

NBER WORKING PAPER SERIES

THE RETURN TO EDUCATION IN THE MID-20TH CENTURY:  
EVIDENCE FROM TWINS

James J. Feigenbaum  
Hui Ren Tan

Working Paper 26407  
<http://www.nber.org/papers/w26407>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
October 2019

For detailed feedback and helpful suggestions, we thank Ran Abramitzky, Michael Andrews, Brian Beach, Kevin Lang, Erika Lee, Bob Margo, Ross Mattheis, Christopher Muller, Ee Cheng Ong, Jessica Pan, Hanna Schwank, Kelvin Seah, and Marybeth Train. All errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2019 by James J. Feigenbaum and Hui Ren Tan. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Return to Education in the Mid-20th Century: Evidence from Twins  
James J. Feigenbaum and Hui Ren Tan  
NBER Working Paper No. 26407  
October 2019  
JEL No. J2,J3,N3,N32

### **ABSTRACT**

What was the return to education in the United States at mid-century? In 1940, the correlation between years of schooling and earnings was relatively low, less than it had been in 1915 or than it would be in later decades. In this paper, we estimate the causal return to schooling in 1940, constructing a large linked sample of twin brothers to account for differences in unobserved ability and family background. Though imperfect, the twins identification strategy allows us to compare the return to education to recent studies implemented similarly. We find that the return to education was relatively low in 1940, with each additional year of schooling increasing labor earnings by approximately 4%. Returns to education were evident both within and across occupations and were higher for sons born to lower SES families.

James J. Feigenbaum  
Department of Economics  
Boston University  
270 Bay State Road  
Boston, MA 02215  
and NBER  
jamesf@bu.edu

Hui Ren Tan  
National University of Singapore  
huiren@nus.edu.sg

## 1 Introduction

The American labor market was at a crossroad of major economic events and transitions in the middle of the 20th century. With the first generation of workers educated in America’s rapidly expanding high school system entering the labor force, the human capital century was well underway (Goldin and Katz 2009). While the aftershocks of the Great Depression were still reverberating, the economic and social upheaval of WWII had yet to hit the country. The Great Compression—a period of declining wage dispersion that saw earnings inequality by education, experience, and occupation all shrink—was also just beginning (Goldin and Margo 1992). An important feature of any economy is the relative price paid for skilled labor or educated workers and is particularly informative at a moment when both supply and demand for educated labor were changing dramatically.

What was the return to schooling at mid-century? Mincerian regressions of earnings on years of schooling suggest that the returns may have been low: Goldin and Katz (2009) trace out a U-shaped pattern in the skill premium that reaches its nadir at mid-century. The high school premium was lower in 1940 than it had been in 1915, while the college premium was lower in 1940 than it is today.<sup>1</sup> These Mincerian education coefficients, however, are difficult to interpret causally, particularly in the presence of ability or selection bias.

In this paper, we find that the return to schooling in 1940 was positive but smaller in magnitude compared with the late 20th century, consistent with the trends Goldin and Katz (2009) document. To identify the causal effect of education on earnings, we construct a large linked sample of twin brothers who grew up in the same families and compare their education and earnings in adulthood. If twins have similar innate characteristics or abilities, then a within-twins comparison can help address confounding effects of unobserved differences.

---

<sup>1</sup>Differences in data make these comparisons across time challenging. The 1915 estimate of the rate of return is based on the 1915 Iowa State Census (Goldin and Katz 2000), while estimates in 1940 are based only on labor and wage earnings, rather than income, limited by the questions asked on the 1940 census. Comparisons over time are also tricky because of the growth in educational attainment: high school and college graduates in the past are in very different parts of the education distribution compared with the present.

There are two main advantages to using the within-twins estimator in a historical setting. First, the same approach has been applied to more recent data to estimate the causal return to schooling. To the extent that biases in the twins approach are similar in the past and present, this allows us to compare our estimates with recent twin studies and to study changes over more than 50 years. Recent work identifying the return to education from twins has shown that selection bias is relatively small but whether that holds in the past as well—which might allow us to gauge how similar correlational returns are to causal returns—is unclear without our study. Second, the availability of the complete historical census records enables us to construct samples of twins that are both larger in size and wider in geographic coverage than contemporary studies, many of which rely on small samples from local surveys or state-specific registers.<sup>2</sup>

We undertake our analysis, however, recognizing the limitations of the twins methodology. Some of these weaknesses are common across twins papers, while others are specific to the historical setting and data. One general critique of all twins studies is that even identical twins are not exactly the same; after all, the method only “works” if twin brothers complete different years of education from one another. Such differences could compromise the identifying assumption of the within-twins estimator. Both the size of our twins sample and the availability of names in the historical censuses allow us to introduce new tests on the extent of ability bias in our estimates, tests that are not always feasible with contemporary data. While we cannot completely rule out ability differences between twins, the evidence as a whole suggests that our baseline findings are unlikely to be driven by such issues.

Studying twins in a historical setting presents its own distinct challenges, some of which we are able to address in robustness checks. First, because we need to link twins from one census—when they are children—to 1940—when the twins are adults and unlikely to be living in the same household as their twin sibling—without unique individual identifiers, this

---

<sup>2</sup>As we will detail in the data section, our sample of twins is an order of magnitude larger than the contemporary American samples analyzed by, for example, Ashenfelter and Krueger (1994) or Ashenfelter and Rouse (1998). Our sample is drawn from the full population enumerated in the census across the entire country.

necessarily introduces some incorrect matches and measurement errors in our analysis. Reassuringly, we show that our results are robust to alternative linking methods.<sup>3</sup> Second, we can only link men—a common constraint in the census linking literature—and are thus able to estimate the return to education from male-male twin pairs only. Third, contemporary twins studies often exploit a twin’s report of his or her sibling’s education to address measurement error in self-reported schooling. Such information is not available in our setting, though we provide suggestive evidence that measurement error in education is unlikely to be driving our results. Fourth, it is not possible to distinguish monozygotic (MZ) and dizygotic (DZ) twins in the historical data. MZ twins are the preferred study population as they share both the same genes and environment. However, because many same-sex twins can be wrong about their identical-fraternal status—for example, nearly one-fifth of twins were wrong in Add Health (Conley and Fletcher 2017, Chapter 2)—MZ and DZ status may not be well-measured in contemporary studies either. Beyond the genetic-similarity that leads researchers to study MZ twins, we introduce measures of twin name similarity that may capture how similarly parents plan to treat their twins. This is a measure of “identicalness” in twin nurture rather than nature.<sup>4</sup>

We find that the return to education in the US was positive in 1940, but smaller than estimates for more recent periods. Contemporary returns tend to be relatively large across different identification strategies (Card 1999), including within-twins estimates (Ashenfelter and Krueger 1994; Ashenfelter and Rouse 1998; Rouse 1999; Behrman and Rosenzweig 1999), quarter of birth instruments (Angrist and Krueger 1991), or distance to college instruments

---

<sup>3</sup>Bailey et al. (2017) and Abramitzky et al. (2019) discuss the current state of the field in historical census linking. Bailey et al. (2017) emphasizes the unrepresentativeness of linked samples relative to the starting samples which we address here using their preferred inverse propensity weighting solution. Abramitzky et al. (2019) describe the most common, current generation linking methods and conclude that, while methods differ in their propensity to minimize false positive matches and false negative non-matches, research conclusions are generally robust across the methods. In this paper, we show that our results are robust to variants of the main machine learning based linking method (Feigenbaum 2016) that we use. In addition our results are similar when studying twins linked by the Abramitzky et al. (2012) method.

<sup>4</sup>Another challenge to any estimation of the return to education in 1940 is the earnings measure in the 1940 Federal Census. As we describe in more detail in the data section, census respondents were asked to report their wage and salary earnings, so we do not measure total income.

(Card 1993). Our smaller historical results, juxtaposed with the larger contemporary returns, are consistent with the U-shaped trend in Goldin and Katz (2009). We also find that the positive returns in the past are driven both by the access to “better” jobs from more education and from higher earnings within occupations, with the latter driving two-thirds of the effect.

Our large sample of twins also enables us to estimate the return to education across four important background characteristics, shared by twins but differing in our sample. First, we find that the returns are highest for the older cohorts in our sample. These differences are modest and could be driven by cohort-specific differences or interactions between experience and education. Second, returns are higher for twin sons whose fathers were farmers or were of lower socio-economic status. Third, the return to schooling varies by family immigration history: twin brothers with more foreign-born grandparents earn relatively less for the same level of education. Fourth, the return to schooling is reasonably similar across birth regions.<sup>5</sup>

Studying alternative economic measures beyond earnings, we see returns to education in occupational status, non-wage earning status, and home ownership, as well as increases in labor supply. Twins with more education are also more likely to migrate and to have more children, though the magnitudes of these effects are relatively small.

Our paper is closest to Clay et al. (2016), who exploit early 20th-century changes in compulsory schooling laws (CSLs) to instrument for actual education with compelled years of schooling, differencing out cohort and state-of-birth fixed effects.<sup>6</sup> To estimate the CSL effects, Clay et al. (2016) undertook a massive data collection, compiling and updating CSLs across states from 1880 to 1930. They find returns of 6.4 to 7.9% based on the 1940 census, which overlaps with the lower range of estimates for more recent cohorts. The CSL-based identification strategy also enables Clay et al. (2016) to estimate returns to education throughout the earnings distribution with quantile regressions. And, though Clay et al.

---

<sup>5</sup>Because location choice during adulthood is endogenous to education, we do not estimate differences in the return to education across 1940 labor markets.

<sup>6</sup>Ward (2019a) studies the return to internal migration in the early 20th century. To benchmark the importance of geographic mobility, he compares migration with the return to schooling, which he estimates to be 5.5% using a within-brothers (not twins) analysis. These results line up with our finding that the return to education was positive but smaller in 1940 than today.

(2016) focus only on white men, the CSL method could enable estimates of the return to education for women or minority groups who are difficult to accurately link across censuses. However, unlike the twins approach, it is difficult to make a direct comparison between their results and contemporary studies. There are two reasons for this. First, Stephens and Yang (2014) show that the return to schooling estimated in contemporary CSL studies can be eliminated by allowing birth year effects to vary by region. This would deviate from most other recent estimates of the return to education. Second, intertemporal comparisons are challenging with the CSL approach because education levels have changed over time. When educational attainment in the population rises faster than CSL thresholds—as dramatically occurred in the US over the 20th century (Goldin and Katz 2009)—the CSL compliers or those on the margin of treatment changes as well. Clay et al. (2016) show that the early CSLs primarily affected people in the later years of common school with around six, seven, or eight years of education. Compliers of more recent CSLs will instead be in high school. In contrast, because twinning is closer to random—at least prior to recent advances in fertility treatment (Kulkarni et al. 2013)—the study population in the past and present may be more comparable and the estimated treatment effect is less local as we see people throughout the educational distribution.

The paper proceeds as follows. In the second section, we describe our historical sample of twins, including details on the complete count census data we draw on and the census to census linking method we use. In section three, we present our estimated return to education in 1940 and several robustness tests. We also show that the returns to education in 1940 were driven more by effects of education within occupations rather than across occupations. Exploiting our large sample of twins, section four explores how the return to schooling in 1940 varies by cohort, family SES, family immigration history, and geography. In the fifth section, we estimate the effects of education on migration and other measures of economic status for our sample of twins. We conclude in the sixth section.

## 2 Historical Sample of Twins

To estimate the return to education using twins, we need information on the completed schooling and adult earnings of twin pairs. In this section, we describe our data and linking procedure.

### 2.1 Complete Count Census Data

Our main variables of interest, years of schooling and weekly earnings, come from the 1940 Federal Census. We estimate the returns to education in 1940 and not another year for three reasons. First, education and earnings were not recorded in federal censuses before 1940.<sup>7</sup> Second, full access to the complete census enumerations, including the names that are essential for linking individuals across censuses, is restricted for privacy for 72 years after a given census was taken. The 1940 records are thus the latest available full counts. Third, the return to schooling in the middle of the twentieth century is inherently interesting. Mincerian estimates of education returns in 1940 are low, possibly due to the influx of educated workers coming out of the high school movement (Goldin 1998), even as demand for skilled workers was growing during the middle of the human capital century.

The 1940 census measures educational attainment as the “highest grade of school completed.” Enumerators were instructed to ask the education question of everyone and not to include half years or unfinished grades. The average worker had about 9 years of schooling in 1940 (Goldin and Katz 2009), while 28.2% were high school graduates and 4.3% were college graduates.<sup>8</sup> In practice, years of education is top-coded at five or more years of college, but this affects just 1.2% of the population.

Our main variable of interest is the log of weekly earnings. We construct weekly earnings by dividing annual labor earnings by the reported number of weeks worked. We note five

---

<sup>7</sup>While some states, notably Iowa in 1915 and 1925, did record education and earnings in state censuses before 1940, these censuses are not useful for us. We would only observe outcomes for pairs of twins who both live in the same state as adults, potentially biasing the sample in problematic ways. Along these lines, Feigenbaum (2018) documents the biases to estimating intergenerational mobility in a sample restricted to children who do not move out of state.

<sup>8</sup>We calculated high school and college graduate shares from IPUMS 1% samples.



important points about this measure. First, both annual earnings and weeks worked refer to the 1939 calendar year, asked when the census was conducted in April 1940. Second, unlike later censuses, the 1940 census asked only about annual earnings from wages.<sup>9</sup> To the extent that education also affected business earnings or other forms of earning, we will miss this.<sup>10</sup> Third, wage responses were top-coded at \$5,000, though only 1% of our twins sample was top-coded.<sup>11</sup> Fourth, with just one year of labor earnings, we have a noisy measure of permanent earnings, which may be the more relevant concept when estimating the return to schooling. This is a common issue in the twins literature. We thus complement our baseline analysis with alternative measures of economic status, including occupation scores and home ownership (Table 12). These alternatives are not perfect, but the evidence converges in the same direction. Fifth, the number of weeks worked may also be measured with error.<sup>12</sup>

## 2.2 Locating Twins in the Census

We identify twin brothers in the 1900, 1910, and 1920 censuses. Twin brothers are defined as any pair of male siblings living in the same household who have the same last name, age in years, birthplace, and relationship to the household head.<sup>13</sup> As is common in most analyses with linked historical data, we focus on men because women tend to change their last names upon marriage in this period, making women difficult to track across censuses. Thus, our sample excludes any girl-girl or boy-girl twin pairs. By using the full counts, we are able to observe the universe of boy-boy twins in each base year. Our initial sample size compensates for the rarity of twin births and imperfections in record linkage, both of which shrink the ultimate sample size.

[Figure 1 about here.]

---

<sup>9</sup>Enumerators asked for the “amount of money wages or salary received (including commissions).”

<sup>10</sup>As we show in Table 12, twins with more education were more likely to report at least \$50 in non-wage earnings in 1939, our only crude measure of such earnings in the 1940 census.

<sup>11</sup>Technically, wages should be top-coded at \$5,001 in the 1940 census, but the share of respondents reporting exactly \$5,000 suggests that in practice top-coding was done at \$5,000.

<sup>12</sup>The prompt for weeks worked was: “Number of weeks worked in 1939 (equivalent full-time weeks).”

<sup>13</sup>We use the complete count census data prepared by Integrated Public Use Microdata Series (IPUMS) (Ruggles et al. 2010). In the IPUMS data, these are siblings matching on `serial`, `famunit`, `age`, `bpl`, `relate` and `namelast`.

From the full counts, we identify approximately 900 thousand children, aged 0 to 25, in each decennial census who have a twin sibling. Of these twinned children, more than 200 thousand are in boy-boy pairs in each census.<sup>14</sup> We plot the rate of twins overall in Figure 1 and describe the rate and number of twins by census year in Table 1. As the figure makes clear, we are more likely to identify twins among younger children or those born closer to the enumeration year. This pattern reflects our procedure for identifying twins: children need to be residing in their childhood households with their same-age siblings before they can be tagged as twins, but older children are more likely to have left home, leading us to miss out twins at older ages.

Is the frequency of twinning high or low historically? As we document in Table 1, we identify approximately 15 twins per 1,000 people. This is likely to be an underestimate, as our sample includes everyone under 25 and we are unable to pick out twins if either twin has left the childhood household. Per Figure 1, the twin rate is between 16 and 18 twins per 1,000 among younger children. These are still lower than the contemporary rate of 33.3 twins per 1,000 (Martin et al. 2018). There are two reasons for this. First, the advent of fertility treatments raised the likelihood of twin births substantially. Kulkarni et al. (2013) estimate that by 2011, 36% of twin births in the US were due to fertility treatments such as in-vitro fertilization (IVF). Second, child mortality rates are lower today than they were in the early 20th-century US (Preston and Haines 1991). Because our twin identification requires both twins to be in their childhood household when the decennial census is taken, children born as twins but with a deceased sibling will *not* be identified as twins.

[Table 1 about here.]

---

<sup>14</sup>As we collect twins in decennial censuses using an age range of 0 to 25, it is possible that we will include a twin pair more than once in our universe of 200 thousand boy-boy pairs. For example, twins born in 1899 would be 1 in 1900, 11 in 1910, and 21 in 1920, at risk of being observed two or three times. To address this over-sample, we present robustness results in Table A.3. Our estimated return to education estimated in the full baseline sample lines up with the returns estimated in alternative samples built from non-overlapping ages in each of the three initial-year censuses.

## 2.3 Record Linkage

We construct the linked sample of twins using the machine learning approach introduced by Feigenbaum (2016). This procedure begins by searching for the space of all potential matches based on name, birthplace, race, and implicit year of birth.<sup>15</sup> A random subsample is then drawn and manually matched or “trained” by a human researcher. Humans tend to be reasonably good at identifying links, even on messy data, but the rules they use to do so are opaque and difficult to write down. The machine learning method makes the implicit importance of various record features in determining a match explicit, capturing the weights on different features as covariates in a probit model. The resulting estimates are used to generate probabilistic scores for all potential matches. To be considered a true match, these scores need to be the sufficiently high both in absolute terms and relative to any alternative options.<sup>16</sup> This produces a preliminary linked sample of 312,369 individuals with a match rate of 45%.<sup>17</sup> Later, we show that our results are robust to altering how conservative the machine learning procedure is, and to using the more classic approach to record linkage outlined by Abramitzky et al. (2012).

## 2.4 Baseline Sample

We impose three restrictions on the preliminary linked sample. First, only twin pairs where both brothers can be linked to 1940 are kept, since the return to schooling will be identified

---

<sup>15</sup>In all waves, the census records age in years rather than year of birth. We can estimate likely year of birth from the age, but given that censuses are taken on different dates in each wave (June 1 in 1900, April 15 in 1910, January 1 in 1920, and April 1 in 1940), this adds some noise to the linking process. We use the year and month of birth question, asked only in 1900 during our sample, to validate our twins construction and show that our conclusions are unlikely to be driven by imprecision in twin tagging (see Appendix A.2).

<sup>16</sup>In our baseline sample, we use a relatively strict threshold to tamp down on false positives. To choose the two hyperparameters governing matching—how absolutely good a link has to be and how relatively good a link has to be—we use 10-fold cross validation, picking hyperparameters that maximize the out of sample weighted sum over accuracy (positive predictive value (PPV)) and recall (true positive rate (TPR)) with a weight of 3 on PPV and 1 on TPR. In robustness, we vary this weight from 1 (least conservative on accuracy but most likely to recall true matches) to 10 (most conservative).

<sup>17</sup>Our match rate is comparable to the success rates of other studies that also create their own historical linked samples. For example, Abramitzky et al. (2012) match 29 percent of Norwegian men from the 1865 Norwegian census to the 1900 Norwegian and US censuses using an iterative matching procedure.

from within-twins variation.<sup>18</sup> Second, we limit the sample to twin pairs where both brothers are wage and salary workers in 1940. As stated earlier, the 1940 census records earnings from wages but not other sources of income. Consequently, the earnings of self-employed persons are likely to be understated.<sup>19</sup> Third, to ensure that our results are not driven by outliers, both brothers are required to have worked a positive number of weeks in the previous year, a positive number of hours in the preceding week, and to have earned at least \$6 a week.<sup>20</sup> The final dataset comprises 38,652 individuals or 19,326 pairs of twins.

Our twins sample is large compared with contemporary work using twins to estimate the return to schooling. Ashenfelter and Krueger (1994) and Ashenfelter and Rouse (1998), for example, have 149 and 340 pairs of twins, respectively. Drawing on data from the UK, Bonjour et al. (2003) study 214 pairs. Studies based on Scandinavian registry data tend to have larger samples of twins. Bingley et al. (2009), for instance, have information on 4,809 pairs of Danish twins, while the data in Isacsson (2004) comprise 6,210 pairs of Swedish twins.

How does the sample of twins compare with the general population? We compare our sample both as children (Table 2) and as adults (Table 3) and find that while our analysis sample of linked twins is slightly whiter and significantly less likely to be foreign-born than the general population, twins are not an extreme subset of the population.

[Table 2 about here.]

First, we compare twins and their families in 1900, 1910, and 1920 to their cohort-mates in Table 2. To draw cohort-mates, we pull a 1% random sample from all boys under 25 who

---

<sup>18</sup>Twin pairs where one or both members have missing information on years of education are also dropped.

<sup>19</sup>In Appendix Figure A.4, we show that our results are robust to including these non-wage earners, as well as relaxing other restrictions on the sample. In Table 12, we also show that there is a return to education on the extensive margin of having non-wage income when we regress the indicator for more than \$50 dollars of non-wage earnings in 1940 on education. Twin pairs where one or both brothers were employed in emergency relief work in 1940 are excluded.

<sup>20</sup>The \$6-a-week threshold follows Goldin and Margo (1992), who focus on individuals earning more than one-half the minimum wage on a full-time basis. Clay et al. (2016) make similar sample restrictions studying the effects of compulsory schooling laws on labor earnings in 1940 to exclude non-wage earners.

are not the head or head’s spouse in the 1900, 1910, and 1920 censuses.<sup>21</sup> We note three differences here. First, twins tend to be younger, likely because the likelihood of identifying twins declines with age as individuals leave their childhood households. Second, though only 86% of twins in the full population (column 1) are white, our linked and analysis sample is significantly whiter, rising to 95% of the twin pairs we will include in our analysis. This may be driven by the relative difficulty of linking non-white records across censuses. Third, and unsurprisingly, twins are likely to have more siblings than a random age-mate from the population. These three differences, coupled with other small differences in Table 2, suggest caution when extrapolating the returns for twins to the general population.

We also find twins to be broadly similar to their cohort-mates during adulthood. Table 3 compares our twins with individuals in a 1940 1% IPUMS sample. For consistency, the comparison group is restricted to male wage and salary workers aged 17-68, and the same sample restrictions are imposed on wages, weeks worked, and hours worked as in the baseline. Linked twins are more likely to be married and less likely to be foreign-born. The latter could reflect the fact that a substantial share of immigrants Americanize their names (Biavaschi et al. 2017), which lowers the odds of matching them across censuses by name.

[Table 3 about here.]

In the Appendix, we show that the education and earnings distribution for our linked twins overlaps closely with the distributions in the 1940 1% IPUMS sample (Figures A.2 and A.3).

### **3 Returns to Schooling in the First Half of the 20th Century**

Comparing twin brothers born in the late 19th and early 20th centuries, we find that a return to schooling of about 4%, which is about half of what contemporary twin studies in the US find. In this section, we describe our empirical strategy more formally, present the main results, and show that our results are robust to a variety of concerns about linked data,

---

<sup>21</sup>We also limit the sample to individuals not in group quarters.

measurement error, and the twins method. In the final subsection, we present some evidence that the returns to education accrued both within and across occupations.

### 3.1 Empirical Strategy

The twins identification strategy has a long history in labor economics as a method of estimating the causal return of education. Naive comparisons of earnings and schooling may overstate the rate of return, as people with more schooling may also have higher unobserved ability. An imperfect solution is to compare siblings within a household to reduce the bias from nurture-induced differences in ability. The within-twins comparison goes one step further. Monozygotic twins have near-identical genetic makeups, allowing researchers to “control” for nature and nurture.<sup>22</sup> Dizygotic twins are no more genetically related than any pair of siblings but are of the same age. They will thus be subject to common cohort-specific time-varying shocks and share a more similar family environments than siblings of different ages.

To determine how an additional year of schooling affects earnings, we implement the following regressions with our linked sample of twins:

$$\log W_{ih} = \alpha + \beta^{OLS} \cdot Schl_{ih} + X_{ih} + \varepsilon_{ih} \quad (1)$$

$$\log W_{ih} = \alpha + \beta^{FE} \cdot Schl_{ih} + \gamma_h + X_{ih} + \varepsilon_{ih} \quad (2)$$

for individual  $i$  and household  $h$ . The outcome is the log weekly wage,  $\log W$ , and the main independent variable is the years of schooling,  $Schl$ , based on the highest grade completed. Both variables are measured in 1940. The first regression is a simple OLS benchmark that uses variation across and between all twin brothers. The second regression implements our twins design by adding in family fixed effects,  $\gamma_h$ , exploiting only variation between twins to estimate the return to education.

---

<sup>22</sup>Recent studies have questioned how identical monozygotic twins are. Fraga et al. (2005) document epigenetic differences between MZ twins and Bruder et al. (2008) locates genetic differences between MZ twins.

What controls do we include? We partition controls,  $X$ , in two groups. The first is a vector of predetermined characteristics: a quadratic in age, a race dummy, and an indicator for nativity status. As these controls will not vary between twins, they are effectively dropped in our preferred specification. The second group of covariates are: full-time employment status, employment tenure, marital status, number of children, and region of residence, all drawn from the 1940 census. These are endogenous controls as they may be outcomes of schooling itself and are measured or determined contemporaneously with earnings. We include the second set of controls in some specifications for consistency with the existing twins literature, much of which predates Angrist and Pischke (2009) formalizing the concept of “bad” controls. Our focus, however, will be on models without endogenous controls.

The key identifying assumption of the within-twins approach is that twin brothers have the same innate ability.<sup>23</sup> We conduct several indirect tests for the validity of this assumption after presenting the baseline results.

We weight all regressions with our linked sample to account for differential difficulty in linking records census to census. Given the lack of unique individual identifiers in the historical records and the limited covariates that are available for matching, historical linked samples are necessarily imperfect representations of the underlying populations. Following Bailey et al. (2017), we use inverse propensity weights to adjust for observable differences between matched and unmatched persons. We describe the weighting procedure in Appendix A.1.

## 3.2 Main Results

What is the return to education in 1940? Figure 2 summarizes the results, plotting twin differences in earnings against twin differences in education: an additional year of schooling raises earnings by 0.043 log points, an increase of more than 4%. To estimate the return to

---

<sup>23</sup>Conley and Fletcher (2017, Chapter 4), comparing cohorts born in the US from 1920 to 1955, document that the predictive power of polygenic scores on educational attainment is fairly stable. If anything, genes may have been more important for the oldest cohorts in their sample. This suggests that the need to difference out genes—a key reason for the twins design—is as relevant historically as in recent data.

education more formally, we turn to the regression specified in (1) and (2).

[Figure 2 about here.]

We estimate that the return to schooling in the first half of the 20th century are around 4 to 5% (Table 4). Columns (1) to (3) are based on the 1940 1 percent IPUMS sample presented earlier in Table 3, while columns (4) to (8) use the sample of linked twins. We compare results from both datasets to determine if the rates of return are likely to differ between twins and the general population. There are two takeaways from Table 4. First, the OLS coefficients from the two datasets are reasonably similar. In the columns with matching controls—none, the set of exogenous controls, or the set of both exogenous and “bad” controls—the coefficients on years of education are nearly identical. If the direction and magnitude of bias in the OLS results are similar for the two groups, the actual return to education may be similar for both populations. In either dataset, the estimated return is lowest when the bad controls are added. Of course, this should be viewed cautiously given the difficulty in interpreting regressions with bad controls (Angrist and Pischke 2009). Second, the estimates are slightly smaller when attempting to address ability bias with the twins-fixed-effects specification. Our preferred model in column (7) indicates a 4.4 percent increase in wages for each additional year of schooling on average, down from 5.0 or 5.6 percent without the fixed effect, without or with the “good” controls. Again, this effect shrinks when the set of potentially endogenous controls are included in column (8).

[Table 4 about here.]

Our findings suggest a lower return to schooling at mid-century than has been previously documented. Clay et al. (2016), exploiting changes in compulsory schooling laws during the first half of the 20th century, estimate a return with a lower bound of 6.4%. Contemporary twins estimates, on the other hand, hover around 10% (Ashenfelter and Rouse 1998; Behrman and Rosenzweig 1999; Rouse 1999). That our results differ from Clay et al. (2016)’s despite



focusing on a similar period may not be surprising since the group of compliers in the CSL “experiment” is likely to be different from ours. This difference could work in our favor as the use of twins allows for a direct comparison with the estimates for more recent times, many of which are based on twin studies. Such a comparison suggests that the value of schooling has increased over time. Goldin and Katz (2008) show that the return to education were high in 1915, but collapsed thereafter before rising again in the late 20th century. Our lower rates fit well with the middle part of the Goldin and Katz narrative.<sup>24</sup>

### 3.3 Sensitivity Analysis

Are our baseline results robust? Three potential issues that may distort our findings are imperfections in record linkage, measurement error in years of schooling, and any remaining differences between twins that the fixed effects estimation has not accounted for. We address these issues in turn and continue to find that the return to schooling was positive in the past but smaller compared with the present.

#### 3.3.1 Robustness to Alternative Linking Methods

Our estimated return to education is relatively stable across two of the most commonly used linking methods, as well as variants of those methods that change the relative conservatism of the linking, robustness tests recommended by Abramitzky et al. (2019) for historical census linking research.

We use two methods to link our sample of twins ahead to 1940.

The first, presented through out the paper as our baseline, is the machine learning ap-

---

<sup>24</sup>There is some debate in the literature as to whether or not the return to schooling exhibited the U-shaped pattern described in Goldin and Katz (2009) is a separate issue. The trends in Goldin and Katz (2009) are not causal and only begin in the early 20th century. Jovanovic and Rousseau (2005) put two separate series of wage ratios together: the 1870-1894 wage ratio of urban skilled to unskilled workers from Williamson and Lindert (1980), and the 1939-1995 ratio of clerical to manufacturing production wages from Goldin and Katz (1999). This amplifies the U-shape nature of the skill premium. Other work by Katz and Margo (2014) suggests that the return to schooling was on an upward trend for much of the 19th century, leaving open the possibility of a continuous rise in the skill premium over the course of US history. We cannot speak to whether the return to education was linear or U-shaped over time, as education attainment is only recorded from the 1940 census on and implementation of a twins design for earlier periods is beyond the scope of this paper.

proach (Feigenbaum 2016), described in Section 2. The second is the simpler algorithmic approach, often referred to as ABE after Abramitzky et al. (2012). The ABE method is similar to Long and Ferrie (2013) and matches individuals with a deterministic procedure based on names, birthplace, and age. The ABE procedure begins by searching for exact matches. If an exact match cannot be found, an age difference of 1 year is allowed, still requiring exact name and birthplace matching. The age window is then expanded to 2 years if a match still cannot be found. Both methods allow us to change the relative conservatism. In the machine learning method, this is done by varying the relative weight on false positives versus false negatives. For ABE, we report both the classic version and a more conservative method described in Abramitzky et al. (2019) that imposes two further restrictions on the classic approach: (i) names must be unique within a 5-year age band in both the base and target datasets, and (ii) the age must match exactly.

[Figure 3 about here.]

Figure 3 shows that our results are reasonably robust across all variants of the machine learning method, from the most to least conservative, as well as when we use both the classic and conservative ABE procedures. Even the result least in line with the rest—the conservative ABE method, based on a much smaller sample of twins—actually accords with our main conclusion, that the return to education in 1940 was positive but relatively low.

### 3.3.2 Measurement Error in Education

Measurement error in years of schooling is unlikely to have distorted our results. Such errors pose a common challenge in the twins literature and may be magnified when comparing twins with similar levels of education. Rouse (1999) finds that 8 to 12% of the variation in reported schooling across twins stems from errors. Measurement error could attenuate or amplify the estimated returns, depending on their nature (classical or non-classical). One solution used in the twins literature, first introduced by Ashenfelter and Krueger (1994), is to instrument each twin’s own reported years of schooling with the levels of education that

the other twin reports for them. Unfortunately, the information required to implement this method—namely asking one twin their sibling’s years of education—is not available in our census data.

As an alternative means of reducing the degree of measurement error, we focus on twins who completed key milestones in education. We assume that respondents can remember key events in the past more clearly—for instance, whether or not they finished high school or college. There may thus be less reporting error for those attaining these levels of education. To implement this, we narrow our sample to twin pairs where both members completed one of the following education levels exactly: common school graduate, high school graduate, college graduate, or five or more years of college.

Using only twins who reached key milestones in education, we find rates of return that are similar to the baseline. Table 5 presents the results. Columns 1 and 2 are based on the random sample of men in 1940 and columns 3 and 4 are based on the pooled OLS regression. With these specifications, the point estimates are sensitive to the types of controls that are included—incorporating the good controls for age, race, and foreign-born status increase the returns a bit. Our preferred within-twins estimates in column 5, though a bit larger than our findings in Table 4 still reinforces the conclusions from the preceding analysis: the return to schooling was positive at mid-century, but it was smaller than the returns estimated today.

[Table 5 about here.]

### **3.3.3 Unobserved Differences**

If twins are indeed identical, why should either twin complete more or fewer years of schooling? This is a major critique of using twins to estimate the returns to schooling: even identical twins are not exactly the same. Any remaining unobserved differences between twins will not be accounted for in the fixed effects model. Bound and Solon (1999), for example, make the case that there are differences within twin pairs before, during, and after birth. While this might be true, we provide suggestive evidence here that such differences

are unlikely to compromise our key findings.

For the within-twins estimates to be less biased than the OLS results, the endogenous variation in schooling within families needs to be smaller than that between families. Since ability is unobserved, Ashenfelter and Rouse (1998) and Bonjour et al. (2003) assess this by taking several potential correlates of ability—such as marital status, self-employment, union coverage, test scores, spouse’s education—and comparing the associations between these covariates and schooling both within and between families. Using contemporary data, they find that the within-family correlations are much weaker than those between families, suggesting that much of the variation in unobserved ability is between families.

We follow the test in Ashenfelter and Rouse (1998) and Bonjour et al. (2003). We focus on several potential correlates of ability: marital status, full-time employment status, the number of children one has, spouse’s years of education, and spouse’s labor force participation all measured in 1940. The correlations between these characteristics and years of schooling are always statistically significant both within and between families (see Table A.4 in the Appendix) except for whether or not one’s spouse is in the labor force. Nonetheless, the within-family correlations are always smaller in magnitude than the between-family correlations. This suggests that the rates of return based on the family fixed effects model may be less biased than the returns from the OLS model.

This exercise, however, is not ideal in our setting for three reasons. First, many correlates of ability are not available in our historical data, most importantly test scores. Second, while we can construct other correlates, an important outcome like self-employment does not occur in our baseline analysis sample of twins because we only observe earnings in the 1940 census for men reporting labor earnings. Finally, while marital status and spousal education are reported, they may be less reflective of a person’s ability in the first half of the 20th century. For example, over 75 percent of our linked twins were married by 1940 while the corresponding figure in Ashenfelter and Rouse (1998)’s twins data is lower at 64 percent.

We propose two ways of assessing the unobservable differences between twin brothers, both of which are unique to our historical setting.

First, we gradually restrict the sample of twins to those with smaller differences in years of schooling and find little change in the estimated return to education. The narrower the gap in schooling, the more similar twins are likely to be in terms of unobserved ability. Consequently, the estimated rate of return will be less biased. Figure 4 plots the results from this exercise—each marker represents the coefficient from a separate twin fixed effect regression, where the maximum difference in schooling ranges from 1 to 17 along the x-axis, with 95% confidence intervals. Behind the point estimates, we plot the sample size corresponding to each regression. The return to schooling is consistently around 4% across the span of differences in years of schooling. This suggests that our baseline estimates are unlikely to be severely biased by within-twins differences in unobserved ability. Of particular interest are the results with bandwidths of 4 years or less in either direction—these categories encompass the majority of twin pairs as depicted by the sample size bars. With the smallest bandwidth of a year’s difference in schooling, the point estimate is 4.2 percent and is significant at the 1 percent level. The confidence intervals are very wide in this case, which is not surprising given the large amount of noise in wages within this narrow band hinted at in Figure 2. Nonetheless, the extreme upper bound of the 95% confidence interval is still slightly lower than the return to schooling in contemporary twins studies.

[Figure 4 about here.]

Second, we restrict our sample to twins who were more likely to have received similar treatment from their parents and continue to find rates of return that are consistent with our baseline. Focusing on those who are likely to have received equal resources and treatment from their parents may help to eliminate differences in the nurture aspect of unobserved ability. We propose using the similarity of first names as an indicator of parents’ intentions to treat their twins similarly. Appendix A.3 provides suggestive evidence that first names do carry valuable information on the intentions of parents: twins with more similar names

are more likely to have the same school enrollment status in 1900, 1910, and 1920.<sup>25</sup> Names that are more similar may thus identify children who were raised with comparable resources and in similar environments.<sup>26</sup> We use three measures of name similarity: first names that begin with the same letter, first names that have a Jaro-Winkler string distance of less than 0.2, or names that belong to the same Soundex phonetic group.<sup>27</sup> These metrics are used to divide the sample of twins into three subsets and we re-estimate equations (1) and (2). Table 6 presents the resulting coefficients. Though some subsamples are quite small, the twin fixed effects estimates in the even columns are broadly similar to the baseline.

[Table 6 about here.]

### 3.4 Return to Education Within or Across Occupations?

Education can translate into higher earnings in two key ways. First, more education could enable workers to work in “better” occupations. Second, within a given occupation, better educated workers may be more productive and thus earn more. Which channel is driving the return to schooling? To shed light on this, we add occupation fixed effects of varying coarseness to the twins specification to account for variation in earnings across occupations.

About two-thirds of the return to education appears to be higher earnings within an occupation (Table 7). In our baseline regression, column 1, we see that a year of education increases weekly earnings by 0.044 log points. In columns 2, 3, and 4, we include increasingly narrow fixed effects for occupations, based on the three-digit occupation code (`occ1950`) from

---

<sup>25</sup>We also show in Appendix Table A.2 that children who were twins in 1940 are more likely to be in the same grade in school when their names are more similar.

<sup>26</sup>Three famous twins in our period are illustrative examples of similarly names twins with similar career paths. Roy and Ray Grimes, born as twins on September 11, 1893 in Bergholz, OH, both grew up to be professional baseball players, though Ray was significantly more successful. Lee and Lyn Wilde were actresses and signers born October 10, 1922 as Marion Lee Wilde and Mary Lyn Wilde. Finally, Ann Landers and “Dear Abby”, the famed advice columnists were born twins in Sioux City, IA, on July 4, 1918 and named Esther Pauline Friedman and Pauline Esther Friedman, the first name of one as the middle name of the other. In 1940, all sets of twins had the same occupation. Because we restrict to male twin pairs, only the Grimes twins are in our data.

<sup>27</sup>The Jaro-Winkler string distance is a measure of string dissimilarity or the edit distance between two strings. It weights disagreements early in strings more harshly than disagreements towards the ends of strings and is often used to compare names in the record linkage literature. The Jaro-Winkler string distance ranges from 0 to 1. Strings that match have a Jaro-Winkler distance of 0.

IPUMS.<sup>28</sup> The return falls to 0.030 at the one-digit level, to 0.029 at the two-digit level, and to 0.028 at the three-digit level. This differs somewhat from Goldin and Katz (2009), who perform a similar analysis using the 1915 Iowa sample and conclude that about half of the return to schooling is within occupation and the other half is between occupation.

[Table 7 about here.]

Another way of showing that more schooling improves both the type of occupation and earnings within an occupation is to estimate the rate of return with occupation scores rather than earnings. Occupation scores, constructed by IPUMS based on the median earnings in 1950 for each occupation, are a common proxy for socio-economic status. These scores eliminate any variation within occupations. Twins with more education have higher occupation scores, as shown in column 5 of Table 7, but the relative elasticity is lower than the results with earnings. These results underscore that workers in 1940 benefited from education both via entry into better paying occupations and via higher pay within occupations.

In this section, our analysis of twins suggested that the return to education in 1940 was positive, though not especially large compared to the return to education in the late 20th or early 21st century. One benefit of our twin-based identification strategy and our large sample of twins is that we can estimate the return to education in many different contexts in 1940 which we turn to in the next section.

## 4 Heterogeneity in the Return to Education

With the twins identification strategy and our a large sample, we can estimate the return to schooling for different subsets of the population in 1940. This section documents a higher return for older cohorts<sup>29</sup> and for those whose fathers were farmers or had lower socio-economic

---

<sup>28</sup>As an example, occupations in the 000s are “professional, technical”; the 080s include economists, psychologists, statisticians and actuaries, and miscellaneous social scientists; and economists are code 081. While most 1 and 3 digit occupation categories are sensible, not all 2 digit occupation codings are economically meaningful: bookbinders (502) are grouped with cabinetmakers (505). For the codes, see: <https://usa.ipums.org/usa-action/variables/occ1950>

<sup>29</sup>Because we only observe outcomes in 1940, cohorts and ages are collinear and we cannot separately estimate the return to education for people born in 1900 from people who are 40 years old because they are

status.<sup>30</sup> The returns were also higher for twins with fewer foreign-born grandparents, suggesting that education alone did not offset the costs of nativism in the early 20th-century labor market. Finally, we find comparable returns across different regions.

## 4.1 Heterogeneity by Cohort

Could the average effects estimated in the preceding section mask differences in the return to schooling across cohorts? We construct our sample by searching for twin pairs in the 1900, 1910, and 1920 censuses and our twins are born across five decades. Figure 5 groups cohorts into 5-year bins and repeats our baseline analysis. The oldest twins in our sample were born between 1871-1875, while the youngest had birth years from 1916-1920.<sup>31</sup>

[Figure 5 about here.]

The point estimates in Figure 5 suggest that the returns to schooling were higher for the older cohorts, though relative imprecision of the point estimates makes it difficult to distinguish the coefficients for each cohort from each other. Simply fitting a line through the point estimates yields a downward slope in the estimated return of -0.003 per 5-year binned cohort.

The differences in the return to schooling by cohort could be interpreted in two ways. First, the downward trend in the return to schooling is consistent with Goldin and Katz (2008), who provide descriptive evidence that the high school and college wage premium fell from 1910 to 1950, before rising thereafter. Educational attainment rose steadily in the early twentieth century, with each new cohort of American workers having about 0.8 more years of schooling per decade (Goldin and Katz 2009). Skilled workers were thus more scarce

---

always the same set of twins in our data. The heterogeneity-by-cohort results may thus suggest that the return to education were higher for workers with more experience or who entered the labor market earlier in the 20th century or who were educated in the late 19th century.

<sup>30</sup>Table 3 shows that our sample is almost exclusively American-born whites. Because census to census linking rates are lower for African Americans, we do not explore heterogeneity in the return to education by race.

<sup>31</sup>There were only 38 pairs of twins born between 1871-1875, explaining the wide confidence intervals around the leftmost point plotted in Figure 5. Appendix Figure A.5 provides the sample size for each cohort.



among the older cohorts when they entered the labor market. Second, the trend in Figure 5 could reflect lower returns to education for individuals with less labor market experience or who are younger. It is not possible to distinguish between the two interpretations as we only observe earnings and education at one point in time, which makes cohort and age collinear.

## 4.2 Heterogeneity in the Return to Education by Family Status and History

Education could be important for enhancing intergenerational mobility (Card et al. 2018) or facilitating assimilation among immigrants. This subsection considers the heterogeneity in the return to schooling by family socio-economic status (SES) and immigration history.

Schooling can weaken intergenerational economic links if the return to education is higher for children from lower SES families. We find rates of return that are higher for the sons of farmers and for sons whose fathers were in the bottom half of the SES distribution, but such differences are not large. Table 8 stratifies the sample by the father’s occupation and SES and then estimates the return to education using twins.<sup>32</sup> We split the sample into three groups: the twin sons of farmers, the twin sons with fathers in the top half of the occupation score distribution, and twin sons of fathers in the bottom half of the occupation score distribution.<sup>33</sup>

[Table 8 about here.]

Our results suggest that the return to education at mid-century was higher for sons from households lower on the SES ladder. This is consistent with the quantile IV results

---

<sup>32</sup>We use occupation scores to split fathers by SES because occupation scores are the best measure of status we see. Recall, we collect the fathers in the 1900, 1910, and 1920 censuses with occupations but not earnings or education. Occupation scores, a measure produced by IPUMS that assigns occupations ranks based on median earnings by occupation in 1950, have been used by economists as a substitute measure of SES. These scores offer a useful but crude measure of SES for samples without earnings. See Saavedra and Twinam (2018) for a review of recent studies using occupation score when earnings or income are not available.

<sup>33</sup>We separate out farmers for two reasons. First, farmer is the most common occupation for fathers in our sample, not surprising given the declining but still-important place of agriculture in the US economy in the early 20th century. Second, as we note, we use occupation score of the father to determine family SES. Occupation scores may be especially weak for the large and highly varied category of farmers, ranging from sharecroppers and tenant farmers to those owning and farming huge numbers of acres. Feigenbaum (2018) documents in Iowa 1915 the huge amount of variance of actual earnings among farmers even within the same state—farmers who would all be coded with the same occupation score.

in Clay et al. (2016) that show the returns were monotonically decreasing in quantile.<sup>34</sup> However, the differences in returns across the even columns in Table 8 are not statistically significant at traditional levels. Our results may also reflect that the educational attainment of twin sons depends on their fathers' status—on average, sons of farmers had 9 years of schooling in 1940 compared to 9.6 and 10.3 years for sons of non-farmer fathers below or above median occupations score—though these relatively small differences are unlikely to explain the findings in Table 8.

Many proponents of mass education in the early 20th century believed that the American school system could facilitate the assimilation of immigrants who arrived during the Age of Mass Migration. Our sample contains very few foreign-born twins and twins with foreign-born parents. Thus, to study if the return to schooling varies by immigration history, we focus on grandparents. Specifically, we divide white twins in our sample into three groups: those with no foreign-born grandparents, those with 1-3 foreign-born grandparents, and those whose grandparents are all foreign-born.<sup>35</sup>

[Table 9 about here.]

The return to education was lower for children with more foreign-born grandparents. Table 9 presents the results. We focus on the even columns based on the twins fixed effects specification. The rates of return are highest for twins with four American-born grandparents and lowest for twins with four foreign-born grandparents.<sup>36</sup> Put differently, the labor market

---

<sup>34</sup>Parman (2012) shows the return to education among farmers in the early 20th century. Because farmers in 1940 do not report wages or salaries, we cannot assess these findings. However, for sons of farmers who moved to a new, wage-earning professions, column 2 suggests that education was valuable to them.

<sup>35</sup>We limit the sample to whites because most African American twins had four American-born grandparents, inducing a high correlation between grandparent nativity and race and complicating the interpretation. Of the foreign-born grandparents in our sample, 31.7% were born in Germany, 15% in Ireland, 6.9% in England, with 5% born in Russia, Canada, and Italy and smaller shares from other countries. We are able to track the immigration history of grandparents because we observe twins in the census as children in 1900, 1910, and 1920. In all three census waves, census enumerators asked everyone for their birthplace, as well as the birthplace of their mother and father. We see the twins' parents and use their answers about their parents here. This limits our sample to twins whose mother and father are both present in the household. 85.4% of our sample meet this criterion.

<sup>36</sup>The return in column (2) is statistically significantly larger than the returns in column (4) and (6), tested by stacking the data and estimating interaction coefficients.

offers smaller rewards for twins with more recent family immigrant histories. Clay et al. (2016), exploiting variation in compulsory schooling laws by state and cohort, also find higher returns to education among children with two American-born parents, compared to children with one or more foreign born parents. While their results are somewhat statistically imprecise, perhaps because the first stage effects of compulsory schooling on the educational attainment of children with foreign-born parents are weak (Lleras-Muney and Shertzer 2015), taking their results and ours together point to a lower return in the American labor market for education among the children and grandchildren of the foreign born.<sup>37</sup>

The lower return to schooling for third generation American twins could be explained by both segregation and language or cultural fluency. Residential segregation of immigrants in urban and rural areas was high in the early 20th century (Eriksson and Ward 2018). This implies that the children or grandchildren of immigrants were unlikely to attend the same schools as the American-born. Consequently, the type and quality of schooling acquired by both groups may have been different, generating variation in the value of education across groups. This echoes the findings on the black-white gap in education returns (Card and Krueger 1992; Carruthers and Wanamaker 2016, 2017). The lower returns could also reflect poorer language or cultural fluency. If the importance of formal education is mediated by language or culture, then those with a better command of English or American culture will have relatively more effective (useful) units of human capital, all else equal. However, using an individual fixed effects model, Ward (2019b) finds that the return to English fluency in the early 20th century was small compared to recent years. Residential segregation may thus be the more important of the two explanations.

---

<sup>37</sup>There are also differences in average educational attainment between twin sons, depending on their family immigration history, though the pattern is non-linear. White twins with four American-born grandparents averaged 9.7 years of schooling in 1940, less than the 10.4 years of schooling among White twins with 1 to 3 American-born grandparents. However, twins with four foreign-born grandparents had only 9.6 years of schooling in 1940 on average.

### 4.3 Returns to Education Across Regions

While Goldin and Katz (2009) describe the 20th century the human capital century, investments in education and the use of human capital in production are not uniform across the country. Could there be spatial variation in the return to schooling? Since stratifying the sample on location in 1940 would be post-treatment, we examine the rate of return between twin pairs raised in different parts of the country.

Our estimates suggest little spatial variation in the return to schooling. We divide the sample into four census regions, based on where twins were living when first observed in the 1900, 1910, and 1920 censuses. The returns are broadly similar across regions, as shown in Figure 6. Though the point estimates are higher for twins from the South, these differences are not statistically significant.

[Figure 6 about here.]

What might explain this apparent geographic similarity? It is not the differential shares of non-white twins across the regions, as we plot the regional returns for the full twins sample and the subset of white twins. Another unlikely hypothesis is that the quality of schooling was similar across localities. Goldin and Katz (2008), for example, document large differences in the rate of progress of the high school movement across regions during the first half of the 20th century. Given the vast differences in labor market structures across the country and huge variation in the supply of educated workers across states and regions (Goldin and Katz 2009, p. 204), it also does not seem plausible to attribute the regional consistency to equal rewards to human capital regardless of locality.<sup>38</sup> We leave an investigation of the lack of spatial variation in returns to future research with better-suited empirical strategies.

---

<sup>38</sup>Within our sample of twins, differences in the supply of educated workers are dramatic across census regions. Southern-born white twins averaged only 9 years of schooling, less than the 9.8 years in the Northeast, 10 in the Midwest, or 10.7 in the West.

## 5 The Return to Education on Other Outcomes

While the primary goal of this paper is to determine the earnings return to education, an increase in schooling may also affect other outcomes. In this section, we leverage the same twin comparison design and sample to estimate the causal effect of education on a host of outcomes in 1940, including migration, marriage, and family fertility. We also show that there are returns to education on economic dimensions beyond simple labor earnings, including non-wage earning and on the extensive margin of labor supply.

In the early twentieth century, education played an important role in geographic mobility, as we show across a variety of outcomes in both panels of Table 10. As before, we focus on the results with twin family fixed effects in the even columns. Brothers with more schooling have a higher propensity of leaving the counties where they grew up (Panel A, column 2). These moves tend to go beyond state borders (Panel A, column 6). The association between education and geographic mobility during this period has been documented elsewhere in the literature. Margo (1990), for example, argues that better-educated African Americans were more likely to move because schooling lowered the cost of migration through increasing knowledge of distant opportunities or enhancing the ability to assimilate.<sup>39</sup> We also find that as adults, the twin with more schooling is also more likely to reside in urban and more populous areas rather than on farms (Panel B). Table 10 limits the sample to our main analysis set of twins, but as we show in Table A.7, when we replicate on the complete sample of linked twins we find very similar results about the effect of education on migration and location choice.

[Table 10 about here.]

Do differences in education levels alter family structures? Table 11 implements regressions (1) and (2), replacing the original wage outcome with three measures of family status in 1940:

---

<sup>39</sup>Our findings also complement those in the development economics literature. Jensen and Miller (2017), for instance, find evidence that parents in India strategically underinvest in their children’s education in order to discourage them from migrating to urban areas and to remain on the farm instead.

whether or not the twin is married, whether or not the twin has children in the household, and how many children are in the household, all measured in 1940. The fixed effects regressions in the even columns show no impact of schooling on the likelihood of marriage. Those with more years of schooling, however, are less likely to have children and have fewer children overall.<sup>40</sup> The point estimates, however, are small compared with the outcome means. 53% of our sample have children but another year of schooling lowers this by less than a percentage point. The effects are also small when compared to other shocks to human capital during this period. In a quite different context—the large scale school building program in the US South targeted at African Americans from 1913-1932 known as the Rosenwald Rural Schools Initiative—Aaronson et al. (2014) find that African-American women who acquired more schooling had substantially lower fertility, particularly along the intensive margin.

[Table 11 about here.]

Finally, we explore if schooling influenced economic dimensions other than earnings. These results are notable on their own, but they also serve as a check on our main estimates. In the 1940 census, respondents provided not just their wages, but they also reported their occupation, whether or not they earned more than \$50 in non-wage earnings, and homeownership, as well as hours and weeks worked. We use the within-twins comparison to study how education affects these outcomes.

Education raises occupational status. We showed a positive effect on the occupation scores from IPUMS in column 5 of Table 7—each additional year of schooling increases occupation scores by about 2 percent. Panel A of Table 12 shows similarly-sized positive effects on three variants of occupation scores. First, we use the LIDO scores from Saavedra and Twinam (2018).<sup>41</sup> While there is no variation in the IPUMS scores within a given occupation, LIDO scores allow for within-occupation differences. Specifically, Saavedra and

---

<sup>40</sup>Appendix Table A.8 replicates this analysis on the complete sample of linked twins and finds similar results for fertility. However, we find that while education correlates with a lower probability of marriage in column 1, a small but positive effect is observed when using the twins fixed effects model.

<sup>41</sup>The LIDO acronym stands for lasso-adjusted industry, demographic, and occupation scores.

Twinam (2018) use lasso regressions and the 1950 census to generate scores for each occupation by industry, race, sex, age, and region. Column 2 shows that another year of education raises LIDO scores by just more than 2%. Second, we turn to the occupation scores calculated by Olivetti and Paserman (2015). These are based on the 1900 Cost of Living survey and impute the occupational status of farmers with data on both farm owners and tenants. Column 4 suggests a rate of return of over 2.6% with this outcomes. Finally, we generate our own occupation scores based on the 1940 complete counts and analyze use them in columns 5 and 6.<sup>42</sup> With this measure, the rate of return is about 3%.

[Table 12 about here.]

There also is a return to non-occupation-based economic outcomes, as we show in Panel B of Table 12. Here we expand the sample to all linked twins, rather than the subset used in our baseline analyses. Only 1% of our twins sample has a top-coded earnings level (the census top-coded earnings at \$5000). But each year of education increased the odds of being top-coded by 0.3 points. These effects are relatively large: an extra four years of schooling, the difference between a high school and college graduate twin, is more than double the baseline rate of top-coding.

Education also increases the odds that a twin reports non-wage earnings in excess of \$50. Only 13% of our twins sample reports non-wage earnings.<sup>43</sup> But an additional year of education increases that rate by 0.3% (Panel B, column 4), suggesting that education played a very important role in determining who worked for a wage or earned capital income in 1940.

We also see a small effect of education on homeownership (Panel B, column 6), though even the difference between a twin with a college degree and a common school education

---

<sup>42</sup>While the IPUMS scores are based on the 1950 data, we use the 1940 census to match the year we observe the twins as adults. We construct the median wage earners for each occupation among men who are wage and salary workers aged 16 to 64 and who worked at least 35 hours and 40 weeks in the preceding year.

<sup>43</sup>In our baseline specification, we restricted the sample to twin pairs who were both working full time and had non-zero weekly earnings. In Table 12 we relax that restriction to include the full sample of linked twins.

is less than five percentage points off a sample base of 46%. These small effects could be partially explained by the increased urbanization rate of better educated twins, as we showed in Table 10. The small effects on homeownership also stand in contrast with the period post-WWII, which saw a dramatic rise in both education attainment and homeownership among US veterans (Bound and Turner 2002; Fetter 2013). After WWII, however, there were generous benefits for both education and housing under the GI Bill.

Education also increased labor supply (Panel C, Table 12). With more education, twins were more likely to work full-time (defined as working more than 40 weeks in 1939), more weeks in the year, and hours per week. Unlike the baseline results for weekly wages, however, the point estimates are relatively small, each being between 1 and 3 percent of the respective outcome means.

As documented in this section, the returns to education in 1940 did not just accrue via more earnings. We have also shown that twin siblings with more education were also more likely to migrate (to cities and more populous places) and to have higher socio-economic status, as well as work more. Interestingly, many of these effects of education on other outcomes were relatively small.

## 6 Conclusion

What was the return to schooling in mid-century America? Constructing a large linked sample of twins and exploiting within-twins variation in education and earnings, we estimate a return of around 4% in 1940. We provide evidence that suggests that our results are not severely biased by imperfections in record linkage, measurement error in the years of schooling, or possible differences in ability or selection within twin-pairs.

While positive, our estimated return to schooling is substantially smaller than estimates of the return today. This is consistent with Goldin and Katz (2009) as 1940 sits at the nadir of the U-shaped trend in skill premium they trace. However, our paper cannot speak to the decline in skill premium before 1940. Beyond quantifying the causal effect of years of schooling further back in time, future research could also consider the historical returns in



other countries and evaluate if the value of education in the US was the exception or the norm.

Taking advantage of the large size of our dataset, we are also able to document heterogeneous effects of education along several dimensions. Older cohorts, for instance, are observed to have a higher return to schooling, while the returns are lower for those with more foreign-born grandparents.

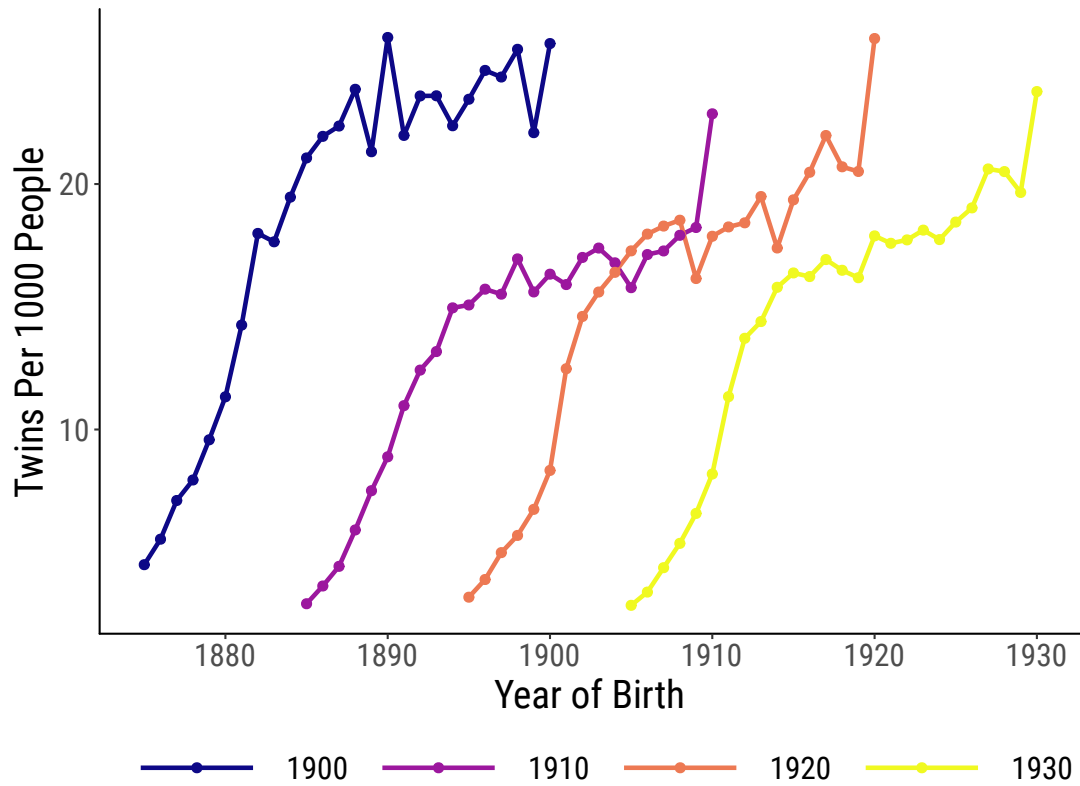
## References

- Aaronson, Daniel, Fabian Lange, and Bhashkar Mazumder (2014). Fertility transitions along the extensive and intensive margins. *American Economic Review* 104(11), 3701–24.
- Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson (2012). Europe’s Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration. *American Economic Review* 102(5), 1832–1856.
- Abramitzky, Ran, Leah Platt Boustan, Katherine Eriksson, James J Feigenbaum, and Santiago Pérez (2019). Automated Linking of Historical Data. Working Paper 25825, National Bureau of Economic Research.
- Angrist, Joshua D. and Alan B. Krueger (1991). Does Compulsory School Attendance Affect Schooling and Earnings? *The Quarterly Journal of Economics* 106(4), 979–1014.
- Angrist, Joshua D. and Jorn-Steffen Pischke (2009). *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton University Press.
- Ashenfelter, Orley and Alan Krueger (1994). Estimates of the Economic Return to Schooling from a New Sample of Twins. *American Economic Review* 84(5), 1157–1173.
- Ashenfelter, Orley and Cecilia Rouse (1998). Income, Schooling, and Ability: Evidence from a New Sample of Identical Twins. *Quarterly Journal of Economics* 113(1), 253–284.
- Bailey, Martha, Connor Cole, Morgan Henderson, and Catherine Massey (2017). How Well Do Automated Methods Perform in Historical Samples? Evidence from New Ground Truth. Technical report, National Bureau of Economic Research, Cambridge, MA.
- Behrman, Jere R. and Mark R. Rosenzweig (1999). “Ability” Biases in Schooling Returns and Twins: A Test and New Estimates. *Economics of Education Review* 18(2), 159–167.
- Biavaschi, Costanza, Corrado Giulietti, and Zahra Siddique (2017). The Economic Payoff of Name Americanization. *Journal of Labor Economics* 35(4), 1089–1116.
- Bingley, Paul, Kaare Christensen, and Ian Walker (2009). The returns to observed and unobserved skills over time: Evidence from a panel of the population of danish twins. *Danish National Institute for Social Research*.
- Bonjour, Dorothe, Lynn F. Cherkas, Jonathan E. Haskel, Denise D. Hawkes, and Tim D. Spector (2003). Returns to Education: Evidence from U.K. Twins. *The American Economic Review* 93(5), 1799–1812.
- Bound, John and Gary Solon (1999). Double trouble: On the value of twins-based estimation of the return to schooling. *Economics of Education Review* 18(2), 169–182.
- Bound, John and Sarah Turner (2002). Going to war and going to college: Did world war ii and the gi bill increase educational attainment for returning veterans? *Journal of labor economics* 20(4), 784–815.

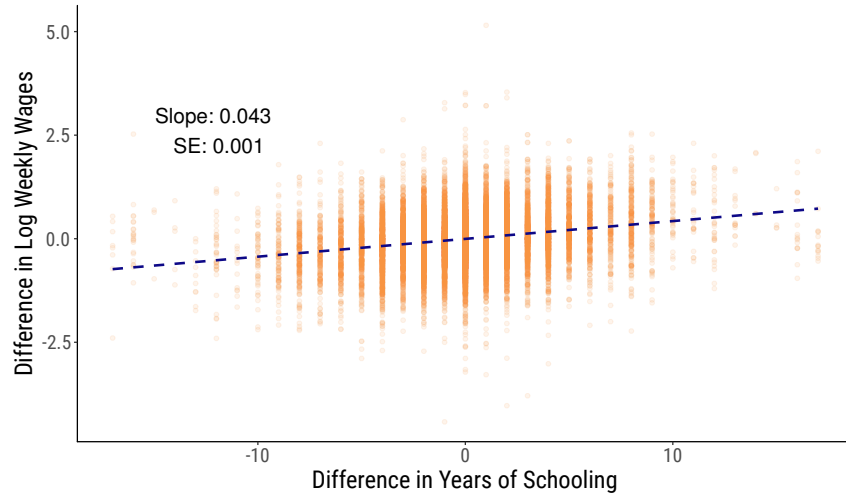
- Bruder, Carl E. G., Arkadiusz Piotrowski, Antoinet A. C. J. Gijsbers, Robin Andersson, Stephen Erickson, Teresita Diaz de Ståhl, Uwe Menzel, Johanna Sandgren, Desiree von Tell, Andrzej Poplawski, Michael Crowley, Chiquito Crasto, E. Christopher Partridge, Hemant Tiwari, David B. Allison, Jan Komorowski, Gert-Jan B. van Ommen, Dorret I. Boomsma, Nancy L. Pedersen, Johan T. den Dunnen, Karin Wirdefeldt, and Jan P. Dumanski (2008). Phenotypically Concordant and Discordant Monozygotic Twins Display Different DNA Copy-Number-Variation Profiles. *The American Journal of Human Genetics* 82(3), 763–771.
- Card, David (1993). Using Geographic Variation in College Proximity to Estimate the Return to Schooling. Working Paper 4483, National Bureau of Economic Research.
- Card, David (1999). Chapter 30 - The Causal Effect of Education on Earnings. In O. C. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3, pp. 1801–1863. Elsevier.
- Card, David, Ciprian Domnisoru, and Lowell Taylor (2018). The Intergenerational Transmission of Human Capital: Evidence from the Golden Age of Upward Mobility. Working Paper 25000, National Bureau of Economic Research.
- Card, David and Alan B. Krueger (1992). School Quality and Black-White Relative Earnings: A Direct Assessment. *The Quarterly Journal of Economics* 107(1), 151–200.
- Carruthers, Celeste K. and Marianne H. Wanamaker (2016). Separate and Unequal in the Labor Market: Human Capital and the Jim Crow Wage Gap. *Journal of Labor Economics* 35(3), 655–696.
- Carruthers, Celeste K. and Marianne H. Wanamaker (2017). Returns to school resources in the Jim Crow South. *Explorations in Economic History* 64, 104–110.
- Clay, Karen, Jeff Lingwall, and Jr. Stephens, Melvin (2016). Laws, Educational Outcomes, and Returns to Schooling: Evidence from the Full Count 1940 Census. Working Paper 22855, National Bureau of Economic Research.
- Conley, Dalton and Jason Fletcher (2017). *The Genome Factor: What the Social Genomics Revolution Reveals about Ourselves, Our History, and the Future*. Princeton: Princeton University Press.
- Eriksson, Katherine and Zachary A Ward (2018). The Ethnic Segregation of Immigrants in the United States from 1850 to 1940. Working Paper 24764, National Bureau of Economic Research.
- Feigenbaum, James J. (2016). Automated Census Record Linking: A Machine Learning Approach.
- Feigenbaum, James J. (2018). Multiple Measures of Historical Intergenerational Mobility: Iowa 1915 to 1940. *The Economic Journal* 128(612), F446–F481.

- Fetter, Daniel K (2013). How do mortgage subsidies affect home ownership? evidence from the mid-century GI bills. *American Economic Journal: Economic Policy* 5(2), 111–47.
- Fraga, M. F., E. Ballestar, M. F. Paz, S. Ropero, F. Setien, M. L. Ballestar, D. Heine-Suner, J. C. Cigudosa, M. Urioste, J. Benitez, M. Boix-Chornet, A. Sanchez-Aguilera, C. Ling, E. Carlsson, P. Poulsen, A. Vaag, Z. Stephan, T. D. Spector, Y.-Z. Wu, C. Plass, and M. Esteller (2005). From The Cover: Epigenetic differences arise during the lifetime of monozygotic twins. *Proceedings of the National Academy of Sciences* 102(30), 10604–10609.
- Goldin, Claudia (1998). America’s Graduation from High School: The Evolution and Spread of Secondary Schooling in the Twentieth Century. *The Journal of Economic History* 58(02), 345–374. 00404.
- Goldin, Claudia and Lawrence F Katz (1999). The Returns to Skill in the United States across the Twentieth Century. Working Paper 7126, National Bureau of Economic Research.
- Goldin, Claudia and Lawrence F. Katz (2000). Education and Income in the Early Twentieth Century: Evidence from the Prairies. *The Journal of Economic History* 60(3), 782–818.
- Goldin, C. and R. A. Margo (1992). The Great Compression: The Wage Structure in the United States at Mid-Century. *The Quarterly Journal of Economics* 107(1), 1–34. 00682.
- Goldin, Claudia Dale. and Lawrence F. Katz (2008). Long-Run Changes in the Wage Structure: Narrowing, Widening, Polarizing. *Brookings Papers on Economic Activity* 2007(2), 135–165.
- Goldin, Claudia Dale and Lawrence F. Katz (2009). *The Race between Education and Technology*. Harvard University Press.
- Isacsson, Gunnar (2004). Estimating the economic return to educational levels using data on twins. *Journal of Applied Econometrics* 19(1), 99–119.
- Jensen, Robert and Nolan H Miller (2017). Keepin’em down on the farm: Migration and strategic investment in children’s schooling. Technical report, National Bureau of Economic Research.
- Jovanovic, Boyan and Peter L. Rousseau (2005). Chapter 18 - General Purpose Technologies. In P. Aghion and S. N. Durlauf (Eds.), *Handbook of Economic Growth*, Volume 1, pp. 1181–1224. Elsevier.
- Katz, Lawrence F. and Robert A. Margo (2014). Technical Change and the Relative Demand for Skilled Labor: The United States in Historical Perspective. *Human Capital in History: The American Record*, 15–57.
- Kulkarni, Aniket D., Denise J. Jamieson, Howard W. Jones, Dmitry M. Kissin, Maria F. Gallo, Maurizio Macaluso, and Eli Y. Adashi (2013). Fertility Treatments and Multiple Births in the United States. *New England Journal of Medicine* 369(23), 2218–2225.

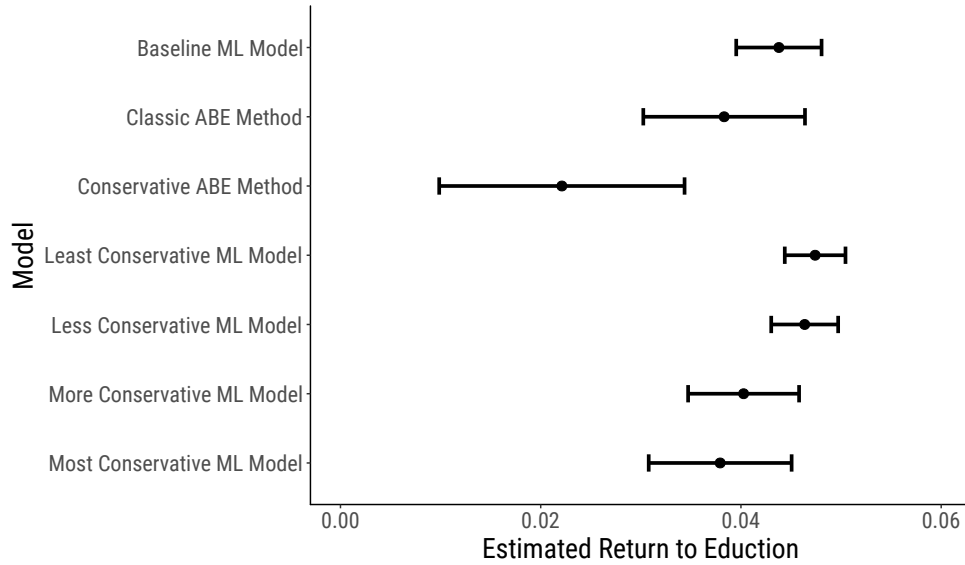
- Lleras-Muney, Adriana and Allison Shertzer (2015). Did the Americanization Movement Succeed? An Evaluation of the Effect of English-Only and Compulsory Schooling Laws on Immigrants. *American Economic Journal: Economic Policy* 7(3), 258–290.
- Long, Jason and Joseph Ferrie (2013). Intergenerational Occupational Mobility in Great Britain and the United States Since 1850. *The American Economic Review* 103(4), 1109–1137.
- Margo, Robert A. (1990). *Race and Schooling in the South, 1880-1950: An Economic History*. University of Chicago Press.
- Martin, Joyce A., Brady E. Hamilton, Michelle J.K. Osterman, Anne K. Driscoll, and Patrick Drake (2018). Births: Final data for 2017. *National Vital Statistics Reports* 67(8), 1–49.
- Olivetti, Claudia and M. Daniele Paserman (2015). In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1940. *American Economic Review* 105(8), 2695–2724.
- Parman, John (2012). Good schools make good neighbors: Human capital spillovers in early 20th century agriculture. *Explorations in Economic History* 49(3), 316–334.
- Preston, Samuel H. and Michael R. Haines (1991). *Fatal Years: Child Mortality in Late Nineteenth-Century America*. NBER Series on Long-Term Factors in Economic Development. Princeton, N.J: Princeton University Press.
- Rouse, Cecilia Elena (1999). Further estimates of the economic return to schooling from a new sample of twins. *Economics of Education Review* 18(2), 149–157.
- Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek (2010). *Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]*. Minneapolis, MN: Minnesota Population Center [producer and distributor].
- Saavedra, Martin Hugo and Tate Twinam (2018). A Machine Learning Approach to Improving Occupational Income Scores. SSRN Scholarly Paper ID 2944870, Social Science Research Network, Rochester, NY.
- Stephens, Melvin and Dou-Yan Yang (2014). Compulsory Education and the Benefits of Schooling. *American Economic Review* 104(6), 1777–1792.
- Tan, Hui Ren (2019). More is Less? The Impact of Family Size on Education Outcomes in the United States, 1850–1940. *Journal of Human Resources* 54(4), 1154–1181.
- Ward, Zachary (2019a). Internal Migration, Education and Upward Rank Mobility: Evidence from American History.
- Ward, Zachary (2019b). The low return to English fluency during the Age of Mass Migration. *European Review of Economic History*.
- Williamson, Jeffrey G and Peter H Lindert (1980). *American inequality: A macroeconomic history*. Academic Press.



**Figure 1:** Twins per 1000 people in the complete count censuses, 1900-1930. We construct estimates of the number of twins in the United States in each decennial census from 1900 to 1930. To identify pairs of twins, we focus on two people under 25 living in the same household with the same age, birthplace, relationship to the household head, and last name. To calculate the number of twins per 1000 people, we count up the number of people under 25 as the denominator.

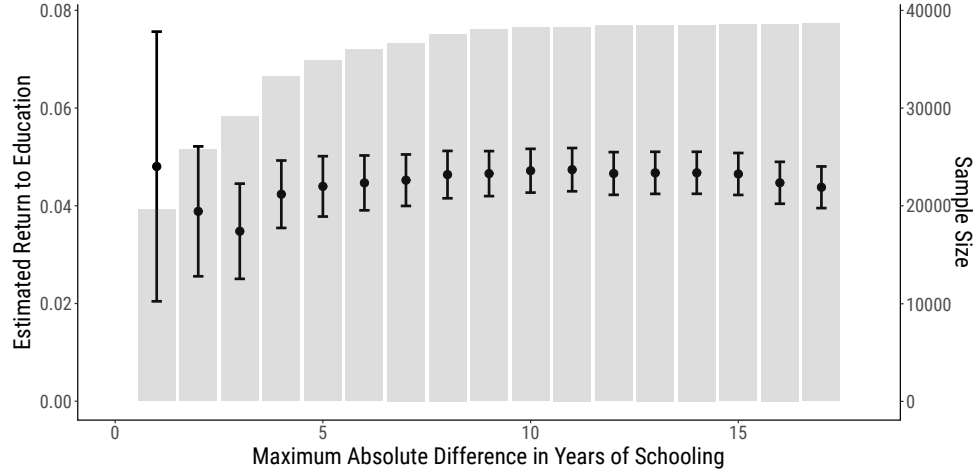


**Figure 2:** Returns to Schooling in the Early 20th Century. Data are from the pooled 1900-1940, 1910-1940, and 1920-1940 linked twins samples. The sample is restricted to wage and salary male workers with a weekly wage of at least \$6, who worked a positive number of weeks in the previous year, and a positive number of hours in the preceding week, according to the 1940 census. The best linear fit is shown. Robust standard errors are displayed.

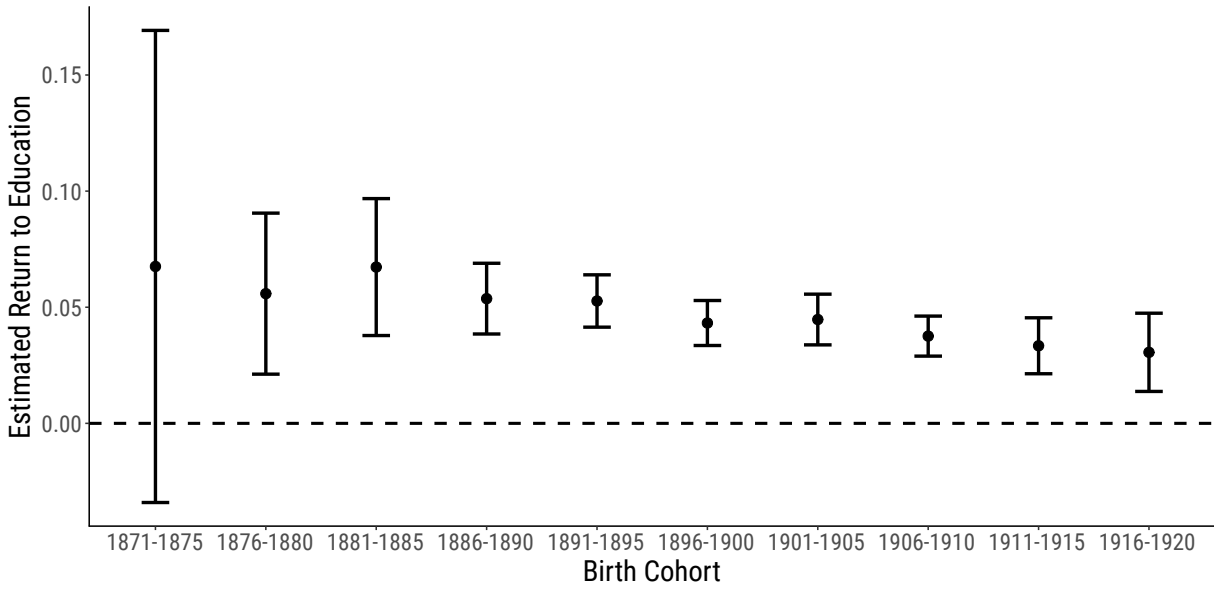


**Figure 3:** Robustness of Estimated Return to Schooling in the Early 20th Century Across Different Linking Methods. Data are from the pooled 1900-1940, 1910-1940, and 1920-1940 linked twins samples. We plot the return to education with 95% confidence intervals. The sample is linked with two methods: the machine learning approach (Feigenbaum 2016) and the algorithmic approach (ABE after Abramitzky et al. (2012), but very similar to Long and Ferrie (2013)). In the machine learning method, we vary the relative weight on false positives versus false negatives. For ABE, we report both classic version and a more conservative method described in Abramitzky et al. (2019). The sample is restricted to wage and salary male workers with a weekly wage of at least \$6, who worked a positive number of weeks in the previous year, and a positive number of hours in the preceding week, according to the 1940 census. Following Bailey et al. (2017), we use inverse propensity weights to adjust for observable differences between matched and unmatched persons in our census linked sample.

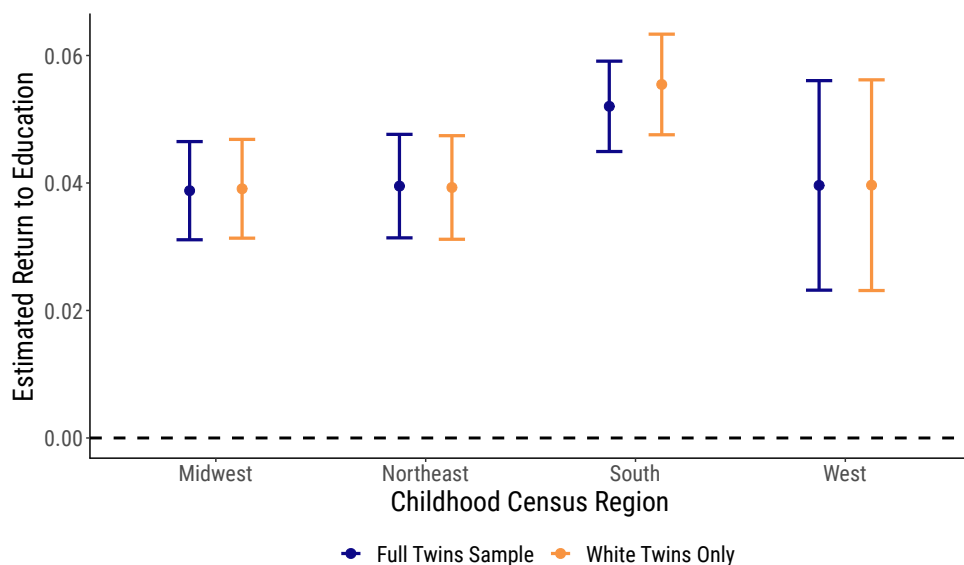




**Figure 4:** Returns to Schooling Are Still Positive Even as the Sample is Narrowed to Twin Pairs With More Similar Years of Schooling, First Difference Results. Data are from the pooled 1900-1940, 1910-1940, and 1920-1940 linked twins datasets. Each marker represents the returns to schooling from separate regressions of the within-twin difference in log weekly wages on the within-twin difference in years of schooling. The sample is restricted to wage and salary male workers with a weekly wage of at least \$6, who worked a positive number of weeks in the previous year, and a positive number of hours in the preceding week. The sample size ranges from 39,770 twins when the maximum difference in schooling is 17 years in either direction, to 20,670 twins when the maximum difference is only 1 year in either direction. 95% confidence intervals are shown, based on robust standard errors. Following Bailey et al. (2017), we use inverse propensity weights to adjust for observable differences between matched and unmatched persons in our census linked sample.



**Figure 5:** Returns to Schooling by Cohort. For each 5-year bins of cohorts in our linked twins data, we estimate the return to education, replicating our main specification with twin family fixed effects. Each point represents a different regression and is plotted with 95% confidence intervals. There is a downward trend in the returns across cohorts.



**Figure 6:** Returns to Schooling in the Early 20th Century did not vary across region of childhood. We split our sample into four census regions based on where the twins were living when we observe them as children—in the household with their parents and their twin. Data are from the pooled 1900-1940, 1910-1940, and 1920-1940 linked twins samples. The sample is restricted to wage and salary male workers with a weekly wage of at least \$6, who worked a positive number of weeks in the previous year, and a positive number of hours in the preceding week, according to the 1940 census. We estimate returns to schooling by region for the full sample and the subsample with only white twins. Following Bailey et al. (2017), we use inverse propensity weights to adjust for observable differences between matched and unmatched persons in our census linked sample.

**Table 1:** Twins in the United States Census, 1900-1930

Census Year	Twins per 1000				Number of Twins	
	All	Boy-Boy	Girl-Girl	Boy-Girl	All	Boy-Boy
1900	19.3	5.4	5.4	8.5	820,292	230,212
1910	13.9	4.1	4.2	5.7	691,500	203,668
1920	15.8	4.6	4.7	6.5	860,674	252,088
1930	14.8	4.5	4.5	5.8	897,078	273,196

*Note:* We construct estimates of the number of twins in the United States in each decennial census from 1900 to 1930. To identify pairs of twins, we focus on two people under 25 living in the same household with the same age, birthplace, relationship to the household head, and last name. To calculate the number of twins per 1000 people, we count up the number of people under 25 as the denominator.

**Table 2:** Families with Twins Are Similar to American Population, 1900-1920

	Twins in 1900-1920				IPUMS 1% Samples
	All	Linked	Weighted	Analysis Sample	1900, 1910, and 1920
Age	9.6 (6.7)	9.6 (6.6)	9.6 (6.7)	8.9 (6.3)	11.0 (7.1)
White	0.86 (0.35)	0.93 (0.25)	0.92 (0.27)	0.95 (0.22)	0.88 (0.32)
Number of Siblings	4.2 (2.2)	4.1 (2.2)	4.2 (2.2)	4.0 (2.1)	2.8 (2.3)
Father's Occupation Score	17.22 (11.36)	17.70 (11.92)	17.58 (11.71)	18.73 (12.66)	17.92 (12.00)
Urban	0.35 (0.48)	0.36 (0.48)	0.35 (0.48)	0.48 (0.50)	0.41 (0.49)
Farm	0.44 (0.50)	0.44 (0.50)	0.45 (0.50)	0.31 (0.46)	0.37 (0.48)
Foreign Born	0.02 (0.16)	0.02 (0.12)	0.02 (0.13)	0.02 (0.13)	0.05 (0.22)
Foreign Parents	0.47 (0.80)	0.47 (0.78)	0.47 (0.79)	0.54 (0.82)	0.51 (0.82)
Foreign Grandparents	1.46 (1.78)	1.55 (1.78)	1.55 (1.79)	1.74 (1.80)	1.54 (1.79)
Farmer Father	0.43 (0.49)	0.43 (0.49)	0.44 (0.50)	0.29 (0.45)	0.37 (0.48)
Number of Observations	693926	145914	145914	38652	676978

*Note:* Means with standard deviations below in parentheses. In the first four columns, we summarize the twins identified in 1900, 1910, and 1920. Column 2 presents the linked twins, unweighted, while column 3 uses inverse propensity weights to account for differences between census matched and unmatched twins. Column 4 limits the sample to our analysis sample, focusing only on the twins who were wage and salary male workers with a weekly wage of at least \$6 who worked a positive number of weeks in the previous year (1939) and a positive number of hours in the previous week. The twins sample includes only twins who both fit our sample criteria. Finally, column 5 presents summary statistics for boys under the age of 25 in 1900, 1910, and 1920 in their parents' households, based on three 1% random samples.

**Table 3:** Linked Twins Are Similar to American Population in 1940

	Linked Twins		1940 1% Sample
	Unweighted	Weighted	
Years of Schooling	9.63 (3.20)	9.52 (3.21)	9.18 (3.46)
Weekly Earnings	30.98 (27.13)	30.87 (27.54)	29.16 (26.72)
Log of Weekly Earnings	3.27 (0.55)	3.27 (0.55)	3.20 (0.58)
Works Fulltime	0.82 (0.39)	0.81 (0.39)	0.78 (0.42)
Years of Experience	18.5 (9.7)	18.9 (9.9)	19.5 (12.5)
Age	36.6 (9.4)	37.0 (9.6)	37.5 (12.2)
White	0.96 (0.20)	0.95 (0.22)	0.93 (0.26)
Foreign-born	0.02 (0.13)	0.02 (0.13)	0.13 (0.34)
Married	0.75 (0.43)	0.76 (0.43)	0.73 (0.44)
Number of Children	1.2 (1.5)	1.2 (1.5)	1.1 (1.5)
Number of Observations	38652	38652	191110

*Note:* In columns 1 and 2 we summarize our linked sample of twin brothers, linking twins from the 1900, 1910, and 1920 censuses to 1940. Column 1 are the linked twins, unweighted, while column 2 uses inverse propensity weights to account for differences between census matched and unmatched twins. In column 3, we use a random 1% sample of the 1940 census. We limit the twins to our analysis sample, focusing only on the twins who were wage and salary male workers with a weekly wage of at least \$6 who worked a positive number of weeks in the previous year (1939) and a positive number of hours in the previous week. The twins sample includes only twins who both fit our sample criteria. Similarly, the 1940 1% sample is limited to employed men aged 17-68 who are wage and salary workers and we make all the same sample restrictions based on wages, weeks worked, and hours worked to both samples.

**Table 4:** Returns to Education: Baseline

	1940 1% Sample			Twins, Pooled			Twins, Family FE	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Years of Education	0.046*** (0.000)	0.055*** (0.000)	0.038*** (0.001)	0.050*** (0.001)	0.056*** (0.001)	0.036*** (0.002)	0.044*** (0.002)	0.035*** (0.003)
Good Controls	No	Yes	Yes	No	Yes	Yes	No	No
Bad Controls	No	No	Yes	No	No	Yes	No	Yes
Twin Family FE	No	No	No	No	No	No	Yes	Yes
Observations	191110	191110	191110	38652	38652	38652	38652	38652
Adjusted R2	0.08	0.30	0.34	0.09	0.23	0.29	0.42	0.44
Y Mean	3.20	3.20	3.20	3.27	3.27	3.27	3.27	3.27

*Note:* All columns present regressions of the log of weekly earnings in 1939 on years of education, drawing on the 1940 census. In columns 1, 2, and 3, we use a random 1% sample of the 1940 census. In columns 4, 5, and 6, we turn to our linked sample of twin brothers, linking twins from the 1900, 1910, and 1920 censuses to 1940. In columns 7 and 8, we include twin family fixed effects, forcing the comparisons of education and earnings to be between twin brothers. With the twin family fixed effects, the “good” controls—age, age-squared, race, and nativity—are subsumed because they cannot vary between twins. In all cases our sample is restricted to wage and salary male workers with a weekly wage of at least \$6 who worked a positive number of weeks in the previous year (1939) and a positive number of hours in the previous week. The twins sample includes only twins who *both* fit our sample criteria. Following Bailey et al. (2017), we use inverse propensity weights to adjust for observable differences between matched and unmatched persons in our census linked sample.

**Table 5:** Returns to Education: Restricted to Twins with Milestone Education Outcomes

	1940 1% Sample		Twins, Pooled		Twins, Family FE
	(1)	(2)	(3)	(4)	(5)
Years of Education	0.044*** (0.001)	0.061*** (0.001)	0.047*** (0.002)	0.062*** (0.002)	0.056*** (0.004)
Good Controls	No	Yes	No	Yes	No
Twin Family FE	No	No	No	No	Yes
Observations	97100	97100	13094	13094	13094
Adjusted R2	0.05	0.28	0.06	0.23	0.46
Y Mean	3.28	3.28	3.32	3.32	3.32

*Note:* All columns present regressions of the log of weekly earnings in 1939 on years of education, drawing on the 1940 census. In columns 1 and 2, we use a random 1% sample of the 1940 census. In columns 3 and 4, we turn to our linked sample of twin brothers, linking twins from the 1900, 1910, and 1920 censuses to 1940. In column 5, we include twin family fixed effects, forcing the comparisons of education and earnings to be between twin brothers. With the twin family fixed effects, the “good” controls—age, age-squared, race, and nativity—are subsumed because they cannot vary between twins. In addition to the restrictions on our sample described in Table 4, we also restrict our sample to twins who *both* attained a milestone number of years of education: 8, 12, 16, or 16+. In all cases our sample is restricted to wage and salary male workers with a weekly wage of at least \$6 who worked a positive number of weeks in the previous year (1939) and a positive number of hours in the previous week. Following Bailey et al. (2017), we use inverse propensity weights to adjust for observable differences between matched and unmatched persons in our census linked sample.



**Table 6:** Returns to Education: Twin Pairs with Similar First Names

	Same First Letter		Jaro-Winkler Distance $\leq 0.2$		Same Soundex	
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	0.053*** (0.002)	0.041*** (0.005)	0.063*** (0.005)	0.045*** (0.010)	0.057*** (0.009)	0.053*** (0.026)
Good Controls	Yes	No	Yes	No	Yes	No
Twin Family FE	No	Yes	No	Yes	No	Yes
Observations	8262	8262	1644	1644	312	312
Adjusted R <sup>2</sup>	0.21	0.45	0.28	0.48	0.25	0.40
Y Mean	3.22	3.22	3.16	3.16	3.17	3.17

*Note:* All columns present regressions of the log of weekly earnings in 1939 on years of education, drawing on our linked sample of twin brothers, linking twins from the 1900, 1910, and 1920 censuses to 1940. We restrict the sample to pairs of twins who, based on their given names, we believe may be more likely to have been treated alike by their parents and families. In columns 1 and 2, we limit the sample to twins whose names start with the same letter. In columns 3 and 4, we limit to twins whose first names are quite close in string distance. In columns 5 and 6, we limit to twins whose names have the same phonetic score (using Soundex). In the even columns, we include twin family fixed effects, forcing the comparisons of education and earnings to be between twin brothers. With the twin family fixed effects, the “good” controls—age, age-squared, race, and nativity—are subsumed because they cannot vary between twins. In all cases our sample is restricted to wage and salary male workers with a weekly wage of at least \$6 who worked a positive number of weeks in the previous year (1939) and a positive number of hours in the previous week. The twins sample includes only twins who *both* fit our sample criteria. Following Bailey et al. (2017), we use inverse propensity weights to adjust for observable differences between matched and unmatched persons in our census linked sample.

**Table 7:** Returns to Education: Within or Across Occupations

	Baseline	Occupation Code FEs			Log Occscore
	(1)	(2)	(3)	(4)	(5)
Years of Education	0.044*** (0.002)	0.030*** (0.002)	0.029*** (0.002)	0.028*** (0.002)	0.019*** (0.001)
Occupation 1 Digit FE	No	Yes	No	No	No
Occupation 2 Digit FE	No	No	Yes	No	No
Occupation 3 Digit FE	No	No	No	Yes	No
Twin Family FE	Yes	Yes	Yes	Yes	Yes
Observations	38652	38652	38652	38652	38652
Adjusted R2	0.42	0.49	0.50	0.52	0.26
Y Mean	3.27	3.27	3.27	3.27	3.27

*Note:* Columns 1 to 4 present regressions of the log of weekly earnings in 1939 on years of education, drawing on our linked sample of twin brothers, linking twins from the 1900, 1910, and 1920 censuses to 1940. In column 5, we use the same sample but take the log of occupation score (using the standard IPUMS occscore variable) as the outcome. Column 1 duplicates our baseline results from Table 4. In Columns 2, 3, and 4, we add fixed effects for occupation, using the three-digit occupation code from IPUMS. The reduction in the return to education with the inclusion of these increasingly narrow occupation fixed effects suggests that some of the return to education in 1940 was driven by education changing (upgrading) occupation. But because the coefficient in column 4 is still sizable (nearly two-thirds of the coefficient in column 1), there is also substantial returns to education within occupations. In all columns, we include twin family fixed effects, forcing the comparisons of education and earnings to be between twin brothers. Our sample is restricted to wage and salary male workers with a weekly wage of at least \$6 who worked a positive number of weeks in the previous year (1939) and a positive number of hours in the previous week. The twins sample includes only twins who *both* fit our sample criteria. Following Bailey et al. (2017), we use inverse propensity weights to adjust for observable differences between matched and unmatched persons in our census linked sample.

**Table 8:** Returns to Education: Heterogeneity by Father’s Occupation and SES

	Farmer Father		Non Farmer Father					
	All		All		$\leq$ Median Occscore		$>$ Median Occscore	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Years of Education	0.049*** (0.002)	0.049*** (0.004)	0.043*** (0.001)	0.041*** (0.003)	0.045*** (0.002)	0.044*** (0.004)	0.040*** (0.002)	0.039*** (0.004)
Good Controls	Yes	No	Yes	No	Yes	No	Yes	No
Twin Family FE	No	Yes	No	Yes	No	Yes	No	Yes
Observations	10189	10189	25468	25468	12950	12950	12518	12518
Adjusted R2	0.15	0.40	0.13	0.41	0.15	0.42	0.10	0.38
Y Mean	3.13	3.13	3.32	3.32	3.28	3.28	3.36	3.36

*Note:* All columns present regressions of the log of weekly earnings in 1939 on years of education, drawing on our linked sample of twin brothers, linking twins from the 1900, 1910, and 1920 censuses to 1940. We split the data by the occupation and occupational status of the twins’ fathers. In the first two columns, we focus on twins whose father was a farmer. In the remaining six columns, we focus on twins whose father was not a farmer, splitting these twins by their fathers’ occupation scores at the median in columns 5 and 6 versus 7 and 8. In even columns, we include twin family fixed effects, forcing the comparisons of education and earnings to be between twin brothers. With the twin family fixed effects, the “good” controls—age, age-squared, race, and nativity—are subsumed because they cannot vary between twins. Our sample is restricted to wage and salary male workers with a weekly wage of at least \$6 who worked a positive number of weeks in the previous year (1939) and a positive number of hours in the previous week. The twins sample includes only twins who *both* fit our sample criteria. Following Bailey et al. (2017), we use inverse propensity weights to adjust for observable differences between matched and unmatched persons in our census linked sample.

**Table 9:** Returns to Education: Heterogeneity by Family Immigration History

	Number of Foreign-Born Grandparents					
	None Foreign-Born		1-3 Foreign-Born		All Foreign-Born	
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	0.055*** (0.002)	0.050*** (0.003)	0.044*** (0.002)	0.043*** (0.005)	0.038*** (0.002)	0.036*** (0.004)
Good Controls	Yes	No	Yes	No	Yes	No
Twin Family FE	No	Yes	No	Yes	No	Yes
Observations	14493	14493	7154	7154	11369	11369
Adjusted R2	0.14	0.39	0.10	0.43	0.07	0.37
Y Mean	3.22	3.22	3.33	3.33	3.36	3.36

*Note:* All columns present regressions of the log of weekly earnings in 1939 on years of education, drawing on our linked sample of twin brothers, linking twins from the 1900, 1910, and 1920 censuses to 1940. We split the data by the family immigration history of the twins. Because we observe twins in 1900, 1910, and 1920 in their childhood homes, we can use the census questions about mother and father’s birthplaces to identify the birthplaces of most twins’ grandparents. The sample includes only white twins to avoid conflating race with grandparent and parent immigration history, as nearly all black twins in the linked sample had four American-born grandparents. We also drop some pairs of twins from our baseline sample if both their mothers and fathers did not reside in their childhood household in the census. The return to education computed between twins with more foreign-born grandparents was lower. In even columns, we include twin family fixed effects, forcing the comparisons of education and earnings to be between twin brothers. With the twin family fixed effects, the “good” controls—age, age-squared, race, and nativity—are subsumed because they cannot vary between twins. Our sample is restricted to wage and salary male workers with a weekly wage of at least \$6 who worked a positive number of weeks in the previous year (1939) and a positive number of hours in the previous week. The twins sample includes only twins who *both* fit our sample criteria. Following Bailey et al. (2017), we use inverse propensity weights to adjust for observable differences between matched and unmatched persons in our census linked sample.

**Table 10:** Effect of Education on Migration

	Panel A. Migration					
	Moved Out of County		Moved Within State		Moved Out of State	
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	0.006*** (0.001)	0.011*** (0.002)	-0.008*** (0.001)	-0.004** (0.002)	0.015*** (0.001)	0.015*** (0.002)
Good Controls	Yes	No	Yes	No	Yes	No
Twin Family FE	No	Yes	No	Yes	No	Yes
Observations	38652	38652	38652	38652	38652	38652
Adjusted R2	0.01	0.50	0.01	0.37	0.02	0.42
Y Mean	0.60	0.60	0.32	0.32	0.28	0.28

	Panel B. 1940 Location Choice					
	Urban		Log Size of Place		Farm	
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	0.021*** (0.001)	0.015*** (0.002)	0.120*** (0.005)	0.095*** (0.011)	-0.010*** (0.001)	-0.009*** (0.001)
Good Controls	Yes	No	Yes	No	Yes	No
Twin Family FE	No	Yes	No	Yes	No	Yes
Observations	38652	38652	38652	38652	38652	38652
Adjusted R2	0.02	0.36	0.02	0.43	0.01	0.27
Y Mean	0.69	0.69	9.99	9.99	0.08	0.08

*Note:* All columns present regressions of a migration or residential outcome in 1940 on years of education, drawing on our linked sample of twin brothers, linking twins from the 1900, 1910, and 1920 censuses to 1940. In even columns, we include twin family fixed effects, forcing the comparisons of education and earnings to be between twin brothers. With the twin family fixed effects, the “good” controls—age, age-squared, race, and nativity—are subsumed because they cannot vary between twins. Our sample is restricted to wage and salary male workers with a weekly wage of at least \$6 who worked a positive number of weeks in the previous year (1939) and a positive number of hours in the previous week. The twins sample includes only twins who *both* fit our sample criteria. In Appendix Table A.7, we replicate this analysis on the complete sample of linked twins and find very similar results. Following Bailey et al. (2017), we use inverse propensity weights to adjust for observable differences between matched and unmatched persons in our census linked sample.

**Table 11:** Effect of Education on Marriage and Fertility

	Married		Any Children		Number of Children	
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	-0.007*** (0.001)	0.001 (0.002)	-0.016*** (0.001)	-0.008*** (0.002)	-0.085*** (0.003)	-0.052*** (0.007)
Good Controls	Yes	No	Yes	No	Yes	No
Twin Family FE	No	Yes	No	Yes	No	Yes
Observations	38652	38652	38652	38652	38652	38652
Adjusted R2	0.04	0.33	0.04	0.24	0.05	0.22
Y Mean	0.75	0.75	0.53	0.53	1.17	1.17

*Note:* All columns present regressions on marriage or fertility outcomes in 1940 on years of education, drawing on our linked sample of twin brothers, linking twins from the 1900, 1910, and 1920 censuses to 1940. In even columns, we include twin family fixed effects, forcing the comparisons of education and earnings to be between twin brothers. With the twin family fixed effects, the “good” controls—age, age-squared, race, and nativity—are subsumed because they cannot vary between twins. Our sample is restricted to wage and salary male workers with a weekly wage of at least \$6 who worked a positive number of weeks in the previous year (1939) and a positive number of hours in the previous week. The twins sample includes only twins who *both* fit our sample criteria. In Appendix Table A.8, we replicate this analysis on the complete sample of linked twins and find similar results for columns 3, 4, 5, and 6. The marriage result in the full sample, however, does show a positive effect on the probability of being married from an additional year of education. Following Bailey et al. (2017), we use inverse propensity weights to adjust for observable differences between matched and unmatched persons in our census linked sample.

**Table 12:** Effect of Education on Alternative Economic Outcomes

	Panel A. Occupation Scores					
	Log of Occupation Scores, Alternative Measures					
	Saavedra and Twinam		Olivetti and Paserman		1940 Occupation Median	
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	0.025*** (0.001)	0.023*** (0.002)	0.031*** (0.001)	0.026*** (0.002)	0.035*** (0.001)	0.031*** (0.002)
Good Controls	Yes	No	Yes	No	Yes	No
Twin Family FE	No	Yes	No	Yes	No	Yes
Observations	31108	31108	31020	31020	31108	31108
Adjusted R2	0.17	0.38	0.13	0.32	0.11	0.34
Y Mean	3.24	3.24	6.42	6.42	7.14	7.14
	Panel B. Alternative Measures of Economic Status					
	Top-Coded Earnings		> \$50 Non-Wage Earnings		Owns Home	
	(1)	(2)	(3)	(4)	(5)	(6)
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	0.003*** (0.000)	0.003*** (0.000)	-0.002*** (0.000)	0.003*** (0.001)	0.006*** (0.000)	0.006*** (0.001)
Good Controls	Yes	No	Yes	No	Yes	No
Twin Family FE	No	Yes	No	Yes	No	Yes
Observations	145914	145914	145914	145914	145914	145914
Adjusted R2	0.01	0.13	0.02	0.19	0.02	0.21
Y Mean	0.01	0.01	0.32	0.32	0.46	0.46
	Panel C. Labor Supply					
	Works Fulltime		Weeks of Work		Hours of Work	
	(1)	(2)	(3)	(4)	(5)	(6)
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	0.015*** (0.000)	0.015*** (0.001)	0.584*** (0.015)	0.664*** (0.034)	0.763*** (0.018)	0.829*** (0.042)
Good Controls	Yes	No	Yes	No	Yes	No
Twin Family FE	No	Yes	No	Yes	No	Yes
Observations	145914	145914	145914	145914	145914	145914
Adjusted R2	0.02	0.15	0.02	0.16	0.02	0.17
Y Mean	0.69	0.69	40.04	40.04	35.62	35.62

*Note:* In Panel A, we use three alternative measurements of occupation score where all outcomes are in logs and should be interpreted as semi-elasticities with respect to education. The scores in columns 1 and 2 are from Saavedra and Twinam (2018) and are based on earnings data from 1950 in lasso-adjusted industry, demographic, and occupation cells. In columns 3 and 4, the scores are from Olivetti and Paserman (2015) and are based on the 1900 Cost of Living Survey with imputations of farmer's incomes. The sample size is slightly lower because one occupation code—545, airplane mechanics and repairmen—could not be imputed based on the 1900 data. In columns 5 and 6, we create our own occupation scores based on the 1940 complete count census—IPUMS uses 1950—taking the median of male wage and salary workers aged 16 to 64 who worked at least 40 weeks in the preceding year and at least 35 hours in the preceding week. In Panels B and C, we relax the sample criteria and include all pairs of twins who were both merged ahead to 1940. In Panel B, we estimate the effect of education on the probability the twin has earnings top-coded (earned \$5000 or more), earned more than \$50 in non-wage earnings (our only measure of non wage and salary earnings in the 1940 census), and owned a home. In Panel C, we estimate the effect of education on the probability that the twin works fulltime (more than 40 weeks in 1939), weeks of work, and hours of work. Following Bailey et al. (2017), we use inverse propensity weights to adjust for observable differences between matched and unmatched persons in our census linked sample.

## A Appendix

### A.1 Inverse Propensity Weights

Given the lack of unique individual identifiers in the historical censuses and the limited covariates that are available for matching, any resulting linked samples are necessarily imperfect representations of the underlying populations. Bailey et al. (2017) propose using inverse propensity weights to adjust for observable differences between matched and unmatched persons. They construct these weights in two steps:<sup>44</sup>

- Step 1: Run a probit regression of link status (whether an individual is matched) on the following variables: an indicator for those with a middle name; the length of first, middle, and last names; polynomials in the day of birth and age; an index for how common the first and last names are; whether or not one has siblings and the number of siblings; and the length of the names of one’s parents.
- Step 2: Inverse propensity scores are then computed as  $\frac{1-p}{p} \frac{m}{1-m}$ , where  $p$  is the predicted likelihood of being matched based on the estimated probit coefficients and  $m$  is the actual match rate.

To make our linked samples more representative of the underlying populations, we adapt the Bailey et al. (2017) reweighting procedure, with some minor adjustments:

1. We do not include polynomials for the day of birth as this information is not available in any of the historical censuses we use. Much of Bailey et al. (2017) piece uses the Longitudinal, Intergenerational Family Electronic Micro-Database (LIFE-M) sample, which is based on a random draw of birth certificates from Ohio and North Carolina. This dataset contains more individual-level information than the historical censuses used in our paper, one of which is the day of birth.

---

<sup>44</sup>See footnote 33 of Bailey et al. (2017).



2. We do not control for the presence of siblings (but control for the *number* of siblings), as our starting sample of twin brothers automatically guarantees the presence of at least one sibling.
3. We use the names of parents, but because not all twins in our sample have both parents residing with them in a given census year we interact these terms with indicators for parent in the household.
4. We use a quadratic (second-order polynomial) in both age and year of birth because we are linking children from multiple censuses.
5. For the degree of commonness of first and last names, we use the log of the number of people in the 1900, 1910, and 1920 censuses with that first or last name.

## A.2 Using Birth Month in 1900 to Identify Twins

We use age to identify twins in the census because exact birth date (or birth month or birth year) are not available in most historical US censuses, including the 1910 and 1920 Federal Censuses. However, in 1900, the census did ask respondents their month of birth and year of birth.<sup>45</sup> Using these fields, we can examine whether or not our somewhat imperfect method of twin pairing is affecting our results. We find that the estimated return to education in the set of twins we are most confident about in 1900—those with the same age *and* birth month—is even lower than the return overall, suggesting that our conclusions are not likely to be an artifact of any errors in twin tagging.

We start by looking at the twins we identify in the 1900 census. Figure A.1 indicates that some of our twin pairs have different birth months and may not be actual twins. Overall, 55% of the twins in our baseline sample have the same birth month. Put differently, about half of the twins identified in the main text are potentially incorrect. This is consistent with Tan (2019), which also compares the twin rates in 1900 under the two different definitions. However, because there may be error in birth month—either in the enumeration or the transcription—some of the 45% of twin pairs with mismatched birth months may still be true twins. Further, because the likelihood of twins misclassification varies systematically with various household characteristics (Tan 2019), the overall direction of the bias, if any, is unclear. The effect of such misclassification on our results thus remains an empirical issue.

[Figure A.1 about here.]

Table A.1 shows that the return to education among all twins tagged using our usual twin procedure in the 1900 census was 0.053 log points. That this is larger than our baseline estimate is not surprising, given the trend we found in the return to education across cohorts (see Figure 5). The twins in 1900 are born between 1875 and 1900, with most born between 1890 and 1900.

---

<sup>45</sup>The 1910 census only records the month of birth for persons residing in Alaska.

[Table A.1 about here.]

What do the results in Table A.1 tell us about our imperfect method of identifying twins? When we have access to both age and birth month and the values all agree between twins, as in column 4, we estimate a return to education of 0.049 log points. When different birth months are recorded for twins in the census, as is the case in column 6, the return is higher: 0.056 log points. This suggests that, if anything, our imperfect method of tagging twins induces an upward bias in the estimated return to education. Thus, our main conclusion—that there was a positive return to education in 1940 but it was smaller than the return in more recent twin studies—seems unlikely to be an artifact of any errors in twin tagging.

### A.3 Twin Name Similarity

In the paper, we suggested that parents who give twins similar names also intend to treat them more similarly. Is this true? If the nurture of similarly named twins is more similar, then the identifying assumption of the twins fixed effect method—that twins have the same unobserved ability—may be more plausible. This appendix shows that twins with more similar names *do* tend to have more similar education outcomes as children, the only possible measure of parental investment available in the census.

To implement this exercise, we consider the school enrollment of twin brothers in the 1900 to 1920 censuses—the sources for our baseline sample—as well as twin brothers in the 1940 census for whom we observe both enrollment and years of schooling (highest grade completed). In all cases, we limit the sample to boy-boy twin pairs aged 7 to 17 to target children of school-going age.<sup>46</sup>

Table A.2 shows that while twins in general are very likely to have the same educational outcomes in childhood—school enrollment or current grade—those with similar names are even more likely to agree on enrollment or grade. Differences between twins with more similar names and twins with less similar names are always positive, suggesting that similarly named twins have more similar education outcomes during childhood. These differences are statistically significant at the 5% level or better in all comparisons.

[Table A.2 about here.]

---

<sup>46</sup>That all pairs of twins aged 1 or 2 are both not in school is not particularly informative about the nurture process.

## A.4 Additional Figures and Tables

[Figure A.2 about here.]

[Figure A.3 about here.]

[Figure A.4 about here.]

[Figure A.5 about here.]

[Table A.3 about here.]

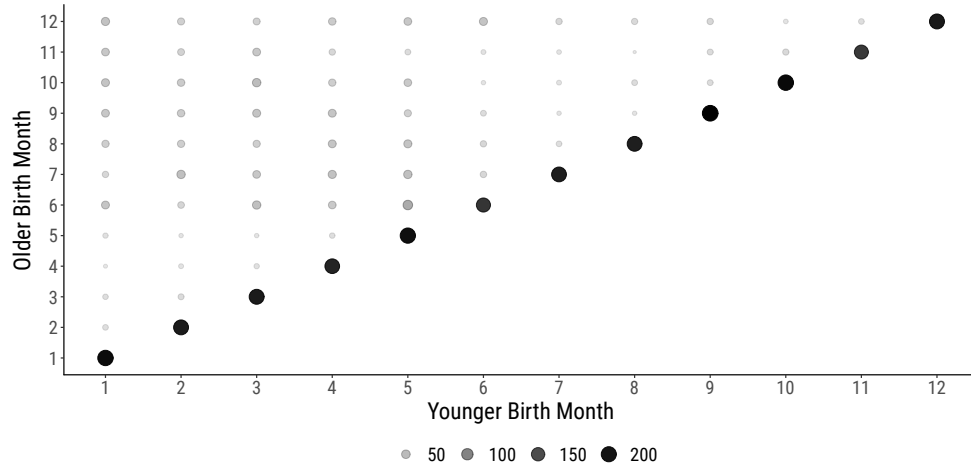
[Table A.4 about here.]

[Table A.5 about here.]

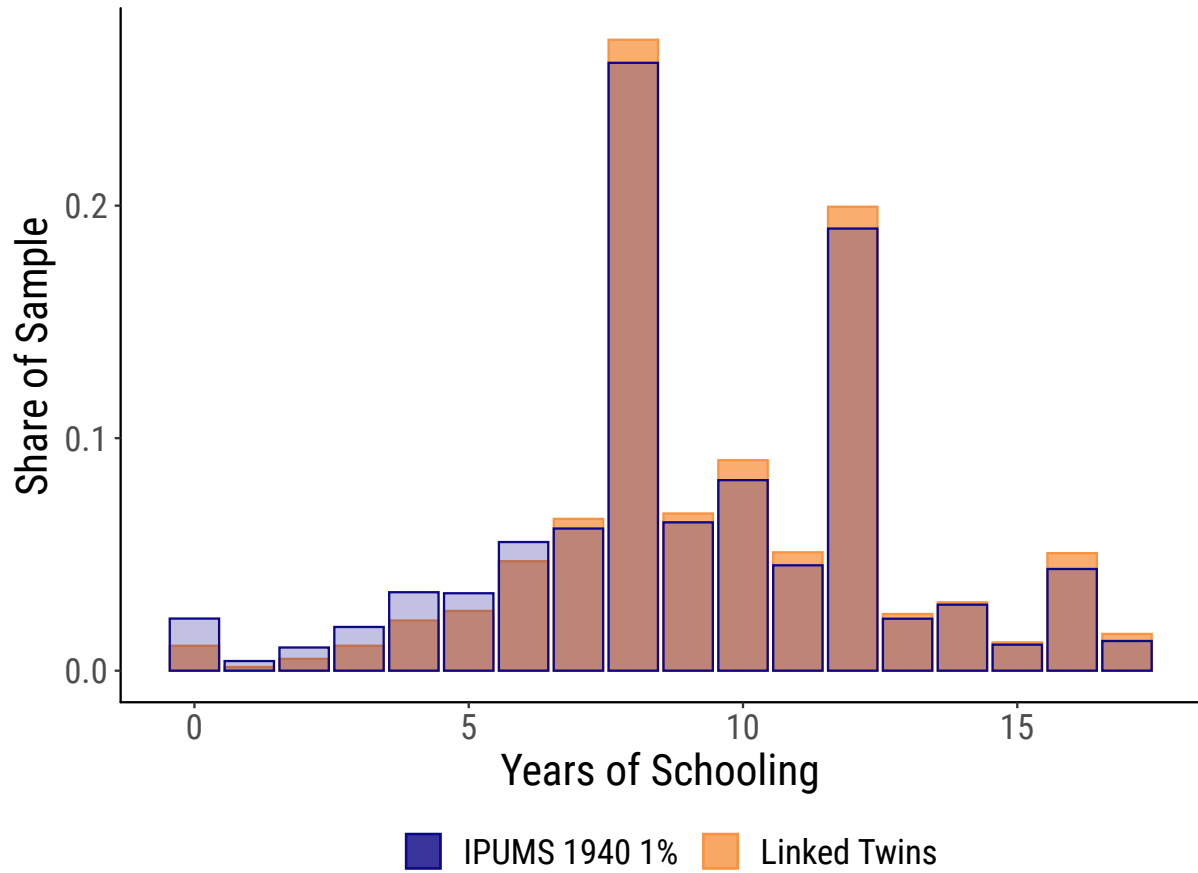
[Table A.6 about here.]

[Table A.7 about here.]

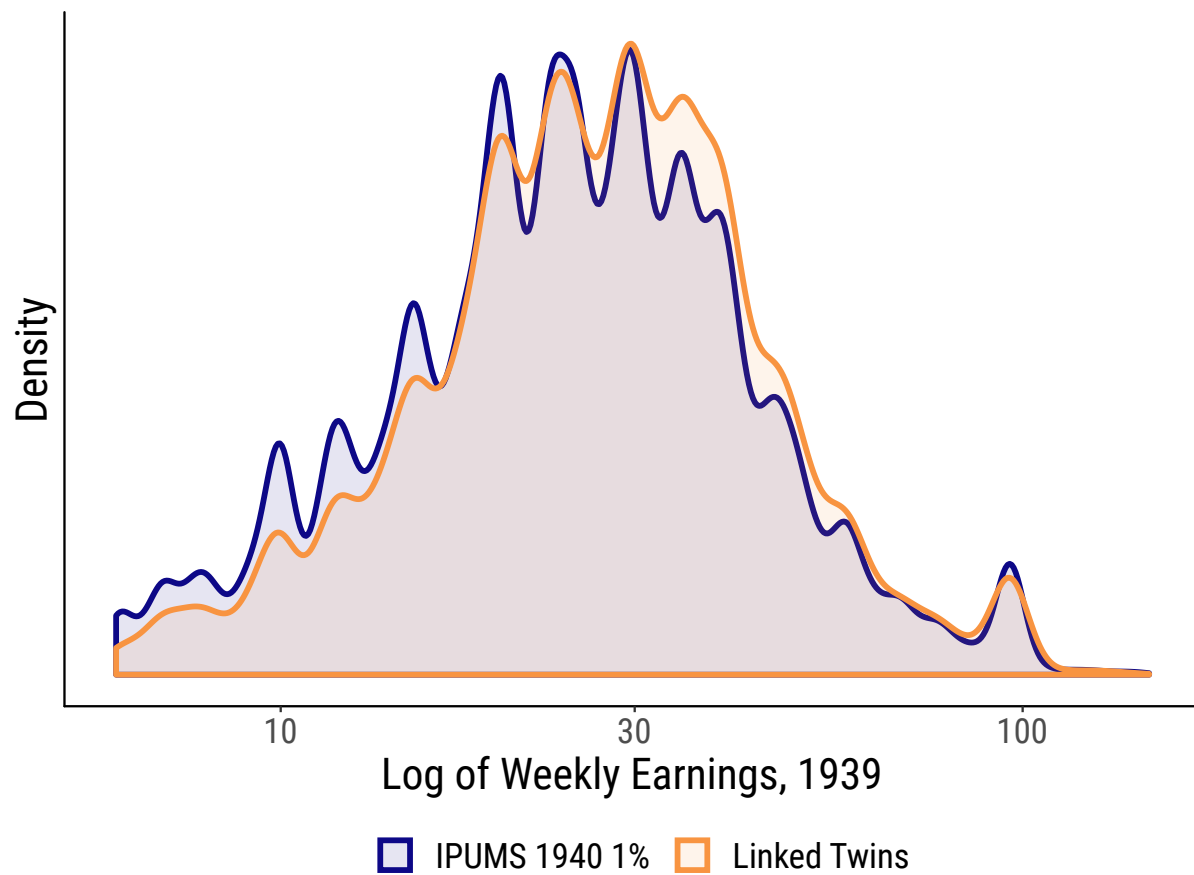
[Table A.8 about here.]



**Figure A.1:** Birth months of twins in 1900 tagged using age. Because birth month is not recorded in 1910 or 1920, we tag twins as any pair of boys living in the same household in the same family of the same age and birthplace and relationship to the head of household and with the same lastname. In 1900, we can test our procedure. This graph suggests there may be some twins in our data who are not actually twins, though the birth month variable could also be recorded with error or noise in the census. Both the size and shading of the circles indicates the number of observations at each point.

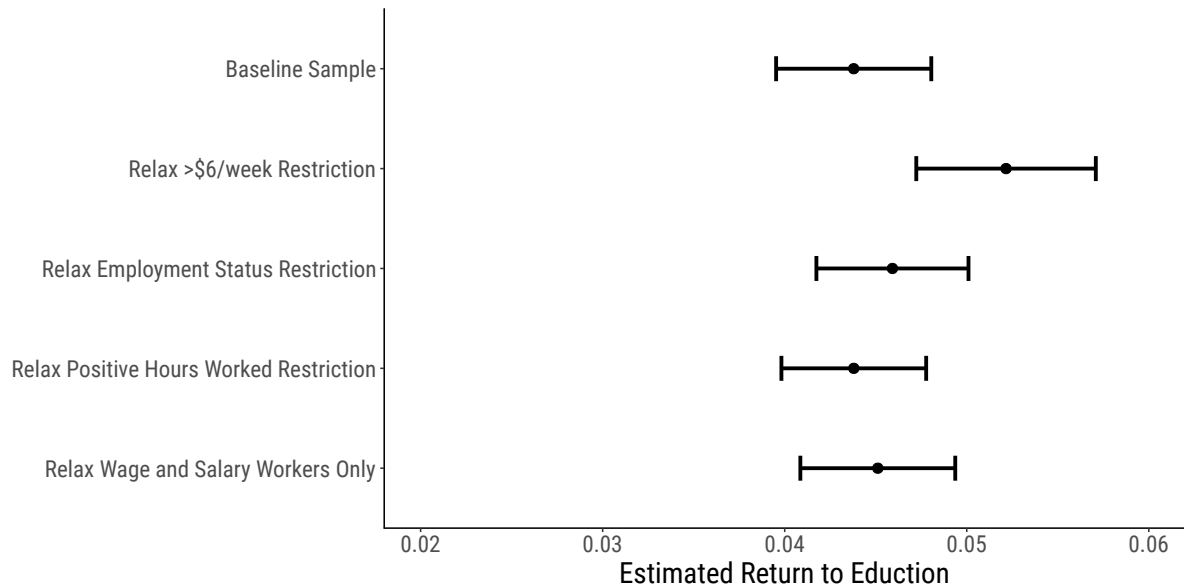


**Figure A.2:** Distribution of Years of Education in the Linked Twins Sample and the 1940 Census. Our sample of linked twins are slightly more likely to be common school (8 years), high school (12), or college (16) graduates compared to random cohort-mates in 1940 and less likely to report no schooling or 6 years of less of schooling. This difference is small and is likely driven by the smaller foreign-born and African-American shares in the linked sample (see Table 3).

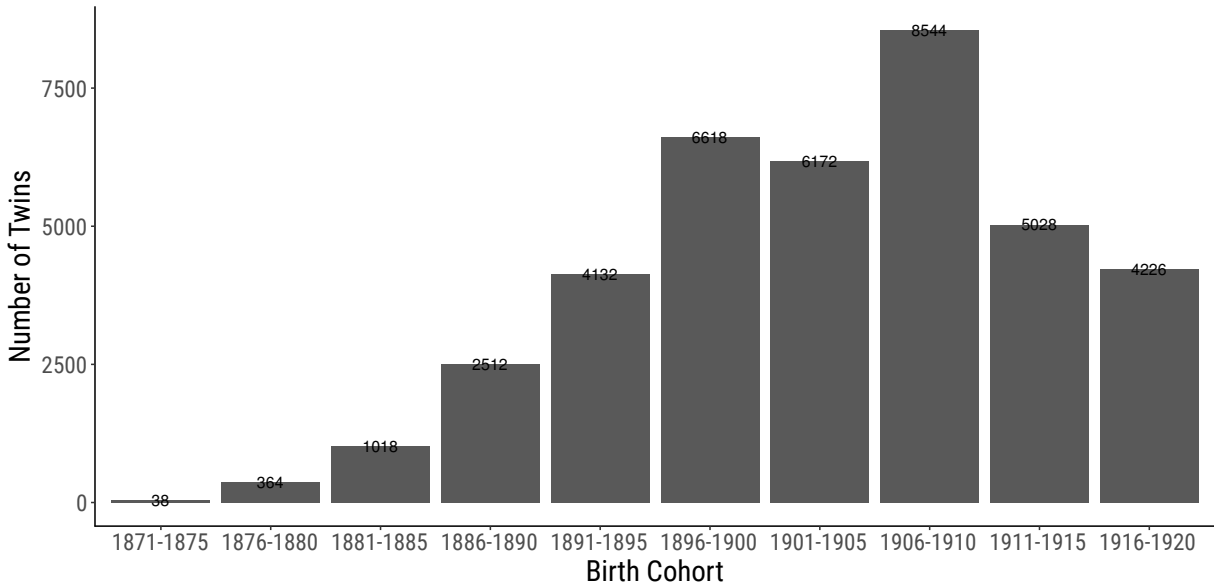


**Figure A.3:** Distribution of Log Weekly Earnings in the Linked Twins Sample and the 1940 Census.





**Figure A.4:** Robustness of the estimated return to education by sample restrictions. Our baseline sample is restricted to wage and salary male workers with a weekly wage of at least \$6 who worked a positive number of weeks in the previous year (1939) and a positive number of hours in the previous week. The twins sample includes only twins who *both* fit our sample criteria. This figure presents estimated returns to education as we relax these restrictions one by one. We cannot relax the positive weeks worked in 1939 restriction because our outcome is the log of weekly earnings, undefined when weeks worked is missing or zero. Following Bailey et al. (2017), we use inverse propensity weights to adjust for observable differences between matched and unmatched persons in our census linked sample.



**Figure A.5:** Number of twins in the baseline analysis sample in each cohort. These sample sizes explain some of the variation in confidence intervals in Figure 5. Because we only observe twins in 1900, 1910, and 1920 and can only identify twins when still living in their childhood homes (with their twin), we see very few twins who are over 20 years old.

**Table A.1:** Returns to Education: Twins Found in 1900 by Birth Month Agreement

	Baseline		Same Birth Month		Different Birth Month	
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	0.070*** (0.002)	0.053*** (0.004)	0.070*** (0.003)	0.049*** (0.006)	0.069*** (0.003)	0.056*** (0.006)
Twin Family FE	No	Yes	No	Yes	No	Yes
Observations	8546	8546	4668	4668	3878	3878
Adjusted R2	0.15	0.31	0.15	0.32	0.15	0.29
Y Mean	3.40	3.40	3.41	3.41	3.39	3.39

*Note:* All columns present regressions of the log of weekly earnings in 1939 on years of education, drawing on our linked sample of twin brothers, linking twins from 1900 to 1940. Columns 1 and 2 include all twins, where twin status is identified based on children in the same household in the same family with the same age, birthplace, relationship to head of household, and same last name. Columns 3 and 4 exploit the birth month variable, only available in the 1900 Census and limit the sample to twins with the same recorded birth month. Column 5 and 6 include twins with different recorded birth months. In the even columns, we include twin family fixed effects, forcing the comparisons of education and earnings to be between twin brothers. With the twin family fixed effects, the “good” controls—age, age-squared, race, and nativity—are subsumed because they cannot vary between twins. Our sample is restricted to wage and salary male workers with a weekly wage of at least \$6 who worked a positive number of weeks in the previous year (1939) and a positive number of hours in the previous week. The twins sample includes only twins who *both* fit our sample criteria. Following Bailey et al. (2017), we use inverse propensity weights to adjust for observable differences between matched and unmatched persons in our census linked sample.

**Table A.2:** Twins with Similar Names Have Similar Educational Outcomes in Childhood

Panel A. Same First Letter				
Census	Outcome	Not Similar Names	Similar Names	Difference
1900	Both Twins Attend School	0.891	0.919	0.028 (0.003)
1910	Both Twins Attend School	0.930	0.965	0.035 (0.003)
1920	Both Twins Attend School	0.863	0.940	0.077 (0.003)
1940	Both Twins Attend School	0.919	0.961	0.042 (0.002)
1940	Twins in Same Grade	0.668	0.798	0.130 (0.004)
Panel B. Jaro-Winkler Distance $\leq 0.2$				
Census	Outcome	Not Similar Names	Similar Names	Difference
1900	Both Twins Attend School	0.894	0.939	0.045 (0.006)
1910	Both Twins Attend School	0.935	0.974	0.039 (0.004)
1920	Both Twins Attend School	0.874	0.956	0.082 (0.004)
1940	Both Twins Attend School	0.927	0.966	0.039 (0.003)
1940	Twins in Same Grade	0.692	0.819	0.127 (0.005)
Panel C. Same Soundex				
Census	Outcome	Not Similar Names	Similar Names	Difference
1900	Both Twins Attend School	0.896	0.917	0.021 (0.010)
1910	Both Twins Attend School	0.937	0.972	0.034 (0.006)
1920	Both Twins Attend School	0.882	0.921	0.039 (0.008)
1940	Both Twins Attend School	0.931	0.954	0.022 (0.005)
1940	Twins in Same Grade	0.706	0.806	0.100 (0.010)

*Note:* Rows are based on the universe of boy-boy twins in the given census year aged 7 to 17. We study twins in 1900, 1910, and 1920 as in our main analysis, as well as twin children in 1940 when we can observe current grade in school. Panel A defines twins with similar names as twins whose names starts with the same letter. Panel B defines twins with similar names as twins whose names are within 0.2 in Jaro-Winkler string distance. Panel C defines twins with similar names as twins whose names have the same soundex code. First names are measured in the childhood census and all twins—not just those we were able to link to 1940—are included. The difference column represents the difference between similar and not similar named twins with standard errors reported in parentheses below. In all cases, the differences are positive, suggesting that twins with more similar names are more similar as children in educational outcomes.

**Table A.3:** Returns to Education: Robustness to Uniqueness in the Initial Sample of Twins

	Baseline Sample		Unique Cohorts 1		Unique Cohorts 2	
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	0.044*** (0.002)	0.035*** (0.003)	0.045*** (0.003)	0.037*** (0.003)	0.043*** (0.003)	0.036*** (0.003)
Bad Controls	No	Yes	No	Yes	No	Yes
Twin Family FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	38652	38652	25264	25264	25024	25024
Adjusted R2	0.42	0.44	0.43	0.45	0.43	0.45
Y Mean	3.27	3.27	3.22	3.22	3.22	3.22

*Note:* All columns present regressions of the log of weekly earnings in 1939 on years of education, drawing on the 1940 census. In our baseline sample of twins, we attempt to link all twins aged 0 to 25 in the 1900, 1910, and 1920 censuses. However, we might collect data on the same pair of twins more than once: a twin pair born in 1899 could be observed in 1900 at age 1, 1910 at age 11, and 1920 at age 21. To show that our results are robust to any potential double counting of twins as we pool over censuses, we define two robustness samples that partition the set of twins by birth year across censuses. In columns 3 and 4, we draw twins born 1875 to 1900 from the 1900 census, twins born 1901 to 1910 from the 1910 census, and twins born 1911 to 1920 from the 1920 census. In columns 5 and 6, we draw twins born 1875 to 1899 from the 1900 census, twins born 1900 to 1909 from the 1910 census, and twins born 1910 to 1920 from the 1920 census. In all cases, we estimate the return to education to be in line with our main findings in Table 4. All columns include twin family fixed effects; even columns add the “bad” controls. In all cases our sample is restricted to wage and salary male workers with a weekly wage of at least \$6 who worked a positive number of weeks in the previous year (1939) and a positive number of hours in the previous week. The twins sample includes only twins who *both* fit our sample criteria. Following Bailey et al. (2017), we use inverse propensity weights to adjust for observable differences between matched and unmatched persons in our census linked sample.

**Table A.4:** Ability Bias Test

	Panel A. Full Sample	
	Correlation with Schooling	
	Overall	Across Families
Married	−0.0726***	0.0073
Works Full-time	0.1132***	0.0780***
Number of Children	−0.1823***	−0.0894***
Panel B. Sample with Spouses in 1940		
	Correlation with Schooling	
	Overall	Across Families
Works Full-time	0.1374***	0.0845***
Number of Children	−0.1927***	−0.1132***
Spouse's Years of Education	0.6126***	0.4300***
Spouse in Labor Force	−0.0041	0.002

*Note:* Ability bias test using correlates of ability following Ashenfelter and Rouse (1998) and Bonjour et al. (2003). We correlate years of schooling with possible proxies for ability. We find stronger correlations between families than when we difference the measures within twin pairs and calculate within family correlations.

**Table A.5:** Returns to Education: Within or Across Industries?

	Baseline	Industry Code FEs		
	(1)	(2)	(3)	(4)
Years of Education	0.044*** (0.002)	0.042*** (0.002)	0.038*** (0.002)	0.037*** (0.002)
Industry 1 Digit FE	No	Yes	No	No
Industry 2 Digit FE	No	No	Yes	No
Industry 3 Digit FE	No	No	No	Yes
Twin Family FE	Yes	Yes	Yes	Yes
Observations	38652	38652	38652	38652
Adjusted R2	0.42	0.46	0.48	0.48
Y Mean	3.27	3.27	3.27	3.27

*Note:* All columns present regressions of the log of weekly earnings in 1939 on years of education, drawing on our linked sample of twin brothers, linking twins from the 1900, 1910, and 1920 censuses to 1940. Column 1 duplicates our baseline results from Table 4. In Columns 2, 3, and 4, we add fixed effects for industry, using the three-digit occupation code from IPUMS. The small reduction in the return to education with the inclusion of these increasingly narrow industry fixed effects suggests that a small part of the return to education in 1940 was driven by education changing (upgrading) industry. But in contrast to our finding that about one-third of the return to education comes occupational upgrading (Table 7), significantly less comes from industrial upgrading. In all columns, we include twin family fixed effects, forcing the comparisons of education and earnings to be between twin brothers. Our sample is restricted to wage and salary male workers with a weekly wage of at least \$6 who worked a positive number of weeks in the previous year (1939) and a positive number of hours in the previous week. The twins sample includes only twins who *both* fit our sample criteria. Following Bailey et al. (2017), we use inverse propensity weights to adjust for observable differences between matched and unmatched persons in our census linked sample.

**Table A.6:** Effect of Education on Labor Supply: All Linked Twins vs Main Analysis Sample

Panel A. Labor Supply, All Linked Twins						
	Works Fulltime		Weeks of Work		Hours of Work	
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	0.015*** (0.000)	0.015*** (0.001)	0.584*** (0.015)	0.664*** (0.034)	0.763*** (0.018)	0.829*** (0.042)
Good Controls	Yes	No	Yes	No	Yes	No
Twin Family FE	No	Yes	No	Yes	No	Yes
Observations	145914	145914	145914	145914	145914	145914
Adjusted R2	0.02	0.15	0.02	0.16	0.02	0.17
Y Mean	0.69	0.69	40.04	40.04	35.62	35.62

Panel B. Labor Supply, Main Analysis Sample Only						
	Works Fulltime		Weeks of Work		Hours of Work	
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	0.014*** (0.001)	0.014*** (0.002)	0.456*** (0.018)	0.449*** (0.043)	0.280*** (0.019)	0.212*** (0.047)
Good Controls	Yes	No	Yes	No	Yes	No
Twin Family FE	No	Yes	No	Yes	No	Yes
Observations	38652	38652	38652	38652	38652	38652
Adjusted R2	0.02	0.13	0.02	0.16	0.01	0.16
Y Mean	0.82	0.82	46.15	46.15	43.95	43.95

*Note:* In Table 12, we estimated the effects of education on labor supply. Our sample included any pair of twins we were able to match both to 1940. We duplicate that analysis in Panel A here. In Panel B, we return to our main analysis sample—the sample we use to estimate the earnings return to education in Table 4 for example. That sample, recall, was limited to pairs of twins who both earned \$6 per week in 1940 and worked a positive number of weeks and a positive number hours. While such sample selection could make sense for estimating returns to education—an intensive margin question—this selection on the dependent variable could be problematic when studying the effects of education on labor supply, especially when labor supply is measured as either hours worked or weeks worked. Indeed, we see some evidence of that in this table as the relative magnitudes of the odd columns (without twin fixed effects) and the even columns (with twin fixed effects).

Following Bailey et al. (2017), we use inverse propensity weights to adjust for observable differences between matched and unmatched persons in our census linked sample.



**Table A.7:** Effect of Education on Migration, Full Sample

	Panel A. Migration					
	Moved Out of County		Moved Within State		Moved Out of State	
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	0.002*** (0.000)	0.008*** (0.001)	-0.009*** (0.000)	-0.006*** (0.001)	0.011*** (0.000)	0.013*** (0.001)
Good Controls	Yes	No	Yes	No	Yes	No
Twin Family FE	No	Yes	No	Yes	No	Yes
Observations	145914	145914	145914	145914	145914	145914
Adjusted R2	0.01	0.42	0.01	0.29	0.02	0.34
Y Mean	0.61	0.61	0.34	0.34	0.27	0.27
	Panel B. 1940 Location Choice					
	Urban		Log Size of Place		Farm	
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	0.032*** (0.000)	0.022*** (0.001)	0.183*** (0.002)	0.122*** (0.005)	-0.028*** (0.000)	-0.018*** (0.001)
Good Controls	Yes	No	Yes	No	Yes	No
Twin Family FE	No	Yes	No	Yes	No	Yes
Observations	145914	145914	145914	145914	145914	145914
Adjusted R2	0.06	0.37	0.05	0.43	0.05	0.36
Y Mean	0.52	0.52	9.00	9.00	0.26	0.26

*Note:* All columns present regressions of a migration or residential outcome in 1940 on years of education, drawing on our linked sample of twin brothers, linking twins from the 1900, 1910, and 1920 censuses to 1940. In even columns, we include twin family fixed effects, forcing the comparisons of education and earnings to be between twin brothers. With the twin family fixed effects, the “good” controls—age, age-squared, race, and nativity—are subsumed because they cannot vary between twins. Our sample include all twins who were *both* linked to 1940, without any further restrictions. The results are very similar to Table 10. Following Bailey et al. (2017), we use inverse propensity weights to adjust for observable differences between matched and unmatched persons in our census linked sample.

**Table A.8:** Effect of Education on Marriage and Fertility, Full Sample

	Married		Any Children		Number of Children	
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	-0.003*** (0.000)	0.005*** (0.001)	-0.011*** (0.000)	-0.002*** (0.001)	-0.075*** (0.001)	-0.038*** (0.003)
Good Controls	Yes	No	Yes	No	Yes	No
Twin Family FE	No	Yes	No	Yes	No	Yes
Observations	145914	145914	145914	145914	145914	145914
Adjusted R2	0.05	0.30	0.04	0.21	0.04	0.22
Y Mean	0.72	0.72	0.52	0.52	1.28	1.28

*Note:* All columns present regressions on marriage or fertility outcomes in 1940 on years of education, drawing on our linked sample of twin brothers, linking twins from the 1900, 1910, and 1920 censuses to 1940. In even columns, we include twin family fixed effects, forcing the comparisons of education and earnings to be between twin brothers. With the twin family fixed effects, the “good” controls—age, age-squared, race, and nativity—are subsumed because they cannot vary between twins. Our sample include all twins who were *both* linked to 1940, without any further restrictions. The results are similar to Table 11. Following Bailey et al. (2017), we use inverse propensity weights to adjust for observable differences between matched and unmatched persons in our census linked sample.