ARTICLE IN PRESS

JID: YEXEH [m3GeSsc;January 2, 2023;12:20]

Explorations in Economic History xxx (xxxx) xxx

Contents lists available at ScienceDirect

Explorations in Economic History

journal homepage: www.elsevier.com/locate/eeh

Research Paper

The long-run effects of childhood exposure to market access shocks: Evidence from the US railroad network expansion[☆]

Jeff Chan[*]

Wilfrid Laurier University, Canada

ARTICLE INFO

JEL classification:
F14
J13
N31
N71
O18
R23

Keywords:
Market access
Railroads
Linked census
Long-run effects

ABSTRACT

In this paper, I use the expansion of the US railroad network from 1900 to 1910 and the resulting spatial variation in increased market access to investigate whether economic shocks that occur during childhood have long-run ramifications on later-life outcomes, and the channels through which such effects manifest. I link individuals across the 1900, 1910, and 1940 full-count US Censuses and incorporate an instrumental variable strategy to help isolate the causal effect of market access. I find that, in the short run, sons are less likely to be literate and have more siblings. In the long-run, these sons then become less likely to be well-educated and earn lower incomes. The results of this paper shed light on the mechanisms through which railroad-induced market access and other economic shocks during childhood can impact individuals even in later life.

1. Introduction

Researchers across multiple fields in economics have studied the effects of economic shocks on individuals, firms, and regions. Market openness, unemployment, and other similar shifters are, as a result, well-understood in terms of how they affect economic outcomes. What is less well-understood, however, is how such shocks can manifest themselves in differences to outcomes later on in affected individuals' lives; this has in part been due to data limitations, which have only relatively recently been overcome. Trade economists, for example,[1] have used various empirical methodologies and longitudinal datasets for workers and regions to trace one-time shocks' persistent effects. Missing in this literature, however, is how market openness can affect people who were exposed to such shocks during childhood, and how those shocks can be transmitted through time to affect later-life outcomes. While such a study would be challenging to conduct using modern-day settings given data and time frame limitations, nineteenth century America provides both a setting in which market access, driven by railroad expansions, changed greatly in a short amount of time and recent data innovations, in terms of the availability of full-count Census microdata, may allow researchers to obtain a better understanding of how children can be affected by trade-type shocks even into their working-age lives.

☆ I thank Stephen Craig, Byron Lew, and David Rosé for providing helpful comments. I am grateful to participants at the 2022 Canadian Economics Association Meeting and the 2022 NARSC Meeting for their feedback. Ginny Hanson provided superb research assistance. Suggestions from the editor, Karen Clay, and two referees dramatically improved the paper. A prior version of this paper circulated as "The Long-Run Effects of Market Access Shocks from U.S. Railroad Network Expansions". Replication files are available at: https://doi.org/10.3886/E183563V1

* Corresponding author.
E-mail address: jchan@wlu.ca

[1] See Autor et al. (2014), Kovak and Morrow (2021), and Dix-Carneiro and Kovak (2017).

https://doi.org/10.1016/j.eeh.2022.101503
Received 13 June 2022; Received in revised form 18 December 2022; Accepted 20 December 2022
Available online xxx

Please cite this article as: J. Chan, The long-run effects of childhood exposure to market access shocks: Evidence from the US railroad network expansion, Explorations in Economic History, https://doi.org/10.1016/j.eeh.2022.101503

The construction and expansion of the railroads in the mid-nineteenth to early twentieth century is often considered one of America's landmark achievements, not only in transport infrastructure. Vast swathes of the interior and eventually the West coast were opened up to settlement and further in-migration as a result. This paper focuses on the latter stages of this expansion during the 1900–1910 period and investigates how the resulting shocks to market access induced by this railroad expansion led to changes in the short and long-run impacts of affected sons. Specifically, I examine how shocks to market access from 1900 to 1910 impacted sons' outcomes, as well as those of their parents, in 1910 using linked Census waves. I then further connect sons to later-life outcomes in the 1940 Census to investigate the long-run effects of market access shocks on men who were exposed as children. To help address issues of endogeneity, I use an instrumental variable that leverages variation in market access coming only from two sources: improvements in transport connectivity holding population levels constant, and connections to markets that are at least 500 km away.

I find that, in the short run, male children exposed to market access shocks from 1900 to 1910 have worsened outcomes. I find that male children experience a decreased likelihood of being literate by 1910 and a higher likelihood of living on a farm and being employed in 1910 in response to larger increases to market access. I also find that their mothers were more likely to have had more children. These results are not driven by a migration response.

In the long run, the worsened short-run outcomes translate into lower education and incomes for exposed children. Specifically, I find that 1900–1910 market access shocks led to lower income and occupational scores, while being more likely to have non-wage income above $50. Exposed individuals were also less likely to have completed high school and less likely to be in manufacturing employment. As in the short-run, I find no migration response for affected sons in the long-run up until 1940.

I conduct further analyses to disentangle the most important mechanisms through which market access affects 1940 income, and to verify the robustness of my results. I find that the additional inducement of remaining on a farm in 1910 is most important, although occupation choice and human capital accumulation also play important roles. I also find that market access shocks primarily occurred in less populated and rural areas; combined with the negative effects found in the rest of the paper, this suggests that market access could have contributed to a widening of the rural-urban divide in America. The main results are robust to a series of robustness checks, such as the dropping of the 10 largest counties in 1900, alternative linking procedures, and including a lagged measure of market access shocks. I also exploit sons who migrated to a new county by 1940 and show that migrants carry their original 1900 counties' market access shock with them, with the original shocks exhibiting significant effects on 1940 outcomes even after controlling for market access shocks in their new counties.

The findings in this paper shed additional light on the impacts of the changes in market access brought on by the American railroad network expansion, adding to the existing work on the effects of market access shocks due to railroads during the 19th and 20th centuries. Donaldson and Hornbeck (2016), Chan (2022b), and Hornbeck and Rotemberg (2021), most particularly, studied the role that market access played in American economic growth within the agricultural and manufacturing sectors across counties, finding that increased market access generally results in expansions to economic activity, in-migration, and increased land values. Another example is (Zimran, 2020), who examines market access from earlier American railroad expansions prior to the Civil War and link them to declines in health as proxied by height. This recent literature on railroad-induced market access in the US follows from work on India by Donaldson (2018). In comparison to this earlier work on railroad-induced market access, I instead use individual-level linked panel data to find that effects of market access on individuals who are exposed as children. My findings also contrast with this earlier work, which often finds positive effects on regional outcomes; I instead show that affected sons are negatively affected both in the short and the long run in terms of education and income. This paper therefore suggests that a more close-in view of the effects of economic shocks is a useful counterpart to regional-level analyses in providing a complete picture of the effects of such shocks.

A parallel track of research to that examining railroad-induced market access shocks instead directly investigates the impacts of railroads themselves on economic outcomes, particularly during the nineteenth century. Papers in this field include Atack (2016) and Atack et al. (2010) for the U.S., and Berger (2019); Hornung (2015), and Tang (2014) for Norway, Prussia, and Japan, respectively. Particularly notably in this area, work by Costas-Fernandez et al. (2022) uses linked English-Welsh Censuses in the 19th century to examine how proximity to train stations affect intergenerational mobility; in contrast, my paper uses actual incomes and education of affected male children in later life, and strives to link later-life outcomes to contemporaneous changes occurring within affected families in the short-run. More generally, I build on this existing body of work in my paper by focusing on the long-run impacts of market access shocks from railroad network expansions on the later-life outcomes of individuals exposed to these shocks in childhood by using linked Census data to track male children over their lifetimes up to 1940. This paper therefore contributes to the railroads literature by showing that railroad-based shocks can also have later-life negative consequences on those initially exposed as children.

There has also been a considerable amount of work looking into the historical effects of other types of transportation infrastructure beyond railroads to which my paper relates as well. Dalgaard et al. (2020), for example, study the long-run effects of the Roman road network and find that the location of Roman roads is predictive of modern road locations and economic activity. Similarly, a body of work has developed on the effects of canals on economic activity. Feyrer (2021) shows that the closure of the Suez canal had a negative effect on trade for affected country pairs. Using a market access methodology, Maurer and Rauch (2021) find that the increase to market access afforded to U.S. counties thanks to the opening of the Panama canal spurred population increases and had positive effects on a host of other economic outcomes. A pair of papers (Brooks et al., 2021 and Ducruet et al., 2022) have linked the containerization revolution in the mid-20th century with expansion of ports that benefited most from this new shipping technology. Finally, a substantial literature in urban economics has studied the effects of highway construction, particularly in the U.S. Jaworski and Kitchens (2019) focuses on the Appalachian Highway system and finds that it contributed over 53 billion dollars to the U.S. economy. A common theme across all these papers is that transportation infrastructure can have significant effects on economic activity and development, both historically and echoing into the modern day. My paper fits into the literature on transportation infrastructure and adds to this existing work by using individual-level panel data to show that expansions to transport infrastructure can have long-run consequences on men who were exposed as children; this notably stands in stark contrast to the aggregated positive effects.

I also add to a growing number of papers in economic history which make use of recent advances in linking individuals in historical full-count Censuses across different years. Specifically, I make use of algorithms and matches generated from work by Abramitzky et al. (2021) and Price et al. (2021). Both papers use combinations of information from the full-count Censuses, such as names and dates of birth, with machine learning to probabilistically link individuals across Censuses. These links are publicly available from Abramitzky et al. (2020), via their website.[2] These kinds of linking methods have been used in a variety of papers in economic history for various topics, including the intergenerational effects of abolition on white Southern slave owners (Ager et al., 2021), the impacts of immigration restrictions on local labour markets (Abramitzky et al., 2019), the long-run effects of the boll weevil (Ager et al., 2017), and the returns to education (Feigenbaum and Tan, 2020). My work uses similar Census linking methods to the above papers to instead estimate the long-run and intergenerational impacts of market access shocks from railroad expansions. In contrast to these papers, I examine a shock that has been shown to have positive effects on regional, aggregated outcomes and show that the use of linked, individual-level data can reveal that people affected by these historical large-scale shocks can be more negatively affected than aggregate data suggest.

Next, Section 2 outlines a conceptual framework and describes potential mechanisms, Section 3 describes the data and methods used, Section 4 goes over the empirical results, and Section 5 concludes.

2. Conceptual framework and potential mechanisms

In this section I highlight several potential mechanisms for why market access might affect later-life outcomes, particularly income, for exposed sons. I focus on three potential mechanisms in this discussion: 1) delaying the rural-urban transition, 2) differences in occupational choice, and 3) worsened investment in human capital.

At the beginning of the twentieth century, the majority of Americans still lived in rural areas; only 39.6% of the population lived in urban areas. By 1940, this share had increased to 56.5%. One reason this so-called structural transformation occurred in countries around the world in the nineteenth and twentieth centuries is the expansion of railroad networks and increased connectedness to faraway markets. A series of papers have connected railroads to urban growth in settings such as Sweden (Berger and Enflo, 2017), Japan (Yamazaki, 2019), and the American Midwest (Atack et al., 2010). Using access to world markets via rail connections in 19th century Argentina as variation in market access, Fajgelbaum and Redding (2022) provide a theoretical explanation for why market access might increase urbanization; in their model, regions with greater access to markets develop a comparative advantage in traded goods production, which increases urban growth in these regions. Market access could therefore increase the speed of the rural-urban transition.

On the other hand, much of this prior evidence comes from aggregate population statistics of urban versus rural populations. It is entirely possible that, in aggregate, urban populations grow due to in-migration or changes in fertility while families become more likely to remain on farms or in rural areas in response to market access shocks. This would be consistent with evidence from Donaldson and Hornbeck (2016) and Chan (2022b), who find that agricultural land values and output increase in response to market access. This increased demand for farming output could lower the relative value of moving to cities and towns. This mechanism is similar to that proposed by Bloom et al.. In their model, Bloom et al. argue that innovation increases from increased import competition because firms no longer value the firm activities that produce outputs competing with now-cheaper international products as highly. The resources used to produce these products therefore have a greater incentive to be re-allocated towards innovation. Similarly, in my setting, if market access increases the value of the output produced by farms and in their current occupations, this could trap sons on farms and in a more limited set of occupations than they otherwise could attain.

Knowing whether the rural-urban transition is affected positively or negatively by market access is important because rural and urban areas have very different economic opportunities for sons. Boustan et al. (2018) shows that, during the early twentieth century, the urban wage premium was 20–35%; in other words, those working in cities earned a substantial premium over rural workers. Rural areas might also have a worse selection of jobs available to workers. Lin (2011), for example, finds that the emergence of new occupations in the United States takes place in cities with higher concentrations of college graduates and with a greater variety of industries; such characteristics are not as commonly found in the rural United States during my sample period of 1900 to 1940. This implies that if market access induces sons to stay on farms or in areas with more limited variety in the types of industries or occupations that can employ them, they may subsequently be limited to a smaller set of occupations in their later lives with potentially lower earnings potential.

Market access could also, separately from geographical location, have an effect on occupational choice. On the one hand, a recent literature suggests that career and occupation switching is pro-cyclical. Carrillo-Tudela et al. (2016), for example, show that occupation switches in the UK happened largely during economic expansions, and such switches also often came with wage increases. This would imply that market access might increase occupational switching and increase occupation scores, thereby increasing incomes by 1940 for sons if areas receiving market access shocks subsequently go on to a sustained period of economic expansion. On the other hand, market access shocks could induce sons to enter or remain in occupations that suddenly receive higher premia but may not be an engine of future growth. For example, the American ice industry was at its largest at the turn of the twentieth century,

[2] More information on the linking conducted in this paper is available in the Appendix.

but would be soon rendered largely obsolete by the rise of refrigeration in the 1930s.[3] If market access induces economic booms, this could therefore inadvertently lead sons to become trapped in occupations or industries that become less high-paying by 1940.[4] Market access might also therefore have an effect on occupational choice, creating another channel through which market access affects income.

One final mechanism could be that changes to family composition induced by market access could lead to worsened investments in human capital accumulation in children and subsequently lower incomes in later life. Social scientists have long studied whether a quantity-quality trade-off exists for children, where families trade off having more children at the expense of investing in human capital in lower amounts per child. Becker and Lewis (1973) provides a theoretical underpinning of the forces behind the quantity-quality trade-off; in their model, families have a finite amount of resources and an increase in children leads to fewer resources left to invest in each child. Since (Becker and Lewis, 1973), economists have looked for evidence in favour of a quantity-quality trade-off.Bleakley and Lange (2009), in one example, find evidence of such a trade-off in the American South during the early twentieth century. If fertility is affected by market access, there could therefore be subsequently worsened human capital accumulation in children, and thus worsened incomes later in life.

One might be concerned that the U.S. educational context during this period may not have allowed for a quantity-quality trade-off to exist. I therefore provide some context for U.S. education in the early twentieth century here. At the beginning of the twentieth century, the United States was still far from attaining widespread high school attendance and completion. According to Goldin and Katz (2008), only 18% of Americans aged 15–18 (and therefore of high school age) were actually attending high school in 1910.[5] This is in part because the so-called high school movement, where compulsory high school laws were enacted across the country, had not taken off yet. Goldin and Katz (2011), in fact, name 1910 as the starting date of the high school movement. During this movement, the enrolment rate of 14–17 year olds in high school soared from just over 10% in 1910 to over 70% by the end of the movement in 1940. Part of the cause of this improvement in high school educational attainment in America was the adoption of compulsory high school education laws as well as legislation that strengthened the state's ability to combat child labour. As the majority of these changes and other factors that led to this shift in high school educational attainment occurred after my cohort of sons had exited school age, they would not have had as compelling a reason to attend high school as later generations; this is reflected in the relatively low high school completion rate of 28.5% in my sample. Educational attainment is therefore a choice that sons in my sample and their families were able to make, leading to the potential quantity-quality trade-off I observe in the data. This is also consistent with the findings supporting a quantity-quality trade-off model found by Bleakley and Lange (2009) for the American South in 1910, a similar context to mine.

Market access could also affect educational attainment in other ways beyond the quantity-quality trade-off. There is a significant literature that links economic booms to worsened educational attainment. Mosquera (2022) finds that natural resource booms lowered post-secondary educational attainment, as low-skilled jobs in the resource sector became relatively more attractive. Similarly, Lee (2021) finds that exposure to Mexican import competition from the North American Free Trade Agreement in 1994 led to U.S. workers enrolling more in community colleges. Market access, by increasing economic opportunity, could induce sons to enter the labour force instead of attending school; this short-run decision could have led to a narrower set of occupations available to sons by 1940 as well as lower income even within the same occupation.

3. Data and methodology

This section summarizes the data used for this paper.[6]

3.1. Data

I make use of the complete-count Censuses for the years 1900, 1910, and 1940 from IPUMS (Ruggles et al., 2021). In addition to ensuring that the data are easily accessible, IPUMS has done significant work to ensure that variables such as industry and occupation are consistently defined, as much as possible, for Census waves. The complete-count Censuses provide several major improvements over the previously available small samples, particularly for this project. First, the much larger sample sizes coupled with the county-level geographical identifiers ensure that each county in the U.S. is sufficiently well-represented enough to obtain enough cross-county variation in estimating the impacts of market access at the individual level. Most importantly, however, the much larger sample sizes ensure that a large enough sample of individuals survives the linking procedures across multiple Census waves; this is important because even the best linking procedures are probabilistic and imperfect, with nontrivial numbers of individuals dropped with each link made. Having the universe of individuals in the United States as a starting basis results in a final linked sample of individuals across the 1900,1910, and 1940 Censuses that is still of a size of several million.

I also make use of county-level variables originally derived from the Census and Census of Agriculture from Haines and ICPSR (2010). These variables are used in part to construct measures of market access. Calculating market access also requires

[3] For a more in-depth discussion of the American ice industry, see Cummings (1949).

[4] This is also similar to the "trapped factors" mechanism for innovation described by Bloom et al..

[5] In 1910, the sons in my sample would be roughly 10–18. A significant proportion of my sample would therefore have been of high school age in 1910, with the remainder of my sample reaching that age very shortly after.

[6] Replication files are available as (Chan, 2022a) and are available at: https://doi.org/10.3886/E183563V1.

county-to-county transport costs which vary as the railroad network changes over time. These costs come from the (Donaldson and Hornbeck, 2016) decadal transport costs database.[7]

The Donaldson and Hornbeck (2016) county-to-county transport costs were originally constructed using digitized historical railroad maps from Atack (2016), which are publicly available from Jeremy Atack's website.[8]

In a robustness check, I also use imputed income measures taken from Saavedra and Twinam (2020).

3.2. Linking census over time

A critical requirement of this paper is that individuals need to be linked across Censuses over time in order to link market access shocks in childhood to later-life outcomes. This requirement is satisfied by two relatively recent data innovations: the release of cleaned complete-count U.S. Censuses from 1850 to 1940 and publicly available linking variables to link individuals across these waves from the Census Linking Project.

Linking variables from the Census Linking Project (Abramitzky et al., 2020) are used to facilitate the linkages of individuals across Census waves.[9]. The Project provides publicly accessible linkages for individuals for any pairwise combination of complete-count Censuses between 1850 and 1940. For each Census pair, various linking methods are used and for each method, linkages between individuals in each Census are provided that allow researchers to use these links combined with the publicly available individual ID variables in the IPUMS data to create linked Census samples across years. I use these link variables from the Project to create a linked sample of individuals across the 1900, 1910, and 1940 samples. To link the 1900 and 1910 Censuses, I use the methodology used in Price et al. (2021). To then link the 1910 and 1940 Censuses, I used the links from the Abramitzky-Boustan-Eriksson method using NYSIIS standardized names as described in Abramitzky et al (2021). Further information on these methods and their choice are also provided in the Appendix.

3.3. Market access measure

I utilize the same measure of market access as is commonly used in the literature, such as in Donaldson and Hornbeck (2016) and Hornbeck and Rotemberg (2021). Specifically, I make use of the following county-level measure of market access:

$$MA_{ct} = \sum_{d} (\tau_{cdt})^{-\theta} * N_{dt} \tag{1}$$

Market access of a county c in year t is given by taking the product of the inverse of the transport cost between county c and destination county d in year t, τ_{cdt},[10] scaled up by the trade elasticity θ and county d's market size in year t N_{dt} (as proxied by its population) and then summing up over all possible counties d. The analysis will use a trade elasticity of 8.22 estimated in Donaldson and Hornbeck (2016), who study railroad-induced market access from 1870 to 1890 and use similar data.

The market access measure, as it uses the (Donaldson and Hornbeck, 2016) county-to-county transport data that is based on 1890 county boundary definitions and population data at the county level in 1900 and 1910 from Haines and ICPSR (2010), exists only for those counties that existed in 1900.[11] This is consistent with the literature (see Donaldson and Hornbeck, 2016) and explains the missing coverage in market access for Oklahoma (which did not become a state until 1907) in Fig. 1. This omission affects only a very small proportion of people living in the United States at the time; Oklahoma had a population of under 800,000 in 1900, compared to a U.S. population of over 76 million. This omission therefore has a very minimal effect on the representativeness of my sample.

In Figs. A.1 to A.3 in the Appendix, I map the railroad network in 1900 (Fig. A.1), the railroad network in 1910 (Fig. A.2), and the change in the railroad network between 1900 and 1910 (Fig. A.3).[12] The map shows that, by 1900, the US railroad network was already significantly developed and dense owing to the massive improvements completed in the preceding several decades. Nonetheless, the map of the network in 1910 shows further densification of the network in just 10 years; this densification seems to occur in all regions in the United States and shows no clear congregation in one area. To further highlight this point, in Fig. A.3 I display the rail segments that came into operation between 1900 and 1910. The map shows that there are improvements to the rail network made in every region in the US, with no obvious geographical clustering or concentration. This stands in stark contrast to the map of the change in market access in Fig. 1, where the changes are largely concentrated in the west. This implies that changes to market access are not necessarily driven by the expansion of the rail network locally but rather in improvements to the network in distant parts of the country; this distant improvement in networks is the basis for the instrumental variable strategy employed in this paper.

The patterns shown in the map line up well with the historical context in which the late stages of the American railroad expansion occurred. The 1910s are generally seen as the period in which the network's extent reached its peak, with railroad mileage itself

[7] The data has been made generously freely available on Richard Hornbeck's website.

[8] https://my.vanderbilt.edu/jeremyatack/data-downloads/

[9] Data are publicly available at https://censuslinkingproject.org/data/

[10] I follow the lead of Donaldson and Hornbeck (2016) and Chan (2022b) and define τ_{cdt} as one plus the lowest-cost route's freight rate cost, divided by the average price of transported goods. This average price is assumed to be 35, based on the approach of Fogel (1964).

[11] Specifically, I keep only observations for which a FIPS county code can be assigned.

[12] These maps use the Jeremy Atack GIS shapefiles of the railroad network (Atack, 2016), and use dates in that dataset where each rail segment was in operation by.

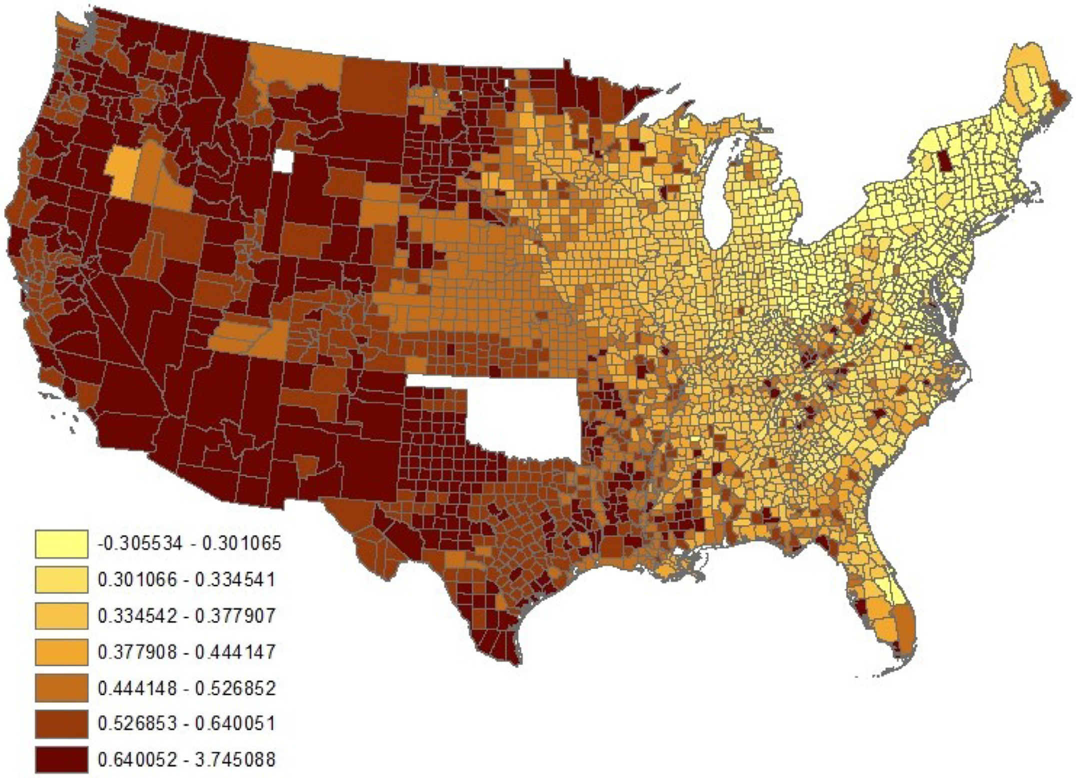

Fig. 1. Change in (asinh) market access, 1900–1910. Figure displays 7 quantiles of the change in market access (transformed by inverse hyperbolic sine) from 1900 to 1910. Darker shades reflect larger changes to market access. Empty sections of the map reflect counties for which data are not available.

peaking in 1916. The expansions in the network therefore represented completions of various lines and densification of coverage in areas that already had railroad connections. One example of this completion is the last transcontinental railroad line in 1910, the Western Pacific railroad terminating in Oakland, California.[13] Even the completion of another transcontinental line was more of an exception during this period, however. As (Wolmar, 2013) states in his book, the period from 1896 to 1916 was largely characterized by the main network of railroad lines being completed already and secondary lines being constructed. This account accords well with the pattern of new rail segments seen in Fig. A.3.

3.4. Specification

The main empirical analysis will rely on two specifications. One will address the short-run implications of 1900–1910 market access exposure on male individuals who were children at the time. I will then estimate a much longer-run set of results by linking children from 1900 to the 1940 Census and examining their outcomes 40 years onward. The short-run specification will take the form:

$$y_{i,c00}^{1910} = \beta_0 + \beta_1 * \Delta asinh(MA_{00-10,c00}) + \beta_2 X_{i,c00} + \gamma_{s(c00)} + \epsilon_i \tag{2}$$

The dependent variable, $y_{i,c00}^{1910}$, represents, for an individual i that resided in county $c00$ in 1900, an outcome in 1910. This specification focuses on the short-run implications of market access shocks on male children and their parents. In the case of some outcomes, such as the change in number of children birthed by the mother or the change in occupational score of the father, $y_{i,c00}^{1910}$ takes the form of a first-differenced variable with the value in 1900 subtracted from the value of the variable in 1910. For other variables, such as whether the father is in the manufacturing sector in 1910 and other binary indicators that do not lend themselves readily to a first-differenced version, $y_{i,c00}^{1910}$ is instead the 1910 value of that variable. In the case of these variables that are in levels, I additionally control for the 1900 value of that variable. The exceptions are the case of the child's outcomes in 1910, since the very young age of the sons in 1900 removes a substantial number of observations.[14]

[13] Vance et al. (2020).

[14] This is especially true for the literacy outcome. The inclusion of 1900 literacy as a control severely reduces the sample size, by several orders of magnitude, to several thousand.

The main coefficient of interest is β_1, which represents the impact of the change in transformed market access from 1900 to 1910 on individuals based on their 1900 county of residence ($\Delta asinh(MA_{00-10,c00})$). Market access is transformed using the inverse hyperbolic sine transformation, which is functionally similar to the log transformation but also allows for zeroes.[15] I also control for a vector of 1900 control variables for each individual, including age (and age squared), foreign-born status, and a non-white racial identity indicator. Cubic terms in the 1900 county of residence's latitude and longitude are included as controls; these controls, as in Donaldson and Hornbeck (2016), help mitigate concerns about unobservables that might be present in counties with similar latitude or longitude, such as climate, suitability for certain crops, and east-versus-west or north-versus-south differentials in growth trends. Finally, I include state fixed effects, for an individual's 1900 state of residence. Standard errors are clustered by 1900 state of residence.

The second specification investigates the long-run implications of childhood exposure to market access shocks:

$$y^{1940}_{i,c00} = \beta_0 + \beta_1 * \Delta asinh(MA_{00-10,c00}) + \beta_2 X_{i,c00} + \gamma_{s(c00)} + \epsilon_i \quad (3)$$

The specification is identical to that represented by the above equation, with the exception that the outcomes $y^{1940}_{i,c00}$ now represent individual outcomes in 1940. Importantly, the 1940 Census is the first time that income information is available, making feasible a study of the long-run impacts of market access in a similar way to studies examining long-run effects of trade shocks such as Autor et al (2014) and Devlin et al (2021). I also investigate the impacts of market access on other long-run employment and educational outcomes, such as occupational score, employment status, and educational attainment. Unlike in specification (2), all outcome variables are measured in levels, since 1940 variables either were not reported in earlier Census waves (such as income and educational attainment), or because the individuals would have been children in 1900. I include the same set of control variables as in the first specification, and also cluster standard errors by 1900 state of residence.

3.5. Instrumental variable

Estimation of the specification in (3) using OLS, even with the controls and fixed effects, may be subject to upward and downward bias. For example, one possibility is that market access shocks are driven by unobservable positive economic shocks; in other words, counties with larger market access increases from 1900 to 1910 may simply have been undergoing economic booms which also drove improvements to outcomes. In this example, the market access coefficient might be subject to upward bias. On the other hand, there may also be sources of downward bias. One channel of downward bias is correlation of market access over time. Market access shocks from 1900 to 1910 are correlated with changes to market access from the earlier period of 1870 to 1900. If counties that received earlier market access shocks already adjusted their outcomes in the 1870–1900 period, they may be subsequently less likely to increase those same outcomes in the 1900–1910 period. This correlation in market access over time could therefore be introducing downward bias in the estimates.

In order to help address concerns of endogeneity, I make use of an instrumental variable of the following form:

$$MAIV_{ct} = \sum_{d \in D_{500km}} (\tau_{cdt})^{-\theta} * N_{d,1870} \quad (4)$$

The instrument largely mirrors the main market access measure, with two key exceptions. First, population is instead fixed to 1870 levels ($N_{d,1870}$) instead of being allowed to vary from 1900 to 1910. Second, instead of calculating market access using the set of all other counties as potential destinations, the instrument uses only those counties that are at least 500 km away (D_{500km}).

This IV helps resolve at least two issues of note. First, populations might be endogenous and population growth might be correlated over time within counties. Correlated population growth, in particular, could lead to market access shocks being correlated over time and changes to market access therefore also being correlated over different time periods. My IV helps resolve this issue by using only 1870 population levels in the construction of market access; this means that changes to market access will be driven purely by changes to transport costs induced by changes to the railroad network's structure.[16] One other issue that the IV helps address is spatial correlation in shocks that could be driving both increased connectivity to nearby counties and local economic outcomes. The IV mitigates these concerns by using only variation in market access coming from other counties that are at least 500 km away, as measured by centroid-to-centroid distances. The instrument therefore mitigates issues of both upward and downward bias, as described earlier.

In addition to mitigating issues of endogeneity, the instrument is highly predictive of changes in market access from 1900 to 1910. The correlation between the original change in transformed market access and the instrument is 0.9614, suggesting that this instrument is very strong. In addition, the correlation between the instrument and the change in market access from 1870 to 1900 is only 0.2775, suggesting that the IV uses contemporaneous changes in the railroad network to obtain exogenous variation in market access from 1900 to 1910 that is separate from changes in market access that occurred from 1870 to 1900. Table A.1 in the Appendix presents the first-stage results from the IV regressions for the main 1940 outcomes table. The IV is highly statistically significantly positively associated with changes in market access, with a first stage F-statistic of approximately 1000. This value suggests a very strong instrument.[17]

[15] I also show in the Appendix that the main results are robust to taking logs of the market access measure instead.

[16] The data confirms that removing population changes from market access dramatically reduces the correlation between the change to market access in 1870–1900 and 1900–1910 from 0.4352 to 0.2734.

[17] The strength of the instrument also suggests that the key variation in market access come from reductions in transport costs, and not changes to the sizes of markets.

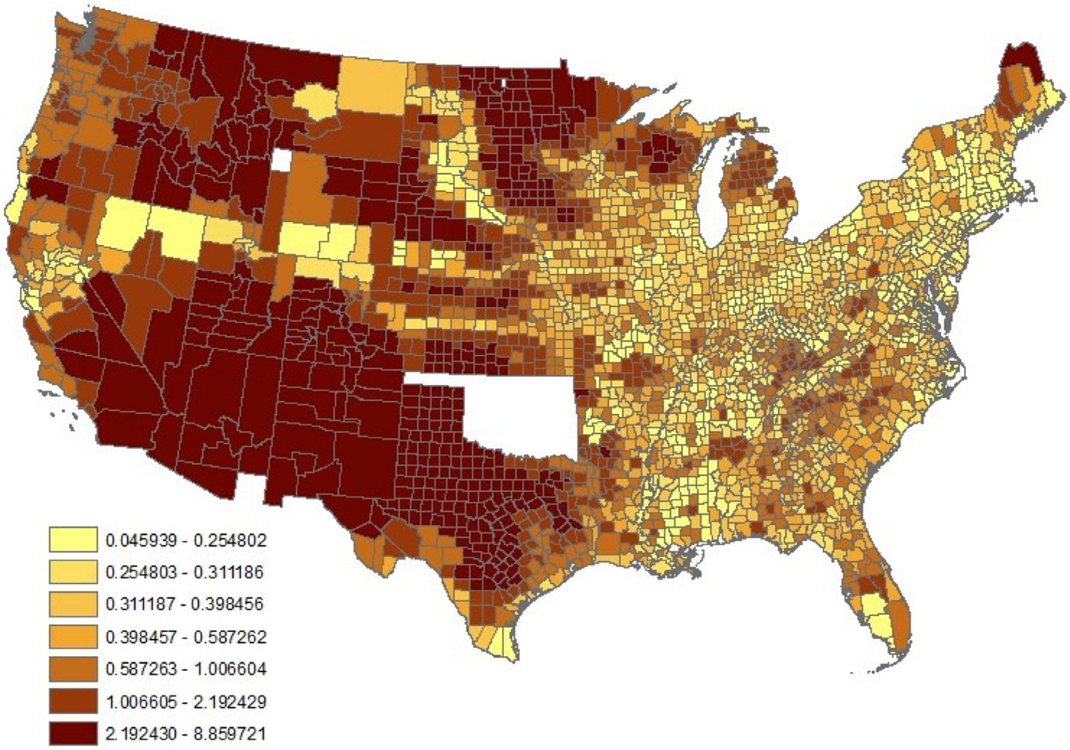

Fig. 2. IV: change in (asinh) market access with 500km donut and 1870 population, 1900–1910. Figure displays 7 quantiles of the change in the instrumental variable for market access (transformed by inverse hyperbolic sine) from 1900 to 1910. The instrument removes market access from any counties within 500 km of that county and fixes population levels to 1870 levels, leveraging transportation infrastructure improvements only for time variation. Darker shades reflect larger changes to market access. Empty sections of the map reflect counties for which data are not available.

3.6. Data description

I first display the spatial variation in the market access shock from 1900 to 1910, using the main market access measure that uses contemporaneous populations for each county. The map in Fig. 1, shows the change in inverse hyperbolic sine-transformed market access from 1900 to 1910.[18] Counties that were more westward clearly exhibited larger increases to market access, although considerable variation exists even within state; this is important, as the specifications will rely on within-state variation in the market access shock.

When fixing population levels of each county to their 1870 levels and also removing variation in market access coming from nearby (within 500 km) counties to construct the instrument, the dispersion and within-state variation in the market access shock significantly increases. This variation is shown in Fig. 2, with darker colours again representing counties that have larger positive shocks, this time for the instrument. Fig. 2 illustrates two features about the instrument. First, the pattern across space is broadly consistent with that from Fig. 1, which aligns with the strength of the instrument. Second, the removal of contemporaneous population levels in the construction of the market access measure removes a significant amount of the spatial correlation in market access across adjacent counties; this suggests that the instrument is of use in helping to mitigate the issue of geographic spillovers from neighbouring counties in the main market access measure.

To show the impacts of the various sample selection choices on the size of the sample, Table 1 presents the cumulative impact of including the various sample selection criteria, one by one. Starting from the full-count 1900 Census, which has over 75 million observations, by far the most costly sample restriction comes from the observations lost by keeping only people who can be linked across all three Census waves: 1900, 1910, and 1940. The linkings leave only approximately 3.7 million individuals out of the 75 million from the 1900 Census. Out of these 3.7 million who can be linked across all 3 Censuses, over 1.4 million were 8 or below in 1900; this age-based sample restriction limits the analysis to individuals who would have been approximately 18 or below by 1910. Another sample restriction is to keep only individuals who can be matched to both fathers and mothers in both 1900 and 1910; this is to ensure that parental outcomes can be fully explored as a mechanism through which their sons are impacted in the long-run by market access shocks. A final restriction is to keep only whites in the sample, due to the under-representation of nonwhites in the

[18] In the map, the change in transformed market access is binned into 7 quantiles, with darker colours representing more positive changes.

Table 1
Sample restrictions.

Sample Restriction	Obs.
Full-Count 1900 Census	75,824,712
Linking across 3 Census years (1900, 1910, 1940)	3,727,644
Only those aged 8 or below in 1900	1,469,556
Can be matched to mother and father in 1900 and 1910 with non-missing MA, dropping non-whites	1,137,094

Notes: this table shows the observation counts (second column) that remain after the listed sample restriction in the first column is implemented. The restrictions are implemented in sequential order, with the last row representing the last of the sample restrictions and the corresponding number of observations in the final estimation sample.

Table 2
Summary statistics.

Variable	Obs	Mean	Std. Dev.	Min.	Max.
Panel A: 1900 variables					
age	1,137,094	3.803	2.529	0	8
non-immigrant status	1,137,094	0.994	0.075	0	1
school enrolment	1,136,996	0.205	0.404	0	1
father in mfg.	1,137,094	0.108	0.310	0	1
father's asinh(occ. score)	1,137,094	3.365	1.077	0	5.075
mother's no. of children	1,133,351	4.550	2.711	1	25
mother's no. of surviving children	1,132,417	3.951	2.264	1	25
Panel B: 1910 variables					
literacy	1,116,661	0.981	0.136	0	1
father in mfg.	1,137,094	0.146	0.353	0	1
Δ father's asinh(occ.score)	1,137,094	0.164	1.353	-5.075	5.075
Δ mother's no. of children	1,057,174	1.800	1.802	0	27
Δ mother's no. of surviving children	1,037,170	1.458	1.518	-1	26
changed counties, 1900–1910	1,134,810	0.248	0.432	0	1
changed states, 1900–1910	1,137,094	0.101	0.302	0	1
Panel C: 1940 variables					
asinh(income)	705,646	7.804	0.870	0.881	9.211
asinh(occ. score)	1,137,094	3.715	0.905	0	5.075
non-wage income over $50	1,100,050	0.396	0.489	0	1
employed	1,137,094	0.898	0.303	0	1
mfg.	1,137,094	0.192	0.394	0	1
completed high school	1,137,094	0.285	0.451	0	1
changed counties, 1900–1940	1,134,118	0.601	0.490	0	1
changed states, 1900–1940	1,137,094	0.296	0.456	0	1
Panel D: Market access variables					
$\Delta asinh(MA_{00-10,c00})$	1,137,094	0.414	0.217	-0.305	4.839
$\Delta asinh(MAIV_{00-10,c00})$	1,137,094	0.069	0.136	-0.648	3.652
$\Delta asinh(MA_{00-10,c40})$	1,089,732	0.449	0.281	-0.305	4.839
$\Delta asinh(MAIV_{00-10,c40})$	1,089,732	0.078	0.165	-0.648	3.652

Notes: Panels A, B, and C report the summary statistics for the main variables of interest in 1900, 1910, and 1940, respectively. Panel D shows summary statistics for the main right hand side variable of interest (the change in transformed market access) as well as the instrumental variable. All summary statistics reported only for observations in the estimation sample. asinh() refers to the inverse hyperbolic sine transformation.

linked sample relative to the full-count Census; this issue is discussed further below. This final restriction leaves us with a sample of approximately 1.1 million individuals who can be traced in 1900, 1910, and 1940.[19] The resulting sample is also geographically diverse across the states.[20]

Finally, Table 2 displays summary statistics for the main variables used in this paper. In the top panel of Table 2, I present characteristics of my linked sample in 1900. The sample is restricted to be 8 or under in 1900, and the mean age is correspondingly low at 3.8 years. This young age also explains the low school attendance of the sample; only 20% of the sample is enrolled in school in 1900. The sample is also overwhelmingly (over 95%) non-foreign born. The top panel also summarizes the probability that the

[19] Some specifications have slightly fewer individuals due to missing values for variables.

[20] Specifically, there are 522,820 individuals from the Midwest Census region, 291,160 from the Northeast, 303,524 from the South, and 68,153 from the West. Locations are measured according to individuals' 1900 state of residence, and Census regions are defined using current-day definitions and which region the modern-day equivalent states belong to.

Table 3
Childhood outcomes, 1910.

Dep. Var.:	(1) $\mathbb{1}$(in school) OLS	(2) $\mathbb{1}$(literate) OLS	(3) $\mathbb{1}$(employed) OLS	(4) $\mathbb{1}$(in school) IV	(5) $\mathbb{1}$(literate) IV	(6) $\mathbb{1}$(employed) IV
$\Delta asinh(MA_{00-10,c00})$	−0.0133	−0.0156***	0.0208*	−0.00335	−0.0237***	0.0393***
	(0.0106)	(0.00573)	(0.0110)	(0.00703)	(0.00799)	(0.0130)
Observations	1,137,094	1,116,661	1,137,094	1,137,094	1,116,661	1,137,094
R-squared	0.272	0.053	0.294	0.272	0.053	0.294
Mean of Dep. Var.:	0.826	0.981	0.328	0.826	0.981	0.328

Notes: *** denotes 1% significance, ** denotes 5% significance, * denotes 10% significance. All specifications include, as controls, a foreign-born dummy, latitude and longitude cubics for the 1900 county of residence, age and age squared in 1900, and 1900 state of residence fixed effects. The right hand side variable of interest, $\Delta asinh(MA_{00-10,c00})$, is the inverse hyperbolic sine transformed change in market access (MA) in an individual's 1900 county of residence. asinh() refers to the inverse hyperbolic sine transformation. $\mathbb{1}()$ refers to a dummy variable equal to 1 if the condition within the brackets is satisfied; in these dummies, "in school" is whether the son is in school in 1910, "literate" is whether the son is literate in 1910, and "employed" is whether the son is employed in 1910. OLS refers to columns where coefficients are estimated using OLS. IV refers to columns where an instrumental variable regression is employed. For variables in levels, controls for initial 1900 levels of the right hand side variable are included as controls. Standard errors are clustered at by 1900 state of residence.

father is employed in manufacturing as well as the inverse hyperbolic sine transformed occupation score. Finally, mothers of the sons in the sample had approximately 4 children in 1900, including the child in my sample.

The next panel of Table 2 shows the outcome variables used in the short-run results. Given that the mean age of my sample in 1910 is now in the teens, it is not surprising that the large majority of the sample is in school; approximately 82% of the individuals in my sample are enrolled in school by the 1910 Census. On average, fathers tended to have upgraded their occupations according to their occupation score, and more fathers were employed in manufacturing in 1910 relative to 1900. Mothers were more likely to have had more children in the intervening 10 years as well. Finally, a non-trivial amount of migration occurred from 1900 to 1910, with almost 25% of individuals having moved counties and almost 10% having moved states.[21]

The third panel of Table 2 shows the summary statistics for 1940 outcomes of the now-grown sons, who are now between 40 and 48. The transformed income and occupational scores are summarized. A non-trivial proportion of the sample, almost 40%, earns non-wage income over $50. A high proportion of the sample is also employed, at almost 90%, with 19% of the sample being associated with the manufacturing sector. High school completion is quite low, at only 28% of the sample having achieved a high school level of education or higher by 1940. Finally, a large proportion of these individuals were observed in counties that differed from their original 1900 county of residence (approximately 60%), with almost 30% of individuals being observed in a different state than in 1900.

The last panel of Table 2 describes the market access shock variables for 1900–1910. In particular, the main market access shock variable shows that, although the period examined is later than that of Donaldson and Hornbeck (2016), market access is still increasing quite significantly during this later era.

The various necessary sample restrictions may lead to the final sample in this paper being different in observables to the population in 1900. In order to investigate these differences, I take the full-count 1900 Census and keep those aged 8 or below, but otherwise make no further sample restrictions; this allows for comparison to Table A.2 in the Appendix, which shows 1900 variable summary statistics for the linked sample (but not dropping nonwhites). The summary statistics for those 8 and under in the full-count Census are presented in Table A.3 in the Appendix. The summary statistics broadly line up with those from the linked sample used in the main analysis. The only exception is that the linked sample is overwhelmingly white, with only 4.1% being non-white. In contrast, 13.7% of the full 1900 Census is non-white. Put together, the comparison of summary statistics suggests that the linked sample may not be representative of non-whites during this period, but otherwise looks quite similar to the under-8 population in 1900. I therefore drop nonwhites from the main estimation sample in the paper.[22]

4. Results

4.1. Short-run results

I first examine the impacts of 1900–1910 market access shocks on short-run outcomes for exposed male children. Table 3 presents the impacts on school attendance, literacy, and employment in 1910. Columns 1, 2, and 3 show OLS estimates of market access's impact on attendance, literacy, and employment while columns 4, 5, and 6 present the corresponding IV estimates. Given the large

[21] For a more thorough discussion of the migration variables and their construction, please see the Appendix.
[22] In a robustness check, I confirm that adding nonwhites back into my estimation sample does not impact my results.

proportion of children in school in 1910, at over 82%, and the small, precisely estimated coefficients in columns 1 and 4, I conclude that there is no economically or statistically significant effect of market access on school attendance in 1910. On the other hand, I do find that increases to market access had a negative effect on literacy by 1910. Taking the IV estimate in column 5 as the preferred estimate, I find that the mean increase in market access from 1900 to 1910 decreased the probability of being literate (both for reading and writing) by approximately 1 percentage point, against a mean of 98.1% of the sample being literate in 1910; this implies that market access had a small, but significant contemporaneous effect on education outcomes. Finally, column 6 presents one potential reason for why education outcomes might already be worsened by 1910: rising opportunity costs of investing in human capital; specifically, column 6 shows that market access has a positive effect on a son being employed in 1910. Keeping in mind that the sons in my sample would be 10–18 in 1910, the results in Table 3 imply that market access may have induced young men and boys to work at the potential expense of human capital accumulation and worsened later-life outcomes.

Given that the sons in my sample were 0 to 8 in 1900, market access shocks that propagated via labour markets or other economic channels would have also likely transmitted to the sons via their parents. I therefore examine the impacts of market access on parental outcomes in 1910; these results are shown in Table 4. I again focus on the IV estimates, in columns 6–10. I first show, in columns 6 and 7, that market access shocks from 1900 to 1910 have little impact on the father's contemporaneous employment. There is no statistically significant effect on the change in occupation score,[23] and only a very marginally significant and economically small impact on being employed in manufacturing. The mean increase in market access would result only in less than a half percentage point increase in being in manufacturing, from a baseline 14.6% in the sample in 1910.

Column 8 of Table 4 shows that market access increases the probability that the family lives on a farm in 1910, even conditional on living on a farm on 1900. I highlight this result because this is consistent with one of the proposed mechanisms from the framework section: delayed rural-urban transitions. As Chan (2022b) found, market access increases agricultural output and land used in affected counties. This increased agricultural activity may have induced families to remain in agriculture and stay on their farms, at the expense of being able to take advantage of living in urban areas and the new types of occupations available there.

There is also a statistically significant increase in the number of children that have ever been had and that have survived by the mother. Columns 9 and 10 show, more specifically, that positive changes in market access lead to increases in the number of children ever had by the mother, and the number of children had by the mother who have survived. Evaluated at the mean change in market access of 0.414, the corresponding change in surviving children by mothers is approximately 0.136;[24] in other words, the mean increase in market access from 1900 to 1910 induced mothers to have an extra 0.136 children who survived between 1900 and 1910. The finding that market access increases fertility is consistent with work by Ager et al. (2020) who find that negative agricultural shocks can reduce fertility; market access, as in Chan (2022b), if viewed as a positive agricultural shock can therefore increase fertility, potentially at the expense of quality of education per child.

This finding therefore suggests that one mechanism through which sons may have been affected in both the short and the long run may have been through family compositional changes. If households have more children, this could impact the per-child level of resources and parental care due to a quantity-quality tradeoff,[25] leading to worsened educational and later-life outcomes. Closer to the context of my paper, Bleakley and Lange (2009) use the eradication of hookworm disease in the American South in 1910 as a natural experiment and find support for the quality-quantity trade-off model.[26]

4.2. *Long-run results*

Table 5 investigates the possibility that the previous short-run results may have had longer-run impacts by looking at the role that 1900–1910 market access shocks play on these sons' later life outcomes in the 1940 Census. I again focus on the IV estimates in columns 7–12. I find that market access shocks from 1900 to 1910 led to worsened wage income and lower occupation scores by 1940.[27] In addition, individuals are also less likely to be employed or be in manufacturing, although these estimates are modest in size and only marginally statistically significant. Finally, consistent with the literacy results in the short-run table and with the family compositional changes leading to lower per-child resources, I find that individuals affected by the 1900–1910 shocks had lower

[23] In the main results that use occupation score (including the later sons' 1940 outcomes), I include occupation scores of zero in the sample. In Table A.19 of the Appendix, I exclude any observations where the occupation score values are zero (or where the occupation score is zero in either period for the change in father's occupation score from 1900 to 1910). The results in Table A.19 are very similar to the main results, showing that the inclusion or exclusion of zeroes in the occupation score has very little to no effect on my results.

[24] This value is obtained by multiplying 0.414 by the coefficient for market access shock in the change in the number of children specification in column 9 of Table 4.

[25] While some work using modern-era data such as Black et al. (2005) have shown that quality-quantity tradeoffs are weaker in such contexts, papers in historical contexts such as Fernihough (2017), who examines the quality-quantity tradeoff in early 1900s Ireland, have suggested the presence of a negative relationship between education and family size.

[26] Specifically, Bleakley and Lange (2009) find that hookworm disease's eradication led to declines in fertility and increased school attendance in the American South. This is consistent with a quality-quantity type model since hookworm affected mostly children and increased the cost of human capital investment.

[27] The sample size for wage income is smaller because I follow Clay et al. (2021) and other papers in this literature by dropping those with incomes of zero and including only those sons who are employed in 1940.

Table 4
Parental Outcomes, 1910.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dep. Var.:	Δ asinh(father's occ. score) OLS	𝟙(father mfg.) OLS	𝟙(farm) OLS	Δ no. children OLS	Δ no. surviving children OLS	Δ asinh(father's occ. score) IV	𝟙(father mfg.) IV	𝟙(farm) IV	Δ no. children IV	Δ no. surviving children IV
$\Delta asinh(MA_{00-10,c00})$	0.00423	0.00210	–0.0212	0.268***	0.192***	–0.0499**	–0.0107	0.0396**	0.329***	0.274***
	(0.0209)	(0.0102)	(0.0248)	(0.0729)	(0.0537)	(0.0200)	(0.00680)	(0.0167)	(0.0741)	(0.0686)
Observations	1,137,094	1,137,094	1,137,094	1,057,174	1,037,170	1,137,094	1,137,094	1,137,094	1,057,174	1,037,170
R-squared	0.003	0.185	0.409	0.059	0.067	0.002	0.185	0.406	0.059	0.067
Mean of dep. var.:	0.164	0.146	0.426	1.800	1.458	0.164	0.146	0.426	1.800	1.458

Notes: *** denotes 1% significance, ** denotes 5% significance, * denotes 10% significance. All specifications include, as controls, a foreign-born dummy, latitude and longitude cubics for the 1900 county of residence, age and age squared in 1900, and 1900 state of residence fixed effects. The right hand side variable of interest, $\Delta asinh(MA_{00-10,c00})$, is the inverse hyperbolic sine transformed change in market access (MA) in an individual's 1900 county of residence. asinh() refers to the inverse hyperbolic sine transformation. 𝟙() refers to a dummy variable equal to 1 if the condition within the brackets is satisfied; in the dummies, "father mfg" is whether the son's father is in manufacturing in 1910. OLS refers to columns where coefficients are estimated using OLS. IV refers to columns where an instrumental variable regression is employed. For variables in levels, controls for initial 1900 levels of the right hand side variable are included as controls. Standard errors are clustered at by 1900 state of residence.

Table 5
Sons' outcomes, 1940.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dep. Var.:	asinh(income) OLS	asinh(occ. score) OLS	$\mathbb{1}$(nonwage inc.) OLS	$\mathbb{1}$(employed) OLS	$\mathbb{1}$(mfg) OLS	$\mathbb{1}$(HS) OLS	asinh(income) IV	asinh(occ. score) IV	$\mathbb{1}$(nonwage inc.) IV	$\mathbb{1}$(employed) IV	$\mathbb{1}$(mfg) IV	$\mathbb{1}$(HS) IV
$\Delta asinh(MA_{00-10,c00})$	–0.0318	–0.00139	–0.00630	–0.0102***	–0.00395	–0.0189	–0.175***	–0.0615***	0.0223*	–0.00762**	–0.0122*	–0.0502***
	(0.0554)	(0.0241)	(0.0135)	(0.00376)	(0.00819)	(0.0122)	(0.0454)	(0.0189)	(0.0134)	(0.00358)	(0.00688)	(0.0133)
Observations	705,646	1,137,094	1,100,050	1,137,094	1,137,094	1,137,094	705,646	1,137,094	1,100,050	1,137,094	1,137,094	1,137,094
R-squared	0.049	0.011	0.028	0.002	0.024	0.015	0.049	0.011	0.028	0.002	0.024	0.015
Mean of dep. var.:	7.804	3.715	0.396	0.898	0.192	0.285	7.804	3.715	0.396	0.898	0.192	0.285

Notes: *** denotes 1% significance, ** denotes 5% significance, * denotes 10% significance. All specifications include, as controls, a foreign-born dummy, latitude and longitude cubics for the 1900 county of residence, age and age squared in 1900, and 1900 state of residence fixed effects. The right hand side variable of interest, $\Delta asinh(MA_{00-10,c00})$, is the inverse hyperbolic sine transformed change in market access (MA) in an individual's 1900 county of residence. asinh() refers to the inverse hyperbolic sine transformation. $\mathbb{1}()$ refers to a dummy variable equal to 1 if the condition within the brackets is satisfied; in the dummies, "nonwage inc." is whether the son has nonwage income in 1940, "employed" is whether the son is employed in 1940, "mfg" is whether the son is in manufacturing in 1940, and "HS" is whether the son has a high school education (or better) in 1940. OLS refers to columns where coefficients are estimated using OLS. IV refers to columns where an instrumental variable regression is employed. For the income regressions, the sample includes only those with non-zero incomes and are employed. Standard errors are clustered at by 1900 state of residence.

educational attainment by 1940; specifically, I find that the probability of having a high school education or better is significantly lower with exposure to larger market access shocks in childhood.[28]

In Table 6, I investigate further 1940 outcomes that focus on the family and dwelling characteristics of the individuals in my sample. For brevity, I focus only on the IV results in columns 6–10. I do not find that childhood market access shocks statistically significantly affect an individual's likelihood of being married; shocks do, however, increase the number of children had. Similar to their own parents, market access shocks also seem to influence the size of affected children's own families. I then move on to studying the effect of market access on the status of dwellings. First, column 8 shows that increased market access from 1900 to 1910 increases the probability that affected sons live on farms in 1940; this suggests that childhood exposure to market access may have slowed the transition from agriculture to manufacturing, or similarly the transition from rural to urban. Finally, in column 10 I focus on only on those individuals who own their own dwelling.[29] The results show that house values are lower for those who received a market access shock during childhood; this is consistent with the increased likelihood of living on a farm relative to an urban setting. Taken together, Table 6 implies that childhood market access shocks can result in changes to family composition and housing differences in later life.

4.3. Mechanism

In this subsection, I present additional evidence on what mechanisms are operating to affect later-life outcomes in 1940 from Table 5.

One potential mechanism for my results could be that market access may have driven shifts in or out-migration from initial 1900 counties of residence, both in the short run (by 1910) or in the longer run (by 1940). For example, if families are induced to out-migrate to other areas due to rising land values from market access, this may have pushed families to periphery areas that ended up doing less well economically, leading to worsened outcomes. Conversely, if families were induced to stay in areas hit with shocks instead of migrating to other areas as they otherwise might have done, this may have trapped families in regions that then declined over time. I test this mechanism by estimating whether 1900–1910 market access shocks led to out-migration to other counties or states by 1910 or 1940. The results, shown in Table 7, clearly illustrate that there is no migration response to market access shocks and that migration is not an important mechanism driving my main results.

I also conduct further analyses to determine whether migration could be confounding the main results. First, in Table A.4 of the Appendix, I test a longer-run definition of migration with a dummy equal to 1 if an individual moved in either 1900–1910 or 1910–1940; this accounts for individuals who may have moved from 1900 to 1910 but moved back to their original county by 1940. The results in Table A.4 show that, using this longer-run definition, market access still has no statistically significant effect on migration across states or counties. Next, in Table A.5 I show that the main long-run results from Table 5 remain qualitatively unchanged if I use only a sample of individuals who never migrate from their 1900 county of residence, either in 1910 or 1940. Taken together, Tables A.4 and A.5 suggest that migration is not a major concern in this paper's main analyses.

Despite the lack of differential migration behaviour in sons affected by market access shocks, the presence of individuals who moved counties by 1940 from their original 1900 county of residence provides an opportunity to test for one method of transmission for market access. More specifically, I take advantage of movers by using them as a subsample and estimating the effects of 1900–1910 market access shocks based on both the 1900 and 1940 county of residence. If market access shocks based on the original 1900 county of residence drives my results, then this would imply that childhood exposure is most important. On the other hand, if market access shocks based on 1940 county of residence is more important, then sons that have moved to other counties by 1940 may simply have adopted their recipient counties' market access shocks. I test these channels in Table 8.

First, in Panel A of Table 8 I use my main market access shock measure by 1900 county and estimate my main outcomes for movers only. The results largely echo the results from the larger main sample, although the results for manufacturing employment and nonwage income are weaker. Nonetheless, the patterns, particularly for income and occupation score, are very similar to those found in the main analysis. Next, in Panel B, I replace the market access shock with my alternative measure based on the 1940 county of residence. The results are also quite similar to my main results, suggesting that it is possible that movers adopt their new counties' market access shocks instead of carrying their childhood market access shocks with them. I therefore test which of the two shocks matters most by including both in the same regression; these results are presented in Panel C. The horserace-type regressions show that, for income, both types of market access shocks have negative effects and have similar magnitudes. For employment and high school education the original 1900 county is what matters most, while for nonwage income and manufacturing employment the 1940 county of residence's market access shock appears to be more important. Overall, Table 8 shows that both types of market access shocks are important to son's outcomes in 1940, although importantly for my main analysis market access shocks by 1900 county of residence remain robust to the inclusion of the 1940 county market access shock; sons do therefore carry their childhood exposure with them to their new counties. One potential concern in this exercise is multicollinearity between the two market access shock variables; if the two variables are highly correlated then it may not be possible to separately identify the two shocks' effects. The correlation, while positive, is only 0.35, suggesting that multicollinearity is not a substantial issue.

[28] It should be noted that the inclusion of 1900 state of residence fixed effects helps control for differential adoption of mass compulsory high schooling across states as documented by Goldin and Katz (2011).

[29] column 9 showed that there was a slight positive effect on owning their own dwelling, although the magnitude is relatively small and the coefficient is only marginally significant at the 10% level.

Table 6
Family and dwelling outcomes, 1940.

Dep. Var.:	(1) 𝟙(ever married) OLS	(2) no. children OLS	(3) 𝟙(farm) OLS	(4) 𝟙(own dwell.) OLS	(5) asinh(house value) OLS	(6) 𝟙(ever married) IV	(7) no. children IV	(8) 𝟙(farm) IV	(9) 𝟙(own dwell.) IV	(10) asinh(house value) IV
$\Delta asinh(MA_{00-10,c00})$	–0.00443 (0.00522)	0.140* (0.0726)	–0.0107 (0.0279)	–0.0187 (0.0328)	–0.163** (0.0688)	0.00532 (0.00346)	0.295*** (0.0921)	0.0651*** (0.0221)	0.0170* (0.0103)	–0.310*** (0.0753)
Observations	1,137,094	1,137,094	1,137,094	1,137,094	314,264	1,137,094	1,137,094	1,137,094	1,137,094	314,264
R-squared	0.005	0.021	0.068	0.015	0.094	0.005	0.021	0.068	0.015	0.093
Mean of dep. var.:	0.896	1.809	0.238	0.526	8.340	0.896	1.809	0.238	0.526	8.340

Notes: *** denotes 1% significance, ** denotes 5% significance, * denotes 10% significance. All specifications include, as controls, a foreign-born dummy, latitude and longitude cubics for the 1900 county of residence, age and age squared in 1900, and 1900 state of residence fixed effects. The right hand side variable of interest, $\Delta asinh(MA_{00-10,c00})$, is the inverse hyperbolic sine transformed change in market access (MA) in an individual's 1900 county of residence. asinh() refers to the inverse hyperbolic sine transformation. 𝟙() refers to a dummy variable equal to 1 if the condition within the brackets is satisfied; in the dummies, "ever married" is if the son was ever married by 1940, "farm" is whether the son lives on a farm in 1940, and "own dwell." is whether the son owns their dwelling in 1940. OLS refers to columns where coefficients are estimated using OLS. IV refers to columns where an instrumental variable regression is employed. Standard errors are clustered at by 1900 state of residence

Table 7
Migration.

Dep. Var.:	(1) $\mathbb{1}$(migrate, 1910) OLS	(2) $\mathbb{1}$(migrate, 1940) OLS	(3) $\mathbb{1}$(mig. states, 1910) OLS	(4) $\mathbb{1}$(mig. states, 1940) OLS	(5) $\mathbb{1}$(migrate, 1910) IV	(6) $\mathbb{1}$(migrate, 1940) IV	(7) $\mathbb{1}$(mig. states, 1910) IV	(8) $\mathbb{1}$(mig. states, 1940) IV
$\Delta asinh(MA_{00-10,c00})$	–0.00107	–0.0169	0.00478	0.00848	–0.00160	0.0114	–0.00876	–0.00811
	(0.0159)	(0.0235)	(0.00891)	(0.0120)	(0.0108)	(0.0141)	(0.00692)	(0.0118)
Observations	1,134,810	1,134,118	1,137,094	1,137,094	1,134,810	1,134,118	1,137,094	1,137,094
R-squared	0.029	0.028	0.025	0.046	0.029	0.028	0.025	0.046
Mean of dep. var.:	0.248	0.601	0.101	0.296	0.248	0.601	0.101	0.296

Notes: *** denotes 1% significance, ** denotes 5% significance, * denotes 10% significance. All specifications include, as controls, a foreign-born dummy, latitude and longitude cubics for the 1900 county of residence, age and age squared in 1900, and 1900 state of residence fixed effects. The right hand side variable of interest, $\Delta asinh(MA_{00-10,c00})$, is the inverse hyperbolic sine transformed change in market access (MA) in an individual's 1900 county of residence. asinh() refers to the inverse hyperbolic sine transformation. $\mathbb{1}()$ refers to a dummy variable equal to 1 if the condition within the brackets is satisfied; "migrate,1910" is if the son changed counties between 1900 and 1910, "migrate,1940" is if the son changed counties between 1900 and 1940, "mig. states,1910" is if the son changed states between 1900 and 1910, and "mig. states,1940" is if the son changes states between 1900 and 1940. OLS refers to columns where coefficients are estimated using OLS. IV refers to columns where an instrumental variable regression is employed. Standard errors are clustered at by 1900 state of residence.

Table 8
Movers' Outcomes, 1940.

Dep. Var.:	(1) asinh(income) OLS	(2) asinh(occ. score) OLS	(3) 𝟙(nonwage inc.) OLS	(4) 𝟙(employed) OLS	(5) 𝟙(mfg) OLS	(6) 𝟙(HS) OLS	(7) asinh(income) IV	(8) asinh(occ. score) IV	(9) 𝟙(nonwage inc.) IV	(10) 𝟙(employed) IV	(11) 𝟙(mfg) IV	(12) 𝟙(HS) IV
Panel A: 1900–1910 shock based on 1900 county of residence												
$\Delta asinh(MA_{00-10,c00})$	–0.0319	–0.0172	–0.00305	–0.00642**	–0.00321	–0.0244**	–0.113***	–0.0457***	0.00266	–0.00673*	–0.00455	–0.0485***
	(0.0324)	(0.0139)	(0.00634)	(0.00317)	(0.00472)	(0.0100)	(0.0315)	(0.0136)	(0.00747)	(0.00393)	(0.00527)	(0.0124)
Observations	460,413	702,298	676,769	702,298	702,298	702,298	460,413	702,298	676,769	702,298	702,298	702,298
R-squared	0.045	0.008	0.018	0.002	0.021	0.016	0.045	0.008	0.018	0.002	0.021	0.016
Panel B: 1900–1910 shock based on 1940 county of residence												
$\Delta asinh(MA_{00-10,c40})$	0.0564**	0.0401***	0.0292***	–0.0159***	–0.0409***	0.0669***	–0.143***	–0.0329***	0.0660***	–0.00134	–0.0544***	–0.00888
	(0.0249)	(0.0109)	(0.00779)	(0.00465)	(0.00766)	(0.0157)	(0.0184)	(0.00893)	(0.00587)	(0.00530)	(0.00870)	(0.00896)
Observations	436,326	662,374	637,987	662,374	662,374	662,374	436,326	662,374	637,987	662,374	662,374	662,374
R-squared	0.045	0.008	0.018	0.002	0.020	0.018	0.041	0.007	0.018	0.002	0.020	0.016
Panel C: 1900–1910 shocks based on 1900 and 1940 counties of residence												
$\Delta asinh(MA_{00-10,c00})$	–0.0266	–0.0137	–0.0100	–0.00647**	–9.51e-05	–0.0240**	–0.101***	–0.0401***	–0.00761	–0.00768**	8.82e-05	–0.0455***
	(0.0345)	(0.0151)	(0.00643)	(0.00296)	(0.00483)	(0.0110)	(0.0346)	(0.0142)	(0.00787)	(0.00365)	(0.00538)	(0.0138)
$\Delta asinh(MA_{00-10,c40})$	0.0568**	0.0403***	0.0294***	–0.0158***	–0.0409***	0.0672***	-0.141***	–0.0317***	0.0662***	–0.00113	–0.0544***	–0.00760
	(0.0247)	(0.0109)	(0.00781)	(0.00465)	(0.00765)	(0.0157)	(0.0183)	(0.00877)	(0.00582)	(0.00529)	(0.00869)	(0.00899)
Observations	436,326	662,374	637,987	662,374	662,374	662,374	436,326	662,374	637,987	662,374	662,374	662,374
R-squared	0.045	0.008	0.018	0.002	0.020	0.018	0.040	0.007	0.018	0.002	0.020	0.016
Mean of dep. var.:	7.804	3.715	0.396	0.898	0.192	0.285	7.804	3.715	0.396	0.898	0.192	0.285

Notes: *** denotes 1% significance, ** denotes 5% significance, * denotes 10% significance. All specifications include, as controls, a foreign-born dummy, latitude and longitude cubics for the 1900 county of residence, age and age squared in 1900, and 1900 state of residence fixed effects. $\Delta asinh(MA_{00-10,c00})$ is the inverse hyperbolic sine transformed change in market access (MA) in an individual's 1900 county of residence. $\Delta asinh(MA_{00-10,c40})$ is the inverse hyperbolic sine transformed change in market access (MA) in an individual's 1940 county of residence. asinh() refers to the inverse hyperbolic sine transformation. 𝟙() refers to a dummy variable equal to 1 if the condition within the brackets is satisfied; in the dummies, "nonwage inc." is whether the son has nonwage income in 1940, "employed" is whether the son is employed in 1940, "mfg" is whether the son is in manufacturing in 1940, and "HS" is whether the son has a high school education (or better) in 1940. OLS refers to columns where coefficients are estimated using OLS. IV refers to columns where an instrumental variable regression is employed. For the income regressions, the sample includes only those with non-zero incomes and are employed. All regressions estimated using only individuals who live in a different county in 1940 than they did in 1900. Standard errors are clustered at by 1900 state of residence.

Table 9
Effects of family and personal characteristics on 1940 income.

Dep. Var.:	(1) asinh(income)	(2) asinh(income)	(3) asinh(income)	(4) asinh(income)	(5) asinh(income)	(6) asinh(income)	(7) asinh(income)
$\mathbb{1}(\text{farm}_{10})$	–0.358***						–0.183***
	(0.0178)						(0.00878)
$\mathbb{1}(\text{employed}_{10})$		–0.201***					–0.0332***
		(0.0124)					(0.00351)
asinh(occ. score$_{40}$)			0.967***				0.831***
			(0.0278)				(0.0240)
$\mathbb{1}(\text{HS}_{40})$				0.582***			0.384***
				(0.0183)			(0.00772)
$\mathbb{1}(\text{literate}_{10})$					0.565***		0.309***
					(0.0244)		(0.0127)
no. oth. surviving children						–0.0391***	–0.00830***
						(0.00170)	(0.000651)
Observations	705,646	705,646	705,646	705,646	692,246	699,198	685,915
R-squared	0.083	0.057	0.253	0.142	0.055	0.061	0.309
Mean of dep. var.:	7.804	7.804	7.804	7.804	7.804	7.804	7.804
1 SD mediated effect of MA:	–0.059	–0.003	–0.025	–0.012	–0.006	–0.004	–0.064

Notes: *** denotes 1% significance, ** denotes 5% significance, * denotes 10% significance. All specifications include, as controls, a foreign-born dummy, latitude and longitude cubics for the 1900 county of residence, age and age squared in 1900, and 1900 state of residence fixed effects. The left hand side variable is the inverse hyperbolic sine transformation of the son's income in 1940. asinh() refers to the inverse hyperbolic sine transformation. $\mathbb{1}()$ refers to a dummy variable equal to 1 if the condition within the brackets is satisfied; "HS" is whether the son has a high school education (or better) in 1940, "literate$_{10}$ is whether the son was literate in 1910. "occ. score$_{40}$ is the occupation score of the son in 1940. All specifications estimated using OLS. The sample includes only those with non-zero incomes and are employed. Standard errors are clustered at by 1900 state of residence. The 1 SD mediated effect of MA row at the bottom of the table refers to the effect that a 1 standard deviation increase in market access has on asinh(income), via the mediating variable(s)'s coefficient in this table and the effect of the MA shock on the mediating variable(s) in question from the coefficient on MA in previous tables showing the effect of MA on those mediating variables.

The analysis in this paper suggests that market access can have negative impacts on individuals' incomes. The results also suggest that this might occur via differences in occupational choice, educational attainment, and family composition that occur due to market access shocks. I verify whether this mechanism is plausible in Table 9 by regressing transformed 1940 income against various individual and family characteristics that are affected by market access. Specifically, previous results showed that market access had a positive effect on the probability of living on a farm in 1910, being employed in 1910, and the change in the number of surviving children by the mother between 1900 and 1910. Market access also had a statistically significant negative effect on the likelihood of being literate in 1910, 1940 occupation score and high school attainment by 1940. I find that each of these variables affects income in the expected way, and that the impact of market access shocks would affect these characteristics in a way that would be consistent with worsened income by 1940. For example, sons with mothers who had a greater increase in surviving children from 1900 to 1910 had lower 1940 incomes which, when combined with the fact that market access increases the number of children had by mothers, allows me to conclude that market access may have an effect on income by impacting earlier-life family structure. In the last column (column 7) of Table 9, I include all variables in one regression to help determine which plays the most important role in explaining 1940 income. The coefficients for all variables remain highly statistically significant and of the same sign as when each is included on its own.

To more clearly evaluate each mediating variable's effect on income in Table 9, the last row of the table reports the effect of a one standard deviation increase in the size of a market access shock on each variable (taken from the market access coefficients where each of those variables is the dependent variable, reported in prior tables). I then take that change in the mediating variable and multiply it by the coefficient in Table 9 to get the implied effect on income.[30] The results of this exercise show that farm status in 1910 has the largest single effect on 1940 income, with an implied decrease of -0.059 from a one standard deviation larger market access shock. For comparison, the effect of a one standard deviation larger market access shock on income, using the coefficient from column 7 of Table 5, is 0.073. This implies that remaining on farms was the leading mechanism of market access's effect on 1940 incomes. Occupation score and the education variables also play a non-trivial role, however, implying that occupation choice and human capital accumulation are also ways in which sons' later life incomes were impacted.

In the final column of Table 9, I include all mediating variables and use the coefficients in this last column to calculate the total, combined effect of a one standard deviation larger market access shock on income, via these mediating variables. The overall implied decrease in (transformed) income is -0.064. This value is approximately 87% of the overall effect of market access on income, suggesting that these mechanisms explain much of the overall effect together. Of all the individual variables, 1910 farm status remains very important; the individual effect remains large at -0.030, compared to the original -0.059 from column 1. Occupation choice also remains important in explaining the loss in income, given that the coefficient on occupation score in column 7 remains largely of the

[30] For example, in column 1, I first take one standard deviation of asinh(market access), which is 0.414, and multiply by the coefficient on farm status in 1910, which is 0.396 from Table 4. I then take this product and multiply by the farm coefficient in Table 9, obtaining -0.059.

same magnitude compared to column 3. The implied effect of a one standard deviation larger market access shock on income via occupation score is 0.021, compared to the original effect of -0.025 from column 3. The education variables (high school attainment and literacy) also remain collectively important in explaining the effect on income. Taken together, Table 9 provides some clear indications for the mechanisms through which later-life income is affected by early-life exposure to market access shocks and finds important roles for delayed rural-urban transition, occupation choice, and human capital attainment. I therefore find supporting evidence for all three mechanisms discussed in the framework section.

4.4. Robustness checks

In this subsection, I discuss a series of robustness checks that help rule out potential issues, sources of bias, or mechanisms.

One concern might be that, even with the use of the IV, there may be lingering issues of endogeneity of market access. In particular, expansions in the railroad network from 1900 to 1910 could have been influenced by within-county forces. For example, New York financiers, faced with a positive shock within the region, might then decide to invest in improving railroad access to distant markets by investing in railroads. In order to combat this concern, I drop the top 10 counties by 1900 population from the sample and re-estimate the main 1940 long-run results. The results, reported in Table A.6, clearly show that the main results are robust to dropping these top 10 counties. The estimates, in fact, appear to be even more precisely estimated than in the main table. This implies that the endogeneity of railroad expansion was not likely to have been a leading driver of my main estimates.

In Table A.7, I drop all counties that reported non-zero cotton acreage planted in the 1900 Agricultural Census. This effectively removes all counties that have any reliance on cotton. This robustness check accounts for the effects of the boll weevil, a recurring pest that wreaked havoc on cotton during the period and have been shown by Ager et al. (2017) to have adverse effects on the agricultural sector in the South; if market access shocks are correlated with cotton reliance in counties then my main results could be picking up long-term effects of the boll weevil. The results in Table A.7 are broadly consistent with those obtained using my main sample, although the nonwage and employed specifications are no longer significant at conventional levels. The income, occupation score, and high school attainment coefficients are still significant at the 5% level, and are similar in magnitude to the main results in Table 5. Taken together, the results show that my findings hold in non-cotton counties, mitigating concerns of the boll weevil confounding them.

Another historical shock that occurred during the sample period was the Spanish flu of 1918, which killed tens of millions worldwide. If counties that received market access shocks also suffered disproportionately from Spanish flu years later, then any long-term scarring effects on sons' 1940 outcomes could be the result of the Spanish flu and not market access. To mitigate this concern, in Table A.8, I use only a subsample of sons who moved counties between 1900 and 1910. These sons should not have been exposed to the Spanish flu, having moved to other counties. The table's results are very similar to my main table, although the smaller sample size means that the coefficients on nonwage income, employment, and manufacturing are no longer significant; the income, occupation score, and high school attainment coefficients remain significant at the 1% level. Table A.8 therefore provides supporting evidence that my results are not driven by exposure to the Spanish flu.

The period from 1900 to 1940, particularly near the end of this time, saw the increasing electrification of America. While even by 1940 the country had not come close to becoming completely electrified, notably in rural areas, this transformation allows for the possibility that counties that saw market access shocks may have subsequently had different rates of adoption for electricity. Arguably, this could be itself part of the longer-run scarring effect of market access on children. Nonetheless, to help address this possibility, I split the sample into areas that were more versus less likely to be electrified even by the end of the sample period. To perform this splitting exercise, I refer to Lewis (2018). In his paper, Lewis (2018) shows via maps that by 1930, near the end of my sample period, electrification had already begun but was largely concentrated in the Northeast region and in California. I therefore define the Northeast and California as being high-likelihood areas for electrification and the rest of the U.S. as low-likelihood areas. I then re-estimate my main 1940 results with these two subsamples. The results are reported in A.9. First, Panel A shows the non-electrified areas. The results from the main table remain robust, with similar magnitudes and levels of significance. I then re-estimate the 1940 results for only the Northeast and California in Panel B. The magnitudes of the coefficients remain similar to those from the main analysis, although the significance is notably lower likely due to the much smaller sample sizes. Taken together, the results from A.9 show that even conditioning on using samples that were likely to have similar levels of electrification by the end of my sample period, my main results are robust. I therefore conclude that electrification is not the major channel through which my results operate.

In Table A.10, I compile my linked sample using a different linking procedure for linking the 1910 and 1940 full count Censuses and report the main 1940 outcome results again with this new sample.[31] Specifically, instead of using the Abramitzky-Boustan-Eriksson method with NYSIIS standardized names, I instead make use of matches using exact names only.[32] This matching procedure results in a smaller sample, as the observation counts in Table A.10 clearly demonstrate. Nonetheless, the main patterns present in the results from Table 5 are qualitatively unchanged with this smaller sample. This suggests that the main results are not especially sensitive to alternative Census linking methods.

[31] I maintain the use of the 1900–1910 linkage using the (Price et al., 2021) procedure. This is due to this method's superior matching quality from the incorporation of additional hand-matched information from Census Tree. For more information on the linking methods used in the main sample, see the Appendix.

[32] This serves to eliminate one data cleaning step which may introduce additional false matches, at the expense of sample size and the elimination of true matches due to misspellings over time.

Another concern could be that, even with the instrumental variable approach, the market access coefficients on the 1900–1910 shocks may be partly capturing lingering impacts from earlier, 1870–1900 changes to market access. In Table A.11 of the Appendix, I therefore re-estimate the 1940 outcome results but additionally control for the change in market access from 1870 to 1900. Due to the high correlation between the raw measures of market access from 1870 to 1900 and that for 1900–1910, I present only the IV estimates. The results clearly show that the coefficients in the main analysis are largely unchanged in both magnitude and statistical significance. Furthermore, the coefficients on the 1870–1900 market access measure are consistently very small and statistically insignificant. I therefore conclude that the results in this paper are the impacts of 1900–1910 market access shocks, and not largely explained by any correlation to earlier market access shocks from earlier railroad expansions.

Tables A.12 to A.18 of the Appendix further demonstrate additional robustness exercises and provide further supplementary detail for the main results. In Table A.12, I show that taking logs of the market access measure instead of transforming it via the inverse hyperbolic sine makes no qualitative difference in the main results for sons' outcomes in 1940. Next, Table A.13 re-runs the main results for son's outcomes in 1940 but includes nonwhites in the sample and adds a nonwhite dummy as a control variable.[33] The results in Table A.13 are virtually identical in magnitude, sign, and significance to the main table of results, showing that the inclusion or exclusion of nonwhites is not driving my results. Table A.14 re-estimates my main 1940 outcomes but instead uses Conley standard errors, which allow for spatial correlation of observations. I set the range to 100 km, and measure distances using the latitudes and longitudes of each 1900 county's centroid. The standard errors are generally even smaller than the state-clustered standard errors I use in the rest of the paper, suggesting that clustering by state is the more conservative option.

One might be concerned that market access shocks might be correlated with initial county characteristics; if so, my results could be picking up differential trends in counties with differing characteristics. In order to test this, I regress several 1900 county characteristics against (instrumented) 1900–1910 market access changes. The results of this test are presented in Table A.15. Two characteristics are derived from my estimation sample: the 1900 share of non-immigrants and the 1900 average age of sons. I also test the inverse hyperbolic sine of 1900 population, the 1900 urban population share, and the 1900 inverse hyperbolic sine of agricultural land and building value per capita. The latter variable in particular was used by Donaldson and Hornbeck (2016) as their outcome measure and is likely to be correlated with agricultural productivity and other unobservables of concern. I find that non-immigrant share and land and building value per capita are not associated with market access shocks, but that average age of sons, population, and urban population share are. However, the coefficient for age is small in magnitude, and I already control for age in my main specification. To test if population, urban share, and land value are important omitted variables, in Table A.16 I re-estimate my main table of 1940 outcomes with these 3 variables included; controlling for the urban share, in particular, is similar to the suggestion of controlling for distance to the nearest urban center.I find that the main results are largely robust to their inclusion, although the magnitudes do decrease somewhat. For example, the coefficient for asinh(income) decreases from -0.175 in the original specification to -0.097, although the coefficient remains statistically significant at the 1% level. Other coefficients are similarly diminished in magnitude, and the coefficients in the employment and manufacturing specifications are no longer significant at conventional levels (although these two variables coefficients for market access were less statistically significant in the original table as well). Overall, the table shows that initial conditions do matter, but do not eliminate the main findings in the paper.

In Table A.17, I aggregate the sample to the 1900 county of residence level. This aggregation requires that variables be changed in some cases. Specifically, the dummy variables are replaced with the share of my sample that has that characteristic and the income, occupation score, and age variables are instead average values of those variables. I weight each county in my sample by the number of individuals from my main sample, to best mimic my main specification. The table shows that the main results remain very similar to the main individual-level analysis in Table 5; with the exception of employment, all coefficients remain statistically significant, of the same sign, and of similar magnitude. This implies that aggregating my sample to the county level does not eliminate the large majority of my results.

In Table A.18, I present the same specification as in the main table showing sons' outcomes in 1940, but I display more of the control variables' coefficients for reference.

In Table A.19 in the Appendix, I further explore the robustness of the occupation score measure. For brevity, I focus only on the discussion of the IV results in the text. In columns 4 and 5, I re-estimate the change in the father's occupation score from 1900 to 1910 and the son's occupation score in 1940, respectively. In this table, however, I drop zeroes in the occupation score. The results are remarkably similar to the main tables' coefficients, implying that the exclusion of zeroes in occupation scores would not affect the estimated coefficients for the effect of market access on occupation score, quantitatively or qualitatively. In column 6, I consider whether an alternate measure of occupation score would affect my main results. Specifically, I use the LIDO score measure from Saavedra and Twinam (2020).[34] The LIDO measure is a predicted income score generated by taking additional characteristics such as demographics and location into account and using machine learning to agnostically determine the best predictors of 1950 income. when I use this LIDO measure in column 6 instead of occupation score, the resulting coefficient on market access is very similar in magnitude to that obtained in my occupation score regressions, and remains highly statistically significant. This implies that my occupation score results are robust to the use of alternate predicted income measures.

One concern is that market access shocks from 1900 to 1910 may have induced young sons to be more likely to enter military service, particularly during World War I. This service may have subsequently have had negative effects on economic outcomes such as

[33] Recall that in the main sample, nonwhites were dropped since the linked sampled under-represented nonwhites.

[34] This measure is also very similar in spirit to the income scores used in Collins and Wanamaker (2022), who also impute incomes in U.S. data using occupation, demographics, and location.

income, explaining my results. To test this potential mechanism, in Table A.20 I estimate the effects of 1900–1910 market access on veteran status, which is reported in the 1940 Census. Unfortunately this variable is only available in so-called sample-line men, which is a very small proportion of the full-count Census. The sample size is correspondingly much smaller than in my other specifications, at 36,139 men. Nonetheless, the results for OLS and IV show that market access has little effect on veteran status; in both specifications, the coefficient on market access is insignificant and very small in magnitude. I therefore find evidence that World War I, at least via affecting men in my sample through veteran status, is unlikely to explain why market access can have negative effects on sons' later-life outcomes.

5. Conclusion

In this paper, I find that increased access to markets brought on by expansions in the railroad network from 1900 to 1910 had both short and long run effects on white men exposed to this shock while children. In the short run, these sons were less likely to be literate and more likely to live on farms and be employed in 1910 while their mothers were more likely to have more children. These short-run effects translate into worsened long-run outcomes as well: affected sons from 1900, by 1940, have lower incomes and are less well-educated.

My results imply that economic shocks which are seen as positive and lead to economic growth for affected regions, such as improved access to markets, can have unexpected negative consequences for individuals. More importantly, these consequences may not be immediately obvious for years, highlighting the importance of studies, especially in economic history, which track individuals over time. Notably, many of the outcomes affected, such as education and family composition, are not observable in most linked administrative datasets; this study therefore highlights the value of using historical linked Censuses to examine the long-run impacts of economic shocks. Finally, I show in this paper that work which has shown positive effects of market access from railroads during the American railroad expansions of the 19th and 20th centuries, such as (Donaldson and Hornbeck, 2016), Chan (2022b), and Hornbeck and Rotemberg (2021), may only be seeing part of the complete picture. Overall regional growth may be partly offset by losses, both contemporaneous and long-run, at the individual level. Future research will hopefully shed further light on the costs of changes in market access and other economic shocks on people over time and across generations.

Data availability

The data are available at the following DOI link: https://doi.org/10.3886/E183563V1.

Appendix A

A1. Additional figures and tables

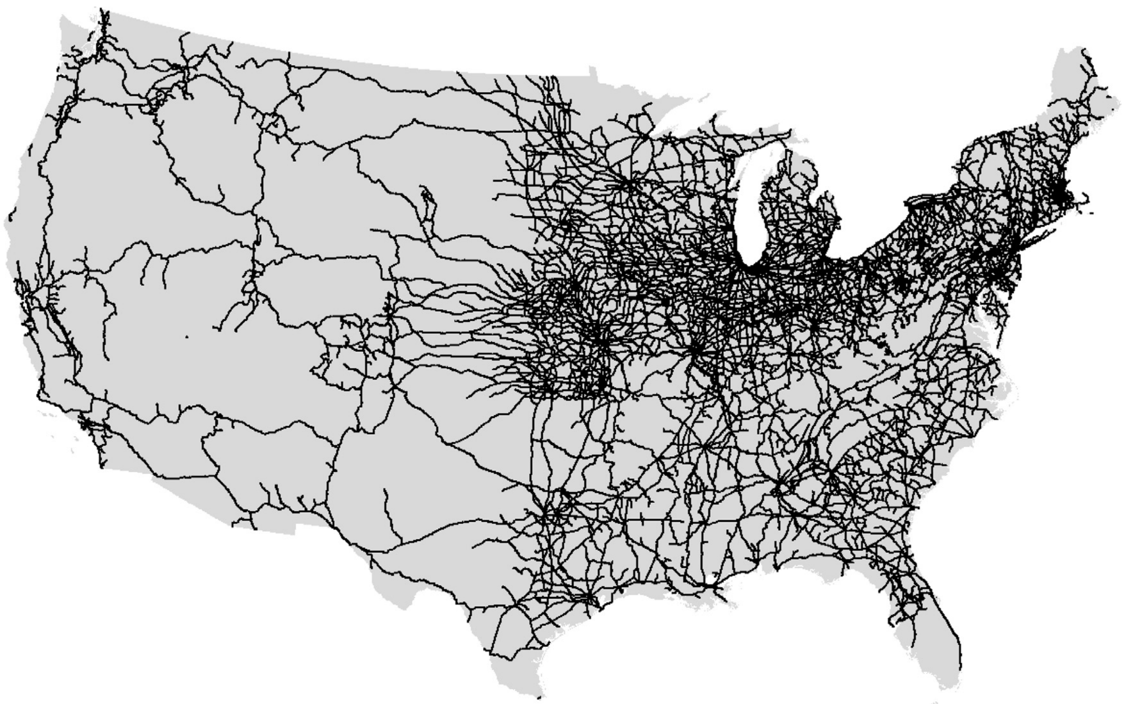

Fig. A.1. Railroad network, 1900. Figure displays the American railroad network in 1900, using only rail segments in operation by 1900.

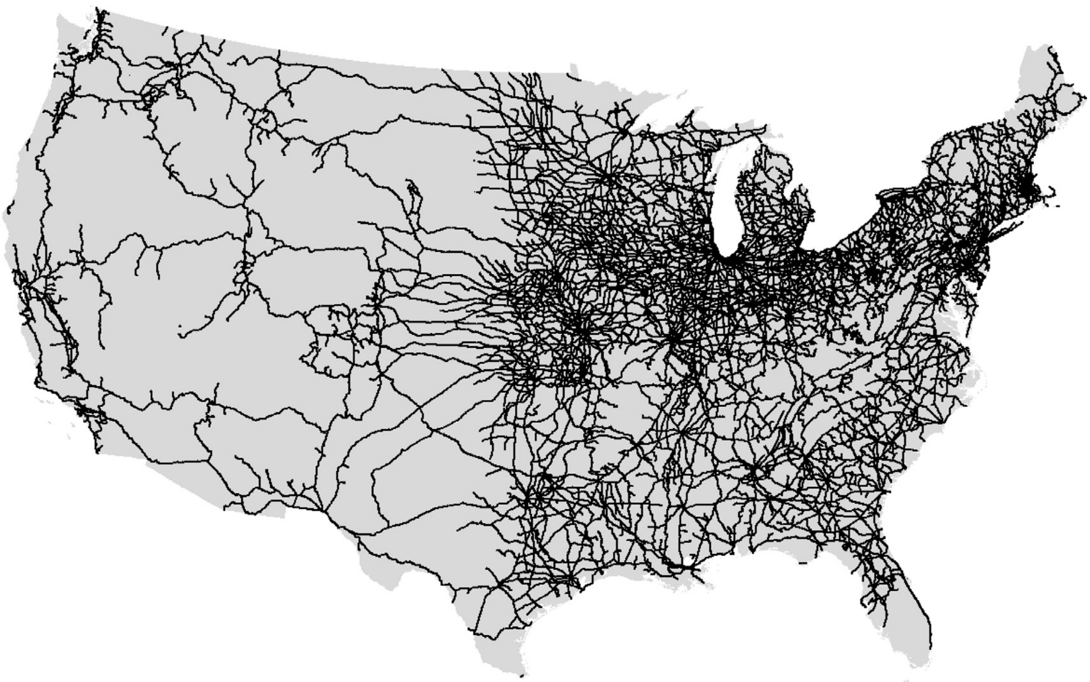

Fig. A.2. Railroad network, 1910. Figure displays the American railroad network in 1910, using only rail segments in operation by 1910.

Fig. A.3. Additions to railroad network, 1900–1910. Figure displays the change in the American railroad network between 1900 and 1910, displaying only rail segments that came into operation between 1900 and 1910.

Table A.1
Sons' outcomes, 1940, first stage.

	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta asinh(MAIV_{00-10,c00})$	1.145***	1.135***	1.135***	1.135***	1.135***	1.135***
	(0.0423)	(0.0364)	(0.0363)	(0.0364)	(0.0364)	(0.0364)
F-test, excluded instrument	733.23	972.89	979.57	972.89	972.89	972.89
Observations	705,646	1,137,094	1,100,050	1,137,094	1,137,094	1,137,094
R-squared	0.868	0.881	0.882	0.881	0.881	0.881
Mean of dep. var.:	0.414	0.414	0.414	0.414	0.414	0.414

Notes: *** denotes 1% significance, ** denotes 5% significance, * denotes 10% significance. All specifications include, as controls, a foreign-born dummy, latitude and longitude cubics for the 1900 county of residence, age and age squared in 1900, and 1900 state of residence fixed effects. The right hand side variable of interest, $\Delta asinh(MA_{00-10,c00})$, the inverse hyperbolic sine transformed change in market access (MA) in an individual's 1900 county of residence, is the dependent variable in all specifications. The main right hand side variable is $\Delta asinh(MAIV_{00-10,c00})$, the instrumental variable. asinh() refers to the inverse hyperbolic sine transformation. For the first stages for the income regressions, the sample includes only those with non-zero incomes and are employed.All columns estimated using OLS. Standard errors are clustered at by 1900 state of residence.

Table A.2
Summary statistics, sample in 1900 (including nonwhites).

Variable	Obs	Mean	Std. Dev.	Min.	Max.
age	1,185,657	3.800	2.529	0	8
non-immigrant status	1,185,657	0.995	0.074	0	1
non-white status	1,185,657	0.041	0.198	0	1
school enrolment, 1900	1,185,657	0.201	0.400	0	1
father in mfg., 1900	1,185,657	0.105	0.306	0	1
father's asinh(occ. score), 1910	1,185,657	3.359	1.068	0	5.075
mother's no. of (other) children, 1910	1,181,660	4.608	2.758	1	25
mother's no. of (other) surviving children, 1910	1,180,658	3.987	2.288	1	25

Notes: Ther rows report the summary statistics for the main variables of interest in 1900. All summary statistics reported only for observations in the estimation sample, and also includes nonwhites. asinh() refers to the inverse hyperbolic sine transformation.

Table A.3
Summary statistics, 1900 census.

Variable	Obs.	Mean	Std. Dev.	Min.	Max.
age	16,335,085	3.982	2.576	0	8
non-immigrant status	16,335,085	0.989	0.103	0	1
non-white status	16,335,085	0.137	0.344	0	1
school enrolment	16,335,085	0.197	0.398	0	1
father in mfg.	14,865,640	0.102	0.303	0	1
father's asinh(occ. score)	14,865,640	3.331	1.076	0	5.075
mother's no. of (other) children	15,457,741	4.860	2.922	1	30
mother's no. of (other) surviving children	15,286,207	4.118	2.354	1	29

Notes: summary statistics for the entirety of the full-count Census, conditional on meeting the demographic sample restrictions (i.e. aged 0–8, male, white), is displayed.

Table A.4
Migration, both periods.

Dep. Var.:	(1) $\mathbb{1}$(mig. counties) OLS	(2) $\mathbb{1}$(mig. states) OLS	(3) $\mathbb{1}$(mig. counties) IV	(4) $\mathbb{1}$(mig. states) IV
$\Delta asinh(MA_{00-10,c00})$	–0.0136 (0.0230)	0.0109 (0.0135)	0.0103 (0.0136)	–0.00719 (0.0126)
Observations	1,137,094	1,137,094	1,137,094	1,137,094
R-squared	0.031	0.047	0.031	0.047
Mean of dep. var.:	0.618	0.313	0.618	0.313

Notes: *** denotes 1% significance, ** denotes 5% significance, * denotes 10% significance. All specifications include, as controls, a foreign-born dummy, latitude and longitude cubics for the 1900 county of residence, age and age squared in 1900, and 1900 state of residence fixed effects. The right hand side variable of interest, $\Delta asinh(MA_{00-10,c00})$, is the inverse hyperbolic sine transformed change in market access (MA) in an individual's 1900 county of residence. asinh() refers to the inverse hyperbolic sine transformation. $\mathbb{1}()$ refers to a dummy variable equal to 1 if the condition within the brackets is satisfied; "migrate" is if the son changed counties between 1900 and 1910 or between 1910 and 1940, and "mig. states" is if the son changed states between 1900 and 1910 or between 1910 and 1940. OLS refers to columns where coefficients are estimated using OLS. IV refers to columns where an instrumental variable regression is employed. Standard errors are clustered at by 1900 state of residence.

Table A.5
Sons' Outcomes, 1940, only stayers.

Dep. Var.:	(1) asinh(income) OLS	(2) asinh(occ. score) OLS	(3) 𝟙(nonwage inc.) OLS	(4) 𝟙(employed) OLS	(5) 𝟙(mfg) OLS	(6) 𝟙(HS) OLS	(7) asinh(income) IV	(8) asinh(occ. score) IV	(9) 𝟙(nonwage inc.) IV	(10) 𝟙(employed) IV	(11) 𝟙(mfg) IV	(12) 𝟙(HS) IV
$\Delta asinh(MA_{00-10,c00})$	–0.0505	0.0280	–0.0138	–0.0189**	–0.00522	–0.00710	–0.401***	–0.0969***	0.0689***	–0.00866	–0.0310**	–0.0538***
	(0.130)	(0.0511)	(0.0342)	(0.00875)	(0.0219)	(0.0205)	(0.0922)	(0.0338)	(0.0252)	(0.00625)	(0.0130)	(0.0176)
Observations	245,233	434,796	423,281	434,796	434,796	434,796	245,233	434,796	423,281	434,796	434,796	434,796
R-squared	0.095	0.024	0.065	0.005	0.039	0.021	0.092	0.023	0.065	0.005	0.039	0.021
Mean of dep. var.:	7.804	3.715	0.396	0.898	0.192	0.285	7.804	3.715	0.396	0.898	0.192	0.285

Notes: *** denotes 1% significance, ** denotes 5% significance, * denotes 10% significance. All specifications include, as controls, a foreign-born dummy, latitude and longitude cubics for the 1900 county of residence, age and age squared in 1900, and 1900 state of residence fixed effects. The right hand side variable of interest, $\Delta asinh(MA_{00-10,c00})$, is the inverse hyperbolic sine transformed change in market access (MA) in an individual's 1900 county of residence. asinh() refers to the inverse hyperbolic sine transformation. 𝟙() refers to a dummy variable equal to 1 if the condition within the brackets is satisfied; in the dummies, "nonwage inc." is whether the son has nonwage income in 1940, "employed" is whether the son is employed in 1940, "mfg" is whether the son is in manufacturing in 1940, and "HS" is whether the son has a high school education (or better) in 1940. OLS refers to columns where coefficients are estimated using OLS. IV refers to columns where an instrumental variable regression is employed. The sample consists only of sons who stayed in their original 1900 county of residence. For the income regressions, the sample includes only those with non-zero incomes and are employed. Means of dependent variables are for the full sample. Standard errors are clustered at by 1900 state of residence.

Table A.6
Sons' outcomes, 1940, removing effect of largest 10 counties in 1900.

Dep. Var.:	(1) asinh(income) OLS	(2) asinh(occ. score) OLS	(3) $\mathbb{1}$(nonwage inc.) OLS	(4) $\mathbb{1}$(employed) OLS	(5) $\mathbb{1}$(mfg) OLS	(6) $\mathbb{1}$(HS) OLS	(7) asinh(income) IV	(8) asinh(occ. score) IV	(9) $\mathbb{1}$(nonwage inc.) IV	(10) $\mathbb{1}$(employed) IV	(11) $\mathbb{1}$(mfg) IV	(12) $\mathbb{1}$(HS) IV
Panel A: Dropping 10 Largest Counties												
$\Delta asinh(MA_{00-10,c00})$	–0.0657	–0.0128	–0.00105	–0.00920**	–0.00112	–0.0269**	–0.183***	–0.0642***	0.0240*	–0.00739**	–0.0121*	–0.0514***
	(0.0434)	(0.0195)	(0.0133)	(0.00376)	(0.00857)	(0.0120)	(0.0426)	(0.0180)	(0.0136)	(0.00368)	(0.00659)	(0.0133)
Observations	619,861	1,017,922	985,196	1,017,922	1,017,922	1,017,922	619,861	1,017,922	985,196	1,017,922	1,017,922	1,017,922
R-squared	0.043	0.010	0.025	0.002	0.025	0.014	0.042	0.010	0.024	0.002	0.025	0.014
Panel B: Dropping 10 Largest Counties and Controlling for Distance to Nearest Top 10 County												
$\Delta asinh(MA_{00-10,c00})$	–0.0650	–0.0125	–0.00114	–0.00916**	–0.00101	–0.0268**	–0.183***	–0.0640***	0.0239*	–0.00739**	–0.0120*	–0.0514***
	(0.0433)	(0.0195)	(0.0133)	(0.00375)	(0.00857)	(0.0120)	(0.0426)	(0.0180)	(0.0136)	(0.00367)	(0.00658)	(0.0133)
Observations	618,144	1,015,415	982,774	1,015,415	1,015,415	1,015,415	618,144	1,015,415	982,774	1,015,415	1,015,415	1,015,415
R-squared	0.043	0.010	0.025	0.002	0.025	0.014	0.042	0.010	0.024	0.002	0.025	0.014
Mean of dep. var.:	7.804	3.715	0.396	0.898	0.192	0.285	7.804	3.715	0.396	0.898	0.192	0.285

Notes: *** denotes 1% significance, ** denotes 5% significance, * denotes 10% significance. All specifications include, as controls, a foreign-born dummy, latitude and longitude cubics for the 1900 county of residence, age and age squared in 1900, and 1900 state of residence fixed effects. The right hand side variable of interest, $\Delta asinh(MA_{00-10,c00})$, is the inverse hyperbolic sine transformed change in market access (MA) in an individual's 1900 county of residence. asinh() refers to the inverse hyperbolic sine transformation. $\mathbb{1}()$ refers to a dummy variable equal to 1 if the condition within the brackets is satisfied; in the dummies, "nonwage inc." is whether the son has nonwage income in 1940, "employed" is whether the son is employed in 1940, "mfg" is whether the son is in manufacturing in 1940, and "HS" is whether the son has a high school education (or better) in 1940. OLS refers to columns where coefficients are estimated using OLS. IV refers to columns where an instrumental variable regression is employed. The sample drops the 10 largest counties, by 1900 population. Panel B includes, as an additional control for all regressions, the inverse hyperbolic sine transformed transport cost in 1900 to the nearest top-10 largest county (by population, in 1900), according to the Donaldson-Hornbeck (2016) transport cost database. For the income regressions, the sample includes only those with non-zero incomes and are employed. Means of dependent variables are for the full sample. Standard errors are clustered at by 1900 state of residence.

Table A.7
Sons' outcomes, 1940, non-cotton counties.

Dep. Var.:	(1) asinh(income) OLS	(2) asinh(occ. score) OLS	(3) 𝟙(nonwage inc.) OLS	(4) 𝟙(employed) OLS	(5) 𝟙(mfg) OLS	(6) 𝟙(HS) OLS	(7) asinh(income) IV	(8) asinh(occ. score) IV	(9) 𝟙(nonwage inc.) IV	(10) 𝟙(employed) IV	(11) 𝟙(mfg) IV	(12) 𝟙(HS) IV
$\Delta asinh(MA_{00-10,c00})$	0.00750	0.0174	–0.0125	–0.0101**	–7.54e-05	–0.0173	–0.135**	–0.0431**	0.0169	–0.00673	–0.00845	–0.0540***
	(0.0619)	(0.0276)	(0.0158)	(0.00484)	(0.00983)	(0.0134)	(0.0547)	(0.0216)	(0.0143)	(0.00468)	(0.00895)	(0.0142)
Observations	608,125	960,561	929,474	960,561	960,561	960,561	608,125	960,561	929,474	960,561	960,561	960,561
R-squared	0.037	0.008	0.029	0.002	0.022	0.015	0.037	0.008	0.029	0.002	0.022	0.015
Mean of dep. var.:	7.804	3.715	0.396	0.898	0.192	0.285	7.804	3.715	0.396	0.898	0.192	0.285

Notes: *** denotes 1% significance, ** denotes 5% significance, * denotes 10% significance. All specifications include, as controls, a foreign-born dummy, latitude and longitude cubics for the 1900 county of residence, age and age squared in 1900, and 1900 state of residence fixed effects. The right hand side variable of interest, $\Delta asinh(MA_{00-10,c00})$, is the inverse hyperbolic sine transformed change in market access (MA) in an individual's 1900 county of residence. asinh() refers to the inverse hyperbolic sine transformation. 𝟙() refers to a dummy variable equal to 1 if the condition within the brackets is satisfied; in the dummies, "nonwage inc." is whether the son has nonwage income in 1940, "employed" is whether the son is employed in 1940, "mfg" is whether the son is in manufacturing in 1940, and "HS" is whether the son has a high school education (or better) in 1940. OLS refers to columns where coefficients are estimated using OLS. IV refers to columns where an instrumental variable regression is employed. The sample drops any counties where positive cotton acreage was reported in the 1900 Agricultural Census. For the income regressions, the sample includes only those with non-zero incomes and are employed. Means of dependent variables are for the full sample. Standard errors are clustered at by 1900 state of residence.

Table A.8
Sons' outcomes, 1940, 1900–1910 movers.

Dep. Var.:	(1) asinh(income) OLS	(2) asinh(occ. score) OLS	(3) 𝟙(nonwage inc.) OLS	(4) 𝟙(employed) OLS	(5) 𝟙(mfg) OLS	(6) 𝟙(HS) OLS	(7) asinh(income) IV	(8) asinh(occ. score) IV	(9) 𝟙(nonwage inc.) IV	(10) 𝟙(employed) IV	(11) 𝟙(mfg) IV	(12) 𝟙(HS) IV
$\Delta asinh(MA_{00-10,c00})$	–0.0114	0.000335	–0.000778	–0.00402	–0.00295	–0.0136	–0.105***	–0.0422***	0.00930	–0.00583	–0.00614	–0.0379***
	(0.0333)	(0.0168)	(0.00628)	(0.00418)	(0.00595)	(0.0123)	(0.0307)	(0.0161)	(0.00723)	(0.00501)	(0.00531)	(0.0142)
Observations	179,078	281,931	272,872	281,931	281,931	281,931	179,078	281,931	272,872	281,931	281,931	281,931
R-squared	0.044	0.010	0.019	0.002	0.022	0.014	0.044	0.010	0.019	0.002	0.022	0.014
Mean of dep. var.:	7.804	3.715	0.396	0.898	0.192	0.285	7.804	3.715	0.396	0.898	0.192	0.285

Notes: *** denotes 1% significance, ** denotes 5% significance, * denotes 10% significance. All specifications include, as controls, a foreign-born dummy, latitude and longitude cubics for the 1900 county of residence, age and age squared in 1900, and 1900 state of residence fixed effects. The right hand side variable of interest, $\Delta asinh(MA_{00-10,c00})$, is the inverse hyperbolic sine transformed change in market access (MA) in an individual's 1900 county of residence. asinh() refers to the inverse hyperbolic sine transformation. 𝟙() refers to a dummy variable equal to 1 if the condition within the brackets is satisfied; in the dummies, "nonwage inc." is whether the son has nonwage income in 1940, "employed" is whether the son is employed in 1940, "mfg" is whether the son is in manufacturing in 1940, and "HS" is whether the son has a high school education (or better) in 1940. OLS refers to columns where coefficients are estimated using OLS. IV refers to columns where an instrumental variable regression is employed. The sample includes only sons who migrated counties between 1900 and 1910. For the income regressions, the sample includes only those with non-zero incomes and are employed. Means of dependent variables are for the full sample. Standard errors are clustered at by 1900 state of residence.

Table A.9
Sons' outcomes, 1940, by likelihood of electrification.

Dep. Var.:	(1) asinh(income) OLS	(2) asinh(occ. score) OLS	(3) 𝟙(nonwage inc.) OLS	(4) 𝟙(employed) OLS	(5) 𝟙(mfg) OLS	(6) 𝟙(HS) OLS	(7) asinh(income) IV	(8) asinh(occ. score) IV	(9) 𝟙(nonwage inc.) IV	(10) 𝟙(employed) IV	(11) 𝟙(mfg) IV	(12) 𝟙(HS) IV
Panel A: Excluding Northeast and California												
$\Delta asinh(MA_{00-10,c00})$	–0.113**	–0.0353**	0.00997	–0.00746**	–0.00285	–0.0345***	–0.188***	–0.0674***	0.0283*	–0.00674*	–0.0124*	–0.0489***
	(0.0420)	(0.0171)	(0.0129)	(0.00358)	(0.00896)	(0.0118)	(0.0464)	(0.0192)	(0.0148)	(0.00367)	(0.00699)	(0.0134)
Observations	487,869	826,376	800,876	826,376	826,376	826,376	487,869	826,376	800,876	826,376	826,376	826,376
R-squared	0.038	0.008	0.018	0.001	0.024	0.011	0.038	0.008	0.018	0.001	0.024	0.011
Panel B: Northeast + California												
$\Delta asinh(MA_{00-10,c00})$	0.0922	0.0735**	–0.0287	–0.0189***	–0.0320	0.0241	–0.109	–0.0290	–0.0234	–0.0193**	–0.0331	–0.0605
	(0.0650)	(0.0276)	(0.0267)	(0.00578)	(0.0375)	(0.0168)	(0.154)	(0.0476)	(0.0299)	(0.00845)	(0.0383)	(0.0424)
Observations	217,777	310,718	299,174	310,718	310,718	310,718	217,777	310,718	299,174	310,718	310,718	310,718
R-squared	0.033	0.003	0.013	0.002	0.007	0.018	0.032	0.003	0.013	0.002	0.007	0.018
Mean of dep. var.:	7.804	3.715	0.396	0.898	0.192	0.285	7.804	3.715	0.396	0.898	0.192	0.285

Notes: *** denotes 1% significance, ** denotes 5% significance, * denotes 10% significance. All specifications include, as controls, a foreign-born dummy, latitude and longitude cubics for the 1900 county of residence, age and age squared in 1900, and 1900 state of residence fixed effects. The right hand side variable of interest, $\Delta asinh(MA_{00-10,c00})$, is the inverse hyperbolic sine transformed change in market access (MA) in an individual's 1900 county of residence. asinh() refers to the inverse hyperbolic sine transformation. 𝟙() refers to a dummy variable equal to 1 if the condition within the brackets is satisfied; in the dummies, "nonwage inc." is whether the son has nonwage income in 1940, "employed" is whether the son is employed in 1940, "mfg" is whether the son is in manufacturing in 1940, and "HS" is whether the son has a high school education (or better) in 1940. OLS refers to columns where coefficients are estimated using OLS. IV refers to columns where an instrumental variable regression is employed. For the income regressions, the sample includes only those with non-zero incomes and are employed. Standard errors are clustered at by 1900 state of residence.

Table A.10
Sons' Outcomes, 1940, Alternate 1910–1940 Linking Method.

Dep. Var.:	(1) asinh(income) OLS	(2) asinh(occ. score) OLS	(3) 𝟙(nonwage inc.) OLS	(4) 𝟙(employed) OLS	(5) 𝟙(mfg) OLS	(6) 𝟙(HS) OLS	(7) asinh(income) IV	(8) asinh(occ. score) IV	(9) 𝟙(nonwage inc.) IV	(10) 𝟙(employed) IV	(11) 𝟙(mfg) IV	(12) 𝟙(HS) IV
$\Delta asinh(MA_{00-10,c00})$	–0.0329	0.00438	-0.00714	–0.0107***	–0.00621	–0.0142	–0.186***	–0.0629***	0.0218*	–0.00929***	–0.0126*	–0.0497***
	(0.0592)	(0.0281)	(0.0133)	(0.00328)	(0.00911)	(0.0132)	(0.0462)	(0.0191)	(0.0127)	(0.00344)	(0.00762)	(0.0141)
Observations	667,605	1,069,062	1,034,246	1,069,062	1,069,062	1,069,062	667,605	1,069,062	1,034,246	1,069,062	1,069,062	1,069,062
R-squared	0.049	0.011	0.029	0.002	0.024	0.015	0.049	0.010	0.029	0.002	0.024	0.015
Mean of dep. var.:	7.804	3.715	0.396	0.898	0.192	0.285	7.804	3.715	0.396	0.898	0.192	0.285

Notes: *** denotes 1% significance, ** denotes 5% significance, * denotes 10% significance. All specifications include, as controls, a foreign-born dummy, latitude and longitude cubics for the 1900 county of residence, age and age squared in 1900, and 1900 state of residence fixed effects. The right hand side variable of interest, $\Delta asinh(MA_{00-10,c00})$, is the inverse hyperbolic sine transformed change in market access (MA) in an individual's 1900 county of residence. asinh() refers to the inverse hyperbolic sine transformation. 𝟙() refers to a dummy variable equal to 1 if the condition within the brackets is satisfied; in the dummies, "nonwage inc." is whether the son has nonwage income in 1940, "employed" is whether the son is employed in 1940, "mfg" is whether the son is in manufacturing in 1940, and "HS" is whether the son has a high school education (or better) in 1940. OLS refers to columns where coefficients are estimated using OLS. IV refers to columns where an instrumental variable regression is employed. The sample, instead of the linking procedure used in the main sample, instead uses an alternate linking method for the 1910–1940 Censuses (the ABE method using exact name matches). For the income regressions, the sample includes only those with non-zero incomes and are employed. Standard errors are clustered at by 1900 state of residence. Means of dependent variables are for the original full sample.

Table A.11
Sons' outcomes, 1940, controlling for 1870–1900 shocks.

Dep. Var.:	(1) asinh(income) IV	(2) asinh(occ. score) IV	(3) $\mathbb{1}$(nonwage inc.) IV	(4) $\mathbb{1}$(employed) IV	(5) $\mathbb{1}$(mfg) IV	(6) $\mathbb{1}$(HS) IV
$\Delta asinh(MA_{00-10,c00})$	–0.176***	–0.0625***	0.0221*	–0.00725**	–0.0121*	–0.0506***
	(0.0433)	(0.0179)	(0.0133)	(0.00361)	(0.00687)	(0.0131)
$\Delta asinh(MA_{70-00,c00})$	0.000755	0.00217	0.000278	–0.000777	–0.000189	0.000981
	(0.0156)	(0.00741)	(0.00387)	(0.000819)	(0.00233)	(0.00419)
Observations	705,646	1,137,094	1,100,050	1,137,094	1,137,094	1,137,094
R-squared	0.049	0.011	0.028	0.002	0.024	0.015
Mean of dep. var.:	7.804	3.715	0.396	0.898	0.192	0.285

Notes: *** denotes 1% significance, ** denotes 5% significance, * denotes 10% significance. All specifications include, as controls, a foreign-born dummy, latitude and longitude cubics for the 1900 county of residence, age and age squared in 1900, and 1900 state of residence fixed effects. The right hand side variable of interest, $\Delta asinh(MA_{00-10,c00})$, is the inverse hyperbolic sine transformed change in market access (MA) in an individual's 1900 county of residence. asinh() refers to the inverse hyperbolic sine transformation. $\mathbb{1}()$ refers to a dummy variable equal to 1 if the condition within the brackets is satisfied; in the dummies, "nonwage inc." is whether the son has nonwage income in 1940, "employed" is whether the son is employed in 1940, "mfg" is whether the son is in manufacturing in 1940, and "HS" is whether the son has a high school education (or better) in 1940. OLS refers to columns where coefficients are estimated using OLS. IV refers to columns where an instrumental variable regression is employed. In addition to the main market access shock of interest from 1900 to 1910, I additionally control for $\Delta asinh(MA_{70,00,c00}$, the market access shock from 1870 to 1900 in the 1900 county of residence. For the income regressions, the sample includes only those with non-zero incomes and are employed. Standard errors are clustered at by 1900 state of residence.

Table A.12
Sons' outcomes, 1940, logged market access.

Dep. Var.:	(1) asinh(income) OLS	(2) asinh(occ. score) OLS	(3) $\mathbb{1}$(nonwage inc.) OLS	(4) $\mathbb{1}$(employed) OLS	(5) $\mathbb{1}$(mfg) OLS	(6) $\mathbb{1}$(HS) OLS	(7) asinh(income) IV	(8) asinh(occ. score) IV	(9) $\mathbb{1}$(nonwage inc.) IV	(10) $\mathbb{1}$(employed) IV	(11) $\mathbb{1}$(mfg) IV	(12) $\mathbb{1}$(HS) IV
$\Delta\ln(MA_{00-10,c00})$	–0.0318	–0.00139	–0.00630	–0.0102***	–0.00395	–0.0189	–0.175***	–0.0615***	0.0223*	–0.00762**	–0.0122*	–0.0502***
	(0.0554)	(0.0241)	(0.0135)	(0.00376)	(0.00819)	(0.0122)	(0.0454)	(0.0189)	(0.0134)	(0.00358)	(0.00688)	(0.0133)
Observations	705,646	1,137,094	1,100,050	1,137,094	1,137,094	1,137,094	705,646	1,137,094	1,100,050	1,137,094	1,137,094	1,137,094
R-squared	0.049	0.011	0.028	0.002	0.024	0.015	0.049	0.011	0.028	0.002	0.024	0.015
Mean of dep. var.:	7.804	3.715	0.396	0.898	0.192	0.285	7.804	3.715	0.396	0.898	0.192	0.285

Notes: *** denotes 1% significance, ** denotes 5% significance, * denotes 10% significance. All specifications include, as controls, a foreign-born dummy, latitude and longitude cubics for the 1900 county of residence, age and age squared in 1900, and 1900 state of residence fixed effects. The right hand side variable of interest, $\Delta\ln(MA_{00-10,c00})$, is the natural log transformed change in market access (MA) in an individual's 1900 county of residence. asinh() refers to the inverse hyperbolic sine transformation. $\mathbb{1}()$ refers to a dummy variable equal to 1 if the condition within the brackets is satisfied; in the dummies, "nonwage inc." is whether the son has nonwage income in 1940, "employed" is whether the son is employed in 1940, "mfg" is whether the son is in manufacturing in 1940, and "HS" is whether the son has a high school education (or better) in 1940. OLS refers to columns where coefficients are estimated using OLS. IV refers to columns where an instrumental variable regression is employed. For the income regressions, the sample includes only those with non-zero incomes and are employed. Standard errors are clustered at by 1900 state of residence.

Table A.13
Sons' outcomes, 1940, including nonwhites.

Dep. Var.:	(1) asinh(income) OLS	(2) asinh(occ. score) OLS	(3) $\mathbb{1}$(nonwage inc.) OLS	(4) $\mathbb{1}$(employed) OLS	(5) $\mathbb{1}$(mfg) OLS	(6) $\mathbb{1}$(HS) OLS	(7) asinh(income) IV	(8) asinh(occ. score) IV	(9) $\mathbb{1}$(nonwage inc.) IV	(10) $\mathbb{1}$(employed) IV	(11) $\mathbb{1}$(mfg) IV	(12) $\mathbb{1}$(HS) IV
$\Delta\ln(MA_{00-10,c00})$	–0.0355	–0.00278	–0.00408	–0.00974**	–0.00462	–0.0181	–0.178***	–0.0618***	0.0249*	–0.00677*	–0.0128*	–0.0493***
	(0.0546)	(0.0235)	(0.0133)	(0.00369)	(0.00798)	(0.0121)	(0.0445)	(0.0184)	(0.0134)	(0.00353)	(0.00669)	(0.0132)
Observations	734,593	1,185,657	1,146,423	1,185,657	1,185,657	1,185,657	734,593	1,185,657	1,146,423	1,185,657	1,185,657	1,185,657
R-squared	0.073	0.016	0.028	0.003	0.023	0.020	0.073	0.015	0.028	0.003	0.023	0.020
Mean of dep. var.:	7.804	3.715	0.396	0.898	0.192	0.285	7.804	3.715	0.396	0.898	0.192	0.285

Notes: *** denotes 1% significance, ** denotes 5% significance, * denotes 10% significance. All specifications include, as controls, a foreign-born dummy, a nonwhite dummy, latitude and longitude cubics for the 1900 county of residence, age and age squared in 1900, and 1900 state of residence fixed effects. The right hand side variable of interest, $\Delta\ln(MA_{00-10,c00})$, is the inverse hyperbolic sine transformed change in market access (MA) in an individual's 1900 county of residence. asinh() refers to the inverse hyperbolic sine transformation. $\mathbb{1}()$ refers to a dummy variable equal to 1 if the condition within the brackets is satisfied; in the dummies, "nonwage inc." is whether the son has nonwage income in 1940, "employed" is whether the son is employed in 1940, "mfg" is whether the son is in manufacturing in 1940, and "HS" is whether the son has a high school education (or better) in 1940. OLS refers to columns where coefficients are estimated using OLS. IV refers to columns where an instrumental variable regression is employed. For the income regressions, the sample includes only those with non-zero incomes and are employed. Means of dependent variables are for the original, main sample without nonwhites. Standard errors are clustered at by 1900 state of residence.

Table A.14
Sons' outcomes, 1940, conley standard errors.

Dep. Var.:	(1) asinh(income) OLS	(2) asinh(occ. score) OLS	(3) 𝟙(nonwage inc.) OLS	(4) 𝟙(employed) OLS	(5) 𝟙(mfg) OLS	(6) 𝟙(HS) OLS	(7) asinh(income) IV	(8) asinh(occ. score) IV	(9) 𝟙(nonwage inc.) IV	(10) 𝟙(employed) IV	(11) 𝟙(mfg) IV	(12) 𝟙(HS) IV
$\Delta\ln(MA_{00-10,c00})$	–0.0318	–0.00139	–0.00630	–0.0102***	–0.00395	–0.0189	–0.175***	–0.0615***	0.0223*	–0.00762**	–0.0122*	–0.0502***
	(0.037)	(0.016)	(0.010)	(0.0033)	(0.0078)	(0.010)	(0.042)	(0.017)	(0.011)	(0.0036)	(0.0064)	(0.011)
Observations	705,646	1,137,094	1,100,050	1,137,094	1,137,094	1,137,094	705,646	1,137,094	1,100,050	1,137,094	1,137,094	1,137,094
R-squared	0.049	0.011	0.028	0.002	0.024	0.015	0.049	0.011	0.028	0.002	0.024	0.015

Notes: *** denotes 1% significance, ** denotes 5% significance, * denotes 10% significance. All specifications include, as controls, a foreign-born dummy, latitude and longitude cubics for the 1900 county of residence, age and age squared in 1900, and 1900 state of residence fixed effects. The right hand side variable of interest, $\Delta\ln(MA_{00-10,c00})$, is the inverse hyperbolic sine transformed change in market access (MA) in an individual's 1900 county of residence. asinh() refers to the inverse hyperbolic sine transformation. 𝟙() refers to a dummy variable equal to 1 if the condition within the brackets is satisfied; in the dummies, "nonwage inc." is whether the son has nonwage income in 1940, "employed" is whether the son is employed in 1940, "mfg" is whether the son is in manufacturing in 1940, and "HS" is whether the son has a high school education (or better) in 1940. OLS refers to columns where coefficients are estimated using OLS. IV refers to columns where an instrumental variable regression is employed. For the income regressions, the sample includes only those with non-zero incomes and are employed. Standard errors in parentheses are Conley standard errors allowing for spatial correlation within 100 km.

Table A.15
1900 county characteristics and market access.

Dep. Var.:	(1) sh non-immig.	(2) avg age in sample	(3) asinh(pop)	(4) sh urban	(5) asinh(land value pc)
$\Delta\ln(MA_{00-10,c00})$	-0.00359 (0.00294)	-0.0620*** (0.0150)	-0.919*** (0.245)	-0.200*** (0.0596)	0.103 (0.133)
Observations	2734	2734	2734	2734	2734
R-squared	0.472	0.090	0.571	0.473	0.595
Mean of Dep. Var.:	0.997	3.780	10.302	0.135	5.796

Notes: *** denotes 1% significance, ** denotes 5% significance, * denotes 10% significance. Each observation represents a 1900 county of residence. Counties are weighted by the number of individuals from my individual-level sample. Other than the exception (mentioned later in this note) all specifications include, as controls, foreign-born share, latitude and longitude cubics for the 1900 county of residence, average age and average age squared in 1900, and state fixed effects. The right hand side variable of interest, $\Delta\ln(MA_{00-10,c00})$, is the inverse hyperbolic sine transformed change in market access (MA). sh non-immig. refers to the share of sons in my sample that were non-immigrants in 1900. avg age in sample refers to the average age of sons in 1900 in that county. asinh(pop) refers to the asinh-transformed county-level population of that county in 1900. sh urban refers to the share of population that is urban in 1900. asinh(land value pc) refers to the agricultural land and building value per capita in 1900, transformed by the inverse hyperbolic sine. All specifications estimated using IV. Standard errors are clustered at by 1900 state of residence.

Table A.16
Sons' outcomes, 1940, controls for initial 1900 county characteristics.

Dep. Var.:	(1) asinh(income) OLS	(2) asinh(occ. score) OLS	(3) 𝟙(nonwage inc.) OLS	(4) 𝟙(employed) OLS	(5) 𝟙(mfg) OLS	(6) 𝟙(HS) OLS	(7) asinh(income) IV	(8) asinh(occ. score) IV	(9) 𝟙(nonwage inc.) IV	(10) 𝟙(employed) IV	(11) 𝟙(mfg) IV	(12) 𝟙(HS) IV
$\Delta\ln(MA_{00-10,c00})$	–0.0386*	–0.00992	0.0132	–0.00387	–0.00619	–0.0106	–0.0657***	–0.0210**	0.00795	–0.00619*	0.00422	–0.0227**
	(0.0194)	(0.00823)	(0.00899)	(0.00301)	(0.00804)	(0.00973)	(0.0243)					
asinh(pop)	0.0363***	0.00474	0.00631***	0.000938	–0.00131	0.00689	0.0362***	0.00463	0.00626***	0.000916	–0.00121	0.00677
	(0.00825)	(0.00320)	(0.00191)	(0.000799)	(0.00295)	(0.00415)	(0.00833)	(0.00321)	(0.00186)	(0.000794)	(0.00291)	(0.00415)
sh urban	0.414***	0.223***	-0.126***	3.04e-06	0.103***	0.125***	0.412***	0.222***	-0.126***	-0.000149	0.103***	0.124***
	(0.0377)	(0.0164)	(0.00976)	(0.00280)	(0.00954)	(0.0178)	(0.0379)	(0.0167)	(0.00974)	(0.00277)	(0.00970)	(0.0181)
asinh(land value pc)	0.0341***	0.00484	0.0177***	0.00789***	0.00327	0.0197***	0.0328***	0.00428	0.0174***	0.00778***	0.00380	0.0191***
	(0.00842)	(0.00322)	(0.00316)	(0.00164)	(0.00329)	(0.00417)	(0.00834)	(0.00309)	(0.00325)	(0.00163)	(0.00335)	(0.00414)
Observations	705,646	1,137,094	1,100,050	1,137,094	1,137,094	1,137,094	705,646	1,137,094	1,100,050	1,137,094	1,137,094	1,137,094
R-squared	0.071	0.015	0.035	0.003	0.028	0.020	0.071	0.015	0.035	0.003	0.028	0.020
Mean of dep. var.:	7.804	3.715	0.396	0.898	0.192	0.285	7.804	3.715	0.396	0.898	0.192	0.285

Notes: *** denotes 1% significance, ** denotes 5% significance, * denotes 10% significance. All specifications include, as controls, a foreign-born dummy, latitude and longitude cubics for the 1900 county of residence, age and age squared in 1900, and 1900 state of residence fixed effects. Initial condition control variables include asinh(population), the share of urban population, and the inverse hyperbolic sine of agricultural land and building value per capita. The right hand side variable of interest, $\Delta\ln(MA_{00-10,c00})$, is the inverse hyperbolic sine transformed change in market access (MA) in an individual's 1900 county of residence. asinh() refers to the inverse hyperbolic sine transformation. 𝟙() refers to a dummy variable equal to 1 if the condition within the brackets is satisfied; in the dummies, "nonwage inc." is whether the son has nonwage income in 1940, "employed" is whether the son is employed in 1940, "mfg" is whether the son is in manufacturing in 1940, and "HS" is whether the son has a high school education (or better) in 1940. OLS refers to columns where coefficients are estimated using OLS. IV refers to columns where an instrumental variable regression is employed. For the income regressions, the sample includes only those with non-zero incomes and are employed. Standard errors are clustered by 1900 state of residence.

Table A.17
Sons' outcomes, 1940, county level.

Dep. Var.:	(1) asinh(avg. income) OLS	(2) asinh(avg. occ. score) OLS	(3) sh. nonwage inc. OLS	(4) sh. employed OLS	(5) sh. mfg OLS	(6) sh. HS OLS	(7) asinh(avg. income) IV	(8) asinh(avg. occ. score) IV	(9) sh. nonwage inc. IV	(10) sh. employed IV	(11) sh. mfg IV	(12) sh. HS IV
$\Delta\ln(MA_{00-10,c00})$	–0.0581*	–0.0194	0.0124	–0.00620**	–0.0123	–0.0216*	–0.143***	-0.0662***	0.0296*	–0.00535	–0.0146*	–0.0497***
	(0.0318)	(0.0161)	(0.0131)	(0.00290)	(0.0102)	(0.0118)	(0.0384)	(0.0176)	(0.0156)	(0.00347)	(0.00753)	(0.0141)
Observations	2733	2733	2733	2733	2733	2733	2734	2734	2734	2734	2734	2734
R-squared	0.560	0.555	0.646	0.307	0.676	0.480	0.561	0.553	0.646	0.307	0.676	0.482
Mean of dep. var.:	7.934	3.865	0.445	0.900	0.153	0.264	7.934	3.865	0.445	0.900	0.153	0.264

Notes: *** denotes 1% significance, ** denotes 5% significance, * denotes 10% significance. Each observation represents a 1900 county of residence. Counties are weighted by the number of individuals from my individual-level sample. All specifications include, as controls, foreign-born share, latitude and longitude cubics for the 1900 county of residence, average age and average age squared in 1900, and state fixed effects. The right hand side variable of interest, $\Delta\ln(MA_{00-10,c00})$, is the inverse hyperbolic sine transformed change in market access (MA). "sh. nonwage inc." is the share of people with nonwage income in 1940, "sh. employed" is the share employed in 1940, "sh. mfg" is the share employed in manufacturing in 1940, and "sh. HS" the share with a high school education (or better) in 1940. OLS refers to columns where coefficients are estimated using OLS. IV refers to columns where an instrumental variable regression is employed. Standard errors are clustered at by 1900 state of residence.

Table A.18
Sons' outcomes, 1940, showing controls' coefficients.

Dep. Var.:	(1) asinh(income) OLS	(2) asinh(occ. score) OLS	(3) 𝟙(nonwage inc.) OLS	(4) 𝟙(employed) OLS	(5) 𝟙(mfg) OLS	(6) 𝟙(HS) OLS	(7) asinh(income) IV	(8) asinh(occ. score) IV	(9) 𝟙(nonwage inc.) IV	(10) 𝟙(employed) IV	(11) 𝟙(mfg) IV	(12) 𝟙(HS) IV
$\Delta asinh(MA_{00-10,c00})$	–0.0318	–0.00139	–0.00630	–0.0102***	–0.00395	–0.0189	–0.175***	–0.0615***	0.0223*	–0.00762**	–0.0122*	–0.0502***
	(0.0554)	(0.0241)	(0.0135)	(0.00376)	(0.00819)	(0.0122)	(0.0454)	(0.0189)	(0.0134)	(0.00358)	(0.00688)	(0.0133)
age_{00}	0.0128***	–0.00590***	0.0153***	–0.00252***	–0.00486***	–0.000798	0.0128***	–0.00590***	0.0153***	–0.00252***	–0.00486***	–0.000800
	(0.00128)	(0.00106)	(0.000717)	(0.000446)	(0.000696)	(0.000571)	(0.00127)	(0.00105)	(0.000709)	(0.000441)	(0.000688)	(0.000566)
age^2_{00}	–0.00141***	–0.000277**	–0.000648***	–1.05e-06	0.000153**	–0.000260***	–0.00142***	–0.000278**	–0.000647***	–9.79e-07	0.000153**	–0.000261***
	(0.000151)	(0.000118)	(7.65e-05)	(5.67e-05)	(7.24e-05)	(7.16e-05)	(0.000148)	(0.000117)	(7.57e-05)	(5.60e-05)	(7.16e-05)	(7.08e-05)
non-immig.	0.0182	–0.0243*	0.0273**	0.00447	–0.0623***	0.0936***	0.0135	–0.0264**	0.0283***	0.00456	–0.0626***	0.0925***
	(0.0289)	(0.0126)	(0.0107)	(0.00339)	(0.00505)	(0.0171)	(0.0282)	(0.0128)	(0.0102)	(0.00336)	(0.00496)	(0.0170)
Observations	705,646	1,137,094	1,100,050	1,137,094	1,137,094	1,137,094	705,646	1,137,094	1,100,050	1,137,094	1,137,094	1,137,094
R-squared	0.049	0.011	0.028	0.002	0.024	0.015	0.049	0.011	0.028	0.002	0.024	0.015
Mean of dep. var.:	7.804	3.715	0.396	0.898	0.192	0.285	7.804	3.715	0.396	0.898	0.192	0.285

Notes: *** denotes 1% significance, ** denotes 5% significance, * denotes 10% significance. All specifications include, as controls, a foreign-born dummy, latitude and longitude cubics for the 1900 county of residence, age and age squared in 1900, and 1900 state of residence fixed effects. The right hand side variable of interest, $\Delta asinh(MA_{00-10,c00})$, is the inverse hyperbolic sine transformed change in market access (MA) in an individual's 1900 county of residence. asinh() refers to the inverse hyperbolic sine transformation. 𝟙() refers to a dummy variable equal to 1 if the condition within the brackets is satisfied; in the dummies, "nonwage inc." is whether the son has nonwage income in 1940, "employed" is whether the son is employed in 1940, "mfg" is whether the son is in manufacturing in 1940, and "HS" is whether the son has a high school education (or better) in 1940. OLS refers to columns where coefficients are estimated using OLS. IV refers to columns where an instrumental variable regression is employed. For the income regressions, the sample includes only those with non-zero incomes and are employed. Standard errors are clustered at by 1900 state of residence.

Table A.19
Occ. score robustness checks.

Dep. Var.: OLS	(1) Δ asinh(father's occ. score) OLS	(2) asinh(occ. score$_{40}$) OLS	(3) asinh(LIDO$_{40}$) IV	(4) Δasinh(father's occ. score) IV	(5) asinh(occ. score$_4$0) IV	(6) asinh(LIDO$_{40}$)
$\Delta asinh(MA_{00-10,c00})$	0.00221	0.00455	–0.00213	0.00707	–0.0633***	–0.0627***
	(0.00558)	(0.0259)	(0.0220)	(0.00511)	(0.0168)	(0.0189)
Observations	997,029	1,087,386	996,854	997,029	1,087,386	996,854
R-squared	0.001	0.042	0.132	0.001	0.041	0.131
Mean of dep. var.:	0.164	3.715	3.855	0.164	3.715	3.855

Notes: *** denotes 1% significance, ** denotes 5% significance, * denotes 10% significance. All specifications include, as controls, a foreign-born dummy, latitude and longitude cubics for the 1900 county of residence, age and age squared in 1900, and 1900 state of residence fixed effects. The right hand side variable of interest, $\Delta asinh(MA_{00-10,c00})$, is the inverse hyperbolic sine transformed change in market access (MA) in an individual's 1900 county of residence. asinh() refers to the inverse hyperbolic sine transformation. OLS refers to columns where coefficients are estimated using OLS. IV refers to columns where an instrumental variable regression is employed. In all specifications, observations are dropped if the occupation score is zero or in the case of father's occupation score, if the occupation score is zero in either 1900 or 1910. Means of dependent variables are for original, full sample (with zeroes). Standard errors are clustered at by 1900 state of residence.

Table A.20
Veteran status, 1940.

Dep. Var.:	(1) $\mathbb{1}(veteran)$ OLS	(2) $\mathbb{1}(veteran)$ IV
$\Delta asinh(MA_{00-10,c00})$	–0.0143	–0.0128
	(0.0167)	(0.0176)
Observations	36,202	36,202
R-squared	0.056	0.056
Mean of dep. var.:	0.472	0.472

Notes: *** denotes 1% significance, ** denotes 5% significance, * denotes 10% significance. All specifications include, as controls, a foreign-born dummy, latitude and longitude cubics for the 1900 county of residence, age and age squared in 1900, and 1900 state of residence fixed effects. The right hand side variable of interest, $\Delta asinh(MA_{00-10,c00})$, is the inverse hyperbolic sine transformed change in market access (MA) in an individual's 1900 county of residence. asinh() refers to the inverse hyperbolic sine transformation. $\mathbb{1}()$ refers to a dummy variable equal to 1 if the condition within the brackets is satisfied; in the dummies, "veteran" is if the son is a veteran by 1940. OLS refers to columns where coefficients are estimated using OLS. IV refers to columns where an instrumental variable regression is employed. The sample consists only of those sons that are sample-line males in 1940, which is a very small proportion of the full-count 1940 Census. Standard errors are clustered at by 1900 state of residence.

A2. Data appendix

A2.1. County boundaries

All county boundaries for contemporaneous control variables at the county level are those of that year. 1900 county control variables are measured in terms of 1900 county boundaries, for the county in which the individual lived in 1900.

Market access measures are first constructed using 1890 county boundaries, to maintain consistency with the Donaldson and Hornbeck (2016) transportation costs database from which the measures rely on. To construct the measures, county-level population data from ICPSR is converted from 1870, 1900, and 1910 boundaries respectively to 1890 boundaries using a method based on Donaldson and Hornbeck (2016) and their methodology.[35] The 1890 county measures are then converted to 1900 county boundaries using a variant of the county crosswalk procedure in Donaldson and Hornbeck (2016) and Hornbeck (2010). The key difference is that, for each 1900 county, I take a weighted average of the market access for all the 1890 counties which overlap with that 1900 county. Weights are determined by the proportion of the area that each 1890 county makes up, for that 1900 county.

A2.2. Census linking procedures

To link the 1900, 1910, and 1940 Censuses, I make use of publicly available linkages from the Census Linking Project. The Project provides links between individuals based on various methods and allows researchers to create linked samples using the histid variable in the IPUMS Census data which uniquely identifies individuals in each Census. I link individuals across the 3 Census waves used by first linking individuals together across the 1900 and 1910 Censuses, then linking together individuals in the 1910 and 1940 Censuses, and then finally keeping individuals who can be linked from 1900 to 1910 and then 1910–1940. This results in an initial linked sample of individuals of 3.7 million.

For the 1900–1910 Census linkage, I use the Price et al. (2021) linking method. Their procedure combines automated linking methods such as those used by ABE (explained below) with unique information from FamilySearch, a genealogical service. Their method makes use of personal information and linkages uploaded by family members; FamilySearch's site thus allows these family members, who Price et al. (2021) argue have a vested interest in making correct linkages and uploading correct information, to create a crowdsourced dataset that allows for linking across Census waves using a much larger set of matching variables that are much more likely to be correct. Using this crowdsourced information as a training dataset, Price et al. (2021) use a machine learning approach and other methods to link individuals with a higher success rate and degree of accuracy than other methods currently widely used.[36]

In the case of the 1910–1940 Census linkage, I make use of the Abramitzky-Boustan-Eriksson (ABE) method of linking which makes use of NYSIIS standardized names.[37] This procedure uses individuals' full names, year of birth, and birthplace, after some initial data cleaning of names and standardizing of spellings. The procedure looks for exact matches, after which a buffer of 1 and then 2 years is used for year of birth to look for matches. One potential issue present in this method (and indeed all similar methods for linking) is that in the case of multiple matches that occur, a linkage cannot occur since there is no way to know which match is the correct one.

References

Abramitzky, R., Ager, P., Boustan, L., Cohen, E., Hansen, C., 2019. The effects of immigration on the economy: lessons from the 1920s border closure. NBER Working Paper No. 26536.
Abramitzky, R., Boustan, L., Eriksson, K., Feigenbaum, J., Perez, S., 2021. Automated linking of historical data. J Econ Lit 59 (3), 865–918.
Abramitzky, R., Boustan, L., Rashid, M., 2020. Census Linking Project: Version 1.0 [dataset]. https://censuslinkingproject.org.
Ager, P., Boustan, L., Eriksson, K., 2021. The intergenerational effects of a large wealth shock: white southerners after the civil war. Am Econ Rev 111 (11), 3767–3794.
Ager, P., Brueckner, M., Herz, B., 2017. The boll weevil plague and its effect on the southern agricultural sector, 1889–1929. Explor Econ Hist 65, 94–105.
Ager, P., Herz, B., Brueckner, M., 2020. Structural change and the fertility transition. Rev Econ Stat 102 (4), 806–822.
Atack, J., 2016. Historical geographic information systems (GIS) database of u.s. railroads for 1826–1911.
Atack, J., Bateman, F., Haines, M., Margo, R., 2010. Did railroads induce or follow economic growth? urbanization and population growth in the american midwest, 1850–1860. Soc Sci Hist 34, 171–197.
Autor, D., Dorn, D., Hanson, G., Song, J., 2014. Trade adjustment: worker level evidence. Q J Econ 129 (4), 1799–1860.
Becker, G.S., Lewis, H.G., 1973. On the interaction between the quantity and quality of children. J polit Econ 81 (2, Part 2), S279–S288.
Berger, T., 2019. Railroads and rural industrialization: evidence from a historical policy experiment. Explor Econ Hist 74, 101277.
Berger, T., Enflo, K., 2017. Locomotives of local growth: the short- and long-term impact of railroads in Sweden. J Urban Econ 98, 124–138.
Black, S., Devereux, P., Salvanes, K.G., 2005. The more the merrier? the effect of family size and birth order on children's education. Q J Econ 120 (2), 669–700.
Bleakley, H., Lange, F., 2009. Chronic disease burden and the interaction of education, fertility, and growth. Rev Econ Stat 91 (1), 52–65.
Bloom, N., Romer, P., Terry, S., Van Reenen, J., forthcoming. Trapped factors and china's impact on global growth.
Boustan, L., Bunten, D., Hearey, O., 2018. Urbanization in the United States, 1800 deldDel- deliIns–2000. Oxford Handbook of American Economic History.
Brooks, L., Gendron-Carrier, N., Rua, G., 2021. The local impact of containerization. J Urban Econ 126.
Carrillo-Tudela, C., Hobijn, B., She, P., Visschers, L., 2016. The extent and cyclicality of career changes: evidence for the u.k. Eur Econ Rev 84, 18–41.
Chan, J., 2022. Data for "The long-run effects of childhood exposure to market access shocks: Evidence from the US railroad network expansion" (Explorations in Economic History). Tech Rep. Inter-university Consortium for Political and Social Research [distributor]. 2022-12-17. https://doi.org/10.3886/E183563V1
Chan, J., 2022b. Farming output, concentration, and market access: Evidence from the nineteenth century American railroad expansion.

[35] Specifically, I follow their methodology except that I allow intersecting county boundary pieces to come from out-of-state, given the long period examined and the higher likelihood of state boundary shifts making a difference.

[36] In Figure 4 of their paper, Price et al. (2021) show that their augmented approach with the additional FamilySearch data improves on numbers of links generated and the rate at which links agree with their training dataset compared to other approaches including the ABE-type methods.

[37] The procedure is detailed in Abramitzky et al. (2021). Unfortunately, the Price et al. (2021) method using FamilySearch data has not yet been extended to 1910–1940 linking and cannot be implemented by other researchers without the additional data from FamilySearch.

Clay, K., Lingwall, J., Stephens Jr, M., 2021. Laws, educational outcomes, and returns to schooling: evidence from the first wave of US state compulsory attendance laws. Labour Econ 68.
Collins, W., Wanamaker, M., 2022. African American intergenerational economic mobility since 1880. Am Econ J: Appl Econ 14 (3), 84–117.
Costas-Fernandez, J., Guerra, J., Mohnen, M., 2022. Train to opportunity: the effect of infrastructure on intergenerational mobility.
Cummings, R., 1949. The American Ice Harvests: A Historical Study in Technology, 1800–1918. California University Press.
Dalgaard, C., Knudsen, A., Selaya, P., 2020. The bounty of the sea and long-run development. J Econ Growth 25, 259–295.
Dix-Carneiro, R., Kovak, B., 2017. Trade liberalization and regional dynamics. Am Econ Rev 107 (10), 2908–2946.
Donaldson, D., 2018. Railroads of the raj: estimating the impact of transportation infrastructure. Am Econ Rev 108 (4–5), 899–934.
Donaldson, D., Hornbeck, R., 2016. Railroads and american economic growth: a "market access" approach. Q J Econ 131 (2), 799–858.
Ducruet, C., Juhasz, R., Nagy, D., Steinwender, C., 2022. All aboard: The effects of port development.
Fajgelbaum, P., Redding, S., 2022. Trade, structural transformation and development: evidence from argentina 1869–1914. J Pol Econ 130 (5).
Feigenbaum, J., Tan, H., 2020. The return to education in the mid-twentieth century: evidence from twins. J Econ History 80 (4), 1101–1142.
Fernihough, A., 2017. Human capital and the quantity-quality trade-off during the demographic transition. J Econ Growth 22 (1), 35–65.
Feyrer, J., 2021. Distance, trade, and income - the 1967 to 1975 closing of the suez canal as a natural experiment. J Dev Econ 153.
Fogel, R.W., 1964. Railroads and american economic growth. Johns Hopkins Press Baltimore.
Goldin, C., Katz, L., 2008. The Race between Education and Technology. Belknap Press.
Goldin, C., Katz, L., 2011. Mass secondary schooling and the state: The Role of state compulsion in the high school movement. In: Costa, D., Lamoreaux, N. (Eds.), Understanding Long Run Economic Growth. Cambridge University Press.
Haines, M., ICPSR, 2010. Historical, demographic, economic, and social data: The United States, 1790–2002. Interuniversity Consortium for Political and Social Research.
Hornbeck, R., 2010. Barbed wire: property rights and agricultural development. Q J Econ 125 (2), 767–810.
Hornbeck, R., Rotemberg, M., 2021. Railroads, reallocation, and the rise of american manufacturing.
Hornung, E., 2015. Railroads and growth in prussia. J Eur Econ Assoc 13 (4), 699–736.
Jaworski, T., Kitchens, C., 2019. National policy for regional development: historical evidence from appalachian highways. Rev Econ Stat 101 (5), 777–790.
Kovak, B., Morrow, P., 2021. The long-run labour market effects of the canada-u.s. free trade agreement.
Lee, M., 2021. The effect of import competition on educational attainment at the postsecondary level: evidence from NAFTA. Econ Educ Rev 82, 102117.
Lewis, J., 2018. Infant health, women's fertility, and rural electrification in the united states, 1930–1960. J Econ History 78 (1), 118–154.
Lin, J., 2011. Technological adaptation, cities, and new work. Rev Econ Stat 93 (2), 554–574.
Maurer, S., Rauch, F., 2021. Economic geography aspects of the panama canal.
Mosquera, R., 2022. The long-term effect of resource booms on human capital. Labour Econ 74, 102090.
Price, J., Buckles, K., Van Leeuwen, J., Riley, I., 2021. Combining family history and machine learning to link historical records: the census tree data set. Explor Econ Hist 80.
Ruggles, S., Fitch, C., Goeken, R., Hacker, J., Nelson, M., Roberts, E., Schouweiler, M., Sobek, M., 2021. Ipums ancestry full count data: Version 3.0.
Saavedra, M., Twinam, T., 2020. A machine learning approach to improving occupational income scores. Explor Econ Hist 75.
Tang, J.P., 2014. Railroad expansion and industrialization: evidence from meiji Japan. J Econ Hist 74 (3), 863–886.
Vance, J., Shedd, T., Allen, G., 2020. railroad. https://www.britannica.com/technology/railroad. Accessed 24 August 2020.
Wolmar, C., 2013. The Great American Railroad Revolution. PublicAffairs.
Yamazaki, J., 2019. Railroads, technology adoption, and modern economic development: Evidence from Japan.
Zimran, A., 2020. Transportation and health in the antebellum united states, 1820–1847. J Econ History 80 (3), 670–709.