging obstacles, and laying the track. These ties, ballast and the rails might be reused elsewhere and the land itself could be resold, but the grading, cuttings, embankments, bridges, and drainage ditches had few alternative uses—especially in the nineteenth century.² Indeed, railroads (and interstate highways) today still follow many of the routes blazed by the nineteenth century railroads.³

If there were any adjustment costs involved in making improvements—as would seem likely—it would be rational for farmers to improve some land in advance of construction if they anticipated receiving rail access in the near term—that is farmers should be forward-looking rather than merely

² Nowadays, of course, they find recreational uses in the “rails to trails” movement

³ And, of course, thanks to GIS, these can be quite precisely located. One might thus be tempted to begin with modern railroad maps and force the nineteenth century locations to conform to these. On average, this is probably a good strategy even though it would misrepresent earlier track locations that were changed through realignments such as that described above. There are two downsides to this strategy—the potential for copyright infringement since these GIS SHP files are proprietary commercial ventures and the large number of individual segments which comprise the modern rail network. For example, the 177,000 mile rail network from Bureau of Transportation Statistics and distributed ESRI with ArcGIS contains over 131,000 segments. See North American Transportation Atlas Data (NORTAD) CD: “National Railway Network” from http://www.bts.gov/publications/north_american_transportation_atlas_data/

reactive. We return to this point later in our instrumental variable analysis, as it relates to a possible downward bias in our difference-in-differences estimation.

The Data

We use a newly constructed geographic information system (GIS) database on the Midwestern transportation infrastructure in the mid-nineteenth century United States which we have linked to an updated version of the well-known ICPSR county level census database. This GIS-based methodology has several advantages over earlier approaches which have generally involved matching historical transportation maps to county boundary maps by hand (and eye).⁴ Our GIS database was created from geo-referenced digitized historical maps showing the transportation system at different benchmark dates.⁵

Canal data, for example, were taken from Poor (Poor 1860) and Goodrich (Goodrich 1961) as well as digitized maps from the Library of Congress “American Memory” web site such as the Williams’ (Williams 1851) and Disturnell’s 1850 (Burr 1850) maps. Information on navigable rivers was taken from U.S. Army Corp of Engineers GIS data (Vanderbilt University. Engineering Center for Transportation Operations and Research 1999) supplemented by information from Hunter (Hunter 1949) and contemporary newspaper accounts regarding steamboat service on specific rivers. Very little change occurred in the canal and river systems after the 1840s until the Corp of Engineers embarked on

⁴ See, for example, Craig, L. A. P., Raymond B. Weiss, Thomas (1998). "Transportation improvements and land values in the antebellum United States: a hedonic approach." *Journal of real estate finance and economics* **16**(2): 173-189. who visually compared historical maps to county boundaries (which generally did not appear on the historical maps). For a discussion of the hand-matching procedure used by Craig and the pitfalls that can arise, see footnote #2 of Atack, J., F. Bateman, et al. (forthcoming). "Did Railroads Induce or Follow Economic Growth? Urbanization and Population Growth in the American Midwest, 1850-60." *Social Science History*.

⁵ Geo-referencing refers to the process of fixing specific points with known geographic coordinates between the digitized image which was invariably drawn and printed with error (and which may also be subject to parallax error as a result of the digitization process) and geographically accurately rendered base boundary files (a shapefile in ESRI's parlance). Algorithms within the GIS software then distribute the error (the difference) between the historical images and the boundary file across the space between fixed points. In essence, the process treats the historical image as if it were printed on a sheet of rubber which is then stretched over the boundary file with pins holding it in place at fixed reference points between the two. Once done, it is then possible to “trace” features from the historical image onto the geographically coordinated boundary file. The resulting files can then be manipulated and used for computations using the GIS software. Accurate historical county boundary files are freely available from the National Historical Geographical Information System at the University of Minnesota (<http://www.nhgis.org/>) along with a wide variety of historical U.S. census data.

ambitious navigation plans after the Civil War.⁶ Consequently, we have measured access to water-borne transportation as of a single benchmark date – 1850 – rather than by change over time.

The railroad part of our GIS database was created from digitized state-level maps for 1911 from *The Century Atlas* (Whitney and Smith 1911) and digitized by Goldbug.com. The resulting railroad network was then traced backwards to earlier dates using digitized transportation maps from the Library of Congress and other sources such as the maps created by Taylor and Neu (1956) from contemporary travel guides (see below). The underlying premise behind this procedure is that most railroad investment was literally sunk in location-specific grading and other immovable features, a presumption that is strongly supported by the available data.

The Midwestern portion of our GIS database which is the focus of this study covers seven states: Indiana, Illinois, Iowa, Michigan, Missouri, Ohio and Wisconsin. For five of these (Indiana, Illinois, Michigan, Ohio and Wisconsin), information about the spreading rail network was taken from a series of crudely-drawn maps prepared over ninety years ago at the University of Wisconsin by Professor Frederick L. Paxson (Paxson 1914) and his students using extant contemporary travel guides published between 1848 and 1860.⁷ Many of these travel guides seem to have been lost or disintegrated since

⁶ In the Midwestern states we focus on here, for example, the only change of note was the opening of a stretch of the Wabash and Erie canal between Terre Haute and Evansville which diverted traffic from the Wabash River to the canal Cayton, A. R. L. (1996). *Frontier Indiana*. Bloomington, Indiana University Press..

⁷ These travel guides first appeared in the 1840s and include *Disturnell's Guide* Disturnell, J. (1846). *Disturnell's railroad, steamboat, and telegraph book : being a guide through the middle, northern, and eastern states and Canada : also giving the great lines of travel south and west and the ocean steam packet arrangements, containing tables of distances, &c. telegraph lines and charges, list of hotels, express offices, &c. &c.* New York, J. Disturnell., *Doggett's Gazetteer* Doggett, J. (1847). *Doggett's railroad guide and gazetteer for ... with sectional maps of the great routes of travel*. New York, J. Doggett Jr., *Appletons' Guide* Appleton D. and Company. (1848). *Appletons' railway and steam navigation guide for the United States and the Canadas*. New York, D. Appleton & Co., *Dinsmore's Guide* Cobb, C. (1850). *American railway guide and pocket companion for the United States*. New York, C. Dinsmore., *Lloyd's Guide* Lloyd, E. (1857). *Lloyd's American guide : containing new arranged time tables, so simple and correct that a child can understand them, it being universally acknowledged that all other guide books are so complicated that not one in a hundred can understand them : the population, states, and distances to every place on all the railroad routes in the United States and Canadas : photographic portraits of all the railroad presidents and superintendents--men controlling*. Philadelphia, E. Lloyd., *Travelers' Guide* National Railway Publication Company. (1868). *Travelers' official railway guide of the United States and Canada*. Ann Arbor, Mich., printed for the National Railway Publication Co. by University Microfilms. and *The Rand-McNally Guide* Rand-McNally (1871). *The Rand-McNally official railway guide and handbook*. Chicago,.. Some of these were published monthly; others, semi-annually or annually. Each typically went through many editions. All of the guides that we have physically handled are fragile,

Paxson completed his study. This is unfortunate as there is little doubt that they are, or would have been, the single most valuable resource for our research. These guides served the needs of the traveling public, providing up-to-date route maps and timetables. They were published at fairly frequent and regular intervals by a number of different companies and presumably competition among them should have guaranteed that only the best (i.e. the most accurate, complete and useful) survived. Absent access to a complete set of travel guides, our data for Iowa and Missouri are taken from digitized contemporary state maps from the on-line David Rumsey Map Collection (<http://www.davidrumsey.com/>).

While GIS software can produce a wide range of different measures of transportation access, at the present time we use a simple binary variable, ACCESS, which takes on a value of 1 if a railroad crossed or formed a county's border. This reflects the still-preliminary nature of our GIS database. While we are certain that it can create the binary "0-1" access variable very accurately, we are not yet convinced that the original maps and "tracings" of them are sufficiently precise to generate more refined measures of access, such as the number of station stops within a county. However, although this metric is somewhat crude and limited, our rail access variable is still the best current measure of the historical spread of the US railroad system at the county level for the period in question.⁸

The transportation access data have then been linked with the county-level Haines-ICPSR census data using county FIPS codes which are common to both databases. In 1840, the seven states and territorial areas in our study had 391 "counties" (these include, for example, Clayton "county" in the Iowa

especially the multi-page fold-out maps, and not sturdy enough for scanning or copying, although a few guides have been digitized and are available on-line. See, for example, the June 1870 copy of the Travelers' Official Railway Guide at http://cpr.org/Museum/Books/I_ACCEPT_the_User_Agreement/Travellers_Guide_6-1870.pdf from the Central Pacific Railroad Museum. There are also at least two different editions of Appleton's Guide on Google Books such as <http://books.google.com/books?vid=UOM39015016751375> as well as a number of other guides. See http://www.lib.utexas.edu/maps/map_sites/hist_guide_sites.html

⁸ One obvious problem is that the access variable does not capture within-county variation in access and counties differed in size. In our econometric analysis we weight by county land area but we are unable at the present time to adjust for cases in which the railroad lay close to a county boundary but where the adjacent counties do not have rail access. County boundaries, in other words, are arbitrary ways of delineating local economies; the average farmer in a geographically small county in our control group that happened to be adjacent to a large county in the treatment group might have had better access to the railroad than the average person in the large county. If this were common it would produce a downward bias in our estimated treatment effects. Although we do not believe this bias to be important empirically we admit that the question is open.

Territory which then comprised most of what we now know as the state of Minnesota) (Thorndale and Dollarhide 1987). By 1860, however, all of the area under consideration had been organized into states and divided into 623 counties. Because of such boundary changes, we have restricted our analysis to Midwestern counties that were (1) present beginning in 1840, (2) had the same county boundaries (as determined by square mileage in 1850 and 1860), and (3) did not have rail access as of 1850. Taken together these restrictions produce a balanced panel to 278 Midwestern counties. They are distributed as shown in Table 1 while their locations are mapped in Figure 1.⁹

We have divided this panel into a treatment group and a control group. The treatment group did not have rail access in the 1850 but gained it at some point between 1850 and 1860. Our control group is the set of counties in the panel that did not have rail access in either 1850 or 1860. The 1840 data for these same counties then allows us to control for pre-1850s trends in certain variables in our analysis (see below).

By construction, none of the counties in our sample had rail access in 1850. By 1860, however, 195 of the 278 counties, representing almost 72 percent of the land area in the panel, had gained direct access to at least one railroad (See Table 2 and Figure 2).¹⁰ The share of farm acreage that was improved rose by 13 percentage points between 1850 and 1860 across all counties in the panel whereas the logarithm of per acre value of farms, which we use in our robustness checks (see below), increased by 0.796 log points, or 123 percent ($= (\exp[0.796] - 1) \times 100$).

⁹ By “balanced” we mean that the same counties appear in 1850 and 1860; no new counties enter the sample during the 1850s. Balancing ensures that county fixed effects are “differenced away” when we compute the change in economic outcomes from 1850 to 1860; this would not be the case if new counties entered the sample in the 1850s. We restrict our basic analysis to counties with fixed land area because the ICPSR census data are not adjusted for changes in land area over time. Results are qualitatively the same if we do not impose the restriction that the county not have rail access by 1850 and if we do not impose the requirement that county boundaries be the same in 1850 and 1860.

¹⁰ The map shows all railroads in the area as of 1861. The highlighted roads—those with thicker (red) lines—are those built between 1850 and 1860. Almost without exception, those railroads that are NOT highlighted in the East predate 1850 while those in the West post-date 1859.

Table 2 also shows the sample averages for percent improved and the logarithm of per acre farm values conditional on their treatment status. In the treatment counties, the percent of land that was improved rose by 15.6 percentage points from 1850 to 1860 or almost double the 8.4 percentage point increase in the control counties. The difference between these—7.2 percentage points—is the difference-in-differences estimate of the treatment effect of gaining rail access with no other control variables in the regression (see below). This is equal to 18 percent of the mean proportion improved in 1850. As such, the effect is economically large. The equivalent difference-in-differences estimate of the treatment effect of rail access on per acre farm values is 0.072, or about 7.5 percent. Both estimates are statistically significant at the 5 percent level.

Ideally, in a difference-in-differences analysis, the treatment and control counties will be sufficiently “similar” or “matched” such that treatment can be (plausibly) argued to be randomly assigned. However, it is immediately evident from Table 2 that this was not the case with gaining rail access. Counties that gained rail access in the 1850s already had a greater share of their farm acreage improved and more valuable land by 1850 – that is, before gaining rail access.

That the treatment counties already had more of their farm acreage improved and more valuable farms (in part, because the land was more improved, see below) are manifestations of the same general issue –railroad did not arrive randomly. In particular, railroad promoters and investors in the Midwest paid close attention to the quality of agriculture because the profitability of the railroad depended upon it (Fishlow 1965). In Table 3, therefore, we explore some of the correlates, agricultural and otherwise, of gaining rail access in the 1850s using a linear probability model. In column 1 of Table 3, we include just three covariates: the logarithm of the value of agricultural output per improved acre (a measure of the land’s “yield,” in this case revenue productivity); the percentage of total agricultural output in 1840 represented by wheat production; and the change in the percent wheat between 1840 and 1850.¹¹ We

¹¹ The 1840 census reports the value of wheat production and the total value of agricultural output. For 1850 we use an estimate of the value of total agricultural output based on national prices multiplied by quantities; the wheat share

examine wheat in particular because it was the primary “cash crop” of the period and, therefore, likely a key determinant of whether a rail line would be profitable. Each of these variables has a positive and significant effect on the likelihood of gaining rail access; for example, a 10 percentage point increase in the wheat share boosts the probability of gaining rail access by 5.3 percentage points.

However, the significant positive coefficients on the agricultural variables may also be due to other factors that are omitted from the regression. For example, in an earlier paper (Atack, Bateman, Haines, and Margo, forthcoming) examined the relationship between gaining rail access and measures of urbanization, population density and our water transportation dummies using the same panel. Accordingly, in column 2 of Table 3 we include measures of urbanization and population density in 1840 as well as changes in these variables between 1840 and 1850. We also include dummy variables for the presence of different modes of water transportation (river, canals, and abutting the Great Lakes) while in column 3 we also includes dummy variables for each state. Adding additional controls reduces the magnitudes of the coefficients of the agricultural variables but the coefficients on the yield variable and the “pre-trend” (1840-50) in the wheat share remain statistically significant. Moreover (not shown) several of these control variables are themselves statistically significant – for example, the presence of a canal was positively associated with gaining rail access whereas having a navigable river reduced the statistical likelihood of gaining rail access.

(Table 3 about here)

Because gaining rail access was not a random event, we should also control for the factors shown in Table 3 in our difference-in-differences analysis as these factors may have had independent effects on the percentage of improved acreage. Column 1 of Table 4 reports our base DID estimate for the percentage

is therefore the wheat output (in bushels) multiplied by wheat price divided by the estimated value of agricultural output. It is possible that our procedure for estimating agricultural values may overstate the growth in the percent wheat over the 1840s (because 1840 output was probably valued at local prices) but any such bias should be mitigated once we control for state fixed effects (since state-level variation arguably captures the most salient price variation). Our agricultural yield variable also uses the estimated value of agricultural output in its construction (the numerator).

of total farm acres that were improved, taken from Table 3. In column 2 we report the DID estimate of the treatment effect, taken from a regression in which we include interactions between the full set of control variables in Table 3 and the dummy variable for the 1860 census year. Including these control variables reduces the DID treatment effect from 0.072 (base estimate) to 0.048, or 4.8 percentage points. However, even with the additional control variables, the treatment effect is positive: counties that gained rail access in the 1850s increased improved acreage to a significantly greater extent than counties that remained rail-free before the Civil War even after controlling for other factors.

We have made two robustness checks on this finding. First, we estimate the treatment effect of gaining rail access on the per acre value of farms. If the logic underlying the von Thünen model is correct, then the impact of gaining rail access should also be reflected in higher per acre farm values. We have already shown that this was the case with the simple DID calculations in Table 2 but there we have not controlled for the additional factors associated with gaining rail access. Moreover, the size of the treatment effect should be smaller if we also control for the change in the percent improved – again, something that is not reflected in DID estimate in Table 2.

Table 5 shows the difference-in-differences estimates for the per acre value of farm land. In contrast to the results for improved acreage the effect is larger once we add the additional controls. Note that, when we add the change in the percent improved acres to the regression, the treatment effect is reduced in size, as the von Thünen framework would predict; however, controlling for the percent improved does not eliminate the effect of gaining rail access on land values. This suggests that transportation improvements raised the value of unimproved land as well. This, of course, generally should be the case since most unimproved land has an option value – the farmer could always choose to improve it in the future.

Our second check on the robustness of our results is made using an instrumental variable. Although we can control for the factors in Table 3 in our DID analysis it is still possible that the analysis is invalid because the treatment – gaining rail access – is correlated with the error term even after controlling for

observable determinants of gaining rail access. This problem is of particular concern here because the 1840 census reported no acreage statistics whatsoever. Consequently we cannot control for the pre-1850 trend in the percent of land that was improved. This will likely bias the treatment effect towards zero.

To understand why this is so, consider the following thought experiment. Suppose that farmers in a county anticipated gaining rail access in the near future – a reasonable supposition for Midwestern farmers at the start of the 1850s. If the costs of improving acreage were increasing at the margin which was almost certainly the case (the most readily improved land would be improved first), it would make economic sense to incur some of the costs prior to gaining rail access. If this were the case counties that gained rail access in the 1850s would have experienced above average growth in improved acreage somewhat earlier and this would be reflected in a higher average percent improved in 1850 – as we observed was the case (recall Table 2). Conversely, some of the control counties might have anticipated (correctly, as it happens) that they would only gain rail access after 1860. By this same logic, they would have experienced above average growth in improved acreage in the 1850s. To the extent that our control variables capture this process, the bias can be “corrected” but we cannot be sure we have captured all of it. Consequently, the percent improved will grow too slowly in the treatment counties in the 1850s and too fast in the control counties, causing the DID estimate of the treatment effect to be biased downwards.

The appropriate correction for this sort of bias is to estimate the relationship using an instrumental variable—a variable that predicts gaining rail access in the 1850s when we control for other factors but which is otherwise uncorrelated with the outcomes we are examining. Using a variable such as this which isolates plausibly exogenous variation in rail access is similar to what would have been the case if rail access had been randomly assigned. We then predict rail access in 1860 using the instrumental variable and examine the effect of that predicted access on the percent of land that was improved in 1860.

The historical narrative of internal improvements in America, particularly that for canal construction, has assigned an important role to the government in promoting these advances (Goodrich 1961). One

such source of government assistance was the use of the Army Corp of Engineers to conduct surveys for potential transportation routes. Beginning in 1824, the President was granted authority to survey routes for “such roads and canals as he may deem of national importance, in a commercial or military point of view, or necessary for the transportation of the public mail” (U. S. Congress. 1823-4). Although railroads were not mentioned in the original act (hardly surprising since it predates even the Stockton to Darlington Railway, the first steam railroad in the world), it was not long before surveys conducted under this legislation also considered them. For example, in 1825 a survey to “ascertain the practicability of uniting the headwaters of the Kenawha [sic] with the James river and Roanoke river” expressly mentions railroads and they soon came to dominate the surveys with perhaps as many as 61 such surveys being made before the law was repealed by the Jackson administration in 1838 (Haney 1908, p. 277).

Our instrumental variable is based upon these government surveys as reported in *American State Papers* and compiled by Haney (1908, p. 283). For each survey, we have identified the counties at the start and endpoint of the proposed railroad. For example, a 1831 railway survey plotted a route from Portage Summit on the Ohio Canal (near Akron) to the “Hudson River” (we used Albany as the terminus) (Haney 1908, p. 286). In some cases we had to infer the endpoints such as for a 1832 survey for a route between the Mad River and Lake Erie in Ohio (Haney 1908, p. 286). We used Springfield and Sandusky as the termini of this projected railroad. Having identified the starting and ending counties, we then drew a straight line between the center of the “start” and “end” counties. Counties that lay along this straight line received a value of one, while those that did not were coded as zero. That is, if a railroad were built, our instrument presumes that it would be built in a straight line as the shortest distance between the two points.¹²

¹² Our use of a “straight-line” instrument is inspired in part by Bannerjee, Duflo, and Qian (2006) who construct a similar instrument for their study of the impact of rail access in modern China on wages. Of course, many features of topography other than the shortest-distance criterion – grade, hills or mountains, and so on – influenced railroad building but these features of topography likely affected density and urbanization directly, and thus are not candidates for instrumental variables (they fail the exclusion restriction).

These surveys provided valuable information to the general public and prospective railroad promoters alike regarding topography and other factors that would affect potential construction costs. Since the costs of these surveys were borne by the public purse, their existence should have raised the likelihood that a railroad would eventually be built by lowering its private costs. Indeed, George Rogers Taylor (1951, p. 95) even argued “[a]s trained engineers were still very scarce ... the government rendered a uniquely valuable service by making its experts available for such surveys” while Haney (1908, p. 284) observed “it is of some significance that in most cases the routes of these government surveys were early taken by railways ... in the great majority of cases these early surveys have been closely followed” [emphasis added]. Presumably, therefore, they were found to be very useful and so it would seem that our “Congressional Survey” instrument should be well suited to predicting whether or not a county gained rail access.

Table 6 reports estimates using this instrumental variable. The Congressional Survey instrument does quite well in predicting treatment (gaining rail access in the 1850s) even when we control for all of the other variables included in the DID analysis (this is the “first stage” coefficient shown). We have then used this first stage regression to predict the probability of gaining rail access, and used the predicted values of treatment in the second stage of the two-stage least squares (2SLS) regression of 1860 outcomes. The 2SLS coefficient is positive and much larger than our DID estimate (0.192 versus 0.048). The difference between the two is statistically significant at the 6 percent level. The much larger IV coefficient strongly suggests that the anticipatory logic described earlier was present, leading to a downward bias in the estimated treatment effect of gaining rail access on improved acreage.

Concluding Remarks

We have used a novel data set on the antebellum transportation networks in the Midwest derived from applying GIS software to digitized historical maps to estimate the impact of gaining rail access on investments in improving land for agricultural cultivation. Such investment during the nineteenth century

formed a major part of U.S. capital formation at the time. We find that counties that gained early access to the railroad – in our case, prior to 1860 rather than later – experienced significantly greater increases in the percent of farm acres that were improved. Such counties also experienced above average increases in the value of their farms. This was in part because of the rise in the share of improved acreage but also because, with the coming of the railroad, farm land in general was more valuable because of its higher revenue product and locational advantages.

We have used the estimated treatment effects to gauge the overall explanatory power of the railroad as a factor behind these agricultural improvements. If we use the difference-in-differences estimate (0.048, Table 3, column 4), then predicted change in percent improved acreage between 1850 and 1860 due to gains in rail access is 0.034, or 3.4 percentage points ($0.048 \times 0.717 = 0.034$). This predicted increase in turn accounts for 25 percent ($0.034/0.136 = 0.25$) of the increase in percentage of improved acreage over the 1850s. If, however, we used the much larger IV coefficient (0.192), then predicted increase in percent improved acreage is 0.138. Taken at face value, this implies that the coming of the railroad accounts for all of the increase in percent improved acreage over the decade—indeed, that expansion would have been even more rapid but for some (unidentified) retarding factors. Although the IV coefficient is almost significantly different from the DID coefficient at conventional significance levels, it is not significantly different from, say, the average of the DID and IV coefficients. If we use the average of the DID and IV coefficients ($0.12 = (0.048 + 0.192)/2$) we can attribute slightly more than two-thirds (68 percent = $0.12 \times 0.771/0.136$) of the growth in improved acreage to the spread of the railroad. In short, whatever else might have led Midwestern farmers to undertake the back-breaking labor of clearing their land, no other single factor seems likely to be as important as the potential gains from trade deriving from the arrival of the Iron Horse.

State	Number of counties	Percent of Total (%)
Illinois	71	25.5
Indiana	66	23.7
Iowa	17	6.1
Michigan	12	4.3
Missouri	41	14.8
Ohio	57	20.5
Wisconsin	14	5.0

Notes: To be included counties must be (1) present in 1840, 1850 and 1860 (2) have fixed county boundaries (3) did not have rail access in 1850 but gained rail access by 1860 (treatment group) or did not have rail access in both 1850 and 1860 (control group). There are 278 counties in the sample.

Table 2 Sample Statistics			
	1850	1860	Δ (1860 – 1850)
Rail = 1	0	0.717	0.717
% Improved Acres in Agriculture	0.390	0.526	0.136
Treatment Counties	0.411	0.567	0.156
Control Counties	0.337	0.421	0.084
Treatment - Control	0.074	0.146	0.072
Ln (Per Acre Farm Value)	2.119	2.915	0.796
Treatment Counties	2.211	3.028	0.817
Control Counties	1.887	2.632	0.745
Treatment - Control	0.324	0.396	0.072

Source: See Table 1. There are 278 counties in the sample. Treatment = 1 if county gains rail access between 1850 and 1860. Control = 0 if the county does not have rail access before the Civil War. There are 195 treatment counties and 83 control counties. Observations are weighted by square miles prior to calculating sample means.

Table 3: County-Level Agricultural Correlates of Gaining Rail Access in the 1850s: Linear Probability Regression.

	Coefficient	Coefficient	Coefficient
Constant	0.222* (0.089)	0.365* (0.123)	0.456 (0.150)
Log (Agricultural Yield in 1850)	0.325* (0.077)	0.290* (0.078)	0.280* (0.092)
Percent Wheat in 1840	0.533* (0.244)	0.349 (0.253)	0.356 (0.283)
Δ in Percent Wheat, 1850-40	0.684* (0.205)	0.652* (0.272)	0.615* (0.275)
Urban and Population Density Variables Included?	No	Yes	Yes
Water Transportation Dummies Included?	No	Yes	Yes
State Dummies Included?	No	No	Yes
Adjusted R ²	0.111	0.190	0.211

Source: see text and Table 1.

Notes: Unit of observation is the county (N = 278). Dependent variable = 1 if county gains rail access by 1860 (treatment group), 0 otherwise (control group). Agricultural Yield: value of agricultural output/(improved + unimproved acres in agriculture). Percent wheat: value of wheat output/value of total agricultural output. Urban and Population Density variables: percent urban in 1840, ln (population/square miles) in 1840, change in percent urban between 1840 and 1850, change in ln (population/square miles) between 1840 and 1850. Water transportation: canal = 1 if canal existed within county boundary (or part of boundary), river = 1 if navigable river existed within county boundary or part of county boundary, Great Lakes = 1 if county abutted one of the Great Lakes. Observations are weighted by county square miles prior to estimation.

Standard errors are shown in parentheses. *: significant at 5 percent level.

Table 4: Difference-in-Differences Estimates of the Impact of Gaining Rail Access on the Percent of Farm Acres that were Improved: 278 Midwestern Counties, 1850-60

	Coefficient	Coefficient
Rail Access = 1	0.072* (0.015)	0.048* (0.015)
Controls?	No	Yes

The unit of observation is the county. The dependent variable is the percent of farm acres in the county that were improved as measured in the census of agriculture. The coefficient of Rail Access = 1 is the coefficient of a dummy variable for the treatment counties (Treat = 1) interacted with a dummy variable for year = 1860. In column 1 there are no control variables other than fixed effects for counties and year = 1860. In column 2 the regression also includes interaction terms between the control variables in column 3, Table 3 (see the notes to Table 3) and year = 1860. Standard errors shown in parentheses are clustered at the county level. Observations are weighted by square miles prior to estimation. *: significant at the 5 percent level.

Table 5: Difference-in-Differences Estimates of Rail Access on Ln (Value of Farms/Farm Acres): 278 Midwestern Counties, 1850-1860

	Coefficient	Coefficient	Coefficient
Rail Access = 1	0.072* (0.033)	0.120* (0.032)	0.087* (0.030)
% Acres Improved in Agriculture			0.657* (0.164)
Controls?	No	Yes	Yes

The dependent variable is the natural logarithm of the total value of farms divided by total farm acres (improved + unimproved). See notes to Table 4 for controls in columns 2 and 3. All regressions include fixed effects for counties and year = 1860. Standard errors are clustered at the county level. *: significant at the 5 percent level.

Table 6: Instrumental Variable Estimates of the Treatment Effect of Gaining Rail Access on the Percent of Farm Acres that were Improved: 278 Midwestern Counties, 1850-60

	First Stage, IV	Second Stage (2SLS)
Rail Access = 1		0.192* (0.073)
Congressional Survey = 1	0.316* (0.068)	
IV – DID = 0, Significance Level		0.060
Sample	1860	1860
Controls?	Yes	Yes

First Stage shows the coefficient of the Congressional Survey instrumental variable from a linear probability regression of rail access. Controls are the same as in column 3 of Table 3. Second stage shows the coefficient of the predicted value of rail access from a regression of percent improved using the 1860 observations from the panel; control variables are the same as in column 1, above. IV – DID = 0: significance level of a test of the difference between the second stage coefficient (0.192) and the DID coefficient from Table 4, column 2 (0.048).

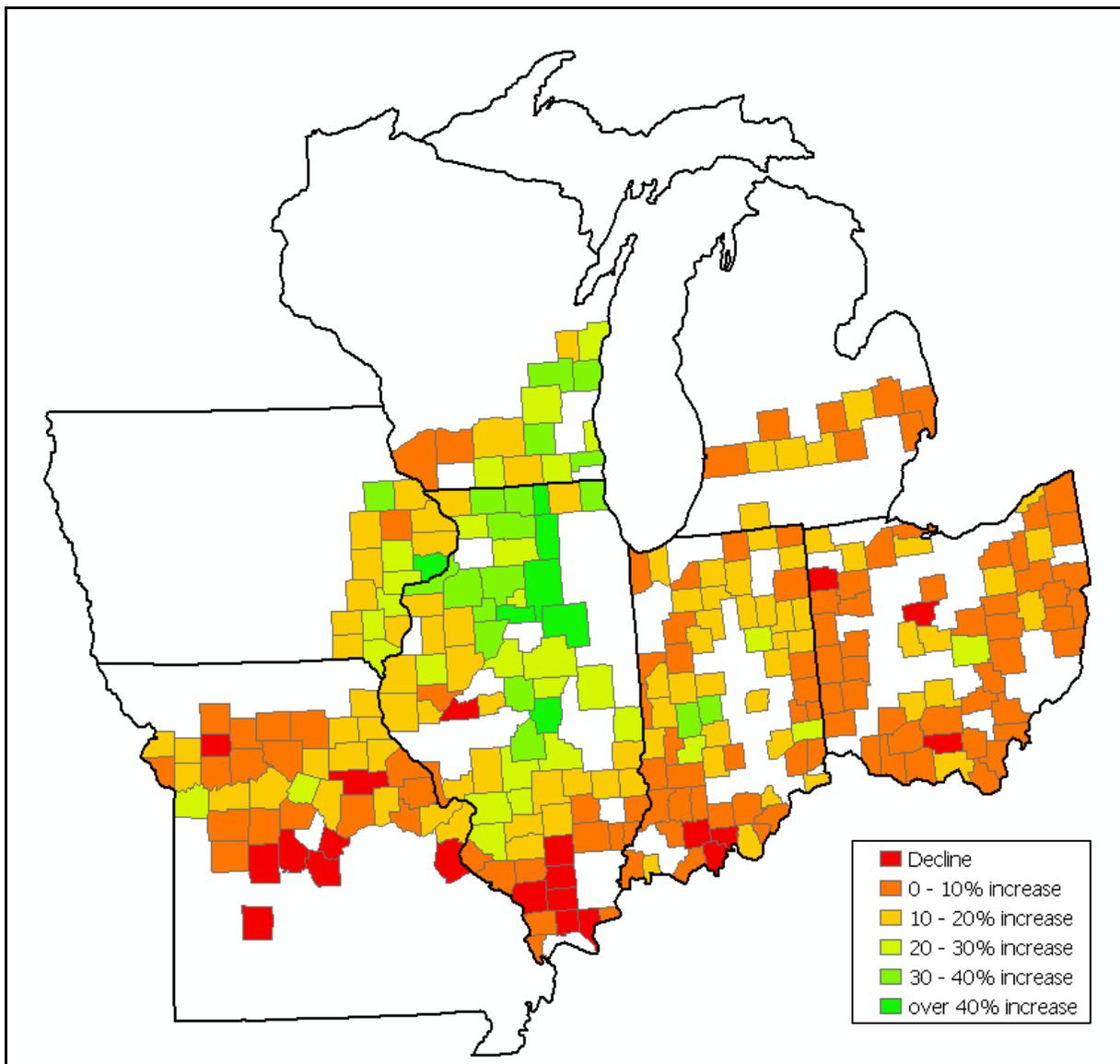


FIGURE 1

Distribution of Counties Existing in 1840 and with Unchanged Boundaries as of 1860, Showing the Change in the Percentage of Farm Land that Was Improved Between 1850 and 1860.

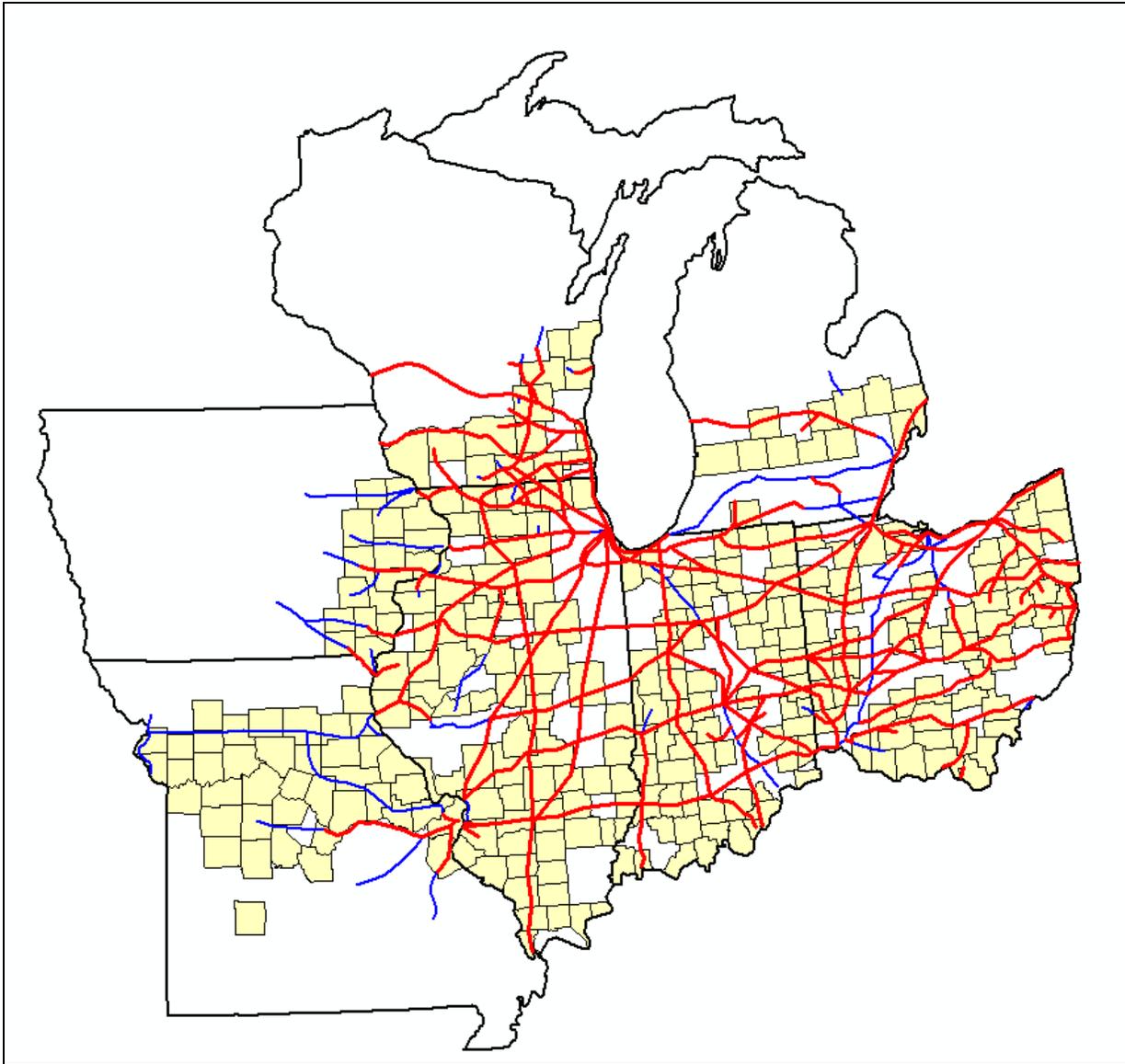


FIGURE 2
Midwestern Railroads Built Between 1850 and 1859 (highlighted in red)
Relative to the Panel of Counties with Stable Boundaries

Bibliography

- Appleton D. and Company. (1848). *Appletons' railway and steam navigation guide for the United States and the Canadas*. New York, D. Appleton & Co.
- Atack, J., F. Bateman, et al. (forthcoming). "Did Railroads Induce or Follow Economic Growth? Urbanization and Population Growth in the American Midwest, 1850-60." *Social Science History*.
- Atack, J., F. Bateman, et al. (2000). Northern Agriculture and the Westward Movement. *The Cambridge economic history of the United States. Volume 2. The Long Nineteenth Century*. S. L. Engerman and R. E. Gallman. Cambridge ; New York, NY, Cambridge University Press: 285-328.
- Burr, H. A. (1850). *Disturnell's new map of the United States and Canada*. New York, Disturnell
- Carter, S. B., S. S. Gartner, et al. (2006). *Historical Statistics of the United States Millennial Edition Online*. New York, Cambridge University Press.
- Cayton, A. R. L. (1996). *Frontier Indiana*. Bloomington, Indiana University Press.
- Cobb, C. (1850). *American railway guide and pocket companion for the United States*. New York, C. Dinsmore.
- Craig, L. A. P., Raymond B., Weiss, Thomas (1998). "Transportation improvements and land values in the antebellum United States: a hedonic approach." *Journal of real estate finance and economics* **16**(2): 173-189.
- Disturnell, J. (1846). *Disturnell's railroad, steamboat, and telegraph book : being a guide through the middle, northern, and eastern states and Canada : also giving the great lines of travel south and west and the ocean steam packet arrangements, containing tables of distances, &c. telegraph lines and charges, list of hotels, express offices, &c. &c.* New York, J. Disturnell.
- Doggett, J. (1847). *Doggett's railroad guide and gazetteer for ... with sectional maps of the great routes of travel*. New York, J. Doggett Jr.
- Gallman, R. E. (2000). Economic Growth and Structural Change in the Long Nineteenth Century. *The Cambridge economic history of the United States. Volume 2. The Long Nineteenth Century*. S. L. Engerman and R. E. Gallman. Cambridge ; New York, NY, Cambridge University Press.
- Goodrich, C. (1961). *Canals and American economic development*. New York,, Columbia University Press.
- Haney, L. H. (1908). *A congressional history of railways in the United States*. Madison, Wis., Democrat Printing Co.
- Hunter, L. C. (1949). *Steamboats on the Western rivers; an economic and technological history*. Cambridge,, Harvard University Press.
- Lloyd, E. (1857). *Lloyd's American guide : containing new arranged time tables, so simple and correct that a child can understand them, it being universally acknowledged that all other guide books are so complicated that not one in a hundred can understand them : the population, states, and distances to every place on all the railroad routes in the United States and Canadas : photographic portraits of all the railroad presidents and superintendents--men controlling*. Philadelphia, E. Lloyd.
- National Railway Publication Company. (1868). *Travelers' official railway guide of the United States and Canada*. Ann Arbor, Mich., printed for the National Railway Publication Co. by University Microfilms.
- Oberly, J. W. (1990). *Sixty million acres : American veterans and the public lands before the Civil War*. Kent, Ohio, Kent State University Press.
- Paxson, F. L. (1914). "The Railroads of the "Old Northwest" before the Civil War." *Transactions of the Wisconsin Academy of Sciences, Arts, and Letters* **17**(Part 1): 247-274.
- Poor, H. V. (1860). *History of the railroads and canals of the United States*. New York,, J.H. Schultz & co.
- Primack, M. (1977). *Farm formed capital in American agriculture, 1850 to 1910*. New York, Arno Press.

- Primack, M. L. (1962). "Land Clearing Under Nineteenth-Century Techniques: Some Preliminary Calculations." *The Journal of Economic History* **22**(4): 484-497.
- Rand-McNally (1871). *The Rand-McNally official railway guide and handbook*. Chicago,.
- Seaman, E. C. (1852). *Essays on the progress of nations, in civilization, productive industry, wealth, and population*. New York, Scribner (Reprinted: Johnson Reprint Corp., 1967).
- Stellingwerff, J. and R. P. Swierenga (2004). *Iowa letters : Dutch immigrants on the American frontier*. Grand Rapids, MI, Wiliam B. Eerdmans Pub. Co.
- Taylor, G. R. (1951). *The transportation revolution 1815-1860*, Holt, Rinehart & Winston.
- Taylor, G. R. and I. D. Neu (1956). *The American railroad network, 1861-1890*. Cambridge,, Harvard University Press.
- Thorndale, W. and W. Dollarhide (1987). *Map guide to the U.S. federal censuses, 1790-1920*. Baltimore, Md., Genealogical Publishing Co.
- Thünen, J. H. v. and P. G. Hall (1966). *Isolated state; an English edition of Der isolierte Staat*. Oxford, New York,, Pergamon Press.
- U. S. Congress. (1823-4). Laws of the U.S. VII, 239. *Acts of 1823-24, Ch. 276*.
- United States. Census Office and J. D. B. De Bow (1854). *Statistical view of the United States, embracing its territory, population--white, free colored, and slave--moral and social condition, industry, property, and revenue; the detailed statistics of cities, towns and counties; being a compendium of the seventh census, to which are added the results of every previous census, beginning with 1790, in comparative tables, with explanatory and illustrative notes, based upon the schedules and other official sources of information*. Washington,, Beverly Tucker, Senate Printer.
- United States. Census Office. 10th census 1880., F. A. Walker, et al. (1884). *Tenth census of the United States, 1880. Volume IV: Transportation*. Washington, Government Printing Office (Reprinted: Norman Ross Publishing Inc., 1991).
- United States. Census Office. and J. C. G. Kennedy (1864). *Agriculture of the United States in 1860 : compiled from the original returns of the eighth census, under the direction of the Secretary of the Interior*. Washington, Government Printing Office (Reprinted by Norman Ross Publ., 1990).
- Vanderbilt University. Engineering Center for Transportation Operations and Research (1999). *National Waterway Network*, U.S. Army Corps of Engineers Navigation Data Center
- Whitney, W. D. and B. E. Smith (1911). *The Century dictionary and cyclopedia, with a new atlas of the world*. New York,, The Century co.
- Williams, W. (1851). *A new map of the United States. Upon which are delineated its vast works of internal communication, routes across the continent &c. showing also Canada and the Island of Cuba*, Philadelphia, Lippincott, Grambo & Co.