Dense Enough To Be Brilliant: Patents, Urbanization, and Transportation in Nineteenth Century America*

Market Access, Information Flows

Elisabeth Ruth Perlman, elisabeth.perlman@alumni.carleton.edu
Click Here For the Current Version.†

October 23, 2017

Abstract

This paper examines the geographic distribution of patenting in the nineteenth century United States as it evolves in response to transportation improvements. I find a robust, statistically significant, and positive effect of increases in local transportation access on patenting. Over the twenty years following the arrival of the railroad in a county, the number of patents per capita doubles. I explore two possible mechanisms behind this increase: a) inventors responded to larger markets afforded by transportation improvements; or b) transportation improved information flows making inventors more productive. I find that patenting weakly responded positively to increased market access, but that local access still matters much more. Using digitized texts of patents, I measure whether any given patent mentions a previous, novel technology within a particular time frame. I find little evidence that the speed of arrival of these new ideas is related to transportation improvements. These results suggest that access to local transportation lowers the effective cost of patenting by forming a nexus around which local agglomerations occur.

*Acknowledgments: I am extremely grateful to Robert Margo, Tom Nicholas, Jeremy Atack, Carola Frydman, Daniele Paserman, Michael Meurer, Michael Haines, Claudia Olivetti, Suresh Naidu, Dan Bogart, Thomas M. Darby, Peter Sims, Matthew C. Berntsen, Henry Gross, Ross Corliss, and the volunteers at the Directory of American Tool and Machinery Patents and to numerous seminar participants for their helpful advice, data, and comments. Finally, to Chloe Hite for excellent research assistance. All errors are my own.

†Address: *http://elisabethperlman.net/papers/Perlman_19thCPatent.html*

1 Introduction

Ever since Marshall (1890), economists have believed that location matters for innovative activity, as co-location facilitates the transfer of innovative ideas (Jaffe et al., 1993). The spread of communication and transportation has a democratizing effect, encouraging people who would not have previously participated in innovative activity to do so (Friedman, 2006). At the beginning of the nineteenth century the world was very poorly connected, but by the end of that century the movement of people, goods, and information among places had increased dramatically.

Before railroads, waterways were by far the most efficient way to transport goods. Moving goods over land, even on the best roads, was extremely costly.[1] By the 1860s, the telegraph had become the fastest way of sending a message, and physical packages moved overland by train or over water via steamship. These changes made transportation faster, cheaper and safer, effectively reducing the distance between locations. This reduction in distance was even more dramatic on the periphery of the transportation and communication network. By the end of the nineteenth century railroad tracks densely intersected much of rural America. This promoted economic growth by linking far-flung factor and product markets, encouraging the exploitation of regional comparative advantage (Fogel, 1964; Atack et al., 2011; Donaldson and Hornbeck, 2013). Railroads also changed the character of the areas around them. They provided loci for new towns—increasing urbanization (Atack et al., 2010), attracting banks (Atack et al., 2014), and encouraging speculators to plat towns (Hudson, 1985).

I investigate how the nineteenth century "transportation revolution," in which the most dramatic change came from the railroad, changed the location of innovative activity in

[1] It has become a commonplace to note that it cost about the same amount to ship goods between London and Boston as to travel 30 miles over land in the United States, or about the distance from downtown Boston to Concord, MA (Howe, 2007).

the United States. I first document that the expansion of improved transportation in a county led to an increase in innovation, as measured by patenting activity. I then test the hypothesis, formulated by Sokoloff (1988), that improving transportation networks created incentives for innovation by facilitating access to larger markets. Following this, I examine the hypothesis that increasing transportation in an area increases information flows, giving potential inventors greater access to new ideas. I use textual analysis on the actual content of about 700,000 patents to construct a new database that allows me to observe the use of phrases over time. I use these data to compute the speed at which novel technology arrives in an area, and I investigate how the speed of arrival relates to the expansion of the transportation network.

I measure innovative activity by the number of patents per capita at the county level from 1790 to 1900. I obtain information on patents from Tom Nicholas (Akcigit et al., 2013) for the 1836-1900 period, and I construct a similar measure for 1790-1836 by collecting and geocodding data from a list of known patents from this period. I link this with new data on the spread of transportation networks from Atack (2013) to construct a measure of local transportation access, the proportion of a county's land area that is within a specified distance of improved transportation.

I find a robust, statistically significant and positive effect of local access on patenting, that suggests that 8% of the increase in patenting over the nineteenth century was due to the spread of transportation. Because the massive increase in patenting that happened mid-century was led by the most developed places, that already had significant transportation access, increased transportation access cannot be the sole explanation. However, transportation had a significant effect for those counties that were not previously well connected; when I restrict my attention to only those counties that were not saturated by transportation in 1850, this estimate doubles. In addition to a positive main effect, increased transportation is associated with decreased concentration of patenting, particularity for more rural places. To address concerns that the documented relationship between transportation access and patenting may be driven by endogeneity in the construction of new transportation, I use straight lines drawn between the prosperous places in 1830 as an instrument for transportation access. This IV specification gives much larger point estimates (they imply that more than 20% of the increase in patenting was due to transportation), but the standard error is such that I cannot reject the hypothesis that the IV estimates and OLS estimates are the same.

Increased market access may lead to patentable innovation as the expected return to research and development is positively related to the size of the relevant market. Numerous studies using modern plant- or firm-level data have noted that exporting firms become more productive after trade liberalization.[2] Extensions to the the influential paper by Melitz (2003) by Bustos (2011) and Lileeva and Trefler (2010), give a clear motivation as to why larger markets might encourage a firm to invest in innovation; larger markets allow more units to be sold, thus providing for greater returns as firms reduce their costs.

Economic historians have also argued that increased market access contributed to the increase in patenting activity in the early nineteenth century. Sokoloff (1988) finds that counties in New York and Pennsylvania along the newly-built canals (particularly the Erie Canal) saw a sharp increase in patenting activity between 1790 and 1846, and attributed this change to the increased market access in these areas.

In order to more directly test the effect of increases in the ability of an area to access larger markets, I calculate a measure of market access inspired by the methodology developed in Donaldson and Hornbeck (2013), which itself builds on earlier studies on this topic (Harris, 1954; **?**). This estimate is the sum of the population of all counties in the United States;

[2] See, among others, Pavcnik (2002); Amiti and Konings (2004); van Biesebroeck (2005); Becker and Egger (2013); Deloecker (2007); Fernandes (2007); Foster et al. (2008); Topalova and Khandelwal (2011).

each counties' contribution to this sum is weighted by the cost of moving a ton of goods from the observation county to it.[3] In Appendix **??** I investigate the direct impact of population density. The correlation of my estimate of market access with patenting per capita is not robust to the inclusion of controls. In particular, the inclusion of the lagged percent of county's farmland that is improved absorbs the relevant variation. Also, when both market access and local transportation access are included in the specification, local transportation access retains a positive, precisely estimated coefficient, but market access does not. It is also notable that when the percent of a county's area that within some distance from transportation is used as a local access measure, the measures that are calculated using shorter distances are more closely related to increases in patenting. This suggests that the impact of access to local transportation comes through localized changes within a county.

In addition to moving goods, transportation facilitates greater communication between areas. This greater exchange of ideas may change where innovation occurs by helping more areas learn about new technologies.[4] Several studies have examined access to a different communication medium, the internet, on the location of innovation, finding that "diffusion of the internet worked against the trend toward increasing geographic concentration of inventive activity" (Forman et al., 2014), and that an increase in communication seemed to allow for greater task specialization (Agrawal and Goldfarb, 2008). Investigating transportation technologies with the movement of people in mind, Agrawal et al. (2014) looks at the effect of highways on patenting. Agrawal et al. (2014) find that not only does increased highway access increase patenting in an area, but it also increases those patents' propensity to cite patents whose inventors are located further away in the same region. These studies suggest

[3] If counties are nodes of a network, and transportation provides connections between them, this measure of market access is a closeness centrality measure (Rochat, 2009).

[4] There is abundant evidence that location matters for innovative activity; see Feldman and Kogler (2010) for a recent survey. One explanation for this importance is that proximity facilitates the exchange of ideas and tacit knowledge (Jaffe et al., 1993; Audretsch and Feldman, 2004). However, there are many other reasons why location might matter for innovation, including credit access, access to skilled labor, or access to knowledge that resists easy diffusion.

that changes in information movement can change the location of innovation.

Finally, in order to investigate the importance of information about new technologies, I use the text of patents as documentation of the movement of idea-use over time and space. I construct a measure of how many new words, from a list of words relating to new technologies, appear in the patent record of any given county. If in a particular county one of these word is observed one year after it is first used in a patent anywhere, then this county is measured as receiving new words at the rate of one new word a year. This measures the use of new technologies in patents originating from a county, telling us how quickly novel technologies become present in inventors' work. In contrast to the strong relationship between local transportation access and patents per capita, increases in local transportation access seem to have no effect on the novelty of innovation (a measure of quality). To understand this pattern in more detail, I examine a sample of patents in newly-connected counties. This examination suggests that newly connected places patented things related to local industry, such as machines to help with the production of cotton or grain. They then slowly started patenting technologies related to the railroad itself—couplers and, in the north, ways to clear snow from the tracks. Following urbanization, they began patenting middle class consumption goods, such as medicine and furniture. This suggest that patenting is driven by concerns that are locally salient.

My results suggest that local transportation access is related to increases in patenting primarily because transportation forms a nexus around which local agglomerations occur. That patent quality does not increase when local transportation access increases but quantity does is a result consistent with these agglomerations facilitating patenting by reducing the effective cost of participating in the formal intellectual property system.

2 Background

2.1 The Patent System

The present system of state-created and -enforced monopolies in intellectual property, patents, developed from of an older tradition of state monopoly grants (Bracha, 2004). In late eighteenth century Britain and its colonies, the process for requesting grants promoting the development of new industries and technologies in a location had been routinized. Still, the 1790 United States Patent Act,[5] which outlined uniform standards for what was patentable and a uniform process for obtaining a patent at a low fee (about $5), was the first of its kind (Khan, 2005). The United State's system became a model for other countries as they introduced or reformed their patent systems during the nineteenth century.

It is striking that, as a new country on the periphery, the 1790 Act specified that a patentee must be "the first and true inventor" anywhere in the world. Most nineteenth century patent systems allowed grants to go to those who were the first to introduce the technology into the country, as colony and then state patents had done (Hrdy, 2013). The United States system fluctuated in its enforcement of this mandate, most notably dropping any attempt to examine patents for novelty in 1793 (but increasing the fee to $30) and then reinstating examinations in 1836 (see appendix A.2 for more details on changes in patent law). Patents can only be issued to individuals (not to corporations), but these intellectual properties can be, and are often, sold. If the rights a patent confers have been sold prior to its being granted, it is "assigned at issue;" this assignment is recorded on the patent specification in the nineteenth century.

The creation of a monopoly over the patented invention is generally justified in two ways:

[5] The United States Constitution gives Congress the power "[t]o promote the progress of science and useful arts, by securing for limited times to authors and inventors the exclusive rights to their writings and discoveries." The 1790 Patent Act was following this mandate.

by the increased incentive to engage in innovation that the monopoly creates and by the value of the information that inventors are forced to disclose as part of the application process. This public disclosure sets patenting apart from other methods that inventors might use to capture gains from innovation, most notably trade secrets (Moser, 2004). The effectiveness of the disclosure requirement varied, though inventors were required to describe their invention so that "a person having ordinary skill in the art" would understand. Until 1871,[6] anyone looking to copy the information contained in the patent would have needed to travel to Washington, DC or pay a substantial reproduction fee in order to read a patent specification. The patent office did not publish summaries of issued patents until 1872. Therefore, investors interested in newly issued patents relied on the summaries provided by private periodicals; the "Journal of the Franklin Institute" published its first issue in 1826, and the magazine "Scientific American" started publication in 1845. Both publications devoted substantial space to new inventions of British origin. The creation of a new, patentable innovation requires learning where the technological frontier is: what problems are interesting, what the existing best solutions are, and what lines of research have been or are being explored. Thus, the role of institutions that disseminate this knowledge is potentially important.

As was the general character of firms in the early nineteenth century, invention was primarily done by individuals. As the century progressed, large firms began to invest in research and development (R&D). It was not until the 1910s, outside of the period of this study, that in-house R&D became the dominant mode of financing innovation (Lamoreaux and Sokoloff, 2005).

[6] After which 22 depository libraries were created around the county, and the patent office distributed copies of issued patents for a reasonable fee.

2.2 Transportation Improvements

Before the twentieth century, transport costs were of utmost importance. Moving goods over land without mechanical power, even on the best roads was extremely costly. In the eighteenth century United States, coastal cities were more closely integrated with Europe than with their hinterlands, not only due to colonial links, but also because of the large cost differentials between sea and overland transport. The placement of cities along the St. Lawrence route to the Atlantic in Eastern Canada is a reflection of the importance of water transportation when they were first settled.

The lack of access to the interior of the country drove investments in transportation infrastructure. In the era before the steam engine, this meant the building of postal roads and efforts to increase ease of travel on rivers.[7] The early part of the nineteenth century saw a large investment in canals. The most notable of these was the Erie Canal in upstate New York, which opened in 1825, but there was also significant construction in both Pennsylvania and Ohio.

Railroad construction in the United States began in 1820. Initial lines were short, linking nearby settlement to mines or waterways. During the 1850s the United States experienced its first great wave of rail expansion when approximately 22,000 miles of track were laid. By 1860, in addition to dense coverage in the Northeast, the railroad network had reached Illinois, Indiana, and Ohio, with significant penetration into Wisconsin and Iowa. The South saw less construction, but it too experienced substantial growth in rail access in the 1850s (Atack et al., 2010). By 1870, the transcontinental railroad had been completed, though the western market it served was mostly limited to the San Francisco Bay area. By 1890, the areas around Portland, Seattle, San Fransisco, and Los Angeles all were connected to the same national rail network that had covered the country east of the Dakotas.

[7]Concurrently there were large investments in turnpikes in Britain (Bogart, 2005).

3 Data

I use patents as a measure of innovative activity. Thought patents are imperfect indicators of technological improvement (Trajtenberg, 1990; Moser, 2004) they are the most accessible and detailed written records of innovation. As such, economists and economic historians have long studied them to probe the economics of technological development.

In order to to connect patenting to changes in transportation, the location of these patents needs to be recorded and geo-located. This requires ether going through 700,000 original patents, or using the yearly lists published by the patent office to link the recorded location with a modern geocode. I construct a Geographic Information System (GIS) database of patents issued from 1790-1836. In 1836 there was a fire in the patent office that burned all the patents that had been issued to that date. In an attempt to recover from the damage this caused, the patent office put out a call for existent information on patents; in 1874 Congress used the information the patent office had received to compile a list of patents issued from 1790-1836 (a 1847 print volume presumably provided much of this information). This 1874 Congressional list has been updated by volunteers, such as Jim Shaw and the maintainers of the Directory of American Tool and Machinery Patents, who have found patents that the 1874 list did not include. I geo-located the patents by merging the town and county information with a database of historical town names from the AniMap 3.0.2 County Boundary Historical Atlas and the United States Board on Geographic Names (part of the Department of the Interior). The rest of the data on patent location comes from Tom Nicholas' dataset of patents issued from 1836-1900 (Akcigit et al., 2013), which has latitude and longitude coordinates of the listed places on these patents.[8]

These geo-located patents are then merged with the National Historical Geographic In-

[8]Based on the consecutive numbering of patents post-1936, the Akcigit et al. (2013) data cover the vast majority of patents granted between 1836-1900, with limited geo-coding errors.

formation System (NHGIS) shape-files of United States county boundaries. This allows patent counts by county to be created. The text of patents was scraped from Westlaw and Google Patents; these respective organizations generated these text files in an automated way (OCR) from the images of the original typeset documents.[9]

This paper uses both contemporaneous county boundaries and a sample of consistent land area counties, harmonized to 1840 boundaries as suggested in Hornbeck (2010).[10] United States Census data are from Haines (2010). Transportation data are from Atack (2013), which are linked with shape-files of United States county boundaries to explore the spread of railroads and canals.

Table 1 shows the summary statistics for each year.[11] Figure 1 shows the total patents in each year, the US population from the census, and the percent of land area in the US that is within 5 miles of the railroad or some form of water transport. The number of patents per person is not at all smooth over time, and there is a large increase in patenting activity that starts in the 1850s, shortly following the sharp increase in rail access.

Figure 2 shows the spread of patenting across the country. It presents the number of patents issued per ten thousand people in three year bins around the named year[12] as well as the spread of the canal and railroad network by county for four benchmark years. Note the increasing area that is involved in patenting, as well as the increase in patenting per person.

Figure 3 shows how the concentration of patenting and population has changed over time. The Herfindahl index of patent concentration falls substantially over the nineteenth century, reaching a low in 1870. This is in contrast to population which rises steadily

[9] Typeset documents only exist for patents post-1836. Due to the fire of 1836 few of the pre-1836 patents, which are written in long-hand, are existent.

[10] More details on this computation see the boundary shifting files available on my website or upon request.

[11] In this tables and in most of the analysis done in this paper, the number of patents refers to those issued in a three year period: the complete named year and the complete year before and after the named year.

[12] Idem.

after 1840. However, it is not so dissimilar from the concentration of urban population. Breaking the country into regions, as is done in the second panel of Figure 3, shows significant heterogeneity among them. The Northeast levels out after 1830, where as the Midwest does so in 1860, and the South has an upward trend between 1860 and 1880. Thus both the Northeast and Midwest show declines in the concentration of location of patents during their transportation booms.

4 Patents per Capita and Local Transportation Access

Starting the analysis of the relationship between transportation access and patenting, I examine the point estimates on the dummy variables for the number of years since the arrival of the canal or railroad in a county from a regression of:

$$\begin{aligned} PatentMeasure_{it} = \alpha + \beta \mathbf{YearsSinceArrivalDummies_{it}} + \varphi \mathbf{X_{i(t-N)}} \\ + \gamma_i + \delta_t + Region \times \delta_t + \varepsilon \end{aligned} \tag{1}$$

where $PatentMeasure_{it}$ is the measure of patenting at the county level, $\mathbf{YearsSinceArrivalDummies_{it}}$ are dummy variables for the number of year since county i received a railroad, $\mathbf{X_i}$ are county level controls (where $t-N$ denotes the use of all previously observed values),[13] $Region$ denotes the nine census divisions, γ_i are county fixed effects, and δ_t are year fixed effects; standard errors are clustered at the county level.

Figure 4 plots these point estimates. The data includes counties that received a railroad as early as 1830 and counties that had not received by 1900; the later group are given their own

[13] Controls that maybe included (when noted) are the percent of the county that is: employed in manufacturing, urban (2,500+), metropolitan (25,000+), literate, in school, born out of state, and foreign born, as well as the percent of farm land that is improved. Each variable is included at lagged values, and interacted with time dummies. More precisely, a variable that is observed in year t takes on the value zero before year $t+1$, and takes on its value in year t for all years following year $t+1$. This variable is interacted with year dummies. This is a more complete way of controlling for observables since many of these variables are only observed for some census years, and often those years are non-consecutive.

non-time varying dummy, while the former are treated as if they received transportation in 1600 (this means that there time varying dummies are entirely distinct from the ones plotted here). The top graph show Equation 1 estimated with no county level controls, while the bottom shows this equation estimated with the full set of county level controls detailed above. Despite the degree to which many of the included controls are assorted with the arrival of the railroad, the picture looks remarkably similar.[14] Similarly Figure 5 shows these estimates for the number of years since counties in the North East received some form of water transportation. Much the same pattern (and magnitude) is seen in the top panels of Figures 4 and 5. However, while the general upward pattern remains after the controls are added, the standard errors are such that one cannot reject the hypothesis that there was no change in patenting after the arrival of water transportation.

This relationship between transportation and patenting can be compared to a number of other things known to have changed with the introduction of rail transportation. Please see Appendix A.1.

Turning to my main specification, a fixed effect specification with year and county fixed effects and pre-trends in transportation access:

$$\begin{aligned} PatentMeasure_{it} = \alpha + \beta TransportMeasure_{it} + \varphi \mathbf{X_{i(t-1)}} + \gamma_i + \delta_t \\ + Region \times \delta_t + T.M._{\cdot i(t-N)} \times \delta_t + \varepsilon \end{aligned} \tag{2}$$

where $PatentMeasure_{it}$ is the measure of patenting at the county level, $TransportMeasure_{it}$ (abbreviated $T.M._{\cdot it}$) is the specified measure of transportation access, $\mathbf{X_{it}}$ are county level controls,[15] γ_i are county fixed effects, δ_t are year fixed effects, and

[14]Figure 6 shows the same estimations preformed with the number of years since any improved transportation arrived in a county (defined as more than 1% of a counties land area being within 1.5 miles of transportation)

[15]Controls that maybe included (when noted) are the percent of the county that is: employed in manufacturing, urban (2,500+), metropolitan (25,000+), literate, in school, born out of state, and foreign born, as well as the percent of farm land that is improved. Each variable is included at lagged values, and interacted

Region are fixed effects at the nine-region level; $t-N$ denotes the use of all previously observed values.[16]

Local transportation access is measured as the percent of a county's land area that is within a specified distance of some form of improved transportation (railroads, canals, steam navigable rivers, or ports). Table 2 shows estimates using the percent of a county's land area that is within 1.5, 5, or 15 miles of transportation and the number of patents per 10,000 people with the fixed effects and pre-trends of transportation access included as controls. Table 3 adds lagged county-level controls one category at time. The largest impacts on the coefficient on local transportation access come from the inclusion of the percent of a county's population engaged in manufacturing, and from the percent of a county's farm land that is improved. This effect is particularly magnified when these two groups of variables are included, causing the size of the estimate to fall to slightly less than half. When all controls are added the effect of a change in the percent of a county's land area that is within 15 miles of transportation is no longer precisely estimated, but the percent of the county's land area that is within five miles (people living in this band would have been able to make a trip to the mode of transportation and back to their homes in a day) and the percent that is within 1.5 miles remain precisely estimated. These estimates imply that a one standard deviation change in local transportation access in 1860 (about 13% more of the county within 1.5 miles of transportation) is associated with 0.1 more patents per ten thousand people, or 5% of 1860's mean; in 1850 this would imply an increase of about 20% of the mean or 0.1 standard deviation in patenting rates.

with time dummies. More precisely, a variable that is observed in year t takes on the value zero before year $t+1$, and takes on its value in year t for all years following year $t+1$. This variable is interacted with year dummies. This is a more complete way of controlling for observables since many of these variables are only observed for some census years, and often those years are non-consecutive.

[16] A variable that is observed in year t is interacted with year dummies such that several new variables are included. One that takes on the value of zero in all years before year t, and takes on its value in year t and then zero thereafter, another one that takes on the value of zero in all years before year $t+1$, and takes on the value from year t in year $t+1$ and then zero thereafter, etc. for all remaining years to $t+N=1900$.

There are many ways to investigate potential heterogeneous effects of increased transportation access; Table 4 follows the specification that Forman et al. (2014) use to examine the impact of increased internet access on patenting rates. I have adapted this equation for a multi-period model using first differences:

$$\begin{aligned} \Delta PatentMeasure_{it} &= \alpha + \beta PatentMeasure_{i(t-1)} + \varphi \Delta TransportMeasure_{it} \\ &+ \zeta \Delta T.M._{it} \times P.M._{i(t-1)} + \eta \mathbf{X_{i(t-1)}} + \delta_t + Region_i \times \delta_t + T.M._{i(t-N)} \times \delta_t + \varepsilon \end{aligned} \tag{3}$$

where the variables are as defined above in Equation 2, one period represents ten years. The first column of Table 4 shows the relationship between the number of patents per capita last period and the increase between the last and this period. The negative coefficient suggests there is a process by which there is regression to the mean—places with more patenting see patenting grow more slowly. The second column presents the impact of increased transportation on changes in patenting, this coefficient is in line with the ones shown in Table 3. The last two columns add the interaction between lagged levels of patents and increased transportation. The main effect of lagged levels of patenting remains negative, while that of increased transportation remains positive. The interaction between these two is negative, suggesting that transportation helps spread out the location of patenting—places with more patenting that receive increased transportation see less of an increase in their patenting rates than other places. The second panel of Table 4 uses the growth rate of patents per capita instead of first differences, with consistent results.[17]

4.1 Instrumental Variables

Given how much smaller the estimates become with the inclusion of county level controls, there is significant reason to worry about reverse causality. Counties that receive the railroad may be positively selected such that counties that received the railroad earlier will patent

[17]Growth rate is computed as $1 - \frac{P_t}{P_{t-1}}^{\frac{1}{10}}$, one period represents ten years.

more.[18] Because these counties are positively selected there is also reason to worry that after this positive section there may be regression to the mean—the estimates of a county's pre-transportation patenting levels may be inflated—and thus the estimates presented above may have downward bias. In addition, though the location of the transportation lines is measured quite precisely, the locations of the places where people might interact with a train or boat (i.e., stops) are not measured. Thus, there may be attenuation bias. To address this, I propose a "straight line" instrument for the spread of the railroad across the country.[19]

Figure 7 shows the lines that are used in the instrument. These originate from the 14 port cities with a customs house or public warehouse that had been built by the US government by 1826 (Congress, 1826), which were the largest ports, and run to the largest city or county in every state in the Midwest or western part of the South in 1830—five years before the first railroad was built in the United States. Lines are restricted so that, with the exception of those originating from New Orleans, only cities with approximately the same latitude are connected. Note the way these lines correspond with where population had settled by the 1860s, heavily in upstate New York, but not proceeding onward to Michigan, across the southern mid-west (Ohio, Indian, and Illinois). This means that most of the variation in the instrument is coming from the industrialized North, and the local treatment effect may be larger than in the more agrarian South. The few lines running north from New Orleans capture those that settled along the Mississippi. The variable is defined at the county level: a county takes on the value of one always if in 1810 more than 1% of the county is within 1.5 miles of water transportation, one in 1850 on if three or more of these lines intersects it, one in 1860 on if two or more of these lines intersect it, a value of one in 1870 if more than

[18] Atack et al. (2010) address the question of if the railroad was built "ahead of demand." They find that it followed economic growth but increased urbanization.

[19] This instrument owes its inspiration to **?**, which uses lines between pre-existing cities in Germany to predict the spread of the railroad, and Michaels (2008), which uses the orientation of a line between the nearest city and a county centroid to predict the presence of an interstate highway.

one line intersects it, and zeros otherwise.[20]

Table 5 shows fixed effects regressions following equation 2 and the first stage and two stage least squares estimation given by:

$$TransportationMeasure_{it} = \alpha + \beta ConnectionLine_{it} + \varphi \mathbf{X_{it}} + \gamma_i + \delta_t + T.M._{i(t-N)} \times \delta_t + \varepsilon \quad (4)$$

and

$$PatentMeasure_{it} = \alpha + \beta \widehat{TransportMeasure}_{it} + \varphi \mathbf{X_{it}} + \gamma_i + \delta_t + T.M._{i(t-N)} \times \delta_t + \varepsilon \quad (5)$$

where $PatentMeasure_{it}$ is the number of patents issued per 10,000 people; $TransportationAccess_{it}$ (abbreviated $T.M._{it}$) is the measure of transportation access which will generally be the percent of the county's area within 5 miles of a railroad; $ConnectionLine_{it}$ is the indicator for whether or not one of the lines described above intersects a county and the year is as described above; $\mathbf{X_{it}}$ are county characteristics; γ_i are county fixed effects, δ_t are year fixed effects, and $t - N$ denotes the use of all previously observed values.[21]

Because the variation in the instrument is in the years 1850 through 1870 Table 5 restricts the sample to 1840-1870.[22] When a full set of controls is used the instrument is not as strong as one might prefer, and the coefficient on local transportation access is very imprecisely estimated. The point estimate of this coefficient is similar to the one without county level controls, which is much more precisely estimated. The predicted growth in patents is consistent across the two tables. All alternate specifications examined suggest even larger coefficients. Using the estimates from Table 5 with controls, about 20% of the change in

[20] Results from a variation on this instrument where counties take on values based on lines from the port cites that run due west and north available upon request or on my website.

[21] All IV regressions are done using xtivreg2 (Schaffer, 2010).

[22] Tables showing the full sample available upon request or on my website.

patenting between 1840 and 1870, can be explained by local changes in transportation access. The fixed effects estimates, by contrast, suggest 2% of the increases in patenting were due to changes in local transportation access.

5 Mechanisms

There are many reasons why an increase in local transportation access may lead to more patenting activity.[23] Demand for innovations may increase once an area is opened up to a larger market, or with more transportation there might be an increase in the supply of novel ideas due to increased information flows. Things having more to do with the propensity to patent, rather than the propensity to innovate might also cause the phenomenon examined above. Below I will explore the relationship between patenting and a measure of market access (Section 5.1) and a measure of information flows (Section 5.2). Some exploration of the relationship between transportation and patenting for different groups of counties can be found in Appendices on my website.

5.1 Demand: Market Access

The hypothesis in Bustos (2011), Lileeva and Trefler (2010), and Sokoloff (1988) was that increased market access leads to more people being willing to pay a fixed cost for innovation.

[23] Sokoloff's main hypothesis falls in line with the intuition gained from the Bustos (2011) and Lileeva and Trefler (2010) extensions to the Melitz (2003) model, which suggest that, upon gaining access to larger markets, it is worthwhile for firms to enter or invest in productivity upgrading. This is an because increase in units sold allows them to recoup a larger investment. Or as Sokoloff (1988) puts it: the effect is likely caused by "individuals and firms choosing to commit additional resources to a search for useful discoveries in response to the increase in the expected return to such investment that stems from their integration into a larger pool of both potential customers and competing suppliers." However, Sokoloff (1988) also gives space to alternative reasons for patenting to increase, noting that "gaining low-cost access to a large market could alter behavior through changes in the prices of goods or in the returns to activities, and thus nurture cultural attitudes more favorable to invention, enhance learning-by-doing, improve the flow of information to potential inventors, raise the amount of resources available for allocation to invention, increase the propensity to invest in general, or foster such changes in methods or in the extent of factor specialization as to facilitate the discovery of possible refinements in technique or other inventions... as well as the improvement in the stock of knowledge... would help account for the patterns in the data."

Above I explored the relationship between local transportation access and patenting, this section explores a measure of transportation access driven by expansions elsewhere in the network. The construction of this measure of market access is described below.

5.1.1 Data: Approximating Market Access

The question of how access to a larger market influences innovation can be more directly addressed by approximating the size of the market that is within easy reach of a location. Equation 9 is the starting point for this approximation. It is worth noting that if counties are thought of as nodes in a network, where each node (county) is connected to the nodes (counties) physically adjacent to it, this market access approximation is a measure of closeness centrality (Rochat, 2009).

Several approximations must be made to apply Equation 9 to the data available for the nineteenth century United States, where there are no good estimates of counties' incomes nor of the trade flows between counties. First is the estimated transportation cost between i and j, which I will discuss below. Second, since GDP is unavailable at the county level, population is used as a crude proxy for income.[24] There are alternate county level measures that one might use instead of population, for instance the access of a county to nearby patenting is also explored in an Appendix available upon request, or on my website.

Market access for a county, MA_i, can be approximated as:

$$MA_i \approx \sum_j pop_j \cdot c_{ij}^{-\theta} \tag{6}$$

where pop_i is the population of location i, c_{ij} is the resistance term between i and j (i.e.,

[24] One can think of the market available to a firm as the number of people that it can reach with its product; people are out of reach if transportation costs make the product unfeasibly expensive. So a firm's reach expands with falling transportation costs. This reach can be estimated by a cost weighted sum of the people in all locations.

the transportation cost between i and j), and $\theta = \sigma - 1$. The same formula is used in Donaldson and Hornbeck (2013). Market access can rise either when c_{ij} falls, or when pop_i increases; over the nineteenth century both the general cost of transportation and the population level will change substantially. While the population that a county contains is an important component of market access, it independently affects the patenting rate.[25] I also explore taking the cost term to be only related to the straight line distance between county centroids. I refer to this as an "as the crow flies measure," that only varies because of population movement, as the relationship between counties is fixed.

Figure 8 shows approximate market access in 1830 and 1870; notice the high computed market access along the coast, the low computed market access in Appalachia, and the increasing importance of the railroad over time.

Transportation cost is computed by a procedure that assumes that each county is only linked to those counties that are physically adjacent to it (or accessible through the network of water transportation if the county contains a port). This can be thought of as a network where counties are nodes and links exist where physically adjacency exits or there is a waterway connection. This procedure will move along the network, starting from a seed county, to compute a transportation cost from every county to that seed county.

The algorithm starts from a reference county and computes approximate costs from this reference to all adjacent counties. The cost of getting from the seed county to each adjacent county is then conditionally updated with a new cost. This new cost adds the cost that was just computed from the reference county to this adjacent county to the cost that was previously computed between the seed county and the reference county. The information on transportation cost to this adjacent county is updated if the newly calculated travel cost is lower than the previous value, or if it had no previous value. If the cost is updated, this

[25] Indeed, since I measure patenting as patents per capita, it mechanically effects this variable.

adjacent county is added to a queue. Counties are popped from the queue one at a time and each, in turn, is taken to be the reference county. This procedure repeats until the queue is empty. This procedure yields the minimum transit costs from the original seed county to all other counties.

The information on the transportation network is created by joining transportation data (Atack, 2013) with the 1840 United States county boundary shapefile from NHGIS. Each county is then denoted as having a railroad, river, canal, or port. If two adjacent counties have one of these features, they are assumed to be connected by that mode of transportation. Starting from the county centroid in question (reference county) I give approximate costs to every county adjacent to it using centroid to centroid distances (if less than 150 miles from each other) and rates taken from Donaldson and Hornbeck (2013), which takes them from Fogel (1964).[26] I restrict the cost between county pairs so that it never increases between two observations, I compare my computed cost for year t to the one from $t-1$ and I assign in year t the lower of the two costs. While the adjacency rules in this procedure are not as precise as using ArcGIS's Network Analyst, the data that are used need not have the level of precision necessary for that program to produce results. All port counties are assumed to be adjacent to all other port counties, with a distance computed by using modern waterway network data, which is compatible with ArcGIS's Network Analyst, from the National Transportation Atlas Databases. Other county adjacency is computed by modifying a tool written by Chieko Maene (Maene, 2011).

[26] Travel along a railroad costs 0.63 cents per ton-mile, a canal 0.50 cents a ton-mile, river or other waterway 0.49 cents per ton-mile, wagon or overland 23.1 cents per ton-mile, and changing mode of transportation 50 cents per ton (e.g., unloading from a rail car onto a river barge). Also following Donaldson and Hornbeck (2013) I take θ to be 3.8. I test both the conventional 1.0 and a suggested 8.22 and I find 3.8 fits my patents per capita data the best.

5.1.2 Results

Tables 9 and 7 use the specification described in Equation 1, using estimated market access[27] as the transportation variable. Table 9 tests the impact of adding different sets of controls to the estimated relationship between market access and patenting. Controlling for the lagged percent of county's farm acres seems, in particular to reduce the coefficient on market access, loading more weight onto a county's own population. Several variables seem to reduce the coefficient on population: the lagged percent of county that is engaged in manufacturing, the lagged percent of the county that lives in an urban area, and the lagged percent of the county that is foreign born and the lagged percent of the county that is born out of state. All of these things are higher in more populous areas, so this apparent correlation is not surprising. Nor is the relationship between market access and farming, as this has long been established (Fogel, 1964; Donaldson and Hornbeck, 2013); it is not clear if the channel of increase in patenting is an income effect, access to larger markets, or some other thing. The inclusion of the percent of the county that is enrolled in school, or the percent of the native born, over 21, white male population that is literate also cause a drop in the coefficient on market access. These have been shown to be related to increased access to transportation (Atack et al., 2012), and are also related to income from the economic growth caused by yields from farming in the Midwest. When all controls are added, nether of the coefficient on market access nor the coefficient on population is large or precisely estimated.

In Table 7, columns (1) and (4) use local transportation access as the variable of interest, replicating columns (3) and (4) of Table 3. The next set of columns, (2) and (5) use estimated market access as the transportation variable. Like local transportation access, the point estimate drops, and it is less precisely estimated when controls are included.[28]

[27] This estimate includes the county's own population in the summation, and $\theta = 3.8$

[28] The same instrument as used above can be used for market access that is transportation cost weighted.

The largest gains in market access are from counties that had no form of improved transportation receiving their first connection to the larger network. Thus, columns (3) and (6) of Table 7 put the local transportation access measure and the estimation of market access in the same regression, thus examining how expansions of the network and local connections relate to patenting while controlling for the other. Transportation access and market access are weakly correlated in the beginning of the nineteenth century, becoming more strongly correlated as the century progresses. Also, as seen in Table 1, the variance in this measure decreases with time. Thus it is somewhat unclear how to interpret the imprecisely estimated negative coefficient in column (3). When the full set of county controls are added in column (6), the only coefficient that does not appear to be different from zero is local transportation access. Overall, this suggests that it is increased transportation in the local area, and not in other locations on the network that is important for the relationship between patenting and transportation.

Exploring the relationship between the previous patenting level in a county and the increase in patenting, Table 8 uses the specification described in Equation 3. Note that as in Table 4, regression to the mean in patenting levels is seen, similarly, increasing market access increases geographic dispersion—places with more patenting have less increase in patenting due to increases in market access. The overall effect of increased market access in column (2) of panel one in Table 8 seems to be negative, however it is positive in panel two and in columns (3) and (4). The average calculated marginal effect of an increase in market access is, in all cases, close to zero, suggesting there is little over all impact of increased market access.

Consistent with the IV estimates of local transportation access, these estimates increase. Because this measure of market access is highly imprecise, attenuation bias may be significant. The point estimate in the IV specification without controls would imply that changes in market access explain 7.5% of the change in patenting between 1840 and 1870, while the point estimate (which is very imprecisely estimated) in in the IV specification with controls would imply that it explains 40% of this increase.

5.2 Information Flows

5.2.1 Data: Speed at Which New Words Arrive

Some patents represent a more important innovative contribution than others; using a pure count variable (as I have done thus far) implicitly assumes that each patent is equivalent.[29] The text of each patent contains a great deal of information about the contribution the patent represents. Automated text analysis allows access to this information (in the about 700,000 patents issued 1836-1900).

I examine the speed at which new words (1-grams) or two-word phrases (2-grams), here collectively referred to as n-grams, move through the patent record. If in a particular county one n-gram appears one year after it is first used in a patent anywhere, then this county is measured as receiving new words at the rate of one new word a year.

I start at the beginning of the existent record (1836), and find the first appearance of an n-gram anywhere in the record, as well as the first appearance in any given county. The number of years between the first appearance of an n-gram anywhere and the n-gram's first appearance in a county, referred to as the time lag, is a measure of how long it takes a new concept to be innovated upon or used in an innovation in any one place. For example, Figure 9 shows the counties in which the word "vulcanized"[30] appeared in by 1850, 1860, and 1880. The word spreads from locations that are recognizable centers of innovation to counties that are dispersed across the county.

To save on computation I do not use every word that has appeared in the patent record,

[29] People have addressed this in a number of ways when working with modern patents, the most popular way is to use patent citations. However, patents did not start citing each other until the twenty century.

[30] The vulcanization of rubber was patented in the US–though not in Britain, where it was patented by someone else in what appears to be case of simultaneous invention–by Charles Goodyear in 1844 (US Patent No. 3633) from New York, NY. This patent did not use the term vulcanized, however an 1848 patent by Charles Goodyear (moved to New Haven, CT) did, as did two other patents, from New York, NY and Worcester, MA, in that year.

but rather a list of about 4,000 n-grams that was generated by a) looking for n-grams that did not appear early on but were common later, b) picking n-grams that are key to important patents of the nineteenth century as found in Brown (1994); United States Patent and Trademark Office (1981); van Dulken (2001) both by hand and algorithmically, and c) finding synonyms of the concepts from the important patents in the Oxford English Dictionary's Historical Thesaurus. The word list can be found on my website.

In a given year, for each county I compute the number of n-grams that made their first appearance in that county in that year. I then compute a speed of word arrival for that county-year as:

$$Speed_{it} = \sum_{T=0}^{N} \frac{\text{Number of New Words}_{itT}}{(T+1)} \tag{7}$$

where N is the number of years of time lag used in the computation (e.g., 10 years), Number of New Words$_{itT}$ is the number of n-grams observed in county i in year t with time lag T[31] (e.g., 3 n-grams first appear in New York, NY, in 1850 with a time lag of 6 years).[32]

Figure 10 shows a map of speeds in each county in 1850 and 1870 computed with $N = 10$, the zeros seen in these figures are places that had only patents that did not use any n-grams that were new to the patent record in the last ten years. Because both patents and new word appearances are rare, I have used three year bins for the analysis in this paper.[33] Further, to compensate for the general downward trend of this statistic over time I have expressed speeds as a ratio. Speeds are always values between zero and one: the computed speed for the county over the largest speed computed for that year. Of particular note is the great increase in patenting between 1850 and 1870, and that the vast majority of those counties

[31] I use $T + 1$ as the denominator of the above sum so that I never divide by zero.

[32] They are: gasometer, plastic, and printing plate.

[33] Thus, $Speed_{i1850}$ is taken to mean $Speed_{i1849} + Speed_{i1850} + Speed_{i1851}$.

are not ones that use words that were new in the past ten years. The highest speed counties seem to be more concentrated in 1870, clustered near areas of high population, with fewer high speed counties along corridors such as the Erie Canal.

This paper uses the ratio of three year bins as described above with both $N = 10$ and all years. The later means that there is no fixed number of years after which a word is no longer considered new,[34] however I count all words that appear before 1842 as "old" words.

5.2.2 Results

Improved transportation did not only have the effect of decreasing freight rates, but decreased the cost and increased the speed of individual travel and the movement of the mails. A local connection to transportation may, thus, increase a county's access to innovative ideas. The increase in patenting observed above may be because people are learning about things near the technology frontier more quickly (see Appendix A.4 for a simple framework to motivate this). I cannot observe the movement of ideas directly, but the text of patents grants contain a great deal of information about the innovation in question, and I can observe the words inventors use to discuss their technologies. The second use of a new word (or two word n-gram) in the patent records suggests that the inventor authoring the patent that contains this word is part of an information network that transmitted knowledge of this new concept. I can glean some understanding of how ideas move by looking at where and when words appear in the patent record. I have constructed a measure how many new words are appear in the patent record of any given county; if in a particular county one new word appears in one year after it is first used, than this county is measured to receive new words at the rate of one new word a year. This measure of the speed of word arrival is discussed in more detail in Section 5.2.1.

[34] This means the maximum time lag possible depends on the year the statistic is computed for, as the data start in 1836 and I consider words that appeared before 1842 as "old", the maximum time lag the data allow for words appearing in a county in 1850 is 8 years, and in 1860 is 18 years.

Figure 11 plots the coefficients from the regression described in Equation 1 (as is done in Figure 5), with two measures of speed used as the dependent variable. One measure only considers words new if they appeared in the last ten years ($N = 10$),[35] while the other considers words new if they appeared any time after 1842 (the record of patent text starts in 1836). Note that including the counties that don't patent as counties observed with a speed of zero changes the precision of the estimate, but does not meaningfully change the point estimate. There appears to be no relationship between the speed of word arrival when $N = 10$, but there seems to be a positive change in the slope when there is no cut off for when a word is considered new.

Table 10 shows the basic fixed effects regression from above (Equation 2), with speed of word arrival as the dependent variable.[36] When county level controls are included, the effect appears to be an imprecisely estimated zero. The second panel in Table 10 shows market access; though positive and precisely estimated, these coefficients imply very small changes in the speed of word arrival.

Table 11 puts the local transportation access measure and the estimation of market access in the same regression (like Table 7) with the speed of word arrival as the measure of patenting. The only coefficient that is consistently positive and precisely estimated is the one on population; the coefficient on local transportation is consistently negative. When controls are added the coefficient on Market Access becomes precisely estimated, it remains positive and small.

One common way of studying technology diffusion is using a distance weighted measure of the "stock" of technology that different countries possess. The cost and distance weights

[35] This means there is no overlap in the words used between consecutive observations.

[36] IV specifications following Equation 5 and using the instrument described in Section 4.1 can be found on my website. The Market Access ones are not shown because the instrument has no power in those regressions.

can be used to measure a county's "patent access," if rather than being used to form a weighted sum of the country's population, they are used to create a weighted sum of the stock of patents issued to inventors in a county. This is explored further in an Appendix on my website.

The arrival of the railroad encourages a county to have an increased number of patents, as shown above, but these patents were not incorporating new technology. However, being close to large populations encouraged patenters to use new technology in their innovations. These effects are likely different for counties in the core than for counties on the periphery, this is explored in Appendices on my website.

5.2.3 Constructing Synthetic Counties

Speed of word arrival is influenced by the content of patents, however, each additional patent represents gives counties a greater chance of mentioning a new technologies. If the data generating process is purely random draws of words from a pool of words, there is a positive probability of drawing a word that is new on every draw. Thus, having more draws will lead to more new words. With this data generating process patent counts would be a sufficient statistic for the speed with which new words arrive. I try to disentangle what aspect of the computed speed of word arrival is the number of patents (draws), rather than the content of theses patents by computing counterfactual relationships produced through random selection.

To test the importance the content of the patents in a county, rather than the sheer number of patents, I construct counterfactual distributions of patents where the number of patents in a county is held fixed, but the patents a randomly chosen. To construct a counterfactual distribution I count the number of patents issued by a county, and then randomly draw that number of patents from the pool of patents issued in the three year period under consideration. I refer to this group of randomly drawn patents as the patents in a synthetic county. This process of creating a simulated distribution was repeated forty times, so that there are at least forty observations for every existent county size, year combination.

The degree to which the speed of word arrival is driven by the number of patents produced by a count are explored in Figures 12 and 13. Figure 12 presents histograms of the speed (with N=100) computed in these randomly drawn counties and observed counties in a year, by the number of patents in these counties. The first column shows counties with only one patent; in 1850 it appears that these two distributions are the same, but in 1880 the real data are completely missing observations in the right tail. Similarly, the 1880 distributions for the speed of word arrival in counties with between 9 and 11 patents or between 99 and 104 patents seem to be missing faster observations. One cannot reject the null, using the Kolmogorov-Smirnov test, that the distribution of speeds of randomly drawn counties with one patent is the same for 1850. However, for 1880 the Kolmogorov-Smirnov test rejects the null at the level 3%, suggesting that patents from real counties contain fewer new words than randomly drawn patents from that year. For next two columns, the pattern is the same: the null cannot be rejected in 1850, but is rejected at the 1% level in 1880.

The full relationship between the number of patents and the speed of word arrival in a county is examined in Figure 13. The blue area represents to 25th to 75th percentile range of speeds for counties with a given number of patents; the red area repents the same for randomly drawn counties. In 1850 the distribution of speeds for real and constructed counties appears the same, whereas in 1880 the distribution of real counties appears slower than the distribution of randomly drawn counties. Again, using the Kolmogorov-Smirnov test, one cannot reject the null in 1850, but can reject it at above the 1% level in 1880. Table 12 shows the P-Values for every year, the real distribution and the simulated one become more disparate as the nineteenth century progresses.

Figure 14 plots histograms of coefficients computed for counterfactual data. Each of the forty observations is derived from a set of synthetic counties, counties with the observed number of patents, but with each of these patents being a random draw from the year in question (see section 5.2.3). The first panel uses speed of word arrival computed with words new in the last ten years, plots the coefficients from the regression from the first column of Table 10, the panel directly below it corresponds to column (3) of Table 10. The two adjacent panels correspond to columns (2) and (4) respectively. While none of the coefficients from the actual distribution of patents are strongly outside the distribution of counterfactual coefficients, they are all clearly on the bottom end of the distribution. This suggests, as do Figures 12 and 13, that the patents in the counties with increasing levels of transportation (which will, in general, be on the periphery) refer to technology that is older than the technology discussed in randomly selected patents.

These results are consistent with transportation leading to a decreased cost of patenting. They are also consistent with a model in which areas innovate on their local production using methods that are distinctive. Gaining access to transportation raises the number of patents a county produces, while the direct impact of market access is unclear.

6 Conclusion

I find a robust, statistically significant, and positive effect of increases in local transportation access on patenting. I sought to establish a richer empirical basis for considering market access as a driver of innovation, starting by establishing the relationship between local transportation access and patenting. Using a large dataset of geo-located patents cross-referenced with a map of the expansion of transportation infrastructure, I show that the arrival of improved transport, primary the railroad, has a positive effect on patenting behavior. Gains are realized slowly over time, suggesting that transportation access causes a trend change in the overall rate of development rather than a sudden innovative shock, and it is increases in very local (within 5 miles) improved transportation access that drive the increase in patenting. This expansion in patenting is largely due to locations registering their first patents.

Today most patenting comes from the Research and Development investments that firms make. Though the patents, by law, are still issued to individual inventors, these individuals have likely contracted to automatically assign their intellectual property to their employer. In the nineteenth century such arrangements were rare, with most inventors self-financing, and only later licensing or assigning the rights to their inventions to others.[37] Thus, these inventions represented individuals investing in a technology that was to be the basis of a business, or in the intellectual property for its value when sold or licensed to others.

Transportation helps encourage general economic development, which in turn leads to more resources available for inventive activity. Because transportation encourages urbanization, it encourages patenting in many ways. Urban areas allow for greater specialization, which might encourage patenting by giving people in those areas better access to the bureaucracy of patenting (e.g., lawyers, machinists, draftspeople), or by encouraging innovation directly. Urban areas also allow for better access to formal credit markets, lifting liquidity constraints for potential innovators. Finally, in an urban area, secrecy may not effectively protect inventions, leading to patenting.[38]

I then test the hypothesis that increased market access due to the transportation revolution was a key driver of innovation for the United States in the nineteenth century (Sokoloff, 1988), as well as the hypothesis that increasing transportation access increased information

[37] Of the 21 important inventions as defined by books on the subject (Brown, 1994; United States Patent and Trademark Office, 1981; van Dulken, 2001) patented before 1860, 3 were assigned at issue: guncotton, the machine gun, and the safety pin. Patents on the vulcanization of rubber, rotary printing press, and automatic sewing machines were among those not assigned at issue.

[38] It is not *a priori* obvious that a transportation link will lead to more innovation in peripheral areas. It might also have increased the importance of being in the center of the network, or lead to human capital flight by providing an easier way to migrate to urban areas where high human capital is better rewarded.

flows that drove patenting.

Increases in the estimate of market access that I create do not show this relationship. When both are included, it is local transportation access that maintains a positive, precisely estimated coefficient. However, because most connections in this period connected less developed places to more developed places, it does not follow that market access was unimportant for innovation. It may mean that in the United States during the nineteenth century, local connections were responsible for the majority of the changes in transportation costs, but it does cast doubt on the market access hypothesis.

In contrast to the strong relationship local transportation access has with patents per capita, increases in local transportation access has no effect on how quickly words representing new technologies appear in a county's patents. These results are consistent with a model where inventors are the heads of firms, producing patents as their product and decreases in transportation costs encourage more people to engage in patenting. However, because larger barriers lead to more positively selected firms (those expecting lower returns will not profit after paying a larger fixed cost, as in Melitz (2003)), these new patenters will, in general, produce patents of a lower quality. Once an area is developed, increased competition from more market access will shift patenting the other direction, with the result being that an area that can more easily reach broad markets will produce higher quality patents.

It appears that local transportation access is related to increases in patenting primarily because transportation forms a nexus around which local agglomerations occur. That patent quality does not increase when local transportation access increases but quantity does is a result consistent with these agglomerations facilitating patenting by reducing the effective cost of participating in the formal intellectual property system.

7 References

Agrawal, A., Galasso, A., and Oettl, A. (2014). Roads and Innovation. CEPR Discussion Papers 10113, C.E.P.R. Discussion Papers. (On page 5.)

Agrawal, A. and Goldfarb, A. (2008). Restructuring Research: Communication Costs and the Democratization of University Innovation. *The American Economic Review*, 98(4):1578–1590. (On page 5.)

Akcigit, U., Kerr, W. R., and Nicholas, T. (2013). The Mechanics of Endogenous Innovation and Growth: Evidence from Historical U.S. Patents. (On pages 3, 10.)

Amiti, M. and Konings, J. (2004). Trade Liberalization, Intermediate Inputs, and Productivity : Evidence from Indonesia. *American Economic Review*. (On pages 4, 42.)

Atack, J. (2013). On the Use of Geographic Information Systems in Economic History: The American Transportation Revolution Revisited. *The Journal of Economic History*, 73:313–338. (On pages 3, 11, 21, 60, 61, 63, 64, 65, 67, 68, 69.)

Atack, J., Bateman, F., Haines, M., and Margo, R. A. (2010). Did Railroads Induce or Follow Economic Growth?: Urbanization and Population Growth in the American Midwest, 1850-1860. *Social Science History*, 34(2):171–197. (On pages 2, 9, 16, 36.)

Atack, J., Haines, M., and Margo, R. A. (2011). Railroads and the Rise of the Factory: Evidence for the United States, 1850-70. pages 162–179. (On pages 2, 36, 44.)

Atack, J., Jaremski, M., and Rousseau, P. L. (2014). American Banking and the Transportation Revolution before the Civil War. *The Journal of Economic History*, 74:943–986. (On pages 2, 36.)

Atack, J., Margo, R., and Perlman, E. (2012). The Impact of Railroads on School Enrollment in Nineteenth Century America. (On page 22.)

Audretsch, D. B. and Feldman, M. P. (2004). Knowledge Spillovers and the Geography of Innovation. *Handbook of Regional and Urban Economics*, 4(04). (On page 5.)

Becker, S. O. and Egger, P. H. (2013). Endogenous Product Versus Process Innovation and a Firm's Propensity to Export. *Empirical Economics*, pages 1–26. (On pages 4, 42.)

Bogart, D. (2005). Turnpike Trusts and the Transportation Revolution in 18th Century England. *Explorations in Economic History*, 42(4):479 – 508. (On page 9.)

Bracha, O. (2004). The Commodification of Patents 1600-1836: How Patents Became Rights and Why We Should Care. *Loyola of Los Angeles Law Review*, 38:177. (On pages 7, 38.)

Brown, T. (1994). *Historical First Patents: the First United States Patent for Many Everyday Things*. Scarecrow Press. (On pages 25, 31.)

Bustos, P. (2011). Trade Liberalization, Exports, and Technology Upgrading: Evidence on the Impact of Mercosur on Argentinian Firms. *The American Economic Review*, 101(February):304–340. (On pages 4, 18, 42, 44.)

Congress, U. (1826). H.R. Doc. No. 93, 19th Cong., 1st Sess. Technical report. (On page 16.)

Deloecker, J. (2007). Do Exports Generate Higher Productivity? Evidence from Slovenia. *Journal of International Economics*, 73(1):69–98. (On pages 4, 42.)

Dhingra, S. (2013). Trading Away Wide Brands for Cheap Brands. *American Economic Review*, 103(6):2554–84. (On page 42.)

Donaldson, D. and Hornbeck, R. (2013). Railroads and American Economic Growth: a "Market Access"Approach. (On pages 2, 4, 20, 21, 22, 36.)

Feldman, M. P. and Kogler, D. F. (2010). Stylized Facts in the Geography of Innovation. *Handbook of the Economics of Innovation*, 01(10):381–410. (On page 5.)

Fernandes, A. M. (2007). Trade Policy, Trade Volumes and Plant-Level Productivity in Colombian Manufacturing Industries. *Journal of International Economics*, 71(1):52–71. (On pages 4, 42.)

Feyrer, J. (2009). Distance, Trade, and Income: The 1967 to 1975 Closing of the Suez Canal as a Natural Experiment. (On page 41.)

Fogel, R. W. (1964). *Railroads and American Economic Growth: Essays in Econometric History*. Johns Hopkins University Press, Baltimore, MD. (On pages 2, 21, 22, 36.)

Forman, C., Goldfarb, A., and Greenstein, S. (2014). *Information Technology and the Distribution of Inventive Activity*, pages 169–196. University of Chicago Press. (On pages 5, 15.)

Foster, L., Haltiwanger, J., and Syverson, C. (2008). Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability? *American Economic Review*, 98(1):394–425. (On pages 4, 42.)

Friedman, T. L. (2006). *The World Is Flat: a Brief History of the Twenty-First Century*. Farrar, Straus and Giroux, New York. Index. (On page 2.)

Haines, M. R. (2010). Historical, Demographic, Economic, and Social Data: The United States, 1790-2002 [Computer file]. (On pages 11, 60, 61, 63, 64, 65, 67, 68, 69.)

Harris, C. D. (1954). The Market as a Factor in the Localization of Industry in the United States. *Annals of the Association of American Geographers*, 44(4):315–348. (On page 4.)

Hornbeck, R. (2010). Barbed Wire: Property Rights and Agricultural Development. *The Quarterly Journal of Economics*, 125(2):767–810. (On pages 11, 60, 61, 63, 64, 65, 67, 68, 69.)

Howe, D. W. (2007). *What Hath God Wrought: The Transformation of America, 1815-1848*. Oxford History of the United States. Oxford University Press. (On page 2.)

Hrdy, C. A. (2013). State Patent Laws in the Age of Laissez-Faire. *Berkeley Technology Law Journal*, 28(1). (On pages 7, 37, 42.)

Hudson, J. C. (1985). *Plains Country Towns*. University of Minnesota Press. (On pages 2, 36.)

Jaffe, A. B., Trajtenberg, M., and Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics*, 108(3):577–98. (On pages 2, 5.)

Khan, B. Z. (2005). *The Democratization of Invention: Patents and Copyrights in American Economic Development*. Cambridge University Press, Cambridge, MA. (On page 7.)

Lamoreaux, N. R. and Sokoloff, K. L. (2005). The Decline of the Independent Inventor: A Schumpterian Story? (On page 8.)

Lileeva, A. and Trefler, D. (2010). Improved Access to Foreign Markets Raises Plant-Level Productivity... for Some Plants. *The Quarterly Journal of Economics*, 125(3):1051–1099. (On pages 4, 18, 42, 44.)

Maene, C. (2011). python ArcGIS Find Adjacent Tool. (On page 21.)

Marshall, A. (1890). *Principles of Economics*. Macmillan and Co., New York. (On page 2.)

Melitz, M. J. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica*, 71(6):1695–1725. (On pages 18, 32, 42, 44.)

Michaels, G. (2008). The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System. *Review of Economics and Statistics*, 90(4):683–701. (On page 16.)

Minnesota Population Center (2011). National Historical Geographic Information System. No citations.

Moser, P. (2004). Determinants of Innovation Evidence from 19th Century World Fairs. *The Journal of Economic History*, 64(2):548–552. (On pages 8, 10.)

Pavcnik, N. (2002). Trade liberalization, exit, and productivity improvements: Evidence from Chilean plants. *The Review of Economics Studies*, 69. (On pages 4, 42.)

Rochat, Y. (2009). Closeness Centrality Extended to Unconnected Graphs: The Harmonic Centrality Index. *Applications of Social Network Analysis*. (On pages 5, 19.)

Schaffer, M. (2010). xtivreg2: Stata Module to Perform Extended IV/2SLS, GMM and AC/HAC, LIML and k-class Regression for Panel Data Models. (Stata .ado). (On page 17.)

Sokoloff, K. L. (1988). Inventive Activity in Early Industrial America: Evidence From Patent Records, 1790–1846. *The Journal of Economic History*, 48(04):813–850. (On pages 3, 4, 18, 31.)

Topalova, P. and Khandelwal, A. (2011). Trade Liberalization and Firm Productivity: the Case of India. *Review of Economics and Statistics*, 93(3):995–1009. (On pages 4, 42.)

Trajtenberg, M. (1990). A Penny for Your Quotes: Patent Citations and the Value of Innovations. *The RAND Journal of Economics*, 21(1):172. (On page 10.)

United States Patent and Trademark Office (1981). *The Story of the U.S. Patent and Trademark Office*. U.S. Department of Commerce, Patent and Trademark Office. (On pages 25, 31.)

van Biesebroeck, J. (2005). Exporting Raises Productivity in Sub-Saharan African Manufacturing Firms. *Journal of International Economics*, 67(2):373–391. (On pages 4, 42.)

van Dulken, S. (2001). *Inventing the 19th Century: 100 Inventions that Shaped the Victorian Age from Aspirin to the Zeppelin*. NYU Press. (On pages 25, 31.)

A Appendices

A.1 Years Since Rail

The spread of the railroad, in particular, and improved transpiration in general, changed the character of the areas around them. They provided loci for new towns—increasing urbanization (Atack et al., 2010), attracting banks (Atack et al., 2014), and encouraging speculators to plat towns (Hudson, 1985). They promoted economic growth by linking far-flung factor and product markets, encouraging the exploitation of regional comparative advantage (Fogel, 1964; Atack et al., 2011; Donaldson and Hornbeck, 2013). Some of these changes can be seen in an event study looking at impact of railroad (or transportation) arrival on county level outcomes. The below figures estimate Equitation 1 using other county characteristics as the outcomes of interest.

Restating the result found in (Atack et al., 2010), Figure 15 plots Equitation 1 estimated on the percent of a county's population that lives in an urban area. While there does appear to be some positive selection on places that have been becoming more urban before the railroad arrives, the growth that happens only once the railroad arrives is dramatic. In contrast, population growth, as seen in Figure 16 levels off upon railroad arival and has no clear break in trend with the overall arrival of transportation. When the logarithum of the number of people living in an urban area is used as the dependent vrable, the patern is very simalur to the precent of the county that lives in an uerban area, see Table 17.

This increase in urbanization corresponds to an increase in the percent of a county's population that is employed in manufacturing, Figure 19, and a decrease in the percent employed in agriculture (data from 1850-1880, 1900 calculated from the 1% sample), Figure 18. At the same time the amount of improved land employed in agriculture increased, Figure 20. Both the prcent of a county employeed in banking, Table 21, and the telegrpah industry, Table 22, appear to have increased, though this is less clear

Two things that don't seem to have responded to the arrival of the railroad that one might have expected to are the percent of the population that is foreign born, Figure 23– perhaps because domestic migration becomes differently more attractive, and the percent of the population that is literate, Figure 24.

A.2 Notable Developments in Patent Law

United States federal patents were introduced in 1790, as allowed by the Constitution. Previously the colonies individually granted patents (Hrdy, 2013) in a manner more typical of the royal monopolies they emulated, with an *ad hoc* process of petitions. When federal patents were introduced, inventors were given the opportunity to apply for federal patents on things they had previously patented at the state level; the last state patent was issued in 1798. In 1790 federal patent applications were examined by the Secretary of State, the Secretary of War, and the Attorney General for both novelty and to see "if they deem the invention or discovery sufficiently useful and important." In practice this meant Thomas Jefferson examined patent applications. Patents were granted for a maximum of 14 years, the board was to determine the grant length of each patent, the fee was between \$4 and \$5. A specification and drawing were required, but a model was encouraged. There was no official channel to appeal decisions. The board of Secretaries (and, as noted above, in particular Thomas Jefferson) had trouble balancing their many other duties and examination of patent applications, which led to the removal of examination in 1793. It was reintroduced in 1836.

In 1793, the requirement that an invention be "sufficiently useful and important" was removed and patents were no longer examined for novelty.[39] The fee was increased to \$30, and aliens were not allowed to obtain patents. In 1800, this requirement was changed so the aliens that have resided in the United States for 2 years and declare an intention of becoming a citizen may receive patents.

The next major change to patent law occurred in 1836, with the patent office established as a distinct bureau that is charged with examining patent application (the patent office consisted of a commissioner, a chief clerk, an examiner, a machinist, two draftsmen, an inferior clerk, and a messenger). Patent office employees are forbidden to acquire any interest in a patent, and a library of scientific works for use by employees of the patent office is created. In addition to the grant length of 14 years, the option to apply for a 7 year extension is made available. The fee remained \$30 for citizens, but foreigners are allowed to hold patents with fees of \$500 for British subjects, \$300 for all others. Applicants must file a specification, a drawing, and a model, and any appeals made be made to three "disinterested persons" appointed by the Secretary of State. That same year a fire in the patent office destroyed many of the previously issued patents.

[39] While the first patents may have been issued with an eye toward the legislative discretion used in chartered monopoly rights, patents were soon seen as themselves a "right" owed to any new invention. This right was due regardless of government judgment on criteria besides novelty and non-obviousness (e.g., utility) (Bracha, 2004).

The requirement of novelty was backed by the declaration in the 1790 Patent Act that the Patentee was supposed to be “the first and true inventor” anywhere in the world. The enablement requirement specified that a patent application disclose a claimed invention in sufficient detail for the notional person skilled in the art to carry out that claimed invention and, vitally for the use of this data, that this description of the invention be made available to the public immediately upon issue. Thus, patents themselves transmit information about new technological ideas. Though the enforcement of these requirements has fluctuated, legislation requires many of the same things today: novelty, non-obviousness, (nominally) utility, and enablement.

In 1839 the patent office is charged with collecting statistics on agriculture; this continues until the Department of Agriculture is created in 1862. In 1842 design patents become available.

The Act of March 2, 1861 extended the patent grant length to 17 years, while removing the possibility of extension. The fee structure was changed so the total was $35 in two payments, $15 at application and $20 on grant, to any person who was a citizen of a country “not discriminating against the US;” by 1924 this had increased to $40, $20 at application and $20 at issue. A permanent board of appeals consisting of three examiners was created. In August of 1861 the Confederate Patent Office granted its first patent; that office granted a total of 266 patents that present day historians are aware of.

In 1870 the filing requirements are changed so that a model only need be provided if requested, until 1880 models were generally requested. On September 24, 1877 a fire in the patent office destroyed many models. The commissioner was also given permission to print copies of patents for the public. In 1871 Congress discontinued its reports on patents issued that year, but distributed individual patents: “for the first time printed patent specifications became available to the public at a nominal charge. Hitherto, in order to study patents, it had been necessary to consult original drawings and specifications in the Patent Office or have copies made at considerable expense." Continuing this movement toward information distribution, in 1872 the Patent Office starts publishing weekly excerpts from patents and law in the Official Gazette of the United States Patent Office.

In 1887 the United States joins the International Convention for the Protection of Industrial Property that had been formed in Paris in 1883. That same year the question of the first inventor of the telephone draws wide public attention.

On June 10, 1898 a Classification Division is formed to reclassify all patents—thought the first classification had been published many years before in 1830. It consisted of sixteen categories; it was updated and expanded to 22 categories in 1847 and then several times thereafter.

A.3 Intuition from Trade Theory

A.3.1 The Gravity Model and Market Access

There are many ways to think about the concept of market access. One of the simplest ways is to understand it as the sum of all markets that a place sells to, including itself. The gravity framework gives a very simple way of estimating the trade flows between places. The empirical predictions made by this framework are some of the most robust in the literature, and many classes of trade models predict a gravity relationship. This framework uses a few simple variables to predict bilateral trade flows, while remaining agnostic on the reason why this relationship fits the data. The basic gravity relationship describes bilateral trade flows as between two places i and j:

$$trade_{ij} = \frac{y_i y_j}{y_w} (\frac{\tau_{ij}}{P_i P_j})^{1-\sigma} \tag{8}$$

where y_i , y_j and y_w are the incomes of i, j, and the world (total market), τ_{ij} is a bilateral resistance term, P_i and P_j are location specific resistance terms. Note that changing τ_{ij} has the same effect on trade as changing P_i or P_j, so lowering transportation cost is qualitatively similar to opening a county to trade. This is often simplified by noting that y_w is a constant for all pairs of counties and thus not needed in the estimation and taking $\frac{\tau_{ij}}{P_i P_j}$ to be the physical distance between the locations, because of the robustness of the empirical evidence relating distance to trade flows.

The size of a market i is the sum of all goods sold in it, the market in i captured by a trading partner j is the sum of its imports to i. Thus one can think of the market available to firms in location j as:

$$MA_j \approx \sum_i trade_{ij}. \tag{9}$$

If one labels the goods sold by j in j as $trade_{jj}$ this is a full description. Thus, market access refers to the areas with which a given region can effectively trade as adjusted for transportation costs.

The resistance term between areas incorporates anything that makes trade less likely including language barriers and cultural differences. By using closing of the Suez Canal as a natural experiment, Feyrer (2009) shows that about half of the resistance term between countries in the 1970s was due to transportation costs. In the nineteenth century United States, where there were no formal trade barriers (except during the Civil War) and no language barriers, one would expect transportation cost to be the largest factor in this term.

For the purpose of the gravity model, however any reduction in τ_{ij} will have the same effect on total trade. Thus, the steep reduction in transport costs over the 19th century should have had an effect on trade analogous to a similar reduction in trade barriers.

A.3.2 Melitz Model and Investment

Recent work in trade has examined linkages between innovation and market openness. One of the facts motivating the Melitz (2003) model was that more productive firms export, while less productive firms only serve the domestic market. In the Melitz (2003) model the difference in productivity between these two classes of firms is due to selection. For the nineteenth century, it is appropriate to think of patents not as the byproduct of a firm's Research and Development (R&D), but as a product in-and-of itself. Inventors were not employed by firms but rather were free agents who licensed and sold their improvements to others or who acted as entrepreneurs themselves.[40]

An isolated county may not engage in patenting due to the high effective cost of obtaining that patent[41] or because a patent is of limited value if its use is restricted to its own county.[42] In any one county, the pool of potential entrepreneurs looking to build on that patent is limited. Furthermore, if these entrepreneurs are effectively restricted to selling the goods they produce locally due to high transportation costs, Bustos (2011) and Lileeva and Trefler (2010) show that the incentive for these entrepreneurs to invest in productivity upgrading is low.

[40] A number of studies (Pavcnik, 2002; Amiti and Konings, 2004; van Biesebroeck, 2005; Becker and Egger, 2013; Deloecker, 2007; Fernandes, 2007; Foster et al., 2008; Topalova and Khandelwal, 2011), using modern plant- or firm-level data, note that exporting firms often do become more productive after trade liberalization, suggesting that the increased market access motivates firms to invest in innovations that lower their marginal cost of production. Also, Bustos (2011) and Lileeva and Trefler (2010) both consider models that examine the decision of a firm in an open economy to invest in process innovation, while Dhingra (2013) examines the choices firms make with regard to the introduction of a new brand (a new product). In this model, trade liberalization decreases the number of brands a firm offers; the model does not consider the question of new product entry via new firm entry. However, these settings do not reflect the nineteenth century innovation process.

[41] Which includes the cost of understanding what technology is novel.

[42] When the federal patent system came into existence in the US there was a provision for inventors to convert their state patents to federal ones, provided they gave up their state patent. Most inventors did so, and very few inventors subsequently applied for a state patent rather than a federal one. Before the federal system was existent it was common for inventors to apply for multipliable state patents, as they recognized the value of a monopoly in only one state was limited (Hrdy, 2013).

When local transportation arrives in a county, this may increase patenting by decreasing the fixed cost of receiving a patent. The new urban center formed near the rail stop may attract drafts-persons, lawyers, patent agents, or greater credit access in the form of a local bank. This local access may also encourage a greater scope of patenting topics. Increased urbanization might allow for greater familiarity with middle class consumptions goods.[43] The railroad itself might even become a topic of innovation. Alternately, when a county is effectively opened to trade by falling transportation costs, inventors have a greater incentive to patent, either because of increasing opportunities to sell or license the patent elsewhere, or because local entrepreneurs might be more interested in using these innovations. Thus, this model motivates the search for increased patenting activity following the introduction of improved transportation infrastructure.

A.4 Innovation and Market Access: A Simple Framework

I develop here a very simple framework to motivate my empirical analysis. Imagine indexing every idea so that each idea is represented by an integer. Each person's knowledge is a vector of zeros and ones, where an entry of one represents that person knowing about or "having" a particular idea. Further, each idea can be combined with any other idea to form a third distinct idea that is also a member of the idea set (so the set is closed).[44]

When people interact they transfer ideas–a random draw from the set of ideas, I_i, that person i has is given to person j. The probability of I_k being transferred is $1/N_{P_i,t}$ if $I_k \in I_{P_i,t}$ and zero otherwise,[45] where N_{P_i} is the number of ideas that person i has. People interact with probability $p(c)$, a function of the cost of interacting. Cost, $c_{P_i,P_j,t}$ is a constant for any

[43] The mail order business, Montgomery Ward, which primarily operated out of Chicago, was founded in 1872. It shipped, generally using freight that delivered packages to a train station (large packages were not allowed to be shipped via US Post until 1913), rural residents a large range of manufactured goods (e.g., bolts of cloth, trunks, and pens).

[44] This is equivalent to the way the integers recombine to form the rationals.

[45] I am assuming that only one idea will be transfered per interaction.

two people at a point in time and is also increasing with the distance between them.

When an idea is transferred from P_i to P_j, P_j has a probability, q, of randomly choosing an idea they currently hold and combining it with the new idea that they just received. If this happens P_j now knows the combination idea. Some fraction of combination ideas will be "novel"—new additions to the stock of ideas and also may be commercially viable.

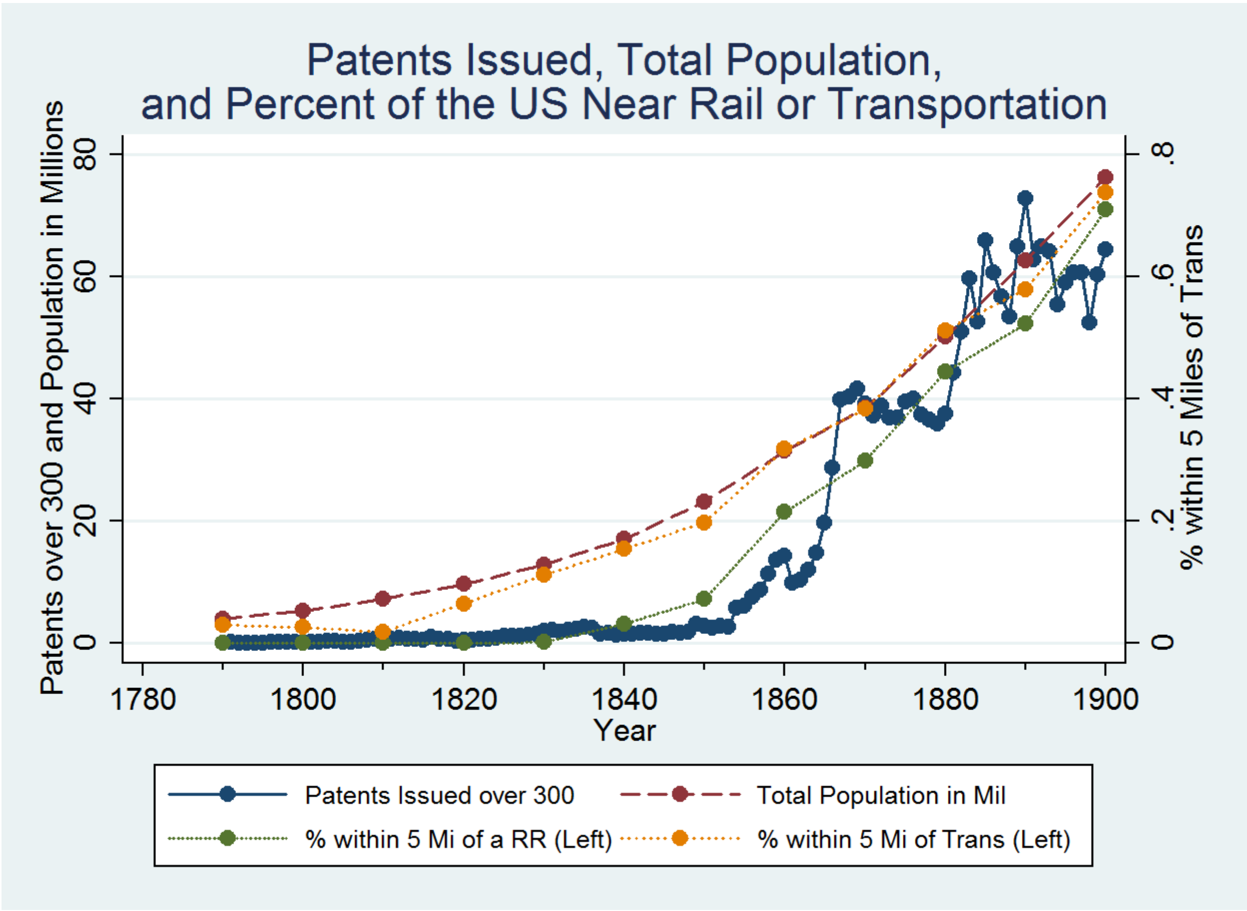
This framework is set up so the rate of new ideas will change if c decreases, that is $\sum_j c_{P_i,P_j,t}$ is a market access, or an access to new ideas. This framework talks only about idea creation and not expression or commercialization of the idea. Yet, it also makes clear that the cost of patenting itself, of knowing how to navigate the system as well as the official cost, are very important to the decision to patent. It is the choice to pay this fixed cost that the Bustos (2011) and Lileeva and Trefler (2010) extensions to the Melitz (2003) model speak to. Thus, lowering the cost of travel might have two important effects on people's propensity to patent, both making it more worthwhile to pay the fixed cost, and by raising the rate of arrival of new ideas by lowering the cost of interacting with others who do not live near by. Increased urbanization, which Atack et al. (2011) showed that railroads encouraged, also will lower the cost of interacting and may lower the costs of participating in the patent system.

B Figures and Tables

B.1 Figures

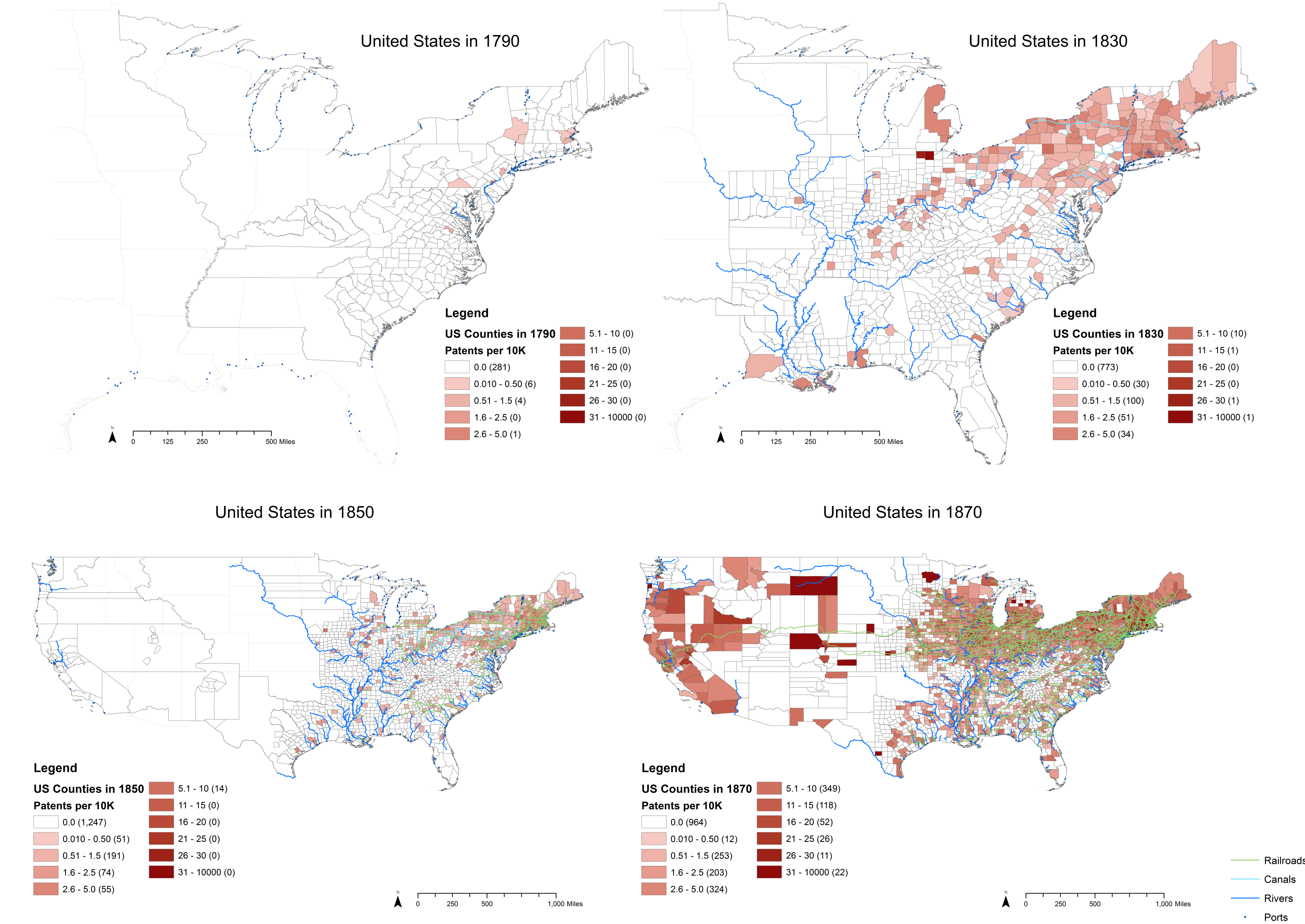
B.1.1 Introduction

Figure 1: Number of Patents Issued Each Year and Total Population over Time



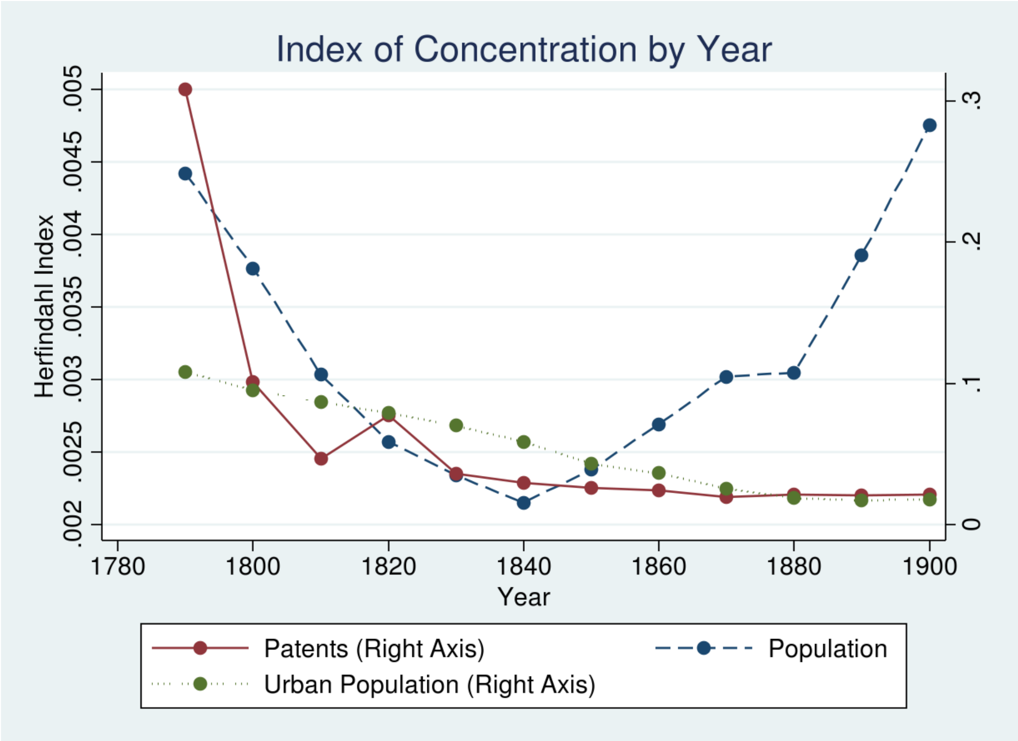
Back to page 11.

Figure 2: Patents per 10,000 People with Transportation

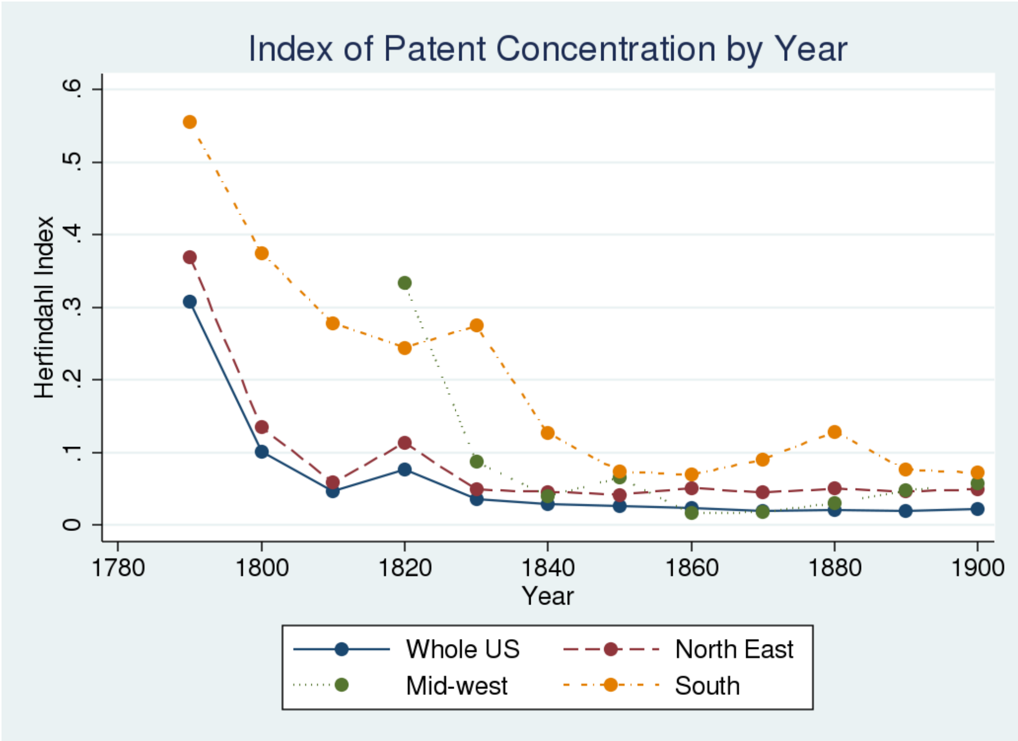


Back to page 11.

Figure 3: Concentration of Patents and Population in Counties by Year, Herfindahl Index



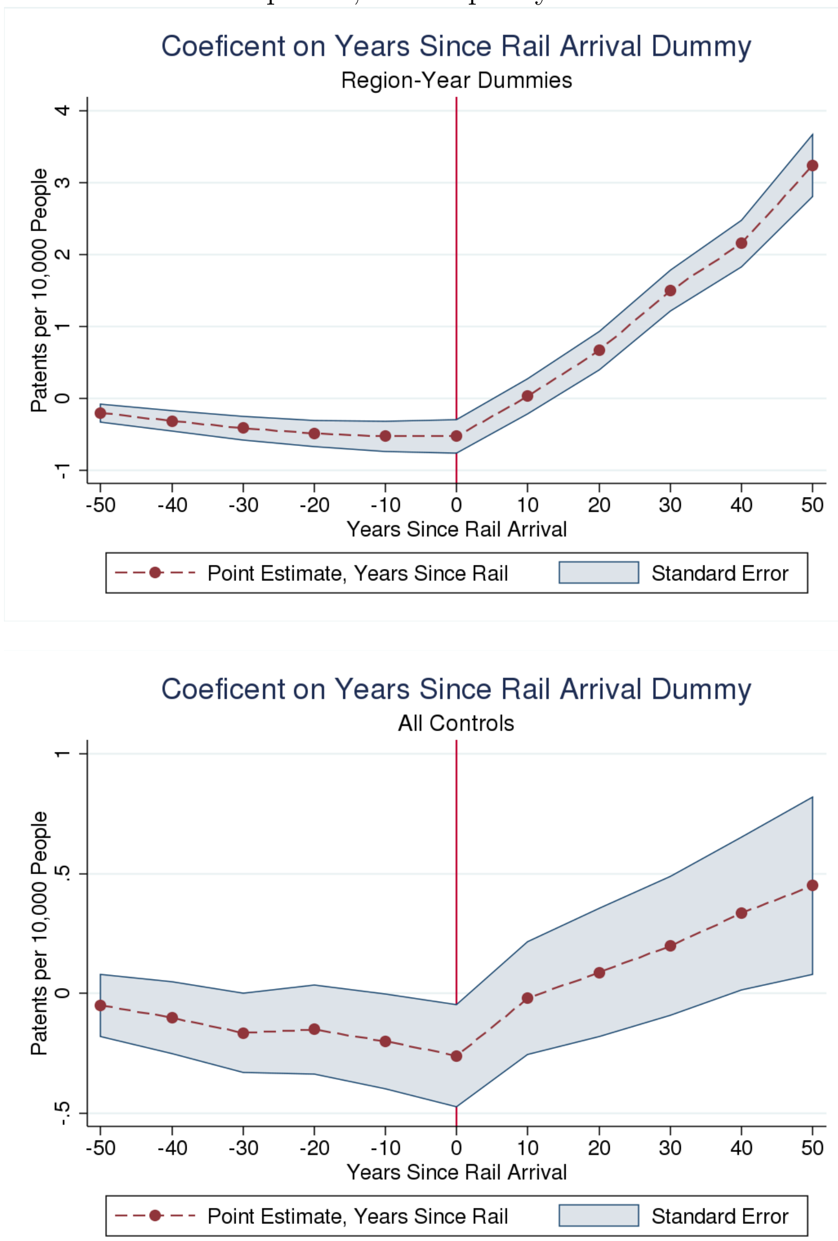
Concentration of Patents by Region in Counties by Year, Herfindahl Index



Back to page 11.

B.1.2 Local Transportation Access

Figure 4: The Mean Patents per 10,000 People by the Years since Railroad Arrival

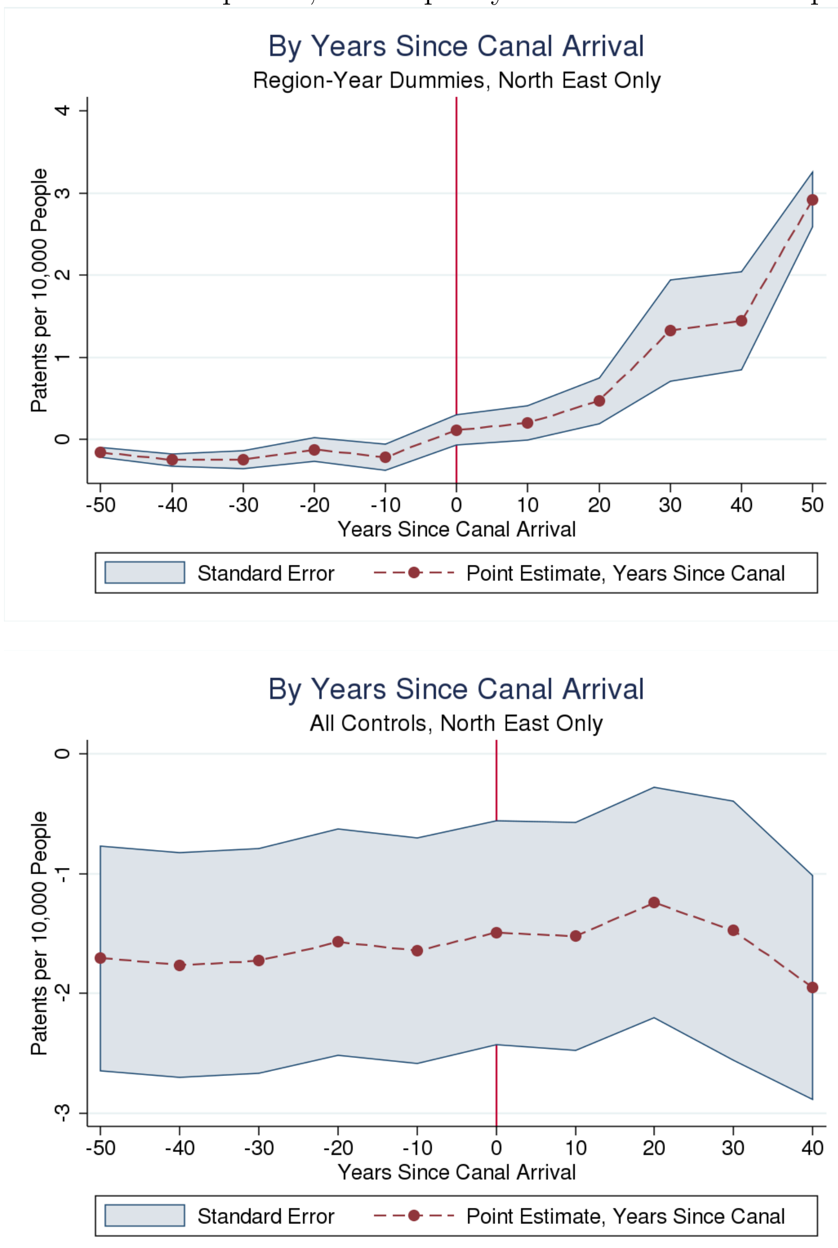


The point estimates and standard errors come from a regression of patents per capita on the dummy variables for the number of years to the arrival of a railroad canal in a county and year and county fixed effects; standard errors are clustered at the county level.

Back to page 12.

B.1.3 Local Transportation Access

Figure 5: The Mean Patents per 10,000 People by the Years since Transportation Arrival

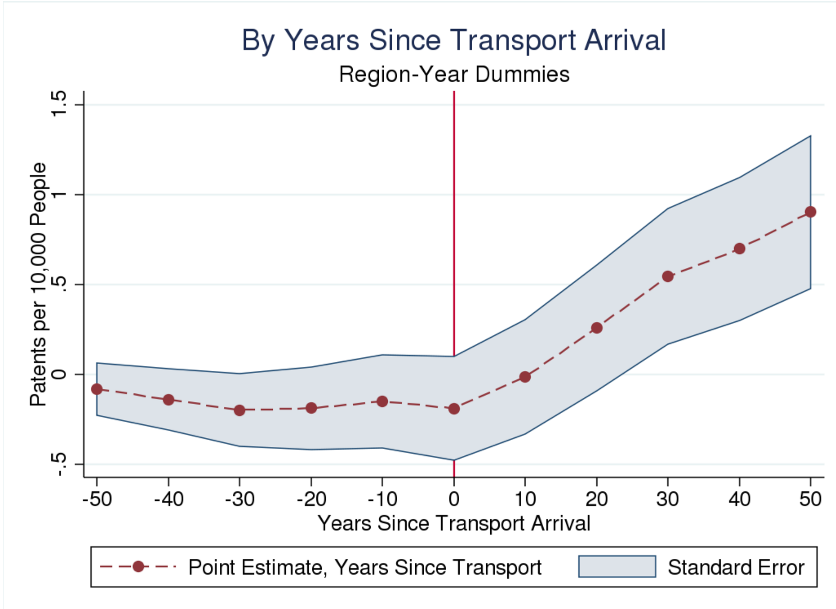


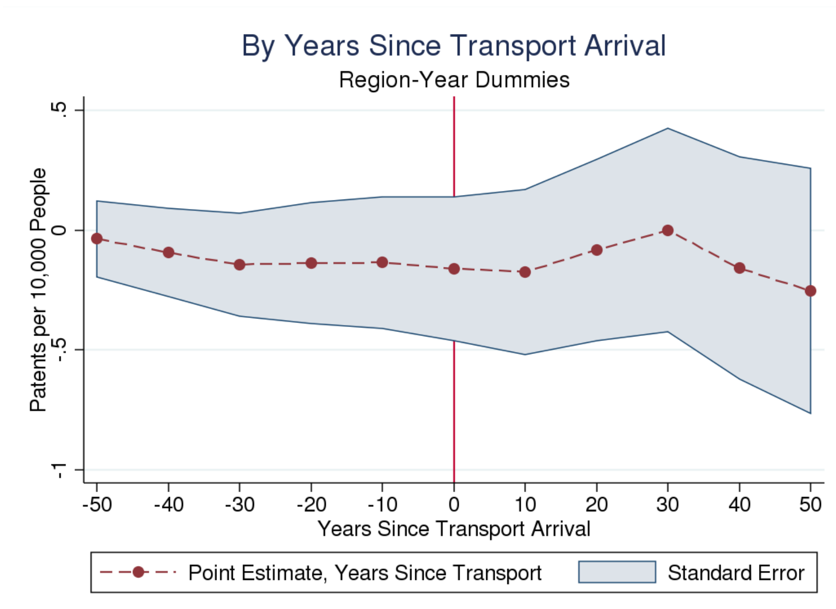
The point estimates and standard errors come from a regression of patents per capita on the dummy variables for the number of years to the arrival of a railroad canal in a county and year and county fixed effects; standard errors are clustered at the county level.

Back to page 13.

B.1.4 Local Transportation Access

Figure 6: The Mean Patents per 10,000 People by the Years since Transportation Arrival



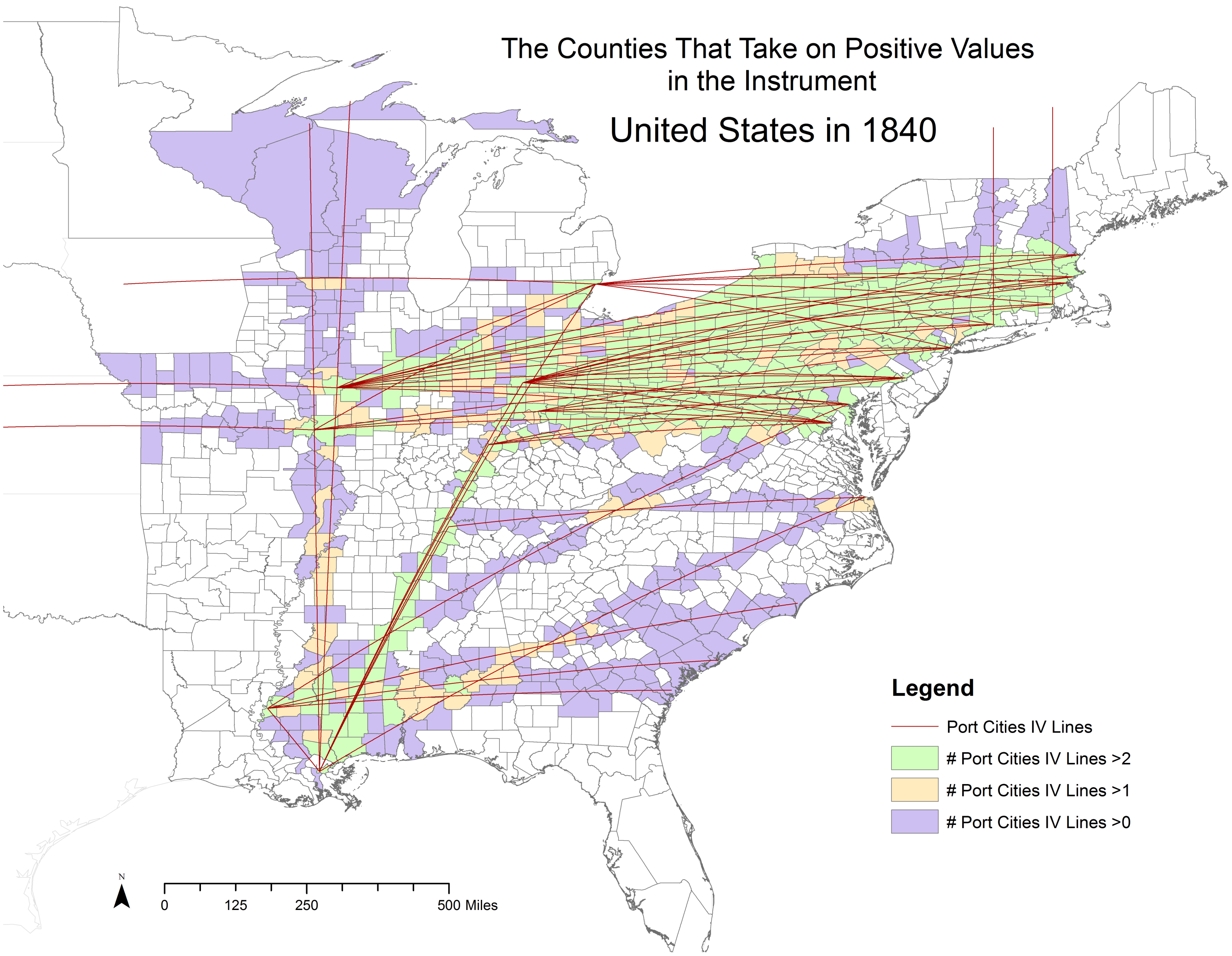


The point estimates and standard errors come from a regression of patents per capita on the dummy variables for the number of years to the arrival of a railroad canal in a county and year and county fixed effects; standard errors are clustered at the county level.

Back to page 13.

B.1.5 Instrument

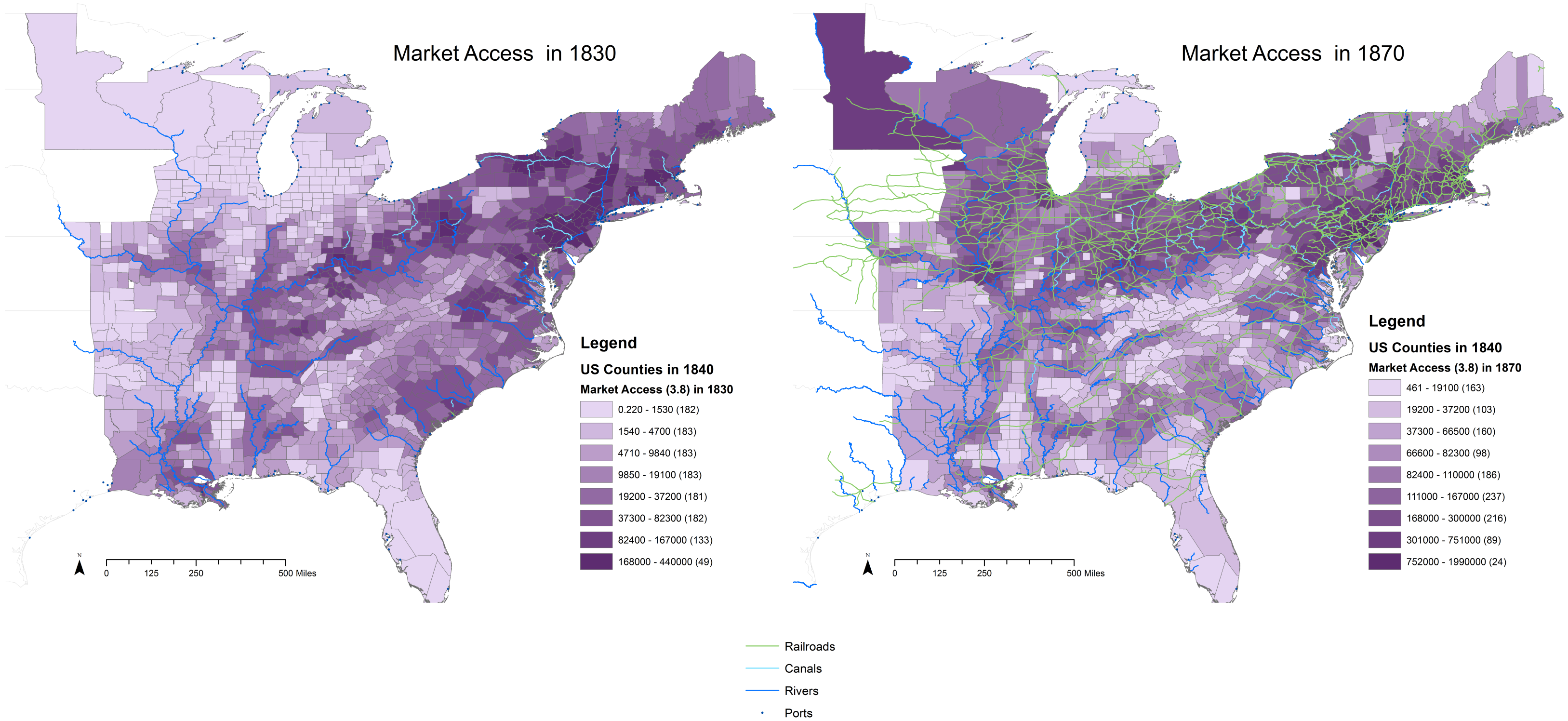
Figure 7: Counties Affected by the Port City Driven Instrument



Back to page 16.

B.1.6 Market Access

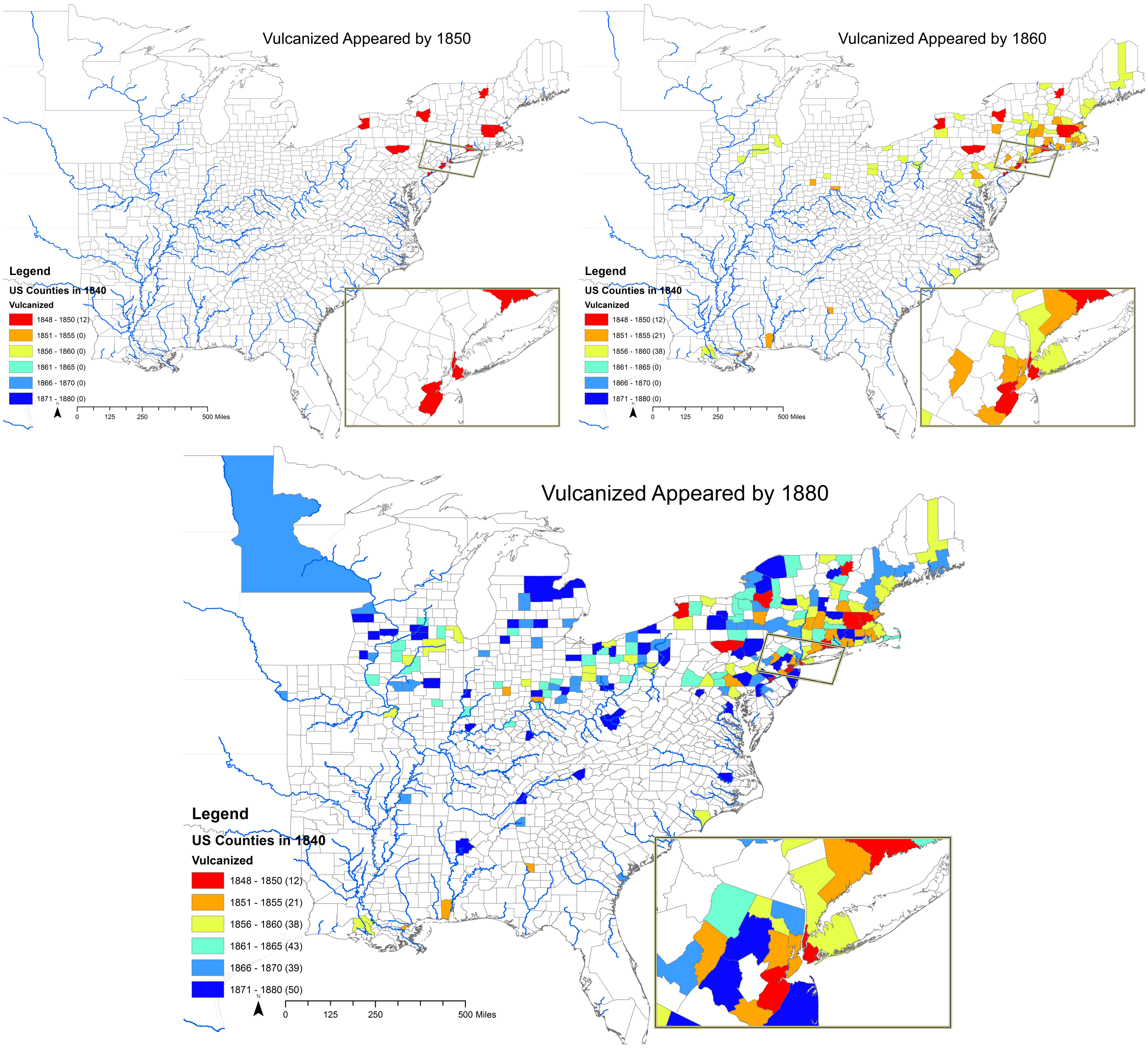
Figure 8: Computed Market Access



Back to page 20.

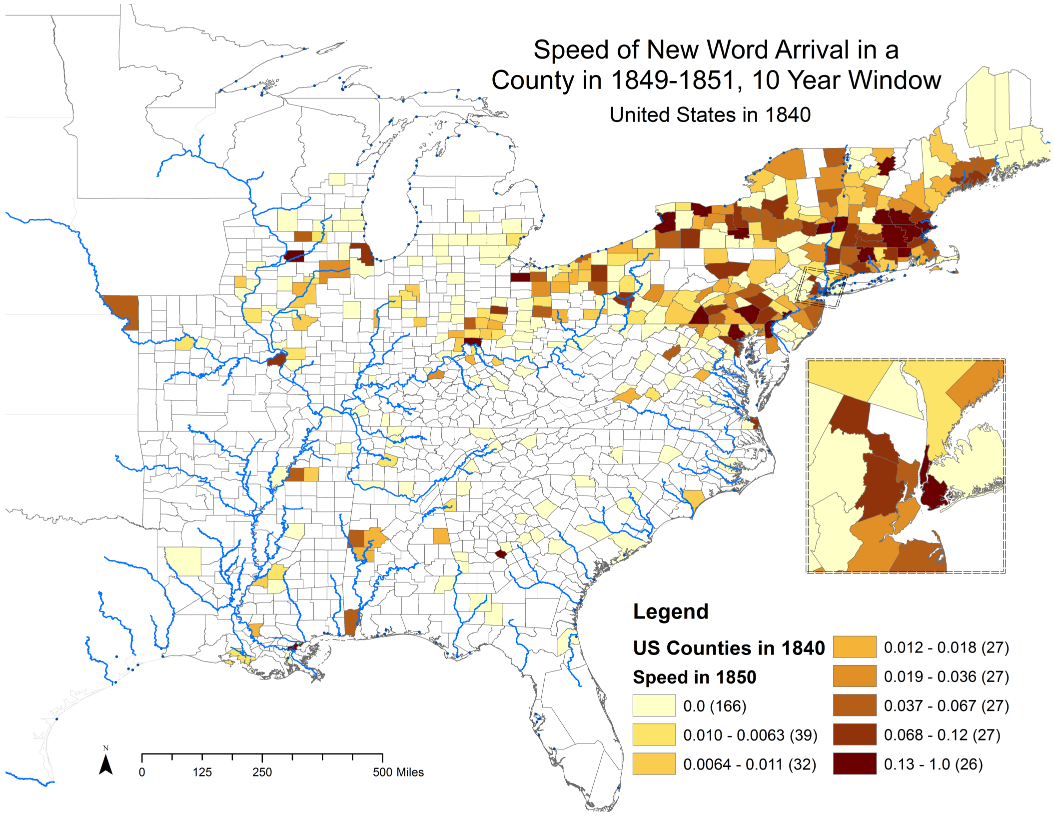
B.1.7 Speed of Word Arrival

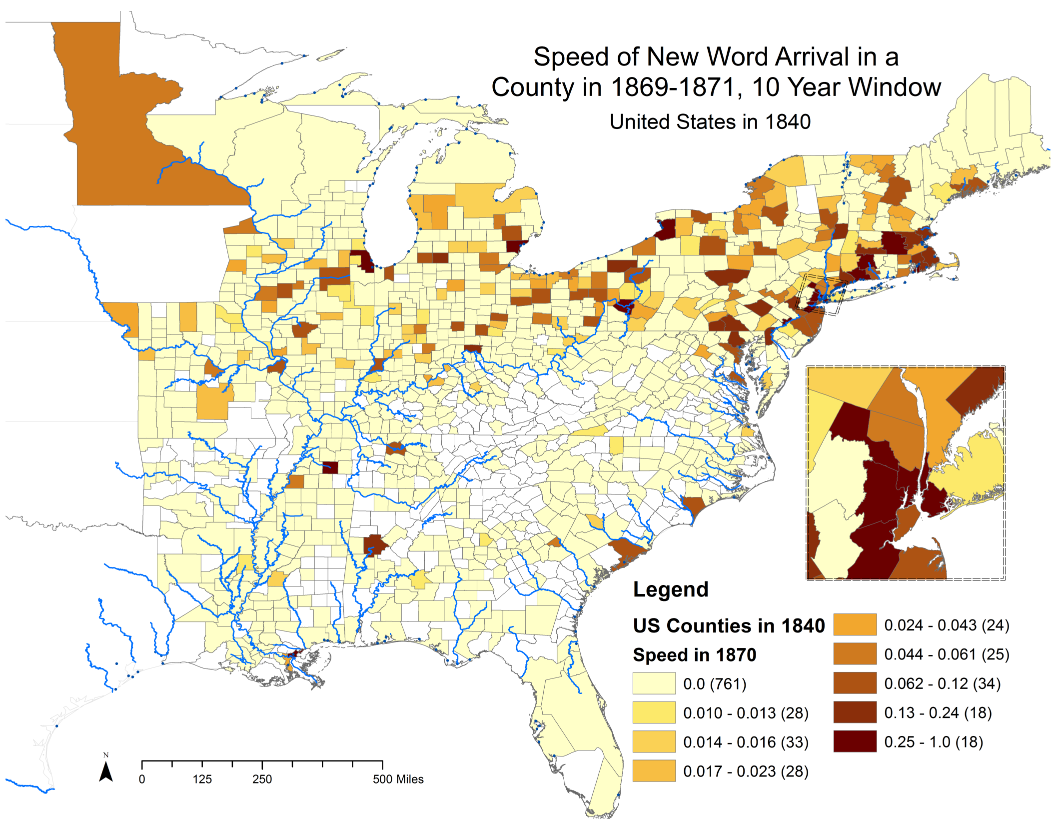
Figure 9: Counties in Which the Word Vulcanized Appeared



Back to page 24.

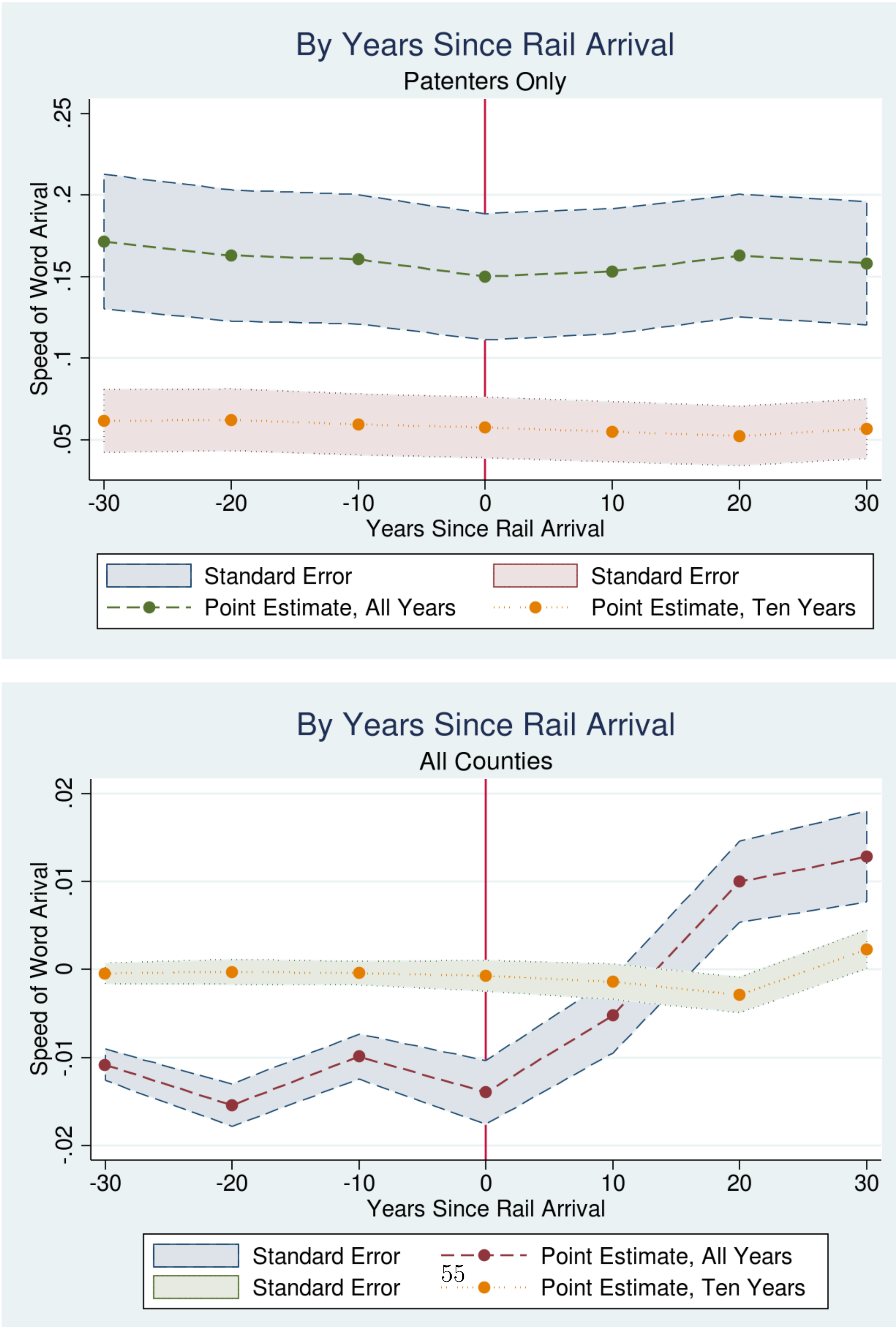
Figure 10: The Speed of New Word Arrival in a County's Patents





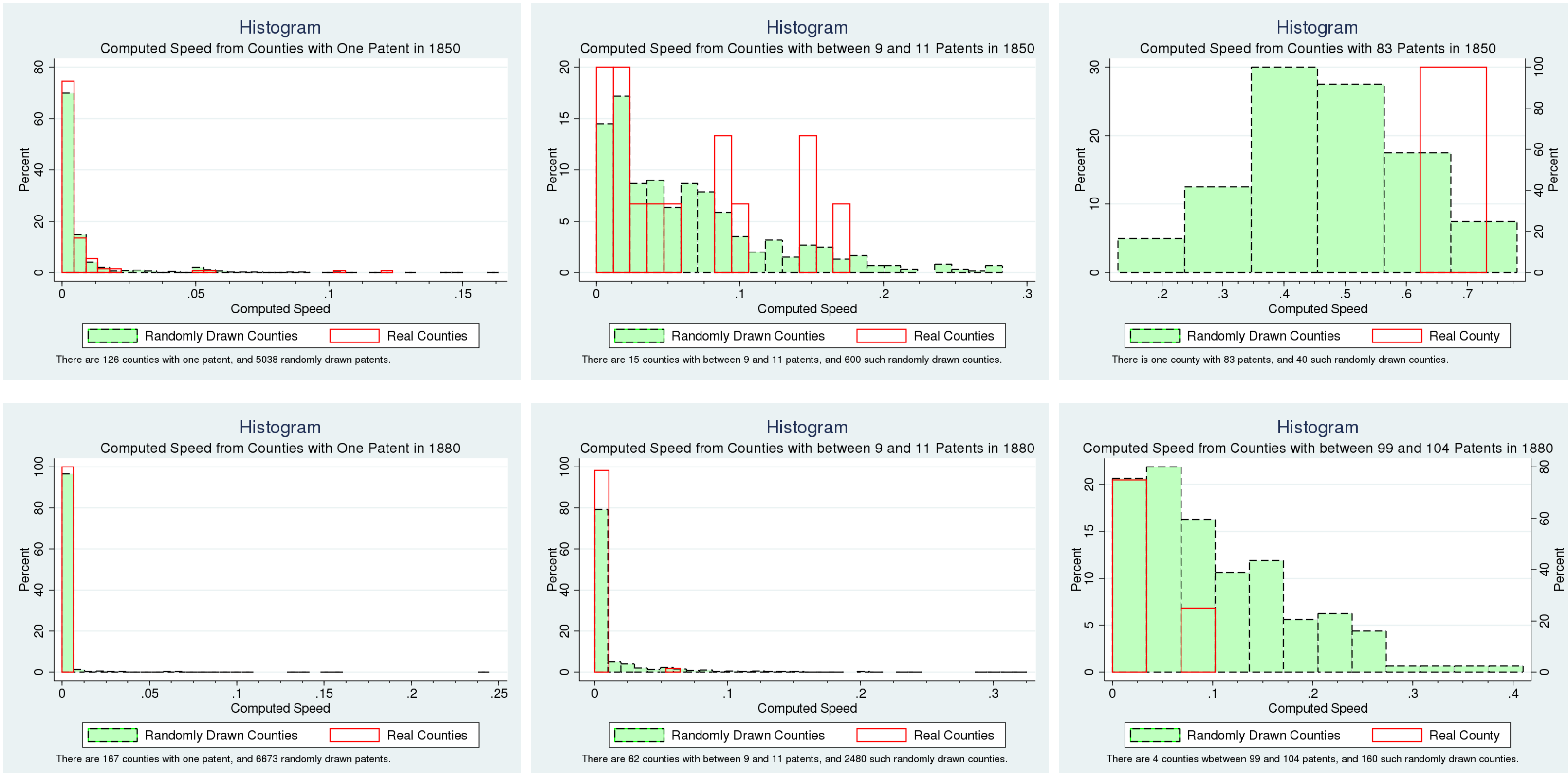
Back to page 25.

Figure 11: The Mean Speed Measured Using Words New in the Last Ten Years or in All Years by the Years to Railroad Arrival



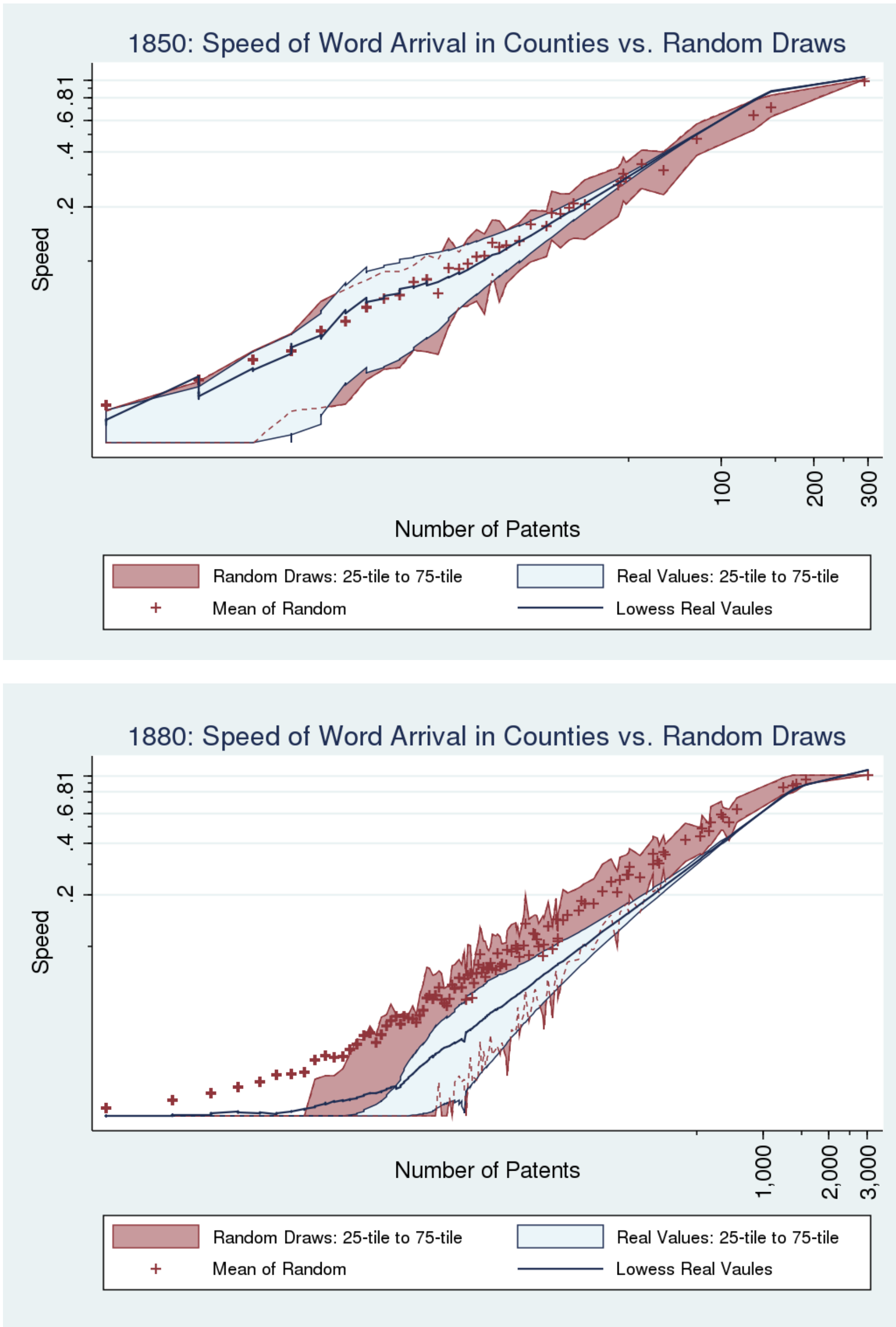
Back to page 27.

Figure 12: Histogram



Back to page 29.

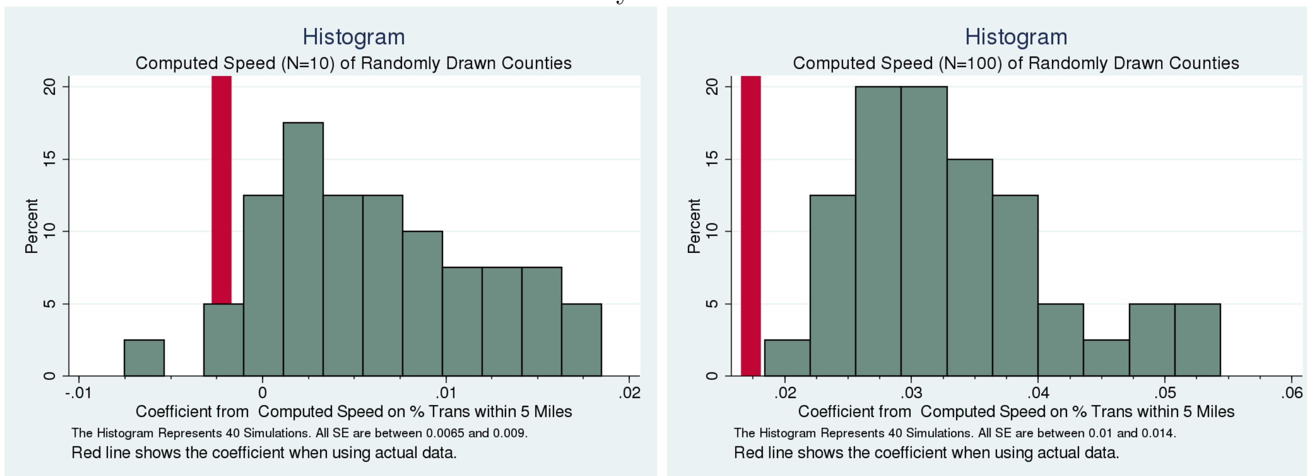
Figure 13: Real vs. Random Draws



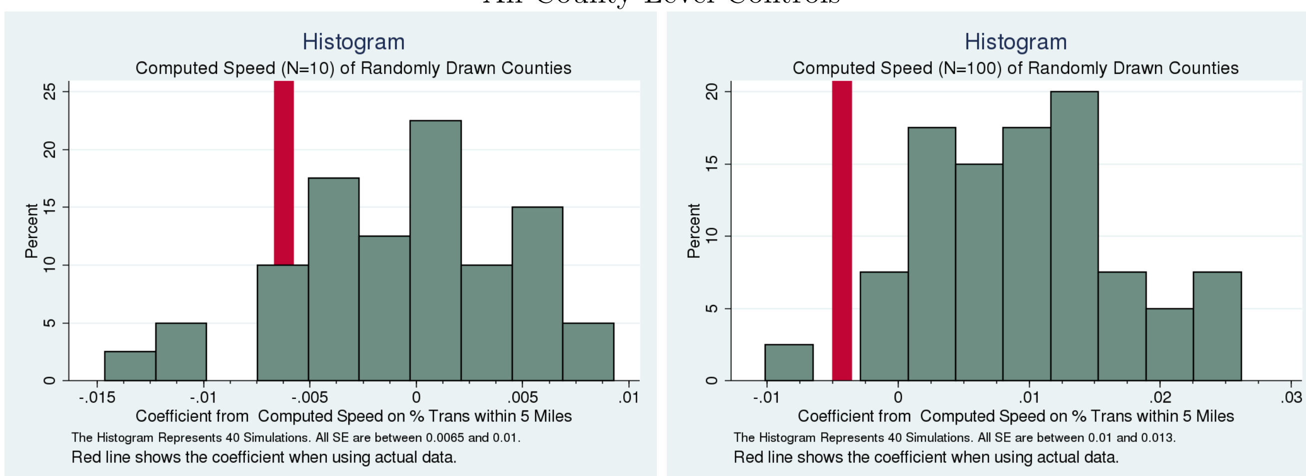
Back to page 29.

Figure 14: Coefficients from the Speed of Word Arrival Computed with Counterfactual Data

No County Level Controls



All County Level Controls



See Table 10 for more detail on the estimation of the real coefficients.

Back to page 30.

B.2 Tables

Table 1: Means by Year, 1790-1900

	1790	1800	1810	1820	1830	1840	1850	1860	1870	1880	1890	1900
Total Population	7,054	7,719	7,274	8,710	10,585	13,363	17,807	23,409	27,733	34,788	41,701	49,693
	(8,471)	(9,694)	(10,337)	(11,837)	(14,356)	(17,661)	(25,405)	(36,546)	(46,763)	(59,165)	(82,606)	(111,868)
Number of Patents	0.0250	0.0227	0.351	0.232	0.962	1.003	1.978	8.642	26.44	25.49	42.12	36.06
	(0.497)	(0.257)	(2.689)	(2.291)	(6.453)	(6.049)	(11.20)	(47.26)	(128.1)	(129.6)	(210.6)	(183.5)
Patent per 10K	0.0397	0.0346	0.144	0.0780	0.314	0.294	0.440	1.851	4.455	3.420	4.640	3.602
	(0.591)	(0.530)	(0.621)	(0.485)	(1.383)	(0.810)	(1.009)	(4.254)	(6.309)	(5.052)	(6.237)	(4.329)
# of NBER Subcategories						2.668	3.255	4.389	6.834	6.220	7.271	6.265
						(2.607)	(3.550)	(4.308)	(5.742)	(5.595)	(6.399)	(5.934)
% Urban, 2500+	0.0141	0.0150	0.0153	0.0149	0.0194	0.0264	0.0428	0.0642	0.0921	0.113	0.152	0.179
	(0.0853)	(0.0910)	(0.0941)	(0.0873)	(0.0980)	(0.112)	(0.136)	(0.158)	(0.178)	(0.192)	(0.215)	(0.226)
% Urban, 25K+	0.00286	0.00308	0.00327	0.00416	0.00390	0.00519	0.0108	0.0150	0.0216	0.0309	0.0414	0.0528
	(0.0481)	(0.0473)	(0.0517)	(0.0567)	(0.0561)	(0.0643)	(0.0875)	(0.101)	(0.117)	(0.139)	(0.160)	(0.179)
% within 1.5 miles of transport	0.00306	0.00360	0.00524	0.0241	0.0436	0.0725	0.0931	0.154	0.189	0.256	0.287	0.366
	(0.0250)	(0.0275)	(0.0373)	(0.0705)	(0.0797)	(0.101)	(0.111)	(0.132)	(0.143)	(0.170)	(0.171)	(0.167)
% within 5 miles of transport	0.0169	0.0189	0.0236	0.0788	0.139	0.214	0.269	0.422	0.495	0.610	0.667	0.786
	(0.0920)	(0.0958)	(0.105)	(0.176)	(0.216)	(0.248)	(0.265)	(0.285)	(0.284)	(0.288)	(0.268)	(0.207)
% within 15 miles of transport	0.0552	0.0626	0.0780	0.206	0.349	0.501	0.599	0.795	0.855	0.912	0.946	0.987
	(0.193)	(0.203)	(0.225)	(0.344)	(0.392)	(0.406)	(0.392)	(0.315)	(0.260)	(0.207)	(0.151)	(0.0678)
% Manufacturing				0.0233		0.0299	0.0193	0.0199	0.0287	0.0271	0.0370	0.0405
				(0.0223)		(0.0332)	(0.0303)	(0.0312)	(0.0366)	(0.0404)	(0.0477)	(0.0455)
% Acres Improved							0.399	0.443	0.474	0.528	0.572	0.582
							(0.174)	(0.197)	(0.208)	(0.217)	(0.209)	(0.200)
% Literate						0.880	0.868		0.887			
						(0.118)	(0.109)		(0.0907)			
% Pop in School						0.0749	0.156		0.147			
						(0.0866)	(0.0835)		(0.0842)			
% Born Out of State							0.226		0.189	0.160		
							(0.191)		(0.148)	(0.123)		
% Foreign Born							0.0522	0.0670	0.0661	0.0596	0.0599	0.0525
							(0.0869)	(0.0947)	(0.0907)	(0.0831)	(0.0859)	(0.0766)

Back to page 11.

B.2.1 Local Transportation Access

Table 2: The Effect of Local Transportation Access on Patents per-Capita, 1790-1900

VARIABLES	(1) Patents per 10K People	(2) Patents per 10K People	(3) Patents per 10K People
% within 1.5 miles of transport	3.366***		
	(0.468)		
% within 5 miles of transport		0.946***	
		(0.152)	
% within 15 miles of transport			0.179*
			(0.0984)
Years	1790-1900	1790-1900	1790-1900
Included County Controls	None	None	None
Region by Year FE	Yes	Yes	Yes
Counties	1249	1249	1249
Observations	13,237	13,237	13,237
R-squared	0.707	0.689	0.667

All specifications control for county dummies, year dummies, and pre-trends.
Robust standard errors in parentheses, standard errors clustered by county.
*** p<0.01, ** p<0.05, * p<0.1

Sources: Patent data as described in the text, U.S. Census Data is from Haines (2010) (county boundaries harmonized to 1840 as in Hornbeck (2010)), transportation data from Atack (2013).

Back to page 14.

Table 3: The Effect of Local Transportation Access on Patents per-Capita with Controls, 1790-1900

1790-1900

VARIABLES	(1) Patents per 10K People	(2) Patents per 10K People	(3) Patents per 10K People	(4) Patents per 10K People
% within 5 miles of transport	0.946***	0.680***	0.429**	0.393***
	(0.152)	(0.138)	(0.168)	(0.151)
Included County Controls	None	People	Economic Activity	All
Region by Year FE	Yes	Yes	Yes	Yes
R-squared	0.689	0.772	0.732	0.786

VARIABLES	(1) Patents per 10K People	(2) Patents per 10K People	(3) Patents per 10K People	(4) Patents per 10K People	(5) Patents per 10K People	(6) Patents per 10K People
% within 5 miles of transport	0.748***	0.842***	0.648***	0.833***	0.817***	0.935***
	(0.149)	(0.137)	(0.164)	(0.147)	(0.164)	(0.153)
County Controls	Manufacturing	Urban	Improved Acres	Migration	Literacy	Schooling
Region by Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Counties	1249	1249	1249	1249	1249	1249
Observations	13,237	13,237	13,237	13,237	13,237	13,237
R-squared	0.726	0.761	0.695	0.715	0.692	0.691

Robust standard errors in parentheses, standard errors clustered by county.

All specifications control for county dummies, year dummies, and pre-trends.

*** p<0.01, ** p<0.05, * p<0.1

Included controls are:

- Manufacturing (Economic Activity): the percent of the county that is employed in manufacturing
- Urban (People): the percent of the county that is urban (2,500+), metropolitan (25,000+)
- Improved Acres (Economic Activity): the percent of farm land that is improved
- Migration (People): the percent of the county that is born out of state, and foreign born
- Literacy (People): the percent of the county that is literate
- Schooling (People): the percent of the county that is in school

Each variable is include at lagged values, and interacted with time dummies. See Footnote 15

Sources: Patent data as described in the text, U.S. Census Data is from Haines (2010) (county boundaries harmonized to 1840 as in Hornbeck (2010)), transportation data from Atack (2013).

Back to page 14.

Table 4: First Differences Patents per Capita and Local Transportation Access with Interactions

VARIABLES	(1) FD Pat per 10K Ppl	(2) FD Pat per 10K Ppl	(3) FD Pat per 10K Ppl	(4) FD Pat per 10K Ppl
Lag Pat per 10K Ppl	-0.344***		-0.337***	-0.484***
	(0.0551)		(0.0607)	(0.0381)
FD % within 5 miles of transport		0.611***	0.917***	0.591***
		(0.127)	(0.216)	(0.130)
Lag Pat per 10K Ppl × FD % within 5 miles			-0.216	-0.196*
			(0.179)	(0.108)
Marginal Effect of FD % 5 mi			0.562	0.269
			(0.176)	(0.162)
z-stat.			3.184	1.655
Included Controls	None	None	None	All
Counties	1249	1249	1249	1249
Observations	11,954	11,954	11,954	11,954
R-squared	0.383	0.280	0.385	0.543

VARIABLES	(1) Pat per 10K Ppl Growth Rate	(2) Pat per 10K Ppl Growth Rate	(3) Pat per 10K Ppl Growth Rate	(4) Pat per p10K Ppl Growth Rate
Lag Pat per 10K Ppl	-0.0185***		-0.0154***	-0.0263***
	(0.00341)		(0.00306)	(0.00453)
FD % within 5 miles of transport		0.110***	0.202***	0.176***
		(0.0318)	(0.0342)	(0.0347)
Lag Pat per 10K Ppl × FD % within 5 miles			-0.0892***	-0.0811***
			(0.0112)	(0.0128)
Marginal Effect of FD % 5 mi			0.0550	0.0427
			(0.0308)	(0.0312)
z-stat.			1.783	1.368
Included Controls	None	None	None	All
Counties	1249	1249	1249	1249
Observations	11,954	11,954	11,954	11,954
R-squared	0.118	0.103	0.124	0.160

Robust standard errors in parentheses, standard errors clustered by county.
*** p<0.01, ** p<0.05, * p<0.1

Included in specifications with controls are, the percent of the county that is: employed in manufacturing, urban (2,500+), metropolitan (25,000+), literate, in school, born out of state, and foreign born, as well as the percent of farm land that is improved. Each variable is include at lagged values, and interacted with time dummies.

Back to page 15.

Table 5: Instrumental Variables: The Effect of Local Transportation Access on Patents per-Capita, 1840-1870

VARIABLES	(1) OLS Patents per 10K People	(2) First Stage Percent Trans 5.0 Miles	(3) IV Patents per 10K People	(4) OLS Patents per 10K People	(5) First Stage Percent Trans 5.0 Miles	(6) IV Patents per 10K People
Fort Instrument		0.235***			0.209***	
		(0.0224)			(0.0234)	
% within 5 miles of transport	1.377***		1.177	0.521		0.272
	(0.414)		(0.968)	(0.435)		(0.860)
log Total Pop				-0.723	0.0103	-0.722
				(0.823)	(0.00850)	(0.701)
Wald Stat.			30.10			25.40
Years	1840-1870	1840-1870	1840-1870	1840-1870	1840-1870	1840-1870
Included County Controls	None	None	None	All	All	All
Region by Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Counties	1229	1229	1229	1229	1229	1229
Observations	4,991	4,991	4,991	4,991	4,991	4,991
R-squared	0.695	0.884	0.168	0.779	0.890	0.042

All specifications control for county dummies, year dummies, and pre-trends.
Robust standard errors in parentheses, standard errors clustered by county.
*** p<0.01, ** p<0.05, * p<0.1

Included in specifications with controls are, the percent of the county that is: employed in manufacturing, urban (2,500+), metropolitan (25,000+), literate, in school, born out of state, and foreign born, as well as the percent of farm land that is improved. Each variable is include at lagged values, and interacted with time dummies. See Footnote 15

Sources: Patent data as described in the text, U.S. Census Data is from Haines (2010) (county boundaries harmonized to 1840 as in Hornbeck (2010)), transportation data from Atack (2013).

Because the variation in the instrument is in the years 1850 through 1870 this table restricts the sample to 1840-1870. Tables showing the full sample are available upon request or on my website.

Back to page 17.

Table 6: The Effect of Market Access on Patents per Capita with Controls, 1790-1900

VARIABLES	(1) Patents per 10K People	(2) Patents per 10K People	(3) Patents per 10K People	(4) Patents per 10K People
log Market Access	0.137	0.210***	-0.0590	0.0264
	(0.101)	(0.0652)	(0.0797)	(0.0695)
log Total Pop	0.303***	-0.246**	0.305***	-0.0272
	(0.114)	(0.123)	(0.0989)	(0.0895)
Included Controls	None	People	Economic Activity	All
Region by Year FE	Yes	Yes	Yes	Yes
R-squared	0.673	0.773	0.732	0.787

VARIABLES	(1) Patents per 10K People	(2) Patents per 10K People	(3) Patents per 10K People	(4) Patents per 10K People	(5) Patents per 10K People	(6) Patents per 10K People
log Market Access	0.139*	0.226***	-0.115	0.299***	0.0915	0.133
	(0.0799)	(0.0663)	(0.0928)	(0.105)	(0.108)	(0.110)
log Total Pop	0.0807	-0.137	0.598***	-0.229	0.343***	0.326***
	(0.106)	(0.0896)	(0.110)	(0.142)	(0.116)	(0.115)
County Controls	Manufacturing	Urban	Improved Acres	Migration	Literacy	Schooling
Region by Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Counties	1249	1249	1249	1249	1249	1249
Observations	13,237	13,237	13,237	13,237	13,237	13,237
R-squared	0.722	0.761	0.685	0.705	0.676	0.675

Robust standard errors in parentheses, standard errors clustered by county.
All specifications control for county dummies, year dummies, and pre-trends.
*** p<0.01, ** p<0.05, * p<0.1

Included controls are:

- Manufacturing (Economic Activity): the percent of the county that is employed in manufacturing
- Urban (People): the percent of the county that is urban (2,500+), metropolitan (25,000+)
- Improved Acres (Economic Activity): the percent of farm land that is improved
- Migration (People): the percent of the county that is born out of state, and foreign born
- Literacy (People): the percent of the county that is literate
- Schooling (People): the percent of the county that is in school

Each variable is include at lagged values, and interacted with time dummies. See Footnote 15

Sources: Patent data as described in the text, U.S. Census Data is from Haines (2010) (county boundaries harmonized to 1840 as in Hornbeck (2010)), transportation data from Atack (2013).

Back to page 22.

Table 7: The Effect of Market Access and Local Transportation Access on Patents per Capita, 1790-1900

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Patents per 10K People	Patents per 10K People	Patents per 10K People	Patents per 10K People	Patents per 10K People	Patents per 10K People
log Market Access		0.137	-0.0391		0.0264	-0.0118
		(0.101)	(0.0804)		(0.0695)	(0.0636)
% within 5 miles	0.865***		0.746***	0.391***		0.389**
of transportation	(0.149)		(0.164)	(0.143)		(0.151)
log Total Pop	0.354***	0.303***	0.349***	0.00957	-0.0272	-0.00174
	(0.0751)	(0.114)	(0.118)	(0.0800)	(0.0895)	(0.0917)
Years	1790-1900	1790-1900	1790-1900	1790-1900	1790-1900	1790-1900
Included County Controls	None	None	None	All	All	All
Region by Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Counties	1249	1249	1249	1249	1249	1249
Observations	13,237	13,237	13,237	13,237	13,237	13,237
R-squared	0.692	0.673	0.699	0.786	0.787	0.789

All specifications control for county dummies, year dummies, and pre-trends.
Robust standard errors in parentheses, standard errors clustered by county.
*** p<0.01, ** p<0.05, * p<0.1

Included in specifications with controls are, the percent of the county that is: employed in manufacturing, urban (2,500+), metropolitan (25,000+), literate, in school, born out of state, and foreign born, as well as the percent of farm land that is improved. Each variable is include at lagged values, and interacted with time dummies. See Footnote 15

Sources: Patent data as described in the text, U.S. Census Data is from Haines (2010) (county boundaries harmonized to 1840 as in Hornbeck (2010)), transportation data from Atack (2013).

Back to page 22.

Instrumented market access available upon request or on my website.

Table 8: First Differences Patents per Capita and Market Access with Interactions

VARIABLES	(1) FD Pat per 10K Ppl	(2) FD Pat per 10K Ppl	(3) FD Pat per 10K Ppl	(4) FD Pat per 10K Ppl
Lag Pat per 10K Ppl	-0.314***		-0.291***	-0.460***
	(0.0562)		(0.0658)	(0.0382)
FD log Market Access		-0.0191	0.136	0.0375
		(0.0805)	(0.0880)	(0.0613)
Lag Pat per 10K Ppl ×			-0.0810	-0.0968***
FD log Market Access			(0.0702)	(0.0276)
Marginal Effect of FD			0.00292	-0.122
			(0.0834)	(0.0629)
z-stat.			0.0350	-1.934
Included Controls	None	None	None	All
Counties	1249	1249	1249	1249
Observations	11,954	11,954	11,954	11,954
R-squared	0.367	0.274	0.369	0.544

VARIABLES	(1) Pat per 10K Ppl Growth Rate	(2) Pat per 10K Ppl Growth Rate	(3) Pat per 10K Ppl Growth Rate	(4) Pat per p10K Ppl Growth Rate
Lag Pat per 10K Ppl	-0.0170***		-0.0141***	-0.0265***
	(0.00328)		(0.00328)	(0.00472)
FD log Market Access		0.0287***	0.0418***	0.0297***
		(0.00789)	(0.00802)	(0.00858)
Lag Pat per 10K Ppl ×			-0.0113***	-0.0106*
FD log Market Access			(0.00415)	(0.00602)
Marginal Effect of FD			0.0233	0.0122
			(0.00877)	(0.0108)
z-stat			2.654	1.133
Included Controls	None	None	None	All
Counties	1249	1249	1249	1249
Observations	11,954	11,954	11,954	11,954
R-squared	0.124	0.110	0.127	0.161

Robust standard errors in parentheses, standard errors clustered by county.
*** p<0.01, ** p<0.05, * p<0.1

Included in specifications with controls are, the percent of the county that is: employed in manufacturing, urban (2,500+), metropolitan (25,000+), literate, in school, born out of state, and foreign born, as well as the percent of farm land that is improved. Each variable is include at lagged values, and interacted with time dummies.

Back to page 23.

Table 9: The Effect of Market Access on Patents per-Capita with Controls, 1790-1900

VARIABLES	(1) Patents per 10K People	(2) Patents per 10K People	(3) Patents per 10K People	(4) Patents per 10K People
log Market Access	0.137	0.210***	-0.0590	0.0264
	(0.101)	(0.0652)	(0.0797)	(0.0695)
log Total Pop	0.303***	-0.246**	0.305***	-0.0272
	(0.114)	(0.123)	(0.0989)	(0.0895)
Included Controls	None	People	Economic Activity	All
Region by Year FE	Yes	Yes	Yes	Yes
R-squared	0.673	0.773	0.732	0.787

VARIABLES	(1) Patents per 10K People	(2) Patents per 10K People	(3) Patents per 10K People	(4) Patents per 10K People	(5) Patents per 10K People	(6) Patents per 10K People
log Market Access	0.139*	0.226***	-0.115	0.299***	0.0915	0.133
	(0.0799)	(0.0663)	(0.0928)	(0.105)	(0.108)	(0.110)
log Total Pop	0.0807	-0.137	0.598***	-0.229	0.343***	0.326***
	(0.106)	(0.0896)	(0.110)	(0.142)	(0.116)	(0.115)
County Controls	Manufacturing	Urban	Improved Acres	Migration	Literacy	Schooling
Region by Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Counties	1249	1249	1249	1249	1249	1249
Observations	13,237	13,237	13,237	13,237	13,237	13,237
R-squared	0.722	0.761	0.685	0.705	0.676	0.675

Robust standard errors in parentheses, standard errors clustered by county.
All specifications control for county dummies, year dummies, and pre-trends.
*** p<0.01, ** p<0.05, * p<0.1

Included controls are:

- Manufacturing (Economic Activity): the percent of the county that is employed in manufacturing
- Urban (People): the percent of the county that is urban (2,500+), metropolitan (25,000+)
- Improved Acres (Economic Activity): the percent of farm land that is improved
- Migration (People): the percent of the county that is born out of state, and foreign born
- Literacy (People): the percent of the county that is literate
- Schooling (People): the percent of the county that is in school

Each variable is include at lagged values, and interacted with time dummies. See Footnote 15

Sources: Patent data as described in the text, U.S. Census Data is from Haines (2010) (county boundaries harmonized to 1840 as in Hornbeck (2010)), transportation data from Atack (2013).

Back to page 22.

B.2.2 Speed of Word Arrival

Table 10: The Effect of Local Transportation Access on the Speed of Word Arrival, 1850-1890

VARIABLES	(1) Speed 10 Years	(2) Speed All Years	(3) Speed 10 Years	(4) Speed All Years
	Local Transportation Access			
% within 5 miles of transport	-0.0106***	-0.0134***	-0.00419	-0.00917**
	(0.00416)	(0.00421)	(0.00409)	(0.00453)
log Count Pat	0.00671***	0.0320***	0.00675***	0.0287***
	(0.000942)	(0.00164)	(0.000888)	(0.00151)
R-squared	0.764	0.777	0.821	0.830
	Market Access			
log Market Access	0.000426	-0.000593	0.00240***	0.000819
	(0.000972)	(0.00161)	(0.000918)	(0.00146)
log Total Pop	0.0174***	0.0263***	0.00476	0.0175***
	(0.00549)	(0.00535)	(0.00397)	(0.00450)
log Count Pat	0.00628***	0.0318***	0.00632***	0.0282***
	(0.000887)	(0.00158)	(0.000881)	(0.00147)
R-squared	0.753	0.772	0.816	0.828
Years	1850-1890	1850-1890	1850-1890	1850-1890
County Controls	No	No	Yes	Yes
Counties	1250	1250	1250	1250
Observations	6,245	6,245	6,245	6,245

Robust standard errors in parentheses,
standard errors clustered by county.
All specifications control for county dummies, year dummies,
and pre-trends.
*** p<0.01, ** p<0.05, * p<0.1

Included in specifications with controls are, the percent of the county that is: employed in manufacturing, urban (2,500+), metropolitan (25,000+), literate, in school, born out of state, and foreign born, as well as the percent of farm land that is improved. Each variable is include at lagged values, and interacted with time dummies. See Footnote 15

Sources: Patent data as described in the text, U.S. Census Data is from Haines (2010) (county boundaries harmonized to 1840 as in Hornbeck (2010)), transportation data from Atack (2013).

Back to page 27.

Table 11: The Effect of Market Access and Local Transportation Access on the Speed of Word Arrival, 1850-1890

VARIABLES	(1) Speed 10 Years	(2) Speed All Years	(3) Speed 10 Years	(4) Speed All Years
log Market Access	0.00105	0.00267	0.00200	0.00267
	(0.00154)	(0.00178)	(0.00129)	(0.00175)
% within 5 miles	-0.0119***	-0.0170***	-0.00740**	-0.0147***
of transport	(0.00416)	(0.00554)	(0.00340)	(0.00547)
log Total Pop	0.0147***	0.0252***	0.00453	0.0162***
	(0.00523)	(0.00516)	(0.00416)	(0.00451)
log Count Pat	0.00640***	0.0311***	0.00653***	0.0283***
	(0.000891)	(0.00158)	(0.000884)	(0.00147)
Years	1850-1890	1850-1890	1850-1890	1850-1890
County Controls	No	No	Yes	Yes
Counties	1250	1250	1250	1250
Observations	6,245	6,245	6,245	6,245
R-squared	0.766	0.783	0.824	0.836

Robust standard errors in parentheses,
standard errors clustered by county.
All specifications control for county dummies, year dummies,
and pre-trends.
*** p<0.01, ** p<0.05, * p<0.1

Included in specifications with controls are, the percent of the county that is: employed in manufacturing, urban (2,500+), metropolitan (25,000+), literate, in school, born out of state, and foreign born, as well as the percent of farm land that is improved. Each variable is include at lagged values, and interacted with time dummies. See Footnote 15

Sources: Patent data as described in the text, U.S. Census Data is from Haines (2010) (county boundaries harmonized to 1840 as in Hornbeck (2010)), transportation data from Atack (2013).

Back to page 27.

Table 12: P-Values from the Kolmogorov-Smirnov Test for Equivalence of Distribution

Group	1850	1860	1870	1880	1890
One Patent Counties	0.865	0.321	0.595	0.031	0.196
Ten Patent Counties	0.795	0.005	0.003	0.000	0.000
Below the 25th percentile	0.865	0.321	0.597	0.042	0.000
Above the 75th percentile	0.066	0.000	0.000	0.000	0.000
Overall	0.395	0.002	0.000	0.000	0.000

All tests report that the real distribution is slower than the simulated one.
The null is that the two distributions are the same.

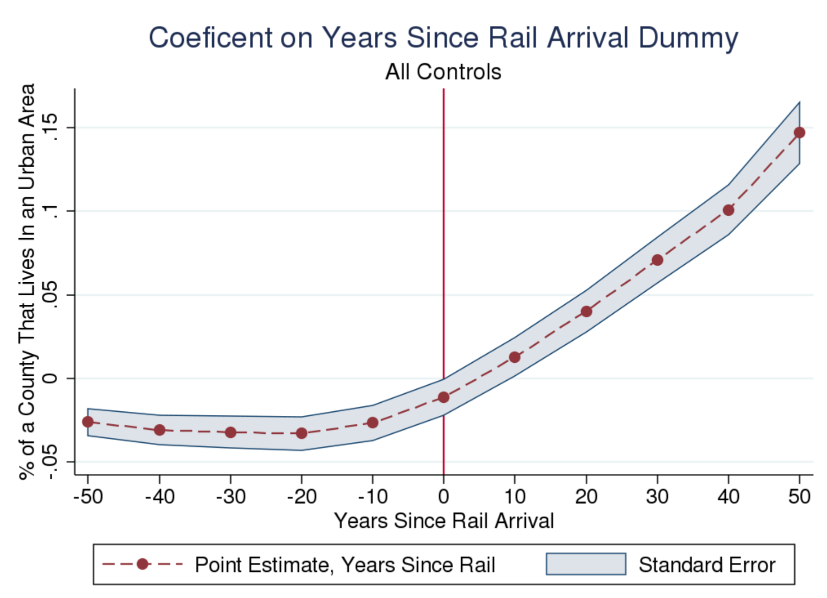
Sources: Patent data from Westlaw and Google.

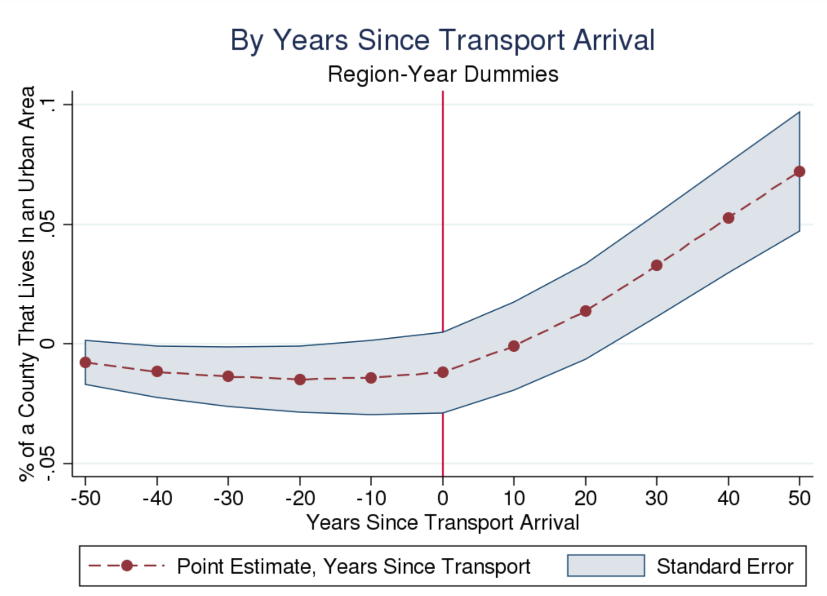
Back to page 29.

C Appendix Figures and Tables

C.1 Appendix Figures

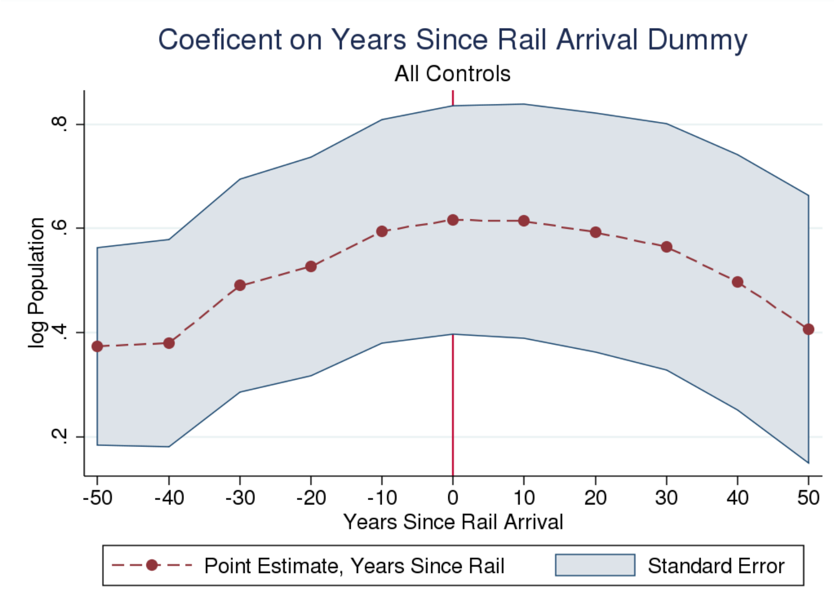
Figure 15: Years since Railroad Arrival





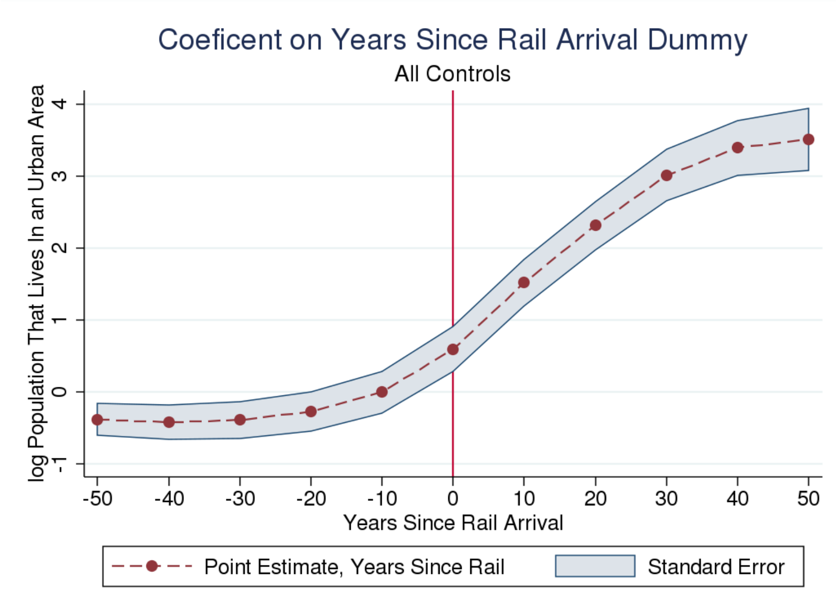
Back to page 36.

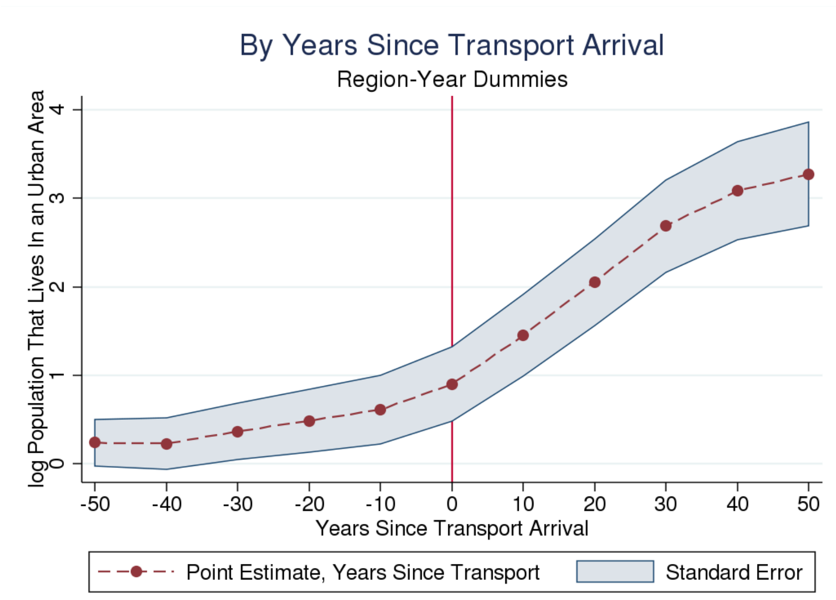
Figure 16: Years since Railroad Arrival



Back to page 37.

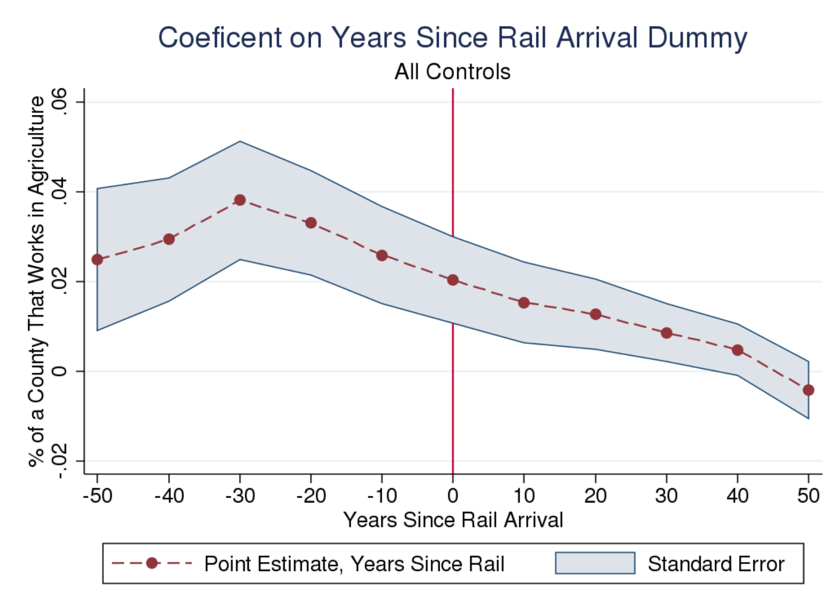
Figure 17: Years since Railroad Arrival

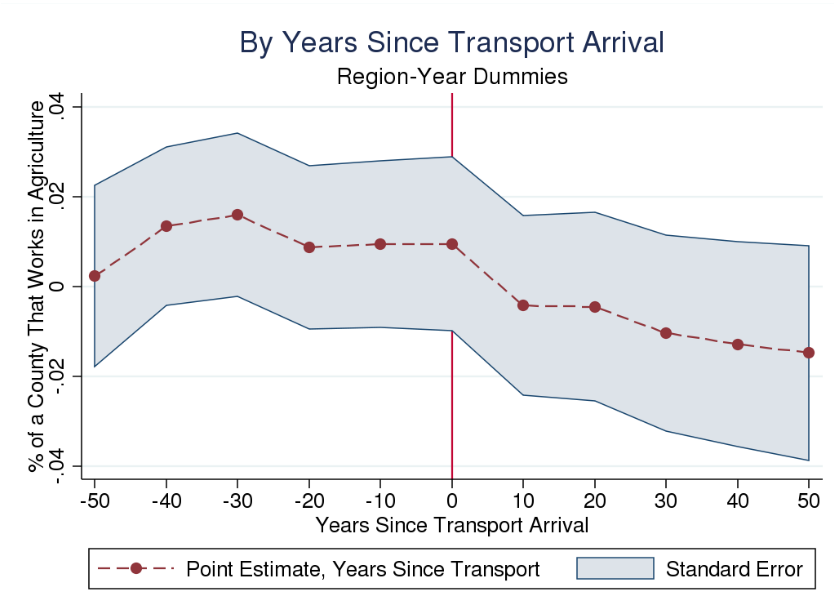




Back to page 37.

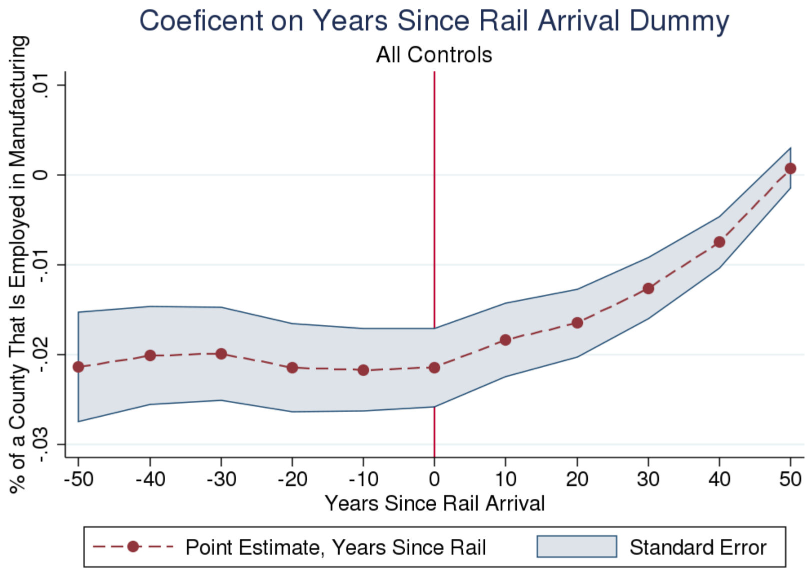
Figure 18: Years since Railroad Arrival

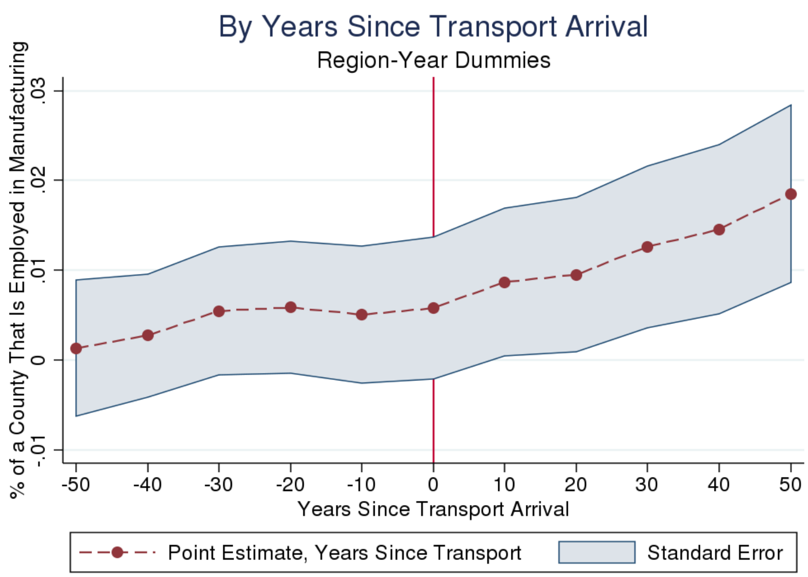




Back to page 37.

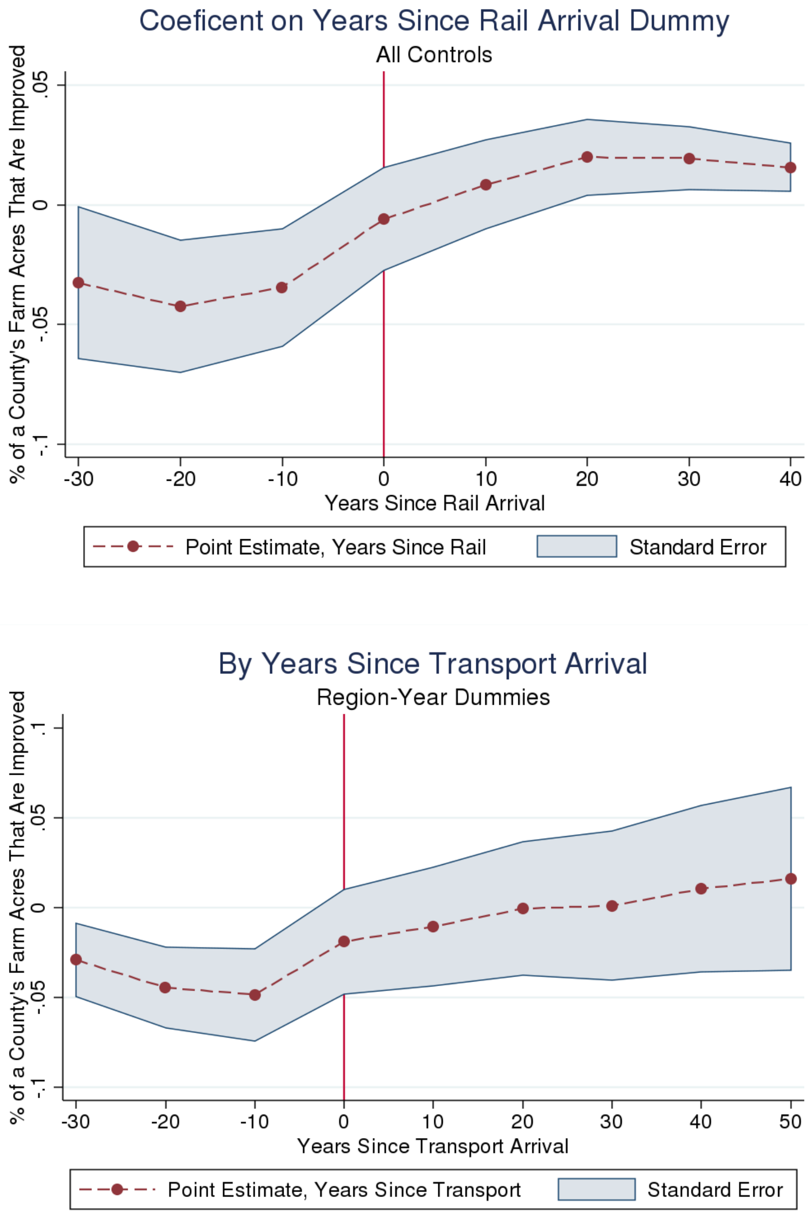
Figure 19: Years since Railroad Arrival





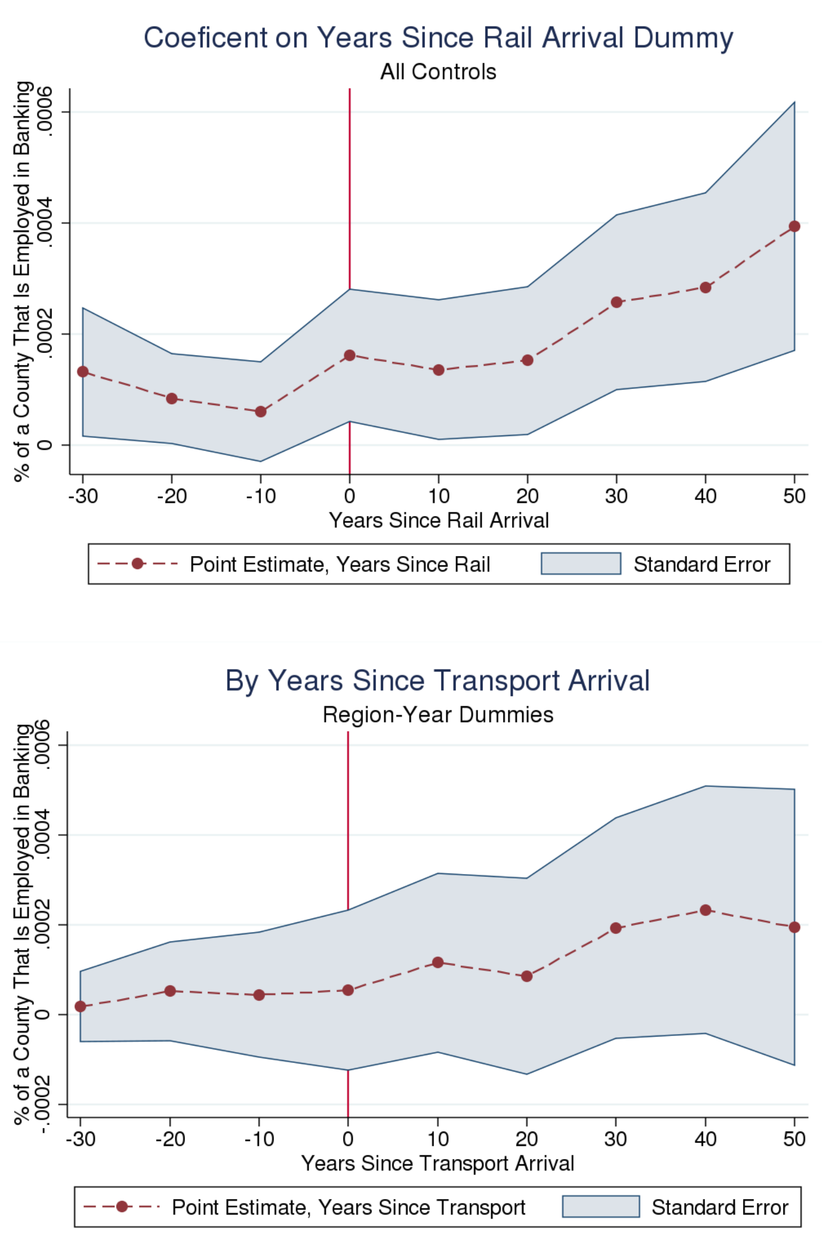
Back to page 37.

Figure 20: Years since Railroad Arrival



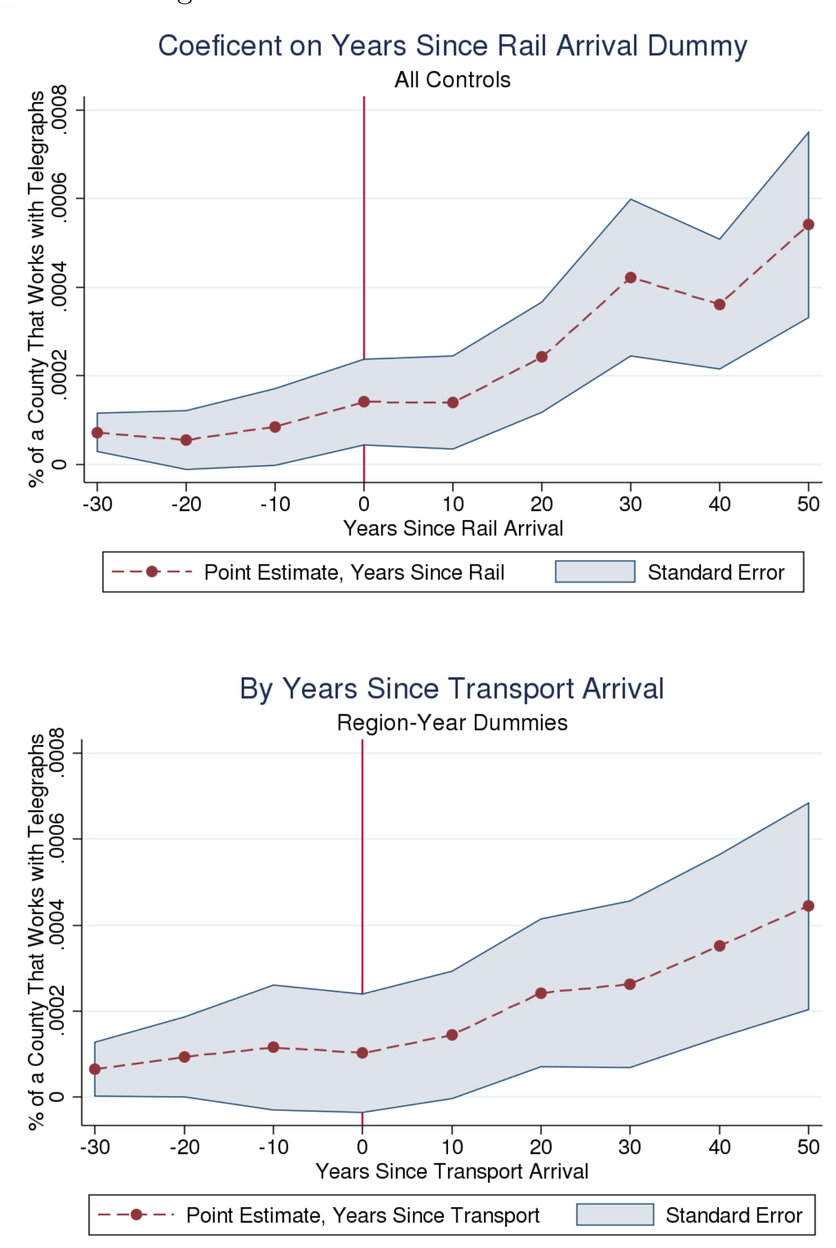
Back to page 37.

Figure 21: Years since Railroad Arrival



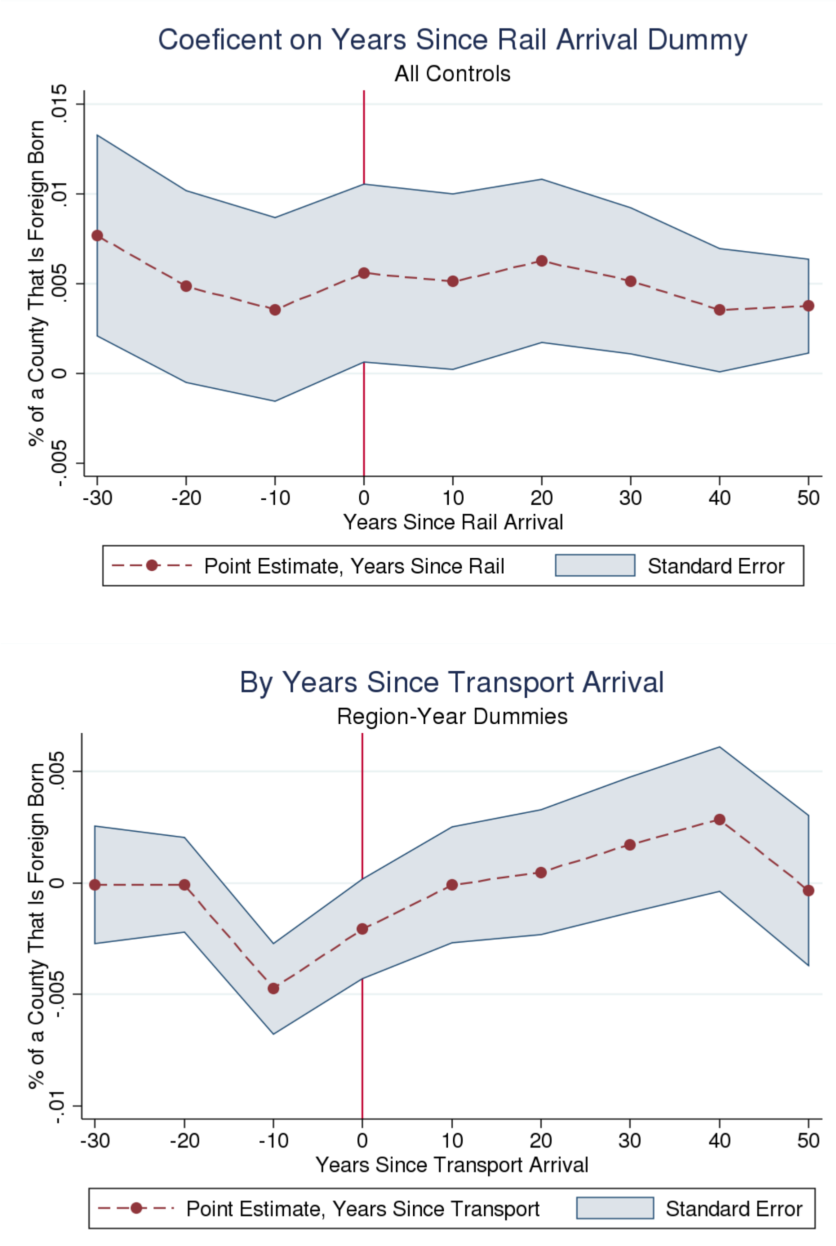
Back to page 37.

Figure 22: Years since Railroad Arrival



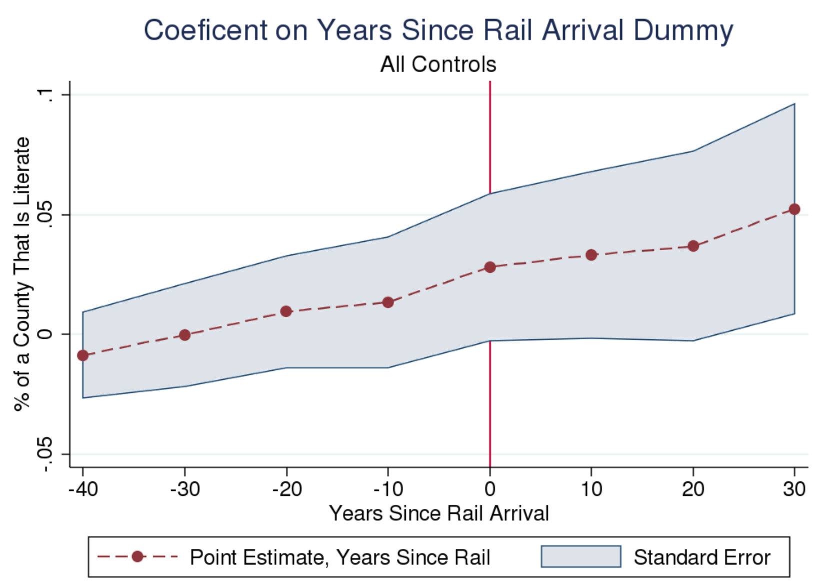
Back to page 37.

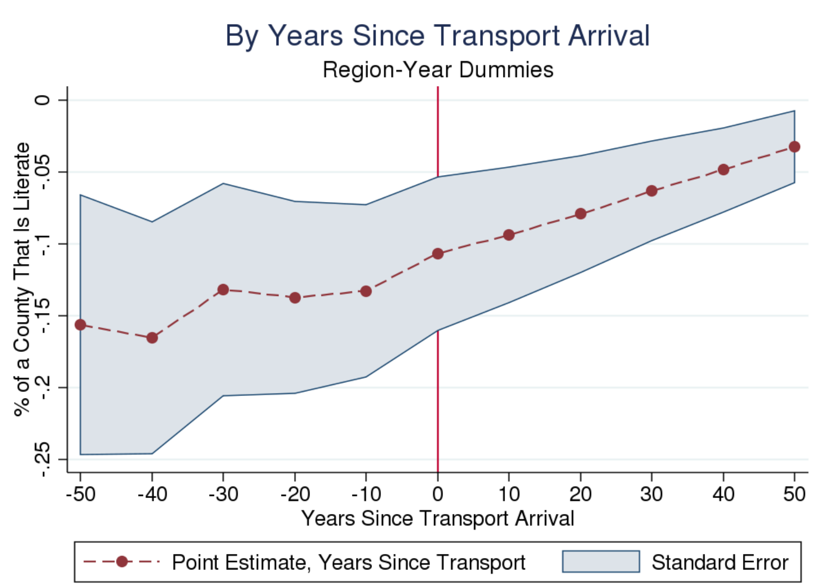
Figure 23: Years since Railroad Arrival



Back to page 37.

Figure 24: Years since Railroad Arrival





Back to page 37.