RAILROADS AND LOCAL ECONOMIC DEVELOPMENT:
THE UNITED STATES IN THE 1850S

Michael R. Haines
Robert A. Margo

Railroads and Local Economic Development: The United States in the 1850s
Michael R. Haines and Robert A. Margo
NBER Working Paper No. 12381
July 2006
JEL No. N51, N71

ABSTRACT

We use county and individual-level data from 1850 and 1860 to examine the economic impact of gaining access to a railroad. Previous studies have found that rail access was positively correlated with the value of agricultural land at a point in time, and have interpreted this correlation as evidence that rail access chiefly benefitted agricultural land owners in the manner predicted by the Hekscher-Ohlin or Von Theunen models. We use a difference-in-difference strategy, comparing changes in outcomes in counties that gained rail access in the 1850s to those that either gained access earlier or did not have access before the Civil War.

Most of the estimated effects are small and the signs are not wholly consistent with either model, under the null hypothesis that agriculture was the chief beneficiary of rail access. For example, we find that rail access appears to have increased urbanization, raised the likelihood of participation in the service sector, decreased agricultural yields, and reduced the share of improved acreage in total land area, opposite to the patterns predicted by either the Heckscher-Ohlin or Von Theunen models.

Michael R. Haines
Department of Economics
217 Persson Hall
Colgate University
13 Oak Drive
Hamilton, NY 13346
and NBER
mhaines@mail.colgate.edu

Robert A. Margo
Department of Economics
Boston University
270 Bay State Road
Boston MA 02215
and NBER
margora@bu.edu

1.0 Introduction

The United States experienced a "transportation revolution" in the nineteenth

century (Taylor 1951). The development of canals and other navigable waterways and,

especially, the railroads, linked together far flung factor and product markets and

stimulated economic growth through division of labor and exploitation of regional

comparative advantage (North 1961; Goodrich 1961; Ransom 1967, 1970; Haites, Mak,

and Walton 1975). Economic historians have measured of the aggregate social savings of

transportation improvements such as railroads (Fogel 1964; Fishlow 1965) but have paid

far less attention to the impact of gaining access at the local economic level, despite the

fact that contemporaries at the time were vitally interested in such effects and that

economic theory provides useful guidance as to what these effects might have been.

In this paper we examine various economic impacts at the local (county) level of

gaining access to a railroad in the 1850s. Our point of departure is previous work (Craig,

Palmquiest, and Weiss 1998; Coffman and Gregson 1998) that has shown that, at a point

in time, proximity to a rail line was positively associated with higher agricultural land

values (see also Fogel 1964; Fishlow 1965). Two inter-related frameworks, the

Heckscher-Ohlin model of international trade, and the Von Theunen model of

agricultural land rents, provide the economic logic to account for the cross-section

patterns. These frameworks predict other economic outcomes besides the rise in land

rents, and it is these ancillary predictions that we primarily focus on.

We measure rail access at the county level using information derived from maps –

essentially, whether a rail line passed through county boundaries in 1850 or 1860 (or

both). Rail access data are linked to county and individual level information on

economic outcomes in 1850 and 1860. We use a "difference in difference" approach, comparing outcomes in a treated group (counties that gain rail access in the 1850s) with a control group, before and after treatment (rail access).

Our analysis yields mixed results on the impact of the rail in light of the two frameworks just mentioned. As in previous work, we find that rail access did lead to higher land prices, both absolutely and relative to farm wages, but the effects were small. We also find that gaining rail access is associated with a decrease in agricultural yields, a smaller share of improved acreage in total land area, a higher proportion urban, and a lower likelihood of participation in agriculture – effects that appear to be opposite to those predicted by either framework.

2.0 Railroads and American Economic Development

Although plans for railroads were first discussed in the United States in the early 1800s, it was not until late 1820s that rail expansion actually took place.[1] The first American railroads were "tramways" using in mining and quarrying, such as the so-called "Granite Railroad" that commenced operations in Quincy, Massachusetts in 1826. But railroads in the modern sense really originate with the struggles of port cities like Baltimore, Boston, and Charleston that lacked adequate inland waterway connections that would enable greater volume of trade with the hinterland. By 1840 some 3,300 miles of track had been laid (of which about 2,800 miles were in operation), the majority of it in New England, the mid-Atlantic, and South Atlantic states, and almost all of it involving trips of short duration ("short line"). Further expansion in mileage took place in the

---

[1] See Taylor (1951, ch. 5) for a concise history of early railroads.

1840s, much of it again in New England, and also in New York. The South and Midwest were largely bypassed during this decade, except for the completion of a rail line linking Savannah and Chattanooga, and a rail line through Ohio from Sandusky to Cincinnati.

It was in the 1850s that the United States experienced its first great wave of rail expansion (Stover 1978). Approximately 22,000 miles of track were laid during the decade, bringing the total mileage on the eve of the Civil War to over 30,000. Although the federal government had been involved in railroad expansion prior to 1850 in an indirect way by providing land surveys free of charge (from 1824 to 1838, when the law authorizing the surveys was repealed), direct subsidies in the form of land grants were first voted in 1850, and later extended several times during the decade. By 1860, in addition to substantial coverage in the Northeast, rail lines crisscrossed Illinois, Indiana, and Ohio, with significant penetration into Wisconsin and Iowa. The South was less well served, but it too experienced substantial growth in rail access in the 1850s. As Stover notes:

> "The decade prior to the Civil War had been one of the most dynamic periods in the history of American railroads. In 1850 a broken skein of short iron lines served the area between Maine and Georgia with a few stray strings of rail connecting the Ohio River and the Great Lakes. By the eve of the Civil War more than 30,000 miles of railroad served quite adequately all the states east of the Mississippi and few areas of substantial population were much removed from the Sound of the locomotive whistle." (1978, p. 232)

Economic historians have devoted considerable attention to measuring the aggregate impact of railroads on nineteenth century American economic growth. Fishlow (1965) focused on the antebellum expansion, particularly on whether railroads were "built ahead of demand" and on total factor productivity growth. Fogel (1964) measured the social savings of the transportation of good by railroad – that is, the savings in

4

resource cost associated with shipping goods by rail as opposed to the next best alternative. In both studies the conclusion was the railroad's impact, while economically significant, was insufficient to deserve the label of "indispensable", as Fogel put it.

Although contemporaries were certainly aware of the aggregate benefits of transportation improvements, many were more concerned with the impact at the local level. Economic historians have not wholly neglected these local effects, although much less attention has been paid to them than to the aggregate social savings.

The two specific studies that motivate our work are Craig, Weiss, and Palmquist (1998) and Coffman and Gregson (1998). Craig, Weiss, and Palmquist used maps to infer whether a county had water or rail access in 1850 and 1860, and then linked their transportation data to census data, which provided information on land values in agriculture and their potential determinants (other than transportation access). Using a hedonic model to motivate their analysis Craig, Weiss, and Palmquist estimated cross-sectional regressions of the log of the average per acre value of farmland on, among other determinants, dummy variables for water and rail access. Controlling for other factors, land values were 15 percent higher in 1850 if the county had rail access, somewhat less (around 8 percent) in 1860. Coffman and Gregson used tax assessment records for Knox County in Illinois in 1855 to measure the impact of distance to a rail line that had recently been completed. Their cross-sectional regression implies that capital gains resulting from rail access equaled about 9 percent of the value of land in the county in 1850.

Both studies interpret their findings in terms of observations from era and modern economic models of the impact of transportation access on local economic activity (see also Atack and Passell, pp. 167, 169). Craig, Weiss, and Palmquist (1998, p. 173) note,

5

for example, that contemporaries "recognized the potential impact of improved

transportation of marketable agricultural production on the value of agricultural land"

and, in some cases, attempted to measure the impact using the rudimentary techniques

then available. Coffman and Gregson (1998, p. 196) explicitly motivate their analysis

using a version of the Von Theunen framework discussed in the next section. However,

neither study investigates the range of predicted impacts that should occur if the

underlying framework that gives rise to increases in land values were correct.

3.0 The Marginal Impact of a Railroad: A Framework for Analysis

In this section we sketch two inter-related frameworks for evaluating the impact

of a railroad on a local economy. The first framework is simply the familiar Hekscher-

Ohlin two sector model of trade, in which there are specific factors to each sector as well

as a mobile factor, labor. The second is the Von Theunen model in which distance to a

central location where trade takes place plays a key role in the determination of land

rents.

As a point of departure, we imagine a local economy with total land area T, a

fraction of which, $\delta A = A$, has been "cleared" or is otherwise available to be used with

labor, L, to producing an agricultural good, $Q_A = F(L_A, T)$. There also exists a local stock

of capital, K, which can be combined with labor to produce a non-farm good, $Q_M = G(L_M, K)$. We assume that F and G are constant returns to scale. Households in this

local economy are endowed equally with amounts of L, K, and T. Initially the local

economy is closed to trade, so equilibrium is defined by local supply equaling local

demand. With conventional utility functions this will ensure an interior solution so that positive amounts of both the farm and non-farm good are produced.

Prior to trade we assume that the local economy has a comparative advantage in producing the farm commodity and would, if it could, import the non-farm commodity. That is, letting p be the external (relative) price of the farm commodity and p* be the local relative price, p* < p – the local economy would experience gains from trade if it could trade with other local economies, near or far.

Next, we imagine that a railroad is built which gives farmers in the local economy access to external trade. Because the relative price of the farm commodity is higher externally, farmers now have an incentive to export their crop and import non-farm goods produced elsewhere.

In the short run, output expands in the farm sector as labor shifts from non-farm to farm production. The magnitude of the increase in output, however, is constrained by the fixed supply of T – there are decreasing returns in the short run. The influx of labor into farm production drives up the marginal product of land. Because both the marginal product of land and the relative price of agricultural goods are now higher, the agricultural yield (output per acre) has increased. As a result, the rental price of land is higher. Conversely, the marginal product of capital declines as labor leaves non-farm production. In the standard model, though, the gross effect on the rental price of land will dominate the net effect on the mobile factor, and the wage relative to the rental price of land will decline.

The Hekscher-Ohlin framework takes the stock of land as given. However, our second framework – the Von Theunen model – makes it clear that the rise in p provides

an incentive to bring additional land into production. To illustrate this point, we assume that each farmer has a single unit of land which is combined with labor to produce the farm good, using a fixed coefficients production technology, $Q = \min (T, L/b)$. Once produced the good must be transported (by wagon, say) to a "central place" where production of the non-farm good and trade (farm for non-farm) occur. The fixed coefficient technology will ensure a linear "bid rent" curve giving the maximum willingness to pay for a unit of land at a given distance d from the central place:

$$R = PQ - wL - cdQ$$

By setting $R = 0$, we define the margin of cultivation, $d^* = (p - wb)/c$. It follows directly that $dd^*/dp > 0$ – the margin of cultivation expands when the relative price of the agricultural good increases, and consequently, land used in agriculture as a share of total land area increases ($d\delta > 0$).

To summarize, our empirical analysis examines the following implications of rail access [d(rail)>0] as derived from these two frameworks:

d(w/r) <0: the wage/rental ratio declines

d(yield) > 0: agricultural yields increase

$d(L_A/L) > 0$: agriculture's share of local labor supply increases

$d\delta > 0$: Land in agriculture/Total land area increases

4.0 Data and Estimation

Our empirical analysis is based on a panel data set of counties for 1850 and 1860. This data set merges the information originally collected by Craig, Palmquist, and Weiss (1998) on railroad access with a revised version of the ICPSR census dataset (Haines and ICPSR 2006) and county-level data on farm wages (Margo 2000). The ICPSR data were augmented by information on urban population, county are, and more complete agricultural census data. Because the wage data are not available for every state (or for every county for states with surviving records) the sample is not fully representative of the antebellum United States in a geographic sense. Within the matched data set we also restrict our attention to counties that did not change land area between 1850 and 1860. The total sample size is 672 counties from 14 states (see Table 1, Panel B).

Margo (2000) collected wage data from the 1850 and 1860 manuscript censuses of social statistics. The unit of observation in the manuscripts in the minor civil division (or, in the case of some cities, the ward); for each county, we constructed an equally weighted average of the sub-county figures. The monthly wage data pertain to contracts in which board was routinely included. However, the census also collected information on the "weekly cost of board to laboring men" from which it is possible to impute a monthly value to board, which we add to the money wage. Hence, our adjusted farm wage is defined as follows:

Adjusted farm wage = reported money wage + 4.3*weekly cost of board

The census also reported the daily wage of common labor, "without board" (that is, contracts in which board was not included). Margo (2000) shows that, once one accounts for the unemployment risk premium associated with daily (as opposed to monthly) work, farm-non farm wage gaps for unskilled labor were negligible within counties on average, consistent with the hypothesis that (free) labor could (and did) move freely between the farm and non-farm sectors at the local level before the Civil War.

Craig, Palmquist, and Weiss used maps to determine whether a county had access to water or rail transportation in 1850 and 1860. Here, "access" means that a rail line or navigable waterway passed through the county boundaries in the given year (or, in the case of water transportation, the county bordered on a navigable waterway, such as the Atlantic Ocean or one of the Great Lakes). Panel A of Table 1 shows that slightly more than 22 percent of the sample counties had access to a railroad, in this sense. Rail access more than doubled between 1850 and 1860 – nearly 45 percent of the sample counties had access to a railroad on the eve of the Civil War.

Rows 3-5 of Panel A show access to the railroad in 1860, conditional on not having access in 1850. Overall, among the 531 counties that lacked access in 1850, 30.1 percent gained access by 1850. Rows 4-5 demonstrate that gaining access was correlated with existing access to water transportation; counties that had access to water transportation were less likely to gain rail access. Below we show that the negative effect of water on gaining rail access survives inclusion of a variety of covariates, as well as state dummies; that is, it is not an incidental consequence of failing to control for other determinants of rail access.

In column 3 of Panel A we weight observations by total population in the county in the relevant census year (in the case of rows 3-5, this is population in 1850). The proportion with rail access in both years increases, an indication that larger counties (in the sense of population) were more likely to have access by 1850 or gain access by 1860 if they did not have it in 1850. A similar effect is not apparent if we weight instead by land area, the implication being that population density was a factor in having access by 1850, or gaining it by 1860. Lastly, the figures in brackets show the proportion of the population with access to either rail or water transportation. In 1850, nearly 71 percent of the American population lived in counties with either rail or water transportation; by 1860, the figure was 82 percent. As the quote from Stover (1978) suggests, the nineteenth century transportation Revolution was largely over by the eve of the Civil War, if not a decade earlier. This fact may help explain why our estimated impacts of rail access are relatively small (see below).

Panel B of Table 1 shows population-weighted rail access by state. In the Northeast, states like Massachusetts and Pennsylvania had near total access by 1860. For the three Midwestern states, access varied substantially before the Civil War. Most of the population of Michigan lived in counties with rail access by 1850, and access grew little during the 1850s. In Indiana, by contrast, only about a third of the population had rail access in 1850 but by 1860, nearly 70 percent. Iowans had no access to rail in 1850, but a third of the state's population did in 1860. The data also show that there was significant variation in rail access in the South in 1850. A little more than three quarters of the population of Georgia resided in counties with rail access in 1850 whereas only 28 percent of the population of Virginia did. But access expanded quite rapidly in some

southern states in the 1850s, almost doubling in Virginia, and more than doubling in Kentucky and Tennessee.

Panel C of Table 1 reports the coefficients of three covariates from linear probability regressions of gaining access to rail by 1860 conditional on not having access in 1850. The covariates are a dummy variable for having water access, percent urban, and the log of population per square mile (population density). The regressions also included state dummies. In column 1 the data (counties) are equally weighted whereas in column 2 they are population weighted. The results in the earlier panels suggest that we should observe a negative effect of water access and a positive effect of population density, and this is confirmed by the regression. We also hypothesized that the presence of a sufficiently large urban population, as indicated by the proportion living in incorporated places of at least 2,500 (the usual census definition of "urban") would enhance the likelihood of gaining access, as the presence of a relatively large "central place" would be complementary to trade – and therefore, provide a convenient place for the railroad to stop to load and unload cargo. This hypothesis is confirmed; an increase in the proportion urban is associated with a positive and highly significant effect on the probability of gaining access. Weighting by population in 1850 slightly attenuates the magnitudes of the coefficients but does not alter their signs. We also estimated regressions excluding the state dummies (not shown) and, again, the substantive findings were not affected. However, F-tests clearly showed that, controlling for the three covariates, the addition of state dummies significantly improved the fit of the regression, indicating that, conditional on the covariates, the likelihood of gaining rail access varied across states.

We investigate the impact of the railroad using a "difference-in-difference" (DID) methodology. Letting y_{it} be an outcome variable and R_{it} be a dummy variable for rail access, with the "i" subscript referring to counties and the "t" subscript referring to years, our basic model is

$$y_{it} = \alpha_i + \beta R_{it} + \delta_t + \varepsilon_{it}$$

where δ is a dummy variable for 1860 and ε is a random error term. Rather than estimate this directly we estimate the equation in first difference form

$$\Delta y = \delta + \beta \Delta R + \gamma$$

where $\gamma = \varepsilon_{i,1860} - \varepsilon_{i,1850}$.[2]

As noted earlier, we excluded two counties that appear to have lost rail access over the 1850s; hence, in our data set, once acquired, rail access is permanent (at least, as of 1860). Therefore, $\Delta R = 1$ if and only if the county did not have rail access in 1850, but acquired it by 1860.

According to the DID method, the "control" counties consist of those that did not have access in 1850 and failed to acquire it as well as counties that had access in 1850. In some specifications (see below) we exclude counties that already had access by 1850 and focus instead on the "population at risk" of obtaining access. The "treatment"

---

[2] Two counties that purport to have gained rail access by 1850 and then lost it by 1860 are excluded. There are seven counties in the sample that claimed to have gained access to water transportation in the 1850s. In the regressions reported in this paper we do not add a dummy for these seven counties but our substantive results would not change if we did or if we excluded them from the sample.

counties are always those that gained access in the 1850s. Thus at issue is whether

outcome variables changed in response to the treatment: access to a railroad. This is a

"before-and-after" method, so we are comparing changes in outcomes between 1850 and

1860 in the treatment group with the analogous changes in the control group.

It would be desirable to include two or more pre-treatment values of y, in order to

If least squares are used to estimate the regression the assumption must be that the

treatment (rail access) and the error term γ are uncorrelated. In standard DID this

assumption can be met in different ways – for example, by pre-selecting the treatment

and control observations so that they are "similar" or by including covariates in the

regression that purge any correlation between rail access and the error term. We have

already shown that Δ R depended on water access, percent urban and population density

in 1850; and, even after controlling for these variables, across states. In certain

specifications, therefore, we include these covariates which, in the context of the

estimation, amount to interactions between the 1860 time dummy and the covariates.

It would be desirable to include two or more pre-treatment values of y, in order to

control for the possibility of pre-existing trends. We cannot include any pre-existing

trends at present, because our data are not yet linked to the 1840 census.[3] However, we

can (and do) include the 1850 values of the outcome variable in some specifications

which, in effect, permits "regression to the mean" in the outcome variables. Table 2

shows the regression results for our quantity outcome variables and Table 3 the results for

our factor price outcome variables.

In Panel A of Table 2 we show the treatment effects of rail access on the percent

urban, again measured as the proportion of the population living in incorporated places of

---

[3] Even if our data were linked to the 1840 census our ability to control for pre-existing trends would be extremely limited, because the 1840 census did not collect information on wages, land values, or acreage in agriculture.

at least 2,500 or more. The idea here is that the percent urban is a proxy for the percent

non-farm; therefore, if the Hekscher-Ohlin framework is correct (and the proportion

urban is, indeed, a reliable proxy) we should observe a negative treatment effect. We do

not – instead, the effect is positive and statistically significant.[4]   The treatment effect is

smaller if we include counties that already had rail access in 1850 in the control group but

still substantial and reliably estimated.

In Panels B and C of Table 2 we report treatment effects of rail access on the log

of agricultural yields, or output per acre. We use two definitions of output. The first

includes the value of crops, orchards and garden produce intended for market. The

second adds to the first the value of slaughtered livestock and household manufactures.

"Acre" here refers to "improved acres". Again, if the Hekscher-Ohlin framework were

correct we should observe positive treatment effects but, instead, the effects are negative,

although statistically insignificant. Counties that gained rail access in 1850 evidently did

not experience rising output per unit of land.[5]

In Panel D we test an implication of the Von Theunen model: the share of land in

agriculture should increase. To be precise, the Von Theunen model predicts that land

---

[4] An alternative interpretation, one consistent with the regression showing that the probability of obtaining
rail access in the 1850s was positively associated with the percent urban in 1850, is that the coming of the
railroad expanded the demand for trade services – a town of a certain size, in other words, facilitated trade.
In future work it may be possible to improve upon this analysis by making use of Craig and Weiss's (1996)
county level estimates of the rural agricultural labor force. These could be combined with estimates of the
urban agricultural and total labor forces, by county, using Craig and Weiss's procedures. However, it
should be noted that Craig and Weiss derived their estimates by assuming agricultural participation ratios
that vary by age, gender, rural and legal status (slave) across states and over time, but not within states.
Thus, for example, a county's number of farm workers will change if the demographic, rural/urban, or
free/slave structure changes but not otherwise. It remains to be seen if this generates sufficient variability
to capture any treatment effects of the rail beyond those measured here by percent urban (or percent slave,
see below).

[5] At present the agricultural output variables are somewhat inconsistently defined; specifically, crop output
is valued at 1860 prices, but orchards, garden, and home manufactures are evaluated in current prices (the
census did not report quantities for these categories of output) and we have yet to adjust these for price
changes. In future work we will be able to restrict the analysis to crop output alone; however, because our
regressions include state dummies, any (state-specific) trends in prices will be controlled for implicitly.

15

under cultivation should increase; if it did increase, this might be a reason why we do not observe an increase in yields. We do not observe land under cultivation, but we do observe improved land. However, we find no evidence of a positive treatment effect – indeed, all of the estimate effects are negative, and several are statistically significant. In interpreting Panel D it is important to keep in mind that, by design, our specification holds total land area constant at the county level.

In Table 3 we report our factor price regressions. We report results for the sub-sample of counties with no rail access in 1850 (results for all counties were similar). The farm wage is adjusted for the imputed value of board, as described earlier. We do not have data on the rental price of land; in its place we use the per acre value of land, computed and adjusted as follows.

The census reported the total value of farms and the number of improved and unimproved acres. We compute the average value per acre as the ratio of the total value to the total number of acres. Next, we estimated cross-sectional regressions for both 1850 and 1860 of the log of the per acre value on the percentage of improved acres and state dummy variables. The residuals from this regression from this regression are what we mean by the "adjusted" per acre value of land. If the residual is positive, land sold for a higher price than what would be predicted based on the proportion of improved acreage and the state in which the county was located.

We find small positive effects of rail access on the farm wage and the adjusted per acre land price but these approach statistical significance only when lagged values of the dependent variables are included. With the lagged value included, the impact of rail access on land price is on the order of 6 percent which, while positive, is less than the

16

impact estimated by Craig, Weiss, and Palmquist (1998). The treatment effect on wages was smaller than on land price, which is consistent with a decline in the wage-rental ratio, as required by the Hekscher-Ohlin model, but the effect is small and not reliably estimated.

In Table 4 we report results of selected regressions for Northern and Southern states separately. The logic of estimating the regression separately by region is that the crop mixes differed significantly, and because the South made use of slave labor. In particular, the use of slave labor in the South allows us to explore another proxy for the share of labor in agriculture, the share of slaves in the total population (ages 10 and over). The overwhelming majority of slaves lived in rural areas and the overwhelming majority of rural slaves were engaged in agriculture. We do find that rail access increased the proportion slave and also appears to have had a larger impact on land values in the South, consistent with Fogel and Engerman's (1980) contention that location rents were larger in the South than in the North. We also find a positive effect of rail access in the North on the share of improved acreage in total land area, but the effect is not statistically significant. The impact on urbanization is also larger in the North (as would be expected from the finding on the slave population share) but in both regions we find no evidence of a positive effect on agricultural yields.

4.0. IPUMS Analysis: Illinois and Indiana

In the previous section we have examined the impact of gaining rail access using county-level data from the published census volumes. In this section, we explore the

17

impact for two states, Illinois and Indiana, using individual-level data from the 1850 and 1860 IPUMS (integrated public use micro-data samples). Indiana is included in our previous analysis but not Illinois because we lack wage data from Illinois in the 1850. Both states are analyzed separately.

There are two advantages to using the IPUMS to study the treatment effects of gaining rail access. First, the IPUMS includes additional information on occupation and sector of employment that is not available (at the county level) in the published volumes. Second, because the IPUMS data are individual-level, it is possible to control for characteristics (such as age) which may influence outcome variables but which, again, cannot be controlled for using published data.

The data consist of repeated cross-sections (1850 and 1860) of males, ages 20-64, who did not attend school in the previous year. Although we have information on labor force participation for individuals ages 15 and over, the census did not record literacy status until age 20, and this is one of the covariates used in the regression (see below and Tables 5 and 6). The results for Indiana are displayed in Table 5 and those for Illinois in Table 6.

Panel A shows rail and water access statistics by states. As noted previously in our discussion of Table 1, Indiana experienced a substantial rise in rail access during the 1850s. The same was true of Illinois: in 1850 only 11 of the 77 Illinois counties included in the analysis had rail access, but by 1860, 54 did.

We explore the impact of rail using four outcome variables, three of which pertain to agriculture, and one to services. We use information on the type of residence (farm = 1), occupation and industry to indicate agricultural participation, and information on

18

industry to indicate participation in the service sector. Panel B shows the sample statistics by state; in both cases, the proportion engaged in agriculture declined while participation in the service sector increased over the 1850s.

Panel C shows the treatment effects of gaining rail access by state when we exclude counties gaining water access. The "baseline" estimates refer to a specification in which there are no covariates; the "with covariate" estimates include a fourth-order polynomial in age, and dummy variables for literacy status, foreign birth, and urban location. Standard errors are clustered by county. Most of the coefficients are not statistically significant at conventional levels but the general patterns are consistent across states: gaining rail access reduced the likelihood of agricultural participation and increased the likelihood of working in the service sector. Including counties that gained water access (Panel D) does not change the substantive findings.

In sum, we find no evidence that individuals in counties that gained rail access were more likely to be engaged in agriculture. Instead, our results suggest that gaining rail access raised the odds the individuals would be engaged in the service sector. To the extent that service sector activities were largely concentrated in urban locations during the antebellum period, the analysis of the IPUMS data is consistent with our finding, using the published census, that railroads promoted urbanization.

4.0 Discussion and Concluding Remarks

We have used county and individual level data to examine the economic impact of gaining rail access in the 1850s. Previous studies have found that rail access raised

land values, and have interpreted this correlation as evidence that rail access chiefly benefited agricultural land owners in the manner predicted by the Hekscher-Ohlin or Von Theunen frameworks elaborated earlier in the paper. In contrast to this previous work, we have used a difference-in-difference strategy, comparing outcomes in a "treated" group of counties – those that gained rail access in the 1850s – to a control group – those that either gained access earlier or did not have access before the Civil War.

Most of the estimated treatment effects are small and the signs are not consistent with either framework. For example, we find that rail access appears to have increased the percent urban and the likelihood of participation in the service sector, decreased agricultural yields, and reduced the share of improved acreage in total land area, opposite to the patterns we would expect if the impact on agriculture were as either framework predicts. Some of our results, of course, are consistent with the frameworks – we find, in particular, positive effects on the price of land, and negative effects on the ratio of farm wages to land price. But the effects that are consistent are generally quite small and frequently statistically insignificant.

If, as appears to be the case, the positive cross-sectional correlation of rail access on land prices is robust, why are our difference-in-difference findings so mixed? One possibility is that the cross-sectional relationship is a consequence of omitted variable bias. That this possibility should be taken seriously is suggested by the fact that, while the sign (positive) of the correlation between rail access and land prices is robust, the size of the correlation is readily influenced by the inclusion or exclusion of covariates (see Craig, Weiss and Palmquist 1998).

A second reason may be that the treatment in question – rail access in 1850s – was not very effective. As we show in Table 1, a very large fraction of the population lived in counties in 1850 that were already served by rail or water. This fraction was actually higher in the North than in the South, which is consistent with our finding that rail access appears to have had a bigger effect on the price of land in the South than in the North. Alternatively, it may be that both rail access and our outcome variables are measured with considerable error, in which case use of DID methods would exacerbate any bias due to measurement error. A related point is that we are unable to control for pre-existing trends. For example, if the percent urban were increasing in the 1840s in counties that acquired rail access in the 1850s, this could account for the positive treatment effect on urbanization that we observe.[6]

Another possibility is that controlling for observable factors that influenced the likelihood of gaining rail access does not adequately purge any correlation between the treatment and the error term – that is, there is selection bias in treatment due to unobservable factors. The solution to such bias would be to use one or more instrumental variables to predict rail access.

Topographical variables are one possible candidate instrument.[7] The idea is that local topography influenced the costs of building a railroad – for example, the presence

---

[6] Another version of this argument is that the treated counties anticipated treatment and thus the changes we would like to observe already occurred in the 1840s. Thus in the 1850s, the control group of counties anticipated that the railroad would eventually be extended to (some) of them, and these counties experienced changes (expanding agricultural yields, or investment in clearing land) prior to treatment.

[7] Two other possible instruments might be the existence of a federal land survey conducted between 1824 and 1838, and the "straight line" method of Banerjee, Duflo, and Qian. (2006). As discussed by Taylor (1951) the federal government conducted numerous land surveys under a law passed in 1824 that was eventually repealed in 1838. The existence of such a survey would lower the costs of building a railroad, and thus might prove to be a useful instrument. The use of pre-existing plans as an instrument has been applied in the case of the federal highway system; see Michaels (2005). The "straight line" method posits that, during the initial phase of rail development, rail are most often built on straight lines between existing urban centers; the instrument in question is a dummy for whether the straight line passes through the county.

of mountainous terrain may have raised the costs, making it less likely that rail access would be provided. At the presence we have only one topographical variable, the presence of water transportation. To explore the potential of this approach, we re-estimated certain regressions treating water access as an instrument (rather than a covariate) for rail access. As Panel C of Table 1 shows, the presence of water access does have a statistically significant, negative effect on the likelihood of gaining rail access in the 1850s – that is, the "first stage" regression appears to be reasonable. When we use water access as an instrument some results appear to be more favorable to our two frameworks – for example, we find a negative treatment effect on percent urban. But none of the more favorable results are statistically significant and, in some cases, the signs of the IV coefficients are still opposite to what the frameworks would predict and are larger in magnitude. We conclude that, while the IV approach has some promise, it will be necessary to construct better instruments than the presence of water transportation.

This paper has a used a pooled sample of counties from 1850 and 1860 to investigate the economic impact of gaining access to a railroad. Previous work has shown that, at a point in time, the presence of railroad increased the value of agricultural land. Our paper takes this correlation as a point of departure and investigates certain ancillary predictions of economic frameworks that predict the rise in land values. We find, surprisingly, that these ancillary predictions are not confirmed – for example, we find no evidence that agricultural yields increase when rail access occurs. Further research is necessary to determine if these findings are robust or whether they are a consequence of various problems and biases in the data.

References

Atack, Jeremy, and Peter Passell. 1994. *A New Economic View of American
    History*. New York: W.W. Norton.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian. 2006, in progress. "The
    Railroad to Success: The Effect of Infrastructure on Economic
    Growth," Department of Economics, Brown University (Power Point
    Presentation).

Craig, Lee and Weiss, Thomas. 1996. "The Nineteenth Century Farm Labor Force and
    Rural Population: County-Level Estimates and Implications," unpublished
    paper, Department of Economics, North Carolina State University.

Craig, Lee, Raymond B. Palmquist, and Thomas Weiss. 1998. "Transportation
    Improvements and Land Values in the Antebellum United States: A
    Hedonic Approach," *The Journal of Real Estate Finance and Economics*
    16: 173-190.

Coffman, Chad and Mary Eschelbach Gregson. 1998. "Railroad Development
    and Land Values," *The Journal of Real Estate Finance and Economics*
    16: 191-204.

Fishlow, Albert. 1965. *American Railroads and the Transformation of the Ante-Bellum
    Economy*. Cambridge, MA: Harvard University Press.

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

Fogel, Robert and Stanley Engerman. 1980. "Explaining the Relative Efficiency of

Slave Agriculture: A Reply," *American Economic Review* 70: 672-690.

Goodrich, Carter (ed). 1961. *Canals and American Economic Development*. New York:

Columbia University Press.

Haines, Michael R. and ICPSR. 2006. Historical, Demographic, Economic, and Social

Data: The United States, 1790-2000. Ann Arbor, MI: Inter-University

Consortium for Political and Social Research.

Haites, Erik; Mak, James; and Gary Walton. *Western River Transportation: The Era*

of Internal Development, 1810-1860. Baltimore, MD: Johns Hopkins

University Press.

Margo, Robert A. 2000. *Wages and Labor Markets in the United States, 1820-1860*.

Chicago: University of Chicago Press.

Michaels, Guy. 2005. "The Effect of Trade on the Demand for Skill: Evidence for the

Interstate Highway System," mimeo, Department of Economics, Massachusetts

Institute of Technology.

North, Douglass C. 1961. *The Economic Growth of the United States, 1790-1860*.

Englewood Cliffs, NJ: Prentice-Hall.

Ransom, Roger. 1967. "Interregional Canals and Economic Specialization in the

Antebellum United States," *Explorations in Entrepreneurial History* 5: 12-35.

Ransom, Roger. 1970. "Social Returns from Public Transport Investment: A Case

Study of the Ohio Canal," *Journal of Political Economy* 78: 1041-1064.

Stover, John F. 1978. Iron Roads to the West: American Railroads in the 1850s.

New York: Columbia University Press.

Taylor, George Rogers. 1951. *The Transportation Revolution, 1815-1860*. New York:

Holt, Rhinehart, and Winston.

Table 1: Railroad Access in 1850 and 1860

Panel A: Full Sample of Counties (Constant Land Area)

	Number of Counties in Sample	Percent with Rail Access (Counties Equally Weighted)	Percent with Rail Access (Counties Population Weighted)
Access in 1850	672	22.4%	46.9% [70.8%]
Access in 1860	672	45.8	66.3 [82.3%]
Access in 1860, No Rail Access in 1850	531	30.1	38.8
1860, No Rail or Water Access in 1850	300	33.7	42.9
1860, Water Access in 1850 but No Rail Access	221	25.2	33.6

Sample consists of counties in merged ICPSR(Haines)-Margo data set with constant land area (square miles) in 1850 and 1860. Equally weighted: unit of observation is the county. Population weighted: counties are weighted by total population in the given year. Sample in row "Gained Access in 1850s" consists of counties with no rail access in 1850. Rail Access = 1 if county had a rail line within its boundaries; water access = 1 if county had a navigable waterway within (or on) its boundaries. [ ]: proportion of population living in counties with rail or water access.

Panel B: Rail Access by State: Population Weighted

State	Number of Counties	Rail Access in 1850	Rail Access in 1860	Percent Change, Rail Access in 1860, No Access in 1850
Massachusetts	11	98.3%	99.4	64.8%
Pennsylvania	58	75.0	89.6	58.4
Indiana	80	34.1	70.8	55.7
Michigan	32	74.1	78.6	17.4
Iowa	46	0	33.0	33.0
Virginia	106	28.5	49.9	29.9
Georgia	22	76.5	84.9	35.7
Louisiana	25	44.3	53.2	16.0
Mississippi	51	5.0	48.4	45.7
North Carolina	63	16.6	40.0	28.1
South Carolina	28	54.6	72.0	38.8
Texas	33	0	15.8	15.8

Kentucky	68	16.7	36.0	23.2
Tennessee	49	23.1	59.5	47.3

See Panel A for sample definition. Column 4: [Column 3 – Column 2]/[100 – Column 2] x 100%

Panel C: Linear Probability Coefficients: Rail Access in 1860, No Access in 1850

	Equally Weighted	1850 Population Weighted
Water Access in 1850 = 1	-0.105	-0.082
	(0.036)	(0.038)
Percent Urban in 1850	0.424	0.273
	(0.148)	(0.124)
Log (Population/Square Miles) in 1850	0.122	0.049
	(0.021)	(0.027)
State Dummies Included?	Yes	Yes
Adjusted R-square	0.229	0.242

Table 2: Treatment Effects of Gaining Rail Access: Percent Urban, Output per Acre, and Improved Acres/Square Mile

Panel A: Urban Population/Total Population

Sample	Number	State Dummies?	1850 Covariates?	Lagged Dependent Variable?	Coefficient
1850 Rail Access = 0	506	No	No	No	0.018 (0.005)
1850 Rail Access = 0	506	Yes	No	No	0.020 (0.005)
1850 Rail Access = 0	506	No	Yes	No	0.018 (0.005)
1850 Rail Access = 0	506	Yes	Yes	No	0.019 (0.006)
1850 Rail Access = 0	506	Yes	Yes	Yes	Na
All	655	No	No	No	0.009 (0.005)
All	655	Yes	No	No	0.011 (0.006)
All	655	No	Yes	No	0.010 (0.005)
All	655	Yes	Yes	No	0.011 (0.005)
All	655	Yes	Yes	Yes	Na

Notes: All: all counties. Percent Urban = population in towns, villages, etc. of 2,500 or more/total population. Counties are equally weighted in estimation. 1850 Covariates: From Table 1, Panel C, Percent urban, log (Population/Square Miles), presence of a navigable waterway. Na: not applicable.

Panel B: Log (Value of Agricultural Output, Definition #1/Improved Acres)

Sample	Number	State Dummies?	1850 Covariates?	Lagged Dependent Variable?	Coefficient
1850 Rail Access = 0	506	No	No	No	-0.036 (0.040)
1850 Rail Access = 0	506	Yes	No	No	-0.024 (0.040)
1850 Rail Access = 0	506	No	Yes	No	-0.027 (0.040)
1850 Rail Access = 0	506	Yes	Yes	No	-0.020 (0.041)
1850 Rail Access = 0	506	Yes	Yes	Yes	-0.019 (0.040)

Sample	Number	State Dummies?	1850 Covariates?	Lagged Dependent Variable?	Coefficient
All	655	No	No	No	-0.033 (0.035)
All	655	Yes	No	No	-0.035 (0.035)
All	655	No	Yes	No	-0.025 (0.035)
All	655	Yes	Yes	No	-0.031 (0.034)
All	655	Yes	Yes	Yes	-0.033 (0.033)

Notes: see Panel A. Definition #1: Value of Crops, orchards, and marketed garden produce.

Panel C: Log (Value of Agricultural Output, Definition #2/Improved Acres)

Sample	Number	State Dummies?	1850 Covariates?	Lagged Dependent Variable?	Coefficient
1850 Rail Access = 0	506	No	No	No	-0.034 (0.032)
1850 Rail Access = 0	506	Yes	No	No	-0.019 (0.032)
1850 Rail Access = 0	506	No	Yes	No	-0.026 (0.033)
1850 Rail Access = 0	506	Yes	Yes	No	-0.018 (0.032)
1850 Rail Access = 0	506	Yes	Yes	Yes	-0.020 (0.030)
All	655	No	No	No	-0.033 (0.029)
All	655	Yes	No	No	-0.032 (0.029)
All	655	No	Yes	No	-0.027 (0.029)
All	655	Yes	Yes	No	-0.030 (0.029)
All	655	Yes	Yes	Yes	-0.036 (0.026)

Notes: see Panel A. Definition #2: definition #1 + value of slaughtered livestock + value of household manufactures.

Panel D: Log (Improved Acres/Square Mile)

Sample	Number	State Dummies?	1850 Covariates?	Lagged Dependent Variable?	Coefficient
1850 Rail Access = 0	506	No	No	No	-0.212* (0.069)

1850 Rail Access = 0	506	Yes	No	No	-0.149* (0.045)
1850 Rail Access = 0	506	No	Yes	No	-0.042 (0.049)
1850 Rail Access = 0	506	Yes	Yes	No	-0.075* (0.039)
1850 Rail Access = 0	506	Yes	Yes	Yes	-0.058 (0.035)
All	655	No	No	No	-0.097 (0.060)
All	655	Yes	No	No	-0.099* (0.039)
All	655	No	Yes	No	-0.057 (0.043)
All	655	Yes	Yes	No	-0.084* (0.033)
All	655	Yes	Yes	Yes	-0.072* (0.030)

Table 3: Factor Price Regressions: Counties with No Rail Access in 1850: Difference-in-Difference Estimates

Dependent Variable	1850 Covariates?	Lagged Dependent Variable?	Coefficient
Log (Farm Wage with value of board imputed)	Yes	No	0.023 (0.022)
Log (Farm Wage with value of board imputed)	Yes	Yes	0.035 (0.019)
Log (Adjusted Per Acre Land Price)	Yes	No	0.036 (0.047)
Log (Adjusted Per Acre Land Price)	Yes	Yes	0.061 (0.036)
Log (Farm Wage with value of board imputed /Adjusted Per Acre Land Price)	Yes	No	-0.013 (0.050)
Log (Farm Wage with value of board imputed/Adjusted Per Acre Land Price	Yes	Yes	-0.027 (0.038)

Adjusted Per Acre Land Price: residual from regression of log(value of farm/total acres) on (Improved Acres/Total Acres) plus state dummies.
Farm Wage with value of board imputed = Log (monthly money wage + 4.3*weekly value of board)

Table 4: North-South Differences: Treatment Effects of Rail Access in 1860, No Access in 1850: Difference in Difference Estimates

Dependent Variable	Sample Size, North	Coefficient, North	Sample Size, South	Coefficient, South
Percent Urban	150	0.049* (0.011)	371	0.005 (0.006)
Slave Population, ages 10+/Total Population, ages 10+	na	Na	371	0.008* (0.004)
Log (Improved Acres/Square Mile)	139	0.046 (0.062)	368	-0.077* (0.034)
Log (Value of Agricultural Output #1/Improved Acres)	140	-0.023 (0.038)	368	-0.008 (0.029)
Log (Value of Agricultural Output #3/Improved Acres)	140	-0.033 (0.038)	368	-0.007 (0.052)
Log (Farm Wage w/value of board imputed/Adjusted Per Acre Land Price)	125	0.006 (0.052)	286	-0.044 (0.050)
Log (Farm Wage w/value of board imputed)	286	-0.013 (0.020)	286	0.061* (0.026)
Log (Adjusted Per Acre Land Price)	140	-0.008 (0.052)	368	0.109* (0.045)

Table 5: Preliminary IPUMS Results for Indiana

Panel A: Sample Statistics: Rail and Water Access in the 1850s, Counties

	Rail = 1	Water = 1
No Access in 1850 or 1860	27	33
No Access in 1850, Access in 1860	28	5
Access in 1850 and 1860	25	42
Total	80	80

Notes: Counties with boundary changes (according to ICPSR coding) or fewer than 10 per year observations are excluded.
Source: 1850, 1860 IPUMS matched to Craig, Weiss, Palmquist (1998). Sample consists of adult males, ages 15-64, not attending school during census year.

Panel B: Sample Statistics, Outcome Variables: Regression Sample (Ages 20-64)

	# Obs	Farm Residence	Agriculture, Occupation	Agriculture, Industry	Services, Industry	Rail = 1	Water = 1
1850	1,983	66.7%	61.8%	61.7%	13.5%	38.2%	44.7%
1860	2,808	61.4	59.2	58.7	15.3	74.6	49.6%

Notes: Farm Residence = 1 if IPUMS; Agricultural occupation = 1 if 3-digit occupation code = 100, 123, 810, 820, or 970 (laborers, n.e.c.) and farm residence = 1; agricultural industry = 1 if 3-digit industry code = 105; Services = 1 if industry code = 500, …, 946.

Panel C: Treatment Effects of Rail: Indiana in the 1850s: Excludes Counties Gaining Water Access

	# Obs	Farm = 1	Agricultural Occupation = 1	Agriculture (Industry) = 1	Services (Industry) = 1
Baseline	4,550	-0.052 (0.042)	-0.053 (0.034)	-0.061 (0.032)	0.040 (0.026)
With Covariates	5,141	-0.050 (0.040)	-0.050 (0.032)	-0.059 (0.033)	0.038 (0.025)

Sample consists of adult males, ages 20-64, not attending school. Five counties gaining water access during the 1850s are excluded. Baseline: treatment plus year (= 1860) dummy. Covariates: baseline plus fourth-order polynomial in age, dummies for literacy, foreign birth, metro area status. Standard errors (in parentheses) are clustered by county.

Table 5: Panel D: Treatment Effects of Rail and Water: Indiana in the 1850s: All Counties in Sample

	# Obs	Farm = 1	Agricultural Occupation = 1	Agriculture (Industry) = 1	Services (Industry) = 1
Baseline, Rail = 1	4,791	-0.050 (0.040)	-0.060 (0.033)	-0.068 (0.034)	0.042 (0.025)
With Covariates, Rail = 1	4,791	-0.049 (0.038)	-0.058 (0.033)	-0.067 (0.033)	0.041 (0.024)

See Panel A; includes five counties gaining water access.

Table 6: Preliminary IPUMS Results for Illinois

Panel A: Sample Statistics: Rail and Water Access in the 1850s, Counties

	Rail = 1	Water = 1
No Access in 1850 or 1860	24	39
No Access in 1850, Access in 1860	42	4
Access in 1850 and 1860	11	34
Total	77	77

Notes: Counties with boundary changes (according to ICPSR coding) or fewer than 10 per year observations are excluded. Source: 1850, 1860 IPUMS matched to Craig, Weiss, Palmquist (1998). Sample consists of adult males, ages 15-64, not attending school during census year.

Panel B: Sample Statistics, Outcome Variables: Regression Sample (Ages 20-64)

	# Obs	Farm Residence	Agriculture, Occupation	Agriculture, Industry	Services, Industry	Rail = 1	Water = 1
1850	1,817	67.1%	61.0%	61.2%	13.4%	23.8%	52.8%
1860	3,616	58.5	55.0	55.3	18.5	78.4	58.0%

Notes: Farm Residence = 1 if IPUMS; Agricultural occupation = 1 if 3-digit occupation code = 100, 123, 810, 820, or 970 (laborers, n.e.c.) and farm residence = 1; agricultural industry = 1 if 3-digit industry code = 105; Services = 1 if industry code = 500, …, 946.

Panel C: Treatment Effects of Rail: Illinois in the 1850s: Excludes Counties Gaining Water Access

	# Obs	Farm = 1	Agricultural Occupation = 1	Agriculture (Industry) = 1	Services (Industry) = 1
Baseline	5,141	-0.033 (0.034)	-0.024 (0.038)	-0.025 (0.037)	0.008 (0.025)
With Covariates	5,141	-0.016 (0.037)	-0.004 (0.038)	-0.005 (0.037)	0.017 (0.026)

Sample consists of adult males, ages 20-64, not attending school. Four counties gaining water access during the 1850s are excluded. Baseline: treatment plus year (= 1860) dummy. Covariates: baseline plus fourth-order polynomial in age, dummies for literacy, foreign birth, metro area status. Standard errors (in parentheses) are clustered by county.

Table 6, Panel D: Treatment Effects of Rail and Water: Illinois in the 1850s: All Counties in Sample

	# Obs	Farm = 1	Agricultural Occupation = 1	Agriculture (Industry) = 1	Services (Industry) = 1
Baseline, Rail = 1	5,433	-0.042 (0.033)	-0.031 (0.035)	-0.031 (0.055)	0.013 (0.056)
With Covariates, Rail = 1	5,433	-0.025 (0.035)	-0.012 (0.036)	-0.012 (0.035)	0.020 (0.025)

See Panel A; includes four counties gaining water access.