The Effect of Industrialization on Fertility and Human Capital in the 19th Century: Evidence from the United States

Ori Katz[1]

November 13, 2016

Abstract

During the period 1850-1900, the United States experienced a rapid process of industrialization. In this paper, I test for its effect on fertility and human capital. Using aerial distance from potential transportation routes as an instrument for industrialization, I find that the share of workers employed in manufacturing in a particular county had a significant negative effect on the ratio of children to adults in that county, and a significant positive effect on the proportion of literate adult males. The effect is robust to alternative specifications and measures of industrialization, and it is not a result of immigration. A heterogeneity analysis suggests that the effect was larger in counties that were more industrialized in 1850, leading to a divergence between them and less developed counties.

[1] Tel Aviv University, The Eitan Berglas School of Economics; email: orikatz@post.tau.ac.il

1. Introduction

During the Second Industrial Revolution, in the late 19^{th} century, fertility rates sharply declined in Western countries while human capital levels increased (Galor, 2005). In this paper, I provide evidence for a causal effect of industrialization on fertility and human capital, using panel data on 1,490 US counties for the period 1850-1900.

In order to identify a causal relationship between industrialization and the outcome variables, I use aerial distance from potential transportation routes between new major cities as an instrument for industrialization. The example of the Illinois counties is shown in figure 1. The population of Chicago grew from about 30,000 in 1850 to about 1.7 million in 1900 and during that period new transportation infrastructures were developed, connecting it to other major cities, such as St. Louis and Cincinnati. The industrialization process itself may be endogenous, and so may be the actual location of transportation infrastructure between major cities. For example, some counties were already more developed or gained access to a railway line because the local population was more educated or more politically connected. However, some counties, such as McLean, Logan and Montgomery, which were located on a straight line connecting Chicago and St. Louis, had a higher probability of gaining access to railways, roads or canals simply because of their location. The transportation infrastructures increased potential profits from industrialization in those counties. Thus, a possible instrument for industrialization is the distance of a county from the nearest straight line connecting between two large cities.

The exclusion restriction assumption is that the distance between a county and the straight line connecting two large cities is uncorrelated with fertility and human capital through any other variable except the measures I present for industrialization. To justify the validity of the instrument, I show that the distances are highly correlated with those to actual railway lines, so that they presumably capture access to transportation routes (but the actual location of transportation routes may be endogenous). Furthermore, I show that there were no pre-treatment differences. For example, prior to the rapid growth of Chicago, counties near the straight line connecting Chicago and St. Louis had no advantage over counties far away from the line. To control for other potential confounding factors, I include the distance to large cities, as well as county and year fixed effects.

The main sources of data are the National Historical Geographic Information System (NHGIS), which includes county-level data and county borders; IPUMS – USA, which includes individual-level data; and the Spatial History Project of the Center for Spatial and Textual Analysis of Stanford University, which includes historical data on cities' locations and populations. Using these sources, I construct a 6-period panel data for all the counties that existed between 1850 and 1900. Most of the results are limited to 1,490 counties east of the 95° line of longitude, because the western counties were not highly populated at the time and the empirical strategy makes less sense for counties that were far away from the largest cities (see figure 2). Industrialization is measured by the share of workers employed in manufacturing, and the main outcome variables are literacy rate and survival fertility (the number of children aged 5-18 per adult). In the main specification, "large cities" are the 10 most populated cities in the United States in each period. However, if a city drops out from the top-10 list its transportation infrastructure would not disappear, nor would its effect on industrialization in the counties lying between major cities. Therefore, I use the minimum distance to a straight line connecting major cities in all previous periods, as well as the current distance. The basic econometric model is as follows:

$$(1)\quad Log(\mathrm{Y}_{\mathrm{i},t}) = \beta_0 + \beta_1 Log(IND_{\mathrm{i},t}) + \beta_2 Log(CITYDIST_{i,t}) + \gamma_t + \delta_{\mathrm{i}} + \epsilon_{\mathrm{i,t}}\ ,$$

where $\mathrm{Y}_{\mathrm{i},t}$ is the outcome for county i at time t, $IND_{\mathrm{i},t}$ is the level of industrialization (which is instrumented by the minimum distances to straight lines connecting major cities), $CITYDIST_{i,t}$ is the distance to the nearest large city, γ_t are year fixed effects and δ_{i} are county fixed effects.

The main results establish that industrialization has a large and significant effect on fertility and human capital. Thus, an increase of 10% in the share of workers employed in manufacturing reduces fertility by about 3.1% and increases literacy by about 2.5%. The results are robust to alternative measures of industrialization, such as the real value of capital invested in manufacturing or the real value of manufacturing output, to various methods of selecting "large cities" and to other possible specifications. Heterogeneity analysis suggests that the effect of industrialization on fertility and human capital was relatively large for counties that were already more industrialized in 1850. If the growth of human capital also increases industrialization, this may imply the existence of a positive feedback loop, which creates a divergence between developed and undeveloped counties.

The theoretical literature suggests several mechanisms for the effect of industrialization on fertility and human capital, including the quality-quantity trade-off (Becker 1960, Becker and Lewis 1974) and the rise in the demand for human capital (Galor and Weil 1999, Galor and Moav 2002). Previous empirical studies focused on the relationship between fertility and human capital (Murphy 2010; Becker, Cinnirella and Woessmann, 2010; Bleakley and Lange, 2009) and on the effect of industrialization on human capital (Franck and Galor, 2015; Pleijt, Nuvolari and Weisdorf, 2016). The main results of this study are in line with those of previous studies and further our knowledge by analyzing the case of the United States, by presenting the effect of industrialization on literacy and fertility at the same time, by using a novel identification strategy combined with a panel data analysis, and by presenting a heterogeneity analysis of the effect according to various attributes.

The paper is organized as follows: The next section surveys the relevant theoretical and empirical literature on the historical development of industrialization, fertility and human capital. Section 3 presents the data and a descriptive analysis of industrialization, fertility and human capital in the United States during the period 1850-1900. Section 4 discusses the empirical strategy and the validation of the instrument. Section 5 presents the main results and some robustness checks. Section 6 examines heterogeneity in the main results. Section 7 concludes.

2. Fertility, Human Capital and the Industrial Revolution

2.1 The First and Second Industrial Revolutions

The first wave of the Industrial Revolution, which started around 1760, was characterized by several important macro inventions, such as the steam engine, and a number of developments in the textile and other industries, which drove prices down. Between 1750 and 1800 the per capita level of industrialization increased by 50% in the United Kingdom and doubled in the United States (Bairoch, 1982). Income per capita increased, as did fertility and urbanization. Human capital played a limited role in the industrialization process during this period, and schooling was mainly motivated by reasons unrelated to the labor market. A large share of the work in the industrial sector could be performed by uneducated workers (Landes, 2003), and industrialization led to a de-skilling process, in which factories employing unskilled labor replaced skilled artisans. The education level in the UK, the most innovative part of the world at the time, was lower than that in other European countries, and most of the technological progress was made by British amateurs and artisans, rather than by professional scientists or engineers (Mokyr 1992).

The second wave of the Industrial Revolution occurred between 1860 and 1913 and is the focus of this paper. According to Rosenberg and Trajtenberg (2004), during the time period 1838-1880 the number of steam engines used for manufacturing in the United States increased from 1,420 to 56,123 and the number of waterwheels and turbines increased from 29,324 to 55,404. The total horsepower used in manufacturing increased during 1828-1900 from under 1,000,000 to about 11,000,000. One difference between the second wave and the first wave is to do with human capital. During this period, capital-skill complementarity emerged: the new machinery required operators, engineers and mechanics, and as the establishments became larger and began to serve more remote markets, the demand for managers and other white-collar nonproduction jobs increased. As a result of these trends, the manufacturing labor force in the United States "hollowed out", as the demand for middle-skilled artisans declined while that for low- and high-skilled jobs increased (Katz and Margo 2013). A key technological change in this process was the rapid diffusion of the new steam engine, which required specialized expertise to install and maintain, though it also increased the division of labor and required unskilled labor (in order to handle the coal). When electric power became available, the demand for educated and skilled labor increased while there was now less need for unskilled jobs involving the movement of raw materials and products around the plant. Katz and Margo (2013) show evidence for this process using establishment-level and individual-level data fpr the United States: the share of white-collar workers (professional-technical workers, managers, clerks and salesman) increased from 6.9% in 1850 to 17.1% in 1900, while the share of skilled blue-collar workers remained about 11%; the share of operators and unskilled workers increased from 28.7% to 36.4%, while the share of workers in agriculture decreased from 52.7% in 1850 to 35.3% in 1900. In this paper, I show that industrialization had an overall positive causal effect on literacy during this period, which may imply a positive net effect of the hollowing-out process.

Through their increasing demand for human capital, capitalists were a significant driving force behind public educational reforms (Galor and Moav 2006). Literacy and schooling rates increased in many countries during this period: the proportion of British children aged 5-14 attending primary school increased from 11% in 1855 to 74% in 1900 (Flora et al. 1983); the average years of schooling of males in the British labor force tripled by the beginning of the 20th century (Matthews et al. 1982, p 573); and British male literacy increased from 75% to 100% and British female literacy increased from 65% to 100% (Clark, 2003). In the United States, the average years of schooling among men born in 1850 was 8.71, as compared to 11 years in 1900 (Hazan, 2009). Similar trends appeared in other countries as well. The lack of good institutions for technical education in the United Kingdom eventually led to a decline in its relative technological advantage, while other countries such as the United States and Germany became technological leaders (Mokyr 1992).

At the same time, fertility rates declined sharply throughout the Western world. Crude birth rates (number of live births per 1,000 individuals) in England declined from 36 in 1875 to 20 in 1920; in Germany from 41 to 26; in Sweden from 31 to 21; and in Finland from 37 to 25 (Galor 2005). The United States had higher fertility rates than other Western countries (as observed already in 1798 by Thomas Malthus), although they also declined during the second wave of the Industrial Revolution: between 1790 and 1860 the population grew at an annual rate of 3%, as comparing to 2.3% during the period 1860-1890 and 1.9% during the period 1890-1910 (Gordon 2016).

2.2 The Demographic Transition: Driving Mechanisms and Empirical Evidence

The general trends of lower birth and death rates and higher investment in human capital are grouped under the term "The Demographic Transition", which was coined in 1929 by the American demographer Warren Thompson. Galor (2012) describes five theoretical mechanisms for the Demographic Transition that appear in the literature: (1) the rise in the level of parental income, which increased the opportunity cost of raising children and promoted investment in "quality" rather than "quantity" (Becker 1960, Becker and Lewis 1974); (2) the decline in infant and child mortality; (3) the rise in the demand for human capital, which increased investment in child education and the cost of raising children (Galor and Weil 1999, Galor and Moav 2002); (4) the decline in the gender gap (Galor and Weil 1996); and (5) the decline in the relative importance of children as "old-age security" with the development of new saving opportunities in the capital markets. In this study, I find no evidence for a larger effect of industrialization on fertility and human capital in counties with higher levels of income in the manufacturing sector or in counties with a higher proportion of women working in manufacturing.

Several studies have attempted to analyze the causal effects behind the demographic transition, most of them based on county-level analysis and instruments for some of the variables (as in this paper). Murphy (2010) studied fertility in 19th-century France. He found that the decline in fertility rates in France appears to precede industrialization, which casts doubt on the role of economic parameters in fertility decisions. Using department-level data (a department is the French equivalent of a county) for 1876-1896 and climatologic data to instrument infant mortality, he finds that female literacy and child schooling are negatively correlated with fertility, while mortality, industrialization, urbanization, and male education do not affect fertility after controlling for other variables. Wealth is positively correlated with fertility, as is the level of religiosity. Becker, Cinnirella and Woessmann (2010) look at 334 Prussian counties in 1849 and instrument fertility by sex ratios and education by landownership inequality and the distance to Wittenberg, where Martin Luther published his "Ninety-five Theses". They find that the causation between fertility and education runs both ways and that education in 1849 predicts fertility decisions in the period 1880-1905. Klemp and Weisdorf (2010) analyze the child quantity-quality tradeoff in England during the period 1580-1871 using exogenous variation in the number of surviving offspring as a result of parental fecundity and parish-level neonatal mortality. They find a significant trade-off: each additional child reduces the chances of literacy among its siblings by 10 percentage points. Bleakly and Lange (2009) use the eradication of hookworm disease in the American South as an exogenous source for the increasing returns on human capital and find a significant decline in fertility associated with its eradication. While these studies are similar to the present one in method, time period and context, they all focus on the relationship between fertility and human capital, rather than the effect of industrialization on fertility and human capital (it may also be that industrialization affects only fertility and human capital is affected by fertility, or the other way around).

Several more relevant papers look at the effect of industrialization. Franck and Galor (2015a) attempts to identify the causal effect of industrialization on human capital investment in 19th century France. Their identification strategy is based on the geographic diffusion of steam engines: the first steam engine used for industrial purposes in France was installed in the Fresmes-sur-Escaut department in 1732. They argue that the distance from this department can be used to predict industrialization in other departments. Using distance from Fresmes-sur-Escaut as an instrument, they find a positive and significant impact of steam engines during the period 1839-47 on the number of teachers, the share of children in primary schools, the share of apprentices in the population and literacy rates in subsequent years. Franck and Galor (2015b) use the same instrument in order to identify the effect of industrialization on fertility in France, and find that the number of steam engines in industrial production in 1860-1865 had a negative impact on fertility in 1870-1930. Pleijt, Nuvolari and Weisdorf (2016) uses exogenous variation in carboniferous rock strata as an instrument for the regional distribution of steam engines in England in 1800. They find that the adoption of steam engines increased the average working skills of the labor force, but did not affect human capital measures such as schooling and literacy.

While also contributing to this effort, the current study differs from the previous literature in several aspects, including geographical scope, time period, industrialization measures and identification strategy. The most important contribution of this paper is the use of panel data, together with a dynamic instrument for industrialization and also county and year fixed effects, rather than a cross-section analysis. Another important contribution is a heterogeneity analysis of the effect of industrialization according to various attributes, including an analysis of the impact of immigration on the effect. The results are in line with those of Franck and Galor (2015a) and Franck and Galor (2015b) who establish a large county-level effect of industrialization on fertility and human capital. However, in contrast to Pleijt, Nuvolari and Weisdorf (2016), I find that industrialization does has an effect on literacy. The difference between the results may be due to the different periods examined.

3. Industrialization, Fertility and Human Capital in the United States during the period 1850-1900

3.1 Scope and Data

The data is taken from the decennial censuses, the Agricultural Census and the Manufacturing Census carried out by the US Bureau of the Census Library throughout the 19th century. As in the case of other historical databases, the data is far from perfect. For example, in the "Remarks on the Tables of Manufacturing Industry" in the 1870 survey, the author describes differences in the methodologies used in the manufacturing surveys of 1860 and 1870, such as the exclusion of the mining industry in 1870 which is partly compensated for by the inclusion of the milling of ores (which was not separated out from mining in 1860). He also describes numerous other practical problems encountered in the collection of the data.[2] Another example is the unavailability of certain age groups in some years. While little can be done to correct these deficiencies, it is worth noting that the main results of the paper are based on a panel analysis which includes fixed effects for counties and years. These fixed effects likely capture at least some of the inconsistencies between different years or between the different methods used by the assistant marshals responsible for collecting the data in each county.

Most of the analysis is based on the county-level data collected by the National Historical Geographic Information System (NHGIS), which also includes county boundaries.[3] The data for population and the location of cities was collected by the U.S. Census Bureau and Erik Steiner, as a part of the Spatial History Project of the Center for Spatial and Textual Analysis at Stanford University.[4] I also use individual-level data collected by IPUMS – USA.[5] Railway data was collected by the "Railroads and the Making of Modern America" project of the Center for Digital Research in the Humanities at University of Nebraska–Lincoln.[6] The CPI measure used to construct consistent panel data on the capital invested in manufacturing and manufacturing output is based on the work of Lawrence H. Officer and Samuel H. Williamson, in "The Annual Consumer Price Index for the United States, 1774-2014".[7] It is important to note that the county-level data is more limited than the IPUMS micro-level data. For example, the IPUMS data can be used to calculate schooling for different age groups for the United States as a whole, although the number of individuals is not large enough to calculate it for each county. In contrast, the NHGIS county-level data includes only a few age groups. Appendix A provides further details regarding the calculation of the main variables.

The sample period is 1850-1900. Data limitations prevented me from going back further than 1850. On the other hand, the empirical strategy, which is based on the emergence of new large cities, did not allow me to go beyond 1900, since there were only minor changes

---

2 The Statistics of the Wealth and Industry of the United States, Bureau of the Census Library, Washington Government Printing Office, 1872, p. 371

3 https://www.nhgis.org/

4 https://github.com/cestastanford/historical-us-city-populations

5 https://usa.ipums.org/usa/

6 http://railroads.unl.edu/

7 https://www.measuringworth.com/uscpi/

in the list of "large cities" after 1880. The analysis is carried out at the county level, and most of the results are limited to 1,490 counties east of the 95° line of longitude whose boundaries remained unchanged during the period. I used only those counties because most of the western counties were sparsely populated at the time (see figure 2), the boundaries and definitions of the western counties changed during the period, and the empirical strategy makes less sense for counties far away from the largest cities. However, the main results generally hold even when western counties whose boundaries remained unchanged and the city of San Francisco are included.

3.2 Industrialization

"Industrialization" is not a well-defined concept. Other empirical papers, such as Franck and Galor (2015) and Pleijt, Nuvolari and Weisdorf (2016), use the number of steam engines as a proxy for industrialization, but it is not clear from a theoretical point of view that this is the correct measure for testing mechanisms based on the wage gap between males and females or on the demand for human capital. In this study, the main measure used for industrialization is the share of the adult male population employed in manufacturing. However, one could argue that the share of workers in manufacturing is less important than the kind of manufacturing activity done in a particular county. For this reason, and also as a robustness test, the paper considers two other measures of industrialization: total capital invested in manufacturing per capita and total value of manufacturing output per capita (see Appendix A for the calculation of the variables). The share of the adult male population employed in manufacturing, total capital invested in manufacturing per capita and total value of manufacturing output per capita are highly correlated. The correlation coefficient between the average capital invested in manufacturing and the average value of manufacturing output during the period 1850-1900 is 0.91; between the capital invested in manufacturing and the share of workers in manufacturing it is 0.9; and between the value of manufacturing output and the share of workers in manufacturing it is 0.94.

The United States experienced a rapid wave of industrialization during the 19th century, which brought it to the global technological frontier (Rosenberg and Trajtenberg, 2004; Gordon, 2016). Table 1 presents summary statistics of the industrialization variables for 1850 and 1900. While the average share of males employed in manufacturing increased by 83% between 1850 and 1900, the average real capital invested in manufacturing increased by 415% and the average real value of manufacturing output increased by 277%. It is important to note that the share of workers in manufacturing, capital invested in manufacturing per capita and manufacturing output per capita for the entire United States are higher than the figures in Table 1, because the North-East and Midwest, which were more industrialized, were also more populated.

Figure 3 presents the geographic distribution of the average share of males employed in manufacturing during 1850-1900. Maps for the other two industrialization variables look very similar. The regional differences that can be seen in figure 3 are significant. For example, the average value of manufacturing output per capita in the counties of New York in 1870 was $122.25 (in nominal terms) and the average share of male workers in manufacturing was 18.72%; in contrast, the average value of manufacturing output per capita was $11.22 in the counties of Alabama and the average share of male workers employed in manufacturing was 3.4%.

3.3 Population, Fertility and Cities

The United States was one of the most fertile societies in the world during the 19th century. Its rate of population growth during 1870-1913 was about 2%, compared to 1.2% in Germany, 0.9% in the UK and 0.2% in France (Gordon 2016). Immigrants were an important part of the story: about half of the population growth in the United States between 1790 and 1920 can be attributed to immigration and the fertility of immigrants who arrived after 1790 (Hanies 2000, p. 155). The population of the United States was about 23.2 million in 1850, as compared to about 15.3 million in the UK and 36.5 million in France, while in 1900 the US population was already about 76.2 million, as compared to about 30 million in the UK, 40.7 million in France and 54.3 million in Germany.

The average population of the sample counties increased by 144% between 1850 and 1900 (see table 1). The population of some counties in New York and Philadelphia was more than one million in the latter part of the period, while some counties in Minnesota and Michigan had populations of only a few hundred. The US population in the 19th century was concentrated mainly in the Northeast and Midwest and expanded westward over the course of the century. Figure 2 presents the population density in each county (not just the eastern counties included in the sample) in 1870.

In this paper fertility is measured by the number of children above age 5 per adult (i.e. survival fertility, as measured by Fernández, 2014). Details regarding its calculation are provided in Appendix A. Apart from limitations in the data, the reason for using survival fertility rather than total fertility or number of births is the high rates of mortality among infants under the age of 5, which changed significantly during 1850-1900 and may have affected fertility decisions (Haines, 1998). Also due to limitations in the data, fertility relates to all adults and not only to white adults or native adults, which could have been more relevant for the purposes of this study. Table 1 presents summary statistics for fertility and population. It can be seen that fertility declined by 33% in the average county between 1850 and 1900. Figure 4 presents the geographical distribution of average fertility during 1850-1900. As can be seen, the regional differences in fertility were large.

One main focus of the analysis is the relationship between industrialization and fertility. Figures 3 and 4 show an overall negative correlation between the two variables. Figure 5 presents a scatter plot of the correlation, with a linear fit line for all of the counties in the sample (plotting fertility against the other industrialization variables yields similar results). The variables in the scatter plot are presented in logarithmic terms, as they are in the analysis that follows, since the relationship between them appears to be non-linear. The elasticity between average fertility and the average share of male workers employed in manufacturing is -0.121. This analysis will attempt to determine whether there is a causal relationship behind this correlation, through the use of panel data, an instrument for industrialization and fixed effects for counties and years.

The empirical strategy presented in the next section is based on the major American cities in each decade. Table 2 presents the top 10 most populated cities in 1850 and 1900. The five most populated cities in 1850 were New York with 515,547 residents, Baltimore with 169,054, Boston with 136,881, Philadelphia with 121,376 and New Orleans with 116,375. By 1900, the map of the largest cities had changed and their populations had increased: New York had 3.4 million residents, Chicago had 1.7 million, Philadelphia had 1.3 million, St. Louis had 575,238 and Boston had 560,892. While all of the largest cities had increased in size, some experienced higher growth than others: between 1850 and 1900 Chicago jumped from 25th place to 2nd place, Cleveland from 42nd to 7th, Buffalo from the 16th to 8th, San Francisco from 24th to 9th and Detroit from 31st to 13th.[8] The rapid growth of these cities and the transportation infrastructure between them provides an exogenous source for industrialization, as described in section 4.

3.4 Human Capital

Fishlow (1966) writes that: "From the earliest time, the United States and her predecessor colonies stood close to or at the very forefront of the world in the educational attainment of the mass of the populace." According to the earliest literacy statistics (for 1840), more than 90% of white adults in the US were literate, a level similar to those in Scotland and Germany and higher than those in England and France. Easterlin (1981) estimates that in 1850 there were 1,800 pupils per 10,000 individuals in the US, as compared to 1,045 in the UK, 930 in France and 1,700 in Germany.

In this study, I use the adult male literacy rate as a measure for the level of human capital in a county, a statistic that is available for most of the counties for the years 1850, 1870, 1880 and 1900.[9] Other possible measures used in the literature, such as schooling, number of schools or occupations, are not available for most of the years and for many of the counties and does not provide a strong first stage.[10] Table 1 presents the proportion of literate adult males in 1850 and 1900. Average literacy rates declined during the period, even if we weight the counties by population. This decline in human capital also appears in other data sources and in other studies (see for example Hazan, 2009).

Figure 6 presents the geographic distribution of average literacy during 1850-1900. Figures 3 and 6 show an overall positive correlation between industrialization and literacy. Figure 7 presents a scatter plot of this correlation. Using other industrialization variables provides

[8] The relative ranking of cities described here is somewhat misleading, since it includes cities that were incorporated into larger cities during the period 1850-1900.

[9] For the years 1850, 1870 and 1900 I use county-level data from NHGIS, while for 1880 I use the 100% individual-level data base from IPUMS. The two sources differ in their definition of "literacy", but due to the use of year fixed effects this inconsistency does not alter the results.

[10] A variable called "total persons in school" is available for 1850, 1870, 1880 and 1890, but a first stage based on the counties and years for which this variable exists does not provide a strong relationship between the distances used as an instrument in this paper and actual industrialization. The correlation between schooling and literacy is high (0.8 in the sample counties in 1870), so that the effect of industrialization may be similar for both.

similar results. The elasticity of literacy with respect to the share of workers employed in manufacturing is 0.079. As can be seen, the literacy rate in many counties was close to 100%. A robustness analysis presented in the following sections excludes those counties and, as expected, finds a larger effect of industrialization on literacy. As with industrialization and fertility, the empirical strategy described in the next section will try to determine whether there is a causal relationship behind this correlation.

4. Empirical Strategy

4.1 The General Framework and the Selection of "Large Cities"

In order to identify a causal relationship between industrialization on the one hand and fertility and human capital on the other, we require an exogenous source for industrialization in a particular county, which in this analysis will be based on the geographic location of counties and cities.[11]

The strategy is best illustrated using the example presented in figure 1. In 1850, Chicago was still a relatively small city with a population of 29,963; however, it grew rapidly to almost 300,000 by 1870. During the same period, the population in St. Louis, another major city, grew from 77,860 to 310,864. The growth of both these cities may have been related to various attributes of their populations (and therefore was endogenous in the sense of our model), but it also affected the Illinois counties located between them. In 1850, Logan County had a population of 5,128 and only 2.7% of its adult males were employed in manufacturing, while the more developed Fulton County had a population of 22,508 and 5.5% of its adult males were employed in manufacturing. However, during the period 1850-1870 the population in Logan County increased by 350% while that of Fulton Country increased by about 70% and while the share of adult males employed in manufacturing increased to 7.8% in Logan Country, in Fulton County it increased to only 5.9%. The fertility rate in Logan County decreased from 3.25 in 1850 to 1.46 in 1870 while that in Fulton County decreased from 3.01 to 1.77. Meanwhile, the adult male literacy rate in Logan County increased from 93.4% to 95.4% while that in Fulton County decreased from 96.1% to 93.3%. What explains the difference in growth paths between those two Illinois counties?

A part of the explanation is the fact that Logan County happens to be on the straight line connecting Chicago and St. Louis (a line which also matches the route of the famous U.S. Route 66, built in 1926). As Chicago and St. Louis grew, new transportation infrastructure between the two cities increased the profits from industrialization in Logan County, and thus increased the share of workers employed in manufacturing, unrelated to the attributes of the local population (and therefore exogenous in the sense of our model). The straight line connecting Chicago and St. Louis was not very important when Chicago was a small city, i.e. in 1850, but became so by 1870. Thus, Logan County was "treated" somewhere during this period. This was not unique to Logan County: the share of adult males employed in manufacturing increased from 1.6% to 10.3% in McLean County, from 3.4% to 7.9% in De

[11] Similar empirical strategies can be found in Michaels (2008) and Atack, Bateman, Haines and Margo (2010).

Witt County, from 5.1% to 14.3% in Will County, from 1.5% to 9.5% in Du Page County, from 7.3% to 11.8% in Menard County, from 3.9% to 9.5% in Macon County and from 2.3% to 13.3% in Montgomery County. Other Illinois counties that experienced a large increase in industrialization, such as Wayne County (0.8% to 14.3%), Wabash County (2.9% to 13%), Marion County (1.2% to 13%), Washington County (0.7% to 8%) and St. Clair County (4.5% to 18.5%), were located near straight lines connecting St. Louis to Cincinnati and to Louisville. At the same time, most of the counties that were farther away from these connecting lines experienced more modest increases and sometimes even a decline in the share of workers employed in manufacturing. The counties near the connecting lines also experienced a relatively large decline in fertility and a large increase in literacy rates.

This was also the case for other states and other large cities. Building on this exogenous variance in industrialization, I propose the following IV for the proportion of male workers employed in manufacturing: the distance between a county and the nearest line connecting between two of the largest cities in each period. I control for year and county fixed effects and for the distance to the nearest large city, since the location of the cities may be endogenous, and counties which are located near them also became more industrialized. The exclusion restriction is that the distance between a county and the straight line connecting two large cities, given the controls, is uncorrelated with fertility and literacy through any other variable except the proportion of males employed in manufacturing. The main results are therefore based on the variation in industrialization level induced by the varying distances to connecting lines between large cities and its effect on fertility and human capital.

The distance between US counties and connecting lines between large cities changed during the second half of the 19^{th} century, thanks to booming new cities such as Chicago and Buffalo, which functioned as transportation hubs. New railways, roads and canals were built in order to transport goods between the new cities and older ones in the east, and "middle counties" benefitted from the transportation infrastructure. Railroads were especially important: in 1840 the railroad mileage in the United States was similar to that of canals, but by 1850 it exceeded that of canals by more than two to one and by 1860 the United States had more miles of railroad than the rest of the world combined (Atack et al. 2010). There is a long-running debate in the literature over the role of railways in the economic growth of the United States during this period. Taylor (1951) argues that the railways advanced economic growth, while Fishlow (1965) claims that the railroad played a more passive role and its growth was endogenous and driven by economic development. Atack et al. (2010) found that railroads had no effect on population density, but did affect the trend of urbanization during that period. For the purpose of this study, it does not matter if railroads caused economic development or followed it - it only matters that to at least some degree the railroads and other types of transportation infrastructure increased the incentives for industrialization. The power of the first-stage regression presented here supports this claim, as do the findings of Atack et al. (2010) regarding the effect on urbanization.

A problem that may arise in using this strategy is the arbitrary selection of "large" cities. Most of the following analysis is based on the 10 most populated cities in each period east of the 95° line of longitude. Limiting the analysis to the eastern part of the United States excludes only one large city, San Francisco. As I show in the following sections, the first-stage regression results and the main results still hold for a longer or shorter list of cities or if we include the western counties and San Francisco.

Another issue that arises involves cities that dropped out of the top-10 list at some point. Assuming that those cities remained relatively large, the railroads or canals leading to those cities did not disappear, nor did their effect on industrialization in the counties between them. Therefore, in the panel analysis I use the minimum distance for all previous periods (in addition to the current one), rather than just the current period. For example, if in 1860 a county was close to a line that connected two top-10 cities, and in 1870 one of those cities dropped to 12th place (and the distance to other connecting lines is larger than in the case of this city), then the distance to the closest connecting line will not change for that county. This effectively means that cities can only enter the list of "large cities", they do not leave the list. Another modification of the top-10 list involves cities that became neighborhoods of other cities during the period 1850-1900. Besides disappearing from the data set, these cities were also very close to the larger cities, so there is no point in drawing a line to connect them. Therefore, the following large cities were omitted for all time periods: Brooklyn (which became part of New York) and Spring Garden, Northern Liberties and Kensington (which became neighborhoods of Philadelphia). Summary statistics regarding the distances to connecting lines in 1850 and 1900 are presented in table 1.

4.2 First-Stage Power

In this section, I establish that the instrument discussed above has a strong first stage. Specifically, I show that the distance between the center of a county and the straight line connecting two large cities has a strong first-stage relationship with the share of adult males employed in manufacturing. Table 3 presents the results of the first-stage regressions which show that these requirements are fulfilled. The regression equation is as follows:

$$(2)\quad Log(IND_{\mathrm{i},t}) = \beta_0 + \beta_1 Log(LINEDIST_{\mathrm{i},t}) + \beta_2 Log(CITYDIST_{i,t}) + \delta_{\mathrm{i}} + \gamma_{\mathrm{t}} + \epsilon_{\mathrm{i,t}}\ ,$$

where $IND_{\mathrm{i},t}$ is the share of adult males employed in manufacturing in county i during period t, $LINEDIST_{\mathrm{i},t}$ is the minimum distance between county i and the nearest connecting line between two of the 10 largest cities in any period $j \leq t$, $CITYDIST_{i,t}$ is the distance to the nearest large city, δ_{i} are county fixed effects and γ_{t} are year fixed effects.

The first stage results are presented for all observations (columns 1-3) and for observations for which we have literacy data (most of the counties in the years 1850, 1870, 1880 and 1900; column 4). As can be seen from the table, the distance has a significant negative effect on industrialization, given the controls. As can be seen from table 8, the results are very similar, both qualitatively and quantitatively, for other industrialization proxies, namely the value of manufacturing output and of capital invested in manufacturing. Since the variables are in logarithm form, the coefficients can be interpreted as elasticities. For example, column 4 indicates that increasing the distance to a connecting line by 10% will decrease the share of male workers employed in manufacturing by about 1.28%. I also present an F-statistic that is robust to clustering, according to Olea and Pflueger (2013). The F-statistic is larger than 10, a common rule of thumb in the empirical literature for a strong first stage.

Figure 8 presents a scatter plot of the relationship between the distances from the connecting lines and the value of manufacturing output. Panel A presents the unconditional relationship, while panel B reports the relationship after controlling for the distance from the nearest city and adding fixed effects for counties and years. The graph clearly shows that the relationship is not driven by outliers. Using the alternative industrialization variables provides a very similar picture. Figure 9 presents a scatter plot of the predicted values of the share of workers in manufacturing according to the first stage against the actual values. The plots and the regression results establish that there is a strong correlation between the distance from a connecting line and actual industrialization. Various robustness tests for these results are discussed in section 5.

4.3 Reduced Form

In this paper, I argue that a causal relationship exists between industrialization and fertility and between industrialization and literacy. However, since "industrialization" is not directly measurable, it is proxied using the share of male workers employed in manufacturing, the value of manufacturing output or the value of capital invested in manufacturing. This may mean that the interpretation of the 2SLS results are problematic, and that the combination of the first stage results together with the reduced form results are of greater interest. The regression equation for the reduced form is:

$$(3)\quad Log(Y_{\mathrm{i},t}) = \beta_0 + \beta_1 Log(LINEDIST_{\mathrm{i},t}) + \beta_2 Log(CITYDIST_{i,t}) + \delta_{\mathrm{i}} + \gamma_{\mathrm{t}} + \epsilon_{\mathrm{i,t}}\ ,$$

where $Y_{\mathrm{i},t}$ is fertility or literacy in county i at time t, and the other variables are defined as above.

Table 4 presents the estimation results, with and without the controls. According to column 3, which includes the controls of the main specification, the distance from a connecting line has a significant positive effect on fertility and a significant negative effect on literacy. Increasing the distance from a connecting line by 10% will increase fertility by 0.28% and decrease literacy by 0.33%. The exclusion restriction assumption is that this effect works only through the effect of the distances on the measures of industrialization.

4.4 Connecting lines and Transportation Infrastructure

A possible concern is that the correlation between the instrument and the industrialization variable is not due to transportation infrastructure, but rather some other missing factor. To address this concern, I present the correlation between the distance from the connecting lines and one of the most important transportation infrastructures, railroads. Data on the location of railroads exists for 1850, 1860 and 1870. The railroads themselves cannot be used as a valid instrument for industrialization, because their location may be endogenous and also because they vary significantly in importance and frequency of use. However, there is a strong correlation between the location of railroads and the connecting lines between major cities. In the sample counties, the correlation between the minimum distance to the nearest connecting line and the distance to the nearest railroad is 0.52 for 1850, 0.5 for 1860 and 0.41 for 1870.

Figure 10 presents maps of the connecting lines between the 10 largest cities and the railways in 1870. Figure 11 presents a scatter plot of the correlation between distance to railways and distance to connecting lines during the period 1850-1870, and also a linear fit and the regression equation for the two variables. The figures present a clear correlation between the location of railways and the location of the connecting lines. However, as can be seen in figure 10, there are many railways leading to distant counties, and figure 11 shows many county-year observations with a relatively short distance to a railway and a relatively large distance to a connecting line. Some of those railways in distant counties may have a low frequency of trains and a small effect on industrialization, while the connecting lines between major cities capture the more important transportation infrastructures with a larger effect on industrialization. Of course, other transportation routes such as canals and roads might have been as important as railways for industrialization at the time; however, lack of data prevents us from including them or from considering the volume of traffic in each railway.

4.5 Pre-Treatment Differences

There is the possibility that counties near or far from connecting lines were already different before the distance from connecting lines became important (given the controls), or that they had different pre-treatment trends. If counties near connecting lines were already different from counties farther away prior to the appearance of new large cities and the transportation infrastructure that connected them, then the proposed IV will not be valid. However, this appears not to be the case and the new transportation infrastructures were usually built in undeveloped areas. For example, Fishlow (1965) mentions a New Orleans Picayune editorial from 1860 claiming that “*nine-tenths of our roads when first traversed by steam pass through long ranges of woodlands in which the ax has never resounded, cross prairies whose flowery sod has never been turned by the plow, and penetrate valleys as wild as when the first pioneers followed upon the trail of the savage*…".

One way to address this concern is to separate the counties into two groups: a "treatment group" and a "control group", thus creating a binary version of the IV. The treatment group includes counties that were far away from the connecting lines in 1850 and then close to the connecting lines in 1870, while the control group consists of counties that were far from the connecting lines in both periods. Differences between the two groups will provide an indication of the validity of the instrument.

Figure 12 presents the analysis for 1850 and 1870.[12] The counties are divided according to whether they were among the farthest 50% or the closest 50% with respect to distance from the connecting lines in each period. This definition provides us with 423 treatment counties that were in the bottom 50% in 1850 and in the upper 50% in 1870 and 1,812 control counties, which were in the bottom 50% in both 1850 and 1870. According to the analysis, treatment and control counties where not statistically different with respect to fertility, literacy or industrialization in 1850, but by 1870 the treatment counties where already more industrialized and had a lower fertility rate and higher literacy rate. The results are robust to the alternative measures of industrialization and to various definitions of the treatment and control groups.

Another way to address this concern is to regress the main variables of interest on both the current distances from connecting lines and the future ones, along with all the controls, in order to determine whether there is any correlation between current variables and future distances. Table 6 presents the results for the following two equations:

(4) $Log(Y_{\mathrm{i},t}) = \beta_0 + \beta_1 Log(LINEDIST_{\mathrm{i},t}) + \beta_2 Log(CITYDIST_{i,t}) + \delta_{\mathrm{i}} + \gamma_{\mathrm{t}} + \epsilon_{\mathrm{i,t}}$,

(5) $Log(Y_{\mathrm{i},t}) = \beta_0 + \beta_1 Log(LINEDIST_{\mathrm{i},t+20}) + \beta_2 Log(CITYDIST_{i,t}) + \delta_{\mathrm{i}} + \gamma_{\mathrm{t}} + \epsilon_{\mathrm{i,t}}$,

where $Y_{\mathrm{i},t}$ are the variables of interest (share of male workers employed in manufacturing, fertility and literacy), $LINEDIST_{\mathrm{i},t}$ is the current minimum distance from the nearest connecting line, $LINEDIST_{\mathrm{i},t+20}$ is the future minimum distance from the nearest connecting line 20 years later, and the other variables are as described above. As can be seen from the table, the current distances are highly correlated with the current variables of interest, while future distances are not.[13] Thus, counties near the connecting lines, which are the ones driving the results, became more developed only after the growth of the connected cities.

5. The Causal Effect of Industrialization on Fertility and Human Capital

5.1 Main Results

Table 7 presents OLS and IV results for the main specification. The econometric model is as follows:

(6) $Log(\mathrm{Y}_{\mathrm{i},t}) = \beta_0 + \beta_1 Log(IND_{\mathrm{i},t}) + \beta_2 Log(CITYDIST_{i,t}) + \delta_{\mathrm{i}} + \gamma_{\mathrm{t}} + \epsilon_{\mathrm{i,t}}$,

where $\mathrm{Y}_{\mathrm{i},t}$ is fertility or literacy in county i at time t, $IND_{\mathrm{i},t}$ is the share of male workers employed in manufacturing, $CITYDIST_{i,t}$ is the distance to the nearest large city, δ_{i} are county fixed effects and γ_{t} are year fixed effects. $IND_{\mathrm{i},t}$ is instrumented by the minimum distance to the nearest connecting line between two of the 10 largest cities in each period.

[12] The graph for literacy does not present 1860 due to lack of data.

[13] These regressions are similar to the first stage and reduced form regressions, but due to the use of future distances the observations are truncated at 1880 so the results are not identical.

Column 2 shows that industrialization has a significant negative effect on fertility: an increase of 10% in the share of male workers employed in manufacturing reduces fertility by about 3.1%. Column 4 shows that industrialization has a significant positive effect on adult literacy rates: an increase of 10% in the share of workers employed in manufacturing increases literacy by about 2.5%. In both models, the OLS coefficient is biased towards zero relative to the IV coefficient. This may be due to measurement errors in the proxy for industrialization or a missing variable that operates in the opposite direction.

5.2 Alternative Measures of Industrialization

Literacy and fertility rates may have been influenced by other characteristics of the industrialization process that are relate to the type of manufacturing activity rather than the share of male workers employed in manufacturing. Table 8 presents first stage results for two alternative measures of industrialization: the real value of capital invested in manufacturing per capita and the real value of manufacturing output per capita. Table 9 presents the main results using these measures.[14]

According to table 8, the first stage is robust to the alternative industrialization measures. The effect of distance from the connecting lines on the real value of capital invested in manufacturing and the real value of manufacturing output is significant at the 1% level, and the F-statistic is above 10. The elasticity of industrialization with respect to the distance to connecting lines is similar for all three measures. For example, for observations with literacy data it is -0.128 for share of employment in manufacturing (see Table 3), -0.127 for manufacturing capital and -0.159 for manufacturing output.

According to table 9, the main results are also robust to the alternative industrialization measures. The IV results indicate that the elasticities of fertility and literacy with respect to the alternative industrialization measures are similar to those for the share of manufacturing employment. A 10% increase in the real value of manufacturing output per capita reduces fertility by 3.3% and increases literacy by 2.5%, while a 10% increase in the real value of manufacturing capital per capita reduces fertility by 2.6% and increases literacy by about 2%. The OLS results are biased towards zero also in the case of the alternative industrialization measures.

5.3 Alternative Methods for Selecting Cities

The previous results are based on the 10 largest cities east of the 95° line of longitude in each period. There is no intrinsic reason for choosing the 10 largest cities as opposed to any other method, such as the top 5% of cities or cities with more than 100,000 residents. It is, however, reasonable to assume that the inclusion of relatively small cities would reduce the power of the first stage, since the transportation infrastructure involved would be less significant in the development of the counties between them.

[14] The reduced form remains the same.

Table 10 presents the results when varying the number of large cities and confirms these assumptions. The results are presented for the 5, 7, 9, 11, 13 and 15 largest cities in each period. Overall, the results are similar both qualitatively and quantitatively to the baseline results reported before, although the power of the first stage is weaker for 13 and 15 cities. Other selection mechanisms, such as cities above some population threshold or the top 5% of cities, lead to similar results: as long as we have between 3 and 12 cities for each period the empirical strategy remains sound and the coefficients are of the same order of magnitude. However, using a larger number of cities usually leads to a weak first stage, which is expected because of the inclusion of less important cities and less important transportation routes.

5.4 Alternative Specifications and Sample Groups

Table 11 presents alternative specifications for the main results: no controls, partial controls, not including counties near large cities, including western counties and the city of San Francisco (which entered the top 10 list in 1870) and including only counties with literacy rates below 95%.

If we include distance to cities as a control and at the same time exclude the fixed effect for counties, then the first stage usually fails, since counties near large cities were much more industrialized than others. Not including year fixed effects appears to work in the technical sense, but the results are in fact strengthened due to the mixing between the actual effect and time trend effects (i.e. we are measuring minimum distance to connecting lines for all previous periods and therefore the distances can only get shorter with time, and in addition counties become more industrialized over time). Not including any controls weakens the effect relative to the main results, while excluding only distance to cities leads to a stronger effect for industrialization, since counties near large cities were also close to connecting lines by definition and became more industrialized when cities developed. When we control for the distance to cities, not including counties near large cities does not affect the main results.

The first stage is weaker and the effect of industrialization is stronger when including the western counties and the city of San Francisco. This is reasonable in view of the low population densities in the counties surrounding the connecting lines between San Francisco and other large cities (even though a railway to San Francisco was constructed by 1869). While they are of the same order of magnitude**,** the changing definitions and borders of western counties and states during the period and the long distances between the counties and connecting lines make these results less reliable than the results based only on the eastern counties.

Finally, including only counties with literacy rates below 95% leads, as expected, to a larger effect of industrialization on literacy.

5.5 Immigration as an Alternative Explanation of the Results

Immigration is a weak point in any study involving county-level variables. The effect of industrialization on fertility and human capital may be the result of the type of immigrants that were attracted by the new economic opportunities offered in industrialized counties, rather than a change in the behavior of native residents.

The individual-level data collected by IPUMS makes it possible to calculate foreign immigration and interstate immigration in each period. The results indicate that the share of foreign immigrants in the United States during the period was about 12%-14%, and the share of interstate immigrants (individuals not born in their current state of residence) was about 20%-26%. The data on foreign immigrants is available at the county level. If immigrants are driving the main results, we would expect that the effect of industrialization will be larger in counties or states with a larger share of immigrants. In order to test this hypothesis at the county level, the following equation is estimated:

$$(7)\quad Log(\mathrm{Y}_{\mathrm{i},t}) = \beta_0 + \beta_1 Log(IND_{\mathrm{i},t}) + \beta_2 HIGH\_IMMI_i * Log(IND_{\mathrm{i},t}) + \delta_{\mathrm{i}} + \gamma_{\mathrm{t}} + \epsilon_{\mathrm{i,t}},$$

where $\mathrm{Y}_{\mathrm{i},t}$ is fertility or literacy in county i at time t, $IND_{\mathrm{i},t}$ is the share of male workers employed in manufacturing, δ_{i} are county fixed effects, γ_{t} are year fixed effects, and $HIGH\_IMMI_i$ is a binary variable indicating whether a county had a relatively high average share of foreign-born residents during the period 1850-1900. $IND_{\mathrm{i},t}$ is instrumented as before by the minimum distance to the nearest connecting line, and the interaction term $HIGH\_IMMI_i * Log(IND_{\mathrm{i},t})$ is instrumented by the interaction between $HIGH\_IMMI_i$ and the log distances.

The results are presented in table 12. In column (1), $HIGH\_IMMI_i$ consists of counties within the top 25% with respect to average share of immigrants, and in column (2) $HIGH\_IMMI_i$ consists of counties within the top 10%. The top-25% counties results show a smaller effect for the share of manufacturing workers on literacy while there is no significant difference in the effect on fertility; however, the F-statistics are low. The top-10% counties results show a significantly smaller effect for the share of manufacturing workers on both literacy and fertility, and the F-statistics are close to 10.

Figure 13 presents the share of interstate immigrants for each state in 1880 – a variable that does not exist in the county-level data. It is mentioned in section 6 that the effect of industrialization was larger in states like New York, Connecticut and Pennsylvania, which had a relatively low level of immigration from other states. In states with high levels of internal migration, such as Iowa, Arkansas and Minnesota, the effect of industrialization is not significantly different from that in other states, while in Florida and Missouri the effect is smaller.

The combined results imply that foreign immigration and interstate immigration did not lead to a larger effect of industrialization on fertility or human capital, and perhaps even the opposite.

6. Heterogeneity

In this section, I decompose the average effect described in the main results in order to determine whether it varies between different groups of counties. Specifically, I use the equation presented in the previous section:

$$(8)\ Log(\mathrm{Y}_{\mathrm{i},t}) = \beta_0 + \beta_1 Log(IND_{\mathrm{i},t}) + \beta_2 GROUP_i * Log(IND_{\mathrm{i},t}) + \delta_\mathrm{i} + \gamma_\mathrm{t} + \epsilon_{\mathrm{i,t}},$$

where $GROUP_i$ is a binary variable indicating whether a county belongs to a particular group of counties that may differ in the effect of industrialization on fertility and human capital, and the other variables are as defined above. $IND_{\mathrm{i},t}$ is instrumented as before by the minimum distance to the nearest connecting line, and the interaction term $GROUP_i * Log(IND_{\mathrm{i},t})$ is instrumented by the interaction between $GROUP_i$ and the log distances. Other specifications of the econometric model such as the inclusion of several groups of counties together provide similar results, but the first stage is usually weaker.

For most of the groups tested, the coefficient of the interaction term was not significant, and in some of them the first stage also failed. Specifically, I could not find significant differences in the effect for counties with high or low shares of female workers in manufacturing, counties with high or low shares of children employed in manufacturing, counties with high or low shares of slaves in the population, relatively urbanized counties, counties in the Northeast, counties in the Midwest, counties in the South, counties in the Confederate states, counties in the slave states and counties in the free states.[15]

However, as can be seen from table 13, the results do indicate that the effect of industrialization varies according to the initial conditions of industrialization, fertility and literacy in 1850. The effects on both literacy and fertility during 1850-1900 were significantly larger in counties that were relatively more industrialized in 1850 and significantly smaller in the less industrialized counties; the effect on fertility was smaller in the low-fertility counties; and the effect on literacy was smaller in the high-literacy counties and larger in the low-literacy counties (though the F-statistic for this last result is somewhat low). These results are robust for various definitions of the groups of counties, and for including several groups in the same regression equation. Figure 14 presents the top-25% industrialized counties in which the effect was large and the bottom 25% in which the effect was small.

While one would expect that the effect on literacy will be smaller in counties with a literacy rate of already close to 100% in 1850, the results imply that fertility also has some "natural boundary", such that low-fertility counties are less affected.

The findings for initial industrialization levels indicate the existence of a positive feedback loop: in more developed counties industrialization had a larger effect on fertility and human capital, which in turn may have encouraged further industrialization, leading a divergence between them and less developed counties. This divergence, which is a familiar

[15] Testing the model for counties in a specific state (against all other counties), I find that the effect on literacy was significantly higher in the states of New York and Connecticut, the effect on fertility was higher in Pennsylvania (at a 10% significance level), the effect on literacy was lower in Ohio, Wisconsin, Missouri, Florida and Alabama, and the effect on fertility was lower in Wisconsin, Missouri, and Alabama.

phenomenon at the country-level during the Industrial Revolution (Galor, 2005), may not be visible at a regional level due to confounding factors such as immigration, unless we use an instrument for industrialization, as in this study.

A possible reason for this divergence is the type of industries established in the different regions. Figure 15 presents the trends in the main variables for county quartiles according to initial industrialization level in 1850. As can be seen, there is a clear divergence in capital invested in manufacturing and in the value of manufacturing output, while there is no divergence in the other variables. Counties that were more industrialized in 1850 developed more capital-intensive industries with higher output value, which generated larger capital-skill complementarity when measured using the identification strategy employed in this study.

Table 14 presents several differences between counties that were in the top 25% in regards to the share of adult males in manufacturing in 1850 and counties that were in the bottom 25%, based on individual-level data. The first 13 rows of the tables presents share of workers in industries which were the top-10 largest in 1850 or 1900. The breakdown of industries was already different in 1850, and some of the differences increased during the period. Specifically, in 1900 top-25% counties had a larger share of workers in the machinery, textile and metal industries, and a lower share in sawmills and grain mills. The last 3 rows of the table shows that in the top-25% counties the share of female employed in manufacturing, the share of literate workers employed in manufacturing and the avergae occupation score were significantly larger in 1900, while the differences in 1850 were smaller.

These results suggest that developed counties in 1850 differed from less developed ones in the type of industrialization that occurred between 1850 and 1900, and not just in the general level of industrialization, as captured by proxies such as the share of male workers employed in manufacturing. Industries that employed a higher share of skilled workers or females may be characterized by a larger effect of industrialization on fertility and human capital.

7. Concluding Remarks

This study establishes a causal effect for industrialization on fertility and human capital in the United States during the period 1850-1900. Using panel data with fixed effects for years and counties and an instrument for industrialization, the analysis showed a large and significant effect at the county level, which is in line with the theoretical literature on the Industrial Revolution and the Demographic Transition. Industrialization is measured by the share of adult male workers employed in manufacturing; fertility is measured by the ratio between the number of children and the number of adults; and human capital is measured by literacy rates among adult males. According to results for the main specification, an increase of 10% in the share of workers employed in manufacturing reduces fertility by about 3.1% and increases literacy by about 2.5%, and alternative specifications or measures for industrialization produce results of a similar magnitude.

The identification strategy is based on the development of new large cities during the period and the new transportation routes that connected them. Residents of counties that happened to be close to a straight line connecting between two large cities had an ex-ante higher probability of gaining access to new roads, railways or canals connecting these cities. The new transportation infrastructure increased the profits from industrialization, such that the distance from a straight line connecting two major cities is an exogenous source for industrialization that is not related to the level of local human capital and other variables. This exogenous source for industrialization, combined with county and year fixed effects, thus provides identification for the causal effect of industrialization on fertility and human capital.

Various concerns regarding the empirical strategy and the results are examined. It was verified that the distance from lines connecting major cities is highly correlated with the routes of actual transportation infrastructure, such as railroads, and that the counties near future connecting lines were no different from other countries before the appearance of the new large cities. The results are robust to various specifications of the variables, the method of selecting the large cities, and the method used to control for the distance from the cities (which may be endogenous). It is shown to be unlikely that immigrants are driving the results, since the effect of industrialization is smaller in counties and states with high proportions of foreign-born or interstate immigrants.

With respect to heterogeneity in the effect of industrialization, no differences were found according to county attributes such as the share of female workers employed in manufacturing, the share of children employed in manufacturing, the share of slaves in the population, and if the country was located in a Confederate state. However the effect was larger for counties that were relatively more industrialized in 1850, and further analysis showed that during the period 1850-1900 those counties developed industries that were more capital-intensive and had a higher share of skilled and female workers. These results may imply that different industries had different effects on fertility and human capital and that the type of industry may matter more than the general level of industrialization, as measured by the various proxies.

While the theoretical literature on the mechanisms behind the Demographic Transition is extensive and there is a growing empirical literature on the relationship between fertility and human capital, there is little empirical evidence for the effect of industrialization on fertility and human capital, and indeed none for the United States. Furthermore, to the best of my knowledge there is no empirical work that considers heterogeneity in the effect of industrialization. This study thus adds to the literature by using a novel identification strategy, based on panel data and an instrument for industrialization, to examine the case of the United States during the second half of the 19th century and by analyzing the heterogeneity of the effect of industrialization according to various attributes.

Although the research on this subject has advanced a great deal during the last decade, the Industrial Revolution and the Demographic Transition—the most dramatic changes in human history since the Neolithic Revolution—are still largely a mystery. The empirical literature, including this study, has not yet pinned down the relative importance of the proposed theoretical mechanisms. A potential direction for future research may be the heterogeneity of the effect of industrialization by industry, which may be related to specific mechanisms through which this effect operates. Figuring out the relative importance of the various proposed mechanisms may enable us to understand exactly what happened to the human race during the last 200 years, why it happened earlier in some parts of the world than others, and whether it will continue in the future.

Figures

Figure1 : Example for the Identification Strategy - Illinois Counties

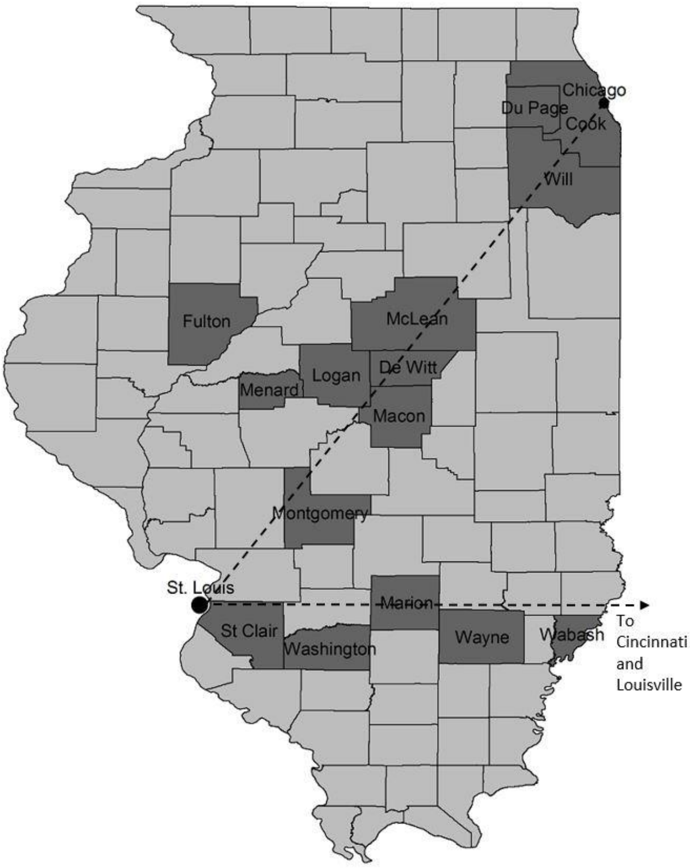


Figure2 : Population Density (individuals per km^2), 1870

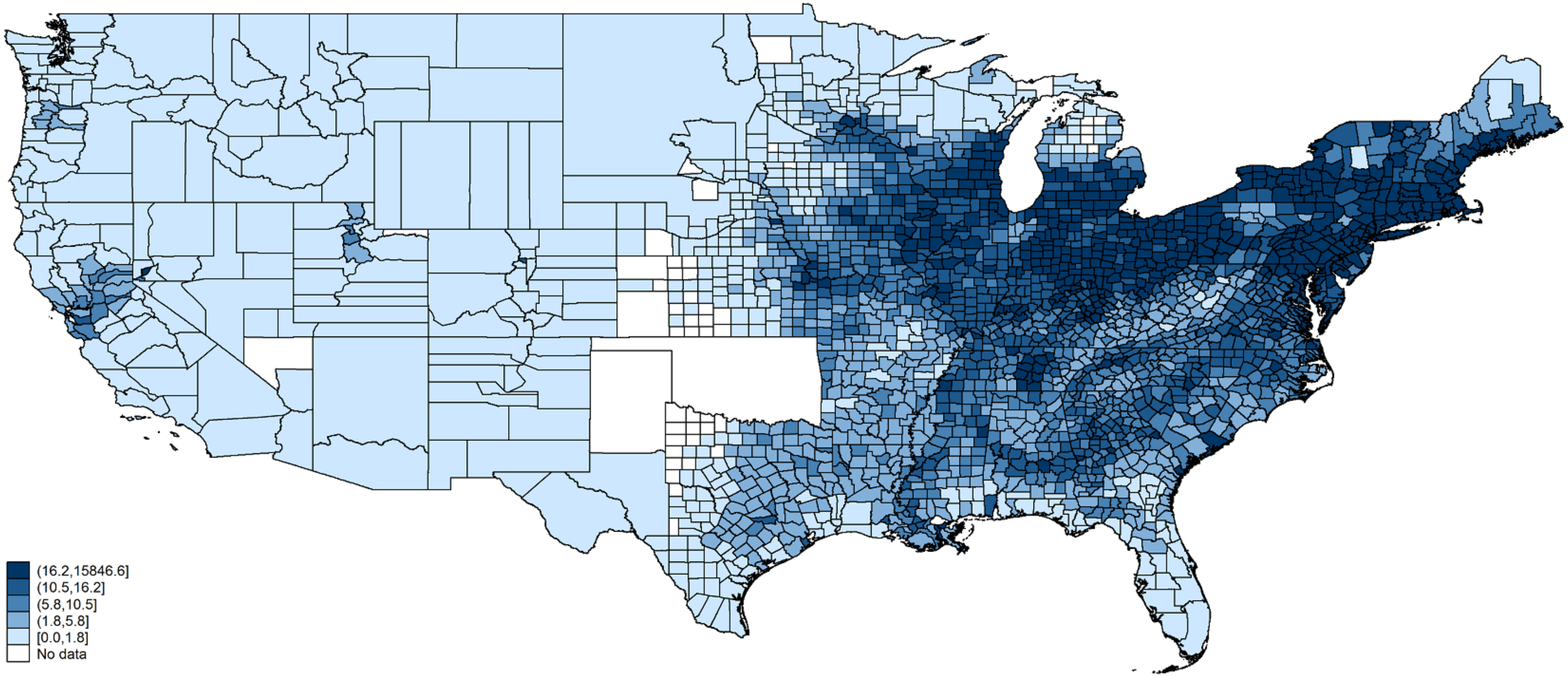


Figure3 : Share of Manufacturing Male Workers, Average 1850-1900

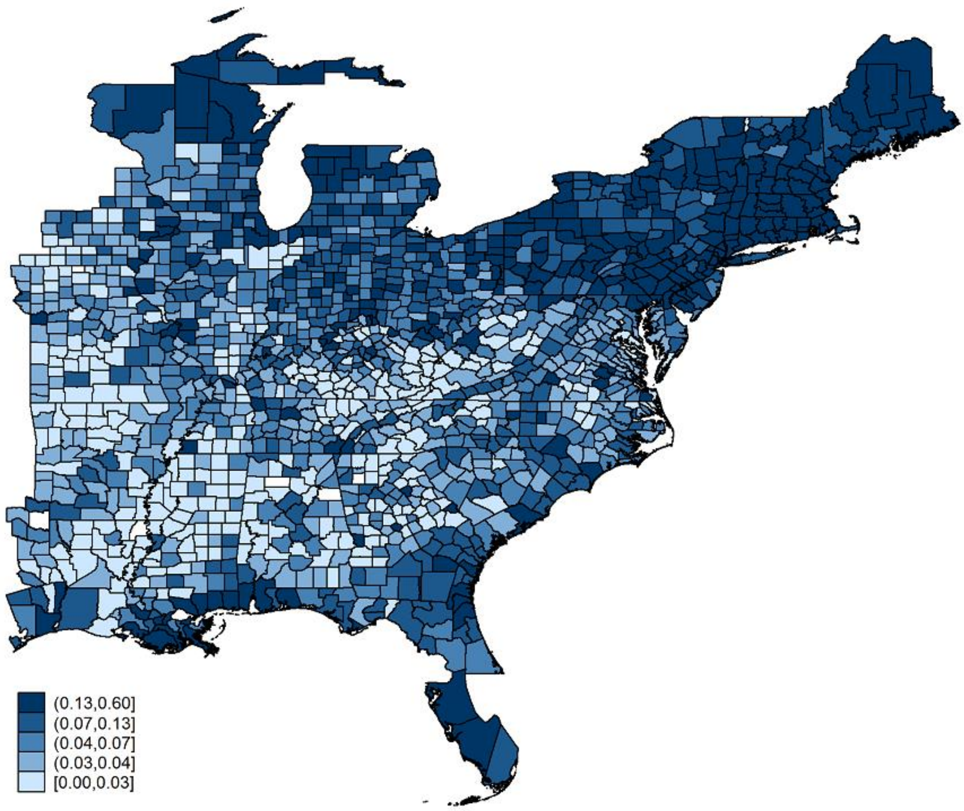


Figure4 : Fertility, Average 1850-1900

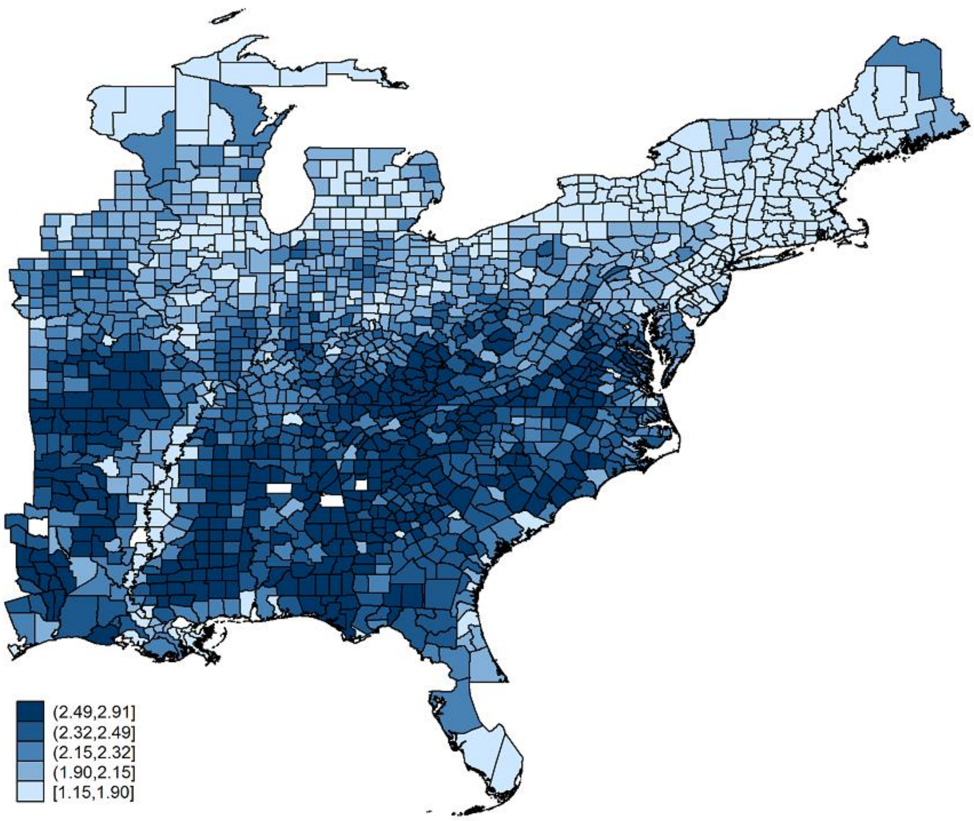


Figure5 : Fertility and the Share of Manufacturing Male Workers, Average 1850-1900

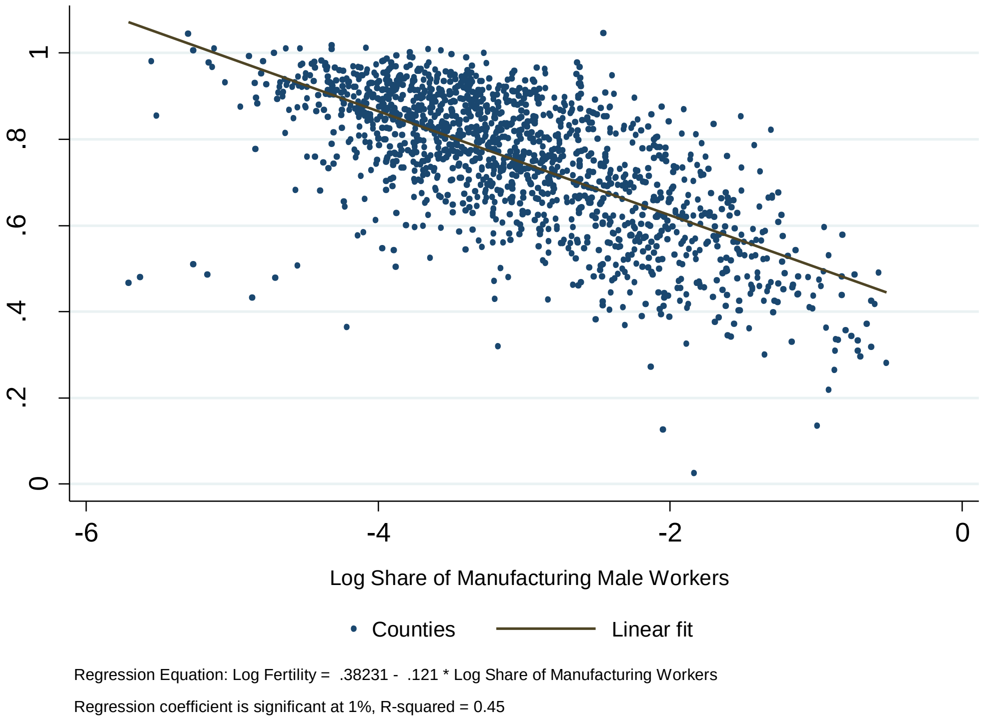


Figure6 : Adult Male Literacy, Average 1850-1900

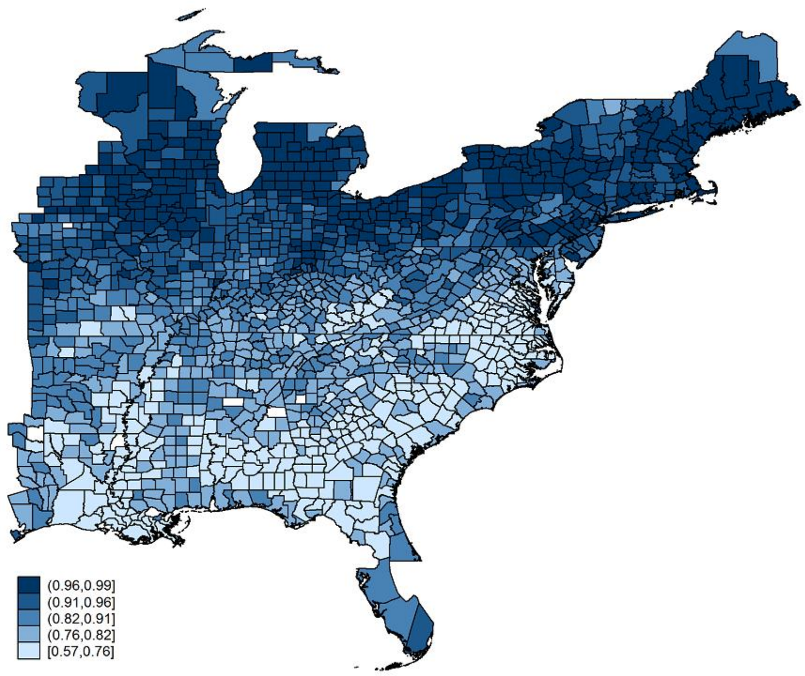
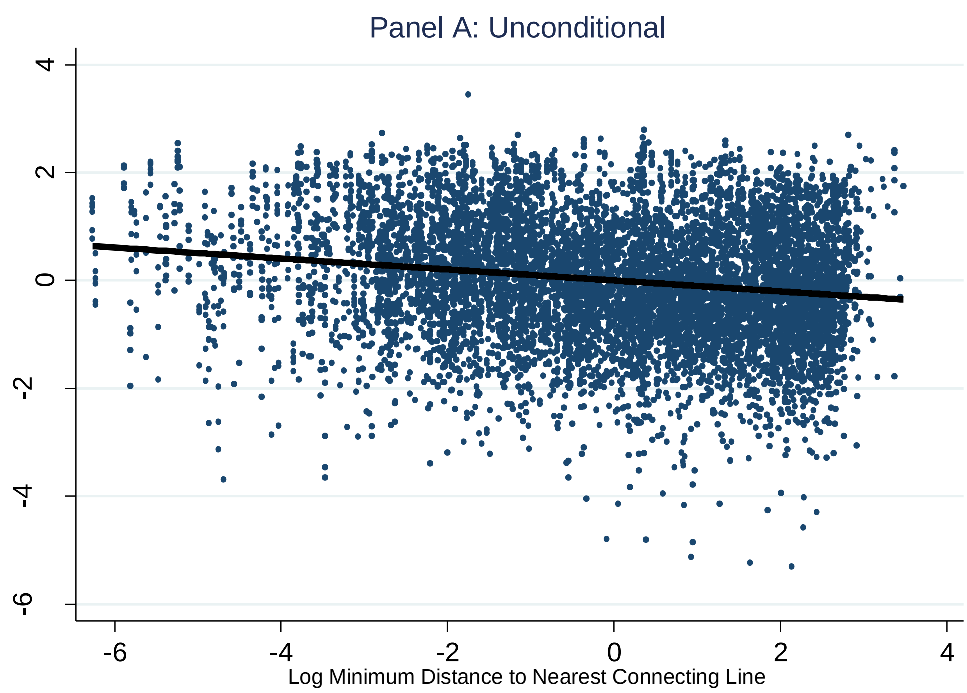


Figure7 : Literacy and the Share of Manufacturing Male Workers, Average 1850-1900

Regression Equation: Log Literacy = .05591 + .079 * Log Share of Manufacturing Workers

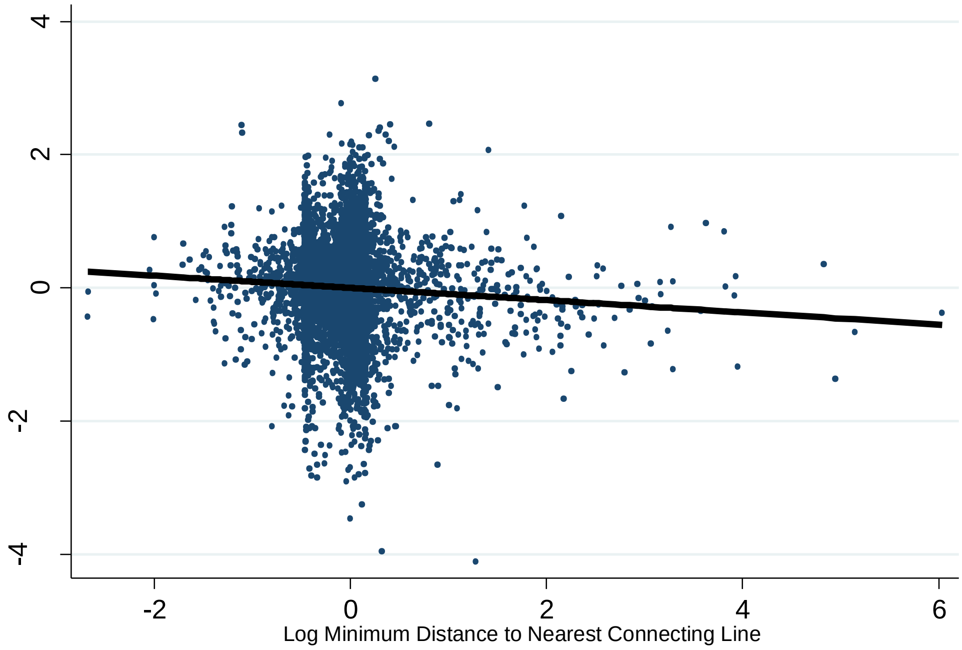
Regression coefficient is significant at 1%, R-squared = 0.23

Figure8 : The Effect of the Distances on Industrialization



coef = -.10154417, (robust) se = .01195915, t = -8.49

Panel B: Conditional on the Controls



coef = -.09220412, (robust) se = .02310386, t = -3.99

Figure9 : Predicted and Actual Industrialization

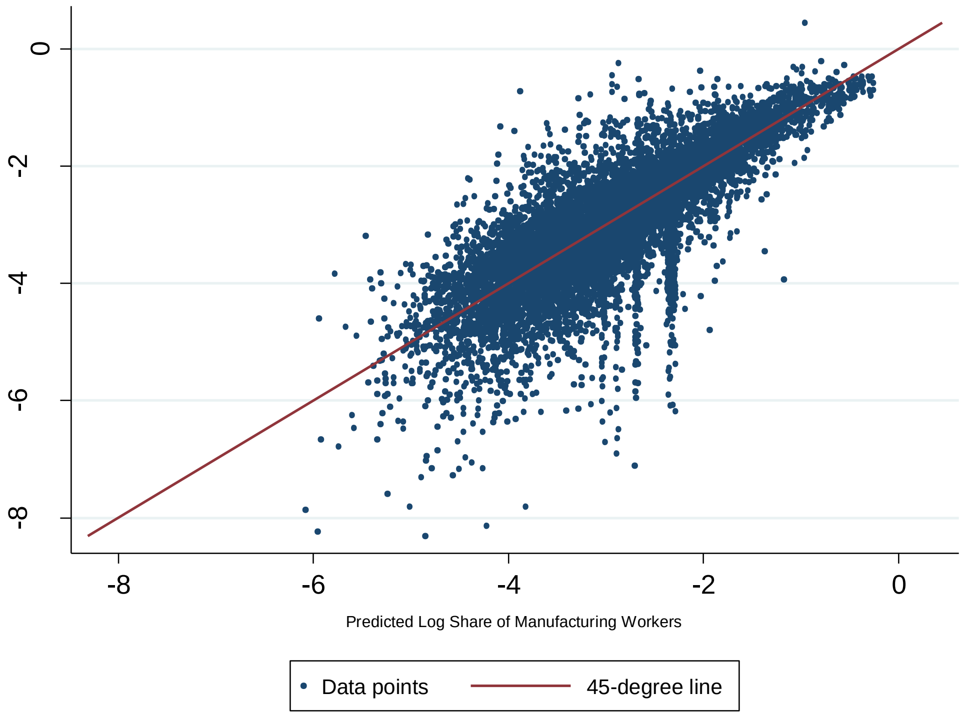


Figure10 : Railways (left) and Connecting Lines (right), Top 10 Cities, 1870

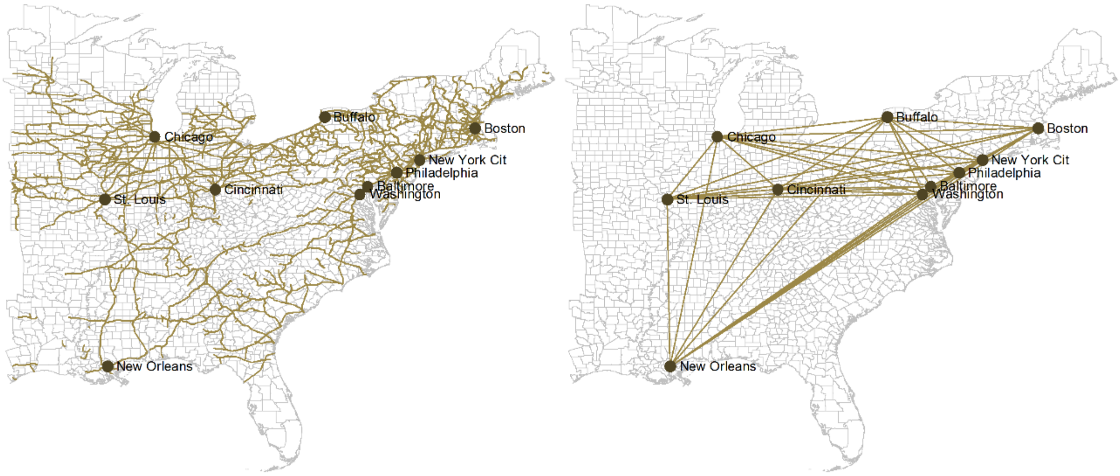


Figure11 : Distance to Railways and Distance to Connecting Lines, 1850-1870

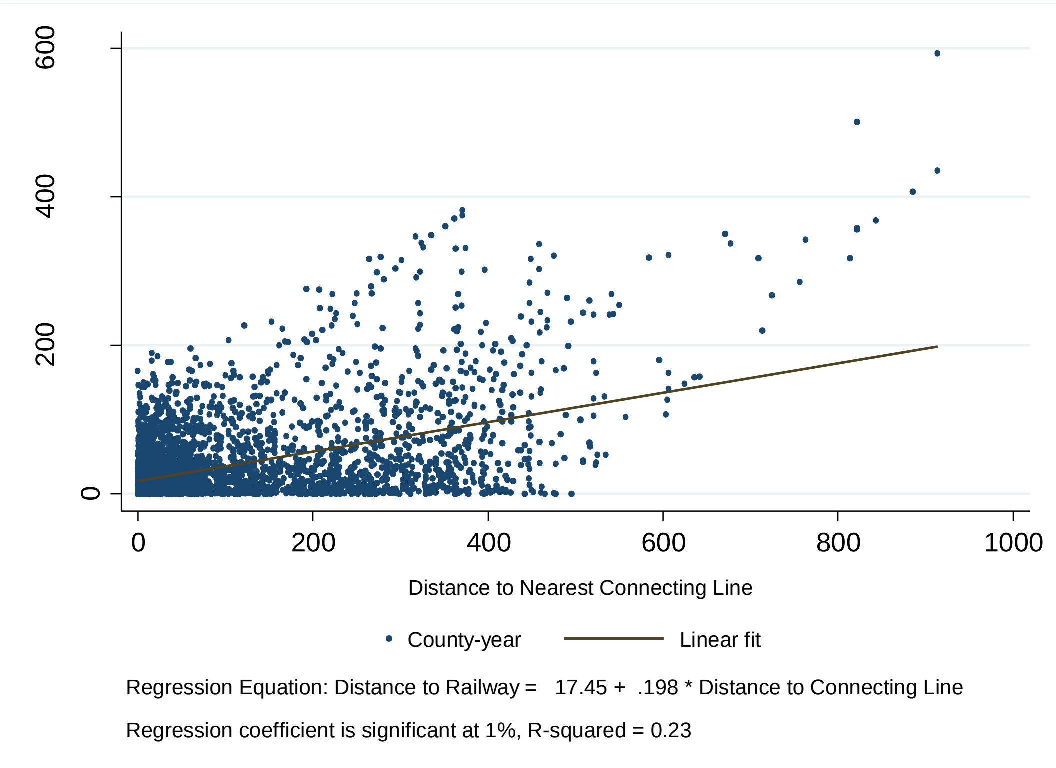
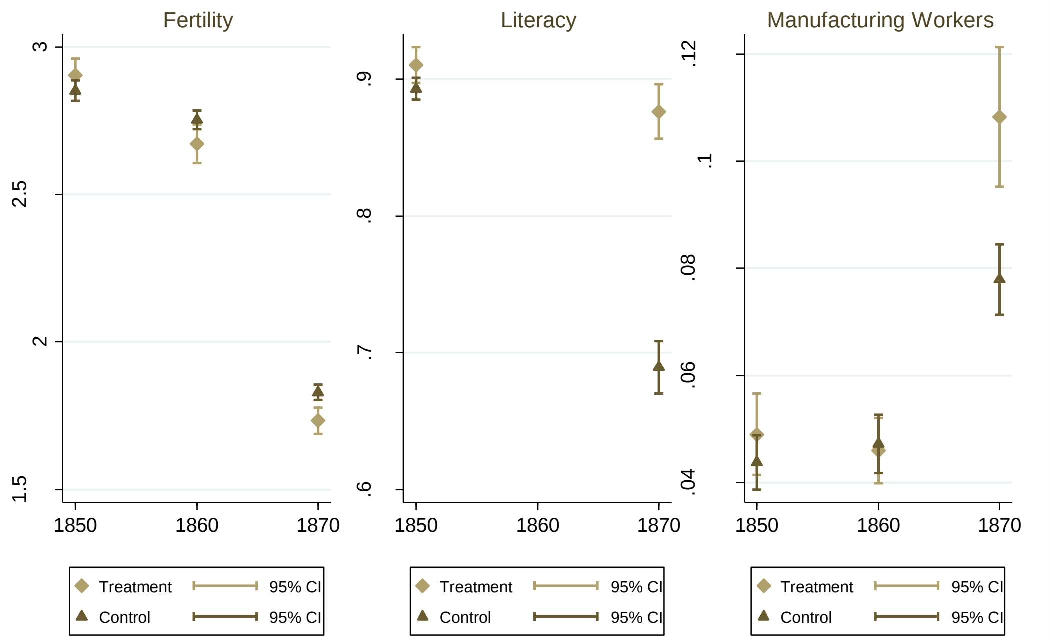


Figure12 : Pre-Treatment Differences



Treatment counties - counties which were far from connecting lines (lower 50%) in 1850 and close to connecting lines (upper 50%) in 1870

Control counties - counties which were far from connecting lines (lower 50%) both in 1850 and in 1870

Figure13 : Share of Interstate Immigrants by States, 1880

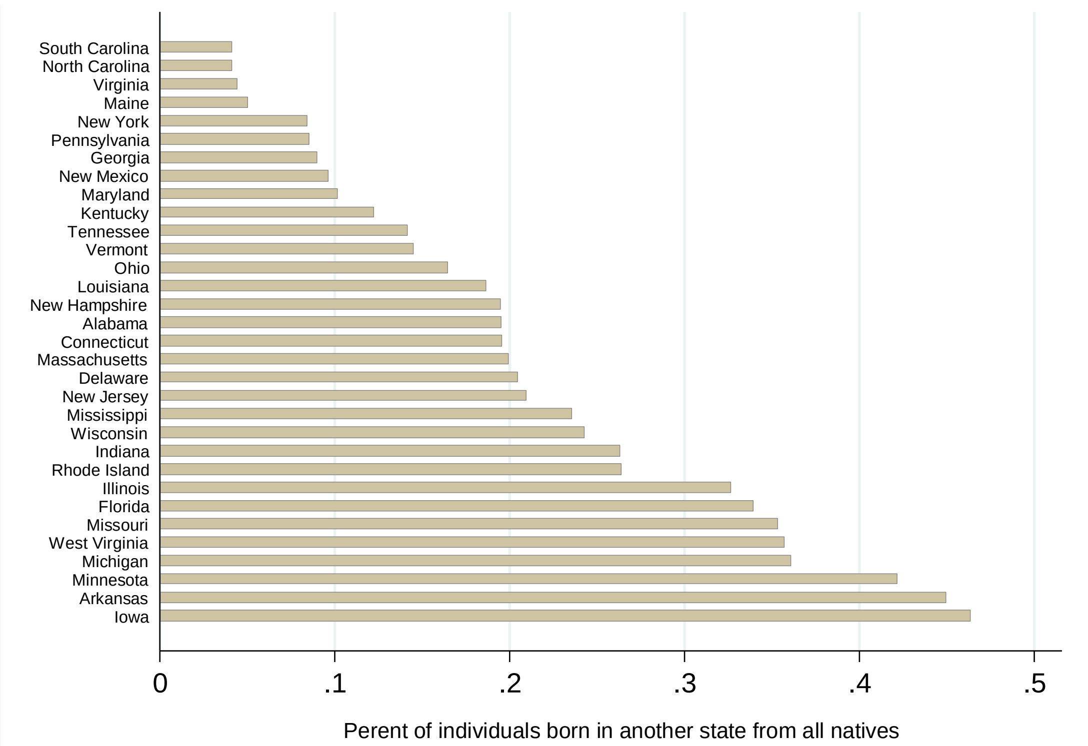


Figure14 : Levels of Industrialization in 1850

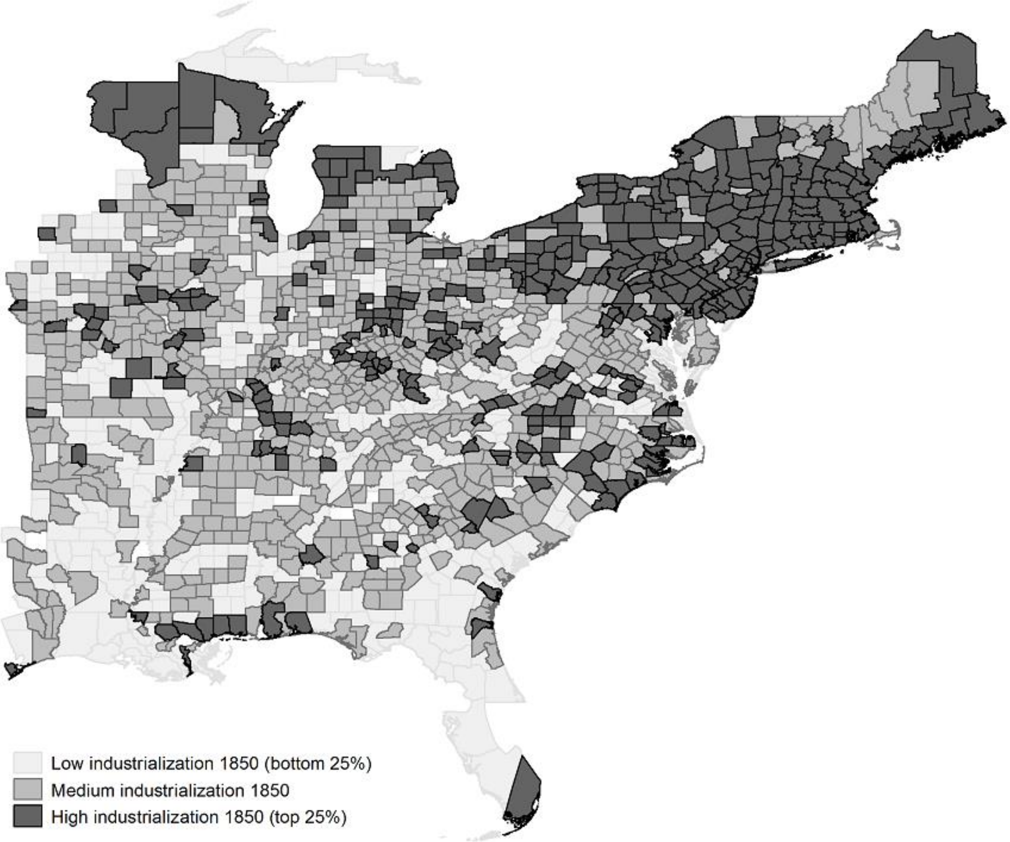
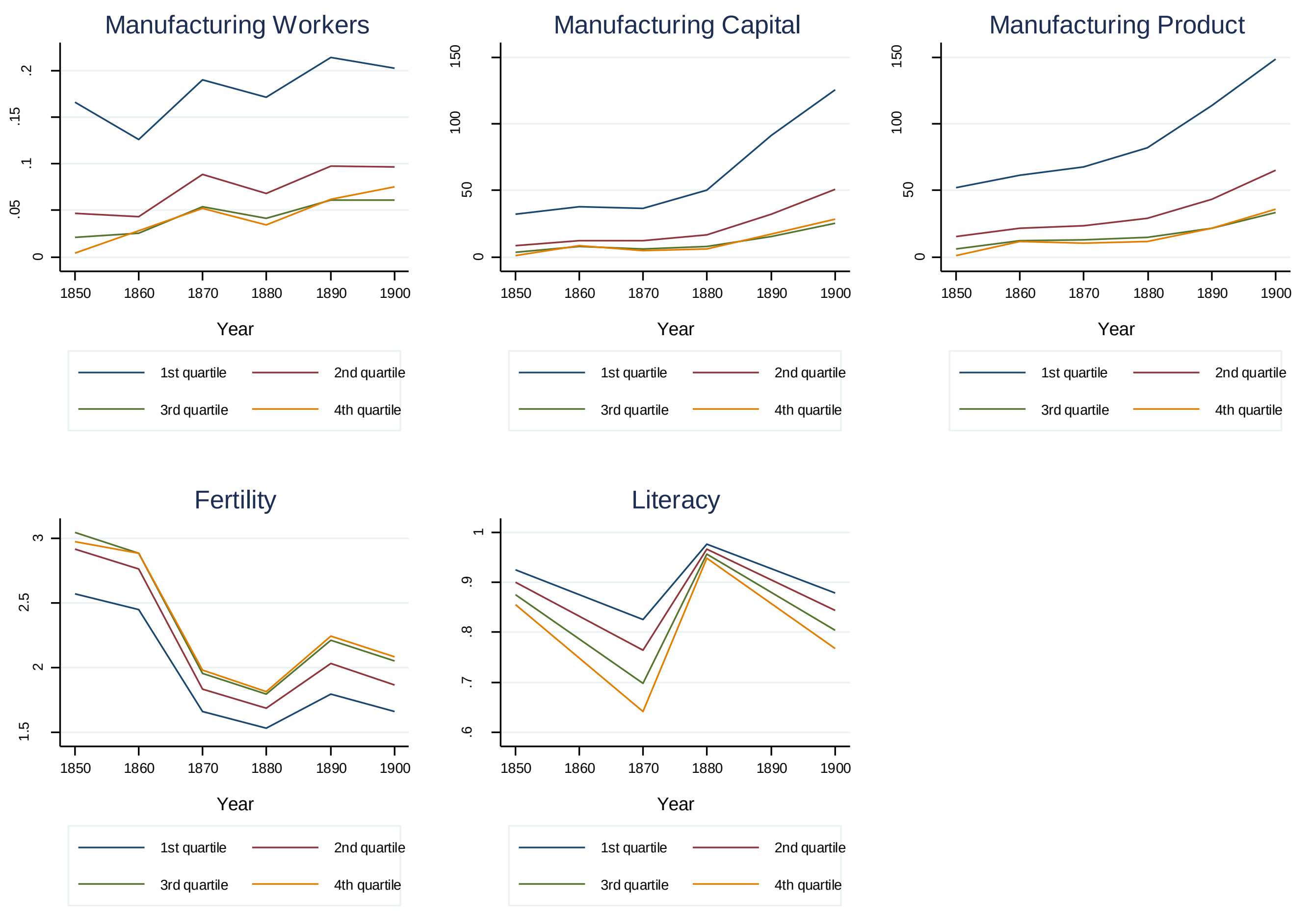


Figure15 : Divergence by Initial Levels of Industrialization



Tables

Table1 : Summary Statistics

	Variable	Mean	Median	Minimum	Maximum	Standard deviation
1850	Fertility (number of children per adult)	2.88	2.94	1.18	4.00	0.42
	Adult male Literacy rate	88.9%	91.8%	19.0%	100.0%	9.8%
	Share of adult male manufacturing workers	6.0%	3.1%	0.0%	81.2%	8.8%
	Capital invested in manufacturing per capita	11.18	4.65	0.00	198.82	18.40
	Value of manufacturing product per capita	18.84	7.58	0.00	402.05	30.70
	Minimum Distance to Nearest Connecting Line (kilometers)	124.03	64.85	0.05	914.07	144.09
	Total population	15,278	10,032	8	515,547	23,874
1900	Fertility (number of children per adult)	1.92	1.89	0.94	3.01	0.42
	Adult male Literacy rate	82.4%	85.6%	32.6%	98.6%	13.8%
	Share of adult male manufacturing workers	10.9%	6.5%	0.3%	60.0%	11.3%
	Capital invested in manufacturing per capita (1850 prices)	57.58	23.08	0.93	564.05	77.96
	Value of manufacturing product per capita (1850 prices)	70.92	34.36	0.55	1,130.30	93.22
	Minimum Distance to Nearest Connecting Line (kilometers)	93.86	31.68	0.05	880.06	124.19
	Total population	37,299	22,023	3,006	2,050,600	96,117

Notes:
1. See appendix A for variables definitions
2. The data is based on 1,490 counties east of the meridian 95° west longitude line
3. The averages are at the county level and do not represent average for all of the United States

Table2 : Top 10 Most Populated US Cities, 1850 and 1900

Rank	1850		1900	
	City	Residents	City	Residents
1	New York City	515,547	New York City	3,437,202
2	Baltimore	169,054	Chicago	1,698,575
3	Boston	136,881	Philadelphia	1,293,697
4	Philadelphia	121,376	St. Louis	575,238
5	New Orleans	116,375	Boston	560,892
6	Cincinnati	115,435	Baltimore	508,957
7	St. Louis	77,860	Cleveland	381,768
8	Albany	50,763	Buffalo	352,387
9	Pittsburgh	46,601	Cincinnati	325,902
10	Louisville	43,194	Pittsburgh	321,616

Note: the table does not include San Francisco (because wetern counties are not included in the main analysis) and cities which became neighbourhoods in other cities

Table3 : First Stage - The Effect of Distances on Industrialization

	(1)	(2)	(3)	(4)
Dependent Variable: Manufacturing Workers	All observations	All observations	All observations	Observations with literacy data
Minimum Distance to Nearest Connecting Line	-0.102***	-0.0333***	-0.0922***	-0.128***
	(0.0120)	(0.0123)	(0.0210)	(0.0217)
Minimum Distance to Nearest Large City		-0.624***	0.112***	0.179***
		(0.0335)	(0.0312)	(0.0354)
F test			14.89	24.18
County Fixed Effects	no	no	yes	yes
Year Fixed Effects	no	no	yes	yes
Observations	8,569	8,569	8,566	5,724

Standard errors are clustered at the county level

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

All variables are in logarithm except the dummies

F test robust to clustering, according to Olea, J. L. M., & Pflueger, C. (2013)

Table4 : Reduced Form – The Effect of Ditances on Fertility and Human Capital

	(1)	(2)	(3)
Panel A: Fertility			
Minimum Distance to Nearest Connecting Line	0.0145***	-0.00306	0.0280***
	(0.00221)	(0.00212)	(0.00312)
Minimum Distance to Nearest Large City		0.0962***	0.0524***
		(0.00485)	(0.00575)
Constant	0.700***	0.231***	0.654***
	(0.00873)	(0.0253)	(0.0308)
Panel B: Literacy			
Minimum Distance to Nearest Connecting Line	-0.0140***	0.00242	-0.0332***
	(0.00173)	(0.00176)	(0.00357)
Minimum Distance to Nearest Large City		-0.0886***	-0.0136**
		(0.00520)	(0.00654)
Constant	-0.137***	0.292***	0.0757**
	(0.00586)	(0.0267)	(0.0356)
County Fixed Effects	no	no	yes
Year Fixed Effects	no	no	yes

Standard errors are clustered at the county level

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

All variables are in logarithm except the dummies

Table5 : Railways and Connecting Lines, 1870

	(1)	(2)
Dependent Variable: Distance to Railway		
Minimum Distance to Nearest Connecting Line	0.183***	0.123***
	(0.0218)	(0.0237)
Minimum Distance to Nearest Large City		0.375***
		(0.0610)
Constant	1.496***	-0.355
	(0.0840)	(0.312)
Observations	1,490	1,490
R-squared	0.045	0.069

Standard errors in parentheses

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

All variables are in logarithm

Table6 : Validation Test - Current and Future Distances, 1850-1880

	(1) Manufacturing Workers	(2) Fertility	(3) Literacy	(4) Manufacturing Workers	(5) Fertility	(6) Literacy
Minimum Distance to Nearest Connecting Line	-0.0688***	0.0277***	-0.0428***			
	(0.0177)	(0.00270)	(0.00408)			
Future Minimum Distance to Nearest Connecting Line				-0.0201	0.0103	-0.0187
				(0.0510)	(0.00664)	(0.0145)
Minimum Distance to Nearest Large City	-0.0396	-0.00686	0.0426***	-0.0777***	0.00896	0.0171**
	(0.0308)	(0.00642)	(0.00868)	(0.0295)	(0.00633)	(0.00837)
County Fixed Effects	yes	yes	yes	yes	yes	yes
Year Fixed Effects	yes	yes	yes	yes	yes	yes
Observations	5,623	5,954	4,462	5,623	5,954	4,462

Standard errors are clustered at the county level

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

All variables are in logarithm except the dummies

Table7 : Main Results

	(1)	(2)	(3)	(4)
	OLS	IV	OLS	IV
Dependent Variable	Fertility	Fertility	Literacy	Literacy
Manufacturing Workers	-0.0227***	-0.315***	-0.00210	0.252***
	(0.00258)	(0.0693)	(0.00464)	(0.0487)
Minimum Distance to Nearest Large City	0.0591***	0.0823***	-0.0204***	-0.0527***
	(0.00567)	(0.0126)	(0.00622)	(0.0127)
Observations	8,566	8,564	5,724	5,721
First Stage F test		14.89		24.18
County Fixed Effects	yes	yes	yes	yes
Year Fixed Effects	yes	yes	yes	yes

Standard errors are clustered at the county level

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

All variables are in logarithm except the dummies

F test robust to clustering, according to Olea, J. L. M., & Pflueger, C. (2013)

Table8 : Alternative Measures for Industrialization - First Stage

	(1) All observations	(2) All observations	(3) All observations	(4) Observations with literacy data
Panel A: Manufacturing Product				
Minimum Distance to Nearest Connecting Line	-0.139***	-0.0257*	-0.0874***	-0.127***
	(0.0147)	(0.0147)	(0.0228)	(0.0233)
Minimum Distance to Nearest Large City		-0.624***	0.0742**	0.206***
		(0.0335)	(0.0354)	(0.0389)
F test			13.35	23.38
Panel B: Manufacturing Capital				
Minimum Distance to Nearest Connecting Line	-0.146***	-0.0452***	-0.111***	-0.159***
	(0.0155)	(0.0158)	(0.0242)	(0.0253)
Minimum Distance to Nearest Large City		-0.553***	0.104***	0.256***
		(0.0349)	(0.0375)	(0.0433)
F test			16.50	28.20
County Fixed Effects	no	no	yes	yes
Year Fixed Effects	no	no	yes	yes
Observations	8,569	8,569	8,569	5,727

Standard errors are clustered at the county level

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

All variables are in logarithm except the dummies

F test robust to clustering, according to Olea, J. L. M., & Pflueger, C. (2013)

Table9 : Alternative Measures for Industrialization - Main Results

	(1)	(2)	(3)	(4)
	OLS	IV	OLS	IV
Dependent Variable	Fertility	Fertility	Literacy	Literacy
Panel A: Manufacturing Product				
Real Manufacturing Product	-0.0340***	-0.332***	-0.00274	0.253***
	(0.00261)	(0.0837)	(0.00464)	(0.0513)
Minimum Distance to Nearest Large City	0.0588***	0.0717***	-0.0202***	-0.0596***
	(0.00565)	(0.0130)	(0.00620)	(0.0144)
Observations	8,569	8,567	5,727	5,724
First Stage F test		13.35		23.38
Panel B: Manufacturing Capital				
Real Manufacturing Capital	-0.0335***	-0.262***	0.00712	0.202***
	(0.00248)	(0.0526)	(0.00448)	(0.0361)
Minimum Distance to Nearest Large City	0.0595***	0.0742***	-0.0220***	-0.0595***
	(0.00572)	(0.0115)	(0.00625)	(0.0130)
Observations	8,569	8,567	5,727	5,724
First Stage F test		16.50		28.20
County Fixed Effects	yes	yes	yes	yes
Year Fixed Effects	yes	yes	yes	yes

Standard errors are clustered at the county level

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

All variables are in logarithm except the dummies

F test robust to clustering, according to Olea, J. L. M., & Pflueger, C. (2013)

Table10 : Different Number of Large Cities

	IV	IV	IV	IV	IV	IV
Number of "Large Cities"	5	7	9	11	13	15
Panel A: Fertility						
Manufacturing Workers	-0.438***	-0.348***	-0.517***	-0.271***	-0.625***	-0.756**
	(0.0679)	(0.0443)	(0.170)	(0.0499)	(0.230)	(0.333)
Minimum Distance to Nearest Large City	0.0306**	-0.00446	0.0494***	0.0745***	-0.0247	-0.0686
	(0.0129)	(0.0149)	(0.0151)	(0.0100)	(0.0299)	(0.0548)
First Stage F test	34.79	42.54	7.197	23.12	5.584	3.920
Panel B: Literacy						
Manufacturing Workers	0.324***	0.207***	0.303***	0.186***	0.574***	0.284
	(0.0625)	(0.0440)	(0.0748)	(0.0347)	(0.221)	(0.186)
Minimum Distance to Nearest Large City	-0.0618***	-0.0551***	-0.0360***	-0.0570***	0.0453	-0.0308
	(0.0127)	(0.0161)	(0.0108)	(0.0106)	(0.0336)	(0.0411)
First Stage F test	24.41	20.17	13.87	35.51	5.058	2.190
County Fixed Effects	yes	yes	yes	yes	yes	yes
Year Fixed Effects	yes	yes	yes	yes	yes	yes

Standard errors are clustered at the county level

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

All variables are in logarithm except the dummies

F test robust to clustering, according to Olea, J. L. M., & Pflueger, C. (2013)

Table11 : Alternative Specifications

Specification	IV No controls	IV Partial controls	IV Partial controls	IV Partial controls	IV Partial controls	IV No counties near large cities	IV Including the west	IV Only counties with literacy < 95%
Panel A: Fertility								
Manufacturing Workers	-0.129***	0.126	0.748	-0.693***	-0.482***	-0.305***	-0.557***	-0.228***
	(0.0158)	(0.0993)	(0.540)	(0.0494)	(0.123)	(0.0673)	(0.140)	(0.0842)
Minimum Distance to Nearest Large City		0.143***	0.424*	0.151***		0.138***	0.113***	0.0849***
		(0.0409)	(0.233)	(0.0257)		(0.0197)	(0.0235)	(0.0231)
First Stage F test	72.14	7.294	2.530	122.3	11.59	15.35	12.94	7.357
Observations	8,566	8,566	8,566	8,566	8,566	7,977	9,055	3,540
Panel B: Literacy								
Manufacturing Workers	0.130***	-0.0656	-0.0218	0.0161*	0.331***	0.258***	0.396***	0.473***
	(0.0162)	(0.0587)	(0.0913)	(0.00936)	(0.0728)	(0.0515)	(0.103)	(0.176)
Minimum Distance to Nearest Large City		-0.114***	-0.0935**	-0.0479***		-0.107***	-0.0694***	-0.0912*
		(0.0254)	(0.0407)	(0.00598)		(0.0206)	(0.0200)	(0.0483)
First Stage F test	82.93	9.545	3.044	135.4	17.94	24.43	12.99	7.357
Observations	5,724	5,724	5,724	5,724	5,724	5,324	6,038	3,540
County Fixed Effects	no	no	no	yes	yes	yes	yes	yes
Year Fixed Effects	no	no	yes	no	yes	yes	yes	yes

Standard errors are clustered at the county level

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

All variables are in logarithm except the dummies

F test robust to clustering, according to Olea, J. L. M., & Pflueger, C. (2013)

Table12 : The Effect of Immigration

	(1)	(2)
Group of Counties	Top 25% immigration	Top 10% immigration
Panel A: Fertility		
Manufacturing Workers	-0.369***	-0.352***
	(0.116)	(0.0828)
Group X Manufacturing Workers	0.106	0.340***
	(0.0825)	(0.0665)
Minimum Distance to Nearest Large City	0.0987***	0.113***
	(0.0218)	(0.0164)
First stage F test	4.683	8.411
Panel B: Literacy		
Manufacturing Workers	0.326***	0.267***
	(0.0969)	(0.0562)
Group X Manufacturing Workers	-0.126*	-0.113**
	(0.0733)	(0.0499)
Minimum Distance to Nearest Large City	-0.0785***	-0.0677***
	(0.0252)	(0.0170)
First stage F test	6.669	14.39
County Fixed Effects	yes	yes
Year Fixed Effects	yes	yes

Standard errors are clustered at the county level

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

All variables are in logarithm except the dummies

Table13 : Heterogeneity by Different Attributes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Group of Counties	Top 25% fertility	Bottom 25% fertility	Top 25% literacy	Bottom 25% literacy	Top 25% manufacturing workers	Top 25% manufacturing capital	Top 25% manufacturing product	Bottom 25% manufacturing workers	Bottom 25% manufacturing capital	Bottom 25% manufacturing product
	1850	1850	1850	1850	1850	1850	1850	1850	1850	1850
Panel A: Fertility										
Manufacturing Workers	-0.297***	-0.315***	-0.302***	-0.312***	-0.209***	-0.244***	-0.247***	-0.301***	-0.304***	-0.294***
	(0.0720)	(0.0683)	(0.0690)	(0.0676)	(0.0390)	(0.0479)	(0.0501)	(0.0482)	(0.0521)	(0.0475)
Group X Manufacturing Workers	-0.0526	0.127**	-0.0629	-0.0259	-0.512*	-0.215**	-0.205**	0.227***	0.223***	0.213***
	(0.0587)	(0.0560)	(0.0566)	(0.0898)	(0.290)	(0.0902)	(0.0821)	(0.0391)	(0.0396)	(0.0365)
Minimum Distance to Nearest Large City	0.0829***	0.0786***	0.0804***	0.0830***	0.0681***	0.0735***	0.0713***	0.0561***	0.0570***	0.0555***
	(0.0124)	(0.0112)	(0.0130)	(0.0133)	(0.0114)	(0.0111)	(0.0113)	(0.00960)	(0.0100)	(0.0100)
First stage F test	9.793	9.167	9.781	5.391	3.740	12.72	12.14	17.24	14.02	14.46
Panel B: Literacy										
Manufacturing Workers	0.250***	0.248***	0.283***	0.231***	0.145***	0.192***	0.195***	0.239***	0.244***	0.241***
	(0.0484)	(0.0490)	(0.0580)	(0.0594)	(0.0292)	(0.0374)	(0.0387)	(0.0369)	(0.0417)	(0.0392)
Group X Manufacturing Workers	0.00377	0.0319	-0.106***	0.400**	0.389**	0.154**	0.141***	-0.166***	-0.128***	-0.130***
	(0.0516)	(0.0434)	(0.0392)	(0.196)	(0.170)	(0.0612)	(0.0530)	(0.0319)	(0.0364)	(0.0350)
Minimum Distance to Nearest Large City	-0.0527***	0.0786***	-0.0575***	-0.0739***	-0.0452***	-0.0449***	-0.0439***	-0.0308***	-0.0365***	-0.0342***
	(0.0127)	(0.0112)	(0.0134)	(0.0208)	(0.0112)	(0.0112)	(0.0113)	(0.00946)	(0.0108)	(0.0109)
First stage F test	18.10	18.34	16.68	5.006	10.01	21.49	20.22	30.18	20.99	20.12
County Fixed Effects	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Year Fixed Effects	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes

Standard errors are clustered at the county level

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

All variables are in logarithm except the dummies

Table14 : Differences between Top-25% and Bottom-25% Industrialized Counties in 1850

Year	1850			1900		
	1st quartile	4th quartile	Difference	1st quartile	4th quartile	Difference
Share of adults employed in each industry:						
Footwear	23.2%	24.9%	0.0167	6.2%	2.9%	-0.0339***
Wood products	6.6%	13.2%	0.0652**	2.3%	3.0%	0.00737***
Furniture	5.3%	11.1%	0.0582**	1.8%	1.0%	-0.00868***
Transportation	4.5%	5.3%	0.00869	2.3%	2.9%	0.00667***
Yarn and fabric	5.5%	0.0%	-0.0552***	12.4%	4.4%	-0.0897***
Grain-mill products	3.2%	7.5%	0.0430**	0.7%	3.1%	0.0241***
Misc machinery	4.7%	1.2%	-0.0349***	8.1%	5.2%	-0.0292***
Leather products	3.3%	5.6%	0.0226	1.0%	1.9%	0.00937***
Printing and publishing	3.6%	2.2%	-0.0146	6.0%	5.2%	-0.00858***
Steel products	3.3%	0.6%	-0.0266***	4.7%	3.3%	-0.0143***
Sawmills and planing mills	2.0%	2.8%	0.00816	2.2%	29.2%	0.270***
Iron and Steel	2.4%	0.6%	-0.0176***	3.6%	1.9%	-0.0179***
Blast furnaces	1.9%	0.6%	-0.0130**	4.0%	1.6%	-0.0248***
Share of females employed in manufacturing	-	-	-	20.4%	8.4%	-0.127***
Share of literate adults employed in manufacturing	85.9%	81.6%	-0.0426	94.7%	87.7%	-0.0759***
Mean Occupation Score	25.74	25.474	-0.266	26.03	25.30	-0.736***

Notes:

1. The 1st quartile include individuals from counties which were in the top 25% in regards to the share of adult males in manufacturing in 1850. The 4th quartile include individuals from counties which were in the bottom 25%.

2. Adults are defined as individuals above age 16

3. Occupation score according to 1950 basis

4. *** p<0.01, ** p<0.05, * p<0.1

Appendix A: Main Variables Definitions

Table A1 presents the calculation of the main variables, based on the NHGIS data. Due to lack of data some variables are calculated differently in each year, and some of the years do not include literacy. However, this inconsistency between the periods is not a problem because of the year fixed effects included in the analysis.

Table A1 : Main Variables Definitions

Variable \ Year	1850	1860	1870	1880	1890	1900
Fertility	Children 5-19 / Females 20-39	Children 5-19 / Females 20-39	Children 5-18 / Males 18-44	Children 5-17 / Males 18-44	Children 5-20 / Males 18-44	Children 5-20 / Males 18-44
Literacy	1 - Percent of males over 20 who cannot read and write	-	1 - Percent of males over 21 who cannot write	Percent of males over 20 who can read and write	-	1 - Percent of Illiterate males over 21
Capital invested in manufacturing per capita	Capital invested in manufacturing / Total population	Capital invested in manufacturing / Total population	Capital invested in manufacturing / Total population	Capital invested in manufacturing / Total population	Capital invested in manufacturing / Total population	Capital invested in manufacturing / Total population
Value of manufacturing product	Value of manufacturing product / Total population	Value of manufacturing product / Total population	Value of manufacturing product / Total population	Value of manufacturing product / Total population	Value of manufacturing product / Total population	Value of manufacturing product / Total population
Percent of manufacturing workers	Employed in manufacturing / Males over 14	Males employed in manufacturing / Males over 14	Males over 16 employed in manufacturing / Males over 21	Males over 16 employed in manufacturing / Males over 21	Males over 16 employed in manufacturing / Males over 21	Males over 16 employed in manufacturing / Males over 21

References

Atack, J., Bateman, F., Haines, M., & Margo, R. A. (2010). Did railroads induce or follow economic growth?. *Social Science History*, 34(02), 171-197.

Bairoch, P. (1982), "International Industrialization Levels from 1750-1980", *Journal of European Economic History* 11: 269-333.

Becker, G. S. (1960). An economic analysis of fertility. In Demographic and economic change in developed countries (pp. 209-240). *Columbia University Press.*

Becker, G. S., & Lewis, H. G. (1974). Interaction between quantity and quality of children. In Economics of the family: Marriage, children, and human capital (pp. 81-90). *University of Chicago Press.*

Becker, S. O., & Woessmann, L. (2009). Was Weber wrong? A human capital theory of Protestant economic history. *The Quarterly Journal of Economics*, 531-596.

Becker, S. O., Cinnirella, F., & Woessmann, L. (2010). The trade-off between fertility and education: evidence from before the demographic transition. *Journal of Economic Growth*, 15(3), 177-204.

Bleakley, H., & Lange, F. (2009). Chronic disease burden and the interaction of education, fertility, and growth. *The review of economics and statistics*, 91(1), 52-65.

Clark, G. (2003). The great escape: The industrial revolution in theory and history. *University of California, Davis, Working Paper*, September.

Easterlin, R. A. (1981). Why isn't the whole world developed?. *The Journal of Economic History*, 41(01), 1-17.

Fernández, R. (2014). Women's rights and development. *Journal of Economic Growth*, *19*(1), 37-80.

Fishlow, A. (1965). American Railroads and the Transformation of the Ante-bellum Economy (Vol. 127). *Cambridge, MA: Harvard University Press*.

Fishlow, A. (1966). Levels of nineteenth-century American investment in education. *The Journal of Economic History*, *26*(04), 418-436.

Flora, P., F. Kraus and W. Pfenning (1983), State, Economy and Society in Western Europe 1815-1975, *vol. 1 (St. James Press, Chicago)*.

Franck, R., & Galor, O. (2015a). The Complementary between Technology and Human Capital in the Early Phase of Industrialization.

Franck, R., & Galor, O. (2015b). Industrialization and the Fertility Decline. *Brown University.*

Galor, O. (2005). From stagnation to growth: unified growth theory. *Handbook of economic growth*, 1, 171-293.

Galor, O. (2012). The demographic transition: causes and consequences. *Cliometrica,* 6(1), 1-28.

Galor, O., & Moav, O. (2002). Natural selection and the origin of economic growth. *Quarterly Journal of Economics*, 1133-1191.

Galor, O., & Moav, O. (2006). Das human-kapital: A theory of the demise of the class structure. *The Review of Economic Studies*, 73(1), 85-117.

Galor, O., & Weil, D. N. (1996). The Gender Gap, Fertility, and Growth. *The American Economic Review*, 374-387.

Galor, O., & Weil, D. N. (1999). From Malthusian stagnation to modern growth. *The American Economic Review*, 89(2), 150-154.

Goldin, C., & Katz, L. F. (1996). The origins of technology-skill complementarity (No. w5657). *National bureau of economic research.*

Goldin, C., & Katz, L. F. (2003). The" virtues" of the past: Education in the first hundred Years of the new republic (No. w9958). *National Bureau of Economic Research.*

Gordon, R. J. (2016). The rise and fall of American growth: The US standard of living since the civil war. *Princeton University Press.*

HAINES, M. R. (1998): "Estimated Life Table for the United States, 1850–1910," *Historical Methods, 31, 149–167.*

Haines, M. R., & Steckel, R. H. (2000). A population history of North America. *Cambridge University Press.*

Hazan, M. (2009). Longevity and lifetime labor supply: Evidence and implications. *Econometrica,* 77(6), 1829-1863.

Katz, L. F., & Margo, R. A. (2013). Technical change and the relative demand for skilled labor: The united states in historical perspective. In Human Capital in History: The American Record (pp. 15-57). *University of Chicago Press*.

Katz, M. B., Burke, C. B., & Hall, P. D. (1983). The role of American colleges in the nineteenth century.

Klemp, M. P., & Weisdorf, J. L. (2010). The child quantity-quality trade-off: evidence from the population history of England. *University of Copenhagen, mimeo.*

Landes, D. S. (2003). The unbound Prometheus: technological change and industrial development in Western Europe from 1750 to the present. *Cambridge University Press.*

Matthews, R. C. O., Feinstein, C. H., & Odling-Smee, J. (1982). British economic growth 1856-1973: the post-war period in historical perspective. *Oxford University Press.*

Michaels, G. (2008). The effect of trade on the demand for skill: Evidence from the interstate highway system. *The Review of Economics and Statistics*, 90(4), 683-701.

Mokyr, J. (1992). The lever of riches: Technological creativity and economic progress. *Oxford University Press.*

Murphy, T. E. (2010). Old Habits Die Hard (Sometimes). Can département heterogeneity tell us something about the French fertility decline. *Bocconi University Innocenzo Gasparini Institute for Economic Research Working Paper, 364.*

Murtin, F. (2013). Long-term determinants of the demographic transition, 1870–2000. *Review of Economics and Statistics*, 95(2), 617-631.

Olea, J. L. M., & Pflueger, C. (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics*, 31(3), 358-369.

Pleijt, A. M. D., Nuvolari, A., & Weisdorf, J. (2016). Human Capital Formation during the First Industrial Revolution: Evidence from the Use of Steam Engines (No. 294). *Competitive Advantage in the Global Economy (CAGE).*

Rosenberg, N., & Trajtenberg, M. (2004). A general-purpose technology at work: The Corliss steam engine in the late-nineteenth-century United States. *The Journal of Economic History*, *64*(01), 61-99.

Taylor, G. R. (1951). The transportation revolution, 1815-60. *Routledge.*