FEDERAL RESERVE BANK *of* ATLANTA WORKING PAPER SERIES

The Evolution of US Educational Mobility over the 20th Century and the Role of Public Education

Martha J. Bailey, A.R. Shariq Mohammed, and Paul Mohnen

Working Paper 2026-1
January 2026

Abstract: We construct two new large-scale datasets to measure relative and upward educational mobility by sex, race, class, and childhood county of residence for cohorts born in 1910−1919 and 1982−1997. We show that both relative and upward educational mobility rose over the 20th century, with historically disadvantaged groups experiencing the largest gains. We also document substantial geographic convergence over the 20th century: both within and across regions, where children live matters much less for their educational mobility today than it did at midcentury. Using a state-border design, we show that greater public investments in primary and secondary education were an important driver of upward educational mobility in the early and late 20th century, but public investments in postsecondary education emerged as a similarly important determinant in the late 20th century.

JEL classification: J62, I24, I28, N32

Key words: education, inequality, intergenerational mobility

https://doi.org/10.29338/wp2026-01

Martha J. Bailey (marthabailey@g.ucla.edu) is a professor of economics and director of the California Center for Population Research at the University of California, Los Angeles, and a research associate at the National Bureau of Economic Research. A.R. Shariq Mohammed (ab.mohammed@northeastern.edu) is an assistant professor of economics at Northeastern University. Paul Mohnen (paul.mohnen@atl.frb.org.with) is a research economist at the Federal Reserve Bank of Atlanta. For helpful comments, the authors thank Dan Aaronson, Sandy Black, Marcus Casey, Dan Fetter, Sasha Indarte, Bhash Mazumder, Matt Staiger, Bryan Stuart, and participants at the ASSA, Atlanta Fed, Cliometrics, EHA, Equitable Growth, NBER Development of the American Economy, NBER Summer Institute, RNIM, SOLE, Philadelphia Fed, UCLA, and University of Pennsylvania. They thank Alex Coblin, Jordan Cammarota, Kelsey Figone, Deniz Gorgulu, David Lee, Avinash Reddy, Joaquín Serrano, Eric Wang, and Mengying Zhang for excellent research assistance. They also thank Jonathan Rothbaum and Hui Ren Tan for sharing their estimates, and James Feigenbaum for sharing the linked Iowa-1940 Census data. They thank Caleb Floyd for guidance with the disclosure process. The LIFE-M project was generously supported by the National Science Foundation (SMA 1539228), the National Institute on Aging (R21 AG05691201), the University of Michigan Population Studies Center Small Grants (R24 HD041028), the Michigan Center for the Demography of Aging (MiCDA, P30 AG012846-21), the University of Michigan Associate Professor Fund, the Michigan Institute on Research and Teaching in Economics (MITRE), the Russell Sage Foundation (1911-19560), and the Washington Center for Equitable Growth. They gratefully acknowledge the use of the Population Studies Center's services and facilities at the University of Michigan (P2CHD041028) and the California Center for Population Research at UCLA (P2CHD041022). The Census Bureau has reviewed this data product to ensure appropriate access, use, and disclosure avoidance protection of the confidential source data used to produce this product. This research was performed at a Federal Statistical Research Data Center under FSRDC Project Number 2548 (CBDRB-FY24-0379, CBDRB-FY25-12440). The views expressed herein are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Atlanta, the Federal Reserve System, or the US Census Bureau.

In the late 19th and early 20th century, universal public education made the U.S. the most widely educated countries in the world (Goldin and Katz 2008). Building on this foundation, the High School Movement catapulted the share of Americans graduating from high school from 10 percent in 1910 to over 50 percent by 1940 (Goldin 1998). After World War II, total postsecondary enrollment had increased by almost 50 percent by 1960, doubling again by 1970, and rising by another 2.3 times by 2019 (U.S. Department of Education 2023a). These remarkable gains in educational attainment were accompanied by a 20-fold increase in public elementary and secondary school spending between 1919 and 2019 (U.S. Department of Education 2023b) and large increases in public post-secondary educational spending (Goldin and Katz 2008; Long 2013).

How did this large expansion in public resources for America's schools affect the evolution of educational opportunity? Theoretically, increased spending may have leveled the playing field, allowing children from a variety of backgrounds to transcend the achievements of their parents. On the other hand, educational spending may have disproportionately benefitted more advantaged groups, amplified inequality, and done little to promote more equal opportunity. To date, a lack of data and measurement challenges have limited conclusions about trends in and determinants of educational opportunity over this important period of U.S. history.

This paper builds two new large-scale datasets spanning 75 years to provide novel evidence on the relationship between public educational spending and the evolution of U.S. educational opportunity. The first dataset includes nearly 1.5 million children born in the U.S. between 1910 to 1919 who completed their schooling between 1930 and 1940; the second includes 2.1 million children born in the U.S. between 1982 to 1997 who completed their schooling between 2002 and 2022. This early 20th century sample is created by linking Social Security application records (SS-5), containing boys, girls, and their parents, to their completed years of education in the full-count 1940 Census—the first to ask about completed education—and the 1920 Census which allows us to connect these outcomes to county of residence in childhood and local public school resources. Our sample of later 20th century cohorts relies upon the U.S. Census Bureau's links of the restricted 2000 Decennial Census, SS-5 records, and the

2001–2022 *American Community Surveys* (*ACS*), which allows us to relate children's completed years of education to their parents' as well as to county of residence in childhood and public school resources.

Our analysis begins with a novel description of both relative and upward intergenerational educational mobility at two points in the 20th century as well as differences across subgroups and by county in both periods. Our measures of educational mobility rely upon the relationship between parents' and children's *relative ranking* in years of schooling rather than educational attainment. Similar to modern estimates of income mobility, we find that the conditional expectation of a child's years of education, given a parent's years of education, is roughly linear in percentile ranks (Chetty et al. 2014a; Fletcher and Han 2019). The slope measures *relative* educational mobility, whereas the intercept measures the *absolute* educational mobility of children of the least educated parents. This linear relationship helps smooth the lumpy educational distribution, which complicates intertemporal and spatial comparisons. (For example, almost one third of the parents of the early 20th century had 8 years of education, whereas roughly one third of the parents of the later 20th century cohorts had exactly 12 years of education, which makes it difficult to find a common percentile to compare in the two distributions.) The rank-rank framework allows us to provide a fuller characterization of educational mobility across the educational distribution (rather than at a single point in the distribution), while minimizing privacy concerns associated with disclosure from the Federal Statistical Restricted Data Center (FSRDC). Following the literature on income mobility, we also present the expected rank of children with their most educated parent at the 25th percentile, which captures *upward* educational mobility.

Our first set of results shows that educational mobility increased over the 20th century, characterized by a flattening rank-rank slope and an increasing expected rank of children whose parents had below-median education. In the early 20th century, a 10-percentile increase in a parent's educational rank was associated with a 4.9-percentile increase in a child's educational rank. In the late 20th century, this association had fallen to 3.9 percentile ranks, indicating less persistence and more mobility across generations. Among children of parents at the 25th percentile of the parental educational distribution, the expected rank is the 36th percentile in the early 20th century and the 38th percentile in the later 20th century. While the magnitudes of changes in upward mobility appear small, these aggregate trends mask rising upward educational mobility among more historically disadvantaged groups, including Black boys and girls (10 and 12 percentile gains, respectively) and Asian boys and girls (12 and 14 percentile gains, respectively). In contrast, the upward mobility of White women barely increased, while the upward mobility of White men fell by 5 percentiles. Another striking pattern is the widening gender gap in upward educational mobility across all racial groups: girls had slightly higher rates of upward mobility in the early 20th century. By the century's end, girls saw a 5 percentile point increase in their upward mobility while boys saw a 1 percentile point decline.

A second set of results characterizes the geography of educational mobility in the early and late 20th century. We find that, historically, upward mobility was uniformly lower throughout the South, with the former slave states almost completely coincident with *little to no* upward educational mobility in the early 20th century. In contrast, counties in the West, Northeast, and Industrial Midwest had very high rates of upward educational mobility. By the later 20th century, upward educational mobility improved for children in almost every county in the South. In contrast, upward educational mobility fell in the West, Northeast, and Industrial Midwest. Put simply, all regions converged towards the national average over the 20th century. While educational opportunity depended on where children grew up in the past, place matters much less today.

A final analysis examines the effect of resources for public education on upward educational mobility using a state-border design. Supporting the internal validity of this approach, we show that a variety of county characteristics are balanced within county pairs on opposite sides of state borders. However, public investments in primary, secondary and post-secondary education vary a great deal across these borders, due to differences in state educational policies. Comparing observationally similar counties on opposite sides of state borders reveals that more K-12 teachers per child and higher salaries for K-12 teachers led to higher upward educational mobility in the early 20th century. Repeating this exercise in the late 20th century reveals that K-12 teacher salaries continue to play an important role, but the number of instructors at public colleges and universities per child has become a similarly important determinant of upward educational mobility.

This paper contributes to the literature by providing new evidence on the evolution of educational opportunity across the 20th century U.S. within a common measurement framework—including differences by sex, race and geography—and characterizing how educational opportunity was shaped by public investments in schools. To date, most historical evidence on educational mobility in the U.S. has focused on persistence estimates at the national level at various points in the 20th century (Couch and Dunn 1997; Hertz et al. 2007; Hout and Janus 2011; Torche 2015; Hilger 2017; Fletcher and Han 2019; Ferrie et al. 2021). Althoff et al. (2025) explore how the spread of universal public primary education during the late 19th and early 20th century reduced the link between children and parents' education (particularly mothers). Closest to our paper is Card et al. (2022) who analyze the upward educational mobility of children ages 16–18 living with their parents in the 1940 Census (born in 1922–1924). They focus on children whose parents had no more than 6 years of education (~25th percentile), and define upward educational mobility as attaining at least an 8th grade education (roughly exceeding 25th percentile). Their analysis documents significant regional differences in upward educational mobility and uses a state-border research design to identify the causal role of educational resources on differences in Black-White upward mobility gaps. Like this paper, we document upward educational mobility in the 1940 Census for a slightly older cohort and also find that teacher salaries play a large, causal role in explaining geographic differences in educational mobility. Our paper broadens this analysis by examining both relative and upward educational mobility over a 75-year period and by examining completed education for individuals who no longer live with their parents. Our paper also adds measures of public investments in post-secondary education, providing a richer characterization of educational inputs beyond primary and secondary school.

A second contribution relates to our geographic comparisons of educational mobility within the U.S., which complements research documenting geographic variation in income and occupational mobility (Chetty et al. 2014b; Connor and Storper 2020; Massey and Rothbaum 2021; Tan 2023). Most studies examining geographic variation in educational mobility have used cross-country comparisons (Black and Devereux 2011). Hertz et al. (2007) find large differences in intergenerational educational correlations across 42 countries, but these correlations have been relatively stable over time. Italy's intergenerational education correlation fell from 0.58 (cohorts 1910–1914) to 0.47 (post-1970) (Checchi et al. 2013), whereas indirect evidence from the U.K. and Spain suggest rising intergenerational educational persistence (Blanden et al. 2004; Güell et al. 2015). By contrast, educational persistence in Germany has remained largely unchanged (Heineck and Riphahn 2009), while educational persistence in Denmark declined for cohorts born in the 1940s through 1960s before rising again for subsequent cohorts (Karlson and Landersø 2025). Recent studies have also examined trends in educational mobility in the context of developing countries (Alesina et al. 2021; Asher et al. 2024). Our geographic comparisons use consistent measures of educational mobility over time to examine changes within a single country over the 20th century, add estimates of upward educational mobility to more standard relative mobility estimates, and relate geographic differences to public investments in education.

Third, our results contribute to an extensive literature on intergenerational occupational and income mobility in the U.S. (Solon 1992; Zimmerman 1992; Ferrie 1996; Mazumder 2005; Long and Ferrie 2013; Mazumder 2016; Bratberg et al. 2017; Feigenbaum 2018; Song et al. 2020; Bailey et al. 2020b; Abramitzky et al. 2021; Collins and Wanamaker 2022; Ward 2023; Jácome et al. 2025; Davis and Mazumder 2025). These studies by necessity limit their analyses—especially in the historical period—to men who are employed or have an occupation, excluding those without an occupation or earnings (e.g., women) or measuring women's mobility according to their husband's outcomes (Eriksson et al. 2023; Espín-Sánchez et al. 2023; Bailey and Lin 2025). While educational mobility is a key input into intergenerational occupational and income mobility (Blau and Duncan 1967; Sewell and Hauser 1975; Hout and DiPrete 2006), data availability has limited research on these interrelationships. Our study of educational mobility provides a complementary lens for a broader set of individuals through which to understand the evolution of economic opportunity in the 20th century U.S.

Finally, our study contributes to a deeper understanding of the determinants of intergenerational mobility. A central finding for the late 20th century is that spatial differences in income mobility are driven by factors that affect children while they are growing up and before they enter the labor market, whereas strong or robust associations with measures of class size or public school expenditures per student are lacking (Chetty et al. 2014b). This paper deepens our understanding of the pre-labor market determinants of income mobility by demonstrating the changing role of public education policy in driving economic opportunity.

I. Measuring U.S. Educational Mobility across Time and Place

Studies of intergenerational mobility aim to capture the degree to which a child's social and economic status are related to her parents' social and economic status. Rather than more standard measures of occupation or income mobility, this study aims to characterize the joint distribution of a child's completed years of education and parent's completed years of education.

We focus on education because it is a key mechanism for intergenerational occupational and income mobility, which have been studied extensively (Blau and Duncan 1967; Sewell and Hauser 1975; Hout and DiPrete 2006). Education both reflects the social and economic status of parents and determines a child's social and economic status, by preparing individuals for jobs, determining social and economic opportunities, and affecting one's mate. Education is also conceptually different: it captures earnings and occupation *potential* as opposed to realized outcomes, which may be especially important if some individuals choose lower status or wage professions (or mates) than their education facilitated.

The measurement of education also alleviates several challenges inherent in using income and occupation. First, education is completed for most adults by their mid-20s and remains stable over their lifetimes. This makes education less susceptible to transitory economic shocks in adulthood, which have limited studies of occupational and income mobility (Solon 1999; Mazumder 2005; Haider and Solon 2006; Mazumder 2018).

Second, education is measured for *all* adults, not just those who are employed, have an occupation, or earn wages. This is especially important for intertemporal comparisons for the broader American population, because wage income and occupation information is missing for a large number of Americans in the early 20th century and especially in certain subgroups. In 1940, only 14 percent of married women participated in the labor force, limiting measurement of their income and occupational mobility (Goldin 1990; U.S. Bureau of Labor Statistics 2023).[1] 17 percent of adult men were farmers or farm laborers in 1940, and much of their in-kind or farm earnings was not reported in the 1940 Census. But the issue of unobserved income is broader than just farmers: 17 percent of prime-age men were self-employed in 1940, and one quarter reported receiving over 50 dollars in non-wage income (although the amount of that income was not asked). Altogether these groups overlap, it is clear that a substantial part of the U.S. population in 1940 did not have an occupation or wage earnings and many had incomplete income information. While considerably more women work today, fewer Americans are farmers, and the measurement of income has improved over the 20th century, differences in missing data for the earlier 20th century limits the interpretation of intertemporal comparisons of the broader population. Using education allows us to characterize the evolution of intergenerational mobility over the 20th century for a much broader set of Americans.

A. Available Measures of Education

The 1940 Census was the first decennial census to ask about educational attainment, defined as the highest year of school or degree completion. Self-reported educational attainment is also available in every census year after 1940 and in every year of the *ACS*, with minimal changes to the definition over time. Since 1990, respondents with no more than high school were classified according to their highest year of education completed, whereas those with more than high school were classified according to their highest degree earned. Because this definitional change applies to everyone in the later 20th century cohorts, we expect it to have little effect on educational *rankings* in our study.

[1] Limited information on women's own incomes or occupations limits analyses of women's intergenerational mobility to indirect methods or analyses of their husbands' or fathers' outcomes (Olivetti and Paserman 2015; Eriksson et al. 2023; Espín-Sánchez et al. 2023; Bailey and Lin 2025).

Despite its advantages, several issues around the measurement of education affect our analysis. First, we measure parents' education in the 1940 Census at older ages than their children, raising concerns about the influence of "educational creep," the well-known tendency for educational attainment for the same birth cohort to rise over time, on our analysis (Margo 1986). To address this concern, we later show that measuring parents' education at different ages has negligible effects on our results. Second, most children have two parents and therefore two measures of parental education, and theory provides little guidance about which measure to use. Our main results use the education of the most educated parent as our primary measure, which has the benefit of allowing us to characterize education when only one parent is observed (especially for late 20th century cohorts, where single-parent households are more common). However, we later show that our results are unchanged by using alternative measures of parental education.

The central challenge to studying intergenerational educational mobility over time is that the completed education of parents and their adult children is not observed for children who no longer co-reside with their parents (around 84 percent of adults over age 21 in the 1940 Census and 88 percent in the 2018–2022 *ACS*). While retrospective survey evidence or longitudinal data could be used for some analyses, their sample sizes are too small to study the evolution of education mobility by county. We address this data limitation by creating two large-scale linked census datasets for the early and late 20th century cohorts and validating estimates derived from these samples using nationally representative surveys.

B. Data for Studying Intergenerational Educational Mobility in the Early 20th Century

Our early 20th century cohorts are based on the public Social Security application (Form SS-5) records contained in the Social Security Numerical Identification (Numident) Files released by the National Archives and Records Administration. These data cover around 40 million individuals who ever applied for a Social Security Number (SSN), died prior to 2007, and whose deaths were not state-reported—the near universe of individuals who died between 1988 and 2007 (Goldstein and Breen 2022). Our estimates suggest that the SS-5 data contain around 50 percent of the 1920 birth cohort

(Appendix Figure A1.A) and have excellent geographic coverage (Appendix Figure A1.B). These data contain full name, exact birth date, sex, race, and birthplace as well as the full birth names of each Social Security applicant's parents, which is crucial for linking children to their parents.[2] Our early 20th century cohorts restrict this sample to children born from 1910 to 1919 because they have high coverage rates in the SS-5, are likely to have at least one parent alive in the 1940 Census, and are old enough in 1940 (ages 21–30) to have completed their education.[3] We also restrict our sample to children born in the U.S. because immigrant children may have completed their schooling abroad.

Appendix Figure A3.A describes how we link these data to other sources. Our methodology follows the supervised machine learning approach of the LIFE-M project (https://life-m.org), which confers the benefits of hand-linking at scale and limits the number of false positives (Bailey et al. 2023). We first link siblings together in the SS-5 data using parent names. Then, we link parent *and* children's information (names, ages, and birth states) as families to the 1900–1930 Censuses. LIFE-M's family-based linking has been found to be highly accurate and increase linking rates (Abramitzky et al. 2025; Ruggles et al. 2025). The 1900–1930 Censuses add information on parents' year and place of birth, which we then leverage to link parents to the 1940 Census: (1) as families, (2) as couples, and (3) as individuals. Linking parents in multiple ways helps us find them in the 1940 Census, irrespective of their household structure and also helps us identify conflicts and cull bad links from the data. We also link SS-5 children to the 1940 Census based on their full name, age (based on the exact date of birth), and birthplace information. We attempt to link girls using both their birth and married names and then reconcile the data when conflicts arise, which again helps cull bad links. (See Appendix A for more details about the linking process.)

[2] Women typically specify their legal surname when filing Form SS-5. For single women, this is their birth name. For married women, it is their husband's birth name. If women change their name after their initial SS-5 filing (e.g., they subsequently get re-married), they usually file an updated SS-5. As a result, women have 1.8 entries in the SS-5 data on average, and 94 percent of women have at least one surname that differs from their father's surname, which is approximately the share of women who had ever been married for these cohorts (Bailey et al. 2014).
[3] Appendix Figure A2 shows that the educational gains past age 20 were small for individuals born in 1910–1919.

Our final early 20th century sample contains around 1.5 million children born in the U.S. between 1910 and 1919 with at least one parent linked to the 1940 Census, and their residence observed in 1920. This is a final match rate of 18.4 percent at 95-percent precision of individuals containing all information of interest relative to all individuals meeting our sample criteria in the SS-5 records (Appendix Table A2). Almost 50 percent of this sample are women. Because linked samples are not generally representative of the population, we use inverse propensity scores to re-weight the sample to match the sex-by-race-by-cohort and education distribution in the 1940 population using the procedure described in Bailey et al. (2020a). These weights help correct for the fact that earlier cohorts are underrepresented in SS-5 records and the fact that Black Americans and less-educated individuals are underrepresented in the final sample because they are harder to link. Appendix Table A3 (column 5) shows that the reweighted linked samples appear well balanced with the population in terms of their age, race, and educational attainment relative. (See Appendix B for details about the creation of these weights). Fetter et al. (2025) show that SS-5 records are not only representative in terms of state of birth but also state of death.

C. Data for Studying Intergenerational Educational Mobility in the Late 20th Century

Appendix Figure A3.B describes how we construct an analogous linked sample for late 20th century cohorts using high-quality Personal Identification Keys (PIKs) created by the Census Bureau and available in the restricted Federal Statistical Research Data Centers (FSRDCs).[4] PIKs allow us to combine restricted SS-5 records, the 2000 Census, and the 2001–2022 *ACS*. We begin by selecting a sample of children ages 3 to 18 in the 2000 Census who are living with at least one biological parent. We use their demographic information (year of birth, sex, race) from the SS-5, which tends to be more reliable than self-reported information. Then, we find these children as adults in the *ACS* after they turn 25, which contains their completed education. The 2000 Long-Form Census contains parents' education for a 1-in-6 sample of the population. For parents not in the 2000 Long-Form Census, we locate them in the *ACS* to obtain their completed education.

[4] The Census Bureau has access to the universe of SS-5 records as well as various administrative sources, which it uses to assign unique PIKs and find the same individuals across data sources (Wagner and Layne 2014).

Our final sample contains 2.1 million children born in the U.S. between 1982 and 1997, for whom we observe the adult child's and at least one parent's educational attainment and the child's county of residence in 2000. The 2.1 million children are around 3.5 percent of the total number of children born in the U.S. in 1982–1997 in the 2000 Census. The low coverage rate is not a reflection of poor PIKing rates but the fact that all children as well as parents who do not appear in the 2000 Long-Form Census are linked to the *ACS*, where each survey year is a 1-percent sample of the population. Our additional requirement that children be at least age 25 means that younger cohorts are slightly underrepresented (since they are age 25 in fewer *ACS* years). Consequently, we similarly construct inverse propensity scores to re-weight the sample to match the sex-by-race-by-cohort distribution in the 2000 population. (See Appendix B for details.)

D. Benchmarking Linked Samples against Nationally Representative Surveys

The methods used to link the data affect inference, either by affecting the rate of false links (type I errors) or the composition of the sample (type II errors) (Bailey et al. 2020b). We therefore benchmark and validate our estimates using two *unlinked* nationally representative surveys: (1) the 1977–2022 *General Social Survey* (*GSS*; Davern et al. 2024), and (2) the 1987–1988 *National Survey of Families and Households* (*NSFH*; Bumpass et al. 2017). These surveys are designed to be representative of the U.S. population, cover both men and women for our cohorts of interest, and contain retrospective information on respondents' completed education as well as the education of their fathers and mothers. We restrict these samples to U.S.-born individuals ages 25 or older at the time of the survey and pool the *GSS* and *NSFH* to increase precision. We follow Jácome et al. (2025) and use normalized sampling weights and further re-weight the sample for the 1910–1919 cohorts to match the 1940 population in terms of demographics and education, which adjusts for longevity bias as well as other sources of imbalance. (See Appendix B for details.)

II. Rank-Rank Estimates of Intergenerational Educational Mobility

The most common measure of educational mobility is *intergenerational educational persistence*, or the slope coefficient from a regression of a child's years of education on a constant and a measure of parental education:

$$Y_i^c = \alpha + \beta X_i^p + \varepsilon_i . \tag{1}$$

The slope coefficient, β, captures persistence in educational attainment across generations, which may reflect both low upward mobility and low downward mobility (Solon 1999; Black and Devereux 2011). The slope coefficient captures the product of the correlation between child and parent outcomes as well as the ratio of the standard deviation of completed years of education in the child's generation relative to the standard deviation of completed years of education in the parent generation. The slope coefficient will, therefore, increase as inequality in the outcome of interest rises, even if the intergenerational correlation does not change. To account for rising inequality across generations, researchers in the income mobility literature have alternatively reported the correlation coefficient or correlation between parent and child ranks (Dahl and DeLeire 2008; Chetty et al. 2014a).

Given the rise in educational inequality across the 20th century, we follow the latter approach and use ordinary least squares regression to summarize the rank-rank relationship of parent and child's educational attainment,

$$R_i = \gamma + \delta P_i + \varepsilon_i , \tag{2}$$

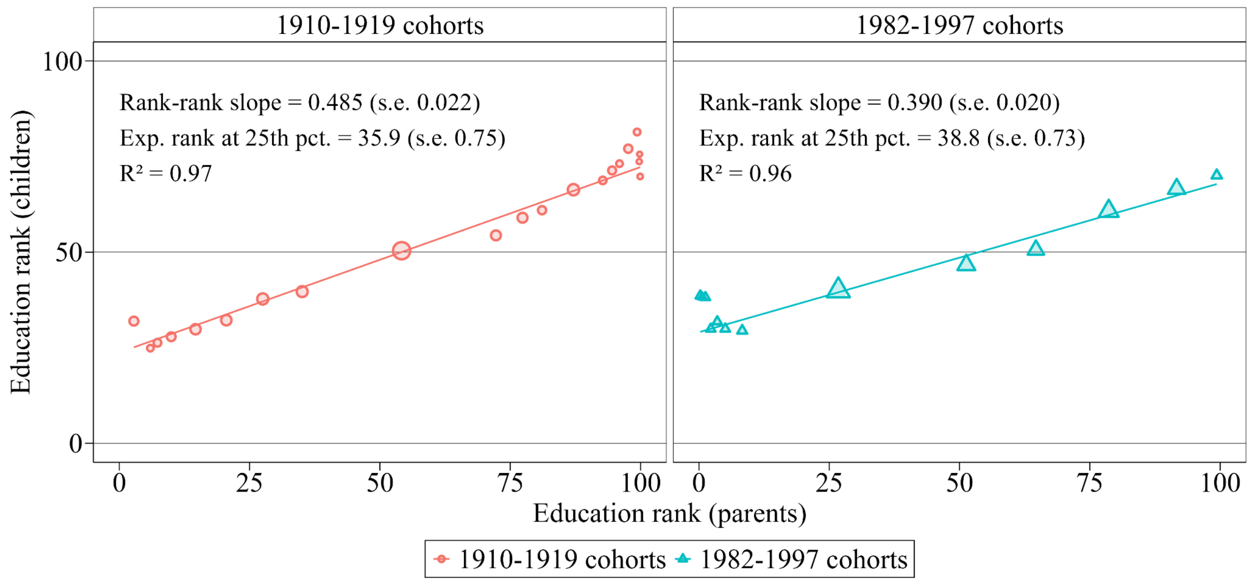
where R_i is the rank of the child's educational attainment and P_i is the rank of the educational attainment of her most educated parent. The rank of the child is determined relative to other children in her birth cohort, and the rank of the parent is determined relative to other parents of children in the relevant birth cohort. The rank-rank slope, δ, measures relative educational persistence (the inverse of relative mobility), and the intercept, γ, measures absolute educational mobility of children with the least educated parents.

Many readers will be familiar with the advantages of using the rank-rank approach for income mobility, but its use in the context of education has some additional benefits. First, this linear parameterization helps smooth the lumpy educational distribution (many individuals have the same number of years of completed education) and parameterize educational mobility at percentiles not directly observed in both time periods. In the early 20th century cohorts, more than a quarter of children report 12 years of education and almost one third of their parents report 8 years of education as their highest level of completion (see Appendix Figure A4 for parent and child educational distributions). This lumpiness means that comparisons at the same percentile of the parental educational distribution is not well defined at many points in our two cohorts of interest, making intertemporal comparisons difficult to interpret. The linear parameterization, however, allows us to predict the expected percentile rank of a child with a parent at the same percentile of educational attainment in both periods. To facilitate comparisons with the income mobility literature, we define upward educational mobility as the expected rank at the 25th percentile of the parental education distribution, or $E[R_i \mid P_i = 25] = \gamma + 25\,\delta$ (Chetty et al. 2014b).

Figure 1 plots these outcomes using the percentile midpoint of each educational category, and Appendix Figure A5 presents the transition matrices for each cohort. Similar to the results for income (Chetty et al. 2014a), the expectation of a child's years of education conditional on parent's years of education appears linear in percentile ranks for both the early and later 20th century cohorts. This finding also accords with Fletcher and Han (2019), who document a linear relationship between child and parental education ranks for cohorts born from the 1960s to 1980s using survey data.

Appendix Table A4 shows that the linear relationship also holds when making different assumptions about percentile ranks; it holds when using the lowest rank within an educational category for parents (lower bound, LB) and the highest rank for children (upper bound, UB); the UB for parents and the LB for children; as well as two other combinations of these ranks. Importantly, the slopes of the fitted lines are statistically indistinguishable from one another (p-value for the test of the joint hypothesis that all slopes are equal is 0.54 for the 1910–1919 cohorts and 0.35 for the 1982–1997 cohorts). Although our measure of absolute mobility in a given period is affected by these choices, our conclusions about *changes* over time in the expected rank at the 25th percentile are not. Our remaining discussion, therefore, focuses on the midpoint definition.

Figure 1. The Relationship Between Child and Parent Ranks in Years of Education, by Cohort



Notes: This graph plots the average education rank of children by the rank of their most educated parent using the midpoint method. Education distributions are shown in Appendix Figure A4. Fitted lines are based on weighted linear regressions. The procedure for creating weights is described in the text and Appendix B.

Because some readers may prefer a direct measure of upward mobility to a predicted measure, we also present the likelihood of reaching the top 60 percent of the education distribution conditional on parents being in the bottom 40 percent. The 40th percentile is one of the few thresholds that can be (approximately) observed for both the 1910–1919 and 1982–1997 birth cohorts and their parents.[5] For children born in 1910–1919, this measure of upward mobility captures the probability of attaining at least 10 years of education conditional on parents having at most 7 years of education. For children born in 1982–1997, this measure captures the probability of attaining at least 13 years of education conditional on parents having at most 12 years of education. This measure is similar to Card et al. (2022), who use younger children (born from 1922 to 1924 in the 1940 Census) residing with parents who have no more

[5] Appendix Figure A4 shows that the education distribution for children born in the early and late 20th century can be divided into the top 64 percent and bottom 36 percent. The corresponding parental education distributions can be divided into the top 61 or 57 percent and the bottom 39 or 43 percent.

than 6 years of education (approximately the 25th percentile), to study the likelihood that children achieve at least 8 years themselves (exceed the 25th percentile). Examining this single point in the distribution is an important point of comparison for our analysis and provides an additional way to validate our linked-sample findings against a slightly younger but unlinked sample. However, we focus our discussion on the linear parameterization, which allows a characterization of relative mobility across the educational distribution (not just at one point) and also more direct comparisons to educational mobility in the later 20th century.

III. The Evolution of Educational Mobility over the 20th Century

The rapid expansion of universal public education over the 20th century increased educational attainment. Whether these growing resources reduced the intergenerational transmission of education from parents to children, opening new opportunities for historically disadvantaged groups, remains an open question. This section uses large-scale linked samples to describe the evolution of intergenerational educational mobility over the 20th century by class, sex, and race/ethnicity.

A. Relative and Upward Educational Mobility in the U.S. over the 20th Century

Table 1 presents estimates of educational mobility for the early (panel A) and later 20th century cohorts (panel B), including two measures of relative mobility (the rank-rank and level-level slopes) and two measures of upward mobility (expected rank at the 25th percentile and the share of children moving from the bottom 40 to the top 60 percent).

Across both the rank-rank and level-level measures, we find that relative educational mobility has risen. During the 75 year period, the rank-rank slope fell from 0.485 to 0.394 while the level-level slope fell from 0.434 to 0.292, both of which are statistically significant at the 1-percent level (panel C, columns 1 and 3). We also find that upward educational mobility rose according to both measures. The expected rank of children of parents at the 25th percentile increased by 1.8 (column 5), and the share of children born to parents in the bottom 40 percent who exceeded the 40th educational percentile rank rose by 6 percentage points (column 7).

Table 1. Educational Mobility Estimates by Cohort and Sample

	Rank-rank slope		Level-level slope		Expected rank at 25th percentile		Bottom 40 to top 60 percent mobility	
	Linked data	Survey data	Linked data	Survey data	Linked data	Survey data	Linked data	Survey data
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. 1910–1919 cohorts								
Estimate	0.485 (0.001)	0.464 (0.019)	0.434 (0.002)	0.410 (0.023)	35.9 (0.05)	36.2 (0.81)	41.6 (0.11)	41.0 (2.03)
N	1,457,003	2,851	1,457,003	2,851	1,457,003	2,851	424,281	843
Panel B. 1982–1997 cohorts								
Estimate	0.394 (0.001)	0.359 (0.021)	0.292 (0.001)	0.282 (0.021)	37.7 (0.03)	39.4 (0.88)	47.6 (0.07)	50.9 (1.83)
N	2,136,000	3,001	2,136,000	3,001	2,136,000	3,001	815,000	1,228
Panel C. Change across cohorts								
Estimate	-0.091	-0.105	-0.142	-0.127	1.8	3.2	6.0	9.9
p-value for rejecting change = 0	0.000	0.000	0.000	0.000	0.000	0.007	0.000	0.000
p-value for rejecting change (linked) = change (survey)	0.622		0.631		0.242		0.154	

Notes: This table shows educational mobility estimates by cohort and sample. The rank-rank slope and expected rank at the 25th percentile come from a regression of children's education rank on parents' education rank. The level-level slope comes from a regression of children's years of education on parents' years of education. Bottom 40 to top 60 percent mobility is defined as the mean probability of reaching the top 60 percent of the cohort-specific education distribution among the subset of children whose parents are in the bottom 40 percent of the cohort-specific parental education distribution (see text for cohort-specific cutoffs). Observations for the 1910–1919 cohorts are weighted by inverse propensity scores that re-weight the sample to match the sex-by-cohort-by-race and education distribution in the 1940 population. Linked data observations for the 1982–1997 cohorts are weighted by inverse propensity scores that re-weight the sample to match the sex-by-cohort-by-race distribution in the 2000 population. Survey data observations for the 1982–1997 cohorts are weighted by *GSS* sampling weights. Robust standard errors in parentheses (calculated using the Delta method for expected rank at the 25th percentile). *p*-values in the bottom two rows respectively test the null hypotheses that changes in educational mobility across cohorts for a given sample are equal to 0, and the null hypotheses that changes in educational mobility are equal across samples.

Comparable statistics from nationally-representative survey data support these findings. Although survey estimates are noisier, both educational persistence and upward educational mobility estimates from the linked sample fall within the 95-percent confidence intervals of the survey estimates. Importantly, we fail to reject the equality of changes over time in the linked and survey samples, which supports our conclusions about changes over time using linked data. Appendix Figure A7 complements these statistics by plotting relative and absolute educational mobility over the 20th century in both survey and linked samples, highlighting the similarity of estimates from both sources.[6]

These findings also survive two additional sensitivity tests. The first examines the sensitivity of our findings to measuring parent's education in 1940 when they are much older than their children. While this is by necessity (nationally representative data on educational attainment are not available earlier in the 20th century), this later age could bias our findings because education tends to increase within a given cohort over time. To test this hypothesis, we use the 1915 Iowa Census which asked about educational attainment 25 years before the 1940 Census. Feigenbaum (2018) studied intergenerational mobility in terms of income, occupation, and education by observing fathers' outcomes in the 1915 Census and obtaining their sons' information by linking them to the 1940 Census. We augmented Feigenbaum's linked sample by additionally linking Iowa fathers to the 1940 Census to observe their educational attainment at a later time. As expected, the sample shows a modest amount of educational creep, with fathers reporting 0.13 more years of education on average in 1940 than in 1915. However, using father's education or educational rank in 1915 or 1940 has a negligible effect on our estimates (Appendix Table A5). Although these results are specific to Iowa (where educational mobility was higher in the early 20th century), they provide the best evidence available that this source of mismeasurement is not driving our early 20th century results.

[6] Appendix Figure A7.A shows that the rank-rank slope initially declined, then changed little for several decades, before declining again more recently. The path for the level-level slope is broadly similar. These estimates are consistent with Ferrie et al. (2021), who report a level-level slope of 0.37 for the 1922–1940 cohorts and 0.36 for the 1955–1990 cohorts using linked census-survey data. Figure A7.B shows that the expected rank at the 25th percentile has changed little over the 20th century.

The second test examines the sensitivity of our findings to using different measures of parents' education. The data show that using the educational attainment of the most educated parent, the least educated parent, or the mean education of the two parents (when present) has modest effects on the rank-rank slope and upward mobility at the 25th percentile (Appendix Table A6).[7] However, these different measures affect both parameters similarly in the early and later 20th century, meaning that our conclusions about changes in educational mobility over time are robust to these alternatives.

Having benchmarked our sample to nationally representative surveys and established the robustness of our measures to a variety of tests, we conclude that both relative and upward educational mobility rose for American children of parents with below median educational attainment over the 20th century. Yet these changes may seem small, given the significant changes to the U.S. educational system, sizable increases in public resources for education, and the large rise in years of schooling over the same period. Because the rank-rank slopes can be decomposed into within-group and between-group components, the observation-weighted average of the subgroup slopes need not equal the overall slope (Greene 2008; Jácome et al. 2025). We next explore heterogeneity in these patterns across different race-sex subgroups and across places.

B. Educational Mobility in the U.S. over the 20th Century, by Sex and Race

Figure 2 summarizes estimates by sex and race/ethnicity in graphical form; Appendix Tables A7–A9 present these results in numerical form, including results for the two alternative educational mobility measures in Table 1, standard errors, statistical tests of changes over time, and comparisons to survey data. All changes discussed in this section are statistically significant unless otherwise noted.

For relative mobility, we find that little has changed among White Americans. Over 75 years, the rank-rank slope fell slightly by 0.02 (from 0.45 to 0.43) for White boys and 0.03 (from 0.44 to 0.41) for White girls. On the other hand, relative mobility has improved measurably among more historically

[7] This finding is closely tied to strong assortative matching on education in the early and later 20th century. In the early 20th century, 41 percent of mothers and fathers have the same years of completed education; 57 percent differ at most by one year, and almost three quarters by no more than two years. The corresponding numbers for late 20th century cohorts are similar.

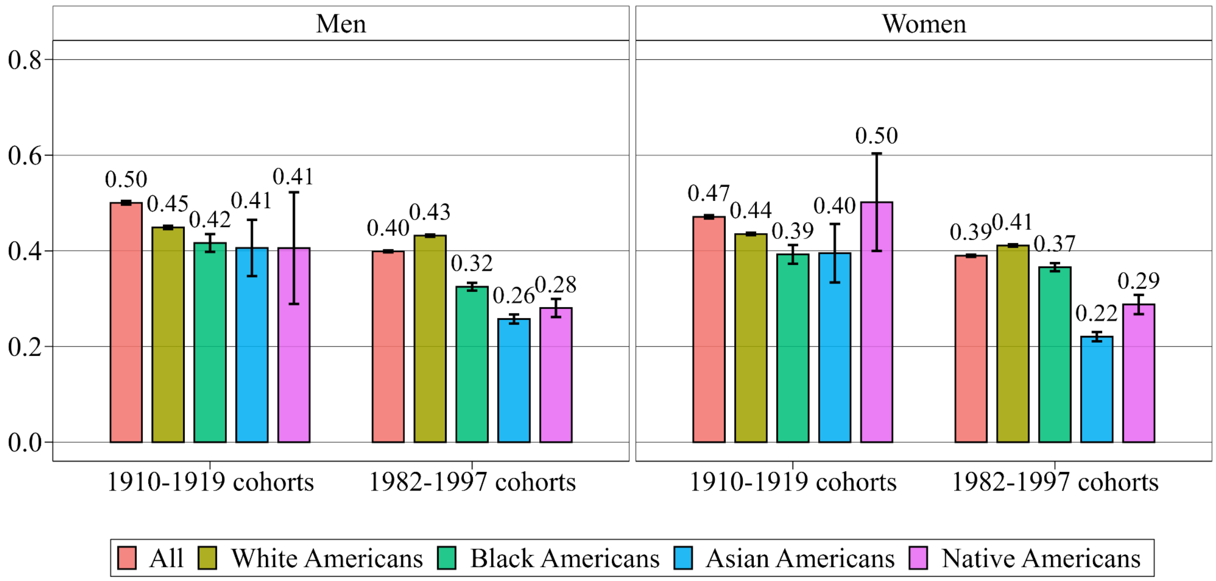
disadvantaged groups: the rank-rank slope fell significantly over the 20th century for Black boys, by 0.09; for Asian boys and girls, by 0.15 and 0.18, respectively; and Native American boys and girls, by 0.13 and 0.21, although only slightly for Black girls, by 0.03.

Changes in upward educational mobility over the 20th century reveal similar patterns, showing declining race/ethnicity gaps and a widening gender gap. Especially notable is that, in the early 20th century, Black boys from disadvantaged parental educational backgrounds fared even worse than their parents on average: the expected educational rank of Black boys with parents at the 25th percentile was the 18th percentile. Black girls were not upwardly mobile either, with their expected educational rank matching their most educated parent. This does not mean that Black children did not exceed their parents' *years* of education: around 50 and 63 percent of Black boys and girls born in 1910–1919 obtained more education than their parents, respectively. The findings imply that this increase in educational attainment was not enough to increase the relative rank of Black children relative to other American children.

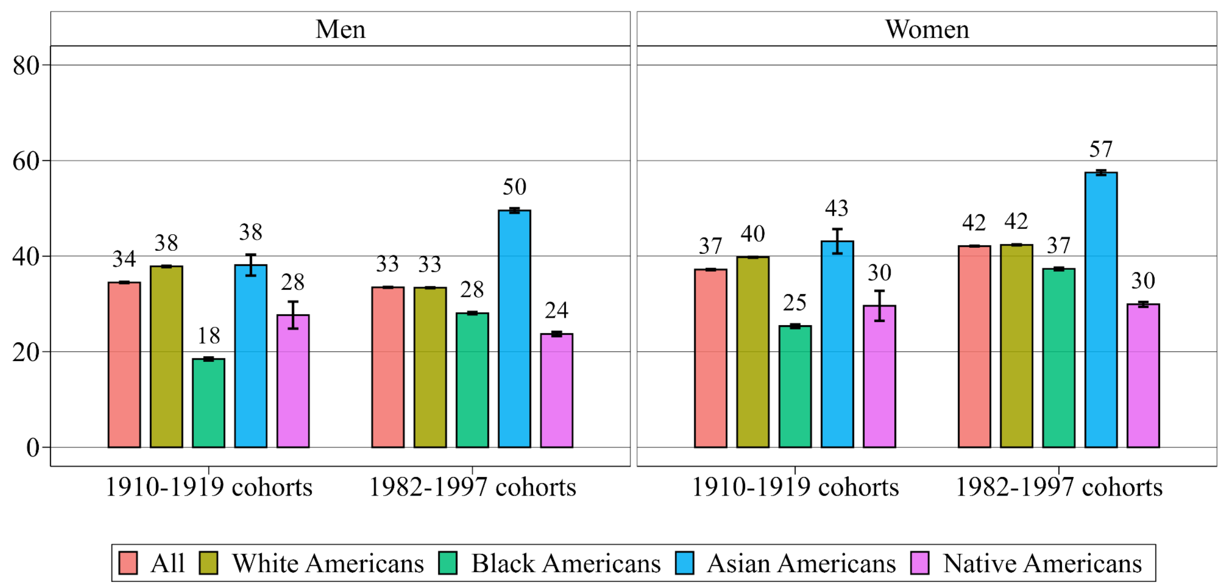
This finding aligns with patterns of change in resources for Black schools, which worsened around the turn of the century before improving in the 1940s (Margo 1990). Over the 20th century, the upward educational mobility of Black Americans improved. By the later 20th century, the expected educational rank at the 25th percentile rose to the 28th percentile for Black boys and to the 37th percentile ranks for Black girls. In contrast, upward mobility among White men *fell* by 5 percentile ranks and nudged up by 2 percentile ranks among White women. Consequently, the Black-White gap in upward educational mobility narrowed sharply, from 20 to 5 percentile ranks for men and from 15 to 5 points among women. Pooling across men and women, the Black-White gap in upward educational mobility declined by 12 percentile ranks over the 20th century (Appendix Table A7). In contrast, Collins and Wanamaker (2022) find that the gap in upward *income* mobility between White and Black men remained persistent for Americans born between 1900 and 1990, highlighting the fact that increasing educational mobility need not translate into income mobility in the presence of labor-market discrimination (Mohammed and Mohnen 2025).

Figure 2. Rank-Rank Slope and Expected Rank at 25th Percentile Estimates by Cohort and Sex or Sex-by-Race/Ethnicity Group

Panel A. Rank-Rank Slope



Panel B. Expected Rank at 25th Percentile



Notes: Figures plot the rank-rank slope and expected rank of children with parents at the 25th percentile using linked samples by cohort and sex or sex-by-race/ethnicity group. 95-percent confidence intervals are based on heteroskedasticity robust standard errors (calculated using the Delta method for expected rank at the 25th percentile). Changes across cohorts within the rank-rank slope and expected rank differ from zero at the 1 percent or 5 percent level (except for Native American girls in Panel B). Appendix Tables A7–A9 present these estimates in tabular form, including standard errors and sample sizes, and Appendix Table A7 compares the estimates by cohort and sex to analogous estimates based on survey data.

A few other patterns in the data stand out. First, while Asian Americans had upward educational mobility rates that were comparable to those of White Americans in the early 20th century, they now enjoy the highest upward mobility of any group by a large margin (Chetty et al. (2020) finds similar results for income mobility). Today, Asian American boys whose most educated parent is at the 25th percentile are expected to reach the 50th percentile and Asian American girls the 57th percentile. Over the 20th century, the upward mobility gap between Asian and White Americans has grown by 17 and 15 percentile ranks for boys and girls, respectively. In contrast, Native Americans had the lowest upward mobility rates of any group in the late 20th century: among children whose most educated parent was at the 25th percentile, Native American boys today appear slightly *downwardly* mobile, with their expected percentile rank at 24 and girls at 30. Earlier in the century, Black boys and girls had the lowest upward mobility within their respective groups, but by the late 20th century Native American boys and girls exhibited the lowest levels of upward educational mobility. In fact, every gender-by-race/ethnicity group has experienced an improvement in upward educational mobility over the 20th century with the exception of Native American and White boys, whose expected rank mobility has fallen by 4 and 5 percentile ranks, respectively. The decline among White men, who make up a large share of the population, has offset the gains among all other groups, resulting in the modest rise we documented above.

Lastly, the gender gap in upward educational mobility, which only slightly favored girls in the early 20th century, grew in favor of girls over the next 75 years. Among White Americans, the gender gap (girls–boys) was around 2 percentile ranks in the early 20th century while today it stands at 9 percentile ranks. Among Black Americans, the initial gender gap was larger at 7 percentile ranks, but it has grown to 9 percentile ranks today. A similar story has unfolded among Asian Americans, with the gender gap rising to 8 percentile ranks, and for Native Americans, to 6 percentile ranks. Because White boys make up a disproportionate share of the male population, declining upward mobility for this group translated into a decline in upward mobility for boys over the 20th century, from 34 to 33 percentile ranks. Together with the 5 percentile rank increase in upward educational mobility among girls, the gender gap expanded from

3 percentile ranks in the early 20th century to 9 percentile ranks at its close—mirroring the growing unconditional gender gap in college attainment (Goldin et al. 2006).

IV. The Role of Public Education Spending in Shaping Intergenerational Educational Mobility over the 20th Century

Rising upward mobility among the more disadvantaged classes and groups could imply that increases in public educational investments lessened the influence of parents on children's educational attainment and created new opportunities for those without good substitutes for public schools (Althoff et al. 2025). But these findings are also consistent with other stories. For example, the Great Migration's relocation of Black Americans out of the South, where school resources were both lower on average and Black schools were comparatively under-resourced, likely improved the educational prospects of Black children (Margo 1990). This story would not be about the causal role of increased educational spending on educational opportunity but about the relocation of Black children to areas with better schools. Similarly, the increasing demand for educated women, especially for clerical and teaching positions, would tend to increase girls' demand for education and their educational mobility (Goldin 1990; Blau and Kahn 2017), even if increasing public school resources had no causal role.

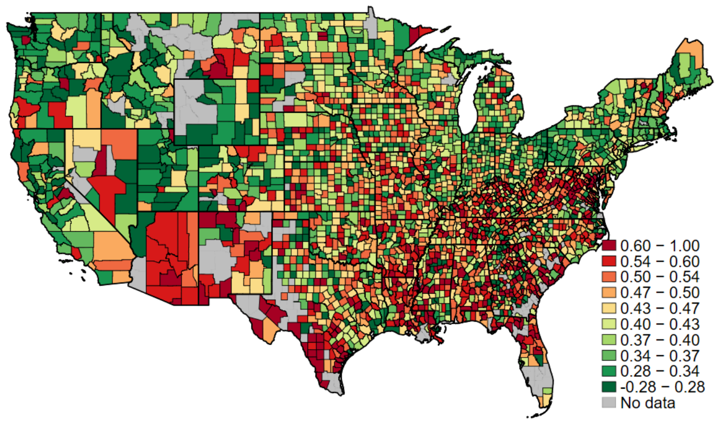
This section investigates the causes of changing educational mobility by first describing its variation across counties of childhood residence in the early and later 20th century. Then, we use a state-border research design to help isolate the causal effect of differences in public education resources—caused by different state policies—on educational mobility.

A. Geographic Variation in Relative and Upward Educational Mobility

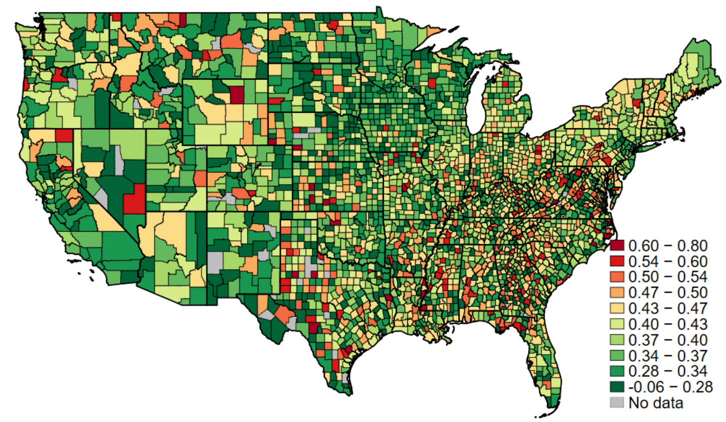
Figure 3 displays the county-level estimates of relative and upward educational mobility for the early and later 20th century cohorts. Because boundaries have changed over time, the maps restrict the sample to counties with stable boundaries, which we define as counties having 95 percent of their geographic area under 1920 boundaries overlap with their area under 2000 boundaries and vice versa, using the crosswalk from Eckert et al. (2020). We also restrict the sample to counties with at least 20 parent-child observations; counties not meeting this threshold are shaded in grey.

Figure 3. The Geography of Educational Mobility by Childhood County of Residence over the 20th Century

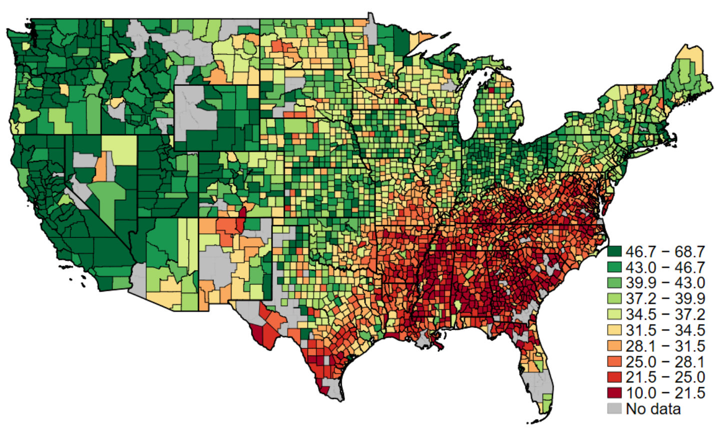
Panel A. Rank-Rank Slope, 1910–1919 Cohorts



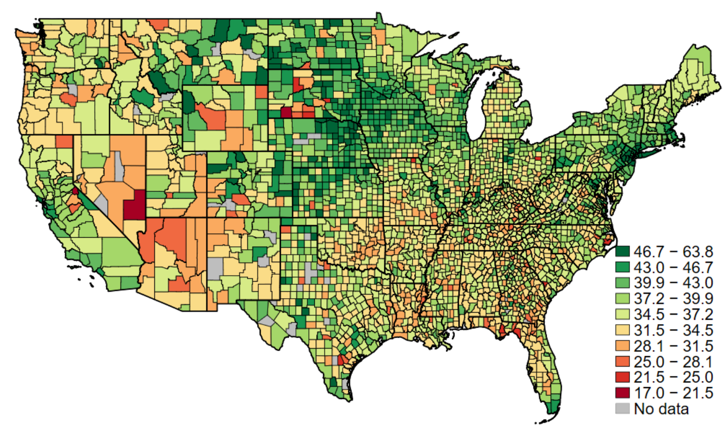
Panel B. Rank-Rank Slope, 1982–1997 Cohorts



Panel C. Expected Rank at 25th Percentile, 1910–1919 Cohorts



Panel D. Expected Rank at 25th Percentile, 1982–1997 Cohorts



Notes: Figures plot educational mobility by children's county of residence in 1920 for the 1910–1919 cohort and by county of residence in 2000 for the 1982–1997 cohort for counties with stable boundaries between 1920 and 2000 (see text for definition) and at least 20 parent-child observations. In all panels, counties are shaded according to a common set of decile bin cutoffs based on the 1910–1919 cohorts.

Disclosing county-level estimates for the 1982–1997 cohorts from the FSRDC required following Census Bureau guidelines to protect the confidentiality of respondents. Following Chetty and Friedman (2019), we infused unweighted county estimates and associated standard errors for the 1982–1997 cohorts with additive Gaussian noise and set the "privacy parameter" to 8. This process results in less than 5 percent of the total variation in the adjusted educational mobility estimates occurring due to noise.[8]

Consistent with research documenting wide geographic differences in income or occupational mobility in the U.S. historically (Connor and Storper 2020; Massey and Rothbaum 2021; Tan 2023), Figures 3.A and 3.C show large geographic disparities in educational mobility in the early 20th century.[9] Our findings contribute new estimates of the geographic distribution of relative educational mobility in the early and late 20th century and also document changes in educational mobility in different areas of the U.S. within a common measurement framework. Relative mobility—measured by the inverse of the rank-rank slope—tended to be lower in the South in the early 20th century and higher in the Northeast, the Midwest, and especially the West.[10] But considerable variation in relative mobility is evident within states, with every state containing counties with some of the lowest and highest rates of relative mobility.

Figure 4 shows that, in the early 20th century, higher rates of relative mobility are generally associated with higher economic development (e.g., share urban, manufacturing share); higher social capital (as measured by the share of religious adherents); lower income inequality; and a smaller share of Black Americans and immigrants. By the later 20th century, however, relative mobility had improved for children growing up in almost all counties in the U.S., and the associated correlations are similar in sign but generally smaller in magnitude.

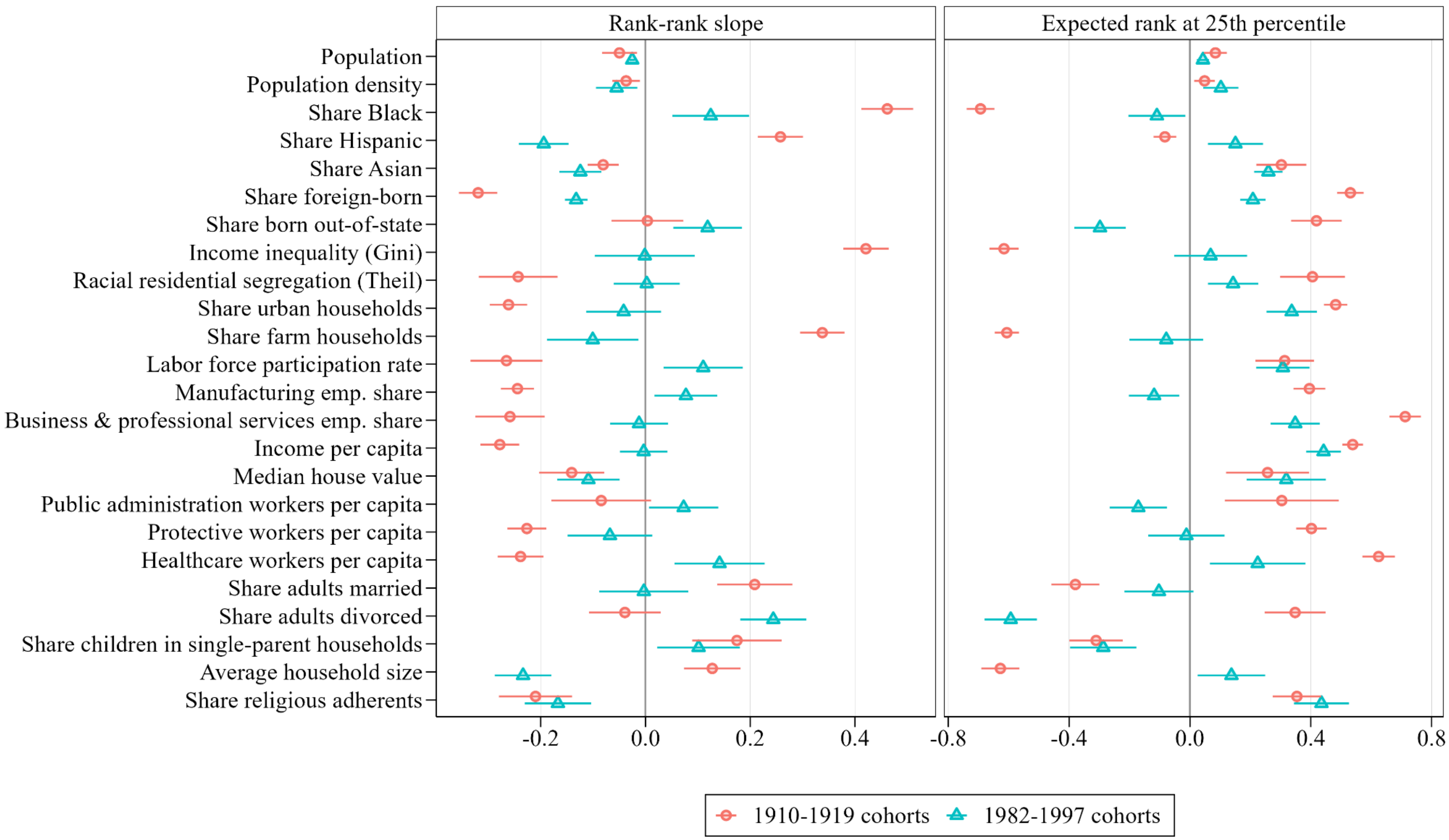
---

[8] County observation counts were separately infused with noise using the Census Bureau's implementation of the Discrete Gaussian Mechanism (Cannone et al. 2022).

[9] Connor and Storper (2020) focus on men born in 1900–1915 linked across the 1920 and 1940 Censuses, while Tan (2023) focuses on White men born in 1892–1910 linked across the 1910 and 1940 Censuses. Both assign occupation-based income scores to both fathers and sons. Massey and Rothbaum (2021) focus on men and women born in 1922–1940 linked across the 1940 Census and tax records from 1974–1979. Fathers' income is based on self-reported wage income in 1940 while children's income is based on Adjusted Gross Income reported in tax data.

[10] The rank-rank slope is negative in a handful of less populous counties, mainly in the early 20th century. To minimize the influence of noise in driving our estimates, we weight our subsequent regression analysis by the inverse of the estimated variance to reduce the influence of small counties and also use alternative measures that shrink the estimates toward the state mean using empirical Bayes methods.

Figure 4. Correlations Between County Educational Mobility Estimates and County Characteristics



Notes: Each dot is the coefficient from a regression of county rank-rank slope or county expected rank at the 25th percentile on a county characteristic (indicated in row label), separately by cohort. Counties are weighted by the inverse of the variance of the relevant county mobility estimates. All variables are normalized to have a mean of 0 and standard deviation of 1 within each cohort. 95-percent confidence intervals are based on robust standard errors. Early 20th century county characteristics are generated from the 1920, 1930 and 1940 full-count Censuses (Ruggles et al. 2024). Modern-day analogs are constructed using 2000 Census data from the National Historical Geographic Information System (NHGIS; Manson et al. 2024) or drawn from Opportunity Insights. See Appendix C for details.

These varied within-state patterns are less evident for upward educational mobility in Figures 3.C and 3.D, as measured by the expected rank at the 25th percentile. In the early 20th century, upward mobility was uniformly lower in the South, with the former slave states almost completely coincident with the lowest expected percentile ranks (shading of orange and red). (Only pockets of northern Missouri, western Texas, and southern Florida showed some upward educational mobility). The lightest shade of orange in these areas indicates little to no upward mobility (an expected percentile rank of 25–28 for a child of a parent at the 25th percentile), and the darkest red indicates downward mobility (an expected percentile rank of 10–22 for a child of a parent at the 25th percentile). That is, outside of a handful of counties, children in former slave states saw *little to no* upward educational mobility in the early 20th century. In contrast, counties in the West, Northeast, and Industrial Midwest had very high upward educational mobility rates in this period as indicated by the light and dark green shading. Historically, opportunities for upward educational mobility depended a great deal on where children grew up. These broad findings are consistent with those of Card et al. (2022), who examine an unlinked sample of children born in the 1920s who resided with their parents in the 1940 Census. The correlation between our measure of upward mobility at the 25th percentile and theirs is 0.77.

By the end of the 20th century, the geography of upward educational mobility had changed, such that the correlation of upward mobility in the early and later 20th century is only 0.25. This sharp regional convergence shows up as less varied shading in Figure 3.D. Over this period, upward mobility improved in almost every county in the South, with the expected percentile rank averaging in the mid-30s for former slave states. This is not just a story of the South catching up. High rates of upward educational mobility in the West, Northeast, and Industrial Midwest fell, currently hovering close to the national average. Similarly, very high rates of upward educational mobility deteriorated in the Industrial Midwest, mirroring its declining economic fortunes. Along with the South, counties in the Great Plains (Iowa, Kansas, Minnesota, Nebraska, and the Dakotas) experienced significant gains in upward educational mobility.[11] As a result, the interquartile range in upward mobility collapsed, from 15 percentiles to 6 percentiles, and cross-county variance declined by 76 percent over the 75 year period (Appendix Figure A9). This pattern of regional convergence also holds for the rank-rank slope and bottom 40 to top 60 percent mobility, with the cross-county variances respectively declining by 50 and 74 percent (Appendix Figures A10 and A11). This geographic convergence means that the place where a child grows up today matters much less for their educational mobility than in the past.

B. The Effect of Public Resources for Education on Educational Mobility

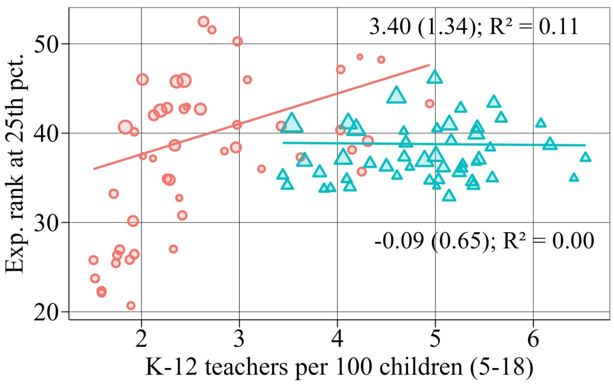
Our final analysis considers whether public resources for education, often called "educational quality" (Card and Krueger 1992), played a causal role in shaping the evolution of educational mobility across time and place. We use three measures of public resources for education. The first two, (1) the number of K-12 teachers per 100 children aged 5 to 18 (Card and Krueger 1992, 1996) and (2) average salaries of K-12 teachers (Card and Krueger 1992, 1996; Card et al. 2022), have been used extensively in the literature.[12]

We also add a third measure which captures public resources for post-secondary education: (3) the number of instructors at post-secondary public colleges and universities per 100 children aged 5 to 18 (Goldin and Katz 2008). This measure includes both two-year community colleges as well as four-year colleges, which may have different effects on educational mobility (Mountjoy 2022). Over the 20th century, publicly funded post-secondary institutions have become an increasingly important part of the U.S. education system. In 1931, enrollment in public post-secondary institutions accounted for 50 percent of total enrollment. By 2021, this number had risen to 73 percent of total enrollment (Goldin 2006; U.S. Department of Education 2023a). In addition, these public institutions are more likely to serve students from more disadvantaged backgrounds.[13]

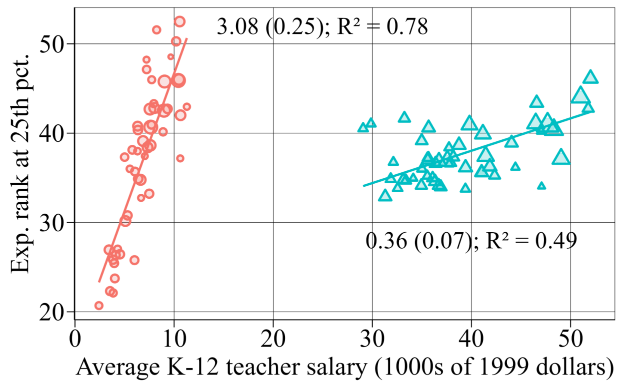
[11] Chetty et al. (2014b) also document that upward income mobility is highest in the Great Plains and lowest in the Southeast in the later 20th century.
[12] Note that kindergarten was not widely publicly supported in the U.S. until the 1970s, so kindergarten teachers only factor into these measures for late 20th century cohorts (Cascio 2009).
[13] In 2022–2023, the average annual tuition and required fees at 4-year public institutions was $9,834 (minimum of in-district and in-state tuition), compared to $40,713 at 4-year private non-profit institutions and $18,241 at 4-year private for-profit institutions. The corresponding figures for 2-year institutions are $4,027, $19,517 and $16,301, respectively (U.S. Department of Education 2023c).

Figure 5. The Relationship Between Expected Rank at 25th Percentile at the State Level vs. State Resources for Public Education Resources

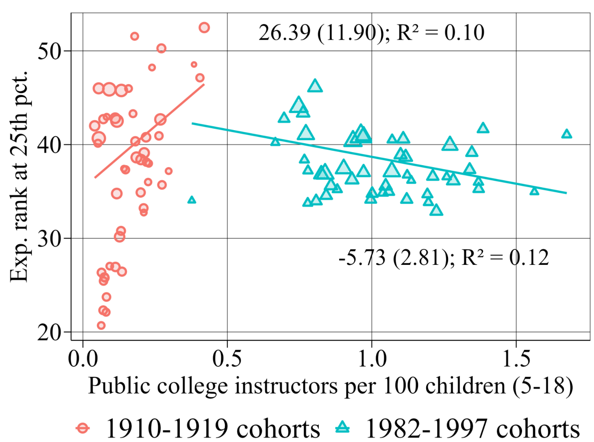
Panel A. K-12 Teachers per 100 Children (5–18)



Panel B. Average K-12 Teacher Salary



Panel C. Public College Instructors per 100 Children (5–18)



Notes: These graphs plot expected rank at the 25th percentile by childhood state of residence for the 1910–1919 and 1982–1997 cohorts against contemporaneous state-level resources for public education. Fitted lines are based on linear regressions where observations are weighted by the number of children in the relevant linked sample (estimated slope coefficients, associated robust standard errors, and regression R^2 also shown). State-level statistics on K-12 education for school year 1919–1920 and statistics on public colleges and universities for school years 1928–1930, which are aggregated at the state level, come from biennial reports compiled by the Bureau or Office of Education (U.S. Bureau of Education 1923; U.S. Office of Education 1931). Analogous state-level statistics on K-12 education for school year 1999–2000 and public post-secondary institutions for school year 2000–2001 are constructed using the Common Core of Data (CCD) and Integrated Postsecondary Education Data System (IPEDS) databases from the National Center for Education Statistics (NCES), or are directly taken from NCES tables (U.S. Department of Education 2022). See Appendix C for details.

We acknowledge these three measures of public educational investment are an incomplete characterization of the broader set of educational policies in both periods. They have the benefit of being well measured and comparable in both periods, and they generally exhibit a strong positive association with upward educational mobility (Figure 5). We, therefore, interpret these measures as consistent proxies of *bundles* of state policies to promote the education of their youth—including one bundle aimed at K-12 education and one bundle expanding access to higher education.

On the other hand, this strong positive relationship between the education measures and educational mobility may reflect omitted variables bias that could threaten the validity of cross-sectional analyses. Although regional convergence in educational mobility has coincided with regional convergence in educational resources (measured by the coefficient of variation or the standard deviation normalized by the mean) (Appendix Figure A12), this relationship could reflect a multitude of unobserved factors, such as the county characteristics in Figure 4.

To narrow the scope for omitted variables bias, we follow the literature leveraging state borders to isolate the role of differing education policies in otherwise highly similar areas (Black 1999; Kane et al. 2006; Knight and Schiff 2019; Card et al. 2022). Consistent with state policy being important in explaining upward educational mobility, we find that two thirds of the cross-county variation in upward educational mobility in the historical period and one third in the contemporary period is explained by state-level factors (i.e., state fixed effects). Similar to Card et al. (2022), our analysis compares county pairs on opposite sides of a state border within the following linear regression framework,

$$y_{cp} = \boldsymbol{Z}_{s(c)}'\theta + \delta_p + \varepsilon_{cp}, \tag{3}$$

where y denotes educational mobility in county c located in state s and part of county pair p, and δ_p is a set of county-pair fixed effects. We assign county pairs using the Census Bureau's 2010 county adjacency crosswalk (U.S. Census Bureau 2010) and use only counties with stable boundaries and at least 20 parent-child pairs.[14] The coefficients of interest, θ, capture the reduced-form effects of state policy via

[14] Appendix Figure A13 presents maps of the counties used in the analysis and provides a visual description of the identifying variation in K-12 teacher salaries across state borders, which also holds for the number of K-12 teachers

three state measures of educational spending, $\boldsymbol{Z}$: the number of K-12 teachers per 100 children aged 5 to 18, average salaries of K-12 teachers, and the number of instructors at post-secondary public colleges and universities per 100 children aged 5 to 18.

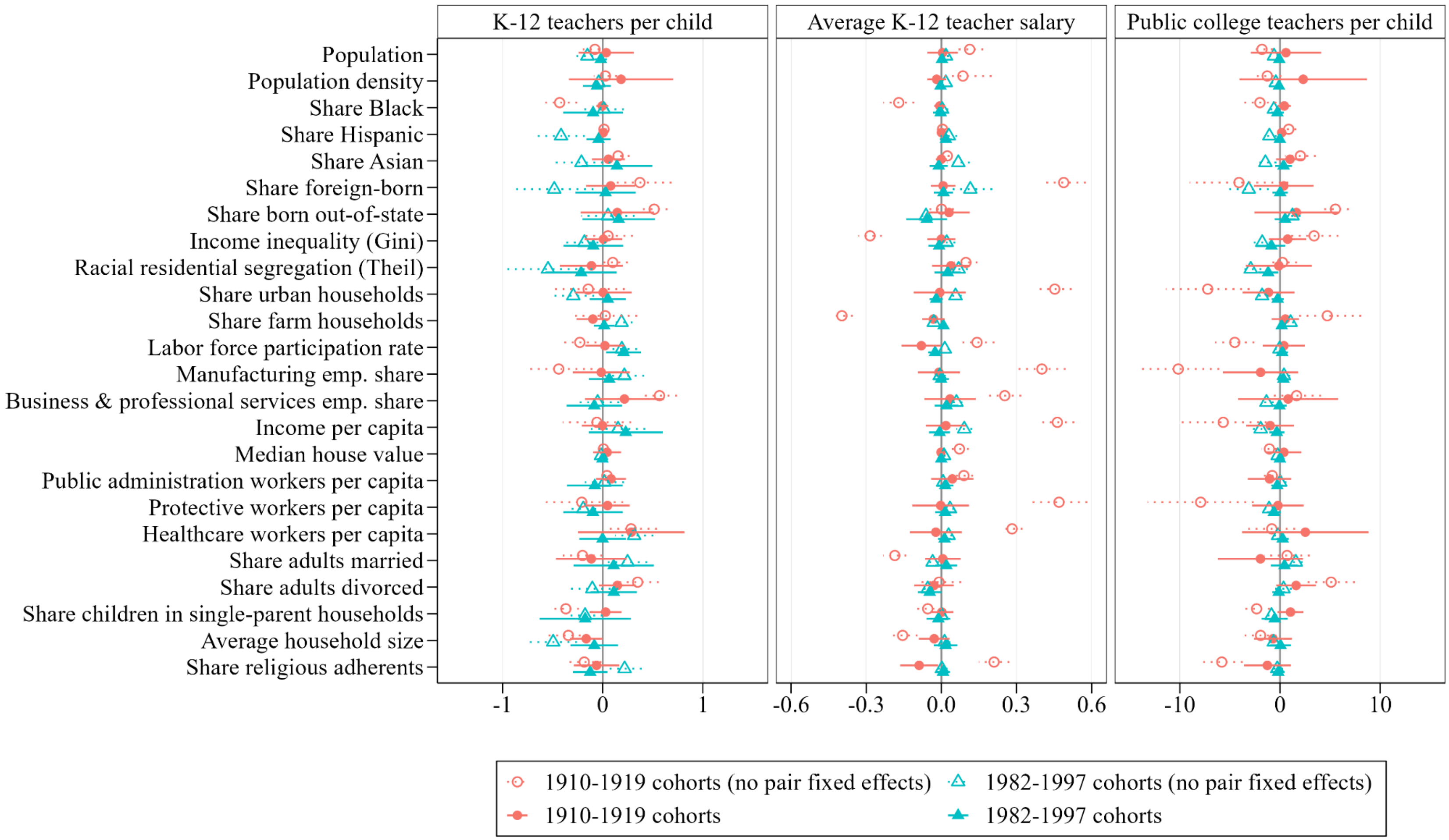
A key identifying assumption is that, within county pairs, the only reason for educational resources differ across state borders is due to differences in state educational policies and not omitted county-level factors, or $\text{cov}(\varepsilon_{cp}, \boldsymbol{Z}_{s(c)} | \delta_p)=0$. We test this assumption by examining whether the rich set of observed county characteristics in Figure 4, $\boldsymbol{X}_c$, are uncorrelated with differences in educational resources within county pairs, or $\text{cov}(\boldsymbol{X}_{cp}, \boldsymbol{Z}_{s(c)} | \delta_p)=0$. In both the early and later 20th century, Figure 6 shows that while county characteristics are highly associated with differences in state educational resources (hollow markers), these correlations are much smaller and statistically indistinguishable from zero after including county pair fixed effects (solid markers). Out of 144 coefficients (24 characteristics × 3 measures of public education resources × 2 time periods), only five are statistically different from zero at the 5 percent level—no more than expected by chance. Although we cannot rule out unobserved factors that affect our results, the fact that adjacent counties on opposite sides of state borders are highly balanced in a wide set of observed county characteristics supports the validity of our research design.[15]

---

and college instructors. Note that counties can be part of multiple county pairs, because they may adjoin more than one county on the other side of the state border.

[15] We also test whether non-education state policies vary across state borders and threaten our interpretation of the results. This is difficult to test, but we assembled a database of candidate policies to provide some evidence on this issue, including women's suffrage laws (Lott and Kenny 1999; Miller 2008), state minimum wage laws in the early 20th century (mostly targeting women's wages) and later 20th century(Marchingiglio and Poyker 2024; Federal Reserve Bank of St. Louis 2024), and compulsory schooling laws (Acemoglu and Angrist 2000; Lleras-Muney 2005; Stephens and Yang 2014). Using the same specification as in (3), we find no evidence that greater K-12 or post-secondary public investments in education predicts having a state-level women's suffrage law prior to 1919 or having a higher minimum wage in either period. This evidence helps assuage concerns that differences in public health investments and child health (Miller 2008) or labor demand are responsible for the relationship between investments in public education and upward educational mobility. However, we find that greater investments in K-12 education systematically predict having a compulsory schooling law. This finding is consistent with states using *bundles* of policies to promote the education of their youth. In what follows, we therefore interpret the estimates as the effect of these policy *bundles* (including the effects of correlated but unmeasured state education policies, such as compulsory education laws) rather than the causal effect of the proxies used for these bundles.

Figure 6. Balance Checks for the State-Border Design



Notes: Each marker (dot/triangle) is the coefficient from a regression of a county characteristic (indicated in row label) on a state public education measure indicated in the panel title. Hollow markers are from regressions that exclude county pair fixed effects, whereas solid markers include county pair fixed effects. Counties are weighted by the inverse of the variance of the county estimates of expected rank at the 25th percentile, which is the estimand targeted in the analysis. All county characteristics are normalized to have a mean of 0 and standard deviation of 1 within each cohort. 95-percent confidence intervals are based on robust standard errors, clustered at the state border level.

Table 2. Impact of State Public Education Resources on Upward Educational Mobility

	Dependent variable: Expected rank at 25th percentile				
	(1)	(2)	(3)	(4)	(5)
Panel A. 1910–1919 cohorts					
K-12 teachers per 100 children (ages 5–18)	2.16 (0.33)	2.68 (0.65)	1.20 (0.72)	1.49 (0.69)	0.80 (0.47)
Average K-12 teacher salary (1,000s of 1999 dollars)	3.20 (0.12)	2.97 (0.17)	0.58 (0.24)	0.91 (0.25)	0.89 (0.19)
Public college instructors per 100 children (ages 5–18)	12.91 (4.53)	9.77 (6.76)	-6.48 (11.11)	-4.64 (10.39)	-2.86 (5.85)
R^2	0.655	0.624	0.925	0.940	0.969
Partial R^2 (K-12)	0.629	0.621	0.039	0.106	0.140
Partial R^2 (post-secondary)	0.017	0.006	0.004	0.002	0.001
Panel B. 1982–1997 cohorts					
K-12 teachers per 100 children (ages 5–18)	1.17 (0.36)	2.36 (0.46)	0.67 (0.64)	0.71 (0.60)	0.18 (0.29)
Average K-12 teacher salary (1,000s of 1999 dollars)	0.41 (0.04)	0.50 (0.08)	0.19 (0.08)	0.23 (0.08)	0.12 (0.04)
Public college instructors per 100 children (ages 5–18)	-0.05 (0.90)	1.36 (1.49)	3.58 (1.72)	3.15 (1.72)	2.07 (0.94)
R^2	0.236	0.325	0.852	0.876	0.947
Partial R^2 (K-12)	0.206	0.288	0.040	0.072	0.034
Partial R^2 (post-secondary)	0.000	0.003	0.037	0.039	0.030
Border sample		✓	✓	✓	✓
County pair fixed effects			✓	✓	✓
Bayesian shrinked estimates				✓	✓
County controls					✓
N	2,881	2,344	2,344	2,344	2,344

Notes: Column 1 shows the estimates from a regression of county-level expected rank at the 25th percentile on state resources for public education. Column 2 estimates the same specification as in column 1 but only for the border-design sample (units of observation are counties-by-border pair). Columns 3–5 show the estimates from our state border regressions (equation 1). Columns 4 and 5 replace the baseline county mobility estimates with Empirical Bayes shrinked versions of those estimates (see text for details). Column 5 additionally includes controls for all county characteristics shown in Figure 4. In all regressions, counties are weighted by the inverse of the variance of the county mobility estimates. Robust standard errors in parentheses. Columns 2-5 additionally cluster at the state border level in columns 2–5.

Table 2 shows the relationship of school quality measures with upward educational mobility. In the first two columns, we start by estimating equation (3) with and without county-pair fixed effects. This reveals that measures of school quality in border counties (column 2) exhibit a similar relationship with upward mobility in the full sample of counties (column 1) in both periods, suggesting the external validity of the findings using the border counties only. After including county-pair fixed effects (column 3), larger investments in K-12 education remain highly related to upward educational mobility in the early 20th century. The estimates imply that a 10-percent increase in the number of K-12 teachers per 100 children relative to the national mean led to a 0.3 increase in the expected rank of children of parents at the 25th percentile, whereas a 10-percent increase in the average K-12 teacher salary relative to the national mean led to a 0.4-percentile increase in upward mobility.[16] The finding that state average teacher salaries are strong predictors of upward educational mobility in the historical period aligns with Card et al. (2022), who document this for slightly younger cohorts.

In the late 20th century, K-12 teachers per child had no significant effect on upward mobility, but a 10-percent increase in K-12 teacher salaries (relative to the national mean of $41,666) increased the educational percentile rank of children by 0.8. Our estimates imply that a 10-percent increase in public college instructors per 100 children (relative to the national mean of 0.96 instructors) raised upward educational mobility by around 0.35 percentile ranks. These findings are more precisely estimated when using Bayesian shrinkage to downweight sampling noise in our county mobility estimates (Armstrong et al. 2022, column 4) and even more so when adding a broad set of county-level controls (column 5).[17] The shift in importance from K-12 to post-secondary public education can be seen in the partial R^2. In the early and later 20th century, K-12 investments explained around 4 percent of the variation in upward educational mobility (column 3). However, the explanatory power of public investments in

[16] The average number of K-12 teachers per 100 children was 2.3 in 1920 (Appendix Table A11). Therefore, a 10-percent increase corresponds to a 0.23 increase, which multiplied by 1.2 yields a 0.3 increase in the expected rank at the 25th percentile. Similarly, the average salary of K-12 teachers was $7,194 in 1920, so a 10-percent increase of 0.719 (in thousands of dollars) multiplied by 0.58 yields a 0.4 increase in expected rank.

[17] We follow the methodology of Armstrong et al. (2022) to construct empirical Bayes confidence intervals without requiring assumptions about the exact distribution. In our case, we shrink county-level educational mobility estimates toward the state mean. Appendix Figure A14 compares the resulting county estimates after shrinkage to the original county estimates.

post-secondary education was very close to zero in the early 20th century but grew to almost 4 percent in the later 20th century. Appendix Tables A12 and A13 reveal qualitatively similar patterns if we look at impacts on the rank-rank slope or bottom 40 to top 60 percent mobility.

V. The Evolution of Educational Opportunity and the Role of Public Funding for Schools

The rapid expansion of universal public education in the U.S. defined the human capital century, setting the stage for American economic leadership in the 20th century (Goldin 2001; Goldin and Katz 2008). This paper provides novel evidence on how these increasing investments in U.S. public education affected the landscape of educational opportunity over the 20th century. By constructing two new large-scale datasets, we document that both relative and upward educational mobility and opportunity increased over the 75 year period we study, with some of largest gains in educational mobility accruing to the most historically disadvantaged socio-economic classes, race and gender groups, and geographic areas. The geographic convergence in educational mobility over the 20th century was astounding: while educational mobility fell in the previously high mobility areas of the West, Northeast, and Industrial Midwest, it rose in historically lower mobility areas of the South and Great Plains. This geographic convergence means that where children live today plays a much smaller role in their educational opportunities than in the early 20th century. Black and Asian Americans experienced significant gains in upward mobility over the 20th century, whereas White Americans did not. The gender gap in upward educational mobility today favors women by 9 percentile ranks, reflecting increasing mobility among women and falling mobility among men.

Our analysis also documents that differences in state investments in public education played an important role in reshaping educational opportunity across the 20th century. Using a Kitagawa-Blinder-Oaxaca decomposition, we find that around half of the convergence between the South and the rest of the U.S. in upward educational mobility across the 20th century is explained by converging resources for public education (Appendix Table A14). Using a state-border design, we show that the number of K-12 teachers per child and average K-12 teacher salaries were important causal determinants of upward educational mobility in the early 20th century. In the late 20th century, teacher salaries continued to play an important role, but the number of instructors at public colleges and universities has become an increasingly important determinant of upward educational mobility. These resources also helped narrow regional gaps in mobility. Just as the high school movement transformed educational opportunity in the early 20th century, access to college education is an increasingly important determinant of educational opportunity today.

Geographic differences in educational mobility share similarities to geographic differences in intergenerational income mobility, both in the early 20th century (Connor and Storper 2020; Massey and Rothbaum 2021; Tan 2023) and the later 20th century (Chetty et al. 2014b). Yet, upward educational and income mobility are not perfectly correlated in either period. For example, the correlation between our county estimates of upward educational mobility for the 1910–1919 cohorts and the county estimates in Massey and Rothbaum (2021) or Tan (2023) for income mobility is around 0.68. For the 1982–1997 cohorts, the correlation of our upward educational mobility estimates with the upward income mobility estimates in Chetty et al. (2014b) is around 0.50, and the correlation between relative educational mobility and relative income mobility is only 0.18. That is, many places in the U.S. are characterized by high rates of educational mobility but low rates of income mobility and vice versa in both periods. These differences reflect differences in the economy and the returns to education, but they also reflect differences in the population studied. Educational mobility can be measured for the near universe of Americans, whereas income mobility can only be measured for those who are employed. In addition, education is a measure of occupational and earnings *potential* rather than their realization, which sometimes reflects personal preferences. In conclusion, the similarities and differences in patterns in educational and income mobility provide important and complementary insights into the evolution of economic opportunity in the United States.

VI. References

Abramitzky, Ran, Leah Boustan, Santiago Peréz, and Elisa Jácome. 2021. "Intergenerational Mobility of Immigrants in the US over Two Centuries." *American Economic Review* 111(2): 580–608.

Abramitzky, Ran, Leah Platt Boustan, Harriet M. Brookes Gray, Katherine Eriksson, Santiago Pérez, Hanna M. Postel, Myera Rashid, and Noah Simon. 2025. “Finding John Smith: Using Extra Information for Historical Record Linkage.” *NBER Working Paper 33999*, National Bureau of Economic Research, Cambridge, MA.

Acemoglu, Daron and Joshua Angrist. 2000. “How Large Are Human-Capital Externalities? Evidence from Compulsory Schooling Laws.” *NBER Macroeconomics Annual, vol. 15*.

Alesina, Alberto, Sebastian Hohmann, Stelios Michalopoulos, and Elias Papaioannou. 2021. “Intergenerational Mobility in Africa.” *Econometrica* 89(1): 1–35.

Althoff, Lukas, Harriet Brookes Gray, and Hugo Reichardt. 2025. “America’s Rise in Human Capital Mobility.” Working paper.

Armstrong, Timothy B., Michal Kolesár, and Mikkel Plagborg‐Møller. 2022. “Robust Empirical Bayes Confidence Intervals.” *Econometrica* 90(6): 2567–2602.

Asher, Sam, Paul Novosad, and Charlie Rafkin. 2024. “Intergenerational Mobility in India: New Measures and Estimates across Time and Social Groups.” *American Economic Journal: Applied Economics* 16(2): 66–98.

Bailey, Martha J., Melanie Guldi, and Brad J. Hershbein. 2014. “Is There a Case for a ‘Second Demographic Transition’: Three distinctive Features of Post-1960 Fertility Decline.” In *Human Capital and History: The American Record,* Edited by Leah Platt Boustan, Carola Frydman, and Robert A. Margo, 273–312. Chicago: University of Chicago Press.

Bailey, Martha J., Connor Cole, and Catherine Massey. 2020a. “Simple Strategies for Improving Inference with Linked Data: A Case Study of the 1850–1930 IPUMS Linked Representative Historical Samples.” *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 53(2), 80–93.

Bailey, Martha J., Connor Cole, Morgan Henderson, and Catherine Massey. 2020b. “How Well Do Automated Linking Methods Perform? Lessons From U.S. Historical Data.” *Journal of Economic Literature* 58(4): 997–1044.

Bailey, Martha J. and Peter Z. Lin. 2025. “Marital Matching and Women’s Intergenerational Mobility in the Late 19th and Early 20th Century US.” In *The Economic History of American Inequality: New Evidence and Perspectives*, edited by Martha J. Bailey, Leah Platt Boustan and William J. Collins, 165–196. Chicago: University of Chicago Press. http://www.nber.org/chapters/c14841

Bailey, Martha J., Peter Z. Lin, A. R. Shariq Mohammed, Paul Mohnen, Jared Murray, Mengying Zhang, and Alexa Prettyman. 2023. “The creation of LIFE-M: The Longitudinal, Intergenerational, Family Electronic Micro-Database project.” *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 56(3), 138–159.

Black, Sandra E. 1999. “Do Better Schools Matter? Parental Valuation of Elementary Education.” *Quarterly Journal of Economics* 114(2): 577–599.

Black, Sandra E. and Paul J. Devereux. 2011. “Recent Developments in Intergenerational Mobility.” *Handbook of Labor Economics, vol. 4*, 1487–1541.

Blanden, Jo, Alissa Goodman, Paul Gregg, and Stephen Machin. 2004. “Changes in Intergenerational Mobility in Britain.” In *Generational Income Mobility in North America and Europe*, edited by Miles Corak, 122–146. Cambridge: Cambridge University Press.

Blau, Peter M. and Duncan Otis Dudley. 1967. *The American Occupational Structure*. New York: John Wiley & Sons.

Blau, Francine D. and Lawrence M. Kahn 2017. “The Gender Wage Gap: Extent, Trends, and Explanations.” *Journal of Economic Literature* 55 (3): 789–865.

Bratberg, Espen, Jonathan Davis, Bhashkar Mazumder, Martin Nybom, Daniel D. Schnitzlein, and Kjell Vaage. 2017. “A Comparison of Intergenerational Mobility Curves in Germany, Norway, Sweden, and the US.” *The Scandinavian Journal of Economics* 119(1): 72–101.

Bumpass, Larry L., James A. Sweet, and Vaughn R.A. Call. National Survey of Families and Households, Wave 1: 1987–1988, [United States]. Inter-university Consortium for Political and Social Research [distributor], 2017-08-31. https://doi.org/10.3886/ICPSR06041.v2

Canonne, Clément, Gautam Kamath, and Thomas Steinke. 2022. “The Discrete Gaussian for Differential Privacy.” *Journal of Privacy and Confidentiality* 12(1).

Cascio, Elizabeth U. 2009. “Maternal Labor Supply and the Introduction of Kindergartens into American Public Schools.” *Journal of Human Resources* 44(1): 140–170.

Card, David, Ciprian Domnisoru, and Lowell Taylor. 2022. “The Intergenerational Transmission of Human Capital: Evidence from the Golden Age of Upward Mobility.” *Journal of Labor Economics* 40(S1): S39–S95.

Card, David and Alan B. Krueger. 1992. “School Quality and Black-White Relative Earnings: A Direct Assessment.” *Journal of Political Economy* 100(1): 1–40.

Card, David and Alan B. Krueger. 1996. “School Resources and Student Outcomes: An Overview of the Literature and New Evidence from North and South Carolina.” *Journal of Economic Perspectives* 10(4): 31–50.
Checchi, Daniele, Carlo V. Fiorio, and Marco Leonardi. 2013. “Intergenerational persistence of educational attainment in Italy.” *Economics Letters* 118(1): 229–232.
Chetty, Raj and John N. Friedman. 2019. “A Practical Method to Reduce Privacy Loss When Disclosing Statistics Based on Small Samples.” *AEA Papers and Proceedings* 109: 414–420.
Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. 2014a. “Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility.” *AER: Papers & Proceedings* 104(5): 141–147.
Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014b. “Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States.” *The Quarterly Journal of Economics* 129(4): 1553–1623.
Chetty, Raj, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter. 2020. “Race and Economic Opportunity in the United States: An Intergenerational Perspective.” *The Quarterly Journal of Economics* 135(2): 711–783.
Collins, William J. and Marianne H. Wanamaker. 2022. “African American Intergenerational Economic Mobility since 1880,” *American Economic Journal: Applied Economics* 14(3): 84–117.
Connor, Dylan Shane and Michael Storper. 2020. “The changing geography of social mobility in the United States.” *Proceedings of the National Academy of Sciences* 117(48): 30309–30317.
Couch, Kenneth A. and Thomas A. Dunn. 1997. “Intergenerational Correlations in Labor Market Status: A Comparison of the United States and Germany.” *Journal of Human Resources* 32(1): 210–232.
Dahl, Molly W. and Thomas DeLeire. 2008. “The Association Between Children’s Earnings and Fathers’ Lifetime Earnings: Estimates Using Administrative Data.” Institute for Research on Poverty Discussion Paper No. 1342-08, University of Wisconsin-Madison.
Davern, Michael, Rene Bautista, Jeremy Freese, Pamela Herd, and Stephen L. Morgan. General Social Survey 1972–2024. [Machine-readable data file]. Principal Investigator, Michael Davern; Co-Principal Investigators, Rene Bautista, Jeremy Freese, Pamela Herd, and Stephen L. Morgan. Sponsored by the National Science Foundation. NORC ed. Chicago: NORC, 2024: NORC at the University of Chicago [producer and distributor]. Data accessed from the GSS Data Explorer website at gssdataexplorer.norc.org
Davis, Jonathan M.V. and Bhashkar Mazumder. 2025. “The Decline in Intergenerational Mobility after 1980.” *The Review of Economics and Statistics,* https://doi.org/10.1162/rest_a_01413.
Eckert, Fabian, Andrés Gvirtz, Jack Liang, and Michael Peters. 2020. “A Method to Construct Geographical Crosswalks with an Application to US Counties since 1790.” *NBER Working Paper 26770*, National Bureau of Economic Research, Cambridge, MA.
Espín-Sánchez, José-Antonio, Joseph P. Ferrie, and Christopher Vickers. 2023. “Women and the Econometrics of Family Trees,” *NBER Working Paper 31598*, National Bureau of Economic Research, Cambridge, MA.
Eriksson, Katherine, Gregory T. Niemesh, Myera Rashid, and Jacqueline Craig. 2023. “Marriage and the Intergenerational Mobility of Women: Evidence from Marriage Certificates 1850–1920.” Working paper.
Federal Reserve Bank of St. Louis. 2024. “Federal and State Minimum Wage Rates, Annual.” https://fred.stlouisfed.org/release/tables?eid=243906&rid=387, accessed on December 11, 2025.
Feigenbaum, James J. 2018. “Multiple Measures of Historical Intergenerational Mobility: Iowa 1915 to 1940.” *The Economic Journal* 128(612): F446–F481.
Ferrie, Joseph P. 1996. “A New Sample of Males Linked from the Public Use Microdata Sample of the 1850 U.S. Federal Census of Population to the 1860 U.S. Federal Census Manuscript Schedules.” *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 29(4): 141–156.
Ferrie, Joseph, Catherine Massey, and Jonathan Rothbaum. 2021. “Do Grandparents Matter? Multigenerational Mobility in the United States, 1940-2015.” *Journal of Labor Economics* 39(3): 597–637.
Fetter, Daniel K,, Lee M. Lockwood, and Paul Mohnen. 2025. “Long-Run Intergenerational Effects of Social Security.” Working paper.
Fletcher, Jason and Joel Han. 2019. “Intergenerational Mobility in Education: Variation in Geography and Time.” *Journal of Human Capital* 13(4): 585–634.
Goldin, Claudia. 1990. *Understanding the Gender Gap: An Economic History of American Women.* New York-Oxford: Oxford University Press.
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(2): 345–374.

Goldin, Claudia. 2001. "The Human-Capital Century and American Leadership: Virtues of the Past." *The Journal of Economic History* 61(2): 263–292.
Goldin, Claudia. 2006. "Enrollment in institutions of higher education, by sex, enrollment status, and type of institution: 1869–1995." Table Bc523-536 in *Historical Statistics of the United States, Earliest Times to the Present: Millennial Edition*, edited by Susan B. Carter, Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright. New York: Cambridge University Press.
Goldin, Claudia and Lawrence F. Katz. 2008. *The Race Between Education and Technology*. Cambridge: Harvard University Press.
Goldin, Claudia, Lawrence F. Katz, and Ilyana Kuziemko. 2006. "The Homecoming of American College Women: The Reversal of the College Gender Gap," *Journal of Economic Perspectives* 20(4): 133–156.
Goldstein, Joshua and Casey Breen. 2022. "Berkeley Unified Numident Mortality Database: Public administrative records for individual-level mortality research," *Demographic Research* 47(5): 111–142.
Greene, William H. 2008. *Econometric Analysis, 6th Edition*. New York: Pearson.
Güell, Maia, José V. Rodríguez Mora, and Christopher I. Telmer. 2015. "The Informational Content of Surnames, the Evolution of Intergenerational Mobility, and Assortative Mating." *The Review of Economic Studies* 82(2): 693–735.
Haider, Steven and Gary Solon. 2006. "Life-Cycle Variation in the Association between Current and Lifetime Earnings." *American Economic Review* 96(4): 1308–1320.
Heineck, Guido and Regina T. Riphahn. 2009. "Intergenerational Transmission of Educational Attainment in Germany: The Last Five Decades." *Jahrbücher für Nationalökonomie und Statistik* 229(1): 36–60.
Hertz, Tom, Tamara Jayasundera, Patrizio Piraino, Sibel Selcuk, Nicole Smith, and Alina Verashchagina. 2007. "The Inheritance of Educational Inequality: International Comparisons and Fifty-Year Trends." *The B.E. Journal of Economic Policy & Analysis* 7(2): 1–46.
Hilger, Nathaniel G. 2017. "The Great Escape: Intergenerational Mobility in the United States, 1930–2010." Working paper.
Hout, Michael and Thomas A. DiPrete. 2006. "What have we learned: RC28's contribution to knowledge about social stratification." *Research in Social Stratification and Mobility* 24(1): 1–20.
Hout, Michael and Alexander Janus. 2011. "Educational mobility in America: 1930s–2000s." In *Whither opportunity? Rising Inequality, Schools, and Children's Life Chances*, edited by Greg J. Duncan, and Richard J. Murnane, 165–185. New York: Russell Sage Foundation.
Jácome, Elisa, Ilyana Kuziemko, and Suresh Naidu. 2025. "Mobility for All: Representative Intergenerational Mobility Estimates over the 20th Century." *Journal of Political Economy* 133(1): 306–354.
Kane, Thomas J., Stephanie K. Riegg, and Douglas O. Staiger. 2006. "School Quality, Neighborhoods, and Housing Prices." *American Law and Economics Review* 8(2): 183–212.
Karlson, Kristian B. and Rasmus Landersø. 2025. "The making and unmaking of opportunity: educational mobility in 20th-century Denmark." *The Scandinavian Journal of Economics* 127(1): 178–212.
Knight, Brian and Nathan Schiff. 2019. "The Out-of-State Distortion." *American Economic Journal: Economic Policy* 11(1): 317–350.
Lleras-Muney, Adriana. 2005. "The Relationship between Education and Adult Mortality in the United States." *Review of Economic Studies* 72(1): 189–221.
Long, Bridget Terry. 2013. "Supporting Access to Higher Education." In *Legacies of the War on Poverty*, edited by Martha J. Bailey and Sheldon Danziger, 93–120. New York: Russell Sage Foundation.
Long, Jason and Joseph Ferrie. 2013. "Intergenerational Occupational Mobility in Great Britain and the United States Since 1850." *American Economic Review* 103(4): 1109–1137.
Lott, Jr, John R. and Lawrence W. Kenny. 1999. "Did Women's Suffrage Change the Size and Scope of Government?" *Journal of Political Economy* 107(6): 1163–1198.
Manson, Steven, Jonathan Schroeder, David Van Riper, Katherine Knowles, Tracy Kugler, Finn Roberts, and Steven Ruggles. 2024. IPUMS National Historical Geographic Information System: Version 19.0 [dataset]. Minneapolis, MN: IPUMS. http://doi.org/10.18128/D050.V19.0
Marchingiglio, Riccardo and Michael Poyker. 2024. "The Economics of Gender-Specific Minimum Wage Legislation." *Journal of Labor Economics*, https://www.journals.uchicago.edu/doi/10.1086/733493.
Margo, Robert A. 1986. "Race, Educational Attainment, and the 1940 Census." *Journal of Economic History* 46(1): 189–198.
Margo, Robert A. 1990. *Race and Schooling in the South, 1880–1950: An Economic History*. Chicago: University of Chicago Press.

Massey, Catherine and Jonathan Rothbaum. 2021. "The Geography of Opportunity over Time." Census Bureau Working Paper SEHSD-WP2021-23, September 8, 2021, https://www.census.gov/library/working-papers/2021/demo/SEHSD-WP2021-23.html.

Mazumder, Bhashkar. 2005. "Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data." *The Review of Economics and Statistics* 87(2): 235–255.

Mazumder, Bhashkar. 2016. "Estimating the Intergenerational Elasticity and Rank Association in the U.S.: Overcoming the Current Limitations of Tax Data." *Research in Labor Economics* 43: 83–129.

Mazumder, Bhashkar. 2018. "Intergenerational Mobility in the United States: What We Have Learned from the PSID." *The Annals of the American Academy of Political and Social Science* 680(1): 213–234.

Miller, Grant. 2008. "Women's Suffrage, Political Responsiveness, and Child Survival in American History." *Quarterly Journal of Economics* 123(3): 1287–1327.

Mohammed, A. R. Shariq and Paul Mohnen. 2025. "Black Economic Progress in the Jim Crow South: Evidence from Rosenwald Schools." *The Journal of Economic History* 85(1): 72–109.

Mountjoy, Jack. 2022. "Community Colleges and Upward Mobility." *American Economic Review* 112(8): 2580–2630.

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.

Ruggles, Steven, Sarah Flood, Matthew Sobek, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Renae Rodgers, and Megan Schouweiler. 2024. IPUMS USA: Version 15.0 [dataset]. Minneapolis, MN: IPUMS. https://doi.org/10.18128/D010.V15.0

Ruggles, Steven, Matt A. Nelson, Matthew Sobek, Catherine A. Fitch, Ronald Goeken, J. David Hacker, Evan Roberts, and J. Robert Warren. 2024. IPUMS Ancestry Full Count Data: Version 4.0 [dataset]. Minneapolis, MN: IPUMS. https://doi.org/10.18128/D014.V4.0

Ruggles, Steven, Julia Drew, Catherine A. Fitch, J. David Hacker, Jonas Helgertz, Matt A. Nelson, Nesile Ozder, Matthew Sobek, and John Robert Warren. 2025. "The IPUMS multigenerational longitudinal panel: progress and prospects." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 58(4): 258–271.

Sewell, William H. and Robert M. Hauser. 1975. *Education, Occupation, and Earnings: Achievement in the Early Career*. New York: Academic Press.

Solon, Gary. 1992. "Intergenerational Income Mobility in the United States." *American Economic Review* 82(3): 393–408.

Solon, Gary. 1999. "Intergenerational Mobility in the Labor Market." *Handbook of Labor Economics, vol. 3*, 1761–1800.

Song, Xi, Catherine G. Massey, Karen A. Rolf, Joseph P. Ferrie, Jonathan L. Rothbaum, and Yu Xie. 2020. "Long-term decline in intergenerational mobility in the United States since the 1850s." *Proceedings of the National Academy of Sciences* 117(1): 251–258.

Stephens, Jr, Melvin and Dou-Yan Yang. 2014. "Compulsory Education and the Benefits of Schooling." *American Economic Review* 104(6): 1777–1792.

Tan, Hui Ren. 2023. "A Different Land of Opportunity: The Geography of Intergenerational Mobility in the Early Twentieth-Century United States." *Journal of Labor Economics* 41(1): 77–102.

Torche, Florencia. 2015. "Analyses of Intergenerational Mobility: An Interdisciplinary Review." *The Annals of the American Academy of Political and Social Science* 657(1): 37–62.

U.S. Bureau of Education. 1923. *Biennial Survey of Education, 1918–1920.* Washington, D.C.

U.S. Bureau of Labor Statistics. 2023. "Women in the Labor Force : A Databook (April 2023)." Report 1103. https://fraser.stlouisfed.org/title/307#673985, accessed on September 24, 2025.

U.S. Census Bureau. 2010. *County Adjacency File*. https://www.census.gov/geographies/reference-files/2010/geo/county-adjacency.html

U.S. Department of Education, National Center for Education Statistics. 2022. "Estimated average annual salary of teachers in public elementary and secondary schools, by state: Selected school years, 1969–70 through 2021–22." *Digest of Education Statistics*, Table 211.60.

U.S. Department of Education, National Center for Education Statistics. 2023a. "Total fall enrollment in degree-granting postsecondary institutions, by attendance status, sex of student, and control of institution: Selected years, 1947 through 2031." *Digest of Education Statistics*, Table 303.10.

U.S. Department of Education, National Center for Education Statistics. 2023b. "Total and current expenditures per pupil in public elementary and secondary schools: Selected school years, 1919-20 through 2020-21." *Digest of Education Statistics*, Table 236.55.

U.S. Department of Education, National Center for Education Statistics. 2023c. "Average total cost of attendance for first-time, full-time degree/certificate-seeking undergraduate students in degree-granting postsecondary institutions, by control and level of institution, living arrangement, and component of student costs: Selected academic years, 2010–11 through 2022–23." *Digest of Education Statistics*, Table 330.40.

U.S. Office of Education. 1931. *Biennial Survey of Education, 1928–1930*. Washington, D.C.

Wagner, Deborah and Mary Layne. 2014. "The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' (CARRA) Record Linkage Software," U.S. Census Bureau Center for Administrative Records Research and Applications Working Paper #2014-01.

Ward, Zachary. 2023. "Intergenerational Mobility in American History: Accounting for Race and Measurement Error." *American Economic Review* 113(12): 3213–3248.

Zimmerman, David J. 1992. "Regression Toward Mediocrity in Economic Stature." *American Economic Review* 82(3): 409–429.

Link to Online Appendix

Online Appendix for

The Evolution of U.S. Educational Mobility over the 20th Century and the Role of Public Education

By Martha J. Bailey, A. R. Shariq Mohammed, and Paul Mohnen

Appendix A. Linking Children and Parents in SS-5 Records to the 1920 and 1940 Censuses Using Supervised Machine Learning

This section describes the process of linking children and parents in SS-5 records to the 1920 and 1940 Censuses using the supervised machine learning approach from Bailey et al. (2023). The final linked dataset is a combination of seven types of links: (A) linking male children in SS-5 records to the 1940 Census, (B) linking female children in SS-5 records to the 1940 Census, (C) linking siblings together across SS-5 records, (D) linking parents and children in SS-5 records as families to households in the 1900–1940 Censuses, (E) linking parents in SS-5 records to the 1940 Census as couples (without age and birthplace information), (F) linking parents in SS-5 records to the 1940 Census as couples (with age and birthplace information), and (G) linking fathers and mothers in SS-5 records to the 1940 Census as individuals.

Each link relies on different pieces of information. Link A uses name, year of birth and birthplace information to link men to the 1940 Census. Link B is analogous to link A, except that women are linked using both their birth name and their married name(s) (if available). Link C uses information on parents' full names to link siblings together across SS-5 records. These links are used to create a crosswalk between SSNs and family identifiers in the SS-5 data, which is then used to reconstruct families in the SS-5 data and link them to households in the 1940 Census using information on parents' names and children's name, age, and birthplace (link D). We also link those families to households in the 1900–1930 Censuses to obtain information on parents' year and place of birth, which we exploit in subsequent links. Link E uses fathers' full name and mothers' first and middle name to link parents to the 1940 Census as couples. Link F is similar to link E, except that it additionally exploits parents' year and place of birth information (for the subset of parents for which this information is available from link D). Lastly, link G also exploits parents' demographic information from link D to separately link fathers and mothers to the 1940 Census as individuals using the same process as links A and B.

Each link above consists of the following four steps: (1) generate a set potential candidates for each record we want to link, (2) make manual linking decisions for a random sample of those records, (3) use the manual linking decisions to train an algorithm to mimic those decisions, and (4) scale up linking to the entire set of records. We briefly describe each step in turn.

Step 1: Generating Sets of Candidate Links

The first step consists of generating a set of potential candidates in the target dataset for humans (or the algorithm) to choose from when attempting to find a link for each "primary" record in the starting dataset. The first two columns in Appendix Table A1 specify the universe of primary records and potential candidates for each type of link. To select promising candidates, we use name similarity scores between primary records and potential candidates. Since there are typically millions of primary and potential records and since generating name similarity scores is computationally intensive, we reduce the dimensionality of the problem by first "blocking" on certain characteristics, such as sex, first or last name initials, year of birth windows, state/continent of birth, or some combination of these. Note that some of these may be conditions we would like links to satisfy anyway. Column 3 in Appendix Table A1 lists the blocking criteria for each type of link, which varies depending on the available information. In practice, we try to impose as few conditions as possible in order to avoid missing potential links due to imperfectly-reported information in historical records (e.g., name misspellings, mis-measured year of birth or birthplace, etc.).[1] Within each block, we compute name similarity scores between each primary-potential record pair, and retain the top 20 or 25 candidates according to these scores. Column 4 in Appendix Table A1 defines the name similarity scores we use to rank potential candidates, which again vary across links depending on the available information.

Step 2: Making Manual Linking Decisions for a Random Sample of Records

Once we have sets of potential candidates for all primary records, the second step consists of making manual linking decisions for a random sample of primary records. Column 5 in Appendix Table A1 lists the information displayed to human "trainers" to aid them in their decisions (this is also what the

[1] For some links we take the *union* of blocks to cast a wide net. For links A and B, we separately block on last name and first name initials, which means that potential candidates need only satisfy one of these conditions. For link D, we block on the union of children's state/continent of birth, which means that potential households need only have one child whose birthplace matches the birthplace of one of the children in the primary family.

algorithm will "see" when making linking decisions). We generally try to provide as much information as possible (subject to all the information comfortably fitting on a computer screen).[2]

Our training process is modeled on the training process described in Bailey et al. (2023). To maximize the quality of the training data, we take two measures: (1) Trainers are instructed to only make links when they are very confident (importantly, trainers have the option to *not* make any link if there is too much ambiguity in the set of potential candidates, and are encouraged to do so when appropriate), and (2) we aggregate decisions from multiple independent trainers. To be more specific, each case is independently reviewed by two trainers. Trainers agree if they select the same link or both decline to make a link. The subset of cases which they disagreed on are shown to three additional trainers, mixed in with a random sample of cases they agreed on (so that the three trainers cannot distinguish between different types of cases). For cases that were considered by five trainers, only links chosen by four out of five trainers are considered links in the final training data. Column 6 in Appendix Table A1 specifies the number of cases that were shown to trainers and the resulting "hand" match rate.

Step 3: Training an Algorithm to Mimic Human Linking Decisions

The third step consists in using the manual linking decisions to train an algorithm to make similar linking decisions. The hand-linked data is randomly split into two sets: a "training" sample (70 or 75 percent of the training data) that the algorithm is allowed to learn from and a "hold-out" sample (25 or 30 percent of the training data) that we use to evaluate the out-of-sample performance of the algorithm. We use the two-stage model described in Bailey et al. (2023) to generate model-based links. In the first stage, the probability that there exists a link in the set of potential candidates for a particular record is modeled using a random forest model and set-level features (intuitively capturing how many promising candidates are contained in the set; the more promising candidates there are, the lower the probability that a true link can be confidently chosen). In the second stage, the probability that any potential candidate in the set is the true link, conditional on a true link being contained in the set, is modeled using a logit-style discrete

[2] Note that some pieces of information, like race and country of birth, are displayed for trainers to optionally take into account with full knowledge that this information is not always accurately reported and does not have to match.

choice model and pair-level features (intuitively capturing how promising a particular candidate is). These probabilities are estimated using ten-fold cross-validation in the training sample, and multiplied to obtain the unconditional link probability for each primary-potential pair. Model-based links are then made using a threshold rule: All pairs with a match probability exceeding the threshold are called links and the remaining pairs are called non-links.

A key decision is how to set this threshold. The higher the threshold the higher the share of links made by human trainers that the algorithm is able to reproduce ("recall rate") but the lower the share of links made by the algorithm that were "correct" according to human trainers ("precision rate"). The advantage of labeled data is that it allows us to control this tradeoff. In practice, we select the threshold associated with a precision rate of 95 percent in the training sample.[3] Column 7 in Appendix Table A1 specifies the out-of-sample recall and precision rates achieved by our models in the hold-out samples.

Link C differs from all other links in that it is an instance of one-to-many linking since individuals can have multiple siblings. Therefore, trainers are allowed to select multiple links, and a random forest model is used to model the probability that any pair of individuals are siblings.

Step 4: Scaling up Linking to the Full Set of Records

The fourth and final step consists in scaling up linking by using the trained algorithm to generate match probabilities in the full sample of records, and generating machine links using the cross-validated probability thresholds associated with a 95-percent precision rate. After generating links in the full sample, we implement a series of post-processing steps to iron out conflicts in the data. For example, it is possible for one record in the target dataset to be linked to multiple primary records in the starting dataset. We eliminate such conflicts either by selecting one link that has a significantly higher match probability than all the other links and dropping the remaining ones, or dropping all the links if no link stands out. Another type of conflict that arises as part of link B is when two name variations of the same female SS-5

[3] The only exceptions are the intermediate links C and D (links to 1900–1930 Censuses) that are used as an input for subsequent links, for which we impose a 97-percent precision rate based on the training sample.

record are linked to two different women in the 1940 Census. We use a similar strategy to eliminate such conflicts.

Appendix B. Details on Construction of Weights

To more closely approximate nationally representative estimates of educational mobility, we construct individual-level weights to balance our linked and survey samples with the population of interest. These weights target (1) sex, (2) race/ethnicity (non-Hispanic White, non-Hispanic Black, other), and (3) birth cohort. For the 1910–1919 cohorts, we additionally balance our samples on years of education as of 1940. This is because linking historical records based on name information generally produces samples that are slightly skewed towards more educated individuals (Bailey et al. 2020b). This stems from the fact that more educated people have longer and less common names, and less educated individuals have shorter and more common names, which makes them harder to link. The survey samples for the 1910–1919 cohorts are also slightly skewed towards more educated individuals, because those surveys took place in the second half of the 20th century and only contain individuals who survived, who also tend to be more educated. These issues are less important for the 1982–1997 cohorts because the Census Bureau uses more information than is available in the historical data, and the Census and *ACS* took place when those cohorts were still relatively young. In contrast, the older individuals in the modern sample have less opportunity to be sampled in the *ACS* after age 25 than the younger cohorts, which means modern samples with completed education tend to be slightly older than the population.

We construct inverse propensity scores weights following the procedure described in Bailey et al. (2020a). This involves pooling the sample to be re-weighted with a random sample of the population, and estimating a probit regression modeling the probability of appearing in the population as a function of common characteristics. For the 1910–1919 cohorts, we include dummy variables for sex-by-race/ethnicity group-by-individual birth year category, and additionally dummy variables for individual years of education. The inclusion of educational attainment dummies allows us to upweight individuals with less education to account for the fact that they are slightly underrepresented in the linked sample. For the 1982–1997 cohorts, we include dummy variables for sex-by-race/ethnicity group-by-individual birth year categories. This accounts for the fact that younger individuals are less likely to appear in our analysis sample because we require that they be age 25 or older in the *ACS*, which means there are fewer *ACS* years in which this requirement can be satisfied. For the later 20th century cohorts, we omit the educational attainment dummies because the PIK'd sample is based on more information and does not underrepresent individuals with less education.

The estimated probabilities are then used to weights such that the mean sample characteristics match the means in the population. If the starting sample already has weights (which is the case for our survey samples for which sampling weights are provided), these can be incorporated in the re-weighting procedure such that the final composite weight achieves balance. For the pooled *GSS/NSFH* sample covering the 1910–1919 cohorts, we first adjust the sampling weights by normalizing them to sum to one within each survey and then multiply them by the share of observations coming from each survey so that more weight is placed on the larger *GSS* sample, similar to Jácome et al. (2025).

Appendix Table A3 compares the population averages in the 1940 Census against the unweighted and weighted means in our 1920-1940 linked sample and in the pooled *GSS/NSFH* survey sample. Comparing the means across columns 1 and 2, we can see that our linked sample is skewed towards later birth cohorts, White Americans, and more educated individuals. Although we cannot observe the population of parents of children born in 1910–1919, we can identify the subset of parents who co-reside with their children in 1940, which is what we use to approximate the mean educational attainment in the population of the most educated parent. This reveals that our linked sample is also skewed towards individuals with more educated parents. The *p*-values in column 3 indicate that we can reject the null hypothesis that the differences in the sample and population means are equal to 0. Similarly, column 6 shows that the pooled *GSS/NSFH* survey sample overrepresents women, later birth cohorts, White Americans, more educated individuals, and individuals with more educated parents.

Columns 4 and 8 show the reweighted means are very close to the population means, and columns 5 and 9 imply that we cannot reject equality (with the exception of the share of non-Hispanic Whites and other race/ethnicity groups in the linked sample, but the differences are very close to 0). Note also that even though we do not target parental education, the weighted means are much closer to the population means, with the difference relative to the population mean declining from 0.74 to 0.10 years in the 1920-1940 linked sample and from 0.99 to 0.07 years in the pooled *GSS/NSFH* sample.

Appendix C. Constructing County Characteristics and State Resources for Public Education in the Historical and Modern Period

County Demographic and Socio-economic Characteristics

We construct various county-level demographic and socio-economic characteristics for the historical period using microdata from the 1920 Decennial Census (unless otherwise indicated). Indicators of demographic structure include total population, population density (information on county area comes from Eckert et al. 2020), demographic composition (share non-Hispanic Black, share Hispanic, share non-Hispanic Asian, share foreign-born, share born out-of-state), income inequality (Gini coefficient), and racial residential segregation (Theil index). Income inequality is based on annual wage income in the 1940 Census. Indicators of economic development include the share of urban households, the share of farm households, the labor force participation rate, the share of the labor force in manufacturing industries, the share of the labor force in business and professional service industries, income per capita, and median house values. Income per capita is based on annual wage income in the 1940 Census, while median house values come from the 1930 Census (both measures are converted to 1999 dollars using the Bureau of Labor Statistics' historical "Consumer Price Index for Urban Consumers" series). Indicators of local public service provision include the number of labor force participants in public administration industries per capita, the number of labor force participants in protective service occupations per capita (e.g., firemen, policemen), and the number of labor force participants in healthcare occupations per capita. Indicators of family structure include the share of adults that are married, the share of adults that are divorced, the share of children that are in single-parent households, and the average household size. Lastly, as a measure of social capital, we construct the share of religious adherents by dividing the number of religious adherents in the 1926 Census of Religious Bodies, obtained via the National Historical

Geographic Information System (NHGIS; Manson et al. 2024), by the total population in the 1920 Census.

We construct modern-day analogs of the above county characteristics using county-level tables from the 2000 Decennial Census published by the Census Bureau and obtained via NHGIS. One minor difference is that the share of people in specific industries (manufacturing, business and professional services, public administration) or specific occupations (protective services, healthcare) are measured using employed individuals rather than labor force participants. Because these measures cannot be computed using county-level data published by the Census Bureau, income inequality and racial residential segregation come from data made available by Opportunity Insights (https://opportunityinsights.org/data/). We also obtained the share of religious adherents in 2000 from Opportunity Insights. Appendix Table A10 shows the means of the county-level characteristics described above, for the subset of counties used in our analysis.

State Resources for Public Education

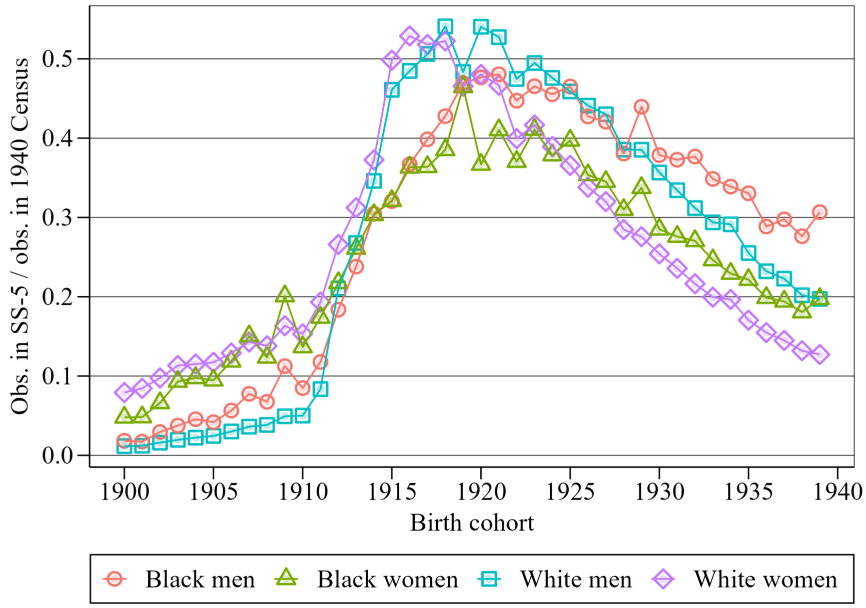
State-level data on the number of K-12 teachers and average K-12 teacher salaries for the historical period are based on data we digitized from the U.S. Bureau of Education's 1919–1920 *Biennial Survey of Education* (U.S. Bureau of Education 1923). Average K-12 teacher salaries are measured using the "average annual salaries of teachers, supervisors, and principals." State-level data on instructors at public colleges and universities for the historical period are based on data we digitized from the U.S. Office of Education's 1928–1930 *Biennial Survey of Education* (U.S. Office of Education 1931). We measure access to post-secondary education in 1930 rather than in 1920 because it is closer to college-going ages of our cohorts of interest (1910–1919), and because there was tremendous growth in public tertiary education during the 1920s (the number of public colleges and universities more than doubled between 1920 and 1930). More specifically, we digitized data on the number of professors and instructors for all 246 publicly-controlled universities, colleges, and professional schools, 134 publicly-controlled teachers colleges, 66 state normal schools, 26 city normal schools, and 47 county normal schools. We then aggregated instructor counts at the state level.

For the modern period, state-level data on K-12 teachers comes from the National Center for Education Statistics' (NCES) Common Core of Data (CCD) state-level files for the school year 1999–2000. Note that NCES measures K-12 teachers in full-time equivalent (FTE) units rather than individuals. State-level data on average K-12 teacher salaries comes from estimates of the average annual salary of teachers in public elementary and secondary schools for the school year 1999–2000 reported by NCES (U.S. Department of Education 2022). State-level data on instructors at public colleges and universities for the modern period are based on institutional-level data from NCES' Integrated Postsecondary Education Data System (IPEDS) database for the school year 2000–2001, which we aggregate at the state level. We focus on publicly-controlled, Title IV-participating post-secondary institutions. Instructors are measured using the number of full-time and part-time faculty involved in instruction, research, or public service. We convert the number of instructors in FTE terms by multiplying part-time faculty members by 1/3.

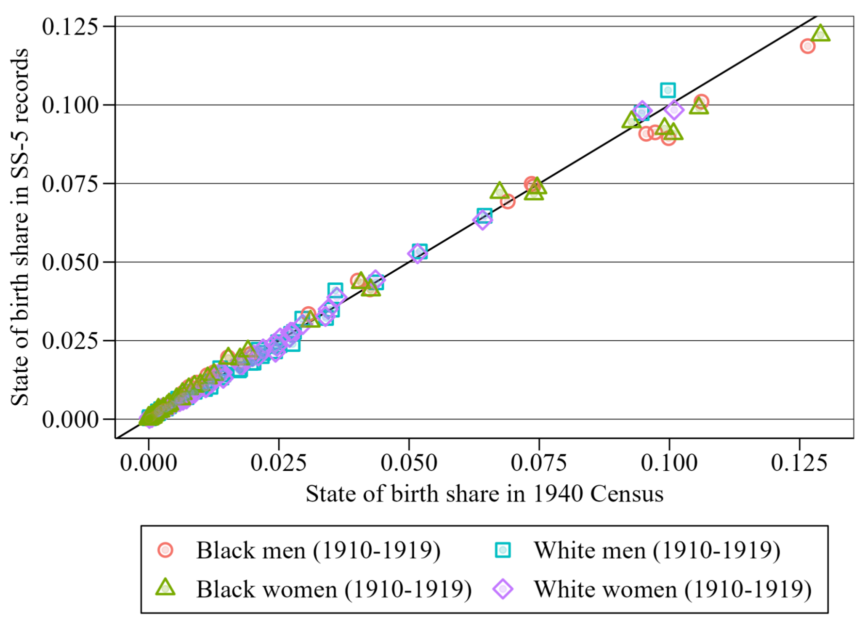
State-level K-12 teachers and public college instructors are normalized by the corresponding number of children aged 5–18, as measured in the 1920 Decennial Census (historical period) or in the 2000 Decennial Census (modern period). Average K-12 teacher salaries are converted to 1999 dollars using the Bureau of Labor Statistics' historical "Consumer Price Index for Urban Consumers" series. Appendix Table A11 displays the resulting statistics for each state.

Figure A1. Cohort and State of Birth Coverage of SS-5 Records by Sex and Race

Panel A. Cohort Coverage

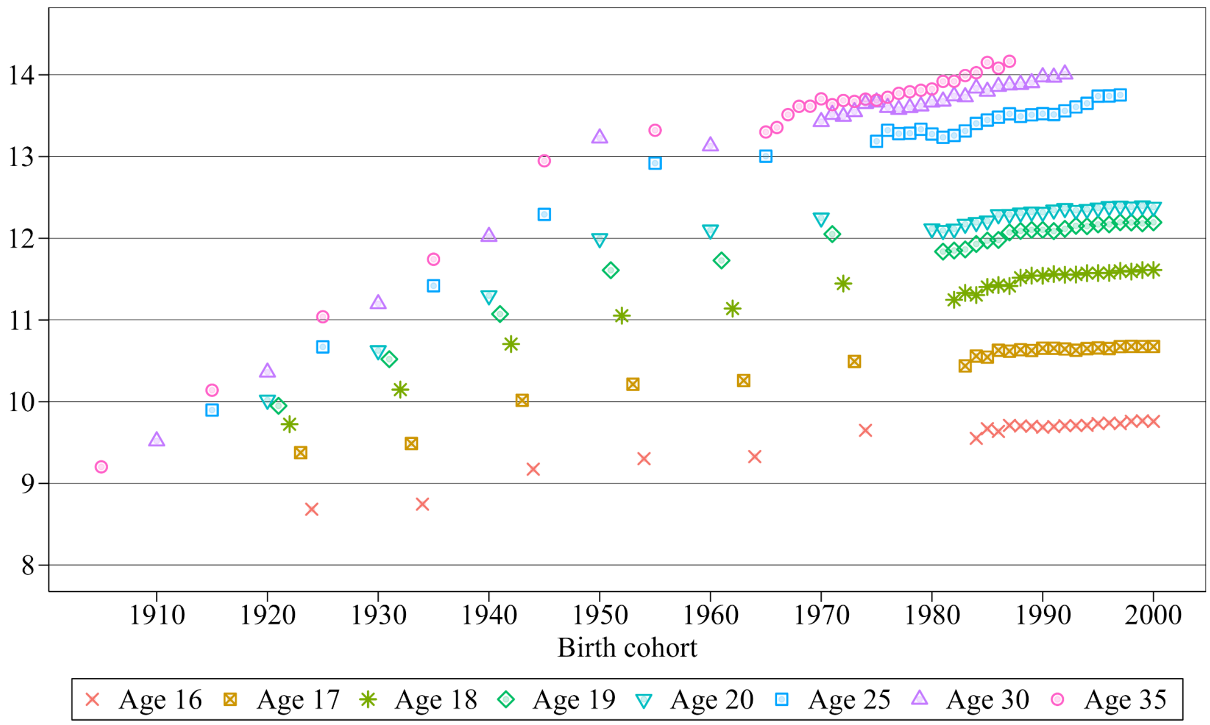


Panel B. State of Birth Coverage



Notes: Panel A plots the number of observations by cohort in SS-5 records divided by the corresponding number of observations in the 1940 Census, separately by sex and race. Panel B plots state of birth shares in SS-5 records against corresponding shares in the 1940 Census among individuals born in the U.S. in 1910–1919, separately by sex and race.

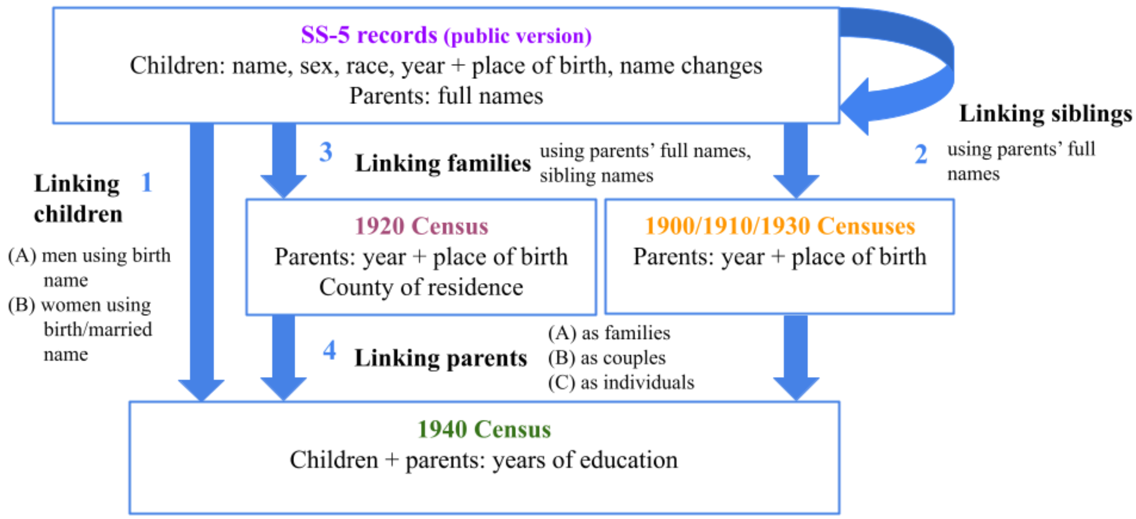
Figure A2. Average Years of Education by Age and Cohort



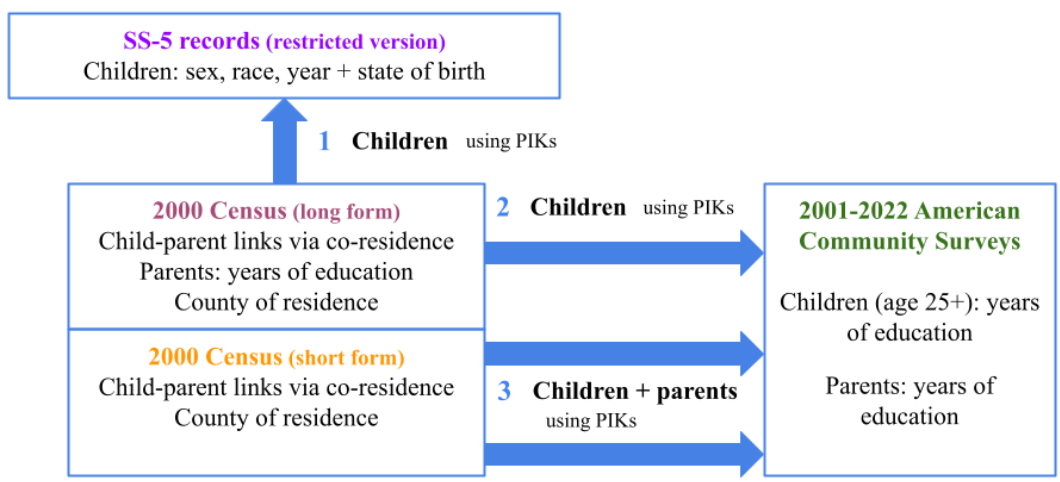
Notes: This graph plots average years of education at specified ages by cohort using data from the 1940–2000 decennial censuses and the 2001–2022 *ACS*. Sample is restricted to individuals born in the U.S. Observations are weighted by census sampling weights.

Figure A3. Overview of Construction of Linked Samples

Panel A. 1920-1940 Census Linked Sample

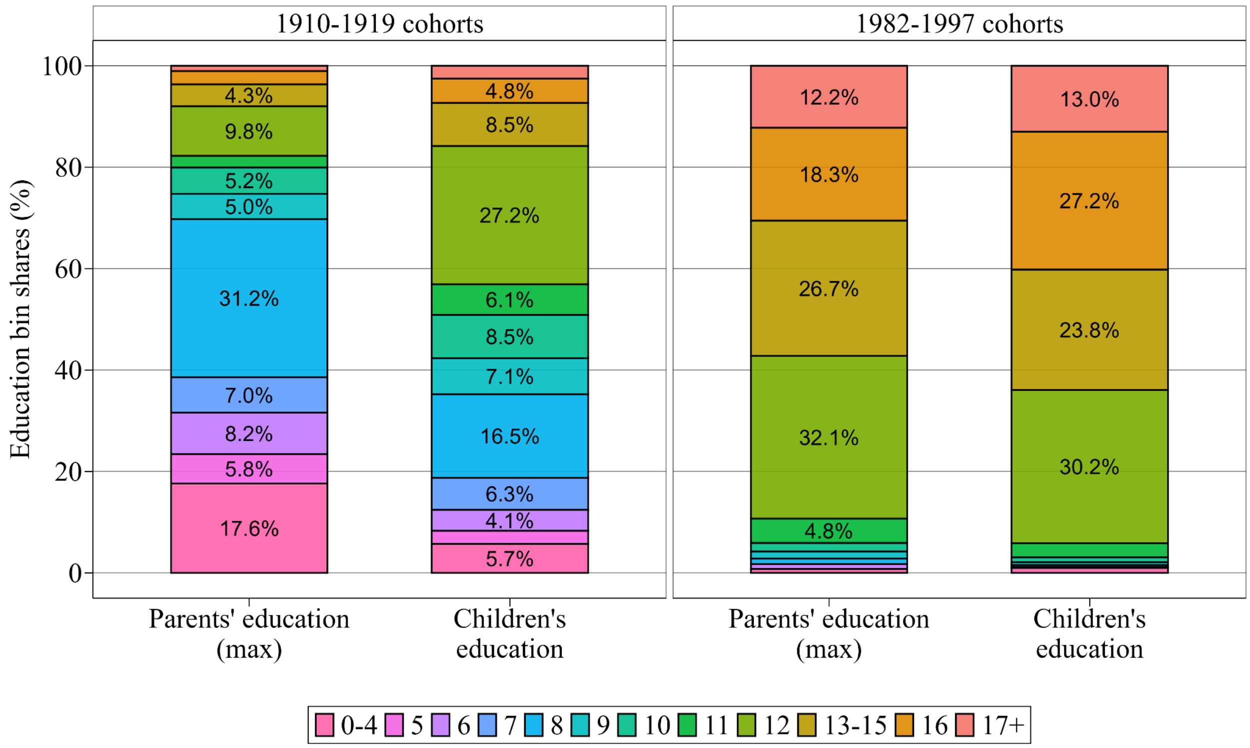


Panel B. 2000 Census-ACS Linked Sample



Notes: PIKs refer to Protected Identification Keys used by the Census Bureau to identify unique individuals across records (Wagner and Layne 2014).

Figure A4. Distribution of Children’s and Parents’ Educational Attainment by Cohort



Notes: This graph plots the distribution of children’s and parents’ education by children’s birth cohort. Parental education is measured for the most educated parent. Children’s education is measured in the 1950 Census for the 1910–1919 cohorts and in the 2022 *ACS* for the 1982–1997 cohorts. Parents’ education is measured in the 1940 Census for the 1910–1919 child cohorts and in the 2000 Census for the 1982–1997 child cohorts. The sample is restricted to individuals born in the U.S. Children are weighted by sample-line weights in the 1950 Census and by person weights in the 2022 *ACS*. Parents are weighted by household weights in the 2000 Census.

Figure A5. National Education Transition Matrix by Cohort

Panel A. 1910–1919 Cohorts

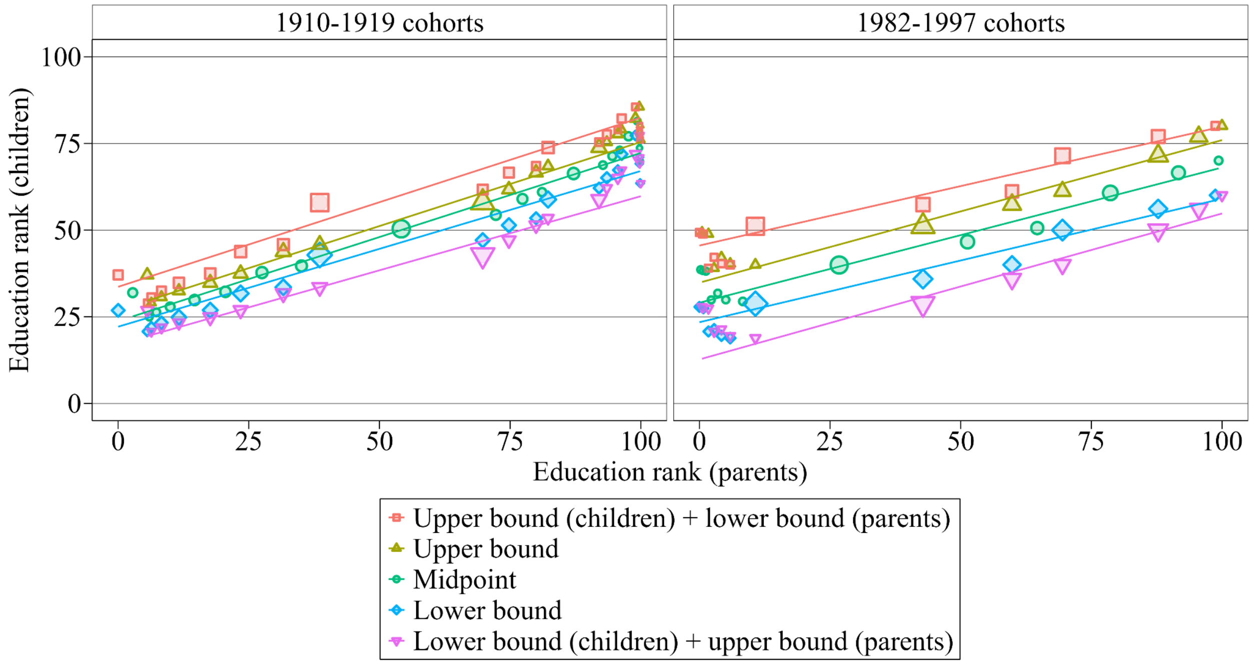
Years of education \ Max of parents' years of education	0-4	5-6	7-8	9	10	11	12	13-15	16	17+
17+	0.4%	0.4%	0.7%	1.0%	1.4%	1.6%	2.6%	4.0%	7.1%	15.2%
16	1.1%	1.2%	2.4%	3.7%	5.0%	5.6%	9.7%	13.7%	24.2%	23.0%
13-15	2.1%	2.7%	5.7%	8.7%	11.0%	12.9%	16.7%	24.7%	26.9%	25.3%
12	12.1%	16.5%	29.6%	35.4%	39.5%	40.3%	41.9%	33.3%	24.7%	19.3%
11	3.8%	5.1%	6.7%	7.9%	8.0%	9.0%	5.8%	4.6%	3.1%	2.8%
10	7.4%	9.1%	10.7%	10.5%	10.6%	9.1%	6.7%	5.4%	3.8%	3.6%
9	7.5%	8.8%	8.4%	9.3%	6.6%	5.6%	4.2%	3.3%	2.3%	2.6%
7-8	30.9%	33.8%	28.0%	16.5%	12.8%	10.7%	8.4%	7.0%	4.8%	5.3%
5-6	16.7%	13.6%	4.5%	3.9%	2.8%	2.7%	1.8%	1.7%	1.4%	1.1%
0-4	18.1%	8.7%	3.1%	3.0%	2.3%	2.5%	2.2%	2.1%	1.8%	1.8%

Panel B. 1982–1997 Cohorts

Years of education \ Max of parents' years of education	0-4	5-6	7-8	9	10	11	12	13-15	16	17+
17+	4.0%	3.4%	2.6%	2.5%	2.3%	2.0%	4.7%	7.2%	13.9%	22.1%
16	18.9%	16.7%	11.1%	11.4%	9.0%	7.9%	17.7%	26.7%	45.3%	49.2%
13-15	25.5%	29.0%	22.4%	23.0%	21.5%	22.1%	27.1%	32.1%	23.1%	16.9%
12	41.5%	42.2%	39.7%	47.3%	50.3%	52.2%	45.2%	31.2%	16.5%	10.9%
11	2.5%	3.0%	3.8%	4.2%	5.5%	6.2%	2.0%	1.0%	0.4%	0.3%
10	1.6%	1.8%	3.0%	3.7%	5.0%	4.3%	1.4%	0.7%	0.2%	0.2%
9	1.3%	1.2%	3.1%	3.2%	3.2%	2.5%	0.7%	0.3%	0.1%	0.1%
7-8	0.7%	0.9%	12.8%	3.0%	1.5%	1.0%	0.4%	0.2%	0.1%	0.1%
5-6	0.3%	0.6%	0.3%	0.3%	0.2%	0.2%	0.1%	0.1%	0.0%	0.0%
0-4	3.8%	1.2%	1.2%	1.5%	1.5%	1.5%	0.8%	0.5%	0.3%	0.3%

Notes: For both panels, each cell reports the percentage of children attaining a level of education indicated on the *y*-axis, conditional on parents having the education level indicated in the *x*-axis. Observations for the 1910–1919 cohorts are reweighted to match the sex-by-cohort-by-race and education distribution in the 1940 population. Observations for the 1982–1997 cohorts are reweighted to match the sex-by-cohort-by-race distribution in the 2000 population. Darker cells indicate higher conditional shares. Corresponding percentile ranks can be derived using Appendix Figure A4. The procedure for constructing weights is described in Appendix B.

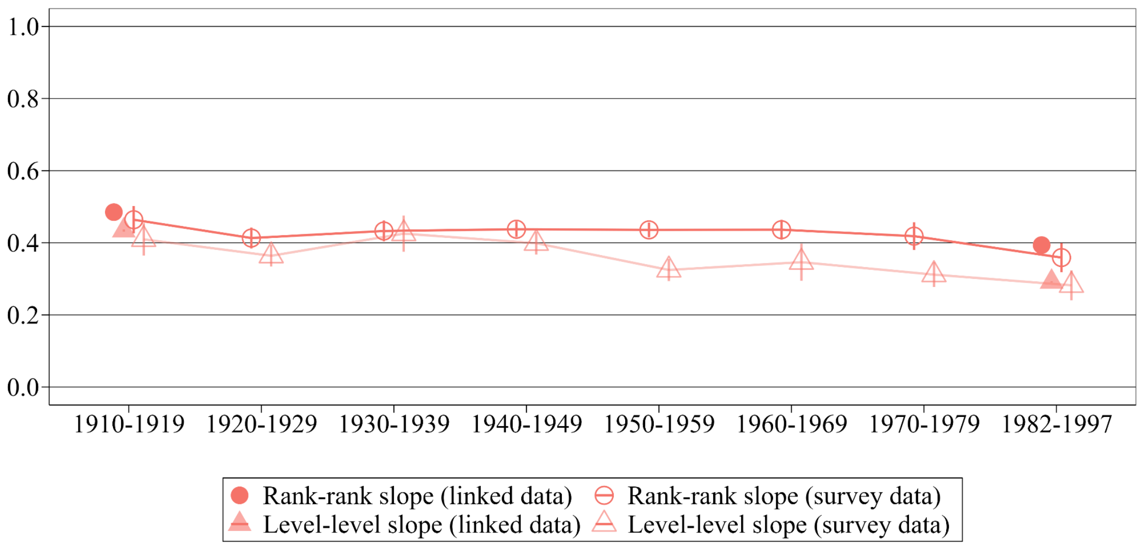
Figure A6. The Relationship Between Child and Parent Education Ranks, by Cohort: Alternative Rank Assignment Methods



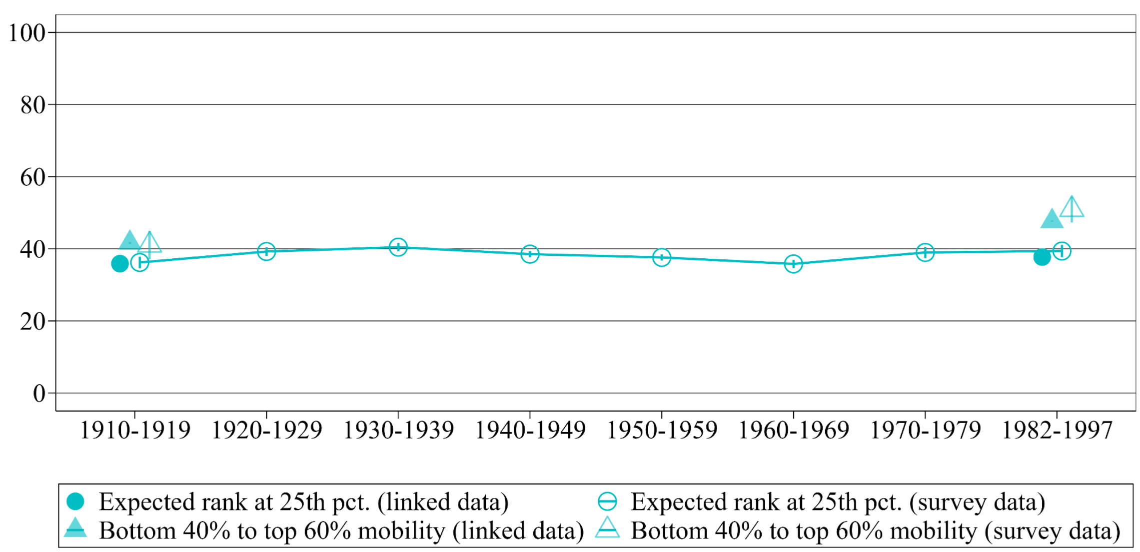
Notes: This graph plots the average education rank of children by parental education rank for different rank assignment methods, separately by cohort. Parental education is measured using the education of the most educated parent. Fitted lines are based on weighted linear regressions (see Appendix Table A4 for detailed results). Observations for the 1910–1919 cohorts are reweighted to match the sex-by-cohort-by-race and education distribution in the 1940 population. Observations for the 1982–1997 cohorts are reweighted to match the sex-by-cohort-by-race distribution in the 2000 population. The procedure for constructing weights is described in Appendix B.

Figure A7. Relative and Upward Educational Mobility over the 20th Century

Panel A. Relative Educational Mobility



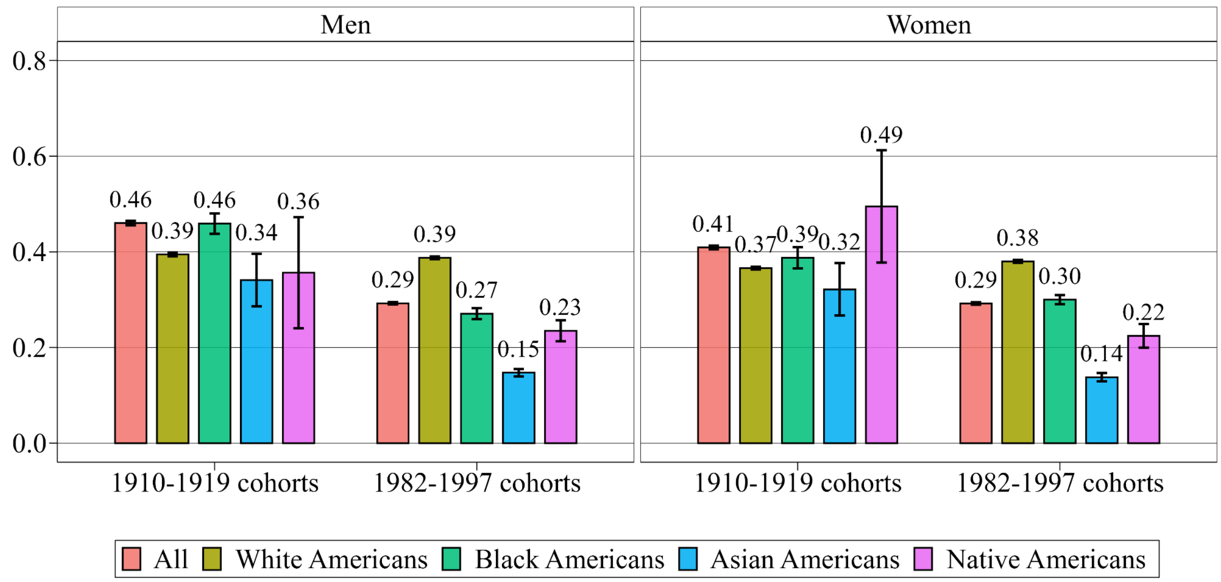
Panel B. Upward Educational Mobility



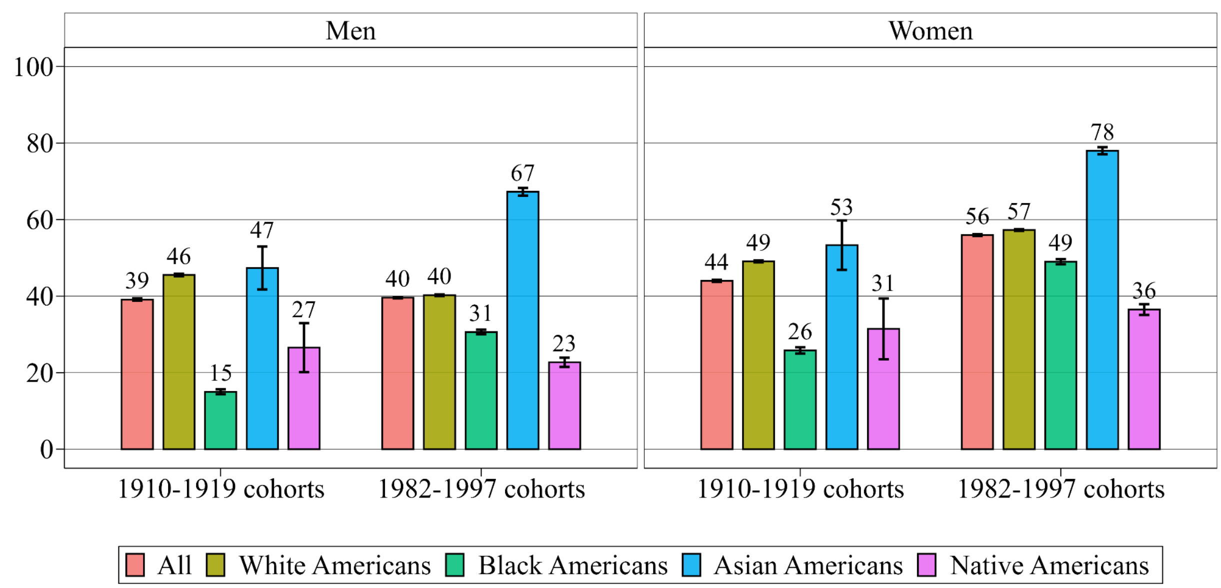
Notes: This graph plots estimates of relative and upward educational mobility by cohort and sample. Linked data for the 1910–1919 cohorts are reweighted to match the sex-by-cohort-by-race and education distribution in the 1940 population. Linked data for the 1982–1997 cohorts are reweighted to match the sex-by-cohort-by-race distribution in the 2000 population. Survey data for the 1910–1979 cohorts are reweighted to match the sex-by-cohort-by-race and education distribution in the nearest decennial census or *ACS* year in which individuals are aged 25 or older (1950 for 1910–1919 cohorts, 1960 for 1920–1929 cohorts, etc.). Survey data for the 1982–1997 cohorts are weighted by *GSS* sampling weights. The procedure for constructing weights is described in Appendix B. 95-percent confidence intervals are based on robust standard errors.

Figure A8. Alternative Measures of Educational Mobility, by Cohort and Sex or Sex-by-Race/Ethnicity Group

Panel A. Level-Level Slope



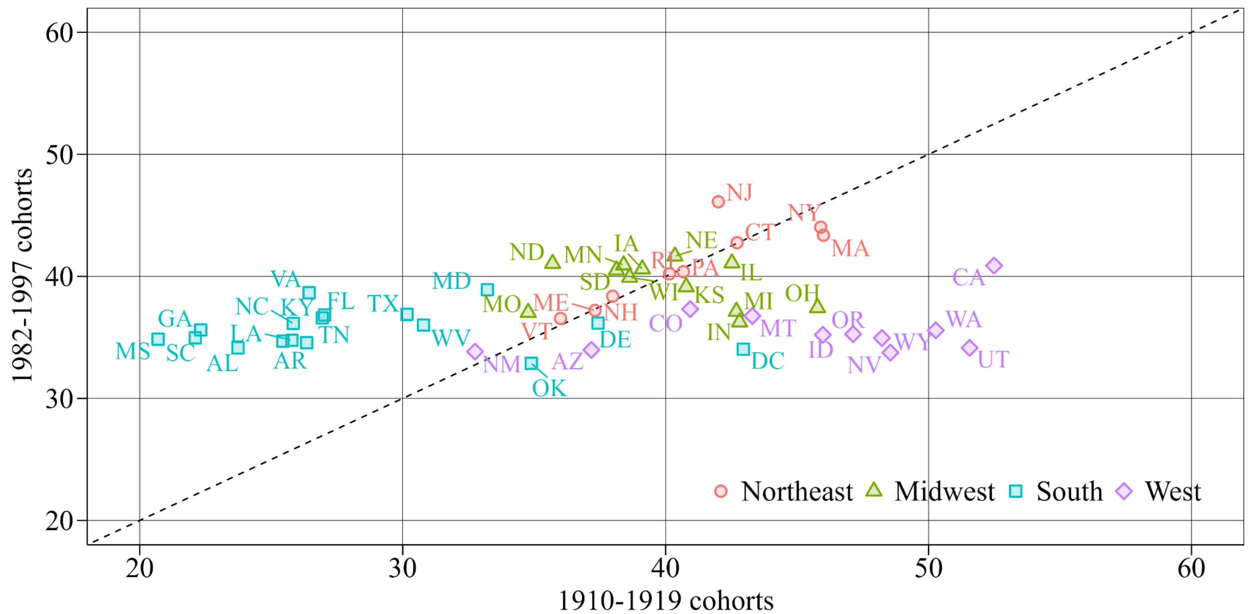
Panel B. Bottom 40 to Top 60 Percent Mobility



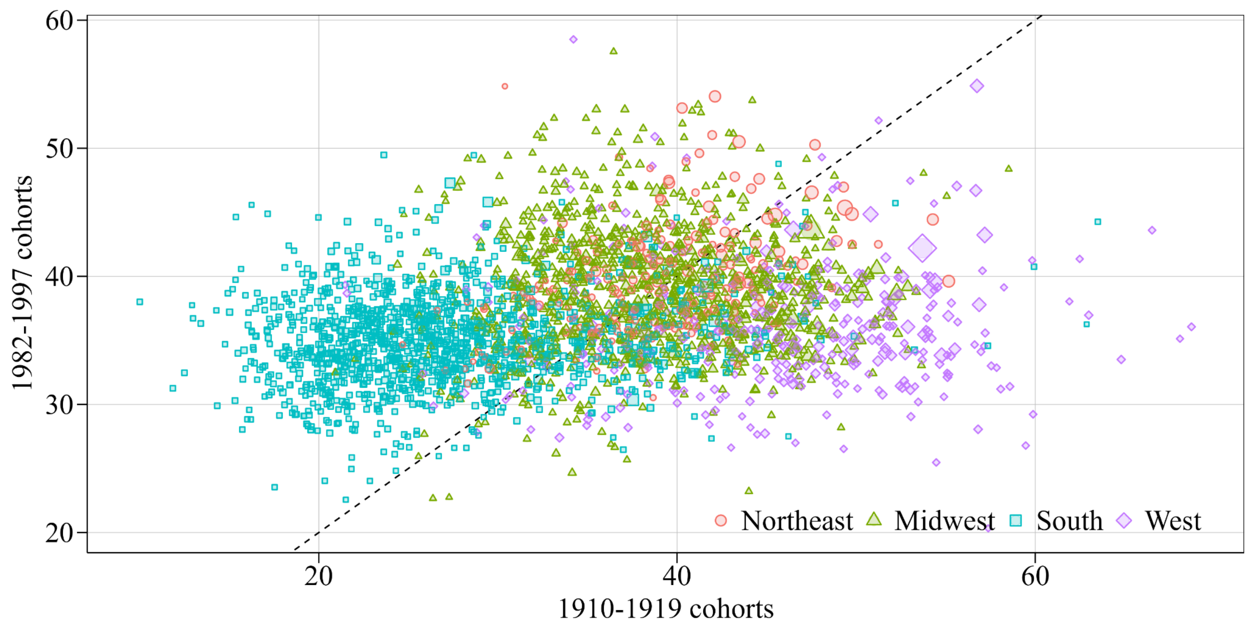
Notes: These graphs plot estimates of the level-level slope and bottom 40 to top 60 percent mobility by cohort and sex or sex-by-race/ethnicity group based on our linked samples. 95-percent confidence intervals are based on robust standard errors.

Figure A9. Exp. Rank at 25th Percentile by Childhood State or County of Residence

Panel A. Exp. Rank at 25th Percentile by Childhood State of Residence, 1910–19 vs. 1982–97 Cohorts



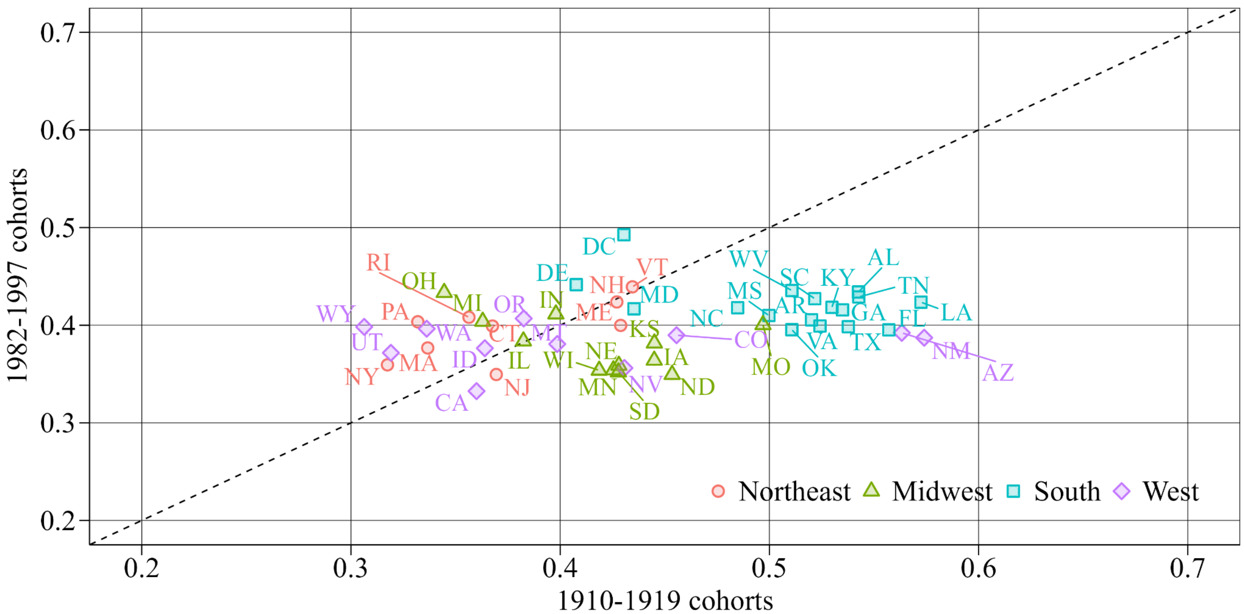
Panel B. Exp. Rank at 25th Percentile by Childhood County of Residence, 1910–19 vs. 1982–97 Cohorts



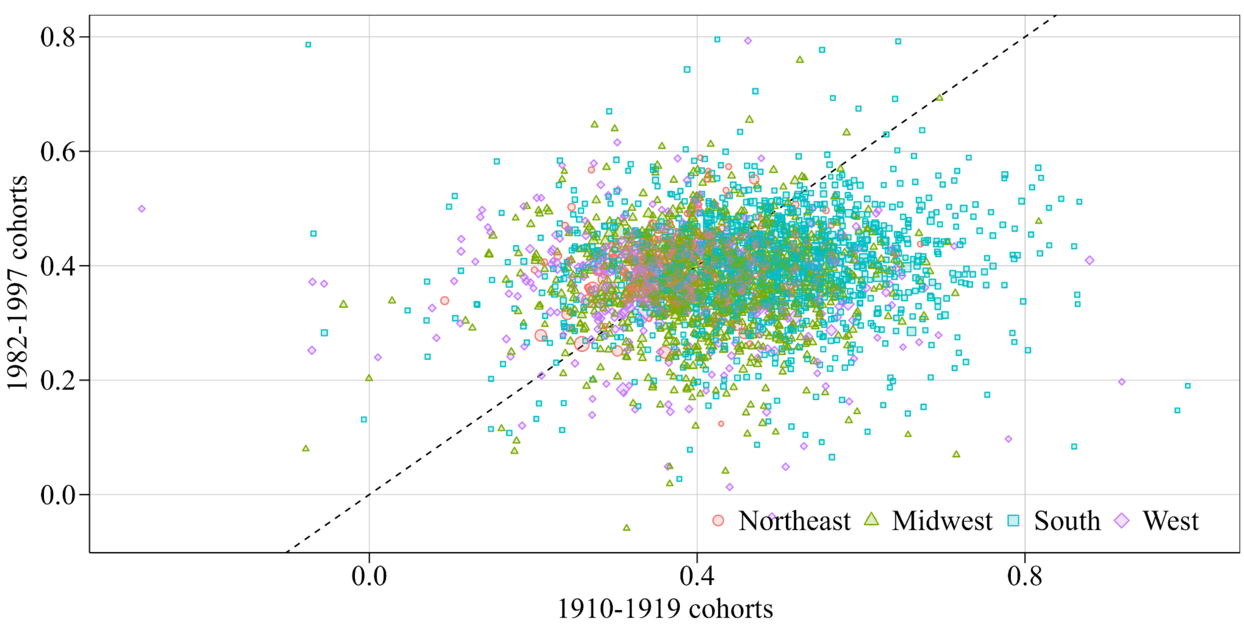
Notes: These graphs plot expected rank at the 25th percentile by 1920 childhood state or county of residence for the 1910–1919 cohorts against expected rank at the 25th percentile by 2000 childhood state or county of residence for the 1982–1997 cohorts.

Figure A10. Rank-Rank Slope by Childhood State or County of Residence

Panel A. Rank-Rank Slope by Childhood State of Residence, 1910–19 vs. 1982–97 Cohorts



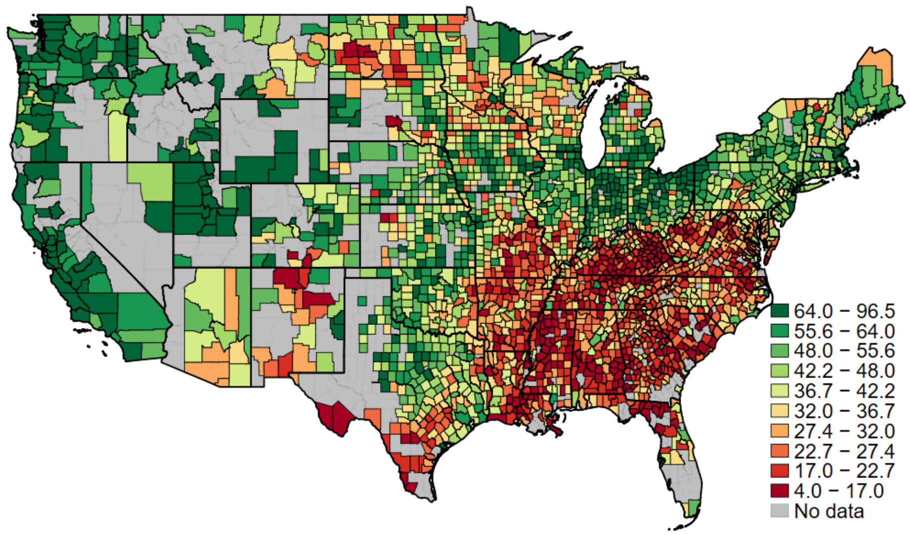
Panel B. Rank-Rank Slope by Childhood County of Residence, 1910–19 vs. 1982–197 Cohorts



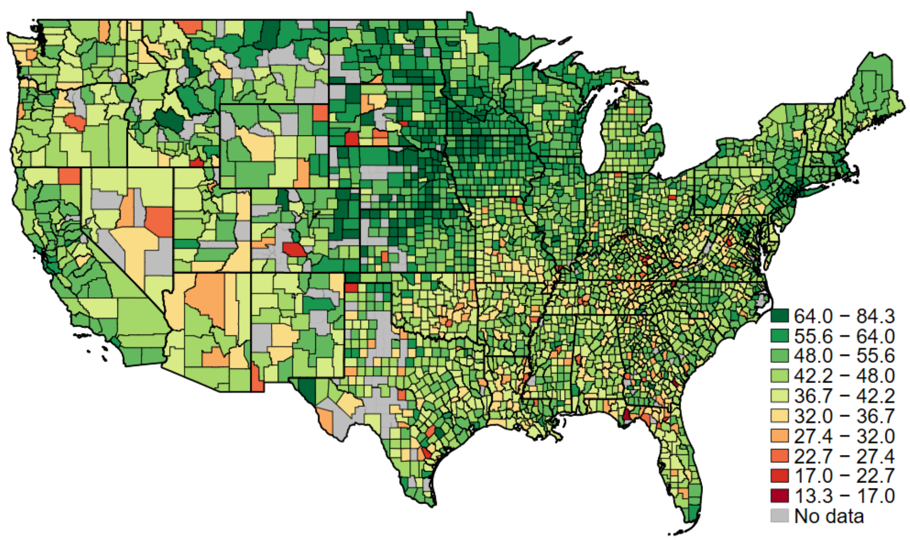
Notes: These graphs plot the rank-rank slope by 1920 childhood state or county of residence for the 1910–1919 cohorts against the rank-rank slope by 2000 childhood state or county of residence for the 1982–1997 cohorts.

Figure A11. Bottom 40 to Top 60 Percent Mobility by Childhood County of Residence over the 20th Century

Panel A. 1910–1919 Cohorts

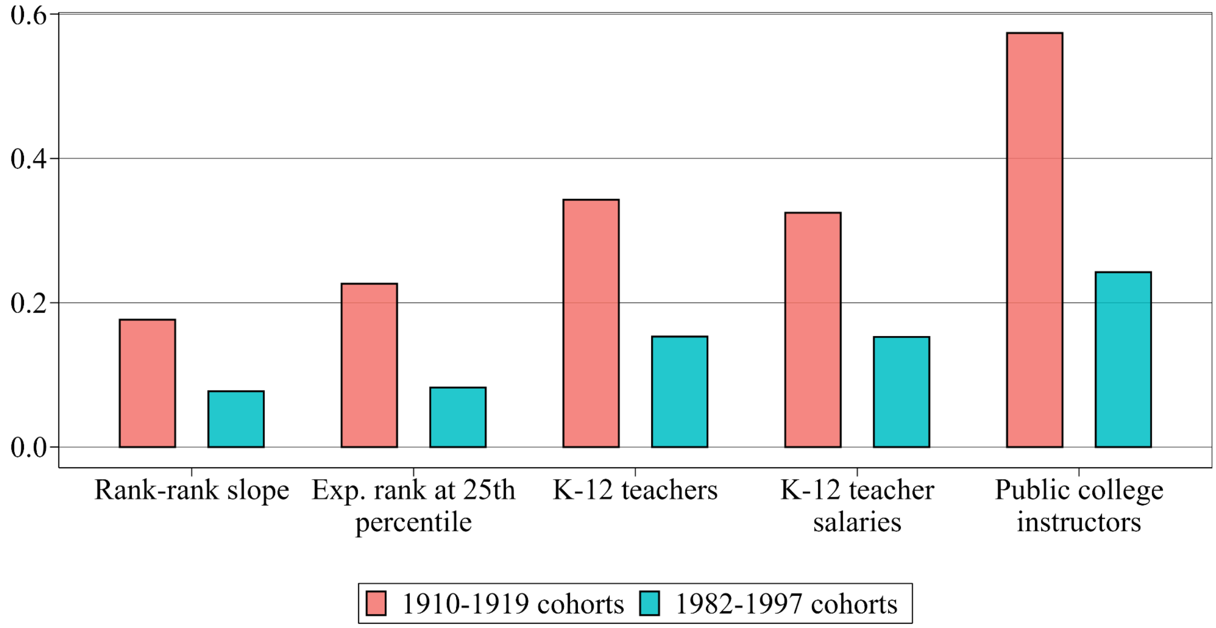


Panel B. 1982–1997 Cohorts



Notes: Figures plot bottom 40 to top 60 percent mobility by children's county of residence in 1920 for the 1910–1919 cohort and by county of residence in 2000 for the 1982–1997 cohort for counties with stable boundaries between 1920 and 2000 (see text for definition) and at least 20 parent-child observations. In both panels, counties are shaded according to a common set of decile bin cutoffs based on the 1910–1919 cohorts.

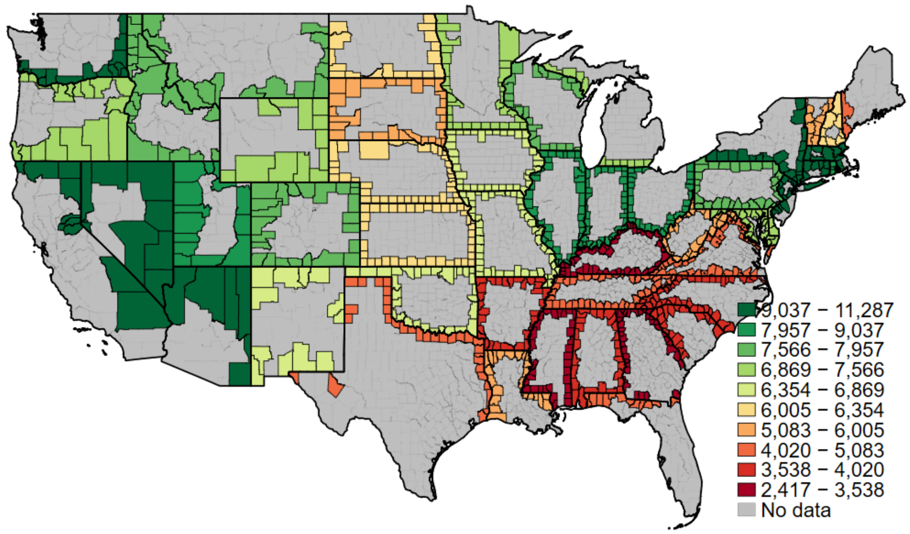
Figure A12. Coefficients of Variation for State Estimates of Educational Mobility and State Resources for Public Education, by Period



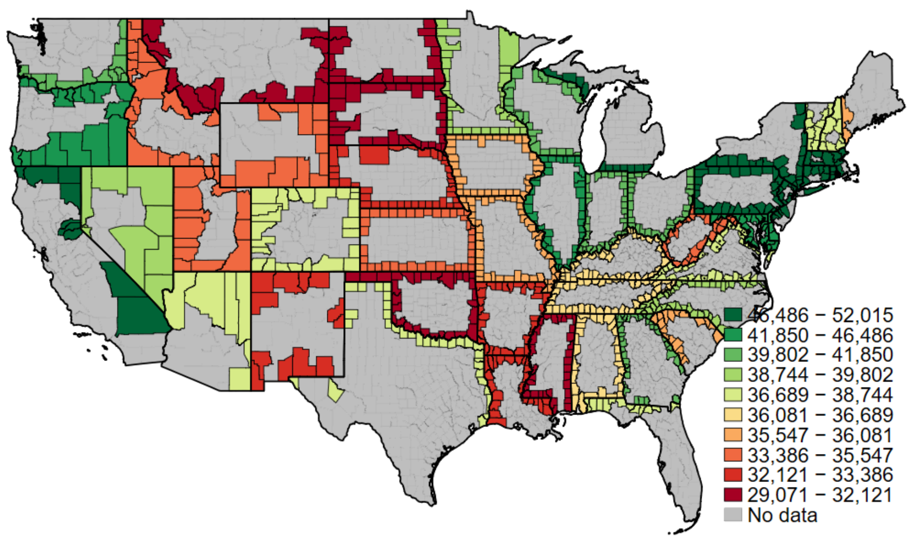
Notes: This graph plots coefficients of variation (standard deviation divided by the mean) associated with state estimates of educational mobility and state resources for public education, separately by period.

Figure A13. State Border Maps: Average K-12 Teacher Salary

Panel A. Average State K-12 Teacher Salaries in 1920

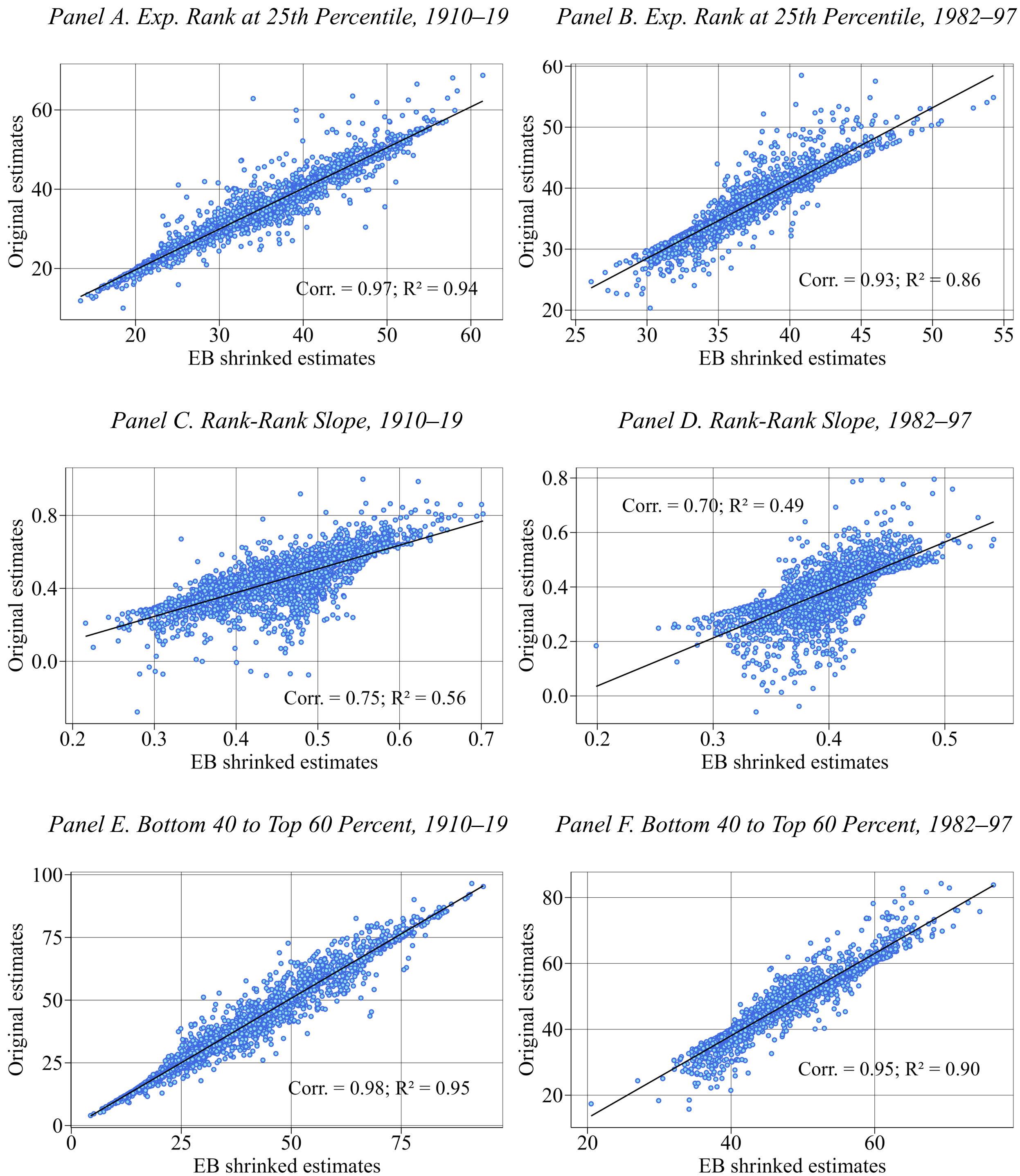


Panel B. Average State K-12 Teacher Salaries in 2000



Notes: These graphs plot average state K-12 teacher salaries in 1920 or 2000 for the subset of the counties in our state border regression samples. Counties are shaded according to which decile bin the associated state K-12 teacher salary falls into.

Figure A14. Original vs. Empirical Bayes Shrinked County Estimates



Notes: These graphs plot the original county educational mobility estimates against empirical Bayes (EB) versions of those estimates. EB estimates shrink toward the state mean and use precision weights proportional to the inverse of the variance of the original county estimates (Armstrong et al., 2022). Fitted lines are based on unweighted regressions of the county estimates on the EB-shrunken estimates (regression R^2 and Pearson correlation coefficient also shown).

Table A1. Overview of Linking Process

	Universe of primary records (starting data)	Universe of potential candidates (target data)	Blocking criteria for potential candidates	Selection criteria for potential candidates	Information displayed to human trainers	Trainer match rate	Model recall / precision rate in hold-out sample	Notes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) Linking male children in SS-5 records to the 1940 Census	Men with non-missing first and last name, year of birth, and birthplace	Men with non-missing first name, last name, age, and birthplace	First or last name initial (union), +/-3-year age window in 1940, state/continent of birth	Jaro-Winkler score for first and last name (top 25)	Full name, age in 1940, state/country of birth, race	42.6% (*N* = 1,500 cases)	76% / 92.5% (25-percent hold-out sample)	
(B) Linking female children in SS-5 records to the 1940 Census	Women with non-missing first and last name (birth or married), year of birth, and birthplace	Women with non-missing first name, last name, age, and birthplace	First or last name initial, +/-3-year age window in 1940, state/continent of birth, marital status (single vs. non-single)	Jaro-Winkler score for first and last name (top 25)	Full name, age in 1940, state/country of birth, race	53.2% (*N* = 500 unique SSNs, *N* = 1,543 cases)	70% / 93.3% (25-percent hold-out sample)	Women linked using birth and married name(s) (within-SSN conflicts resolved ex post if necessary)
(C) Linking siblings together in SS-5 records	Individuals with non-missing year of birth and father last name	Individuals with non-missing year of birth and father last name	Father last name initials (first two), +/-10-year of birth window	Jaro-Winkler score for father and mother first and last name (top 20)	Father and mother full names, year of birth, state/country of birth, race	39.8% (*N* = 493 primaries, *N* = 9,860 cases)	92.1% / 98.9% (30-percent hold-out sample)	One-to-many linking (random forest model)
(D) Linking parents in SS-5 records to the 1900–1940 Censuses as households (with information on children via 97-percent precision sibling linkages)	Families with non-missing father last name and at least 1 child with non-missing birthplace and aged 25 or under in target census year	Households with non-missing father last name and at least 1 child with non-missing birthplace	Father last name initial, children's state/continent of birth (union)	Jaro-Winkler score for father first and last name and mother first name (top 20)	Father full name, mother first and middle name, up to 10 children (first and middle name, age, birthplace), modal race of child(ren)/race of potential father	42.1% (*N* = 985 cases)	92.9% / 92.9% (30-percent hold-out sample)	Training data based on linking families in SS-5 records to households in the 1910 and 1920 Censuses (*N* = 495 cases from 1920 linking; *N* = 490 cases from 1910 linking)

Table A1. Overview of Linking Process (continued)

	Universe of primary records to be linked	Universe of potential candidates	Blocking criteria for potential candidates	Selection criteria for potential candidates	Information displayed to human trainers	Trainer match rate	Model recall / precision rate in hold-out sample	Notes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(E) Linking parents in SS-5 records to the 1940 Census as couples (without age and birthplace information)	Parents with non-missing father first and last name, and mother first name	Couples with non-missing father first and last name, and mother first name	Father last name initials (first two)	Jaro-Winkler score for father first and last name and mother first name (top 20)	Father full name, mother first and middle name, modal race of child(ren)/race of potential father	25.5% (*N* = 990 cases)	58.1% / 95.6% (30-percent hold-out sample)	
(F) Linking parents in SS-5 records to the 1940 Census as couples (with age and birthplace information via 97-percent precision household linkages to 1900–1930 Censuses)	Parents with non-missing father first and last name, mother first name, father year of birth, and birthplace	Couples with non-missing father first and last name, mother first name, father age, and birthplace	Father last name initial, father state/continent of birth, +/-5-year father age window in 1940	Jaro-Winkler score for father first and last name and mother first name (top 20)	Father full name, mother first and middle name, father and mother age in 1940 and birthplace, father race	47.6% (*N* = 500 cases)	84.8% / 98.5% (30-percent hold-out sample)	
(G) Linking parents in SS-5 records to the 1940 Census as individuals (with age and birthplace information via 97-percent precision household linkages to 1900–1930 Censuses)	Parents with non-missing first and last name, year of birth, and birthplace	Men/women with non-missing first and last name, age, and birthplace	First or last name initial (union), +/-3-year age window in 1940, state/continent of birth, marital status (mothers only)	Jaro-Winkler score for first and last name (top 25)	—	—	—	Recycle training data and models from links A and B

Table A2. Sample Sizes and Match Rates by Sex in 1920-1940 Census Linked Sample

	All (1910–1919)	Men (1910–1919)	Women (1910–1919)
	(1)	(2)	(3)
Number of children linked to 1940	3,344,524	1,576,650	1,767,874
Match rate	*0.412*	*0.421*	*0.404*
Number of children linked to 1940 with at least one parent linked to 1940	1,979,754	995,327	984,427
Match rate	*0.244*	*0.266*	*0.225*
Number of children linked to 1940 and 1920 with at least one parent linked to 1940	1,495,542	752,083	743,459
Match rate	*0.184*	*0.201*	*0.170*
Number of children in SS-5 records	8,117,560	3,743,686	4,373,874
Number of children in 1940 Census	21,897,415	10,762,772	11,134,643

Notes: This table displays sample sizes and match rates for our 1920-1940 Census linked sample, pooling women and men in column 1 and separately by sex in columns 2 and 3. Corresponding counts in our starting sample (SS-5 records) and in the reference population (1940 Census) are shown in the bottom two rows.

Table A3. Representativeness of 1920-1940 Linked Sample and Pooled Survey Data: 1910–1919 Cohorts

	Pop. (1940)	1920-1940 linked sample				Pooled *GSS*/*NSFH* surveys			
		Unweighted		Weighted		Survey weights		Weighted	
	Mean	Mean	Diff. (1) – (2) [p-value]	Mean	Diff. (1) – (4) [p-value]	Mean	Diff. (1) – (6) [p-value]	Mean	Diff. (1) – (8) [p-value]
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	0.509	0.497	0.012	0.510	-0.001	0.547	-0.038	0.508	0.001
			[0.000]		[0.253]		[0.000]		[0.961]
Year of birth	1914.59	1915.84	-1.248	1914.59	0.007	1915.17	-0.578	1914.57	0.020
			[0.000]		[0.144]		[0.000]		[0.779]
Non-Hispanic White	0.881	0.970	-0.089	0.882	-0.002	0.902	-0.021	0.885	-0.004
			[0.000]		[0.025]		[0.001]		[0.639]
Non-Hispanic Black	0.105	0.025	0.079	0.104	0.001	0.088	0.016	0.104	0.001
			[0.000]		[0.344]		[0.004]		[0.946]
Other race	0.015	0.004	0.010	0.014	0.001	0.010	0.005	0.011	0.003
			[0.000]		[0.000]		[0.043]		[0.199]
Education	9.87	10.98	-1.110	9.88	-0.005	11.35	-1.473	9.87	0.001
			[0.000]		[0.434]		[0.000]		[0.991]
Education: 1950 rank	48.18	58.06	-9.875	48.20	-0.014	60.16	-11.97	48.16	0.024
			[0.000]		[0.751]		[0.000]		[0.972]
Most educated parent's education	7.75	8.49	-0.742	7.84	-0.095	8.73	-0.988	7.81	-0.068
			[0.000]		[0.000]		[0.000]		[0.491]
Most educated parent's education: 1940 rank	50	55.84	-5.839	50.40	-0.396	57.54	-7.539	50.75	-0.754
			[0.000]		[0.000]		[0.000]		[0.305]

Notes: Population for all outcomes except parental education is a 10 percent random sample of children born in the U.S. in 1910–1919 in the 1940 full-count Census (N = 2,143,755). Population for parental education is a 10 percent random sample of parents co-residing with children born in the U.S. in 1910–1919 in the 1940 Census (N = 466,849). Differences between population and linked sample (N = 1,457,003) or pooled surveys (N = 2,851) are based on linear regressions stacking samples, where the dependent variable is the outcome of interest and the independent variable is a dummy for the population (weights are normalized to sum to 1 in the population and linked sample/survey). *p*-values for the null hypothesis that the difference relative to population mean is 0 (in brackets below) are based on robust standard errors. Columns 4–5 and 8–9 re-weight the relevant samples to match the 1940 population along sex-by-race-by-cohort and years of education. See Appendix B for details.

Table A4. Rank-Rank Slope and Expected Rank at 25th Percentile Estimates from Figure 1

	Rank assignment method				
	(1)	(2)	(3)	(4)	(5)
Panel A. Rank-rank slope					
1910–1919 cohorts	0.488	0.485	0.485	0.449	0.427
1982–1997 cohorts	0.343	0.411	0.390	0.356	0.421
Change	-0.145	-0.074	-0.095	-0.093	-0.006
Panel B. Expected rank at 25th percentile					
1910–1919 cohorts	45.9	39.0	35.9	33.4	27.7
1982–1997 cohorts	54.1	45.1	38.8	32.3	23.2
Change	8.3	6.1	2.9	-1.0	-4.5
R^2 (1910–1919)	0.903	0.958	0.972	0.925	0.900
R^2 (1982–1997)	0.930	0.934	0.958	0.945	0.922
N (1910–1919)	21	21	21	21	21
N (1982–1997)	13	13	13	13	13
Child ranks	Upper bound	Upper bound	Midpoint	Lower bound	Lower bound
Parent ranks	Lower bound	Upper bound	Midpoint	Lower bound	Upper bound

Notes: This table shows the rank-rank slope and expected rank at 25th percentile estimates associated with the fitted regression lines in Figure 1. Note that, unlike the midpoint method under which the fitted line must go through the (50,50) coordinate by construction, alternative rank assignment methods can simultaneously produce a declining rank-rank slope and a declining expected rank at the 25th percentile (columns 4 and 5) as the fitted line can both shift and rotate over time in offsetting ways.

Table A5. The Impact of Educational Creep: Evidence from a Sample Linking the 1915 Iowa Census to the 1940 Census

	Dependent variable:			
	Son's education (1940)		Son's education rank (1940)	
	(1)	(2)	(3)	(4)
Father's education (1915)	0.39 (0.05)			
Father's education (1940)		0.39 (0.03)		
Father's education rank (1915)			0.36 (0.04)	
Father's education rank (1940)				0.37 (0.03)
Constant	7.38 (0.42)	7.31 (0.30)	32.16 (2.32)	31.54 (1.89)
Expected rank at 25th percentile (1915)			32.3 (1.5)	
Expected rank at 25th percentile (1940)				31.6 (1.3)
Mean of son's education (or rank)	10.60	10.60	50.00	50.00
Mean of father's education (or rank)	8.19	8.32	49.63	50.00
R^2	0.092	0.140	0.090	0.127
N	772	772	772	772

Notes: This table shows the results from regressing sons' years of education in 1940 (or education rank) on a constant and fathers' years of education in 1915 or 1940 (or fathers' education rank), using a sample of father-son pairs in the 1915 Iowa Census that are both linked to the 1940 Census. Education ranks for fathers and sons are based on the in-sample distribution of education in 1940 (midpoint method). Robust standard errors are in parentheses (calculated using the Delta method for expected ranks at the 25th percentile).

Table A6. Sensitivity of Educational Mobility Estimates to Alternative Measures of Parental Education

	Measure of parent educational attainment		
	Most educated parent	Least educated parent	Mean of both parents' education, when different
	(1)	(2)	(3)
Panel A. Rank-rank slope			
1910–1919 cohorts	0.464	0.421	0.463
	(0.019)	(0.020)	(0.019)
1982–1997 cohorts	0.359	0.304	0.351
	(0.021)	(0.022)	(0.020)
Change	-0.105	-0.117	-0.112
Panel B. Expected rank at 25th percentile			
1910–1919 cohorts	36.2	37.8	36.5
	(0.8)	(0.8)	(0.8)
1982–1997 cohorts	39.4	41.2	39.6
	(0.9)	(0.9)	(0.9)
Change	3.2	3.4	3.1
R^2 (1910–1919)	0.251	0.210	0.254
R^2 (1982–1997)	0.144	0.102	0.142
N (1910–1919)	2,851	2,851	2,851
N (1982–1997)	3,001	3,001	3,001

Notes: This table shows estimates of the rank-rank slope and expected rank at the 25th percentile in the pooled *GSS/NSFH* survey data, separately by cohort and for different measures of parental education (most educated parent, least educated parent, or mean of both parents' education). Observations for the 1910–1919 cohorts are weighted by inverse propensity scores that re-weight the sample to match the sex-by-cohort-by-race and education distribution in the 1940 population. Observations for the 1982–1997 cohorts are weighted by *GSS* sampling weights. Robust standard errors in parentheses (calculated using the Delta method for expected rank at the 25th percentile).

Table A7. Educational Mobility Estimates by Cohort, Sample, and Sex or Race/Ethnicity

	Rank-rank slope		Level-level slope		Expected rank at 25th percentile		Bottom 40 to top 60 percent mobility	
	Linked data	Survey data	Linked data	Survey data	Linked data	Survey data	Linked data	Survey data
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Men								
1910–1919	0.500	0.439	0.460	0.401	34.5	35.3	39.1	39.0
	(0.002)	(0.031)	(0.002)	(0.036)	(0.08)	(1.30)	(0.16)	(3.10)
1982–1997	0.399	0.375	0.292	0.279	33.5	37.5	39.6	46.7
	(0.001)	(0.031)	(0.001)	(0.034)	(0.04)	(1.35)	(0.10)	(2.80)
Change	-0.102	-0.064	-0.168	-0.123	-1.0*	2.2	0.5*	7.6
Panel B. Women								
1910–1919	0.471	0.487	0.409	0.416	37.2	37.1	44.0	43.0
	(0.002)	(0.022)	(0.002)	(0.027)	(0.06)	(0.95)	(0.15)	(2.57)
1982–1997	0.390	0.348	0.292	0.289	42.1	41.1	56.0	54.8
	(0.001)	(0.028)	(0.001)	(0.027)	(0.05)	(1.15)	(0.11)	(2.41)
Change	-0.081	-0.139	-0.117	-0.127	4.9	4.0	12.0	11.8
Panel C. White Americans								
1910–1919	0.442	0.425	0.380	0.355	38.8	38.7	47.4	45.9
	(0.001)	(0.020)	(0.001)	(0.020)	(0.04)	(0.85)	(0.11)	(2.29)
1982–1997	0.422	0.378	0.384	0.319	37.8	39.8	48.5	50.9
	(0.001)	(0.024)	(0.001)	(0.027)	(0.04)	(1.08)	(0.08)	(2.21)
Change	-0.020	-0.047	0.004	-0.036	-1.1	1.0	1.2	5.0
Panel D. Black Americans								
1910–1919	0.407	0.479	0.425	0.468	22.1	24.9	20.6	23.4
	(0.007)	(0.052)	(0.008)	(0.072)	(0.13)	(1.81)	(0.27)	(3.88)
1982–1997	0.343	0.274	0.284	0.225	32.7	36.1	39.8	49.6
	(0.003)	(0.048)	(0.004)	(0.054)	(0.10)	(1.70)	(0.24)	(3.87)
Change	-0.064**	-0.205	-0.141	-0.242	10.6	11.2	19.2	26.2

Notes: This table shows educational mobility estimates by cohort, sample, and gender or race/ethnicity (see the note below Table 1 for more details). In panel A, the number of observations is 733,288/1,097,600 for the early/late 20th century cohorts in columns 1, 3, and 5; 1,096/1,319 in columns 2, 4, and 6; 208,484/417,700 in column 7; and 333/509 in column 8. In Panel B, the corresponding numbers are 723,715/1,038,100 in columns 1, 3, and 5; 1,755/1,682 in columns 2, 4, and 6; 215,797/397,000 in column 7; and 510/719 in column 8. In Panel C, the corresponding numbers are 1,413,390/1,656,000 in columns 1, 3, and 5; 2,523/2,216 in columns 2, 4, and 6; 396,517/576,000 in column 7; and 670/839 in column 8. In Panel D, the corresponding numbers are 37,109/161,000 in columns 1, 3, and 5; 297/528 in columns 2, 4, and 6; 24,529/75,000 in column 7; and 151/241 in column 8. Robust standard errors are in parentheses (calculated using the Delta method for expected rank at the 25th percentile). ***, **, and * indicate cases where we can reject the null hypothesis that changes in educational mobility over time are equal across samples at the 1 percent, 5 percent and 10 percent level, respectively.

Table A8. Rank-Rank Slope and Level-Level Slope Estimates by Cohort, Sex and Race/Ethnicity

	Men					Women				
	White	Black	Hispanic	Asian	Native	White	Black	Hispanic	Asian	Native
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Rank-rank slope										
1910–1919 cohorts	0.449 (0.002)	0.416 (0.009)	0.399 (0.038)	0.406 (0.030)	0.406 (0.060)	0.435 (0.001)	0.393 (0.010)	0.545 (0.040)	0.395 (0.031)	0.502 (0.052)
N	711,236	18,466	791	1,128	326	702,154	18,643	719	812	285
1982–1997 cohorts	0.432 (0.001)	0.325 (0.004)	0.268 (0.003)	0.258 (0.005)	0.280 (0.010)	0.411 (0.001)	0.366 (0.004)	0.272 (0.004)	0.221 (0.005)	0.288 (0.010)
N	853,000	79,500	99,000	42,000	12,000	803,000	81,500	94,000	36,500	11,900
Change	-0.017	-0.092	-0.132	-0.148	-0.125	-0.024	-0.027	-0.272	-0.175	-0.214
Panel B: Level-level slope										
1910–1919 cohorts	0.395 (0.002)	0.459 (0.011)	0.409 (0.037)	0.341 (0.028)	0.356 (0.059)	0.366 (0.001)	0.388 (0.011)	0.493 (0.039)	0.322 (0.028)	0.495 (0.060)
N	711,236	18,466	791	1,128	326	702,154	18,643	719	812	285
1982–1997 cohorts	0.388 (0.001)	0.271 (0.006)	0.131 (0.003)	0.147 (0.004)	0.235 (0.011)	0.380 (0.001)	0.300 (0.005)	0.132 (0.003)	0.138 (0.004)	0.224 (0.013)
N	853,000	79,500	99,000	42,000	12,000	803,000	81,500	94,000	36,500	11,900
Change	-0.007	-0.188	-0.278	-0.194	-0.121	0.014	-0.087	-0.361	-0.184	-0.270

Notes: This table shows rank-rank and level-level slope estimates by cohort and gender-by-race group (see note below Table 1 for more details). Robust standard errors are in parentheses.

Table A9. Exp. Rank at 25th Percentile and Bottom 40 to Top 60 Mobility Estimates by Cohort, Sex and Race/Ethnicity

	Men					Women				
	White	Black	Hispanic	Asian	Native	White	Black	Hispanic	Asian	Native
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Expected rank at 25th percentile										
1910–1919 cohorts	37.8 (0.07)	18.5 (0.17)	21.9 (0.77)	38.1 (1.12)	27.7 (1.44)	39.8 (0.05)	25.4 (0.20)	26.4 (0.79)	43.1 (1.30)	29.6 (1.60)
N	711,236	18,466	791	1,128	326	702,154	18,643	719	812	285
1982–1997 cohorts	33.4 (0.05)	28.1 (0.13)	34.9 (0.09)	49.5 (0.23)	23.7 (0.23)	42.4 (0.05)	37.3 (0.14)	42.7 (0.11)	57.5 (0.25)	29.9 (0.27)
N	853,000	79,500	99,000	42,000	12,000	803,000	81,500	94,000	36,500	11,900
Change	-4.5	9.6	12.9	11.4	-3.9	2.6	12.0	16.3	14.4	0.3
Panel B: Bottom 40 to top 60 percent mobility										
1910–1919 cohorts	45.5 (0.17)	15.0 (0.32)	16.7 (1.36)	47.4 (2.86)	26.5 (3.28)	49.1 (0.13)	25.8 (0.41)	20.8 (1.65)	53.3 (3.30)	31.4 (4.05)
N	194,325	12,342	597	371	174	202,192	12,187	490	273	129
1982–1997 cohorts	40.2 (0.11)	30.6 (0.32)	41.0 (0.24)	67.3 (0.51)	22.7 (0.62)	57.3 (0.12)	49.0 (0.35)	56.0 (0.25)	78.0 (0.48)	36.5 (0.73)
N	297,000	36,500	61,500	13,000	6,400	279,000	38,500	59,000	11,000	6,400
Change	-5.3	15.6	24.4	19.9	-3.8	8.2	23.2	35.2	24.7	5.0

Notes: This table shows expected rank at the 25th percentile and bottom 40 to top 60 percent mobility estimates by cohort and gender-by-race group (see the note below Table 1 for more details). Robust standard errors in parentheses (calculated using the Delta method for expected rank at the 25th percentile).

Table A10. County Characteristics by Year

	1920		2000	
	Mean	*N*	Mean	*N*
	(1)	(2)	(3)	(4)
Panel A. Demographic structure				
Population	35,662	2,881	92,664	2,881
Population density (people per sq mi)	137	2,881	223	2,881
Share Black	0.11	2,881	0.09	2,881
Share Hispanic	0.01	2,881	0.06	2,881
Share Asian	0.001	2,881	0.008	2,881
Share foreign-born	0.07	2,881	0.03	2,881
Share born out-of-state	0.21	2,881	0.26	2,881
Income inequality (Gini)	0.47	2,880	0.38	2,881
Racial residential segregation (Theil)	0.19	2,881	0.08	2,881
Panel B. Economic development				
Share urban households	0.19	2,881	0.4	2,881
Share farm households	0.52	2,881	0.05	2,881
Labor force participation rate	0.55	2,881	0.61	2,881
Manufacturing employment share	0.11	2,880	0.16	2,881
Business & professional services emp. share	0.04	2,880	0.26	2,881
Income per capita (1999 dollars)	1,981	2,880	17,487	2,881
Median house value (1999 dollars)	22,200	2,880	84,088	2,881
Panel C. Public services				
Public admin. workers per 1,000 people	5.8	2,881	22.5	2,881
Protective workers per 1,000 people	1.1	2,881	8.4	2,881
Healthcare workers per 1,000 people	3.4	2,881	29.7	2,881
Panel D. Family structure & social capital				
Share adults married	0.62	2,881	0.60	2,881
Share adults divorced	0.01	2,881	0.09	2,881
Share children in single-parent households	0.08	2,881	0.25	2,881
Average household size	4.4	2,881	2.5	2,881
Share religious adherents	0.43	2,878	0.53	2,881

Notes: This table reports the means of county characteristics and the number of counties with non-missing data, for the subset of counties used in our analysis. Some county characteristics are measured in 1930 or 1940 instead of 1920. See Appendix C for sources and variable construction details.

Table A11. State Public Education Statistics by Year

	K-12 teachers per 100 children (ages 5–18)		Average K-12 teacher salary (1999 dollars)		Public college instructors per 100 children (ages 5–18)	
	1920	2000	1920	2000	1930	2000
	(1)	(2)	(3)	(4)	(5)	(6)
United States	2.3	4.6	7,194	41,666	0.13	0.96
Alabama	1.5	5.4	4,020	36,689	0.08	1.12
Arizona	2.1	4.1	10,623	36,902	0.3	0.81
Arkansas	1.7	4.9	3,962	33,386	0.07	1.19
California	2.6	3.5	10,565	47,680	0.42	0.96
Colorado	3.0	4.7	7,716	38,163	0.26	1.34
Connecticut	2.4	5.3	9,336	51,780	0.07	0.70
Delaware	2.0	4.7	7,043	44,435	0.14	1.14
District of Columbia	2.5	5.0	11,287	47,076	0.08	0.38
Florida	2.3	3.7	4,302	36,722	0.09	0.82
Georgia	1.6	5.2	3,538	41,023	0.07	0.86
Idaho	3.1	4.6	7,741	35,547	0.16	0.88
Illinois	2.2	4.1	8,978	46,486	0.12	0.77
Indiana	2.3	4.5	8,007	41,850	0.11	0.93
Iowa	4.3	5.4	6,869	35,678	0.21	1.11
Kansas	3.4	5.2	6,321	34,981	0.22	1.35
Kentucky	1.8	4.3	3,430	36,380	0.11	1.21
Louisiana	1.5	5.0	6,005	33,109	0.08	1.00
Maine	3.6	6.5	5,008	35,561	0.15	0.78
Maryland	1.7	4.7	7,492	44,048	0.21	1.10
Massachusetts	2.0	5.6	10,482	46,580	0.05	0.76
Michigan	2.6	4.1	7,566	49,044	0.27	1.07
Minnesota	3.0	5.1	7,326	39,802	0.20	0.97
Mississippi	1.9	4.1	2,417	31,857	0.06	1.04
Missouri	2.3	5.4	6,620	35,656	0.12	0.97
Montana	4.9	5.4	7,957	32,121	0.17	1.12
Nebraska	4.0	5.7	6,354	33,237	0.18	1.39
Nevada	4.2	3.9	9,659	39,390	0.38	0.78
New Hampshire	2.8	5.6	6,304	37,734	0.23	0.77
New Jersey	2.1	5.0	10,648	52,015	0.04	0.80

Table A11. Public Education Statistics by State and Year (continued)

	K-12 teachers per 100 children (ages 5–18)		Average K-12 teacher salary (1999 dollars)		Public college instructors per 100 children (ages 5–18)	
	1920	2000	1920	2000	1930	2000
	(1)	(2)	(3)	(4)	(5)	(6)
New Mexico	2.4	3.9	6,669	32,554	0.21	1.20
New York	2.4	4.6	10,432	51,020	0.09	0.75
North Carolina	1.9	5.1	3,854	39,404	0.07	1.28
North Dakota	4.2	6.1	6,047	29,863	0.27	1.67
Ohio	2.4	5.0	9,037	41,436	0.13	0.90
Oklahoma	2.3	5.1	6,379	31,298	0.20	1.22
Oregon	4.0	3.4	7,226	42,336	0.40	1.37
Pennsylvania	1.8	4.2	7,641	48,321	0.05	0.93
Rhode Island	1.9	4.7	8,887	47,041	0.05	0.67
South Carolina	1.6	5.6	3,854	36,081	0.08	1.06
South Dakota	4.1	5.0	5,781	29,071	0.22	1.07
Tennessee	1.8	5.4	4,103	36,328	0.06	0.84
Texas	1.9	4.9	5,083	37,567	0.13	0.84
Utah	2.7	3.5	8,239	34,946	0.18	1.00
Vermont	3.2	5.3	5,540	37,758	0.23	1.26
Virginia	1.9	6.2	4,535	38,744	0.14	1.12
Washington	3.0	3.8	10,208	41,043	0.27	1.05
West Virginia	2.4	5.3	5,307	35,009	0.13	1.37
Wisconsin	2.3	5.4	7,600	41,153	0.19	1.27
Wyoming	4.4	6.4	7,218	34,127	0.24	1.56

Notes: This table reports measures of resources for public education by state and year for the continental United States. See Appendix C for sources and variable construction details.

Table A12. Impact of State Public Education Resources on the Rank-Rank Slope

	Dependent variable: Rank-rank slope				
	(1)	(2)	(3)	(4)	(5)
Panel A. 1910–1919 cohorts					
K-12 teachers per 100 children (ages 5–18)	-0.013*** (0.004)	-0.020*** (0.007)	-0.004 (0.012)	-0.003 (0.009)	-0.001 (0.008)
Average K-12 teacher salary (1,000s of 1999 dollars)	-0.025*** (0.002)	-0.022*** (0.002)	-0.004 (0.005)	-0.010* (0.005)	-0.012*** (0.004)
Public college instructors per 100 children (ages 5–18)	0.063 (0.053)	0.175** (0.083)	0.127 (0.175)	0.070 (0.152)	0.078 (0.106)
R^2	0.267	0.259	0.762	0.869	0.916
Partial R^2 (K-12)	0.267	0.245	0.005	0.063	0.105
Partial R^2 (post-secondary)	0.002	0.011	0.004	0.003	0.005
Panel B. 1982–1997 cohorts					
K-12 teachers per 100 children (ages 5–18)	0.010** (0.004)	-0.001 (0.007)	-0.004 (0.009)	-0.007 (0.008)	-0.008 (0.005)
Average K-12 teacher salary (1,000s of 1999 dollars)	-0.002*** (0.0005)	-0.002 (0.001)	-0.002 (0.001)	-0.002** (0.001)	-0.002** (0.001)
Public college instructors per 100 children (ages 5–18)	-0.023* (0.012)	-0.021 (0.024)	-0.063** (0.029)	-0.050* (0.027)	-0.030* (0.015)
R^2	0.063	0.024	0.744	0.856	0.908
Partial R^2 (K-12)	0.061	0.022	0.010	0.041	0.036
Partial R^2 (post-secondary)	0.004	0.003	0.036	0.065	0.027
Border sample		✓	✓	✓	✓
County pair fixed effects			✓	✓	✓
Bayesian shrinked estimates				✓	✓
County controls					✓
N	2,881	2,344	2,344	2,344	2,344

Notes: Column 1 shows the estimates from a regression of the county-level rank-rank slope on state resources for public education. Column 2 estimates the same specification as in column 1 but using the border-design sample (units of observation are counties-by-border pair). Columns 3–5 show the estimates from our state border regressions (equation 1). Columns 4 and 5 replace the baseline county mobility estimates with Empirical Bayes shrinked versions of those estimates (see text for details). Column 5 additionally includes controls for all county characteristics shown in Figure 4. In all regressions, counties are weighted by the inverse of the variance of the county mobility estimates. Robust standard errors are in parentheses, clustered at the state border level in columns 2–5. ***$p < 0.01$, **$p < 0.05$, *$p < 0.10$.

Table A13. Impact of State Public Education Resources on Bottom 40 to Top 60 Percent Mobility

	Dependent variable: Bottom 40 to top 60 percent mobility				
	(1)	(2)	(3)	(4)	(5)
Panel A. 1910–1919 cohorts					
K-12 teachers per 100 children (ages 5–18)	2.24** (0.90)	3.73** (1.57)	0.72 (2.01)	1.25 (1.76)	-0.01 (1.30)
Average K-12 teacher salary (1,000s of 1999 dollars)	5.96*** (0.27)	5.44*** (0.34)	1.29** (0.56)	2.00*** (0.56)	1.92*** (0.45)
Public college instructors per 100 children (ages 5–18)	26.83** (10.88)	16.37 (15.11)	-12.27 (22.68)	-8.85 (19.51)	-1.13 (11.81)
R^2	0.548	0.566	0.915	0.931	0.964
Partial R^2 (K-12)	0.514	0.564	0.035	0.102	0.143
Partial R^2 (post-secondary)	0.012	0.003	0.003	0.002	0.000
Panel B. 1982–1997 cohorts					
K-12 teachers per 100 children (ages 5–18)	2.15*** (0.52)	3.23*** (1.17)	1.35 (1.31)	1.39 (1.24)	0.23 (0.63)
Average K-12 teacher salary (1,000s of 1999 dollars)	0.64*** (0.06)	0.64*** (0.18)	0.33** (0.15)	0.39** (0.15)	0.19* (0.10)
Public college instructors per 100 children (ages 5–18)	3.48** (3.47)	3.17 (3.47)	8.00** (4.01)	6.91* (4.02)	4.74** (2.25)
R^2	0.156	0.158	0.821	0.851	0.928
Partial R^2 (K-12)	0.146	0.133	0.032	0.061	0.024
Partial R^2 (post-secondary)	0.006	0.004	0.047	0.051	0.040
Border sample		✓	✓	✓	✓
County pair fixed effects			✓	✓	✓
Bayesian shrinked estimates				✓	✓
County controls					✓
N	2,512	1,862	1,862	1,862	1,862

Notes: Column 1 shows the estimates from a regression of county-level bottom 40 to top 60 percent mobility on state resources for public education. Column 2 estimates the same specification as in column 1 but for the border-design sample (units of observation are counties-by-border pair). Columns 3–5 show the estimates from our state border regressions (equation 1). Columns 4 and 5 replace the baseline county mobility estimates with Empirical Bayes shrinked versions of those estimates (see text for details). Column 5 additionally includes controls for all county characteristics shown in Figure 4. In all regressions, counties are weighted by the inverse of the variance of the county mobility estimates. Robust standard errors are in parentheses, clustered at the state border level in columns 2–5. ***$p < 0.01$, **$p < 0.05$, *$p < 0.10$.

Table A14. Kitagawa-Oaxaca-Blinder Decompositions of Gaps in Upward Educational Mobility: Component Explained by Public Education Resources vs. Unexplained Component

	Kitagawa-Oaxaca-Blinder decompositions	
	Explained component	Unexplained component
	(1)	(2)
Panel A. Decomposing gap in upward mobility between South and non-South for 1910–1919 cohorts		
South vs. non-South (1910–1919 cohorts)	-15.8	
Reference group: South	-12.1	-3.7
% of gap	*76.4%*	*23.6%*
Reference group: non-South	-9.5	-6.3
% of gap	*59.9%*	*40.1%*
Panel B. Decomposing gap in upward mobility between South and non-South for 1982–1997 cohorts		
South vs. non-South (1982–1997 cohorts)	-3.7	
Reference group: South	-2.2	-1.6
% of gap	*58.4%*	*41.6%*
Reference group: non-South	-1.9	-1.8
% of gap	*51.7%*	*48.3%*
Panel C. Decomposing change in gap in upward mobility between South and non-South over time		
Change in South vs. non-South over time	12.1	
Reference group: 1910–1919 cohorts	5.8	6.3
% of gap	*47.8%*	*52.2%*
Reference group: 1982–1997 cohorts	-2.6	14.7
% of gap	*-21.3%*	*121.3%*

Notes: This table decomposes gaps in upward educational mobility, as measured by the expected rank at the 25th percentile, into the component that can be explained by gaps in state resources for public education (explained component) and a residual term capturing everything else (unexplained component) using Kitagawa-Oaxaca-Blinder (KOB) decompositions. Panels A and B show KOB decompositions of the South vs. non-South gap in each period, using either regions as reference groups. Panel C decomposes the change in the South vs. non-South gap over time into two components: (i) the component explained by changes in public education resources between the South and non-South, scaled by the South vs. non-South gap in KOB betas from Panel A or B, and (ii) a residual term. For example, when using the 1910–1919 cohorts as the reference group, the explained component is calculated as the South vs. non-South gap in KOB betas for the 1910–1919 cohorts (estimated separately by region), multiplied by the changes over time in the South vs. non-South gap in public education resources. Counties are weighted by the inverse of the variance of the county mobility estimates.