The Return to Education in the Mid-20th Century: Evidence from Twins*

James J. Feigenbaum† Hui Ren Tan‡

May 31, 2020

Abstract

What was the return to education in the US at mid-century? In 1940, the correlation between years of schooling and earnings was relatively low. 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. We find that each additional year of schooling increased labor earnings by approximately 4 percent, about half the return found for more recent cohorts in twins studies. These returns were evident both within and across occupations and were higher for sons from lower SES families.

*For detailed feedback and helpful suggestions, we thank Ran Abramitzky, Michael Andrews, Orley Ashenfelter, Brian Beach, Kevin Lang, Erika Lee, Bob Margo, Ross Mattheis, Christopher Muller, Ee Cheng Ong, Jessica Pan, Hanna Schwank, Kelvin Seah, and Marybeth Train. Thomas Pearson and Erika Lee provided excellent research assistance. All errors are our own.

†Department of Economics, Boston University, jamesf@bu.edu

‡Department of Economics, National University of Singapore, huiren@nus.edu.sg

The American labor market was at a crossroads of major transitions and economic events 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 in 1940, the economic and social upheaval of World War II (WWII) had yet to hit the country. The Great Compression—a period of declining wage dispersion that saw earnings inequality shrink across education, experience, and occupation groups—was also just beginning (Goldin and Margo 1992).

The relative price paid for educated or skilled workers—an important feature of any economy—is vital to the scholarly understanding of the American economy at mid-century as the supply of skilled workers rose and inequality fell. In recent decades, growth in the skill premium explains about two-thirds of the rise in inequality (Goldin and Katz 2007) and income inequality is much higher at the end of the 20th century than it was at mid-century (Saez 2013; Kopczuk et al. 2010). If slowing growth in the supply of educated workers (Goldin and Katz 2009) drives inequality today, did expansion in the supply at mid-century have the opposite effect? If mass education helped tamp down inequality, lower returns to schooling would be clear evidence. Correlations between education and earnings point in this direction: 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.[1]

However, the trends documented by Goldin and Katz (2009) are based on associations rather than causal estimates—the claim that education returns were lower at mid-century thus needs to be re-investigated. Self-selection may bias simple Mincerian correlations. If high-ability individuals tend to acquire more years of schooling, then any positive association between earnings and education might reflect the higher productivity of such people, independent of schooling. That is, the skill premium could be biased upwards. On the other hand, if the opportunity cost of education pushes high-potential earners out of school earlier, then the estimates would be biased downwards. While studies suggest that the bias in Mincerian regressions is small in the late 20th century (Card 1999), this may or may not have been true further back in time. Beyond selection bias, measurement error in years of schooling might also vary over time, and such error could attenuate the estimated returns. With different forces at play, the overall direction of bias in the

[1] Differences in data make such intertemporal comparisons challenging. The return to education in 1915 is based on the 1915 Iowa State Census (Goldin and Katz 2000), while the 1940 estimate is based only on labor and wage earnings as non-wage income is not recorded in the 1940 census. Intertemporal comparisons are further complicated by the growth in educational attainment: high school and college graduates in the past represent very different parts of the education distribution compared with the present.

Mincerian correlations is unclear, especially during a period of large changes in the labor market. We offer a new look at the return to schooling in the middle of the 20th century, paying particular attention to both selection and measurement error.

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 during adulthood. If twins have similar innate characteristics or abilities, a within-twins comparison can then be used to address the confounding effects of unobserved differences.

Applying the within-twins estimator to a historical setting offers two key advantages. First, the same approach has been used to estimate the return to schooling for more recent cohorts. To the extent that any inherent biases in the twins method are similar in the past and present, this allows us to compare our estimates with twins studies in the late 20th century and to shed light on how the value of education has changed over more than 50 years. 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 other twins studies, many of which rely on small samples from local surveys or state-specific registers.

We undertake our analysis recognizing the limitations of the twins methodology. Some of these weaknesses are common across twins studies, while others are specific to our historical setting and data. A major critique of all twins studies is that even identical twins are not exactly the same. The research design only "works" if twin brothers complete different years of education, but such differences could arise from unobserved differences in characteristics that might 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 to assess 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 and our new tests rely on assumptions about parent behavior, the evidence as a whole suggests that our baseline findings are unlikely to be driven by such confounders.

Studying twins in a historical setting presents its own unique challenges, many of which we are able to address. First, we need to link individuals across censuses to identify twins and construct our key variables. Without unique individual identifiers, the historical census linking process is never perfectly accurate. Reassuringly, our results are robust to alternative linking methods.[2] Second, we can only link men—a com-

[2]Bailey et al. (2019) and Abramitzky et al. (2019) discuss the current state of historical record linkage. Bailey et al. (2019) emphasize 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

mon 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. As such information is not available in the censuses, we use key milestones in education attainment—which respondents may be more likely to remember—to provide alternative 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 the same environment. However, even same-sex twins can be wrong about their identical-versus-fraternal status; for example, nearly one-fifth of twins were wrong in the Add Health longitudinal study (Conley and Fletcher 2017, Chapter 2). Consequently, MZ and DZ status may not be well-measured in recent studies either. Moving beyond genetic similarity, we introduce measures of twin name similarity that may capture how similarly parents intend to treat their twins. This provides a proxy for "identicalness" in nurture rather than nature.[3]

We find that the return to education in the US was positive in 1940 but smaller than the estimates for more recent periods. The positive historical returns arise from education affecting both access to "better" jobs as well as earnings within occupations, with the latter driving two-thirds of the effect. Returns in the late 20th century tend to be relatively large across different identification strategies (Card 1999), including within-twins comparisons (Ashenfelter and Krueger 1994; Ashenfelter and Rouse 1998; Rouse 1999; Behrman and Rosenzweig 1999), the quarter-of-birth instrument (Angrist and Krueger 1991), and the distance-to-college instrument (Card 1993). Our smaller historical results, juxtaposed with the larger returns more recently, are consistent with the U-shaped trend in Goldin and Katz (2009).[4]

Our large sample of twins also enables us to estimate the return to education across two important background characteristics, shared by twins but differing across pairs. First, returns are higher for twin sons whose fathers were farmers or were of lower socio-economic status (SES). Second, the value of schooling varies by family immigration history: twin brothers with more foreign-born grandparents earn relatively less

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. Our results are robust to variants of the main machine learning algorithm based on Feigenbaum (2016) that we use, as well as the iterative approach popularized by Abramitzky et al. (2012).

[3]Another challenge when estimating the return to education in 1940 is that the measure of earnings refers to wage and salary earnings rather than total income.

[4]Although the simple Mincer regressions yield the "correct" answer—that the average return to education was relatively low in 1940—this should not be interpreted as the absence of selection or ability bias. More likely, it reflects the different biases cancelling each other out approximately. Should the biases change over time, then the cancelling result may not hold in all periods of study.

for the same level of education, possibly due to differences in the quality of schools that the descendants of immigrants and the US-born attended.

We also use the twins research design to study the effect of education on outcomes beyond earnings. We show that twins with more schooling had a higher probability of being employed and also supplied more weeks and hours of labor. Additional years of education lowered rates of self-employment and relief employment, as well. Finally, those who had more education were also more likely to migrate but less likely to have children, though the magnitudes of these effects are relatively small.

In estimating the return to education at mid-century, our paper is closest to Clay et al. (2016), who exploit early 20th-century changes in compulsory schooling laws (CSLs) to instrument actual education attainment with compelled years of schooling, differencing out cohort and state-of-birth fixed effects.[5] To estimate the CSL effects, Clay et al. (2016) compile a new dataset of CSLs across states from 1880 to 1930. They find returns of 6.4 to 7.9 percent using 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 the returns throughout the earnings distribution with quantile regressions. In addition, while Clay et al. (2016) focus on white men, their CSL method can be applied to women and minority groups who are difficult to link accurately across censuses. However, unlike the twins approach, it is difficult to compare their historical results with 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 null result deviates 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 the CSL thresholds—as it did in the US over the 20th century (Goldin and Katz 2009)—the CSL compliers or those on the margin of treatment change 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 populations in the past

[5]Two other relevant studies are Ward (2019a) and Parman (2015) which both use brother fixed effects to study labor markets in the early 20th century. Ward (2019a) estimates the return to internal migration in the early 20th-century US. To benchmark the importance of geographic mobility, he compares that return with the return to schooling, which he estimates to be 5.5 percent using a within-brothers (not twins) analysis. These results are consistent with our finding that the return to education was positive but smaller in 1940 than today. Meanwhile, Parman (2015) focuses on the effects of health on education, exploiting variation in exposure to the 1918 Influenza Pandemic.

and present may be more comparable. Treatment effects estimated from twins may also be closer to the population parameter as we observe people across the education distribution.

The paper proceeds as follows. We begin by describing 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. We then present our estimated return to education in 1940 and provide several robustness tests. Beyond these average estimates, we also use our large sample of twins to explore whether there were any heterogeneous returns by family status and immigration history. This is followed by an analysis of how education may have affected other outcomes beyond earnings, such as labor supply, migration, and family structure.

Historical Sample of Twins

Complete Count Census Data

Our main variables of interest, years of schooling and weekly earnings, come from the complete records of the 1940 Federal Census. These data are available though the Integrated Public Use Microdata Series (IPUMS) (Ruggles et al. 2020). We estimate the return to education in 1940 and not another year for three reasons. First, education and earnings were not recorded in federal censuses before 1940.[6] Second, the complete census enumerations with names are essential for linking individuals, but these are only made available 72 years after a given census is taken to protect the privacy of respondents. The 1940 records are thus the latest available full counts. Third, the return to schooling in the middle of the 20th 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 male worker in 1940 had 8.6 years of schooling, with 24.8 percent finishing at least high school and 6.0 percent completing college.[7] In practice, years of education is top-coded at five or more

[6]While some states did record education and earnings in state censuses before 1940, notably Iowa in 1915 and 1925, 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, which is likely to be a non-random subset of the population. Along these lines, Feigenbaum (2018) documents how restricting a sample of children to those who do not move out of state can bias estimates of intergenerational mobility.

[7]We calculate this using the 1940 1 percent sample from IPUMS, restricted to all men aged 25 to 64 who were part of the labor force.

years of college, but this affects just 1.5 percent of the population. Though commonly used in the economic history literature, the measure of education in the 1940 census has its limitations. For example, it does not adjust for school quality, which was considerably worse for African Americans at the time (Margo 1986). Measuring educational attainment based on grade completion is also a distinct concept from years of schooling, particularly for African Americans who took more than a year to complete a grade on average and who often attended ungraded schools (Margo 1986). However, as our baseline sample of twins primarily consists of whites, these quality issues with the education data are unlikely to be crucial to our conclusions.

Our main outcome of interest is the log of weekly earnings.[8] We construct weekly earnings by dividing annual labor earnings by the reported number of weeks worked. We note five 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 only recorded wage and salary earnings. We are thus unable to study how education affected earnings from businesses or other sources and will need to restrict our analysis to wage and salary workers. Third, wage responses were top-coded at $5,000, though only 1 percent of our twins sample was top-coded. 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. Fifth, the number of weeks worked may also be measured with error.

Locating Twins in the Censuses

We identify twin brothers in the complete enumerations of the 1900, 1910, and 1920 censuses (Ruggles et al. 2020). 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. As with most analyses involving linked historical data, we focus on men because women tend to change their last names upon marriage during this period, making it difficult to track them across censuses. Our sample thus 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. This compensates for the rarity of twin births and imperfections in record linkage, both of which shrink the final sample size.

From the full counts, we identify approximately 900,000 children, aged 0 to 25, in each decennial census

[8] Clay et al. (2016) also use log weekly earnings as their outcome of interest. Our key conclusion, that the return to schooling was smaller in 1940 compared with twins estimates from the late 20th century, can also be reached with log hourly earnings.

who have a twin sibling. Of these, more than 200,000 are in boy-boy pairs.[9] We plot the overall twin rate in Figure 1 and describe the rate and number of twins by census year in Table 1. As Figure 1 makes clear, we are more likely to identify twins among younger children or those born closer to the enumeration year. This 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, preventing us from finding older twins.

[Figure 1 about here.]

[Table 1 about here.]

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 lower than the contemporary rate of 33.3 twins per 1,000 people (Martin et al. 2018). There are three reasons for this. First, our sample includes everyone aged 25 or younger but we are unable to pick out twins if either twin has left the childhood household. Second, the advent of fertility treatments raised the likelihood of twin births substantially. Kulkarni et al. (2013) estimate that by 2011, 36 percent of twin births in the US were due to fertility treatments such as in-vitro fertilization. Third, child mortality rates in the US are lower today than they were in the early 20th century (Preston and Haines 1991). Because our method of identifying twins requires both twins to reside in their childhood household when the decennial census is taken, children born as twins but whose twin sibling is deceased will *not* be identified as twins.

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.[10] A random subsample is then drawn and manually matched or "trained"

[9] As we collect twins aged 0 to 25 in each census, it is possible that we will include a twin pair more than once in our universe of 200,000 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. Table A.4 in the Online Appendix shows that this is unlikely to compromise our findings—the return to education estimated with the full sample lines up with the returns based on alternative samples that have non-overlapping ages in each of the three initial-year censuses.

[10] In all waves, the census records age in years rather than year of birth. We can estimate the likely year of birth based on age, but because each census is taken on a different date (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 for our sample, to validate our twins construction and show that our conclusions are unlikely to be driven by imprecision in twins tagging (see Online Appendix A.1).

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 sufficiently high both in absolute terms and relative to any alternative options.[11] This produces a preliminary linked sample of 312,369 individuals with a match rate of 45 percent.[12] 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).

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 from within-twins variation. Second, we limit the sample to twin pairs where both brothers are wage and salary workers in 1940.[13] As stated earlier, the 1940 census records earnings from wages but not other sources of income. Consequently, the earnings of self-employed people are likely understated.[14] 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.[15] The final dataset comprises 38,652 individuals or 19,326 pairs of twins.[16]

Our twins sample is large compared with twins datasets from the late 20th century that have been used to estimate the return to schooling. Ashenfelter and Krueger (1994) and Ashenfelter and Rouse (1998),

[11] In our baseline sample, we use a relatively strict threshold to tamp down on false positives. Two hyperparameters govern our matching—how absolutely and relatively good a link has to be. To choose these hyperparameters, 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. As a robustness check, we will vary the weight on PPV from 1 (least conservative on accuracy but most likely to recall true matches) to 10 (most conservative).

[12] Our match rate is higher than 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.

[13] Twin pairs where one or both brothers were employed in emergency relief work in 1940 are excluded. However, in Table 10 we explore the effects of education on employment, self-employment, and working on emergency relief.

[14] Figure A.4 in the Online Appendix shows that our results are robust to including these non-wage earners and to relaxing other restrictions on the sample.

[15] Clay et al. (2016) use similar sample restrictions to exclude non-wage earners when studying the effects of CSLs on labor earnings in 1940. 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.

[16] Table A.5 in the Online Appendix traces exactly how the census linking and sample restrictions shrink the sample size from the original set of all boy-boy twin pairs in the 1900, 1910, and 1920 censuses.

for example, have 149 and 340 pairs of US twins, respectively. Drawing on data from the United Kingdom, Bonjour et al. (2003) study 214 pairs. Samples based on Scandinavian registry data tend to be larger. Bingley et al. (2009), for instance, have information on 4,809 pairs of Danish twins, while Isacsson (2004) looks at 6,210 pairs of Swedish twins.

We compare our sample both as children and as adults to the general population and find that while our linked twins are more likely to be white and significantly less likely to be foreign-born, they are not an extreme subset of the population.

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 percent random sample from all boys aged 25 or younger who are not the head or head's spouse in the 1900, 1910, and 1920 censuses.[17] We note three differences here. First, twins tend to be younger, possibly because the likelihood of identifying twins declines with age as individuals leave their childhood households. Second, linked twins are more likely to be white—95 percent of our analysis sample is white, compared with 86 percent in the full population of twins. 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 some caution when extrapolating the return to schooling for twins to the general population.

[Table 2 about here.]

We also find that twins are broadly similar to their cohort-mates during adulthood. Table 3 compares our twins with a 1 percent random sample drawn from the 1940 census. For consistency, the comparison group is restricted to male wage and salary workers aged 17 to 68, and the same sample restrictions in the baseline sample are imposed on wages, weeks worked, and hours worked. 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.]

[17] Throughout the paper, we compare twins to various 1 percent random samples from the complete count censuses. We use these samples rather than the full censuses to reduce computational time costs. Specifically, using the IPUMS variable `subsamp`, we draw the 44th random cut of the complete count data in all censuses. We also limit the sample to individuals not in group quarters.

In the Online Appendix, we also show that the education and earnings distributions for our linked twins and the general population in 1940 overlap closely (Figures A.5 and A.6).

The Return to Schooling in the First Half of the 20th Century

Empirical Strategy

The twins identification strategy has a long history in labor economics as a method of estimating the causal return to education. Naive comparisons of earnings and schooling may overstate the rate of return, as individuals with more schooling may also have higher unobserved ability.[18] 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. MZ twins have near-identical genetic makeups, allowing researchers to better "control" for nature and nurture.[19] DZ 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 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:

$$logW_{ih} = \alpha + \beta^{OLS} \cdot Schl_{ih} + X_{ih} + \varepsilon_{ih} \tag{1}$$

$$logW_{ih} = \alpha + \beta^{FE} \cdot Schl_{ih} + \gamma_h + X_{ih} + \varepsilon_{ih} \tag{2}$$

for individual i and household h. The outcome is the log weekly wage, $logW$, and the main independent variable is the years of schooling, $Schl$, based on the highest grade completed. Regression (1) is a simple OLS benchmark that uses variation both across and between all twin brothers. Regression (2) implements the twins design by adding in family fixed effects, γ, exploiting only variation between twins to estimate the return to education.

We partition the set of controls, X, into two groups. The first is a vector of predetermined characteristics:

[18]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.

[19]Recent studies have questioned how identical MZ twins are. Fraga et al. (2005) document epigenetic differences between MZ twins and Bruder et al. (2008) locate genetic differences between them.

a quadratic in age, a race dummy, and an indicator for foreign-born people. As these controls do 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 measured in 1940. These are endogenous controls as they may be outcomes of schooling itself and are recorded 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. We conduct several indirect tests for the validity of this assumption after presenting the baseline results.

When using our linked sample for the analysis, we weight all regressions to account for differential difficulty in linking records from one census to another. Given the lack of unique individual identifiers in the historical censuses and the limited covariates that are available for matching, historical linked samples are necessarily imperfect representations of the underlying populations. Following Bailey et al. (2019), we use inverse propensity weights to adjust for observable differences between matched and unmatched people. We describe the weighting procedure in Online Appendix A.2.

Main Results

Figure 2 summarizes our main findings, plotting within-twins differences in earnings against within-twins differences in education: an additional year of schooling raises earnings by 0.043 log points or more than 4 percent. To estimate the return to education more formally, we turn to the regressions specified in (1) and (2).

[Figure 2 about here.]

We estimate that the return to schooling was around 4 to 5 percent in the first half of the 20th century.[20] Table 4 presents our results. Columns (1) to (3) are based on the 1940 1 percent census sample presented earlier in Table 3, while columns (4) to (8) use the sample of linked twins. We compare the results from

[20] Using indicators for those who completed at least high school or who finished college instead of years of schooling as the explanatory variable, we find increases in weekly earnings of 23.5 and 38.7 percent, respectively.

both datasets to determine whether the returns for the general population and twins are likely to differ. There are two takeaways from Table 4. First, the OLS coefficients are reasonably similar across the two datasets. The coefficients on years of education are nearly identical when comparing columns with the same set of covariates—no controls, exogenous controls, or exogenous and bad controls together. If the direction and magnitude of bias in the OLS results are similar for the two groups, then the actual return to education may also 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 family fixed effects. 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 in the pooled regressions, without or with the exogenous controls, respectively.[21] Again, this effect shrinks when the set of potentially endogenous controls are included in column (8).

[Table 4 about here.]

Our results paint a mixed picture regarding the direction of bias in naive OLS estimates. When predetermined controls are added, from columns (1) to (2) and columns (4) to (5), the point estimates increase in magnitude. This suggests that those with more schooling are negatively selected. However, the fixed effects coefficients are always smaller than the corresponding OLS results, which may instead hint at the more conventional story of positive selection.

Our findings suggest a lower return to schooling at mid-century than has been previously documented.[22] Clay et al. (2016), exploiting changes in CSLs during the first half of the 20th century, estimate a return with a lower bound of 6.4 percent. Twins estimates from the late 20th century, on the other hand, hover around 10 percent (Ashenfelter and Rouse 1998; Behrman and Rosenzweig 1999; Rouse 1999).[23] That our results

[21]Following Hungerford and Solon (1987) and Jaeger and Page (1996), we have also run OLS regressions with indicators for each year of schooling rather than a continuous variable. We find a reasonably linear relation between earnings and education, except for those attaining 16 years of schooling (college graduates) where a larger jump in returns is observed. However, caution is necessary when interpreting these results as they are based on simple OLS models. The same non-linear approach cannot be directly extended to the within-twins specification, which exploits differences in schooling between twins and not the actual level of schooling per se.

[22]We hesitate to take a strong stand on whether or not our estimates offer a lower or upper bound on the return to schooling. As we showed in Tables 2 and 3, while our linked sample is not dramatically unrepresentative of the full population, the share white is higher and twins are more likely to be married and less likely to be foreign-born. Each of these differences in demographics could affect the return to education, but the magnitudes and directions would be speculative.

[23]While we focus on male twin pairs, twins studies from the late 20th century typically pool male and female twins together—this is unlikely to distort a comparison of our results with these studies. Using the sample of twins from Ashenfelter and Rouse (1998), we show in Online Appendix A.3 that the return to schooling is similar even when restricted to only male twin pairs.

differ from those in Clay et al. (2016) despite focusing on a similar period may not be surprising, as 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 cohorts, many of which are based on the twins approach. Such a comparison suggests that the value of schooling has increased over time. Goldin and Katz (2007) show that the return to education was 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.[24]

Sensitivity Analysis

Three potential issues may distort our findings: 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 concerns in turn and provide suggestive evidence that our estimates are unlikely to have been severely biased by these issues.

We begin with possible errors or imperfections in record linkage. 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 alter the relative conservatism of the linking, which Abramitzky et al. (2019) recommend as robustness tests for research using historical record linkage.

We use two methods to link twins. The first, presented throughout the paper as our baseline, is the machine learning approach of Feigenbaum (2016) described above. The second is a 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. This 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 adjust the degree of conservatism. For the machine learning approach, this is done by varying the relative weight on false positives versus

[24]There is some debate on whether the return to schooling exhibited the U-shaped pattern described in Goldin and Katz (2009). 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 onward and implementing a twins design for earlier periods is beyond the scope of this paper.

false negatives. For ABE, we report both the classic method and a more conservative version described in Abramitzky et al. (2019) that further requires names to be unique within a 5-year age band in both the base and target datasets as well as an exact match on age.

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 the classic and conservative ABE procedures. Even the result least in line with the rest—based on the much smaller sample of twins produced by the conservative ABE method—actually accords with our main conclusion, that the return to education in 1940 was positive but relatively low.[25]

[Figure 3 about here.]

While concerns with record linkage are specific to our setting, measurement error poses a common challenge in the twins literature.[26] Rouse (1999), for example, finds that 8 to 12 percent of the variation in reported schooling across twins stems from errors. In the presence of measurement error, the within-twins estimator could potentially increase the noise-to-signal ratio and attenuate the estimated returns (Griliches 1979). This is particularly concerning in our setting, as the smaller returns we estimate compared with the present might simply reflect the bias due to measurement error rather than genuine differences in the value of schooling at different points in time. One solution, first introduced by Ashenfelter and Krueger (1994), is to instrument each twin's own reported years of schooling with the level of education that the other twin reports for them. However, the information required to implement this method—namely asking one twin their sibling's years of education—is not available in the historical censuses.

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. The subset of these twin pairs

[25]The estimated return to education in the two ABE methods have changed slightly from the first working paper version of this manuscript as we now use the links made by Abramitzky, Boustan, and Eriksson directly rather than using our own implementation of their algorithm. While the algorithm is generally straightforward, there were minor differences in the exact implementation that led to slightly different samples and slightly different estimated returns.

[26]One source of measurement error in education is highlighted by Margo (1986): education creep, which is the phenomenon where individuals tend to exaggerate their educational accomplishments as they age. Our twins design, which explicitly compares twins with the exact same age, should minimize the effects of this problem.

where each sibling achieved a different milestone provides the variation needed to identify the return to education.[27]

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. The first four columns show the OLS results, where columns (1) and (2) are based on the random sample of men in 1940 and columns (3) and (4) are based on our twins sample. 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 slightly. Our preferred within-twins estimate 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 for more recent cohorts.[28]

[Table 5 about here.]

Finally, we consider a major critique of using twins to estimate the return to schooling: even identical twins are not exactly the same. Any 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. We provide suggestive evidence 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 data from the late 20th century, 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 consider 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

[27] Among twin pairs where both siblings achieved a milestone level of education, 30.1 percent attained a different milestone.

[28] The results here are unlikely to be substantially biased by sheepskin effects. As mentioned in an earlier footnote, a larger jump in the return to schooling is only observed between those with 15 and 16 years of schooling, that is, when one attains a college degree. However, by construction, twin pairs where one sibling had 15 and the other had 16 years of schooling are not part of the sample of twins who reached key milestones in education.

between these characteristics and years of schooling are almost always statistically significant both within and between families (see Table A.6 in the Online Appendix). Nonetheless, the within-family correlations with years of schooling 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 compared with estimates from the pooled OLS model.

This exercise, however, is not ideal in our setting for three reasons. First, many correlates of ability, such as test scores, are not available in our historical data. Second, while we can construct some correlates, an important outcome like self-employment cannot be used because our sample comprises only wage and salary workers. Third, although marital status and spousal education are reported, we simply do not know how strongly they are associated with ability in 1940. In the case of marriage, over 75 percent of our linked twins were married by 1940 while the corresponding figure for the twins in Ashenfelter and Rouse (1998) is lower at 64 percent. Since relatively more individuals were married in 1940, marital status could be less useful for distinguishing between people of different abilities historically.

We propose two ways of assessing the threat of unobservable differences between twin brothers, both of which are unique to our historical setting and data. Both tests, to some degree, rely on assumptions about how parents treat twin sons who show different levels of ability as children—investing more or less in the "more able" twin to equalize or exacerbate ability differences. After detailing the tests, we return to these assumptions.

First, we gradually restrict the sample of twins to pairs with smaller differences in years of schooling and find little change in the estimated return to education. The smaller the gap in schooling, the more similar twins are likely to be in terms of unobserved ability. Consequently, the estimated rate of return may be less biased. Figure 4 plots the results from this exercise. Each marker represents the coefficient from a separate twin fixed effect regression, with 95 percent confidence intervals, and the maximum difference in schooling ranges from 1 to 17 years along the x-axis. Behind the point estimates, we plot the sample size corresponding to each regression. The return to schooling consistently hovers around 4 percent 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 for twin pairs with education differences of 4 years or less—they comprise 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 percent confidence interval is still slightly lower than the return to schooling in late 20th-century twins studies.

[Figure 4 about here.]

Second, we restrict our sample to pairs of twins who—even more than most twins—might be likely to receive identical treatment or resources from their parents. When we do so, we continue to find rates of return that are consistent with our baseline. Focusing on this subset of twins may help 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. Online Appendix A.4 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.[29] Names that are more similar may thus identify children who were raised with comparable resources and in similar environments. We use three measures of name similarity: first names that begin with the same letter, that have a Jaro-Winkler string distance of less than 0.2, or that belong to the same Soundex phonetic group.[30] Three subsets of twins are identified with these metrics, which we then use to re-estimate equations (1) and (2). Table 6 presents the resulting coefficients. Though some subsamples are quite small, the fixed effects estimates in the even columns are broadly similar to the baseline.

[Table 6 about here.]

While these two exercises strengthen our confidence in our twin-based findings, they require different but strong assumptions. When we compare twins with shrinking differences in actual education, we assume that the gap in education within a twin pair is informative of innate ability. The similar-name subsample analysis requires that names be informative of potential differences in parental investments. In both cases, the metrics—and our tests—may be distorted if parents respond endogenously to differences in ability observed between their children. Parents could respond by equalizing inputs across children, by investing

[29]Online Appendix A.4 also shows that twins with similar first names in 1940 are more likely to be in the same grade in school. Table A.7 in the Online Appendix documents the most common names among twin pairs. Though common names in the population are popular for twins, we also see a large number of twins whose names start with the same letter or rhyme.

[30]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 end 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.

disproportionately in the high-ability twin, or by investing more in the low-ability twin. In this same historical period, Parman (2015) found parents shifted educational investments to healthier siblings. If parents of twins respond similarly then the resulting differences in schooling, while possibly smaller, might not be indicative of smaller differences in unobserved ability. Parents who plan to treat twins identically—signalled with similar naming—could have their plans derailed by the reality of parenting. As we find in Online Appendix A.4, educational investments are more correlated for similarly named pairs. Generally, the weaker any endogenous parental responses, the cleaner these tests are and the more confident we can be in using the twins method to causally identify the return to education.

The 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.

Table 7 presents the results from this exercise. Column (1) reproduces our baseline result—an additional year of education increases weekly earnings by 0.044 log points. Columns (2) to (4) include increasingly narrow fixed effects for occupations, based on the three-digit occupation code (`occ1950`) from IPUMS.[31] 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. About two-thirds of the return to education thus appears to stem from higher earnings within an occupation. 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 occupations while the other half is between occupations.[32]

[Table 7 about here.]

[31] 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 one- and three-digit occupation categories are sensible, not all two-digit occupation codings are economically meaningful: bookbinders (502), for instance, are grouped with cabinetmakers (505). For the codes, see: `https://usa.ipums.org/usa-action/variables/occ1950`.

[32] In contrast to our finding that the return to education arises both within and across occupations, we find little evidence for large returns across industries. Table A.8 in the Online Appendix repeats the analysis of Table 7 but compares twin returns within one-, two-, and three-digit industry fixed effects, based on `ind1950` codes from IPUMS. At baseline, returns are 0.044 and fall only to 0.037 with three-digit industry fixed effects, suggesting only about 15 percent of variation in the return to education is attributable to industry.

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 SES. 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. This underscores that workers in 1940 benefited from education both via entry into better paying occupations and via higher pay within occupations.[33]

Heterogeneity in the Return to Education

With the twins identification strategy and a large sample, we can estimate the return to schooling for different subsets of the population in 1940. This section considers two potential dimensions of heterogeneity: family status and immigration history. Does education enhance intergenerational mobility or facilitate assimilation among immigrants? We find suggestive evidence for the former but not the latter: there are higher returns to education for those whose fathers were farmers or had lower SES, but twins with more foreign-born grandparents had lower returns.

The Return to Education by Family Economic Status

Schooling can weaken intergenerational economic persistence if the return to education is higher for children from lower SES families. We find higher returns 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 of twins by the father's occupation and SES and then estimates the return to education. Specifically, we split the sample into three groups: twin sons of farmers, twin sons with fathers in the bottom half of the occupation score distribution, and twin sons of fathers in the top half of the occupation score distribution. We separate out farmers for two reasons. First, farmers are the most common occupation category for fathers in our sample, which is not surprising given the declining but still-important place of agriculture in the US economy during the early 20th century. Second, we use the occupation scores of fathers to determine family SES. These scores may be especially weak indicators of SES for the large and highly varied category of farmers, which comprises positions ranging from sharecroppers and tenant farmers to those owning and

[33]Table A.9 in the Online Appendix verifies the robustness of this conclusion using alternative versions of occupation scores.

farming large acres of land. Feigenbaum (2018) documents a large variance in actual earnings among farmers in 1915 Iowa but these farmers would all be assigned the same occupation score.

[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 in Clay et al. (2016) that show a monotonic decline in returns by quantile. However, the point estimates in the even columns in Table 8 are not statistically different from each other. Our results may reflect the association between education level and the status of one's father—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 the median occupation score, respectively—though this relation is unlikely to be strong enough to explain the differences in Table 8.

The Return to Education by Family Immigration History

In the early 20th century, proponents of mass education believed that the American school system could facilitate the assimilation of immigrants (Lleras-Muney and Shertzer 2015; Arington 1991). Our sample contains very few foreign-born twins and twins with foreign-born parents. To determine if the return to schooling varies by immigration history, we thus focus on grandparents instead. Specifically, we divide white twins in our sample into three groups: those with no foreign-born grandparents, those with one to three foreign-born grandparents, and those whose grandparents are all foreign-born.[34]

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.[35] Put differently, the labor market offers smaller rewards for twins with more recent family immigrant histories. Clay et al. (2016), exploiting variation in CSLs by state and cohort, also find higher returns among children with two American-born parents, compared to children with one or more

[34]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 percent were born in Germany, 15 percent in Ireland, 6.9 percent in England, 5 percent 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 as children in the 1900, 1910, and 1920 censuses. In all three 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 percent of our sample meet this criterion.

[35]The return in column (2) is not statistically different from the returns in column (4) but is statistically different from that in column (6). We test this by stacking the data and estimating interaction coefficients.

foreign-born parents. While their results are somewhat imprecise, possibly because the first stage effect of compulsory schooling on the educational attainment of children with foreign-born parents is weak (Lleras-Muney and Shertzer 2015), they help reinforce our case that the return to schooling was lower for the children and grandchildren of the foreign-born.[36]

[Table 9 about here.]

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 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 their peers with American-born grandparents. 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 knowledge of 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.

Beyond differences in education returns by family status and immigration history, Online Appendix A.5 also explores if the return to schooling differs across cohorts and regions.[37] We find slightly larger returns to education for older cohorts and show that the value of education is fairly similar across regions.

The Return to Education on Other Outcomes

While the primary goal of this paper is to determine the weekly earnings return to education, additional years of schooling may also affect other outcomes. In this section, we leverage the same twins sample and

[36]There are also differences in average educational attainment between twin sons, depending on their family immigration histories, 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 one to three American-born grandparents. However, twins with four foreign-born grandparents had only 9.6 years of schooling in 1940 on average.

[37]Since our sample almost exclusively comprises American-born whites (Table 3), we do not explore heterogeneity in education returns by race.

approach to estimate the causal effect of education on a host of outcomes in 1940, including labor supply, migration, and family structure.

We begin by studying the effects of educational attainment on labor supply in 1940.[38] To do so, we expand our sample to include all linked twins. Table 10 presents our results. Panel A focuses on the extensive margin of labor supply by using several measures of employment status as the outcome. We find that twins with more schooling are slightly more likely to be employed (excluding relief work), but are less likely to be self-employed or to be employed in relief work. Panel B then studies other dimensions of labor supply that encompass both the extensive and intensive margins. Twins with more education were more likely to work full-time (defined as working more than 40 weeks in 1939), to work more weeks in a year, and to work more hours per week.

[Table 10 about here.]

Education may have also influenced geographic mobility in the early 20th century, as we show across a variety of measures in Table 11.[39] As before, we focus on the estimates with twin family fixed effects in the even columns. Those with more schooling have a higher propensity of leaving the counties they grew up in (Panel A, column (2)), and these moves tend to be across states (Panel A, column (6)). The association between education and geographic mobility during this period has been documented elsewhere in the literature.[40] Margo (1990), for example, argues that better-educated African Americans were more likely to move because schooling lowered the cost of migration through increasing one's knowledge of distant opportunities or by enhancing a person's ability to assimilate. The positive effect on cross-state migration is also consistent with Malamud and Wozniak (2012), who exploit variation in college attainment due to draft avoidance during the Vietnam War to show that more years of college increase the odds of living outside one's birth state. In addition to our migration results, we find that as adults, the twin with more schooling is more likely to reside in urban and more populous areas rather than on farms (Panel B).

[38] Table A.10 in the Online Appendix considers three other dimensions of economic status: whether a person has earnings top-coded (earned \$5,000 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. We find positive effects of schooling on each outcome but these are small in magnitude. For more on the interaction of education and labor supply (and gender) in the mid-20th century labor market, see Acemoglu et al. (2004).

[39] While Table 11 limits the analysis to our main sample of twins, Table A.11 in the Online Appendix shows that we obtain similar results with the complete sample of linked twins.

[40] 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.

[Table 11 about here.]

Finally, to examine differences in family structure, Table 12 implements regressions (1) and (2), replacing the original wage outcome with three measures of family status in 1940: 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, though people with more schooling are less likely to have children and have fewer children overall.[41] The point estimates with the fertility outcomes, however, are small compared with the outcome means. In our sample, 53 percent 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 to 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 12 about here.]

As documented in this section, the return to education in 1940 did not just accrue via more earnings. Twin siblings with more education were also likely to work more, to migrate (to cities and more populous places), and to have smaller families, though the effects on these alternative outcomes are often relatively small in size.

Conclusion

What was the return to schooling in mid-century America? Goldin and Katz (2009) suggest that the returns were lower during this period in history compared with the earlier and later parts of the 20th century. However, their conclusion is based on Mincerian correlations, leaving open the possibility that their results may be biased up or down by selection into education. Our paper provides causal estimates of the return to schooling in 1940 and confirms that the returns were indeed lower relative to today. Given that income

[41] Table A.12 in the Online Appendix 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.

inequality was relatively lower in the middle of the 20th century, the smaller returns we find are also consistent with the rise in income inequality from the 1980s onwards, much of which has been attributed to the growing skill premium (Goldin and Katz 2007).

Constructing a large linked sample of twins and exploiting within-twins variation in earnings and education, we estimate a return of around 4 percent in 1940. While positive, this is substantially smaller than the returns for more recent cohorts. We provide suggestive evidence that our results are not severely biased by imperfections in record linkage, measurement error in the years of schooling, or possible differences in ability within twin pairs.

Taking advantage of the large size of our dataset, we are also able to document heterogeneous effects of education for different subsets of the population. Those with more foreign-born grandparents, for instance, are observed to have a lower return to schooling.

Future research may proceed in several directions. First, quantifying the causal effect of education even further back in time could provide more evidence of the decline in the skill premium Goldin and Katz (2009) trace before 1940. Second, estimating the historical returns in other countries can help shed light on whether the economic value of education in the US during the mid-20th century was the exception or the norm. Third, within the US labor market, our findings relate only to the return to education among men due to our method of linking individuals across censuses. The return to schooling for women remains to be studied, complementing work on the interaction of male and female labor supply at mid-century (Acemoglu et al. 2004). Though the lower rates of female labor force participation in the past complicates estimates of the return to education among women, the impact of schooling on earnings, occupation choice, labor supply, and other outcomes could offer important datapoints for scholars seeking to understand the labor market for women in this era.

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.

Acemoglu, Daron, David H. Autor, and David Lyle (2004). Women, War, and Wages: The Effect of Female Labor Supply on the Wage Structure at Midcentury. *Journal of Political Economy 112*(3), 497–551. 00404.

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.

Arington, Michele (1991). English-Only Laws and Direct Legislation: The Battle in the States over Language Minority Rights. *Journal of Law and Politics 7*, 913–950.

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 (2019). 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.

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 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.

Dahl, Gordon B (2002). Mobility and the return to education: Testing a roy model with multiple markets. *Econometrica 70*(6), 2367–2420.

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.

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 (2007). 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.

Griliches, Zvi (1979). Sibling Models and Data in Economics: Beginning of a Survey. *Journal of Political Economy 87*(5), S37–S64.

Hungerford, Thomas and Gary Solon (1987). Sheepskin effects in the returns to education. *Review of Economics and Statistics 69*(1), 175–177.

Isacsson, Gunnar (2004). Estimating the economic return to educational levels using data on twins. *Journal of Applied Econometrics 19*(1), 99–119.

Jaeger, David A. and Marianne E. Page (1996). Degrees matter: New evidence on sheepskin effects in the return to education. *Review of Economics and Statistics 78*(4), 733–740.

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.

Kopczuk, Wojciech, Emmanuel Saez, and Jae Song (2010). Earnings Inequality and Mobility in the United States: Evidence from Social Security Data Since 1937. *The Quarterly Journal of Economics 125*(1), 91–128.

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.

Malamud, Ofer and Abigail Wozniak (2012). The Impact of College on Migration: Evidence from the Vietnam Generation. *Journal of Human Resources 47*(4), 913–950.

Margo, Robert A (1986). Race, educational attainment, and the 1940 census. *The Journal of Economic History 46*(1), 189–198.

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 (2015). Childhood Health and Human Capital: New Evidence from Genetic Brothers in Arms. *The Journal of Economic History 75*(1), 30–64.

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, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek (2020). *IPUMS USA: Version 10.0 [dataset]*. 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.

Saez, Emmanuel (2013). Striking it Richer: The Evolution of Top Incomes in the United States (Updated with 2012 preliminary estimates).

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.

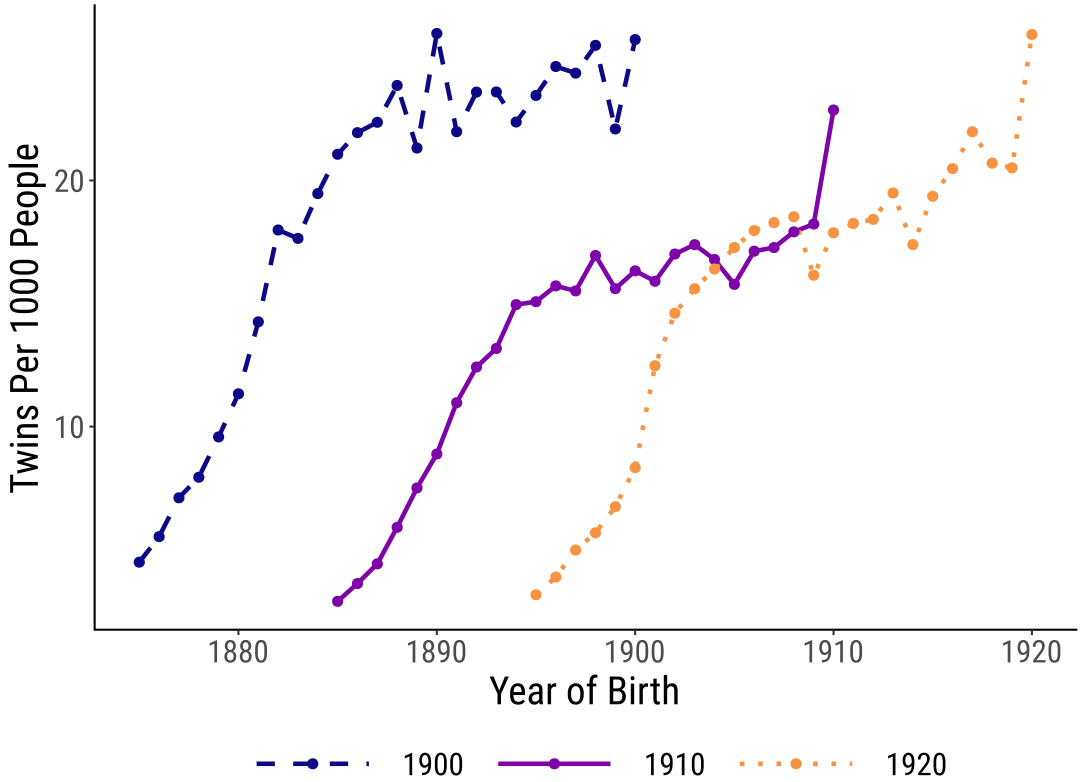

Figure 1: Twins Per 1,000 People in the Complete Count Censuses, 1900-1920. We construct estimates of the number of twins in the US in each decennial census from 1900 to 1920. We identify twins as two people aged 25 or younger living in the same household with the same last name, age in years, birthplace, and relationship to the household head. To calculate the number of twins per 1,000 people, we count up the number of people aged 25 or younger as the denominator.

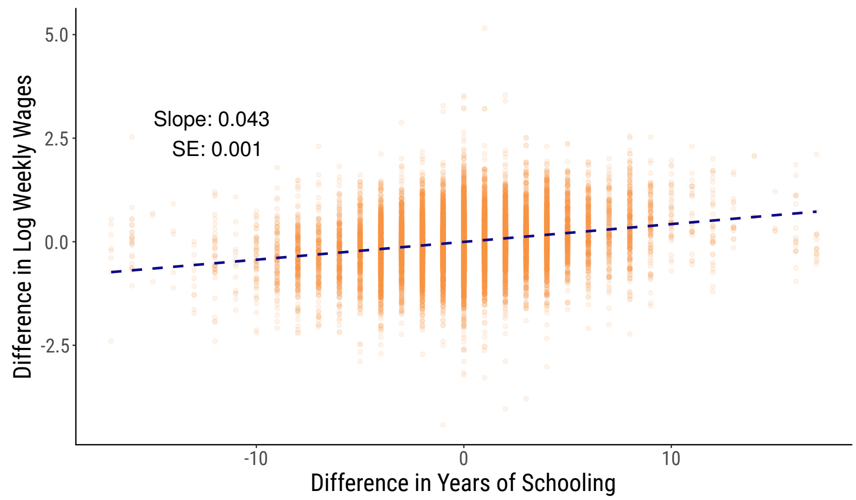

Figure 2: The Return to Schooling in the Early 20th Century. Data are from the pooled 1900-1940, 1910-1940, and 1920-1940 linked twins samples. The sample is restricted to wage and salary male workers with a weekly wage of at least $6, who worked a positive number of weeks in the previous year (1939), and who worked a positive number of hours in the preceding week, according to the 1940 census. The twins sample includes only twins who *both* fit our sample criteria. The best linear fit is shown. Robust standard errors are displayed.

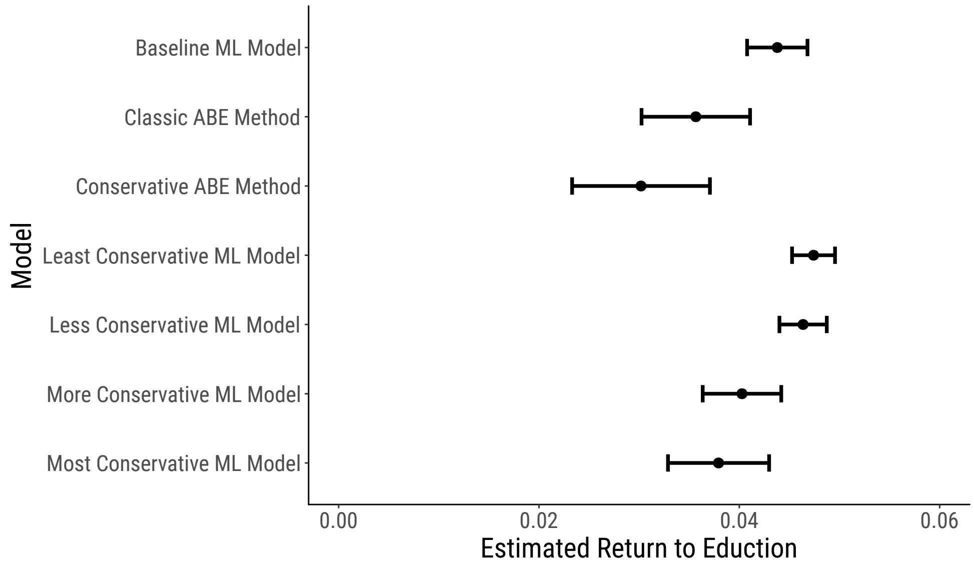

Figure 3: Robustness of the Estimated Return to Schooling in the Early 20th Century Across Different Linking Methods. Data are from the pooled 1900-1940, 1910-1940, and 1920-1940 linked twins samples. We plot the return to education with 95% confidence intervals, based on robust standard errors clustered at the twin-pair level. The sample is linked with two methods: a machine learning approach (Feigenbaum 2016) and an algorithmic approach (ABE after Abramitzky et al. (2012), but 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 the classic method and a more conservative version 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 (1939), and who worked a positive number of hours in the preceding week, according to the 1940 census. The twins sample includes only twins who *both* fit our sample criteria. Following Bailey et al. (2019), we use inverse propensity weights to adjust for observable differences between matched and unmatched people in our census linked sample.

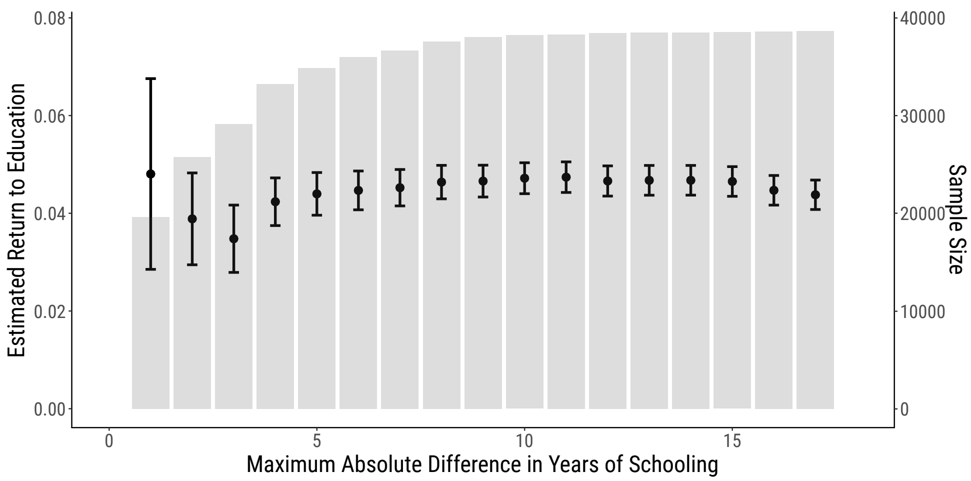

Figure 4: The Return to Schooling is Still Positive Even as the Sample is Narrowed to Twin Pairs With More Similar Years of Schooling, Fixed Effects Results. Data are from the pooled 1900-1940, 1910-1940, and 1920-1940 linked twins datasets. Each marker represents the return to schooling from separate regressions of the log of weekly earnings in 1939 on years of education, with twin family fixed effects. 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 (1939), and who worked a positive number of hours in the preceding week. The sample size, given by the bars, 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 clustered at the twin-pair level. Following Bailey et al. (2019), we use inverse propensity weights to adjust for observable differences between matched and unmatched people in our census linked sample.

Table 1: Twins in the US Censuses, 1900-1920

Census Year	Twins per 1,000				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

Note: We construct estimates of the number of twins in the US in each decennial census from 1900 to 1920. We identify twins as two people aged 25 or younger living in the same household with the same last name, age in years, birthplace, and relationship to the household head. To calculate the number of twins per 1,000 people, we count up the number of people aged 25 or younger as the denominator.

Table 2: Twins and Their Families Are Similar to the American Population, 1900-1920

	Twins in 1900-1920				1% Samples
	All	Linked	Weighted	Analysis Sample	1900, 1910, and 1920
Age	9.6	9.6	9.6	8.9	11.0
	(6.7)	(6.6)	(6.7)	(6.3)	(7.1)
White	0.86	0.93	0.92	0.95	0.88
	(0.35)	(0.25)	(0.27)	(0.22)	(0.32)
Number of Siblings	4.2	4.1	4.2	4.0	2.8
	(2.2)	(2.2)	(2.2)	(2.1)	(2.3)
Father's Occupation Score	17.22	17.70	17.58	18.73	17.92
	(11.36)	(11.92)	(11.71)	(12.66)	(12.00)
Urban	0.35	0.36	0.35	0.48	0.41
	(0.48)	(0.48)	(0.48)	(0.50)	(0.49)
Farm	0.44	0.44	0.45	0.31	0.37
	(0.50)	(0.50)	(0.50)	(0.46)	(0.48)
Foreign-Born	0.02	0.02	0.02	0.02	0.05
	(0.16)	(0.12)	(0.13)	(0.13)	(0.22)
Foreign-Born Parents (#)	0.47	0.47	0.47	0.54	0.51
	(0.80)	(0.78)	(0.79)	(0.82)	(0.82)
Foreign-Born Grandparents (#)	1.46	1.55	1.55	1.74	1.54
	(1.78)	(1.78)	(1.79)	(1.80)	(1.79)
Farmer Father	0.43	0.43	0.44	0.29	0.37
	(0.49)	(0.49)	(0.50)	(0.45)	(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 who worked 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 aged 25 or younger who were residing in their parents' households in 1900, 1910, and 1920, based on three 1% random samples. Standard deviations are in parentheses.

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

	Linked Twins		
	Unweighted	Weighted	1940 1% Sample
Years of Schooling	9.63	9.52	9.18
	(3.20)	(3.21)	(3.46)
Weekly Earnings	30.98	30.87	29.16
	(27.13)	(27.54)	(26.72)
Log of Weekly Earnings	3.27	3.27	3.20
	(0.55)	(0.55)	(0.58)
Works Full-time	0.82	0.81	0.78
	(0.39)	(0.39)	(0.42)
Years of Experience	18.5	18.9	19.5
	(9.7)	(9.9)	(12.5)
Age	36.6	37.0	37.5
	(9.4)	(9.6)	(12.2)
White	0.96	0.95	0.93
	(0.20)	(0.22)	(0.26)
Foreign-Born	0.02	0.02	0.13
	(0.13)	(0.13)	(0.34)
Married	0.75	0.76	0.73
	(0.43)	(0.43)	(0.44)
Number of Children	1.2	1.2	1.1
	(1.5)	(1.5)	(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. 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 who worked a positive number of hours in the previous week. The twins sample includes only twins who *both* fit our sample criteria. In column 3, we use a random 1% sample of the 1940 census. For consistency, the 1940 1% sample is limited to employed men aged 17-68 who are wage and salary workers and we impose the same sample restrictions on wages, weeks worked, and hours worked. Standard deviations are in parentheses.

Table 4: The Return to Education: Baseline

	1940 1% Sample			Twins, Pooled			Twins, Family FE	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Years of Education	0.046***	0.055***	0.038***	0.050***	0.056***	0.036***	0.044***	0.035***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
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 earnings and education 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 who worked 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. (2019), we use inverse propensity weights to adjust for observable differences between matched and unmatched people in our census linked sample. Robust standard errors (clustered at the twin-pair level in columns 4 to 8) are in parentheses.

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

Table 5: The Return to Education: Restricted to Sample with Milestone Education Outcomes

	1940 1% Sample		Twins, Pooled		Twins, Family FE
	(1)	(2)	(3)	(4)	(5)
Years of Education	0.044***	0.061***	0.047***	0.062***	0.056***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)
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 but restrict to men with milestone years of education. 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 earnings and education 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+. Following Bailey et al. (2019), we use inverse propensity weights to adjust for observable differences between matched and unmatched people in our census linked sample. Robust standard errors (clustered at the twin-pair level in columns 3 to 5) are in parentheses.

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

Table 6: The Return to Education: Twin Pairs with Similar First Names

	Same First Letter		Jaro-Winkler Distance ≤ 0.2		Same Soundex	
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	0.053^{***}	0.041^{***}	0.063^{***}	0.045^{***}	0.057^{***}	0.053^{***}
	(0.002)	(0.003)	(0.005)	(0.007)	(0.009)	(0.018)
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 R2	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 first 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 first names have the same phonetic score (using Soundex). In the even columns, we include twin family fixed effects, forcing the comparisons of earnings and education 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 who worked 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. (2019), we use inverse propensity weights to adjust for observable differences between matched and unmatched people in our census linked sample. Robust standard errors clustered at the twin-pair level are in parentheses.

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

Table 7: The Return to Education: Within or Across Occupations?

	Baseline	Occupation Code FEs			Log Occscore
	(1)	(2)	(3)	(4)	(5)
Years of Education	0.044***	0.030***	0.029***	0.028***	0.019***
	(0.002)	(0.001)	(0.001)	(0.001)	(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 scores (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 a substantial return to education within occupations. In all columns, we include twin family fixed effects, forcing the comparisons of earnings and education 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 who worked 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. (2019), we use inverse propensity weights to adjust for observable differences between matched and unmatched people in our census linked sample. Robust standard errors clustered at the twin-pair level are in parentheses.

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

Table 8: The Return 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***	0.049***	0.043***	0.041***	0.045***	0.044***	0.040***	0.039***
	(0.002)	(0.003)	(0.001)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
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 fathers were farmers. In the remaining six columns, we focus on twins whose fathers were not farmers, splitting these twins by their fathers' occupation scores at the median in columns 5 and 6 versus 7 and 8. In the even columns, we include twin family fixed effects, forcing the comparisons of earnings and education 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 who worked 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. (2019), we use inverse propensity weights to adjust for observable differences between matched and unmatched people in our census linked sample. Robust standard errors clustered at the twin-pair level are in parentheses.

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

Table 9: The Return 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***	0.050***	0.044***	0.043***	0.038***	0.036***
	(0.002)	(0.002)	(0.002)	(0.004)	(0.002)	(0.003)
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 twins. Because we observe twins in 1900, 1910, and 1920 in their childhood homes, we can use the census questions about the birthplaces of their parents' parents to determine where their grandparents were born. The sample includes only white twins to avoid conflating race with grandparent and parent immigration histories, as nearly all African-American twins in the linked sample had four American-born grandparents. We also drop twin pairs from our baseline sample if both their mothers and fathers did not reside in their childhood household in the census. In the even columns, we include twin family fixed effects, forcing the comparisons of earnings and education 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 who worked 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. (2019), we use inverse propensity weights to adjust for observable differences between matched and unmatched people in our census linked sample. Robust standard errors clustered at the twin-pair level are in parentheses.

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

Table 10: Effect of Education on Alternative Economic Outcomes

	Panel A. Employment					
	Employed ×100		Self-employed ×100		Employed in Relief Work ×100	
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	1.115***	1.273***	-0.826***	-0.313***	-0.491***	-0.361***
	(0.031)	(0.051)	(0.035)	(0.055)	(0.016)	(0.027)
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	0.24	0.01	0.08
Y Mean	81.84	81.84	24.21	24.21	4.52	4.52

	Panel B. Labor Supply					
	Works Full-time ×100		Weeks of Work		Hours of Work	
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	1.453***	1.518***	0.584***	0.664***	0.763***	0.829***
	(0.036)	(0.060)	(0.015)	(0.024)	(0.018)	(0.030)
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	69.38	69.38	40.04	40.04	35.62	35.62

Note: All columns present regressions of employment (Panel A) or labor supply (Panel B) 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 the even columns, we include twin family fixed effects, forcing the comparisons of outcomes and education 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 further restrictions. Following Bailey et al. (2019), we use inverse propensity weights to adjust for observable differences between matched and unmatched people in our census linked sample. Robust standard errors clustered at the twin-pair level are in parentheses. We scale indicator dependent variables by 100 to simplify the interpretation of effects and standard errors.

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

Table 11: Effect of Education on Migration

	Panel A. Migration					
	Moved Out of County ×100		Moved Within State ×100		Moved Out of State ×100	
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	0.617***	1.111***	-0.845***	-0.430***	1.462***	1.541***
	(0.088)	(0.115)	(0.086)	(0.130)	(0.085)	(0.129)
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	60.47	60.47	32.00	32.00	28.47	28.47
	Panel B. 1940 Location Choice					
	Urban ×100		Log Size of Place		Farm ×100	
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	2.063***	1.516***	0.120***	0.095***	-1.013***	-0.882***
	(0.084)	(0.131)	(0.005)	(0.008)	(0.051)	(0.082)
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	69.14	69.14	9.99	9.99	8.11	8.11

Note: All columns present regressions of migration (Panel A) or residential (Panel B) 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 the even columns, we include twin family fixed effects, forcing the comparisons of outcomes and education 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 who worked 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. (2019), we use inverse propensity weights to adjust for observable differences between matched and unmatched people in our census linked sample. Robust standard errors clustered at the twin-pair level are in parentheses. We scale indicator dependent variables by 100 to simplify the interpretation of effects and standard errors. Table A.11 in the Online Appendix replicates this analysis using the complete sample of linked twins and shows very similar results.

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

Table 12: Effect of Education on Marriage and Fertility

	Married ×100		Any Children ×100		Number of Children	
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	-0.745***	0.107	-1.576***	-0.811***	-0.085***	-0.052***
	(0.074)	(0.110)	(0.087)	(0.144)	(0.003)	(0.005)
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	75.44	75.44	53.29	53.29	1.17	1.17

Note: All columns present regressions of 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 the even columns, we include twin family fixed effects, forcing the comparisons of outcomes and education 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 who worked 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. (2019), we use inverse propensity weights to adjust for observable differences between matched and unmatched people in our census linked sample. Robust standard errors clustered at the twin-pair level are in parentheses. We scale indicator dependent variables by 100 to simplify the interpretation of effects and standard errors. Table A.12 in the Online Appendix replicates this analysis using the complete sample of linked twins and finds very similar results for columns 3 to 6. The marriage result in the full sample, however, does show a positive effect of education on the probability of being married.

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

A Online Appendix

A.1 Using Birth Month in 1900 to Identify Twins

In the main paper, twins were identified based on their age 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 and year of birth.[1] Using these fields, we can examine whether or not our somewhat imperfect method of identifying twins is affecting our results. We find that the return to education estimated with the set of twins we are most confident about in 1900—those with the same age *and* birth month—is even lower than the overall return, suggesting that our conclusions are unlikely to be driven by 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 percent 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), who 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 percent 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, the overall direction of the bias, if any, is unclear (Tan 2019). The effects of such misclassification on our results 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 twins 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 find in the return to education across cohorts (see Figure A.2). The twins in 1900 were born between 1875 and 1900, with most births occurring between 1890 and 1900.

[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

[1] The 1910 census only records the month of birth for people residing in Alaska.

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 schooling. Thus, our main conclusion—that the return to education in 1940 was positive but smaller than the returns for more recent cohorts—is unlikely to be an artifact of errors in twins tagging.

A.2 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. (2019) recommend using inverse propensity weights to adjust for observable differences between matched and unmatched people. They construct these weights in two steps:[2]

- 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 for each person are then computed as $\frac{1-p}{p}\frac{m}{1-m}$, where p is the predicted likelihood of an individual 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 reweighting procedure in Bailey et al. (2019), 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 the analysis in Bailey et al. (2019) 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, one of which is the day of birth.
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 whether the parent resides in the household.

[2] See footnote 33 of Bailey et al. (2019).

4. We use a quadratic (second-order polynomial) in both age and year of birth because we are linking children from multiple censuses.

5. To measure how common first and last names are, we use the log of the number of people in the 1900, 1910, and 1920 censuses with a given first or last name.

A.3 Replication of Ashenfelter and Rouse (1998) by Sex

Ashenfelter and Rouse (1998) estimate the return to education in the US during the late 20th century, exploiting variation in years of schooling between pairs of identical twins. In the paper, we compared our estimated return to education in 1940 with the returns in Ashenfelter and Rouse (1998) and other recent twins studies. However, our census linking strategy can only link male twins across censuses as we need to match individuals on names, among other criteria. In contrast, contemporary studies typically pool both genders together. For comparability, we replicate the results in Ashenfelter and Rouse (1998) for the male-only portion of their sample.[3]

First, we reproduce the results from Ashenfelter and Rouse (1998) in the first three columns of Table A.2. The first column replicates the result in Table II, column (2), of their paper—an OLS regression of log earnings on education with their sample of male and female twins. Included as controls are a quadratic in age, an indicator for females, and a dummy for whites. The second and third columns replicate the results in Table III, columns (4) and (9), first differencing between identical twin pairs, without or with differenced controls, respectively. The controls here are whether twins are covered by unions, married, and their job tenure, all differenced between twins. As we discuss in the body of the paper, these controls may be endogenous or "bad" (Angrist and Pischke 2009) as they could also be affected by schooling.

[Table A.2 about here.]

Second, we show that the results do not change when we include a constant term. Ashenfelter and Rouse (1998) suppress the constant term in their first difference regressions, which we followed in columns (2) and (3). However, including the constants in columns (4) and (5) does not change the results.

Finally, and most importantly for our analysis, we see that the late 20th century return to schooling for men is higher than what we find in 1940. The sample in Ashenfelter and Rouse (1998) is 59 percent female (see their Table I) and because they focus on identical twins all pairs are either male-male or female-female. When we restrict their sample to the 274 identical male twins, we observe reasonably similar results. The OLS return to education (column (6)) is 0.102 compared with the OLS return of 0.110 in the full sample (column (1)). More importantly, the causal returns using twin differences are either similar

[3] We use the replication data for Ashenfelter and Rouse (1998) available at `http://arks.princeton.edu/ark:/88435/dsp01xg94hp567`. We focus on Ashenfelter and Rouse (1998) rather than Ashenfelter and Krueger (1994) because the former is a larger sample that contains the original data in the earlier Ashenfelter and Krueger (1994) study.

(compare columns (4) and (7)) or even larger when estimated only with differences between male twin pairs (compare columns (5) and (8)).

Overall, this replication exercise suggests that our main finding—that the return to education, when estimated using twins, was lower in 1940 than it is in the recent period—is not driven by the sex composition of our sample.

A.4 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 within-twins estimator—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 the enrollment and highest grade completed of twin brothers in the 1940 census. In all cases, we limit the sample to boy-boy twin pairs aged 7 to 17 to target children of school-going age.[4]

Table A.3 shows that while twins in general are very likely to have the same educational outcomes during childhood—school enrollment or grade completed—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 and statistically significant at the 5 percent level or better in all comparisons.

[Table A.3 about here.]

[4]That all pairs of twins aged 1 or 2 are both not in school is not particularly informative about the nurture process.

A.5 Other Dimensions of Heterogeneity

This appendix explores if the return to schooling varied across cohorts and regions. We find some evidence of higher returns for older cohorts and comparable returns across different regions.

The Return to Education by Cohort

We first explore the return to education by cohort. Our sample collects twin pairs in the 1900, 1910, and 1920 censuses and comprises twins born across five decades. Figure A.2 groups cohorts into 5-year bins and repeats our baseline analysis.[5]

[Figure A.2 about here.]

The point estimates in Figure A.2 suggest that the return to schooling was higher for older cohorts, though the relative imprecision of the point estimates makes it difficult to determine if these differences are real or just noise. Simply fitting a line through the point estimates yields a gradient of -0.004 (S.E.=0.001) per 5-year binned cohort.

Differences in the return to schooling by cohort could be interpreted in two ways. First, the downward trend in the return to schooling may reflect the relatively smaller supply of skilled workers among earlier cohorts. Educational attainment rose steadily in the early 20th century, with new cohorts of American workers having about 0.8 more years of schooling per decade (Goldin and Katz 2009). Skilled workers were thus more scarce among the older cohorts when they entered the labor market. Second, the trend in Figure A.2 could reflect a lower return 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.

The Return to Education by Region

While Goldin and Katz (2009) describe the 20th century as the "human capital century", investments in education and the use of human capital in production were not uniform across the country. Could there also be spatial variation in the return to schooling? Since stratifying the sample on location in 1940 would be

[5] We exclude cohorts born 1871 to 1875 as there are only 38 pairs of twins in this interval. Figure A.7 provides the sample size for each cohort.

post-treatment, we examine the rate of return between twin pairs raised in different parts of the country.[6]

Our estimates suggest little geographic 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 of childhood residence, as shown in Figure A.3.[7] Though the point estimates are higher for twins from the South, these differences are not statistically significant.

[Figure A.3 about here.]

This apparent geographic similarity cannot be due to the differential shares of non-white twins across regions, as Figure A.3 plots the regional returns for both 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 (2007), for example, document large differences in the rate at which the high school movement progressed in each region during the first half of the 20th century.[8] Given the vast differences in labor market structures across the country and the huge variation in the supply of educated workers across states and regions (Goldin and Katz 2009, p. 204), it also does not seem reasonable to attribute the regional consistency to equal rewards to human capital regardless of locality.[9] We leave an investigation of the spatial homogeneity in returns to future research with better-suited empirical strategies.

[6]Dahl (2002) develops a correction procedure to estimate the return to education by location even in the presence of endogenous migration. However, adapting this selection correction to our twins fixed effects specification is beyond the scope of our paper.

[7]Figure A.8 shows that the spatial similarity in returns is also observed when twins are assigned to their region of residence in 1940.

[8]The variation in school quality changed substantially during this period. Goldin and Margo (1992) point to geographic narrowing in school quality as one factor behind the Great Compression.

[9]Within our sample of twins, there are sizable differences in the supply of educated workers 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.

A.6 Additional Figures and Tables

[Figure A.4 about here.]

[Figure A.5 about here.]

[Figure A.6 about here.]

[Figure A.7 about here.]

[Figure A.8 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.]

[Table A.9 about here.]

[Table A.10 about here.]

[Table A.11 about here.]

[Table A.12 about here.]

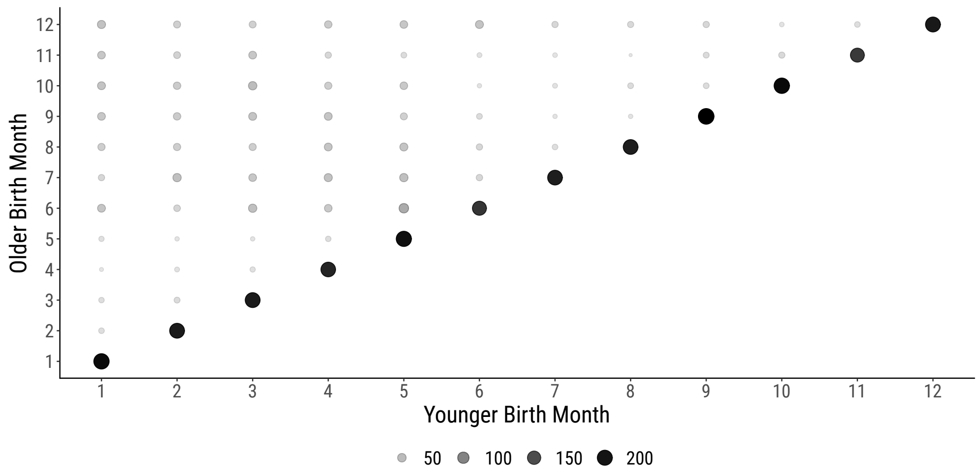

Figure A.1: Birth Months of Twins in 1900 Tagged Using Age. Because birth month is not recorded in 1910 or 1920, we tag twins as any pair of boys living in the same household and family, and who have the same age, birthplace, relationship to the head of household, and last name. In 1900, we can test the accuracy of our procedure. This graph suggests that there may be some twins in our data who are not actually twins, though the birth month variable in the census could itself be recorded with error or noise. Both the size and shading of the circles indicate the number of observations at each point.

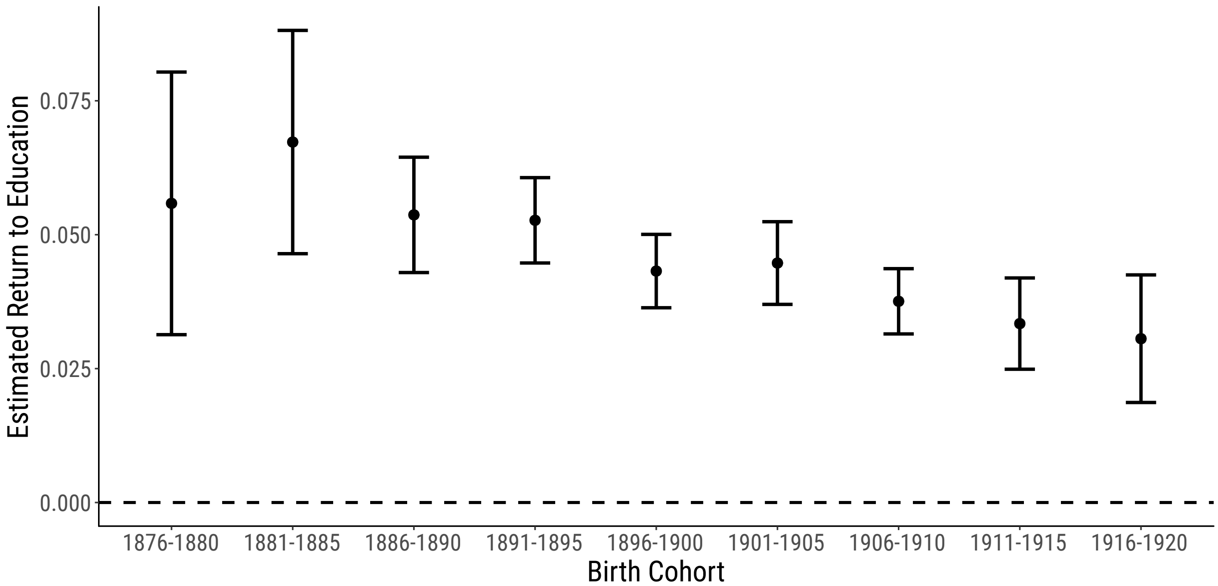

Figure A.2: The Return to Schooling by Cohort. For each 5-year bin of cohorts in our linked twins data, we estimate the return to education, replicating our main specification with twin family fixed effects. Each point represents a separate regression and is plotted with 95% confidence intervals, based on robust standard errors clustered at the twin-pair level. There is some evidence of a downward trend in the returns across cohorts.

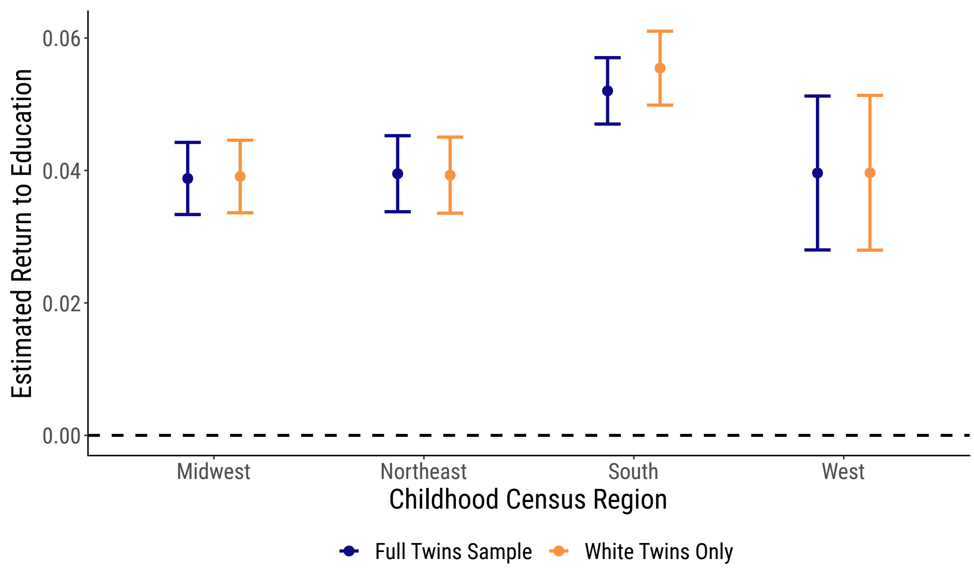

Figure A.3: The Return to Schooling in the Early 20th Century Did Not Vary Across Childhood Census Regions. We split our sample into four census regions based on where the twins were living when we observe them as children—in the household with their parents and their twin. Data are from the pooled 1900-1940, 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 (1939), and who worked a positive number of hours in the preceding week, according to the 1940 census. We estimate the return to schooling by region for the full sample and the subset of white twins. Following Bailey et al. (2019), we use inverse propensity weights to adjust for observable differences between matched and unmatched people in our census linked sample. 95% confidence intervals are shown, based on robust standard errors clustered at the twin-pair level.

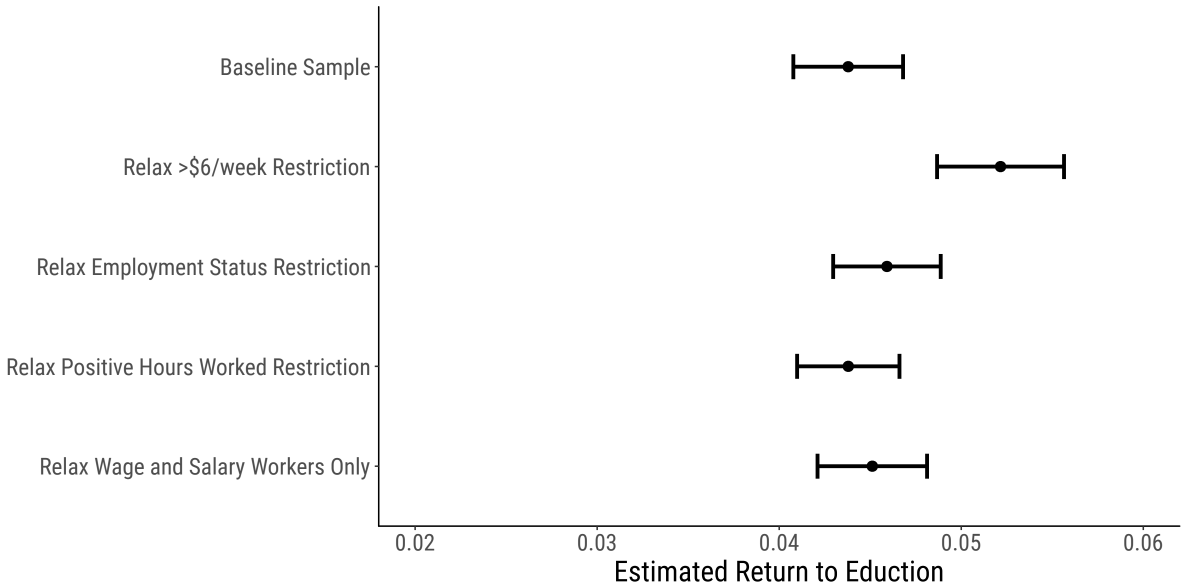

Figure A.4: Robustness of the Estimated Return to Education by Sample Restrictions. Our baseline sample is restricted to wage and salary male workers with a weekly wage of at least $6, who worked a positive number of weeks in the previous year (1939), and who worked a positive number of hours in the previous week. The twins sample includes only twins who *both* fit our sample criteria. This figure presents the estimated returns as we relax these restrictions one by one. We cannot relax the restriction for positive weeks worked in 1939 because our outcome is the log of weekly earnings, which is undefined when weeks worked is missing or zero. Following Bailey et al. (2019), we use inverse propensity weights to adjust for observable differences between matched and unmatched people in our census linked sample. 95% confidence intervals are shown, based on robust standard errors.

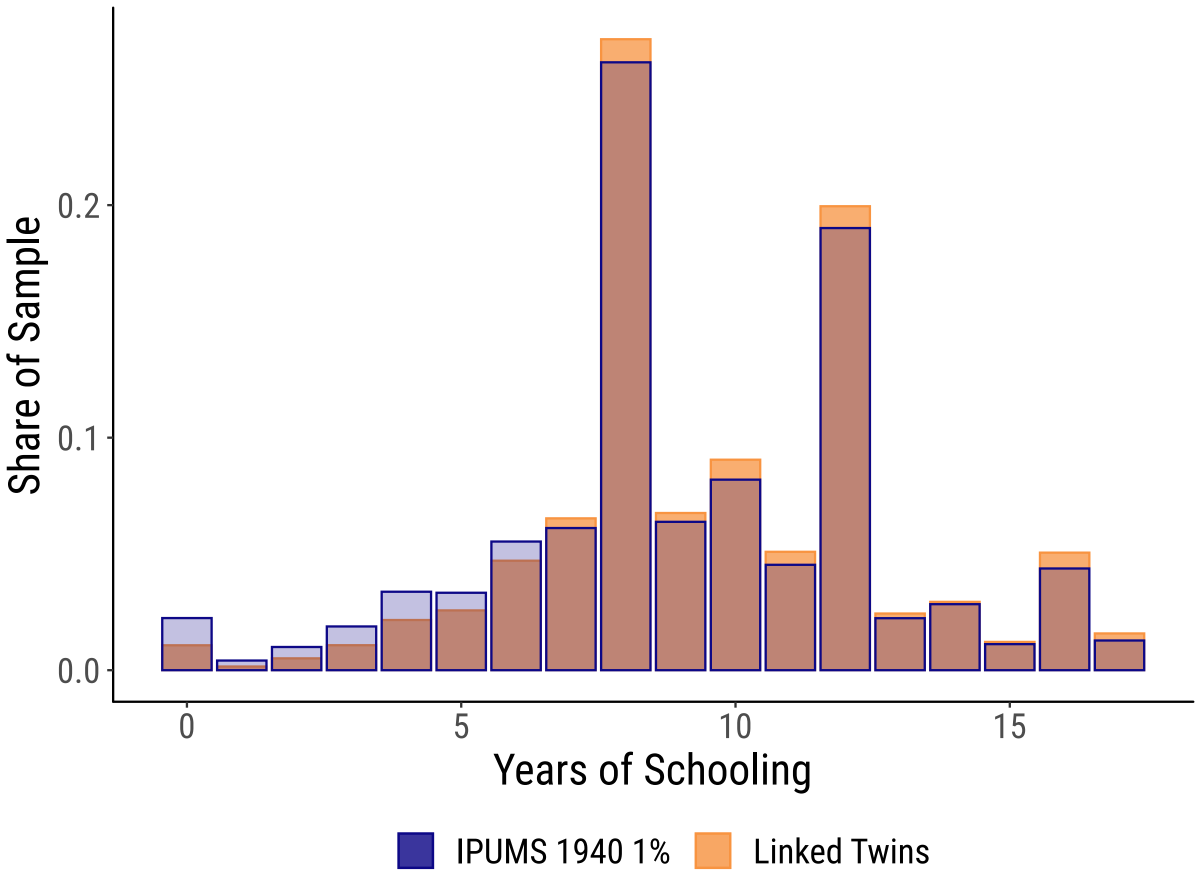

Figure A.5: Distribution of Years of Education in the Linked Twins Sample and the 1940 Census. Our sample of linked twins are slightly more likely to be common school (8 years), high school (12), or college (16) graduates compared with a random sample of cohort-mates in 1940, and less likely to report no schooling or 6 or less years of schooling. This differences are small and are likely to be driven by the smaller shares of foreign-born and African Americans in the linked sample (see Table 3).

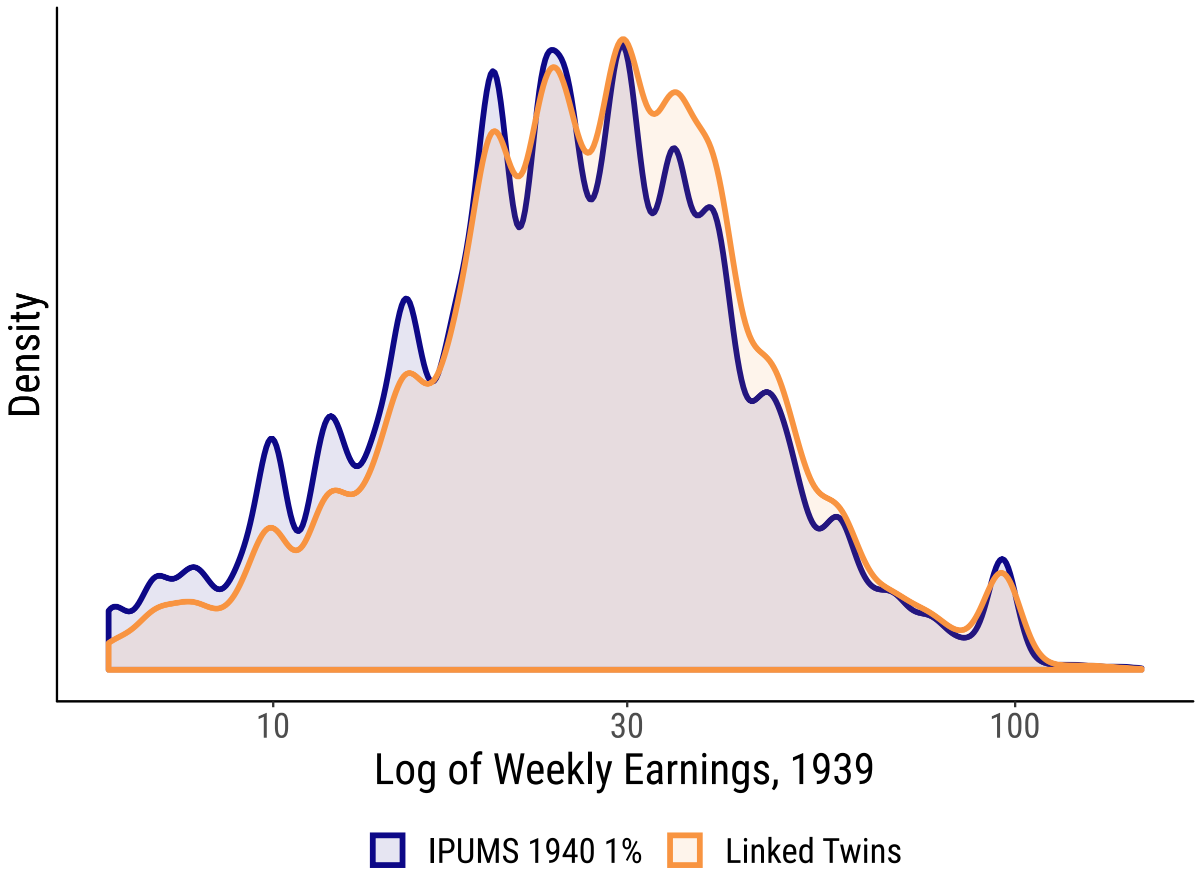

Figure A.6: Distribution of Log Weekly Earnings in the Linked Twins Sample and the 1940 Census.

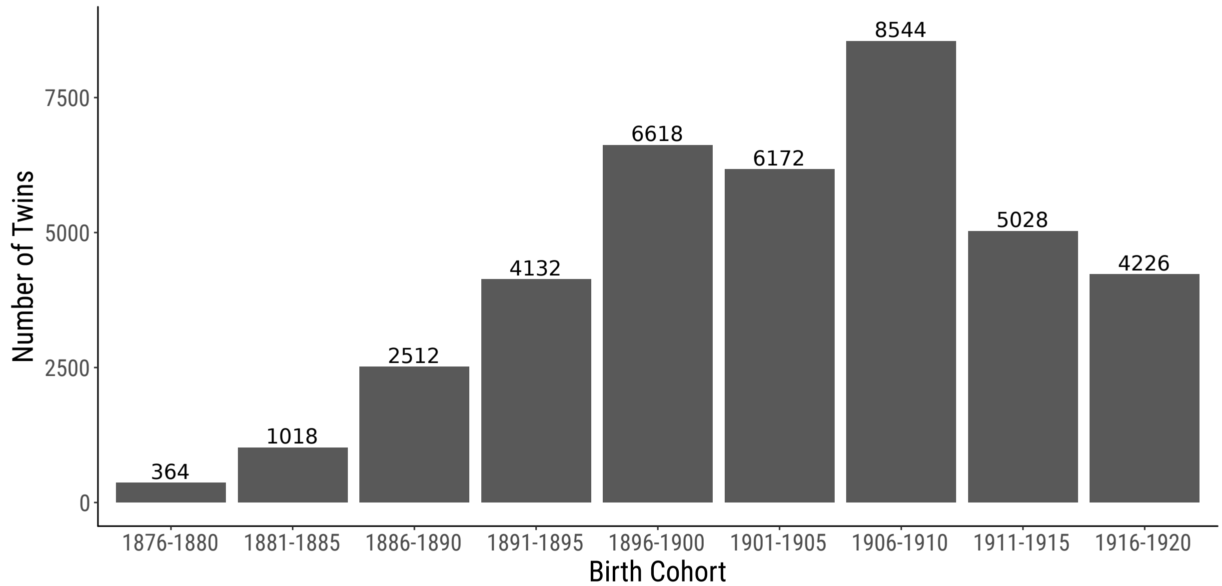

Figure A.7: Number of Twins in the Baseline Sample by Cohort. These sample sizes explain some of the variation in confidence intervals in Figure A.2. Because we only observe twins in 1900, 1910, and 1920 and can only identify twins when they are still living in their childhood homes (with their twin), we see very few twins who are over 20 years old.

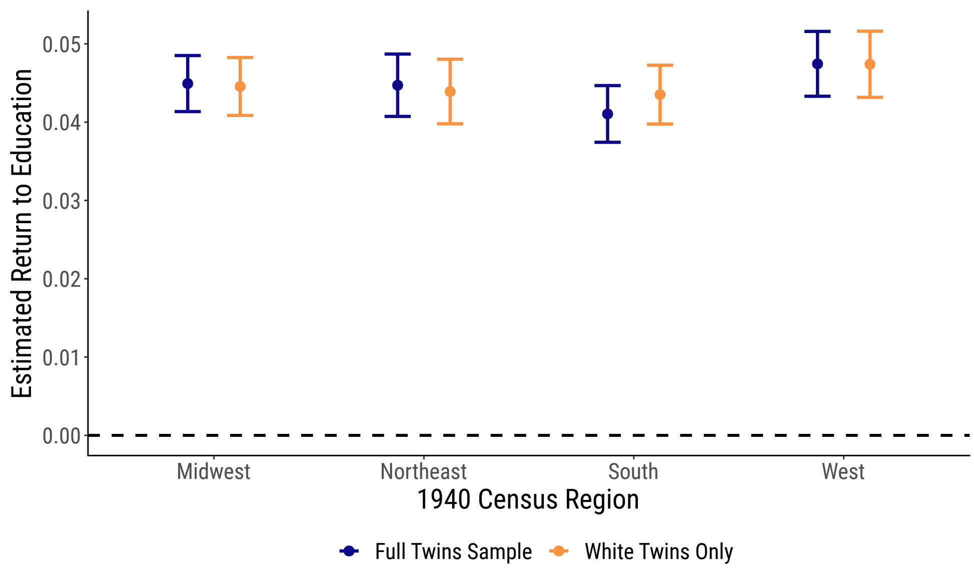

Figure A.8: The Return to Schooling in the Early 20th Century Did Not Vary Across Census Regions in 1940. We split our sample into four census regions based on where the twins were living when we observe them as adults in 1940. Data are from the pooled 1900-1940, 1910-1940, and 1920-1940 linked twins 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 (1939), and who worked a positive number of hours in the preceding week, according to the 1940 census. We estimate the return to schooling by region for the full sample and the subset of white twins. Following Bailey et al. (2019), we use inverse propensity weights to adjust for observable differences between matched and unmatched people in our census linked sample. 95% confidence intervals are shown, based on robust standard errors clustered at the twin-pair level.

Table A.1: The Return 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	0.062***	0.053***	0.065***	0.049***	0.059***	0.056***
	(0.002)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)
Good Controls	Yes	No	Yes	No	Yes	No
Twin Family FE	No	Yes	No	Yes	No	Yes
Observations	8546	8546	4668	4668	3878	3878
Adjusted R2	0.19	0.31	0.18	0.32	0.20	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 and family with the same age, birthplace, relationship to the head of household, and 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 earnings and education 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 who worked 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. (2019), we use inverse propensity weights to adjust for observable differences between matched and unmatched people in our census linked sample. Robust standard errors clustered at the twin-pair level are in parentheses.

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

Table A.2: The Return to Education Among Twins in Ashenfelter and Rouse (1998)

	Original Sample					Men Only		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Education	0.110***					0.102***		
	(0.010)					(0.017)		
Education Twin Differenced		0.070***	0.078***	0.068***	0.077***		0.068**	0.097***
		(0.019)	(0.018)	(0.019)	(0.018)		(0.032)	(0.031)
Constant	-1.095***			0.025	0.022	-1.223***	0.059	0.060
	(0.261)			(0.027)	(0.026)	(0.420)	(0.045)	(0.043)
Good Controls	Yes	No	No	No	No	Yes	No	No
Bad Controls Twin Differenced	No	No	Yes	No	Yes	No	No	Yes
Observations	680	340	333	340	333	274	136	132
Adjusted R2	0.33	0.04	0.17	0.03	0.16	0.28	0.03	0.18
Y Mean	2.44	0.03	0.04	0.03	0.04	2.64	0.07	0.08

Note: All columns present regressions of the log of earnings on years of education, drawing on the replication data for Ashenfelter and Rouse (1998). The first column replicates the result in Table II, column 2. The second and third columns replicate the results in Table III, columns 4 and 9. Ashenfelter and Rouse (1998) suppress the constant term in their first difference results, which we follow in columns 2 and 3. However, including the constants, as we do in columns 4 and 5, does not change the results. To ensure comparability with our male-only sample, columns 6 to 8 restrict the sample to the 274 identical male twins in Ashenfelter and Rouse (1998).

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

Table A.3: Twins with Similar Names Have More Similar Educational Outcomes in Childhood

Census	Outcome	Not Similar Names	Similar Names	Difference
	Panel A. Same First Letter			
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 ≤ 0.2			
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			
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 aged 7 to 17 in the given census years. We study twins in 1900, 1910, and 1920 as in our main analysis, as well as twins in 1940 when we can observe the highest grade completed. Panel A defines twins with similar names as twins whose first names start with the same letter. Panel B defines twins with similar names as twins whose first names are within 0.2 in Jaro-Winkler string distance. Panel C defines twins with similar names as twins whose first names have the same Soundex code. The column with the heading "Difference" represents the difference in educational outcomes between twins with more and less similar names, with standard errors reported in parentheses below. In all cases, the differences are positive, suggesting that twins with more similar names are more similar in terms of educational outcomes during childhood.

Table A.4: The Return 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***	0.035***	0.045***	0.037***	0.043***	0.036***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
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 while the 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 who worked 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. (2019), we use inverse propensity weights to adjust for observable differences between matched and unmatched people in our census linked sample. Robust standard errors clustered at the twin-pair level are in parentheses.

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

Table A.5: Sample Restrictions and Sample Size

	Sample Size			
	1900	1910	1920	Pooled
All Twins in Boy-Boy Pairs	232,188	206,141	255,597	693,926
Linked to 1940	99,343	94,528	118,498	312,369
Both Twins Linked to 1940	49,538	52,232	66,662	168,432
Twins Not Linked to Same 1940 Record	43,880	43,864	58,170	145,914
Neither Twin Missing Years of Schooling	42,252	42,332	56,472	141,056
Neither Twin Missing Earnings	17,038	21,800	32,212	71,050
Both Twins Wage and Salary Workers	10,474	14,798	21,998	47,270
Both Twins Positive Weeks of Work	10,320	14,574	21,660	46,554
Both Twins Positive Hours of Work	9,070	12,936	19,194	41,200
Both Twins Earned at least $6 per week	8,546	12,320	17,786	38,652

Note: In this table, we trace how our sample size shrinks as we link the full set of boy-boy twins identified in 1900, 1910, and 1920 ahead to 1940. That a non-trivial share of our twin pairs link to the same record in 1940 follows from the common practice of twins receiving similar names (see Table A.7). As we show in Figure A.4, our results are robust to all of the final restrictions to eliminate linked twins who did not work positive hours, weeks, earn more than $6 per week, or were not wage or salary workers.

Table A.6: Ability Bias Test

	Panel A. Full Sample	
	Correlation with Schooling	
	Overall	Within Families
Married	-0.072^{***}	0.007
Works Full-time	0.109^{***}	0.088^{***}
Number of Children	-0.182^{***}	-0.089^{***}
	Panel B. Sample with Spouses in 1940	
	Correlation with Schooling	
	Overall	Within Families
Works Full-time	0.131^{***}	0.089^{***}
Number of Children	-0.195^{***}	-0.113^{***}
Spouse's Years of Education	0.609^{***}	0.430^{***}
Spouse in Labor Force	0.003	0.002

Note: Testing for ability bias using several correlates of ability follows the approach in 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.

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

Table A.7: Most Common Twin Pair Names, Boy-Boy Twins Who Were Children in 1900-1920

	Name 1	Name 2
1	James	John
2	John	William
3	Ray	Roy
4	John	Joseph
5	George	John
6	James	William
7	John	Thomas
8	George	William
9	Frank	Fred
10	Floyd	Loyd
11	Frank	John
12	Charles	John
13	James	Joseph
14	Charles	William
15	Floyd	Lloyd
16	Joseph	William
17	James	Thomas
18	Edward	John
19	George	James
20	Richard	Robert

Note: Rankings of name pairs are based on the full set of twins identified in the 1900, 1910, and 1920 censuses. Name 1 is always alphabetically before Name 2. In addition to pairings of very common names (pairs including James, John, William, and George), we also see rhyming name pairs like Floyd and Loyd or Lloyd, as well as names sharing first initials like Ray and Roy or James and John or John and Joseph or Frank and Fred or Richard or Robert. Ronald and Donald, a common pairing in late 20th century twins data collected by Ashenfelter and Rouse (1998), is the 163rd ranked name pairing, still more common than Ronald and Donald on their own.

Table A.8: The Return to Education: Within or Across Industries?

	Baseline	Industry Code FEs		
	(1)	(2)	(3)	(4)
Years of Education	0.044***	0.042***	0.038***	0.037***
	(0.002)	(0.002)	(0.002)	(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 industry 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) industries. But in contrast to our finding that about one-third of the return to education comes from occupational upgrading (Table 7), significantly less comes from industrial upgrading. In all columns, we include twin family fixed effects, forcing the comparisons of earnings and education 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 who worked 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. (2019), we use inverse propensity weights to adjust for observable differences between matched and unmatched people in our census linked sample. Robust standard errors clustered at the twin-pair level are in parentheses.

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

Table A.9: Effect of Education on Different Types of 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***	0.023***	0.031***	0.026***	0.035***	0.031***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
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

Note: All columns present regressions of the log of occupation scores on years of education, drawing on our linked sample of twin brothers, linking twins from the 1900, 1910, and 1920 censuses to 1940. As all outcomes are in logs, they 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 1950 earnings data 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 1901 Cost of Living Survey with imputations of farmers' incomes. In columns 5 and 6, we create our own occupation scores based on the 1940 complete count census—IPUMS uses 1950—taking the median wage of male wage and salary workers aged 16 to 64, who worked at least 40 weeks in the preceding year (1939) and at least 35 hours in the preceding week. In the even columns, we include twin family fixed effects, forcing the comparisons of outcomes and education 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 who worked a positive number of hours in the previous week. The twins sample includes only twins who *both* fit our sample criteria. The sample size is slightly lower in columns 3 and 4 because one occupation code—545, airplane mechanics and repairmen—could not be imputed based on the 1900 data. Following Bailey et al. (2019), we use inverse propensity weights to adjust for observable differences between matched and unmatched people in our census linked sample. Robust standard errors clustered at the twin-pair level are in parentheses.

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

Table A.10: Effect of Education on Alternative Measures of Economic Status

	Top-Coded Earnings ×100		> \$50 Non-Wage Earnings ×100		Owns Home ×100	
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	0.321***	0.261***	-0.163***	0.251***	0.611***	0.617***
	(0.012)	(0.018)	(0.038)	(0.061)	(0.039)	(0.063)
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	1.06	1.06	32.32	32.32	45.85	45.85

Note: All columns present regressions of alternative measures of economic status 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 the even columns, we include twin family fixed effects, forcing the comparisons of outcomes and education 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. We estimate the effect of education on the probability that a twin has earnings top-coded (earned \$5,000 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. Following Bailey et al. (2019), we use inverse propensity weights to adjust for observable differences between matched and unmatched people in our census linked sample. Robust standard errors clustered at the twin-pair level are in parentheses. We scale indicator dependent variables by 100 to simplify the interpretation of effects and standard errors.

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

Table A.11: Effect of Education on Migration, Full Sample

	Panel A. Migration					
	Moved Out of County ×100		Moved Within State ×100		Moved Out of State ×100	
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	0.209***	0.775***	-0.885***	-0.566***	1.094***	1.341***
	(0.039)	(0.052)	(0.039)	(0.060)	(0.037)	(0.056)
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	60.70	60.70	34.03	34.03	26.67	26.67
	Panel B. 1940 Location Choice					
	Urban ×100		Log Size of Place		Farm ×100	
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	3.237***	2.184***	0.183***	0.122***	-2.761***	-1.831***
	(0.038)	(0.060)	(0.002)	(0.003)	(0.034)	(0.052)
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	52.07	52.07	9.00	9.00	25.98	25.98

Note: All columns present regressions of migration or residential 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 the even columns, we include twin family fixed effects, forcing the comparisons of outcomes and education 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 11. Following Bailey et al. (2019), we use inverse propensity weights to adjust for observable differences between matched and unmatched people in our census linked sample. Robust standard errors clustered at the twin-pair level are in parentheses. We scale indicator dependent variables by 100 to simplify the interpretation of effects and standard errors.

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

Table A.12: Effect of Education on Marriage and Fertility, Full Sample

	Married ×100		Any Children ×100		Number of Children	
	(1)	(2)	(3)	(4)	(5)	(6)
Years of Education	-0.289***	0.544***	-1.092***	-0.242***	-0.075***	-0.038***
	(0.036)	(0.053)	(0.039)	(0.063)	(0.001)	(0.002)
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	72.13	72.13	52.42	52.42	1.28	1.28

Note: All columns present regressions of 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 the even columns, we include twin family fixed effects, forcing the comparisons of outcomes and education 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 12. Following Bailey et al. (2019), we use inverse propensity weights to adjust for observable differences between matched and unmatched people in our census linked sample. Robust standard errors clustered at the twin-pair level are in parentheses. We scale indicator dependent variables by 100 to simplify the interpretation of effects and standard errors.

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