LAND OWNERSHIP AND THE COMING OF THE RAILROAD TO THE AMERICAN MIDWEST, 1850-60

PROPRIEDADE DA TERRA E O IMPACTO DA CHEGADA DO CAMINHO DE FERRO AO MIDWEST AMERICANO, 1850-60

Jeremy Atack (U. Vanderbilt, USA EUA)
Robert Margo (U. Boston, USA EUA)

The authors are, respectively, Professor of Economics and History, Vanderbilt University, and Research Associate, National Bureau of Economic Research; Professor of Economics, Boston University, and Research Associate, National Bureau of Economic Research

Jeremy Atack é Professor de Economia e História, Universidade de Vanderbilt (EUA) e investigador associado, NBER National Bureau of Economic Research (EUA).
Robert Mago é Professor de Economia, Universidade de Boston, e investigador associado, NBER National Bureau of Economic Research (EUA).

Abstract Resumo
A central question of nineteenth century American economic history which has preoccupied scholars as diverse as Simon Kuznets, Peter Lindert and Jeffrey Williamson is whether economic development was accompanied by increasing inequality. In this paper we study a particular indicator of economic inequality – the incidence of land ownership among adult men – and its relationship to a putative causal factor in nineteenth American growth – the so-called Transportation Revolution, focusing specifically on the diffusion of the railroad.
We will revise the earlier draft and also make use of a newer version of the transportation database that contains distance-adjusted measures of rail access – for example, the share of a county’s land area that is within a specified distance of a railroad. Preliminary analysis indicates that the new measures are better indicators of rail access than the dummy variable indicator used in our earlier work.

Uma questão fundamental da história econômica do século XIX, que tem preocupado investigadores tão diversos como Simon Kuznets, Peter Lindert e Jeffrey
Williamson, é se o desenvolvimento económico foi acompanhado por uma desigualdade crescente. Neste trabalho estuda-se um indicador de desigualdade – a incidência da propriedade da terra entre homens adultos – e a sua relação putativa com um factor causa do crescimento americano no século dezenove – a chamada revolução dos transportes, com ênfase sobre a questão específica da difusão dos caminhos de ferro.

Faz-se uma revisão de uma análise anterior e usa-se uma nova versão de uma base de dados sobre transportes que usa medidas de distâncias ajustadas pela facilidade de acesso ao caminho de ferro - por exemplo, a fração de terrenos de um concelho que se encontram a uma certa distância da linha de caminho de ferro.

Os resultados obtidos indicam que as novas medidas são melhores indicadores do acesso ao caminho de ferro do que os indicadores anteriormente usados.
Land Ownership and the Coming of the Railroad to the American Midwest, 1850-60

Jeremy Atack, Robert A. Margo

This is a revised version of a paper given at the conference on “Railroads in Historical Perspective”, Foz Tua, Portugal, October 2011. An earlier version was also given at the “New Perspectives on Wage and Wages” conference in Hackenberga, Sweden, October 2010. Comments from Stan Engerman, Tim Hatton, and participants at the Foz Tua and Hackenberga conferences are gratefully acknowledged.

A central question of nineteenth century American economic history which has preoccupied scholars as diverse as Simon Kuznets, Peter Lindert and Jeffrey Williamson is whether economic development was accompanied by increasing inequality. In this paper we study a particular indicator of economic inequality – the incidence of land ownership among adult men – and its relationship to a putative causal factor in nineteenth American growth – the so-called “Transportation Revolution”, focusing specifically on the diffusion of the railroad.¹ Land in nineteenth century America was the single most important source of wealth. In the agriculturally-based economy of nineteenth century America, access to land was a critical determinant of economic and social progress for most of the population, particularly in the Midwest where the farm family provided the bulk of the farm labor. Moreover, the ownership of that land in one form or another was the quintessential “American Dream”, the nation’s principal metaphor of upward mobility. In an economy dominated by agriculture such as the nineteenth century United States, owner-operator status, the final step on the agricultural ladder, was central to the wealth accumulation process and a variety of institutions, most notably the various nineteenth century land acts, were

¹ Holding constant the distribution among those owning land, conventional measures of inequality, such as the Gini coefficient or the Theil index, are decreasing in the share of owners.
structured in such a way as to facilitate the transfer of public land into a wide range of private hands. Even for non-farm households, land was an important store of wealth due to the limited variety of other types of physical and financial assets. The ease or difficulty of acquiring land may also have affected the course of development. The early onset of fertility decline in the nineteenth century, for example, has been linked to variations across space and time in the ease of acquiring land. According to Stanley Engerman and Kenneth Sokoloff, widespread land ownership facilitated a relatively equal distribution of political power, itself associated with institutions (for example, the franchise) that may have been conducive to long run growth.

Nevertheless, published census statistics on tenancy collected from 1880 onwards showed relatively high and rising tenancy rates in the Midwest where land had been settled most recently and under a public policy which increasing emphasized ease of access to this resource (Goldenweiser and Truesdell 1924). Several scholars have also estimated tenancy rates at earlier dates (see, for example, (Bogue 1963; Cogswell 1975; Winters 1977; Attack and Bateman 1987; Atack 1988, 1989)) by comparing individual farm values with the value of real estate reported by the farm’s operator. Unfortunately, these data are available for relatively few locations and only back to 1850—the earliest date at which the census reported the value of real estate that individuals held. They are, however, informative. Tenancy rates rise over time throughout the period for which we have data. In this paper, we look at the real estate wealth that a random sample of individuals held by county and its relation to the pattern of railroad expansion.

Elsewhere, we have reported on how the coming of the railroad influenced land improvement and raised land values. Each of these different forces then plays a role in rising tenancy and reduced upward economic mobility.

Economic historians have long acknowledged that the transportation revolution was an important causal factor in economic growth in nineteenth century America. Although the revolution took many modes, arguably the most visible was the “Iron Horse” – the railroad. Previous work has suggested that the railroad lowered the cost of moving goods, information and people from one location to another, thereby positively affecting growth in a number of ways – for example, by encouraging greater division of labor in manufacturing (Atack, Margo, and Haines, 2011a); by increasing the pace of urbanization which may have raised aggregate efficiency through agglomeration effects (Atack, Bateman, Margo and Haines, 2010); and by encouraging improvements in agricultural land, thereby bringing new resources into production (Atack and Margo, 2011b).

While the coming of the railroad promoted economic development, we show
that it also appears to have reduced the proportion of adult men owning land. The negative effect is strongest among men between the ages of 20 and 39 – the prime ages at which the gradient of ownership and age was steepest – particularly among those in agriculture (Cogswell 1975; Wright 1988; Atack 1989). We also provide evidence, albeit tentative, on a plausible causal mechanism for this finding through the impact on the railroad on farm size. In those counties that acquired rail access in the 1850s, average farm size was higher in 1860 suggesting that “minimum efficient scale” in agriculture increased along with the improved trade opportunities created by the railroad. Larger farms were more expensive; therefore it took longer in these areas to move up the agricultural ladder after the railroad came.

Our analysis of the relationship between land ownership and rail access draws on a new body of data documenting the diffusion of railroads and other transportation media in nineteenth century America derived from digitized historical maps using GIS software (Atack, Bateman, Haines, and Margo 2010). The primary database provides information on transportation access at the county level at census year intervals. For the purposes of this paper we have linked the transportation database to extracts of adult men, ages 20 to 59, from the 1850 and 1860 IPUMS samples. The IPUMS provides individual level information on ownership of real estate, occupation, and various demographic variables, which are used in our econometric analysis.

The analysis in the paper focuses on the Midwest in the 1850s. We restrict our attention to the Midwest for two reasons. First, it was in the Midwest where the bulk of the growth in the rail network occurred in the 1850s making the effects easier to identify. Second, we have developed an identification strategy that is particularly plausible and appropriate for the Midwest in this period (see below). However, analysis of the entire region is complicated by the fact that many county boundaries were redrawn during the period while other counties did not even exist until the 1840s. To minimize these complications, we have restricted our analysis to 278 counties that were in existence by 1840 and which did not change their boundaries during the 1850s.

Our primary measure of rail access is a dummy variable indicating the presence of a rail line within a county. All of the counties in our sample lacked rail access in this sense in 1850 but some would gain access by 1860. Our primary econometric approach is difference-in-differences analysis which compares the change in outcomes between 1850 and 1860 in a set of “treatment” counties – those that gain rail access between 1850 and 1860 – versus a set of “control” counties – those that did not have rail access (as we have defined it) before the Civil War. Other measures of access (such as miles of railroad in the county or
proximity of the centroid of the county to a rail line) yield similar results.

Using this approach we find that rail access reduced the likelihood of land ownership. The effect is small (and insignificant) at the sample means but is much larger (and significant) when we control for individual characteristics and various county level factors associated with the coming of the railroad. It is especially strong among men ages 20-39 who worked in agriculture – controlling for various factors, such men were about 7 percentage points less likely to own real estate in the treatment counties compared with their counterparts in the control counties.

To explore the causal mechanism behind our result we use county level data from the 1860 census on the distribution of farm sizes. Conditional on various factors, counties with rail access in 1860 had relatively fewer small farms than counties in the control group. The negative effect of rail access on farm size is consistent with the hypothesis that were some economies of scale in northern agriculture before the Civil War, and that improved access to trade led to a rise in the “minimum efficient scale” in farming. Larger farms were more costly to acquire than smaller farms and, we surmise, it therefore took longer to move up the agricultural ladder.

We supplement our basic analysis with robustness checks and additional analyses. The principal robustness check addresses a key assumption in our DID analysis 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 effects of the railroad would be biased. In particular we are unable to control for landownership in 1840 and our analysis of the farm size data is cross-sectional (because such data were not published in 1850). It is easy to imagine, therefore, that our results are contaminated by omitted variable bias. To address this concern we re-analyze the 1860 data using an instrument variable (IV) derived from various 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 and well in advance of when we observe the possible effects of the railroad on land ownership. Our IV estimates are significantly larger in absolute value suggesting that if anything, we are understating the impact of the railroad.

THE TRANSPORTATION REVOLUTION

The United States experienced a “transportation revolution” in the nineteenth century (Taylor 1951; Goodrich 1961; Haites, Mak et al. 1975). The elements
of the revolution are well known and include dredging and other improvements to harbors and natural inland waterways, improved all-weather roads for travel by wagon, the building of canals, the marine application of steam, and last, and perhaps of greatest importance, the diffusion of the steam railroad. In this paper, we focus on the railroads rather than other elements of the transportation revolution because, as a practical matter, landownership data from the federal census is available only from 1850 onward. Consequently, landownership information only overlaps to any significant extent with the diffusion of the railroad. For example, the bulk of canal building had been completed by the early 1840s and the western river steamboat and Great Lakes shipping were also fully in place before 1850. Even the steam railroad was well established along the East Coast before 1850. However, the American Middle West was virtually untouched by rail until the late 1840s and the big expansion into the region took place in the 1850s.

American steam railroads followed fast on the heels of the British innovation by Robert Stephenson’s Stockton and Darlington Railway in 1826, emerging first as a competitive response by Atlantic port cities to the growth of the port of New York following the opening of the Erie Canal (Condit 1980). Thus, for example, by 1827, Baltimore had chartered the Baltimore and Ohio Railroad (Stover 1987) and South Carolina had chartered a line between Charleston and Hamburg, just across the Savannah River from Augusta, Georgia (Phillips 1908). Other cities such as Boston soon followed (Poor 1970). By 1840 some 2,800 miles of track were in operation (Carter, Gartner et al. 2006 Series Df874), most of it in New England, the mid-Atlantic, and South Atlantic states, and almost all of it covering short distances with minimal interconnections. Further expansion in mileage took place in the 1840s, much of it again in New England, and the Middle Atlantic states. By 1850, about 9,000 miles of railroad track were in operation (Carter, Gartner et al. 2006 Series Df874).

The 1850s witnessed a substantial wave of rail expansion (Stover 1978). Railroad mileage in operation more than tripled, much of it being built in the Midwest. Indeed, between 1853 and 1856, more than half the track miles built in the U.S. were built in the Midwest and, in 1856, this share reached a remarkable 75 percent, of which almost 40 percent was built in just one state, Illinois (Fishlow 1965; Carter, Gartner et al. 2006 Series Df882). This railroad network was, however, less extensive than the number of miles might indicate as it was split between at least seven different gauges (see (Taylor and Neu 1956; Puffert 1991)), several of which were incompatible with one another (especially between North and South) thus preventing through-haulage. Most of the track in the Midwest was, however, standard gauge (4’ 8 ½”) while that in Ohio was somewhat wider (4’ 10” “Ohio” gauge) but still compatible with the standard
gauge. Outside of Missouri, only one line, the Ohio and Mississippi Railroad which operated track on a 72” gauge from Cincinnati to St. Louis was incompatible, with locomotives and rolling stock operating in the rest of the region. In Missouri, the railroads were divided almost 50:50 between standard gauge in the northern part of the state and a 66” gauge in the southern parts that was incompatible with gauges elsewhere (Taylor and Neu 1956).

The first railroad to operate in the Midwest was the Erie and Kalamazoo Railroad, chartered in 1833, which began steam operation in 1837 between from Port Lawrence, Michigan (now named Toledo, Ohio) at the western end of Lake Erie to a point on the Kalamazoo River which would then (theoretically) provide a link across the state to Lake Michigan (Dunbar 1969; Meints 1992). Shortly thereafter the Mad River and Lake Erie railroad began operation from Sandusky on the shores of Lake Erie in Ohio southwards towards Springfield and Dayton. Expansion of the Midwestern railroad network, however, proceeded slowly until the late 1840s when the pace of construction began to pick up before exploding in the 1850s. Our focus on the Midwest here is explained in part by this dramatic surge in railroad construction in the region precisely at a time when we have more economic, social and demographic information available but also because, as we explain below, we have particularly detailed information about railroad construction in this region during this same period.

Economic historians have devoted considerable attention to measuring the impact of railroads on total output, or what Fogel ((1964); see also (Fishlow 1965; Williamson 1974; Kahn 1988)) called the “social savings” of the railroad. Attention has also been paid to the “backward linkages” that railroad expansion created with other manufacturing sectors (Fogel 1964); whether railroads were “built ahead of demand” (Fishlow 1965); and the impact of transportation improvements on regional economic development (Williamson 1974). Our goal in this paper is to add to this literature by examining another “treatment effect” of the railroad, that on land ownership. To measure this treatment effect requires a data set linking the diffusion of railroads to data on wealth. After surveying the literature on nineteenth century wealth accumulation, we describe our data set and estimation strategy.

WEALTH ACCUMULATION IN NINETEENTH CENTURY AMERICA

Beginning with the eighth census in 1850, the federal government began collecting information on the value of real estate held by individuals. In 1860, these
data were supplemented by questions concerning the value of personal estate in excess of $100 and the questions were repeated in 1870 before being dropped from the census in 1880. Here, we make use of the data on real estate ownership from 1850 and 1860 by individuals who are included in the 1 percent IPUMS sample (http://usa.ipums.org/usa/). These have then been linked to our transportation database.

Economic historians have intensively examined the antebellum wealth data (Pope 1986). With few exceptions, the focus of their work has been on establishing and measuring the correlations between wealth and various individual or household-level characteristics and examining inequality. In contrast, our primary interest in this paper involves the impact of a factor external to households – the railroad – on the incidence of land ownership.

There is no question that land was the principal asset in nineteenth century America. For example, the 1860 census reported the assessed value of real estate at almost $7 billion compared with somewhat more than $5.1 billion for personal estate, a figure which also included the value of slaves, and less than $500 million in bank notes and deposits (United States. Census Office. and Kennedy 1862 193-4) Land was both a store of value and a productive factor. As a savings vehicle individuals would invest over the life cycle as a way of smoothing consumption and to provide consumption in old age (Atack and Bateman 1981). For farmers ownership of land was closely tied to moving up the agricultural “ladder”. Absent any inheritance individuals started on the lower rung of the ladder – wage labor – moving up to some type of tenancy followed by owner-operator status (Atack 1989). For both non-farmers and farmers, therefore, land ownership was strongly related to age. The likelihood of ownership rose with age before reaching a peak. In the data analyzed in this paper, for example, ownership was a steep function of age between age 20 and 39 (see the next section).2

In this paper we concentrate primarily on establishing an empirical link between the railroad and landownership as there are many possible causal pathways and our ability to distinguish among them is very limited. For example, land speculation certainly was common in the antebellum Midwest. From our previous work, we know that the coming of the railroad raised land values and also increased the incentive to improve land (Atack and Margo 2011b). It is possible that once an area became connected to the wider world through the railroad, speculative land acquisition increased. In our data this could register as a decrease in landownership if speculators were, on average, non-residents because, by construction, our data pertain to the real estate holdings of individu-

---

2 Previous work on nineteenth century wealth accumulation also reveals impacts of ethnicity, race, and literacy status. In our primary analyses (see below) we control for race and literacy status as these are pre-determined before the “treatment” of the railroad. We do not control for place of birth initially (e.g. ethnicity) as this arguably is endogenous – the coming of the railroad may affect the likelihood of in-migration. As we discuss later in the paper when we control for place of birth the treatment effect of the railroad is diminished slightly.
als who were resident in the county at the time of the census.

That said, the primary productive use for land at the time was in agriculture and the vast majority of farms were “one farm – one owner” – that is, owner-operated. But not everyone who worked in agriculture was a farm owner – owner-operator status was the final rung, as noted above, on the agriculture ladder.

One way in which the coming of the railroad might have slowed progress up the agricultural ladder, and thus lowered the incidence of landownership, was by affecting the optimal size of land holdings. Ever since Adam Smith, economists have known that improvements in transportation can allow business enterprises to take advantage of economies of scale. Indeed, recent work with our transportation database has shown that, in the case of manufacturing, the coming of the railroad led to an increase in the share of establishments that were “factories’ – that is, firms which employed a sufficiently large number of workers that division of labor in their tasks was likely.

Although the prospects of economies of scale in northern agriculture before the Civil War may have been limited, it does not follow that there were none to be exploited. Certainly a wide range of farm sizes were observed – the census in 1860, for example, reported on farms as large as 500 -1000 acres, even in the Midwest (Anderson 1929; Gates 1945; United States. Census Office. 1990). To be sure, these figures include unimproved acreage as well as land in active production but our point is simply that some farms were large enough to routinely required hired labor alongside that provided by the farm household. Moreover, there may have been important scale effects in agricultural mechanization, which was also taking place at this time and stimulated in part by the increased ease of access to wider markets.

As we document later in the paper, counties that acquired rail access in the 1850s had relatively fewer small farms in 1860 than counties without rail access. If these farms remained under individual ownership then, in effect, some land consolidation must have taken place. A larger farm would be more productive (because of economies of scale) but also more costly. Given the imperfections of the mortgage market at the time, it is plausible that individuals would take longer therefore, to move up the agricultural ladder.

THE DATA

We use our 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. Our GIS-based methodology has substantial advantages of accuracy and consistency over earlier approaches which have generally involved matching historical transportation maps to county boundary maps by hand (and eye).³

Our Midwestern railroad transportation database covers seven states: Illinois, Indiana, Iowa, Michigan, Missouri, Ohio and Wisconsin. Data for five of these states (Indiana, Illinois, Michigan, Ohio and Wisconsin) are based upon a series of maps and a data appendix prepared over ninety years ago at the University of Wisconsin by Professor Frederick L. Paxson (1914) and his students. They were able to use extant contemporary travel guides published between 1848 and 1860 as their principal sources of information.⁴ Since then, unfortunately, many of these travel guides which were cheaply printed ephemera seem to have been lost or disintegrated. This is unfortunate as there is little doubt that they are, or rather would have been, the single most valuable resource for our research. These guides served the needs of the traveling public by providing up-to-date route maps and timetables. Consequently, 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 (that is, the most accurate, complete and useful) would have continued publication for an extended period of time.⁵

The data for Iowa and Missouri are taken from digitized contemporary state maps from the on-line David Rumsey Map Collection (http://www.davidrumsey.com/) for 1856 (Colton 1856a; Colton 1856b), 1857 (Colton 1857) and 1859 (Mitchell 1859b, a) supplemented by national maps from the Library of Congress collection for 1858 (Sage 1858) and 1860 (Colton 1860). There were no railroads in either of these states prior to 1856.

³ See, for example, (Craig, Palmquist et al. 1998) 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, Bateman et al. 2010)

⁴ These travel guides first appeared in the 1840s and include Disturnell’s Guide (Disturnell 1846), Doggett’s Gazetteer (Doggett 1847), Appletons’ Guide (Appleton D. and Company. 1848), Dinsmore’s Guide (Cobb 1850), Lloyd’s Guide (Lloyd 1857), Travelers’ Guide (National Railway Publication Company. 1868) and The Rand-McNally Guide (Rand-McNally 1871). 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, 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://cprr.org/Museum/Books/1_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

⁵ However, as we note below, these sources do not seem to have been infallible—certainly at any moment of time.
Our Midwestern transportation database has been supplemented by a more general but (as yet) temporally less detailed GIS transportation database created from geo-referenced digitized historical maps showing canals, navigable waterways and railroads at different benchmark dates. The canal data were taken from Poor (1860) and Goodrich (1961) as well as digitized maps from the Library of Congress “American Memory” web site such as the maps prepared by Williams (1851) and Disturnell (Burr 1850). 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 (1949) and contemporary newspaper accounts regarding steamboat service on specific rivers. Since very little change occurred in the canal and river systems after the 1840s until the Corp of Engineers embarked on its ambitious navigation plans beginning with the establishment of an “Office of Western River Improvements” in 1866, we have measured access to water-borne transportation as of a single benchmark date – 1850 – rather than by change over time.

The railroad portion of our national GIS database is based on digitized state-level maps for 1911 from The Century Atlas (Whitney and Smith 1911) since these maps appear to be accurately drawn and the rail network was largely completed by that time but had not yet begun to shrink through closures. This 1911 railroad network was then traced back through time 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). Our implicit assumption in this approach 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.

To measure railroad access we use a simple county-level binary variable, ACCESS (Atack, Bateman, Haines, and Margo 2010; Atack and Margo 2011). This variable takes on the value of 1 if a railroad operated within a county’s boundaries or formed a county’s border. This variable should be viewed with

6 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 the 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.

7 These digitized images were purchased from Goldbug.com but the original source may be found in many libraries around the country including Vanderbilt’s.
caution because it does not “cross” county boundaries—a county could lack access in our sense yet parts of it might be accessible to a rail line in the next county over. We are currently in the process of refining our measures of rail access, for example, to estimate the share of a county area within say five or ten miles of a railroad or the area within say five miles of a railroad station or depot. This work is still ongoing but preliminary work with these data indicates that our results hold up regardless of the measure which we use. For the present, though, the dummy variable is still the best current measure of the historical spread of the U.S. railroad system at the county level for this period. Because counties differ in size we weight observations by land area in square miles.

Our transportation access data have then been linked with the county-level Haines-ICPSR census data (http://www.icpsr.umich.edu/icpsrweb/ICPSR/) using the county F(ederal)I(nformation)P(rocessing)S(tandards) 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 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 completely organized into states and was divided into 623 counties. Because of boundary changes, we have restricted our analysis to those counties that were (1) present beginning in 1840, (2) had the same county boundaries (as determined by square mileage in 1850 and 1860). Taken together these restrictions produce a balanced panel to 278 Midwestern counties. They are distributed as shown in Table 1.

Our individual level dependent variable, landownership, is computed

---

8 Geographic proximity to a railroad stations or depots can only be measured beginning in the mid-1850s when maps begin to record this important information. For example, the first digitized map in the Library of Congress collection to mention stations and depots explicitly is by Ensign, Bridgman and Fanning (1856). Paradoxically, it shows many more named places along the right of way than the map from Dinsmore’s travel guide for same year (Fisher 1856) which one might have thought would show prospective travelers their station destinations. Nor does it appear that the Ensign, Bridgman and Fanning map was overly sanguine in its reporting of railroad stations. Another map published two years later (Sage 1858) shows most of the same stations as Ensign, Bridgman and Fanning plus additional ones along the new rights of way as well as a few others scattered between some of the 1856 stations and depots. Beginning in the 1870s, Rand McNally (Morgan 1969) along with other publishers began to produce “commercial” maps for the convenience of shippers, showing destinations to which freight could be consigned making the reporting of depots and stations more reliable and systematic. However, as a practical matter, however, trains could potentially stop and load almost anywhere (albeit inconveniently)—much as they do in parts of Alaska even today—and stations were built only when freight and passenger volume had passed some critical economic threshold.

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.
Table 1: Distribution of Balanced Panel of Sample Counties

<table>
<thead>
<tr>
<th>State</th>
<th>Number of counties</th>
<th>Percent of Total (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illinois</td>
<td>71</td>
<td>25.6</td>
</tr>
<tr>
<td>Indiana</td>
<td>65</td>
<td>23.5</td>
</tr>
<tr>
<td>Iowa</td>
<td>17</td>
<td>6.1</td>
</tr>
<tr>
<td>Michigan</td>
<td>12</td>
<td>4.3</td>
</tr>
<tr>
<td>Missouri</td>
<td>41</td>
<td>14.8</td>
</tr>
<tr>
<td>Ohio</td>
<td>57</td>
<td>20.5</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>14</td>
<td>5.1</td>
</tr>
<tr>
<td>Total</td>
<td>277</td>
<td></td>
</tr>
</tbody>
</table>

Notes: To be included in the sample, counties must be (1) present in all three census years: 1840, 1850 and 1860; (2) have fixed county boundaries in 1850 and 1860; and (3) either not have rail access in 1850 but gained rail access by 1860 (=treatment group) or did not have rail access in either 1850 or 1860 (=control group).

from the IPUMS for 1850 and 1860. The IPUMS provide nationally representative 1 percent random samples in both census years of the free population which we have then restricted to the seven Midwestern states under consideration here and to males between the ages of 20 and 59, which covers the majority of the male labor force. The IPUMS data include county of residence so it is straightforward to link the census data, by year, to our transportation database. There are a total of approximately 12,800 men in the sample overall, about 60 percent of whom come from the 1860 IPUMS, reflecting the overall population growth and the rapid development of the Midwest in the 1850s. The censuses recorded the (self-reported) value of real estate owned by the individual; if this value is greater than zero, landownership is set equal to one, zero otherwise.

Table 2 shows the mean values of landownership by treatment status in 1850 and 1860, and the change in landownership between 1850 and 1860. The difference in differences estimator compares the change between 1850 and 1860 in the counties receiving rail access with that in the control counties. For the entire sample, the estimated effect is negative but small -- gaining rail access reduces the likelihood that an adult male owned land by 2.2 percentage points. The standard error of the estimate is 2.1 percentage points, so the effect is not statistically significant.

However, although the overall effect is small and not statistically significant, it is clear from the remainder of the table that this overall effect masks big differences across age groups. In particular, the negative effect of the railroad is much larger among younger men, particularly those in the age group 30-39. As
Table 2: Difference-in Differences Estimates of the Effects of Gaining Rail Access: Landownership Among Adult Men, Ages 20-69, in the American Midwest, 1850-60

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Ages 20-59</th>
<th>Ages 20-29</th>
<th>Ages 30-39</th>
<th>Ages 40-59</th>
</tr>
</thead>
<tbody>
<tr>
<td>1850</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>0.450</td>
<td>0.228</td>
<td>0.549</td>
<td>0.671</td>
</tr>
<tr>
<td>Control</td>
<td>0.511</td>
<td>0.248</td>
<td>0.614</td>
<td>0.730</td>
</tr>
<tr>
<td>Treatment - Control</td>
<td>-0.061</td>
<td>-0.020</td>
<td>-0.065</td>
<td>-0.059</td>
</tr>
<tr>
<td>1860</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>0.443</td>
<td>0.202</td>
<td>0.521</td>
<td>0.691</td>
</tr>
<tr>
<td>Control</td>
<td>0.526</td>
<td>0.266</td>
<td>0.660</td>
<td>0.748</td>
</tr>
<tr>
<td>Treatment - Control</td>
<td>-0.083</td>
<td>-0.064</td>
<td>-0.139</td>
<td>-0.057</td>
</tr>
<tr>
<td>Difference in Differences</td>
<td>-0.022</td>
<td>-0.044</td>
<td>-0.074</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Figures show percent owning real estate (value of real estate > 0). N = 18,471. Sample consists of males, ages 20-59 from 1850 and 1860 IPUMS linked to Attack-Margo MW transportation database (278 counties). Treatment: county of residence gains rail access between 1850 and 1860. Control: county of residence does not have rail access before Civil War. County means weighted by square miles.

can be seen by examining the sample means, it is between ages 30 and 39 that the gradient between age and landownership was steepest. For men in this age group, the coming of the railroad is associated with approximately a 7 percentage point decrease in the likelihood of owning land, or about a 15 percent reduction in the likelihood of land ownership.

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 rail access as far as our dependent variable is concerned. Compared with the control group, counties that gained rail access in the 1850s had a lower proportion of adult men overall who owned any real estate in 1850.

This lower rate of landownership in the treatment counties is a manifestation of a broader pattern—railroads did not arrive randomly in counties. In general, railroad promoters and investors in the Midwest paid close attention to the economic characteristics of counties because the profitability of the railroad depended upon them (Gates 1934; Fishlow 1965). This is illustrated in Table 3, which shows the correlates of gaining rail access in the 1850s using a simple linear probability model applied to our 1860 sample. The dependent variable
Table 3: County-Level Correlates of Gaining Rail Access in the 1850s: Adult Men, Ages 20-39, Living in 277 Midwestern Counties

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Rail Access = 1</th>
<th>Rail Access = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.371* (0.159)</td>
<td>1.214 (0.561)</td>
</tr>
<tr>
<td>Log (Agricultural Yield in 1850)</td>
<td>0.292* (0.094)</td>
<td>0.290* (0.094)</td>
</tr>
<tr>
<td>Percent Wheat in 1840</td>
<td>0.121 (0.239)</td>
<td>0.117 (0.237)</td>
</tr>
<tr>
<td>Δ in Percent Wheat, 1850-40</td>
<td>0.176 (0.168)</td>
<td>0.175 (0.167)</td>
</tr>
<tr>
<td>Percent Urban in 1840</td>
<td>0.632* (0.295)</td>
<td>0.744* (0.295)</td>
</tr>
<tr>
<td>Ln (Population Per Square Mile in 1840)</td>
<td>0.037 (0.049)</td>
<td>0.037 (0.049)</td>
</tr>
<tr>
<td>Δ in Percent Urban, 1840-1850</td>
<td>0.493* (0.236)</td>
<td>0.491* (0.235)</td>
</tr>
<tr>
<td>Δ in Ln (Population Per Square Mile, 1840-1850)</td>
<td>0.034 (0.059)</td>
<td>0.034 (0.059)</td>
</tr>
<tr>
<td>Water Transportation Dummies Included?</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>State Dummies Included?</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Individual Covariates</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.267</td>
<td>0.270</td>
</tr>
</tbody>
</table>

Source: see text and Table 1.
Notes: Standard errors clustered at the county level. Columns 2, 3: Dependent variable = 1 if county gains rail access by 1860 (treatment group), 0 otherwise (control group). Column 4: Dependent variable = 1 if individual owns real estate, 0 otherwise. 1860 sample is used for regressions in columns 2 and 3; 1850 sample is used for regression in column 4. Agricultural Yield: value of agricultural output/(improved + unimproved acres in farms). 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 the number of square miles in the county.
Standard errors are shown in parentheses. *: significant at 5 percent level.

equals 1 if the individual resided in a county which gained rail access in the 1850s – that is, the county is in the treatment group – and zero if not – the county is in the control group.¹⁰

¹⁰ Given that the 1860 IPUMS was produced by random sampling by using the individual as the unit of observation we are effectively controlling for population size.
In the first column, we include control variables measured at the county level as well as dummy variables for states. State dummies are included because there are numerous possible determinants of gaining rail access that are difficult or impossible to control for directly, given the available data. Three of our variables are indicators of agricultural productivity: the logarithm of the value of agricultural output per acre; the percentage of the value of total agricultural output reported in the 1840 census accounted for by wheat production; and the change in the percent wheat, as just defined, between 1840 and 1850.\footnote{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 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).} We also include the percent urban in 1840, the log of population density in 1840, and changes in these variables over the 1840s. Finally we include dummy variables for access to different modes of water transportation (river, canals, and abutting the Great Lakes).

The results show that counties that were already fairly urban in 1840 or which became more urbanized over the 1840s, as well as counties that had high agricultural productivity (as measured by yields) were more likely to gain rail access over the 1850s, other factors held constant. When we include individual controls for age, literacy status, ethnicity, race, and disability (column 2), our substantive findings with regard to the impact of county characteristics are unchanged.

Because gaining rail access clearly was not a random event, it is important that we control for these important factors in our difference-in-differences analysis (Table 4) in so far as they may have had independent effects on the change in percentage of improved acreage between 1850 and 1860. Column 1 reports our base DID estimate for all males in the IPUMS aged 20-59; by construction, this is the same as in Table 1, panel A. In columns 2 and 3 we add individual and county controls from Table 3. Adding these controls increases the magnitude and statistical significance of the treatment effect. Thus, for example, in column 4, including both county and individual controls, the coming of the railroad reduces the incidence of land ownership by 4.6 percentage points.

In the remaining columns we re-estimate the regression with the full set of controls conditional on an agricultural occupation. The negative impact of the rail access on landownership is slightly larger for farmers than non-farmers for the overall sample, but the difference in the coefficients is very small. However, when we re-estimate for the younger age group, 20-39, the effect is considerably larger for farmers than for non-farmers.
<table>
<thead>
<tr>
<th>Covariates</th>
<th>Sample</th>
<th>Ages 20-59</th>
<th>Agricultural Occupation = 1</th>
<th>Agricultural Occupation = 0</th>
<th>Agricultural Occupation = 1, Ages 20-39</th>
<th>Agricultural Occupation = 0, Ages 20-39</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gain Rail Access?</td>
<td>None</td>
<td>-0.022</td>
<td>-0.046</td>
<td>-0.048**</td>
<td>-0.079</td>
<td>-0.021</td>
</tr>
<tr>
<td></td>
<td>Individual</td>
<td>-0.043*</td>
<td>-0.046</td>
<td>-0.048**</td>
<td>-0.079</td>
<td>-0.021</td>
</tr>
<tr>
<td></td>
<td>Individual + Conty</td>
<td>-0.046</td>
<td>-0.046</td>
<td>-0.048**</td>
<td>-0.079</td>
<td>-0.021</td>
</tr>
<tr>
<td></td>
<td>Individual + Conty</td>
<td>-0.046</td>
<td>-0.046</td>
<td>-0.048**</td>
<td>-0.079</td>
<td>-0.021</td>
</tr>
</tbody>
</table>

All regressions include dummy variables for year and county. Individual: dummy variables for age group (30-39, 40-49, 50-59), Literacy rate, place of birth, disability status, and interactions between age group and year, and age group and treatment status. County: (year = 1860) x (percent urban in 1840, change in percent urban between 1840 and 1850, log of pop density in 1840, change in log pop density between 1840 and 1850, percent wheat in 1840, change in percent wheat between 1840 and 1850, dummy variables for presence of canal, river, or abutting Great Lakes, state). Agricultural occupation = 1, if occupation code for farmer, farm laborer, or living on a farm and laborer not otherwise classified. Observations weighted by square miles in county of residence. Standard errors clustered by county. *significant at 5 percent level; **significant at 10 percent level.
Table 5: Instrumental Variables Estimates

<table>
<thead>
<tr>
<th></th>
<th>Own = 1</th>
<th>Own = 1</th>
<th>% Farms</th>
<th>% Farms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Own = 1</td>
<td></td>
<td>Under 50</td>
<td>Under 50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Acres</td>
<td>Acres</td>
</tr>
<tr>
<td>OLS, Gain RAIL Access?</td>
<td>-0.040*</td>
<td>(0.015)</td>
<td>-0.064*</td>
<td>(0.017)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First Stage, coefficient of</td>
<td>0.227*</td>
<td>(0.058)</td>
<td>0.232*</td>
<td>(0.058)</td>
</tr>
<tr>
<td>Survey IV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSLS, Gain RAIL Access</td>
<td>-0.189*</td>
<td>(0.086)</td>
<td>-0.128*</td>
<td>(0.123)</td>
</tr>
</tbody>
</table>

1860 sample only (N = 11,355). Own = 1 if positive value of real estate reported. Covariates in ownership regression include individual variables (age, place of birth, literacy, race, and disability status) and county variables (state, percent urban in 1840, percent urban 1840-50, log of population density in 1840, change in log of population density 1840-50, percent wheat in 1840, change in percent wheat 1840-50, log of agricultural yield (value per acre) in 1850, dummy variables for presence of canal, navigable river, or abutting Great Lakes). % of Farms under 50 acres: computed from farm size distribution data in 1860 census. First stage: coefficient of survey IV in regression of rail access; full set of individual and county covariates. TSB: second stage, two stage least squares coefficient of rail access. *: significant at 5 percent level.

We explore one possible explanation for the negative effect of rail access on landownership in agriculture – an increase in the “minimum efficient scale” of farms when transportation access improves due to rail access. Although the general opinion among economic historians is that there were limited opportunities for economies of scale in northern agriculture before the Civil, limited does not necessarily mean “constant returns to scale.”

If transportation access led to an increase in farm size, we would expect to see a shift towards larger farms at the county level when the county gained rail access. Unfortunately, we cannot use difference-in-differences methods to investigate this hypothesis because the 1850 census did not report the distribution of farm sizes. However we can estimate a cross-sectional regression using 1860 data, because the 1860 census did report the distribution of farm sizes at the county level.

Row 1 in Table 5 shows the coefficient of rail access in a regression of the share of farms with fewer than 50 acres. The regression is estimated over our IPUMS-linked sample and includes the full set of individual controls in the other columns along with county controls. The coefficient of rail access is -0.064, indicating that, controlling for other factors, the share of small farms decreased by about 6.4 percentage points when a county gained rail access.

The negative effect of rail access on small farms is consistent with the hy-
pothesis that the coming of the railroad led to an increase in farm size, possibly because such farms could capture economies of scale as they took advantage of increased opportunities for trade. But larger farms were, by definition, more expensive farms.

An increase in optimal farm size would not necessarily result in a smaller proportion of landowners among adult men – there could be multiple owners per farm—the census schedules, for example, occasionally listed more than one name for the operator of a farm. However, while such arrangements may have existed, they appear to have been uncommon; rather, the typical organizational structure was “one farm, one owner”. Although larger farms would have likely used less labor per acre, total labor demand would not necessarily decline – and, indeed, if we estimate a difference-in-differences regression of the probability that an adult man in our sample held an agricultural occupation, we find little or no evidence of a railroad effect. Thus the likely explanation of the railroad effect is that, with the coming of the railroad, it took longer for men to move up the agricultural ladder and, thus, at a point in time, the share of men owning land declined. Indeed, the 1937 Presidential commission looking into rising tenancy placed the blame for the persistent long run trend squarely on the rising entry costs in the face of imperfect capital markets (United, Gray et al. 1937).

We have performed a robustness check on our result because, although we can control for various correlates of gaining rail access 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 the observable determinants of gaining rail access. This problem is of particular concern here because the 1840 census reported no statistics on land ownership making it impossible for us to control for the pre-1850 trend in the fraction of adult men owning land. If this trend is correlated with the coming of the railroad – likely, in view of the fact that the treatment and control counties clearly differed in ownership patterns in 1850 – the DID treatment estimates will be biased, even when we control for the factors associated with the coming of the railroad. Further, and as just noted, we have no data on the farm size distribution in 1850 and thus are limited to the cross-sectional analysis discussed above.

One way to deal with this problem is to re-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. For our IV, we have used whether or not a county lay along a straight line between the two termini of federal government transportation surveys conducted during the 1820s and 1830s. Using such a variable which isolates plausibly exogenous variation in rail access such as we propose
here 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 but also arguably for railroads, assigns an important role to the government in promoting these advances (Goodrich 1961). One such important source of government assistance was the assignment 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 the survey law predates even the Stockton to Darlington Railway in England), 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. Soon thereafter, railroad routes came to dominate the surveys with perhaps as many as 61 such surveys being made before the law was repealed by Andrew Jackson’s administration effective in 1838 (Haney 1908, p. 277).

Our specific instrumental variable is derived from the government surveys 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 line. 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 is not otherwise specified). In some cases, we inferred both 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 this is the shortest distance between the two points.12

These U.S. Army Corp of Engineers’ surveys provided valuable information to the general public and prospective railroad promoters alike regarding topog-

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).
raphy 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, 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.” Moreover, as 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]. Indeed, many of the interstate highways today follow these same routes. Presumably, therefore, the surveys were found very useful by the railroad engineers charged with the construction of a line 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 5 reports our estimates using this instrumental variable for the full sample of men (ages 20-59). 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 of the effect of rail on landownership is considerably larger than the OLS estimate, and is statistically significant. We also find a larger negative effect of rail access on farm size, but in this case the IV coefficient is imprecisely estimated.

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 the incidence of land ownership. Real estate was the principal asset in antebellum America (as it is and has been in many parts of the world), and the proportion of adult men owning land is directly related to wealth inequality.

Our primary finding is that that the coming of the railroad was associated with a decrease in the proportion of adult men owning land which later studies
would relate to rising land costs and imperfect capital markets (United, Gray et al. 1937). A significant negative effect is apparent in a difference-in-difference analysis in which we simultaneously control for factors associated with the coming of the railroad. The negative effect is substantially larger when we allow for the coming of the railroad to be endogenous, using an instrumental variable derived from railroad surveys conducted in the 1820s and early 1830s. A plausible mechanism is that the coming of the railroad led to an increase in farm size. Larger farms were more costly than smaller farms; at the margin it took longer for men to move up the agricultural ladder, and some never made it to the top rung. In this way, our analysis shows how a fundamental and far reaching technological change – the diffusion of the railroad – was a factor behind the Kuznets curve in the American case.
Railroads in historical context: construction, costs and consequences

BIBLIOGRAPHY


