Changes in Upward Educational Mobility in the U.S. over the Twentieth Century and the Role of Public Education

By Martha J. Bailey[1,3], A. R. Shariq Mohammed[2], and Paul Mohnen[4]

[1] University of California-Los Angeles [2] Northeastern University [3] National Bureau of Economic Research
[4] Federal Reserve Bank of Atlanta

PRELIMINARY—DO NOT CIRCULATE OR CITE WITHOUT PERMISSION
[CLICK HERE FOR LATEST VERSION]

March 2025

Abstract

We present new facts on the evolution of intergenerational educational mobility in the U.S. over the 20th century. Using two large-scale datasets, we show that upward educational mobility has remained stable between the cohorts entering the labor market in 1940 and 2020, despite large changes in K-12 spending, the provision of public high schools, and the expansion of public colleges and universities. Yet this national perspective obscures considerable variation by subgroup and by county. Upward educational mobility has increased for almost all groups—Black Americans, Hispanic/Latinx Americans, and Asian Americans and for women of all groups. The exception is White men, for whom upward mobility has fallen. We also document considerable geographic convergence, reflecting rising upward mobility in the South and Great Plains and falling upward mobility in the West, Northeast and Industrial Midwest. Comparing counties across state borders, we find that state spending on K-12 education increased upward educational mobility in the early twentieth century, but state spending on public tertiary education increased upward mobility in the later 20th century.

Keywords: education, inequality, intergenerational mobility.

Contact information

Bailey is a Professor of Economics at the University of California, Los Angeles, marthabailey@g.ucla.edu; Mohammed is an Assistant Professor of Economics at Northeastern University, ab.mohammed@northeastern.edu; Mohnen is a Research Economist at the Federal Reserve Bank of Atlanta, paul.mohnen@atl.frb.org.

Acknowledgements

For helpful comments and discussions, we thank Dan Aaronson, Sandy Black, Marcus Casey, Dan Fetter, Sasha Indarte, Matt Staiger, Bryan Stuart, and participants at the ASSA, Atlanta Fed, Equitable Growth, NBER Development of the American Economy, RNIM, System Economic Heterogeneity Conference, UCLA, and University of Pennsylvania. We thank Alex Coblin, Jordan Cammarota, Kelsey Figone, Deniz Gorgulu, David Lee, Avinash Reddy, Joaquín Serrano, Eric Wang, and Mengying Zhang for excellent research assistance. We thank Caleb Floyd for his 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. We 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 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 Census Bureau. 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-P2548-R11099, CBDRB-FY24-0379).

In the late 19th and early 20th century, universal public education made the U.S. one of the most educated countries in the world (Goldin and Katz 2008). Building on its lead in primary education, the High School Movement catapulted the share of Americans graduating from high school from 10% in 1910 to over 50% by 1940 (Goldin 1998). After World War II, college enrollment increased by almost 50% by 1960, with undergraduate enrollments doubling again by 1970, and rising by another 2.3 times by 2019. These remarkable gains in educational attainment were accompanied by expansions in educational spending, with public elementary and secondary school spending rising by over 20 times between 1919 and 2019 (U.S. Department of Education, 2019) and public tertiary spending growing through the GI Bill and financial aid programs (Goldin and Katz 2008; Long 2013).

This paper focuses on how the transformation of American public education affected educational opportunity over the 20th century, both across different groups and communities. Soaring spending, together with the egalitarian principles of American public schools, may have leveled the playing field, allowing children from all backgrounds and communities to transcend the achievements of their parents. On the other hand, disproportionate spending on more advantaged groups or in more advantaged areas may have reinforced historical advantages, providing few opportunities for the less privileged to escape their parents' position. While American educational attainment increased over the 20th century, little research examines how the growth in public spending on schools translated into educational opportunities.

Our analysis examines upward educational mobility over 75 years, documenting these patterns by gender and race/ethnicity and county for cohorts born in the early 20th century and at the end of the twentieth century. For the early 20th century cohorts, we link nearly 1.5 million children, both boys and girls born from 1910 to 1919, and their parents in Social Security application records (SS-5) to the full-count 1940 Census—the first U.S. census to ask about completed education. In addition to allowing us to link both men and women (SS-5 records include women's birth, or "maiden," names), our data contain information on children's childhood county of residence via links to the 1920 Census, which allows us to document wide geographic variation in upward mobility in the midst of the High School

Movement. We compare this early 20th century snapshot with an in-depth look at educational mobility today for over 2.1 million children, born from 1982 to 1997, and their parents in restricted administrative data which links the 2000 Decennial Census and the 2001–2022 American Community Surveys (ACS).

We measure educational opportunity as *upward educational mobility*, or the likelihood that a child exceeds her parents' educational rank in the child's own cohort distribution. Rather than focusing on gains in educational achievement which could occur even if no child was upwardly mobile, this measure documents the likelihood that children shift *upwards* in the educational distribution. Similar to rank-rank measures of income mobility, this measure accounts for the fact that the distribution of educational attainment has changed over time and highlights how parents' relative educational attainment influences their children's relative educational attainment. Our focus is on the share of children born to parents in the bottom forty percent of the education distribution whose own education exceeds that percentile in their cohort's own education distribution. For our early 20th century cohorts, this captures the share of children who complete at least 10 years of schooling among those with parents completing less than 8th grade. For our late 20th century cohorts, this measure captures the share of children who have parents completing no more than 12th grade who themselves complete more than 12th grade.

Our first finding is that upward educational mobility has remained very stable across the 20th century, even as public investments in education have soared. Using our measure at the 40th percentile, upward mobility rose negligibly over the entire period, from 0.45 to 0.48. These findings differ from studies of intergenerational educational persistence which show that educational persistence fell slightly between cohorts born between 1920 and 1990 (Hertz et al. 2007; Hout and Janus 2011; Ferrie et al. 2021). They also differ from recent studies of intergenerational income mobility, which show marked drops in both intergenerational income persistence (Ward 2023; Jácome et al. 2025) and upward income mobility over the same period (Chetty et al. 2017).

Our second finding is that this longer-term stability masks considerable heterogeneity across groups and places. Although the upward educational mobility of White men and women was almost identical for the 1910–19 birth cohorts at around 50 percent, the upward mobility of White men fell to

0.40 as the upward mobility of White women rose to 0.57 for the 1982–97 birth cohorts. Similarly, the upward educational mobility of Black men and women, Hispanic/Latino men and women, and Asian American men and women rose over the twentieth century, with the larger gains favoring women in all cases. These differential changes across groups created a gender gap in upward educational mobility favoring women and resulted in a declining race gap between White Americans and all other groups, consistent with increasing access and rising returns to education for the historically less advantaged groups.

Our third finding is that large geographic differences in upward educational mobility narrowed dramatically over the twentieth century. In the early 20th century, the Northeast, Industrial Midwest, and West saw higher rates of upward educational mobility, whereas rates in the South were generally much lower. By the end of the 20th century, this landscape had changed, with the low-mobility areas in the South converging to higher levels and upward mobility rates falling in the rest of the country. Today, where children live matters much less for their educational opportunity than it did in the early 20th century.

A fourth finding is that differences in state educational resources played an important role in this geographic convergence, but their influence has fallen over time. Leveraging sharp discontinuities in state resources for education across neighboring counties in different states, we find that higher public spending on K-12 education had a significant positive effect on upward educational mobility in the early 20th century, consistent with the findings in Card et al. (2022) for Southern states. We find that a significant portion of regional inequality in the early 20th century can be attributed to differences in state-level investments in primary and secondary education. Over the 20th century, however, this relationship weakened, and geographic differences in K-12 education today explain a negligible share of differences in upward educational mobility. In contrast, the number of faculty per child at public colleges and universities did not predict upward mobility in the early 20th century, but these public investments in tertiary education have significant positive effects on upward mobility today.

This paper contributes to the literature by providing new evidence on the evolution of educational opportunity across the 20th century, uncovering important differences by race, sex and place. While numerous studies have documented more aggregated trends in relative educational mobility using survey and linked census-survey data (Couch and Dunn 1997; Hertz et al. 2007; Hout and Janus 2011; Torche 2015; Fletcher and Han 2019; Ferrie et al. 2021), our large-scale data allow a more comprehensive analysis of how the local geography of educational mobility evolved, revealing how investments in public education shaped educational opportunities over the 20th century.

I. Studies of U.S. Intergenerational Mobility over the 20th Century

A vast literature in economics and other social science disciplines examines the transmission of socioeconomic status across generations. Until recently, relatively few of these studies focused on the early 20th century U.S. or on educational mobility over time. For example, Corak's review of the literature (2006) documents that none of the 41 reviewed studies of intergenerational income mobility in the U.S. used data before 1980. This absence of research reflected considerable data limitations, as very few historical data sources tracked the income or educational outcomes of children and parents to study their relationship.

Occupational and Income Mobility in U.S. History

Recent innovations in data linking have allowed economic historians to link fathers and sons longitudinally to help fill these gaps (Ferrie 1996; Long and Ferrie 2013; Feigenbaum 2018; Song et al. 2020; Abramitzky et al. 2021; Collins and Wanamaker 2022; Ward 2023; Tan 2023). The earliest studies of this topic tended to focus on occupational mobility, because this information was collected in the U.S. Census every year since 1850. The first wave of these studies found that intergenerational occupational mobility in the U.S. was higher in the past compared to today. Ferrie's path-breaking work developed an approach to creating linked U.S. Census samples by focusing on men with uncommon names in the 1850 Census and linking them to males aged 10 and older in the 1860 Census (Ferrie 1996). In a later work, Long and Ferrie (2013) examine intergenerational occupational mobility in Great Britain and the United

States from 1850 onward, finding that while mobility rates were initially higher in the U.S., the U.S. lead over Great Britain was erased by the mid-20th century. They attribute this convergence to declining geographic mobility in the U.S.[1]

Feigenbaum (2018) finds high rates of intergenerational mobility among early-20th century Iowan men. Using linked samples of White men born in all states, Song et al. (2020) find that intergenerational occupational mobility declined after the mid-19th century, with most of this decline occurring in the second half of the 19th century. More recent work by Ward (2023), however, shows that accounting for race and measurement error in occupational status yields much lower rates of intergenerational occupational mobility for the early 20th century than previously believed. After adjusting for both issues, Ward (2023) finds that U.S. occupational mobility *declined* between 1850 and 1940, challenging the conventional narrative that economic opportunities were more equal in the past.

Bailey et al. (2020b) and Jácome et al. (2025) point out several critiques of studies using linked data. First, linked samples are not representative of the U.S. population. Individuals with common names are exceedingly hard to link with the limited information in the Census (which led Ferrie in his early work to omit them entirely). Women are also omitted from most linked samples because they change their names at marriage (Ferrie 1996; Helgertz et al. 2022). In addition, Black Americans are highly underrepresented. For example Ward (2023) links only 1 percent of Black men, and Collins and Wannamaker (2022) link only 3–5 percent of Black men. Second, census enumerators appear to have recorded occupations with substantial error, perhaps owing to transitory changes in occupations (Solon 1992; Ward 2023).

To address these limitations, Jácome et al. (2025) pool multiple smaller retrospective surveys to characterize the long-run evolution of intergenerational mobility for men and women over the 20th century. Their measure of parental income predicted from father's occupation was retrospectively

[1] Research in sociology has long examined intergenerational stratification using occupational status (Blau and Duncan 1967; Hauser et al. 1975; Hauser 1980; Hout 1983). In the U.S., studies of occupational mobility find significant mobility gains for cohorts born in the early 20th century, with little change thereafter (Hauser et al. 2000). See Florencia (2015) for a recent review.

reported by adult children in surveys. They then regress children's own income on their parent's predicted income to recover estimates of the intergenerational income persistence in terms of the elasticities and rank-rank regressions. Accounting for multiple potential sources of measurement error and biases, they find that—for samples including both men and women and nonwhite individuals—intergenerational mobility rose between cohorts born in the 1910s and cohorts born in the 1940s before stagnating or declining later in the 20th century. These findings serve as an important reference for our estimates of educational mobility in the early 20th century.[2]

Studies focused on the later twentieth century use longitudinal data such as the Panel Survey of Income Dynamics (PSID) or National Longitudinal Surveys (NLS) and focus on cohorts born between the 1950s and 1970s (Mazumder 2016; Bratberg et al. 2017; Davis and Mazumder forthcoming). These studies have the advantage of observing parents and children at similar ages, which helps alleviate concerns about life-cycle bias (Haider and Solon 2006; Lee and Solon 2009; Nybom and Stuhler 2016). In addition, they contain multiple observations of parents' and children's incomes which allows the analysis to adjust for transitory income shocks (Solon 1992; Zimmerman 1992; Mazumder 2005). Administrative data from the Internal Revenue Service (IRS) provide an alternative data source to longitudinal surveys. These data have the advantage of very little attrition relative to surveys. However, they are only available after 1996, which limits their cohort coverage to children born in the 1970s and observed at fairly young ages in the Census (whereas their parents are observed at much older ages). IRS data suggest that intergenerational income mobility has been relatively stable since the 1990s, with rank-rank slopes around 0.30, but accounting for the short period of observation may raise these estimates considerably (Mazumder 2016). The literature concludes that the intergenerational elasticity estimate of mobility is around 0.50 today and, in survey data, appears to have increased since 1980 (Davis and Mazumder forthcoming) although they appear stable in IRS data (Chetty et al. 2014a).

[2] They also document the relationship between fathers' predicted income and respondents' educational attainment as a plausibility check on their main findings. These results show that the relationship between father's predicted income and respondents' educational attainment weakened between cohorts born in the 1910s and 1950s.

More recently, several studies have documented geographic differences in intergenerational income mobility using large-scale linked samples. Using IRS data, Chetty et al. (2014b) document considerable variation in income mobility across places in the U.S. for late 20th century cohorts. Subsequent studies have built analogous estimates for early 20th century cohorts using linked census data or linked census-IRS data and concluded that the geography of income mobility has significantly changed over time (Connor and Storper 2020; Massey and Rothbaum 2021; Tan 2023).

Educational Mobility in the 20th Century U.S.

Education is a key mechanism for intergenerational occupational and income mobility (Blau and Duncan 1967; Sewell and Hauser 1975; Hout and DiPrete 2006), but research on intergenerational educational mobility has been more limited by data. Unlike information on occupation, educational attainment was not measured in the Census until 1940.[3] Several studies have used modern survey data with retrospective reports of parents' education to fill this gap. Using the 1984–1989 waves of the PSID, Couch and Dunn (1997) report intergenerational educational correlations of 0.40 to 0.43 for cohorts born around the mid-20th century. Fletcher and Han (2019) exploit longitudinal survey data from the National Center for Education Statistics that covers three sets of cohorts of 8th or 10th graders in 1980, 1988 and 2002 (born in the 1960s, 1970s, and 1980s). Proposing a measure of educational mobility that accounts for the discrete nature of educational attainment, they find that intergenerational educational persistence declined slightly between the first and the last set of cohorts in their sample.

Other studies exploit surveys that include cohorts born over a longer period. Hertz et al. (2007) use the 1994–2000 International Social Survey Programme and find that educational mobility has been relatively stable between cohorts born in the 1930s and cohorts born in the 1970s according to intergenerational educational persistence estimates, while intergenerational correlations exhibit a slight decline over time. Hout and Janus (2011) use the General Social Survey (GSS) and similarly find that

[3] One exception is the 1915 Iowa Census, which measured respondents' educational attainment. Feigenbaum (2018) exploits this data to generate a sample of father-son pairs linked across the 1915 Iowa Census and the 1940 Federal Census. He documents an educational persistence estimate of 0.26 for Iowan men born in 1898–1912.

intergenerational educational persistence has been fairly stable for Americans born between 1910 and 1980.

Two limitations of using retrospective surveys relate to (1) their small sample sizes and (2) longevity bias (survivors may be more educated and potentially differ in their mobility because longevity is positively correlated with education). To address these issues, Hilger (2017) uses large sample sizes in the 1940 to 2010 Decennial Censuses and focuses on children ages 26–29 who reside with their parents at the time of enumeration. Unlike the survey-based studies, he finds an inverted U-shaped pattern: intergenerational educational persistence declined between cohorts born in the 1910s and 1940s, before rising for subsequent cohorts. The downside of conditioning on older children who still co-reside with their parents is that they may also be a highly selected subset of the population, as co-residence rates fall with age. Moreover, to the extent that selection into co-residence changes over time, trends in educational mobility might be misleading.

More recent work by Ferrie et al. (2021) by-passes issues related to co-resident samples by linking respondents across the full-count 1940 Census, Current Population Surveys (CPS) spanning 1973–1990, and the long-form 2000 Census and 2001–2015 American Community Surveys (ACS) to measure educational mobility. They report an intergenerational educational persistence coefficient (correlation) of 0.37 (0.45) for individuals born in 1922–1940 and a coefficient (correlation) of 0.36 (0.42) for individuals born in 1955–1990. That is, consistent with survey-based studies, they find that intergenerational educational persistence has been relatively stable over the 20th century, while intergenerational educational correlations have declined slightly.[4]

In a study closely related to our analysis of the early 20th century, Card et al. (2022) adapt Hilger's co-resident sample strategy to focus on younger children ages 16–18 in the 1940 Census (born in

[4]While most of the literature on education mobility has focused on trends in intergenerational educational persistence or correlations, some studies have also documented trends in the share of children whose educational attainment exceeds that of their parents ("absolute" education mobility). Using GSS data, Hout and Janus (2011) find that absolute education mobility rose between the 1910 and 1950 birth cohorts, and then declined between the 1950 and 1980 birth cohorts. Using various surveys, Hilger (2017) documents a similar inverse U-shaped pattern for Black Americans, while he finds that absolute education mobility for White Americans was initially stable before declining as well.

1922–1924) who live with their parents. Note that because children are younger than 18, the selection into co-residence is less severe than in Hilger's analysis. Moreover, the large sample sizes of the full-count 1940 Census allow them to estimate rates of upward educational mobility by county and by race. Since many of those children have not yet completed their education, they define "upward educational mobility" as having at least 8 years of education conditional on one's parents having at most 6 years of education.[5] They find that White Americans had higher rates of upward mobility than Black Americans on average and document significant regional disparities in upward mobility, which they attribute to differences in school resources and teacher salaries using a cross-state border design. The magnitudes of their estimates suggest that around two thirds of the Black-White gap in upward educational mobility can be explained by racial disparities in school resources.

Althoff et al. (2025) leverage co-resident samples to study the evolution of educational mobility between 1850 and 1950. Since educational attainment is unobserved prior to 1940, they propose a new measure of relative educational mobility that relies on literacy status as a proxy for latent human capital. Their semiparametric framework allows them to isolate the role of mothers' and fathers' human capital. They find that educational mobility rose over this time period according to their measure, driven by a declining influence of mothers' human capital. They attribute this trend to rising investments in public education.

In summary, retrospective surveys and linked data consistently show that intergenerational educational persistence remained stable in the second half of the 20th century. However, little is known about how *upward* educational mobility changed and whether patterns in upward mobility varied across race and sex. In addition, studies have documented how intergenerational income mobility has varied across place in the early and later 20th century, but the changing landscape of upward educational mobility and its links to investment in public schools are not well documented. Our paper contributes to

[5] Appendix Figure A1 plots average years of education at various ages between 16 and 35 for cohorts born between 1905 and 1995, pooling data from the 1940–2000 decennial censuses and the 2001–2022 ACS. Assuming that educational attainment at age 35 approximates completed education and interpolating the observed means within age categories across cohorts, it implies that 16, 17 and 18-year olds in 1940 would roughly go on to complete an additional 2, 1.5 and 1 years of education after 1940 on average, respectively.

the literature by constructing a new large-scale sample linking millions of individuals across administrative and census records to document these changes. This approach allows us to focus on cohorts who have mostly completed their education by 1940 and side-step the limitations associated with co-resident samples or aging cohorts, while also retaining the large sample sizes necessary to characterize the geography of intergenerational educational mobility at the county level in the early 20th century. These data enable us to uncover differential trends by sex and race/ethnicity groups as well as across counties and link them to differences in local public investments in education, providing the most comprehensive evidence to date on the geography of upward educational mobility in the U.S. and the role of local public educational investments over the 20th century.

II. Measuring the Intergenerational Education Mobility over the 20th Century: Data and Methodology

New data—including a novel historical dataset covering the 1910–19 birth cohorts and restricted administrative data from the Census Bureau covering the 1982–97 birth cohorts—are key to this paper's contribution. This section describes the construction of our administrative datasets, retrospective survey data used to benchmark our estimates, and the state- and county-level characteristics used to analyze the determinants of educational mobility. The final subsection describes our methodology for estimating upward educational mobility.

A. Linking SS-5 Records to the 1920 and 1940 Censuses

The 1940 Census was the first decennial census to ask respondents about their educational attainment, and offers the first opportunity to measure educational mobility for early 20th century cohorts at large scale. Ideally, one would link all adult children to their (surviving) parents in the 1940 Census. However, family relationships are only observed for individuals who are part of the same household. Therefore, we link children to their parents—regardless of whether they co-reside or not—via an intermediate set of records that contain parent-child relationships: Social Security application (Form SS-5) records. These records come from the public version of the Social Security Numerical Identification

(Numident) Files released by the National Archives and Records Administration, and 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. In practice, these records cover the near-universe of individuals who died between 1988 and 2007 (Goldstein and Breen 2022).

Panel A of Appendix Figure A2 plots the cohort coverage of SS-5 records by sex and race, as implied by population counts in the 1940 Census. Coverage is low for individuals born before 1910 as most of those individuals died prior to 1988. Coverage then sharply rises as each successive cohort is increasingly likely to have survived until 1988 before gradually declining as cohorts are increasingly likely to have died after 2007. To construct our linked sample we focus on the 1910–19 cohorts for three reasons: (1) they include some of the cohorts with the highest coverage rates, (2) the likelihood that at least one of their parents is still alive in 1940 is high, and (3) they are aged 21–30 in 1940 and have therefore mostly completed their education.[6] Despite the uneven cohort coverage, SS-5 records have excellent geographic coverage. This can be seen in Appendix Figure A2B, which plots state of birth shares by sex and race for our cohorts of interest in SS-5 records against corresponding shares in the 1940 Census.

Each SS-5 record lists the number holder's full name, sex, race, date of birth, and birthplace. Crucially, it also lists the full names of each of the applicant's parents, which not only provides parent-child linkages, but also allows us to infer women's birth (or "maiden") names via their father's surname. In conjunction with their married name, this allows us to find women in the 1940 Census regardless of their marital status at the time.[7] Panel A of Appendix Figure A3 provides an overview of how we construct our linked sample. We start by linking children to the 1940 Census based on their name, age and birthplace information. We then link siblings together in the SS-5 data based on parent names. This allows us to link children and their parents as families to households in the 1900–1930 Censuses

[6] Appendix Figure A1 suggests that the educational gains past age 20 were modest for individuals born in 1910–19.
[7] Women typically specify their married name as their surname if they are married when initially filing Form SS-5, and file another Form SS-5 if they subsequently get (re-)married. As a result, women have 1.8 entries in the SS-5 data on average, and 94% of women have at least one surname that differs from their father's surname.

based on parent names and children's name, age, and birthplace. From the 1900–1930 Censuses, we extract information on parents' year and place of birth, which we then leverage to link parents to the 1940 Census in three complementary ways: (1) as families, (2) as couples, and (3) as individuals. Linking parents in multiple ways helps us find them in 1940 irrespective of their household structure.

We make all the linkages above using the supervised machine learning approach described in Bailey et al. (2023). The advantages of this approach are that it allows us to enjoy the benefits of hand-linking at scale and control the tradeoff between making more links and fewer false links. For more details on how we make all the links described above using this methodology, see Appendix A. Appendix Table A2 shows that we are ultimately able to link around 2 million children born in the U.S. in 1910–1919 *and* at least one of their parents to the 1940 Census, for a match rate of 24.4% at 95-percent precision (relative to human trainers). Since we are interested in geographic differences in educational mobility, we further restrict the sample to individuals linked to their childhood household in 1920 which allows us to estimate educational mobility rates by 1920 county of residence, resulting in a final sample size of around 1.5 million children. Thanks to the availability of married and birth names in SS-5 records, women make up 50% of our linked sample. We construct inverse propensity score weights (IPWs) that re-weight the sample to match the sex-by-race-by-cohort distribution in the 1940 population using the procedure from Bailey et al. (2020a). These weights correct for the fact that earlier cohorts are underrepresented in SS-5 records and the fact that Black Americans are underrepresented because they are harder to link.

B. Linking SS-5 Records to the 2000 Census and American Community Surveys

We construct an analogous linked sample for late 20th century cohorts using restricted Census Bureau data. Specifically, we leverage linkages between SS-5 records, the 2000 Census, and the 2001–2022 waves of the ACS.[8] Panel B of Appendix Figure A3 shows an overview of how we construct this sample. Our starting point are children aged 3–18 who are living with at least one biological parent in

[8] The Census Bureau has access to the universe of SS-5 records and can leverage information from various administrative sources to assign unique personal identifiers called Protected Identify Keys (PIKs) to records, which can be used to link individuals across records (Wagner and Layne 2014).

2000. We start by linking those children to their SS-5 records to obtain their demographic information (year of birth, sex, race), which tends to be more reliably reported in administrative data. For the subset of children observed in the 2000 Long-Form Census (1-in-6 sample of the population), which contains a broad set of questions including education, we can directly observe parental education. We then link children to the most recent wave of the 2001–2022 ACS in which they are 25 years or older to obtain their own education.[9] We augment this initial sample by linking the remaining subset of children *and* their parents in the 2000 Short-Form Census, which covers the entire U.S. population but only includes basic information (demographics, location, household composition), to the 2001–2022 ACS.

Our final sample contains 2.1 million children born in the U.S. in 1982–1997 for which we can observe own education and education for at least one parent. Our linked sample is well-balanced on gender (49% women), but is skewed towards earlier cohorts since later cohorts can appear in fewer waves of the ACS by construction (due to 25+ age restriction). We again construct IPWs that re-weight the sample to match the sex-by-race-by-cohort distribution in the 2000 population. Note that we observe children's county of residence in 2000, which is what we use to characterize the geography of educational mobility for these cohorts.

C. Surveys with Retrospective Information on Parents

To benchmark the educational mobility estimates from linked samples, we use two national surveys that contain information on respondents' educational attainment and retrospective information on their parents' educational attainment: (1) the 1977–2022 waves of the General Social Survey (Davern et al. 2024), and (2) the 1987–1988 wave of the National Survey of Families and Households (NSFH; Bumpass et al. 2017). We selected these surveys because they are representative of the U.S. population, cover both men and women from our cohorts of interest, and contain retrospective information on

[9] Appendix Figure A1 shows that gains in educational attainment between ages 25 and 35 are around 0.65 years on average for the 1982–1987 cohorts.

respondents' father's and mother's education. We restrict samples to U.S. born individuals aged 25 or older at the time of the survey and pool the GSS and NSFH to increase precision.[10]

D. State and County Resources for Public Education

To study the impact of investments in public education on educational mobility, we leverage data from a variety of sources. We digitize state-level statistics on K-12 education for school year 1919–20 and statistics on public colleges and universities for school years 1928–30, which we aggregate at the state and county level, 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 and county-level statistics on K-12 education for school year 1999–2000 and public post-secondary institutions for school year 2000–01 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). We also construct state and county-level population characteristics for the historical period using microdata 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. Appendix B provides more details on the construction of these local characteristics.

E. Measuring Intergenerational Educational Mobility over Time

There are many ways to measure intergenerational mobility (Deutscher and Mazumder, 2023). A common measure of educational mobility in the literature is *intergenerational educational persistence*, or the slope coefficient from a regression of one's own years of education on a constant and a measure of parental education. However, this measure has two limitations. First, it is non-directional, in that lower persistence can reflect greater upward mobility, greater downward mobility, or both. Second, when comparing different subgroups or locations, this measure obscures level differences since they are absorbed by the constant.

[10] We adjust the survey weights by first normalizing them to sum to one within each survey and then multiplying them by the share of observations coming from each survey so that more weight is placed on the larger GSS sample.

In this paper, our primary measure of mobility is *upward educational mobility*, which we define as reaching the top 60 percent of the cohort-specific education distribution conditional on one's parents being in the bottom 40 percent of the cohort-specific parental education distribution.[11] Throughout the paper, we measure parental education as the maximum of fathers' and mothers' years of education. This measure has a directional interpretation, preserves level differences, and because we rank individuals relative to their peers it does not mechanically reflect growth in educational attainment over time. While this measure will miss mobility within the bottom 40 percent and top 60 percent, the upshot is that it is less sensitive to the fact that we do not always observe completed years of education in the data.

Our choice of the bottom 40 vs. top 60 percent is not innocuous: Because the size of different education bins changes over time, in defining a time-consistent mobility measure we are constrained by the configuration of the bins in the data. This is illustrated in Appendix Figure A4. Panel A plots the distribution of education categories by cohort group, as measured in the nearest decennial census or ACS wave in which individuals are aged 31–40 (except for the 1982–97 cohorts, where individuals are aged 25–40 in 2022). Panel B plots the distribution of parental education categories for the same cohort groups as in Panel A, as measured in the nearest decennial census in which children are young and living with their parents. We chose the 40th percentile as the cutoff since for both the 1910–19 and 1982–97 cohorts we can approximately divide the education and parental education distributions into the bottom 40 percent and top 60 percent based on observed values. For children born in 1910–19, upward mobility is defined as the probability of attaining at least 10 years of education conditional on parents having at most 7 years of education. For children born in 1982–97 and their parents, upward mobility is defined as the probability of attaining at least 13 years of education conditional on parents having at most 12 years of education.[12]

[11] This measure is conceptually similar to the measure of upward educational mobility in Card et al. (2022), which can be interpreted as the probability of reaching the top three quartiles conditional on starting in the bottom quartile for their cohorts of interest (1922–1924).

[12] For a more complete picture of education mobility, the appendix presents results for children whose parents are in the top 60 percent of the parental education distribution, which we will refer to as *educational persistence at the top*. This measure can be thought of as the inverse of downward educational mobility (i.e.,. ending up in the bottom 40% of the education distribution conditional on parents being in the top 60%).

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

We begin our analysis by characterizing how educational mobility has evolved at the national level over the 20th century. Figure 1A presents national estimates of educational persistence, for comparison with other studies, and upward educational mobility for early and late 20th century cohorts.[13] Because linked samples may be unrepresentative and generate biased estimates, we begin by benchmarking the estimates based on the linked samples we constructed against corresponding estimates from nationally-representative survey data (Bailey et al. 2020, Jácome et al. 2025). For both educational persistence and upward educational mobility, the linked sample-based estimates align well with the survey-based estimates, falling within the 95% confidence intervals associated with the latter. In the remainder of the analysis, we focus on the estimates from our linked samples, which are much more precisely estimated.

Comparing the educational persistence estimates for early and late 20th century cohorts in the left panel of Figure 1A, our first finding is that educational persistence has declined from around 0.40 for the 1910–19 cohorts to around 0.3 for the 1983–1994 cohorts. Appendix Figure 5, which plots educational persistence for the cohorts in between, shows that educational persistence initially declined from 0.40 to 0.35, then hovered around 0.35 for several decades, before declining to 0.3 more recently. The stability in educational persistence for most of the 20th century is consistent with prior findings in the literature. For example, Ferrie et al. (2021) report an estimate of 0.37 for the 1922–1940 cohorts and 0.36 for the 1955–1990 cohorts using linked census-survey data.

While educational persistence has been described elsewhere in the literature, less is known about upward educational mobility. Examining the right panel of Figure 1A, our second finding is that upward educational mobility has increased only slightly over the 20th century, from 0.446 for the 1910–19 cohorts to 0.476 for the 1982–97 cohorts. Our appendix shows that educational persistence at the top has

[13] For educational persistence, we present estimates for the 1983–1994 cohorts based on an earlier disclosure in which we used a slightly different sample. The construction of this alternative sample is analogous to our baseline sample for the 1982–1997 cohorts, but only leverages the 2001–2019 waves of the ACS (sample size of around 1.3 million individuals).

also been fairly stable over time, declining by less than a percentage point (from 0.777 to 0.769) (Appendix Figure A6, panel A). Therefore, despite significant increases in the level of education in the population and massive changes to the U.S. educational system over the 20th century (Goldin and Katz 2008), absolute educational mobility has barely budged.

This finding is consistent with several possibilities. One is that greater educational investments raised educational attainment overall but did not affect upward educational mobility, primarily because those investments benefitted children from all backgrounds similarly. Alternatively, greater educational investments may have led to greater gains in upward mobility among some groups but declining upward mobility in others. Figure 1B tests these hypotheses by examining upward educational mobility for different gender and race/ethnicity subgroups. The relative stability in upward educational mobility at the national level has masked important shifts across subgroups. The first notable pattern is that the gender gap in upward educational mobility has grown significantly over the 20th century. Among White Americans, the gender gap was around 1 percentage point in favor of women in the early 20th century while today it stands at 17 percentage points. Among Black Americans, the initial gap was larger at 10 percentage points, but it has nearly doubled to 17 percentage points. A similar story has unfolded among Hispanic and Native Americans, with the gap rising by 12 and 10 percentage points respectively. The only exception are Asian Americans for which the gender gap has remained relatively stable at around 21 percentage points. Averaging across all race/ethnicity groups, the gender gap has grown from 3 to 16 percentage points over the 20th century. The growing gender gap in upward educational mobility mirrors the growing unconditional gender gap in college attainment (Goldin et al. 2006).[14]

The second notable pattern is that racial gaps in upward educational mobility have significantly narrowed over time. The Black-White gap among men was close to 30 percentage points in favor of White men in the early 20th century, reflecting significant disparities in educational access and quality in

[14] Columns 2 and 3 in Appendix Table 3 shows that while women were slightly more mobile than men in the early 20th century in terms of educational persistence (0.37 vs. 0.418), this gap has since disappeared (0.289 vs. 0.286). That is, while the rate at which parental education gets transmitted to children is similar for men and women in the recent period, women do significantly better than men in absolute terms (even after conditioning on starting in the bottom 40% of the parental education distribution).

this period (Margo 1990). By the end of the 20th century, this gap was still sizable but much smaller at 9 percentage points. Among women, the Black-White gap fell from 18 to 8 percentage points. Pooling across men and women, the Black-White gap in upward educational mobility has declined from 23 to 8 percentage points over the 20th century. In contrast, Collins and Wanamaker (2022) find that the gap in upward *income* mobility between White and Black men has remained persistent for Americans born between 1900 and 1990. This highlights the fact that educational mobility does not necessarily lead to income mobility, especially when upward income mobility is limited by discrimination and other labor market institutions.

A few other patterns in Figure 1B are worth noting. First, while Hispanic Americans had upward mobility rates comparable to Black Americans in the early 20th century, they have fully closed the gap on White Americans. Second, while Asian Americans had upward mobility rates that were comparable to those of White Americans in the early 20th century, they now enjoy the highest upward mobility rates of any group by quite some margin (0.67 and 0.78 for Asian American men and women, respectively). In contrast, Native Americans have the lowest mobility rates of any group among late 20th century cohorts (0.23 and 0.36 for Native American men and women, respectively).[15] Lastly, every gender-by-race/ethnicity group has experienced an improvement in upward educational mobility over the 20th century with the exception of Native American men and, notably, White men, whose mobility has fallen by nearly 10 percentage points. The decline among White men, who make up a large share of the population, has offset the gains among all other groups, driving the appearance of no change over time.

We also find qualitatively similar, though more muted, patterns for educational persistence at the top (Appendix Figure A6, Panel B). Nevertheless, there are a few noteworthy differences: (1) educational persistence at the top has increased significantly among Black women but remained flat among Black men, and (2) Hispanic Americans have only partially caught up to White Americans in terms of educational persistence at the top.

[15] The ranking of race/ethnicity groups in terms of upward educational mobility for the 1982–1997 cohorts is largely consistent with the corresponding ranking in terms of upward income mobility documented in Chetty et al. (2020) for similar cohorts (1978–1983).

IV. The Changing Landscape of Upward Educational Mobility over the 20th Century

We next characterize how the geography of upward educational mobility has evolved over the 20th century using county of residence in the 1920 Census for the 1910–1919 cohorts and in the 2000 Census for the 1982–1997 cohorts. For comparability over time, we restrict our sample to counties with "stable" boundaries between 1920 and 2000 using Eckert et al's (2020) longitudinal county crosswalk (we define a county as having "stable" boundaries if at least 95% of its area under 1920 boundaries overlaps with its area under 2000 boundaries and vice versa), leaving us with around 95% of all counties in 1920 and 2000.

Disclosing county-level estimates for the 1982–1997 cohorts required following Census Bureau guidelines to protect the confidentiality of respondents. Using the methodology from Chetty and Friedman (2019), we infused unweighted county estimates for the 1982–97 cohorts with additive Gaussian noise and set the "privacy parameter" to 6. This methodology results in less than 0.5% of the total variation in the noise-infused county estimates occurring due to noise. County observation counts were separately infused with noise using the Census Bureau's implementation of the Discrete Gaussian Mechanism (Cannone et al. 2022).

Figure 2 displays maps of county-level estimates of upward educational mobility for the 1910–19 cohorts (Panel A) and the 1982–97 cohorts (Panel B), restricting the sample to counties with at least 20 observations. Counties in Panel A are shaded according to their decile, and counties in Panel B are shaded using the same cutoffs as in Panel A for comparability. Several key patterns stand out. First, similar to the spatial distribution in intergenerational income mobility (Tan 2023), large regional disparities in upward educational mobility marked the early 20th century. Counties in the South had low upward educational mobility rates (mostly below 0.3), while counties in the West had high upward mobility rates (mostly above 0.5). Elsewhere in the Northeast and in the Industrial Midwest, upward educational mobility was also relatively high. These regional patterns are broadly similar to those in Card et al. (2022) (correlation=0.64), although our measure of upward educational mobility, cohort, and sample differ.

Second, the variation in upward educational mobility fell sharply over the 20th century (Panel B), with the more uniform shading revealing regional convergence. While the interquartile range for upward mobility rates at the county level spanned 0.25 points in the early 20th century, this range narrowed considerably to 0.11 by the end of the 20th century, representing a decline in the variance across counties of nearly 70%. In the past, upward educational mobility depended a great deal on where you grew up. But today, geographic convergence means that place matters much less for upward educational mobility (see Appendix Figure A7).

In addition to the variance across place falling, the geographic pattern shifted, which is captured by the fairly low correlation of 0.22 between the county estimates for early vs. late 20th century cohorts. Panel C summarizes these changes in upward educational mobility over 75 years, showing gains in green and losses in red. Notably, much of the gap between the South and the rest of the country has closed, and the West lost its advantage in upward mobility, now hovering close to the national average. Similarly, upward educational mobility deteriorated in the Industrial Midwest, mirroring its declining economic fortunes. Along with the South, counties in the Great Plains (IA, KS, MN, NE, ND, SD) experienced significant gains in upward mobility over time and now enjoy some of the highest rates in the country.

These changes are broadly similar to the geography of upward income mobility documented in the literature. Using slightly different samples and methods to approximate income, Connor and Storper (2020), Massey and Rothbaum (2021), and Tan (2023) all find that the West, Northeast and Industrial Midwest had higher upward income mobility than the South in the early 20th century.[16] For the contemporary period, using linked tax records for individuals born in 1980–1982, Chetty et al. (2014b) document that upward income mobility is highest in the Great Plains and lowest in the Southeast.[17] Yet

[16] 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.

[17] For the geography of educational persistence at the top, we similarly find that large regional disparities and evidence of regional convergence over time (Appendix Figures A8 and A9). The regional patterns are broadly similar, with correlations of 0.6 and 0.5 with county estimates of upward educational mobility for early and late 20th century cohorts, respectively. Here we highlight a few notable differences. In the early 20th century, AZ and NM ranked relatively low in terms of upward educational mobility (0.36–0.39) but relatively high in terms of educational

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 1910–19 cohorts and the county estimates in Massey and Rothbaum (2021) or Tan (2023) for income mobility is around 0.6; the correlation between our estimates for upward educational mobility for the 1982–97 cohorts and the income mobility estimates in Chetty et al. (2014b) is around 0.50.[18] These imperfect correlations underscore the fact that there are many places in the U.S. characterized by high rates of upward educational mobility but low rates of upward income mobility and vice versa, both in the past and today.

V. The Changing Role of Public Education over the 20th Century

What caused these changes in upward educational mobility? Explanations often relate to factors that influence the *ability* of children from low-educated backgrounds to reach higher parts of the educational distribution. This includes changing family and social environments and, crucially, access to schools. In this section, we focus on the role of state educational policies affecting spending on public education. Consistent with the notion that differences in state policy may play an important role, we find that around 50% of the cross-county variation in the historical period and 35% in the contemporary period is explained by state level factors (i.e., state fixed effects).

We consider four key measures of educational access/quality: the number of K-12 teachers per 100 children ages 5 to 18, average salaries of K-12 teachers, and the number of instructors at 2-year and 4-year public colleges and universities per 100 children aged 5 to 18.[19] We focus on access to public post-secondary institutions, which accounted for 50% of total enrollment in 1931 and 73% of total enrollment in 2021 (Goldin 2006; U.S. Department of Education 2023a), since the vast majority of state

persistence at the top (0.84–0.89). Nowadays, Northeastern states like CT, MA, NJ, and NY lead the nation in terms of educational persistence at the top (as opposed to Great Plain states), and some states like DC, MD, and VA are characterized by average or low upward educational mobility but high educational persistence at the top. Strikingly, DC has the lowest rate of upward mobility in the country (0.36) but one of the highest rates of educational persistence at the top (0.8). This implies that the educational outcomes of children from low vs. high-educated backgrounds may be influenced by a different set of local factors.

[18] We thank Jonathan Rothbaum and Hui Ren Tan for sharing their estimates with us.

[19] In the contemporary period, 2-year institutions are primarily community colleges that offer Associate's degrees and other post-secondary credentials. In the historical period, we define 2-year institutions as encompassing "junior colleges," which primarily focused on providing vocational education, and "normal schools," which focused on the training of public school teachers.

funding for higher education is allocated to these institutions. They are also more likely to serve students from lower socioeconomic status backgrounds since they tend to be less selective and have lower tuition rates than private institutions, especially for in-state students. We distinguish between 2-year and 4-year public institutions, because 2-year public colleges have lower tuition rates, different admission policies, and serve different groups of students.[20] While 2-year public institutions only accounted for 5% of total enrollment in 1931 (Goldin 2006), by 2021 that share had risen to 24% (U.S. Department of Education 2023a), highlighting how 2-year public institutions now constitute an important segment of the U.S. higher education landscape. We consider salaries of K-12 teachers since better-quality K-12 education may produce students that are better prepared for post-secondary education. Teacher salaries have also been found to be an important determinant of upward educational mobility in the historical period (Card et al. 2022).

Figure 3 provides suggestive evidence that these state educational investments were crucially important in the early 20th century. The number of kindergarten, primary and secondary teachers (panel A), teacher salaries (panel B), and the number of instructors at 2-year and 4-year public colleges and universities (panels C and D) all appear strongly positively correlated with upward educational mobility at the state level, although these relationships weaken or disappear entirely by the end of the 20th century. However, interpreting these relationships as causal is complicated by the fact that state investments in public education may be correlated with a host of other factors that independently affect educational mobility. Panel A of Figure 4 shows that county-level demographic characteristics, economic inequality, residential segregation, level of economic development, provision of public services, and family structure and social capital are strongly associated with upward educational mobility, both in the past and today.[21]

[20] In 2022–23, 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 2023b).

[21] The coefficients in Figure 4A come from separate, cohort-specific linear regressions of county upward educational mobility on a contemporaneous county characteristic, where both variables are normalized to have a mean of 0 and standard deviation of 1. Counties are weighted by the number of children in the bottom 40% in the relevant sample, and 95% confidence intervals are based on heteroskedasticity robust standard errors. See Appendix B for details on these county characteristics and Appendix Table A7 for descriptive statistics.

Measures of economic development such as higher labor-force participation rates, manufacturing employment shares, or income per capita are positively associated with upward mobility in both periods, whereas greater inequality, higher shares of farm households, and a larger Black population are negatively associated with upward mobility. Because these differences are also strongly correlated with investments in public education (joint significance test of all Figure 4 characteristics with county-level teacher salaries, F-statistic=335 and p-value=0 for the early 20th century cohorts; F-statistic=120 and p-value=0 for the late 20th century cohorts), the correlations in Figure 3 may not reflect the true causal effect of state investments in public education on upward educational mobility.

To ameliorate concerns about these confounding factors, we follow the approach in Card et al. (2022) and evaluate the role of state educational resources using a state-border research design. By comparing neighboring counties on opposite sides of a state border, the design relies on the fact that counties are part of the same local labor market and similar in many other dimensions due to their proximity, educational resources differ sharply due to state policies. (See Panels A and B in Appendix Figure A10 for the counties used in this analysis and a visual description of sharp differences in average K-12 teacher salaries across state borders). We implement this design by estimating the following linear regression, separately for early and late 20th century cohorts,

$$y_{c(s)p} = \boldsymbol{Z}_s'\beta + \delta_{p(c)} + \varepsilon_{c(s)p}, \quad (1)$$

where y denotes upward educational mobility in county c located in state s and part of county pair p, and $\delta_{p(c)}$ is a set of county-pair fixed effects. We assign counties to county pairs using the Census Bureau's 2010 county adjacency crosswalk, and continue to use counties with stable boundaries and at least 20 children in the bottom 40% (as in Figure 2A). 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. The coefficients of interest relate to $\boldsymbol{Z}$, a vector of state educational resources, including the number of K-12 teachers per 100

children ages 5 to 18, the average K-12 teacher salary in thousands of 1999 dollars, and the number of instructors at 2-year and 4-year public colleges and universities per 100 children ages 5 to 18.[22]

A necessary condition for β to be interpreted as the causal effect of these investments on upward educational mobility is that unobserved county-level factors are not systematically related to state educational investments after controlling for county-pair fixed effects, or $\text{cov}(\varepsilon_{c(s)p}, \boldsymbol{Z}_s \mid \delta_{p(c)})=0$. The assumption is fundamentally untestable, so we instead test whether observable county characteristics in Figure 4, $\boldsymbol{X}_c$, are correlated with educational resources, $\boldsymbol{Z}_s$, or $\text{cov}(\boldsymbol{X}_{c(s)p}, \boldsymbol{Z}_s \mid \delta_{p(c)})=0$. Panel B of Figure 4 shows the estimates from variants of equation (1) where upward educational mobility is replaced by an element of $\boldsymbol{X}_c$ as the dependent variable and the key explanatory variable $\boldsymbol{Z}_s$ is the state average K-12 teacher salary. (See Appendix Figure A11 for alternative versions of panel B that use other measures of educational investments as explanatory variables.) Consistent with the research design being internally valid, we find that the relationship of a broad set of demographic, economic, public services, family structure, and social capital measures with the state average teacher salary—*within* county pairs—cannot be distinguished from zero. Out of 24 characteristics, only one is statistically different from zero in the early 20th century and one in the later 20th century, which is no more than expected by chance. Note that when comparing across all counties along borders—and not within county pairs—*most* estimates are statistically different from zero, as captured by the hollow markers and 95-percent confidence intervals with dashed lines. Therefore, consistent with the internal validity of the research design, counties on opposite sides of state borders are balanced in a wide set of county characteristics. Yet, as shown in the second row of panel B, in both periods county-level teacher salaries differ meaningfully across state borders within county pairs, which is key to our research design.[23]

Table 1 shows our main regression results. As suggested by Figure 3, state resources for public education are strongly associated with county upward educational mobility for the 1910–1919 cohorts

[22] 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).

[23] Appendix Figure A11 shows that this is not only true for K-12 teacher salaries, but also our other measures of public education.

(column 1) and the 1982–1997 cohorts (column 5), even though the explanatory power of those resources appears to have diminished over time (with the R^2 declining from 0.45 to 0.17). The subsequent columns explore the extent to which these estimates are causal. First, columns 2 and 6 present estimates from a variant of equation 1 that excludes county pair fixed effects, essentially capturing the association between county upward educational mobility and state public education resources in the subsample of border counties. The fact that the estimates are similar in columns 1 and 2 and columns 5 and 6 suggests that border counties are not dissimilar to the country as a whole—a finding consistent with the external validity of this research design. Columns 3 and 7 use the county border design, adding county pair fixed effects to the specification. In contrast to the associations in columns 1 and 2, comparing counties across state borders reveals that K-12 teachers had no significant effect on upward educational mobility in either period. Average teacher salaries, however, had a positive and significant effect on upward educational mobility in the early 20th century which vanished by the end of the century. To compare the magnitudes of the estimates, a $1,000 increase in average teacher salaries historically (which is equivalent to a 14% increase over the mean of $7,194 in this period) led to a 1.4 percentage point increase in upward mobility, but the same 14% increase ($5,833 relative to the mean of $41,666) led to a statistically insignificant 0.9 percentage point increase in the more recent period.[24] Conversely, tertiary educational investments mattered little for upward educational mobility in the early 20th century but mattered much more for late 20th century cohorts. Our estimates imply that a 14% increase in 2-year public college and university instructors per 100 children relative to the mean raises upward educational mobility by around 0.5 percentage points.[25] Controlling the non-education county characteristics in Figure 4 in columns 4 and 8 does not change this conclusion, except to increase the magnitude (and decrease the standard error) of the effect of instructors at 4-year public colleges for late 20th century cohorts.

[24] The finding that state average teacher salaries are strong predictors of upward educational mobility in the historical period is in line with the findings in Card et al. (2022), who establish this fact for a similar set of cohorts using a similar research design (they focus on the South and additionally instrument for average teacher salaries using state minimum salary laws).

[25] Appendix Figure A12 and Appendix Table A8, which present analogous results to Figure 4 and Table 1 for educational persistence at the top, reveal qualitatively similar patterns for children with high-educated parents. One notable difference is that state investments in tertiary education do not have a significant effect on educational persistence at the top in the contemporary period.

Table 2 summarizes the findings of a Kitagawa-Oaxaca-Blinder (KOB) decomposition of the gap in upward educational mobility between Southern counties and non-Southern counties. For the 1910–1919 cohorts, upward educational mobility was 19.5 percentage points lower in the South than in the rest of the country, and observed differences in public expenditures on education explain around 86% of the gap (panel A). To minimize the role of other factors correlated with public educational resources, column 5 replaces the cross-sectional estimates with the border-design estimates (Table 1, column 3). This exercise suggests that state public education resources alone account for around 32% of the total gap in upward mobility between the South and the rest of the U.S. For the 1982–1997 cohorts (panel B), upward educational mobility was only 7 percentage points lower in the South than elsewhere and public education resources explain around 8% using the border-design estimates (Table 1, column 7). In summary, while differences in public educational resources explain gaps between the South and rest of the U.S. in upward mobility, their importance relative to other factors fell over the course of the 20th century.

VI. Conclusion

Using two new large-scale datasets linking children's educational attainment to that of their parents, we have documented new facts on the evolution of intergenerational educational mobility in the U.S. over the 20th century. At the national level, despite significant improvements in educational attainment and massive investments in public education, upward educational mobility has remained fairly stable over the last 75 years. However, this relative stability has masked a significant rise in the gap between women and men as well as a significant narrowing of the gap between White and Black Americans. The geography of educational mobility has also changed over the 20th century. Parts of the country that were above the national average in the early 20th century, like the West, Northeast, and Industrial Midwest, have experienced falling upward educational mobility, while parts of the country that were below the national average or around it, like the South and Great Plains, have experienced rising mobility. The result has been significant regional convergence over time.

Using a state border design, we find that state resources for primary and secondary education—and teacher salaries in particular—were an important determinant of upward educational mobility in the early 20th century, whereas public resources for tertiary education played little role. Consequently, large regional disparities in K-12 educational resources produced large disparities in upward educational mobility during this period. In contrast, the role of state investments in both 2-year and 4-year public colleges and universities grew increasingly important for upward educational mobility over the 20th century—a finding consistent with the overall increase in the level of education and the fact that access to college remains a hurdle for those from disadvantaged backgrounds, particularly as the cost of college has grown.

Finally, we document that the importance of all public educational resources—both K-12 and tertiary education—fell over the course of the 20th century. Large increases in state resources for education have resulted in near universal access to primary and secondary education, which likely reduced the marginal returns to those resources in terms of educational *attainment*. Increases in resources for tertiary education have not produced offsetting gains. (Note, we do not consider gains in the quality of education, which could be affected by spending but is not easily measured.) The result is that state resources for public education explain a much smaller share of the smaller differences in upward educational mobility today.

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.

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

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., 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.
Bound, John and Sarah Turner. 2002. "Going to War and Going to College: Did World War II and the G.I. Bill Increase Educational Attainment for Returning Veterans?: *Journal of Labor Economics* 20(4), 784–815.
Blau, Peter M. and Duncan Otis Dudley. 1967. *The American Occupational Structure*. John Wiley & Sons.
Bound, John and Sarah Turner. 2002. "Cohort crowding: How resources affect collegiate attainment." *Journal of Labor Economics* 20(4): 784–815.
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.
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, David Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, and Jimmy Narang. 2017. "The fading American dream: Trends in absolute income mobility since 1940." *Science* 356(6336): 398–406.
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.
Corak, Miles. 2006. "Do Poor Children Become Poor Adults? Lessons from a Cross-Country Comparison of Generational Earnings Mobility." *Research on Economic Inequality* 13(1): 143–188.
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.
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. Forthcoming. "The Decline in Intergenerational Mobility after 1980." *The Review of Economics and Statistics.*

Deutscher, Nathan and Bhashkar Mazumder. 2023. "Measuring Intergenerational Income Mobility: A Synthesis of Approaches." *Journal of Economic Literature* 61(3): 988–1036.

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 no. 26770, National Bureau of Economic Research, Cambridge, MA.

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.

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. 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. 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*. 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.

Haider, Steven and Gary Solon. 2006. "Life-Cycle Variation in the Association between Current and Lifetime Earnings." *American Economic Review* 96(4): 1308–1320.

Hauser, Robert M. 1980. "Some Exploratory Methods for Modeling Mobility Tables and Other Cross-Classified Data." *Sociological Methodology* 11: 413–458.

Hauser, Robert M., John N. Koffel, Harry P. Travis, and Peter J. Dickinson. 1975. "Temporal Change in Occupational Mobility: Evidence for Men in the United States." *American Sociological Review* 40(3): 279–297.

Hauser, Robert M., Warren John Robert, Min-Hsiung Huang, and Wendy Y. Carter. 2000. "Occupational Status, Education, and Social Mobility in the Meritocracy." In Arrow, Kenneth J., Samuel Bowles, and Steven N. Durlauf (Eds.). *Meritocracy and Economic Inequality*, Princeton University Press.

Helgertz, Jonas, Joseph Price, Jacob Wellington, Kelly J. Thompson, Steven Ruggles, and Catherine A. Fitch. 2022. "A new strategy for linking U.S. historical censuses: A case study for the IPUMS multigenerational longitudinal panel." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 55(1): 12–29.

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. 1983. *Mobility Tables*. Sage Publications.
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 Duncan, Greg J. and Richard J. Murnane (Eds.). *Whither opportunity? Rising Inequality, Schools, and Children's Life Chances*, 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.
Lee, Chul-in and Gary Solon. 2009. "Trends in Intergenerational Income Mobility." *The Review of Economics and Statistics* 91(4): 766–772.
Lefgren, Lars J., Jaren C. Pope, and David P. Sims. 2022. "Contemporary State Policies and Intergenerational Income Mobility." *Journal of Human Resources* 57(4): 1107–1146.
Long, Bridget Terry. 2013. "Supporting Access to Higher Education." In Bailey, Martha J. and Sheldon Danziger (Eds.). *Legacies of the War on Poverty*, 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.
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
Margo, Robert A. 1990. *Race and Schooling in the South, 1880–1950: An Economic History*. University of Chicago Press.
Massey, Catherine and Jonathan Rothbaum. 2021. "The Geography of Opportunity over Time." Working paper.
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.
Nybom, Martin and Jan Stuhler. 2016. "Heterogeneous Income Profiles and Lifecycle Bias in Intergenerational Mobility Estimation." *Journal of Human Resources* 51(1): 239–268.
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
Sewell, William H. and Robert M. Hauser. 1975. *Education, Occupation, and Earnings: Achievement in the Early Career.* Academic Press.
Solon, Gary. 1992. "Intergenerational Income Mobility in the United States." *American Economic Review* 82(3): 393–408.
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.
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. 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 control and level of institution: 1970 through 2021." *Digest of Education Statistics*, Table 303.25.

U.S. Department of Education, National Center for Education Statistics. 2023b. "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.

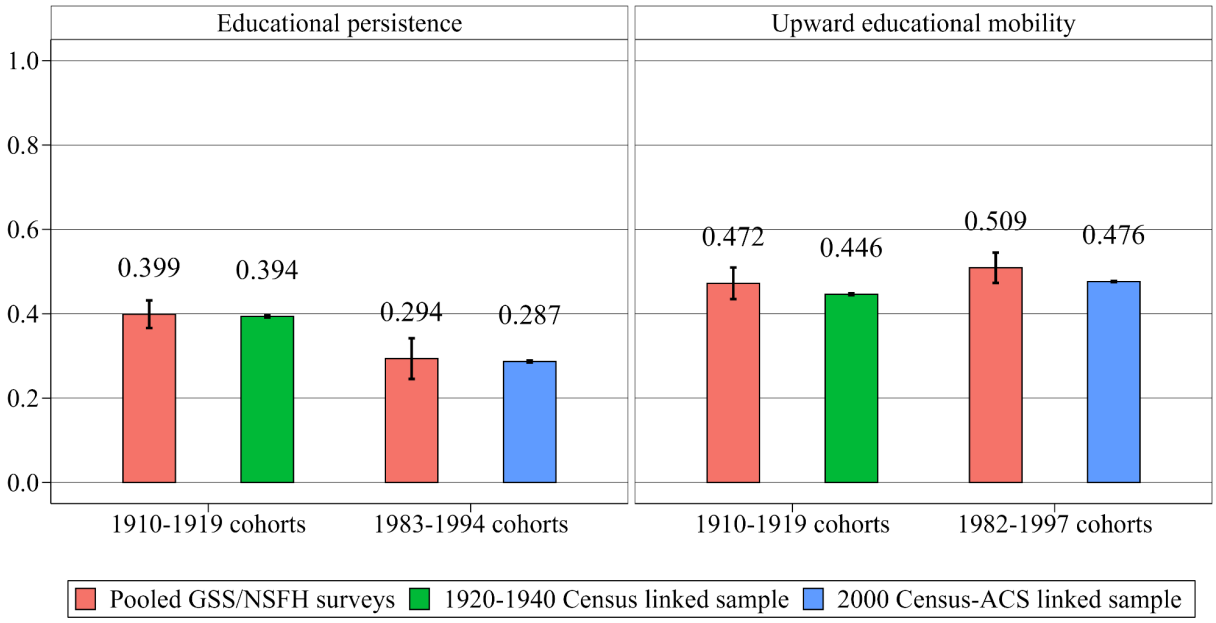
Wagner, Deborah and Mary Layne. 2014. "The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' (CARRA) Record Linkage Software," U.S. Census Bureau Center for Administrative Records Research and Applications Working Paper #2014-01.

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

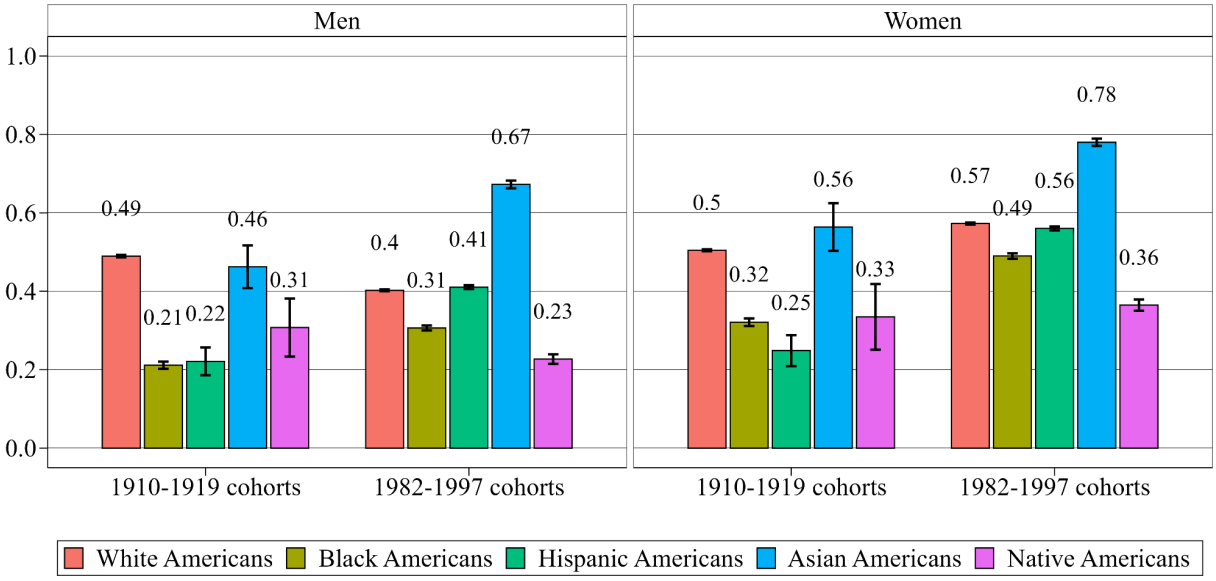
Zimmerman, David J. 1992. "Regression Toward Mediocrity in Economic Stature." *American Economic Review* 82(3): 409–429.

Figure 1. Intergenerational Educational Mobility over the 20th Century

Panel A. Educational Persistence and Upward Educational Mobility by Sample



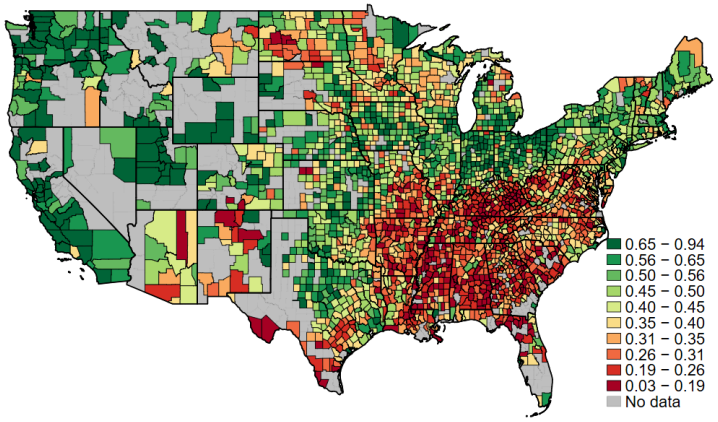
Panel B. Upward Educational Mobility by Sex and Race in Linked Samples



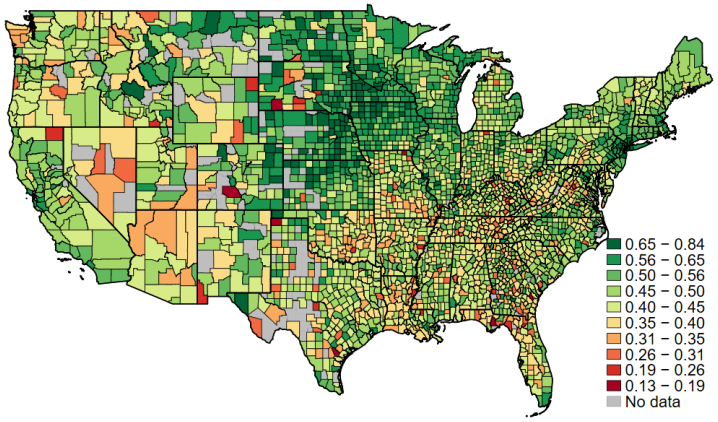
Notes: This graph plots estimates of educational persistence and upward educational mobility by cohort and by sample (panel A), and estimates of upward educational mobility by cohort and by sex and race based on the 1920-1940 Census and 2000 Census-ACS linked samples (panel B). Observations in the pooled surveys are weighted by normalized sampling weights. Observations in the 1920-1940 Census and 2000 Census-ACS linked samples are weighted by inverse propensity scores that re-weight the sample to match the sex-by-cohort-by-race distribution in the 1940 and 2000 population, respectively. 95% confidence intervals based on robust standard errors. See Appendix Tables A3 and A4 for detailed results.

Figure 2. The Geography of Upward Educational Mobility by Childhood County of Residence

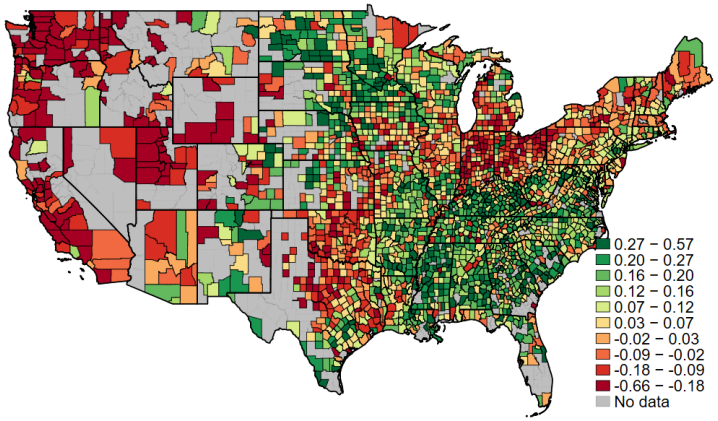
Panel A. 1910–1919 Cohorts



Panel B. 1982–1997 Cohorts

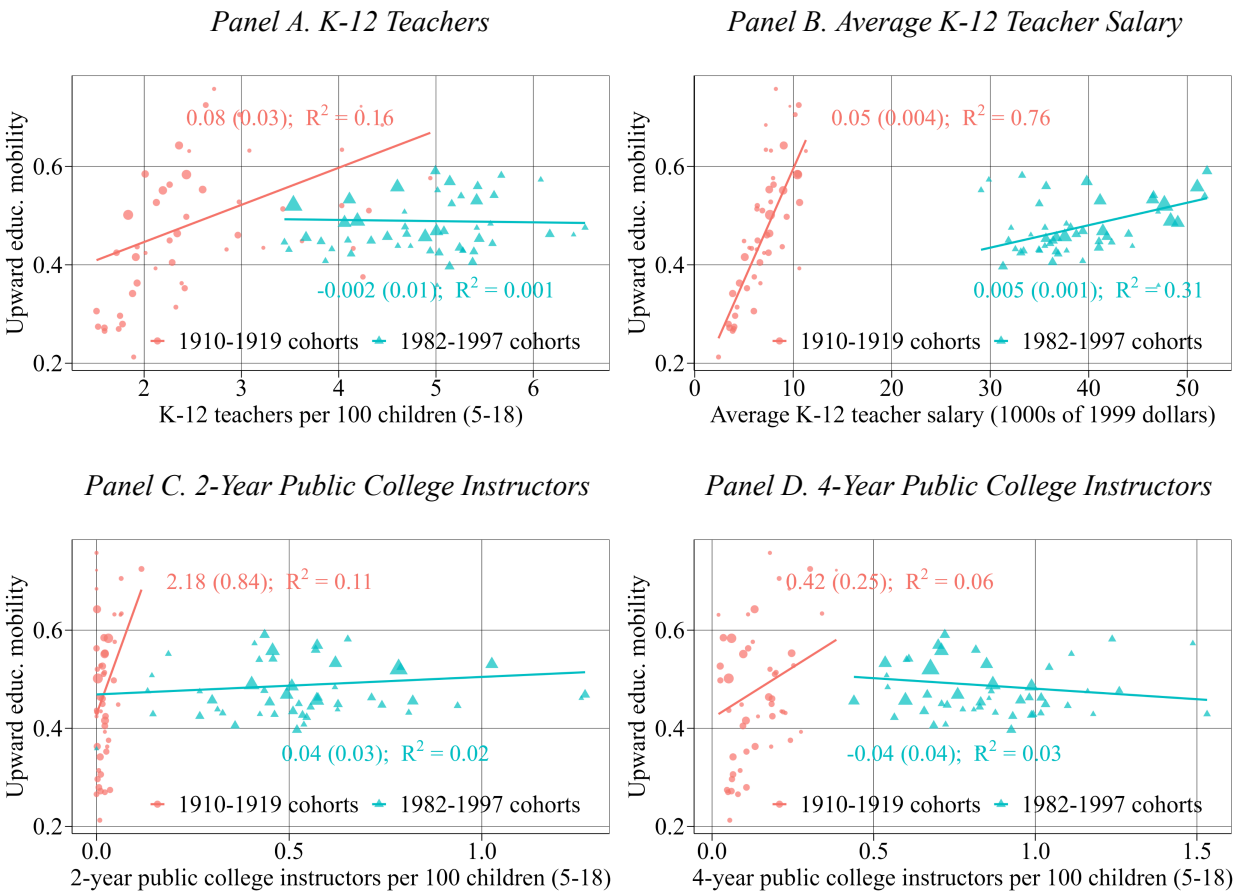


Panel C. Change Between 1910–1919 and 1982–1997 Cohorts



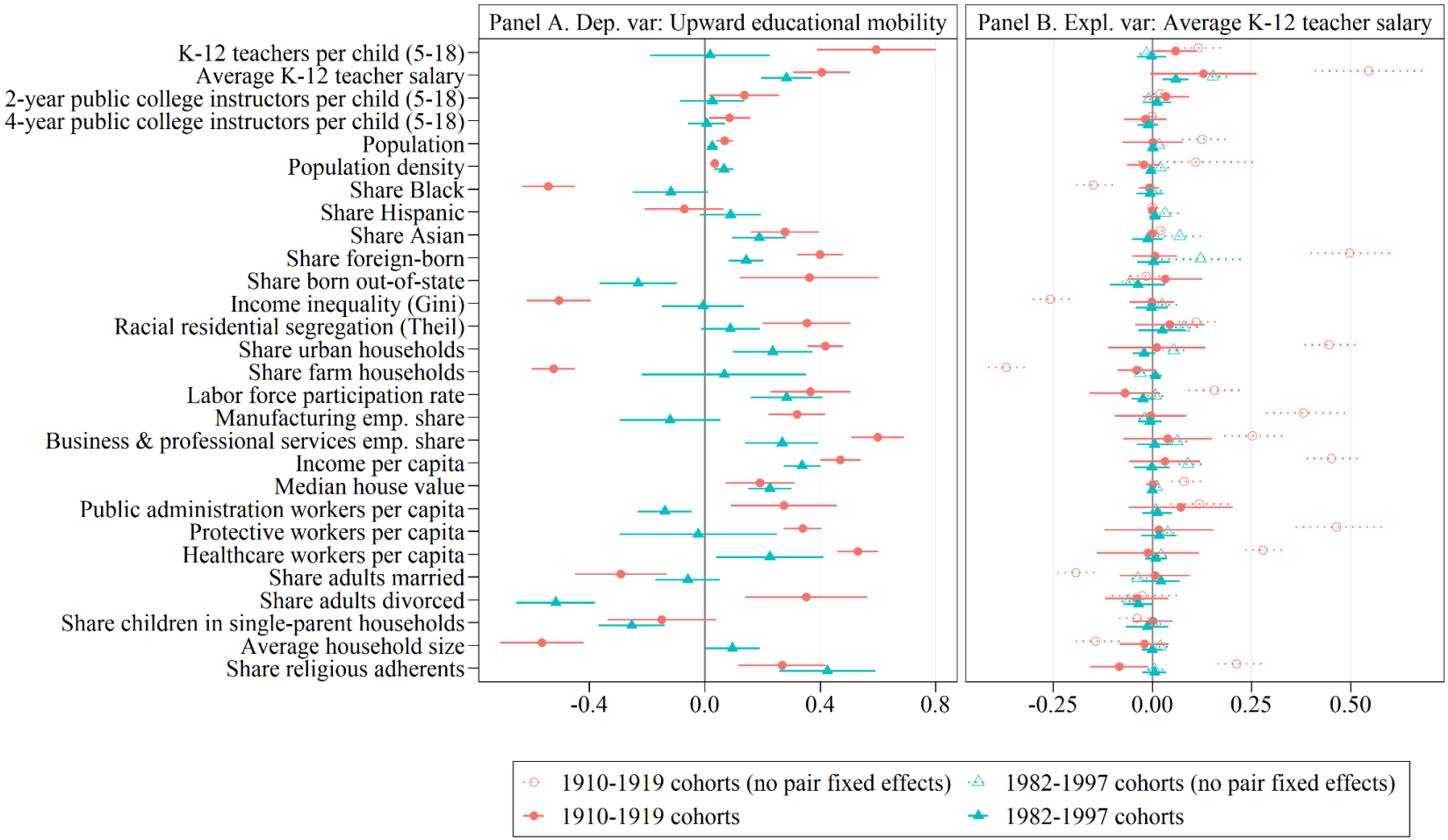
Notes: These graphs plot upward educational mobility by childhood county of residence in 1920 (1910–19 cohorts) or 2000 (1982–97 cohorts). Sample restricted to counties with at least 20 observations and "stable" boundaries between 1920 and 2000 (see text). In panels A and C, counties are shaded according to which decile bin they fall in. In panel B, counties are shaded using the decile bin cutoffs from panel A for comparability.

Figure 3. Upward Educational Mobility vs. Resources for Public Education Resources at the State Level



Notes: These graphs plot upward educational mobility by childhood state of residence for the 1910–1919 and 1982–1997 cohorts against contemporaneous state-level public education statistics (number of K-12 teachers per 100 children ages 5–18 in panel A; average K-12 teacher salary in 1000s of 1999 dollars in panel B; number of 2-year and 4-year public college instructors per 100 children ages 5–18 in panels C and D). Fitted regression lines weighting observations by the corresponding number of children in the bottom 40% in the relevant sample (corresponding estimated slope coefficients and associated robust standard errors as well as regression R^2 shown).

Figure 4. Correlations between Upward Educ. Mobility and County Characteristics and Teacher Salaries Balance Checks



Notes: In panel A, each dot/triangle is the (weighted) correlation coefficient between upward educational mobility and a county characteristic (indicated in row label), separately by cohort. In panel B, each dot/triangle is the coefficient from a variant of equation (1) where upward educational mobility is replaced with a county characteristic (indicated in row label) as the dependent variable and the key explanatory variable is the state-level average K-12 teacher salary (either excluding or including county pair fixed effects). In all regressions, counties are weighted by the number of children in the bottom 40% in the relevant sample. All county characteristics are normalized to have a weighted mean of 0 and standard deviation of 1 within each cohort. 95% confidence intervals based on robust standard errors (clustered at the state border level in panel B).

Table 1. The Changing Relationship Between Upward Educational Mobility and Resources for Public Education

	Dependent variable: Upward educational mobility							
	1910–1919 cohorts				1982–1997 cohorts			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
K-12 teachers per 100 children (ages 5–18)	0.016**	0.033**	0.012	-0.002	0.024***	0.030***	0.012	0.001
	(0.008)	(0.014)	(0.021)	(0.017)	(0.004)	(0.011)	(0.013)	(0.007)
Average K-12 teacher salary (1,000s of 1999 dollars)	0.045***	0.040***	0.014**	0.012***	0.006***	0.006***	0.002	0.001
	(0.003)	(0.003)	(0.006)	(0.004)	(0.001)	(0.002)	(0.002)	(0.001)
2-year public college instructors per 100 children (ages 5–18)	-0.150	-0.384	0.069	-0.283	0.051***	0.052**	0.054*	0.032**
	(0.329)	(0.595)	(0.903)	(0.518)	(0.010)	(0.025)	(0.030)	(0.015)
4-year public college instructors per 100 children (ages 5–18)	0.348***	0.272**	-0.033	0.065	-0.008	-0.016	0.027	0.046**
	(0.103)	(0.133)	(0.278)	(0.193)	(0.014)	(0.031)	(0.034)	(0.021)
County pair fixed effects			✓	✓			✓	✓
County controls				✓				✓
R^2	0.450	0.435	0.860	0.911	0.170	0.169	0.812	0.895
Partial R^2 (educ. vars.)	–	–	0.049	0.038	–	–	0.048	0.039
p-value (educ. vars.)	0.000	0.000	0.068	0.027	0.000	0.000	0.235	0.107
N	2,512 (counties)	1,878 (county × pairs)			2,512 (counties)	1,878 (county × pairs)		

Notes: This table shows regressions of county upward educational mobility on state-level public education resources, separately by cohort. In columns 1 and 5, the sample consists of all counties in the U.S. with stable boundaries and at least 20 children in the bottom 40% in the relevant sample. In all other columns, the sample consists of border counties-by-pairs. The regressions in columns 3 and 7 include county pair fixed effects, and the regressions in columns 4 and 8 additionally control for the non-education county controls listed in Appendix Table A7 (Panels B–E). In all regressions, counties are weighted by the corresponding number of children in the bottom 40% in the relevant sample. Robust standard errors in parentheses (clustered at the state border level in columns 2–4 and 5–8). *** 1%, ** 5%, * 10% significance.

Table 2. Decompositions of Gap in Upward Educational Mobility Between South and Non-South Into Component Explained by Public Educational Resources and Unexplained Component

	Kitagawa-Oaxaca-Blinder decompositions				Border-design estimates
	Reference group: Non-South		Reference group: South		
	Explained $\beta_N \times (X_S - X_N)$	Unexplained $X_S \times (\beta_S - \beta_N)$	Explained $\beta_S \times (X_S - X_N)$	Unexplained $X_N \times (\beta_S - \beta_N)$	Explained $\beta^\star \times (X_S - X_N)$
	(1)	(2)	(3)	(4)	(5)
Panel A. 1910–1919 cohorts (total gap = -0.195)					
Total across all measures	-0.167	-0.028	-0.167	-0.028	-0.061
[95% confidence interval]	[-0.226, -0.107]	[-0.088, 0.032]	[-0.202, -0.131]	[-0.067, 0.012]	[-0.104, -0.019]
Constant	–	-0.087	–	-0.087	–
K-12 teachers	0.000	0.113	-0.039	0.152	-0.008
Average K-12 teacher salary	-0.158	-0.051	-0.118	-0.092	-0.054
2-year public college instr.	0.001	0.018	-0.005	0.024	0.000
4-year public college instr.	-0.010	-0.021	-0.005	-0.026	0.001
Panel B. 1982–1997 cohorts (total gap = -0.070)					
Total across all measures	-0.016	-0.055	-0.021	-0.049	-0.005
[95% confidence interval]	[-0.024, -0.007]	[-0.066, -0.044]	[-0.041, -0.002]	[-0.073, -0.025]	[-0.030, 0.019]
Constant	–	0.188	–	0.188	–
K-12 teachers	0.014	-0.200	-0.002	-0.183	0.005
Average K-12 teacher salary	-0.031	-0.057	-0.020	-0.068	-0.011
2-year public college instr.	0.001	-0.014	0.001	-0.013	0.001
4-year public college instr.	0.000	0.027	0.000	0.027	0.000

Notes: Columns 1–4 present Kitagawa-Oaxaca-Blinder (KOB) decompositions of the gap in upward educational mobility between counties in the South and counties outside the South, separately by cohort and using different reference groups. Column 5 presents a variant of the KOB explained component that replaces the KOB regression coefficients (beta's) with the border-design estimates from Table 1 (columns 3 and 7). 95% confidence intervals for the explained and unexplained components of the gap (in brackets) are based on robust standard errors obtained using the delta method, taking average resources for public education by region as fixed.

Online Appendix for

Changes in Upward Educational Mobility in the U.S. over the Twentieth 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.

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 parent 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 parent 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 in 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 Table A2 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 Table A2 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.).[26] 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 Table A2 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 in making manual linking decisions for a random sample of primary records. Column 5 in Table A2 lists the information displayed to human "trainers" to aid them in their decisions (this is also what the algorithm

[26] 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.

will "see" when making linking decisions). We generally try to provide as much information as possible while ensuring all the information fits on a computer screen.[27]

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 Table A2 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% of the training data) that the computer is allowed to learn from and a "hold-out" sample (25% or 30% 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

[27] 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).

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% in the training sample.[28] Column 7 in Table A2 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 (i.e., individuals can have multiple siblings). Therefore, during the training process trainers are allowed to select multiple links, and a random forest model is used to model the probability that each 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% precision rate. After generating links in the full sample, we implement a series of post-processing steps to eliminate 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 meaningfully higher match probability than all the other links and dropping the remaining ones, or dropping all the links if none objectively stands out. Another type of conflict that arises as part of link B is when two name variations of the same female

[28] The only exceptions are the intermediate links C and D (links to 1900–1930 Censuses) that are used for subsequent links, for which we impose a 97-percent precision rate.

SS-5 record are linked to two different women in the 1940 Census. We use a similar strategy to eliminate such conflicts (either selecting one link that stands out or dropping all of them).

Appendix B. Details on the Construction of State and County Characteristics

State and County Resources for Public Education

State-level data on 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." County-level K-12 teachers are approximated by the number of individuals aged 16 or older in the labor force and who fall under the "teachers (not elsewhere classified)" IPUMS occupation code in the 1920 Decennial Census (the correlation between the state-level analog of this measure and state-level data on K-12 teachers from the Bureau of Education is 0.87). County-level average K-12 teacher salaries are approximated by the average wage and salary income of individuals who fall under the "teachers (not elsewhere classified)" IPUMS occupation code in the 1940 Decennial Census (the correlation between the state-level analog of this measure and state-level data on average K-12 teacher salaries from the Bureau of Education is 0.87).

State and county-level data on instructors at 2-year and 4-year 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 resources for post-secondary education in 1930 rather than 1920 because it is closer to the college-going ages of our cohorts of interest (1910–1919), and there was tremendous growth in public tertiary education during the 1920s (e.g., 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 categorized universities and teachers colleges as being 4-year public institutions, while junior colleges and all normal schools were categorized as being

2-year public institutions.[29] We then assigned each institution to a county based on the reported city and state (using StatsAmerica's City-to-County Finder), and aggregated instructor counts at the state and county level (if a city straddles multiple counties, we assigned it to the county with which it shares the most overlap in terms of area). Note that instructor counts for county normal schools are only provided at the state level. Therefore, the county-level counts implicitly omit these institutions.

For the modern period, state-level data on K-12 teachers comes from 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). County-level data on K-12 teachers are obtained by aggregating counts from CCD school district-level files for the school year 1999–2000 at the county level (NCES assigns each school district to a county). The same data source is used to approximate county-level average K-12 teacher salaries by dividing total expenditures on salaries at the county level by the corresponding number of K-12 teachers.

State and county-level data on instructors at 2-year and 4-year public colleges and universities for the modern period are based on institutional-level data from NCES' IPEDS database for the school year 2000–01, which we aggregate at state and county level (each institution provides a ZIP code, which we map into counties using a crosswalk by the Census Bureau). We focus on publicly-controlled, Title IV participating post-secondary institutions, and leverage the IPEDS classification to identify 4-year institutions and 2-year institutions. Instructors are measured using the number of full-time and part-time faculty involved in instruction, research, or public service.

[29] It is not clear whether teachers colleges and normal schools were closer to 2-year or 4-year institutions, and in practice this may have varied by school and by time period. Our categorization is motivated by the following quote in the Office of Education's report: "In the decade from 1920 to 1930 great changes have taken place in the institutions primarily devoted to the preparation of teachers. Perhaps the most significant of these is the change in the organization of many of the institutions from 2-year normal schools to 4-year, degree-granting professional schools" (as cited in Office of Education 1931, p. 609).

State and county-level K-12 teachers and instructors at 2-year and 4-year public colleges and universities 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 A6 displays the state-level statistics by year. Panel A in Appendix Table A7 shows means of county-level statistics by year, for the subset of counties used in our analysis.

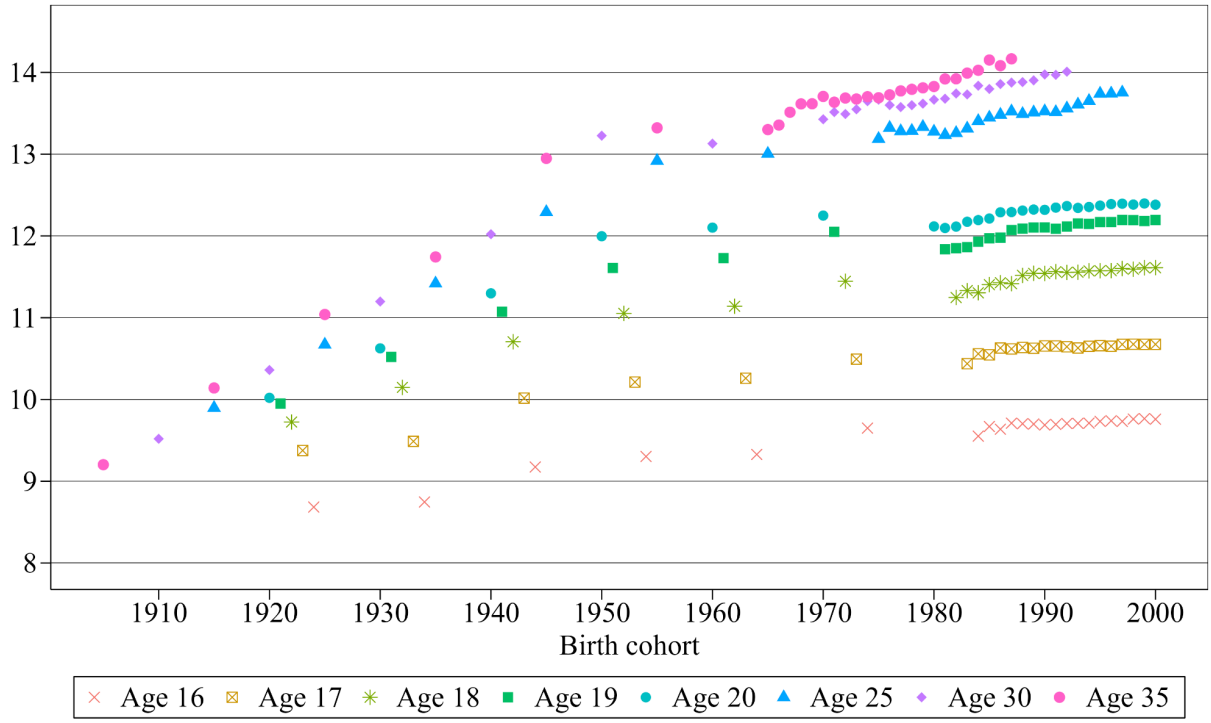
County Demographic and Socioeconomic Characteristics

We construct various county-level demographic and socioeconomic 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 NHGIS) by total population in the 1920 Census.

We construct modern-day analogs of the above county characteristics using county-level data from the 2000 Decennial Census released 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. Panels A–E in Appendix Table A7 show means of the county-level characteristics described in this section for the subset of counties used in our analysis.

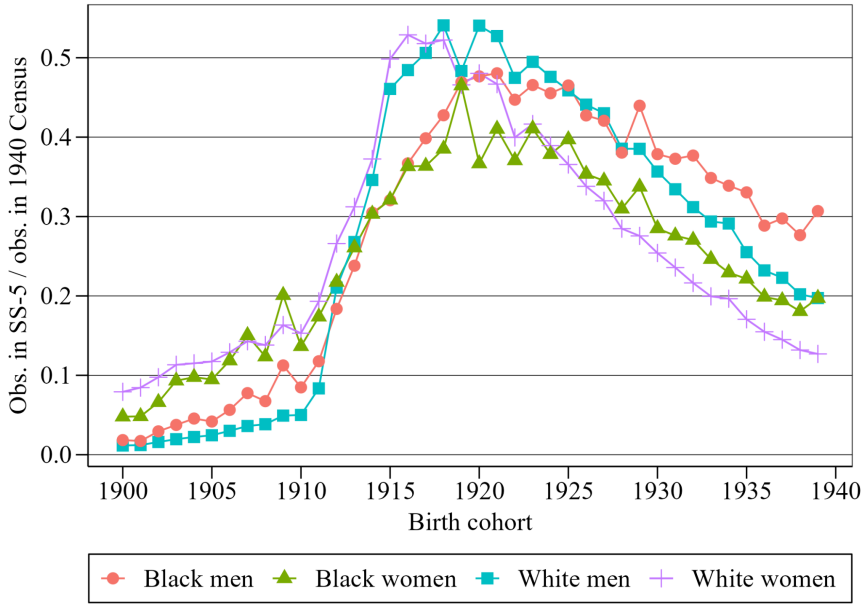
Figure A1. Average Years of Education by Age and Birth Cohort



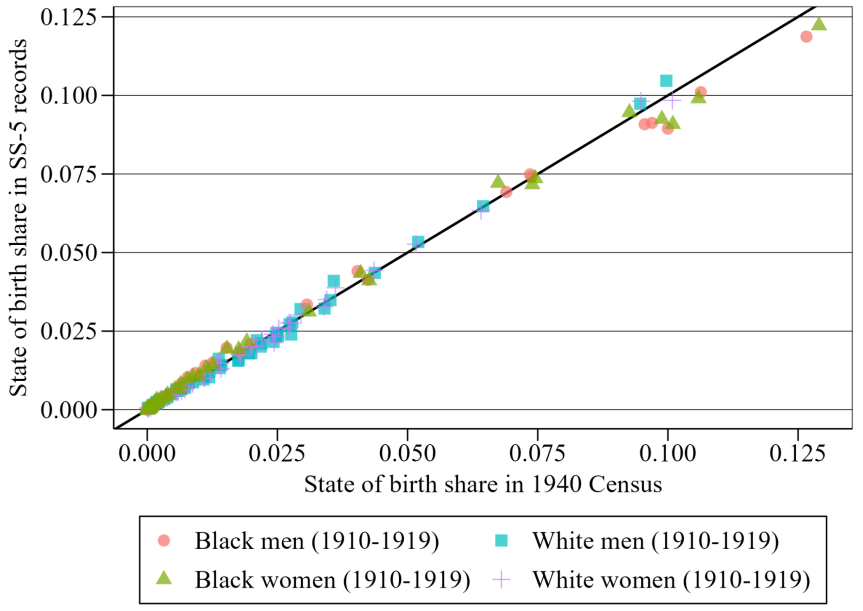
Notes: This graph plots average years of education by cohort at specific ages using data from the 1940–2000 decennial censuses and the 2001–2022 ACS. Observations are weighted by census sampling weights.

Figure A2. Cohort and State of Birth Coverage of SS-5 Records by Sex and Race

Panel A. Cohort Coverage



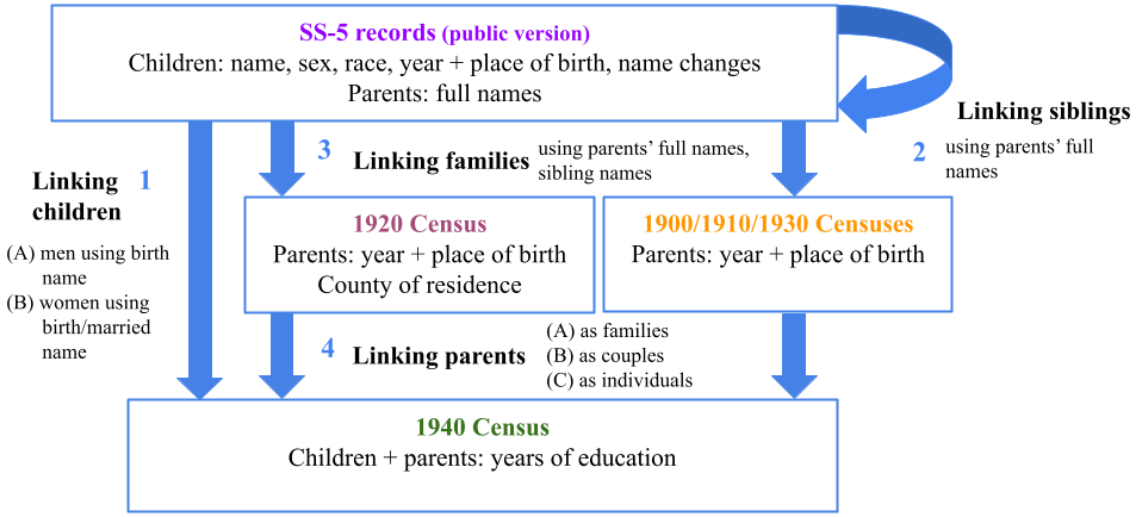
Panel B. State of Birth Coverage



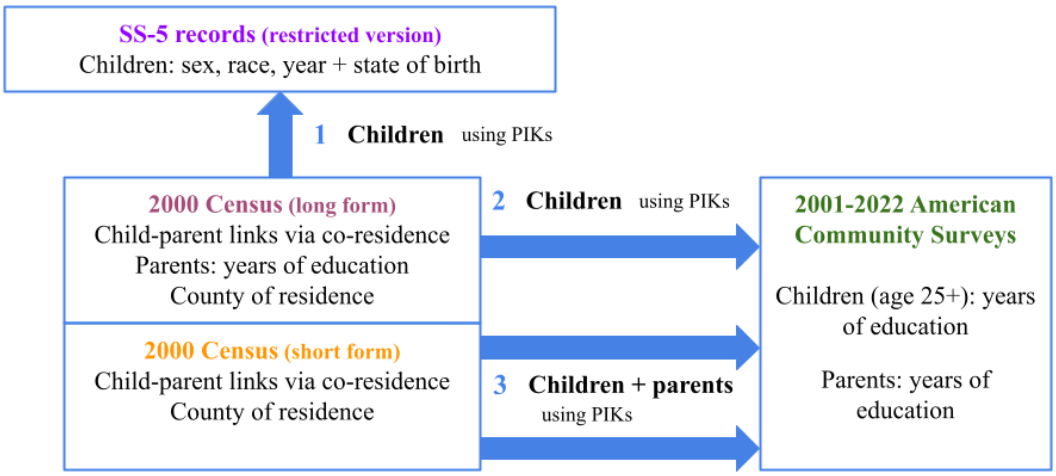
Notes: Panel A plots the number of observations by cohort in SS-5 records divided by the corresponding number of observations in the 1940 Census, separately by sex and race. Panel B plots the state of birth share in SS-5 records against the corresponding share in the 1940 Census among individuals born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

Figure A3. Overview of Linked Sample Construction

Panel A. 1920-1940 Census Linked Sample



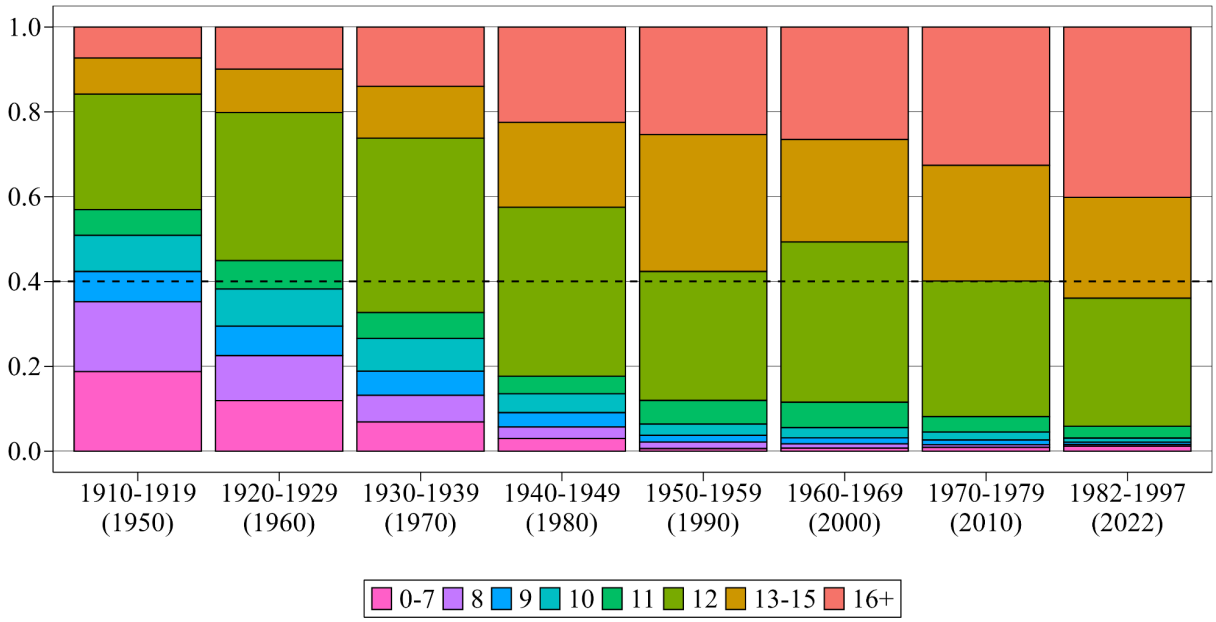
Panel B. 2000 Census-ACS Linked Sample



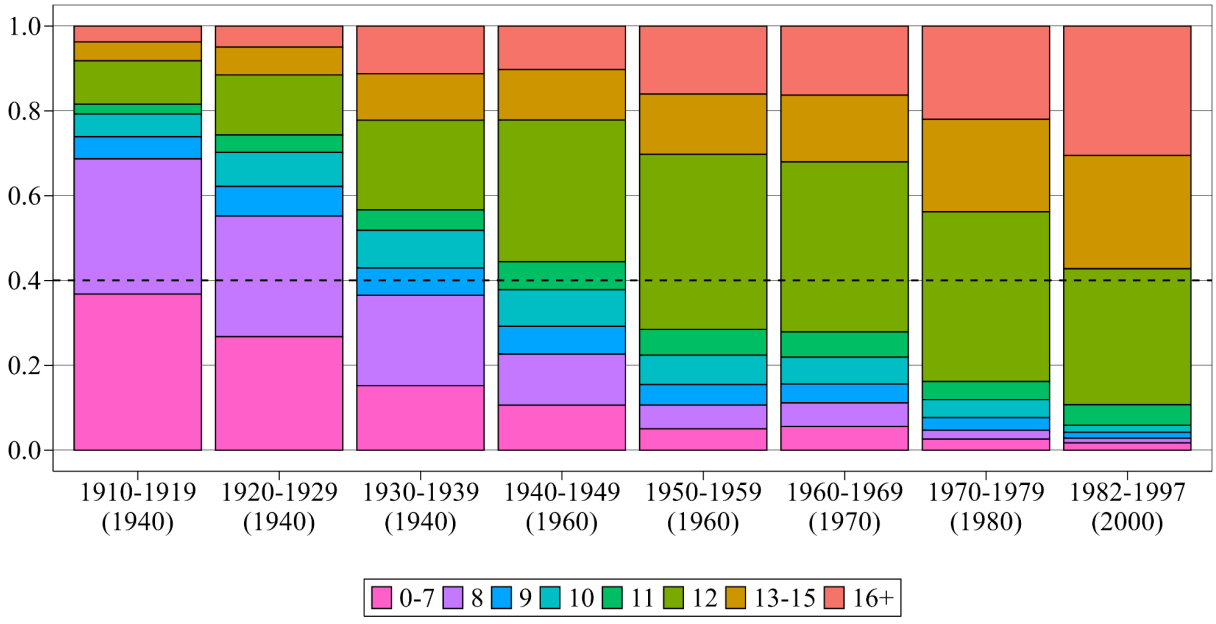
Notes: PIKs refer to Protected Identification Keys used by the Census Bureau to identify unique individuals across records.

Figure A4. Own and Parental Educational Attainment Distribution by Birth Cohort

Panel A. Own Educational Attainment

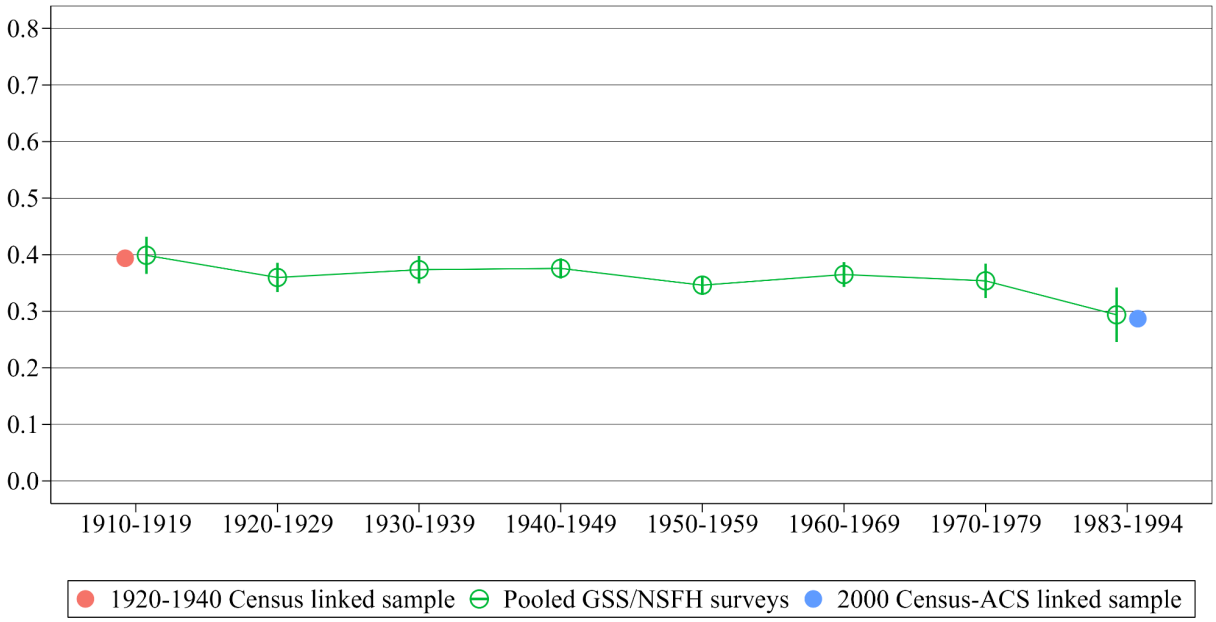


Panel B. Parental Educational Attainment



Notes: Panel A plots the distribution of years of education by cohort group, as measured in the decennial census or ACS wave in which individuals are aged 31–40 (except for 1982–1997 cohorts). Panel B plots the distribution of parental years of education by child cohort group, as measured in the decennial census in which children are aged 1–10 (except for 1910–1919 and 1920–1929 cohorts). Individuals weighted using census sampling weights (individual weights in panel A, household weights in panel B).

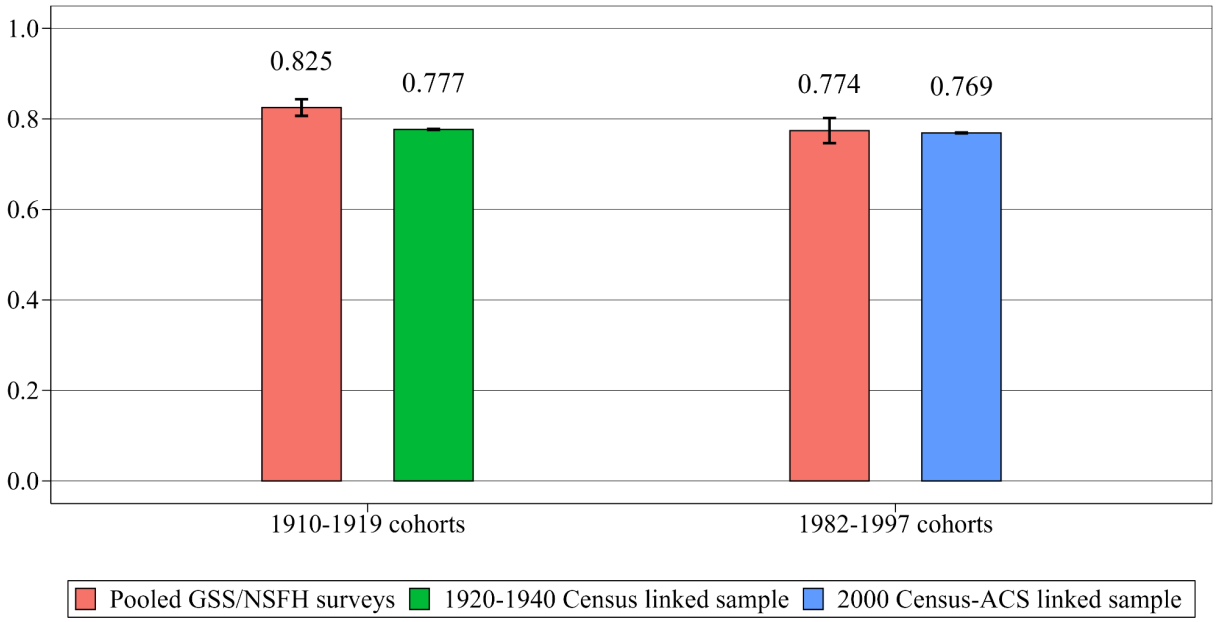
Figure A5. Educational Persistence by Birth Cohorts over the 20th Century



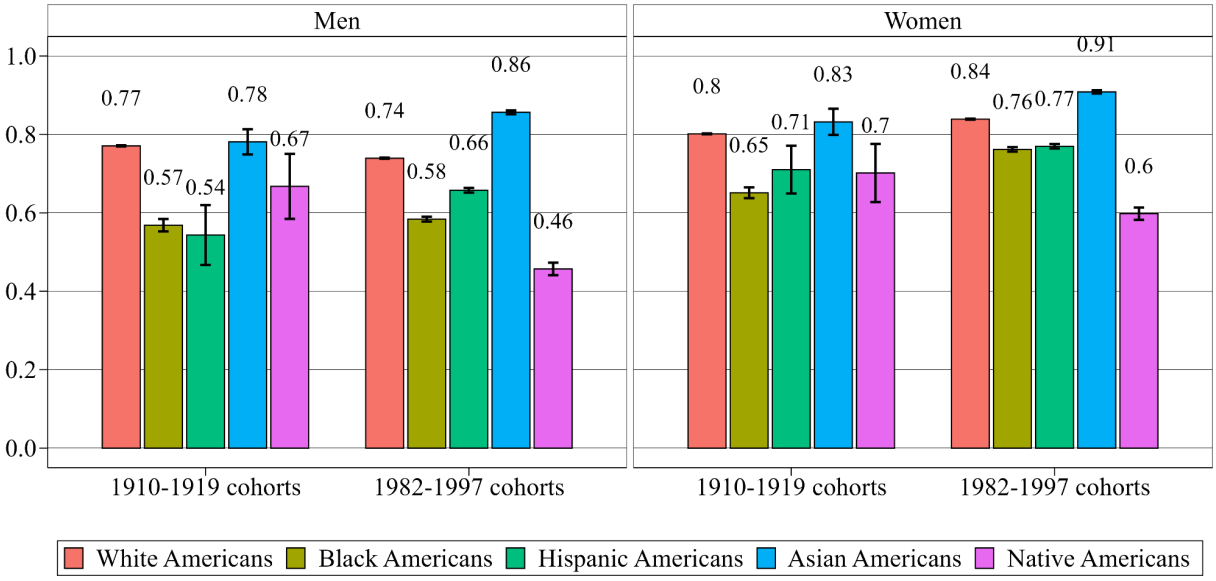
Notes: This graph shows educational persistence estimates (the slope coefficient from a regression of own years of education on parental years of education, a constant and a quadratic in age), for different cohorts and samples. Observations in the pooled GSS/NSFH surveys are weighted by normalized sampling weights. Observations in the 1920-1940 Census and 2000 Census-ACS linked samples are weighted by inverse propensity scores that re-weight the sample to match the sex-by-cohort-by-race distribution in the 1940 and 2000 population, respectively. 95% confidence intervals based on robust standard errors.

Figure A6. Educational Persistence at the Top over the 20th Century

Panel A. Estimates for All in Linked Samples vs. Survey



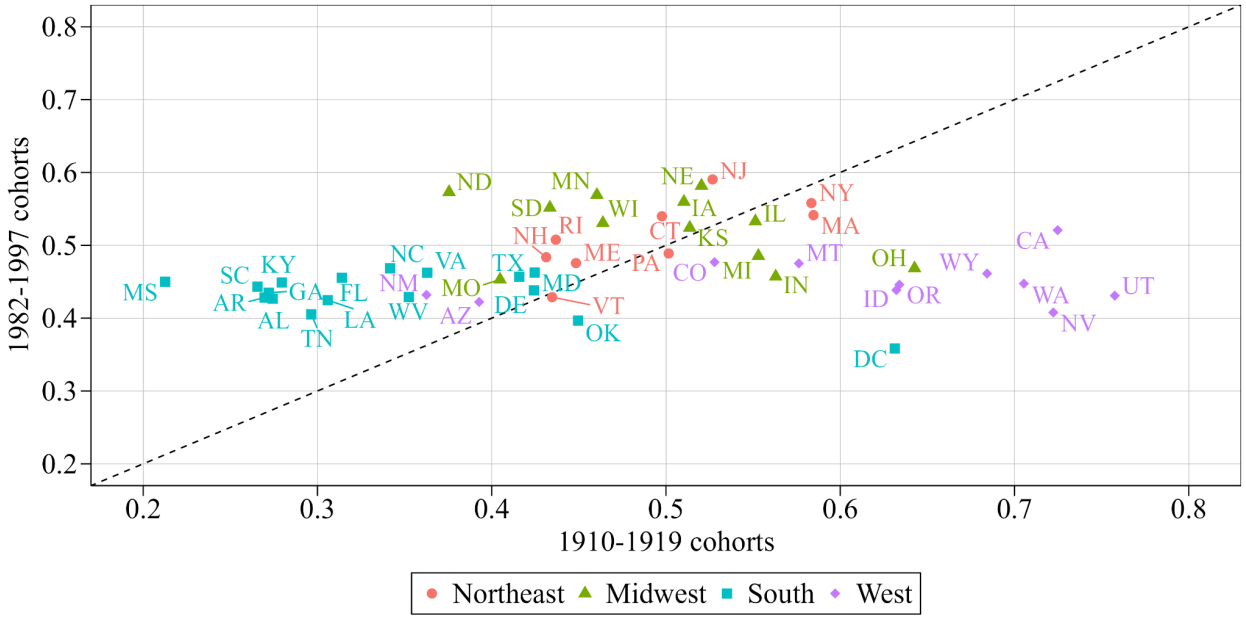
Panel B. Estimates by Sex and Race in Linked Samples



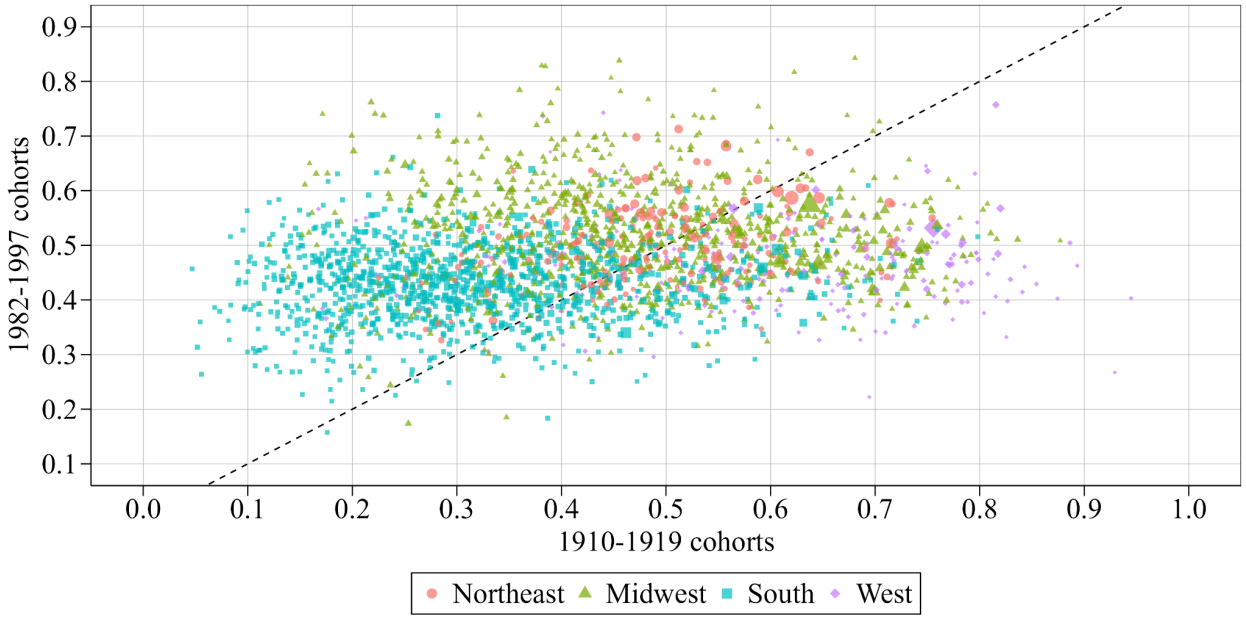
Notes: This graph plots estimates of educational persistence at the top by cohort and by sample (panel A), and estimates of educational persistence at the top by cohort and by sex and race based on the 1920-1940 Census and 2000 Census-ACS linked samples (panel B). Observations in the pooled surveys are weighted by normalized sampling weights. Observations in the 1920-1940 Census and 2000 Census-ACS linked samples are weighted by inverse propensity scores that re-weight the sample to match the sex-by-cohort-by-race distribution in the 1940 and 2000 population, respectively. 95% confidence intervals based on robust standard errors. See Appendix Table A5 for detailed results.

Figure A7. Upward Educational Mobility by Childhood State/County of Residence

Panel A. Upward Educational Mobility by Childhood State of Residence, 1910–19 vs. 1982–97 Cohorts



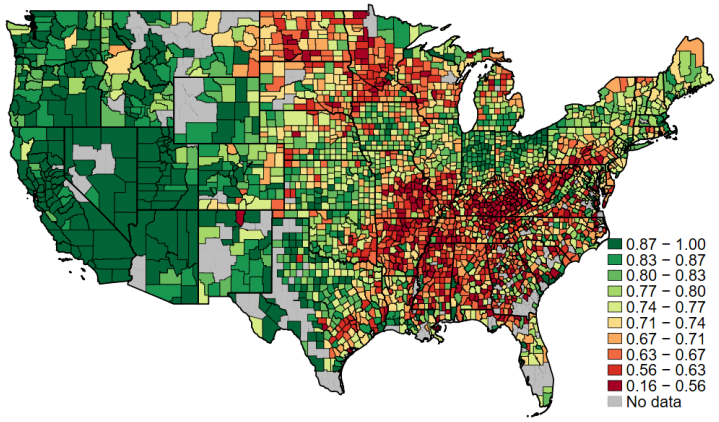
Panel B. Upward Educational Mobility by Childhood County of Residence, 1910–19 vs. 1982–97 Cohorts



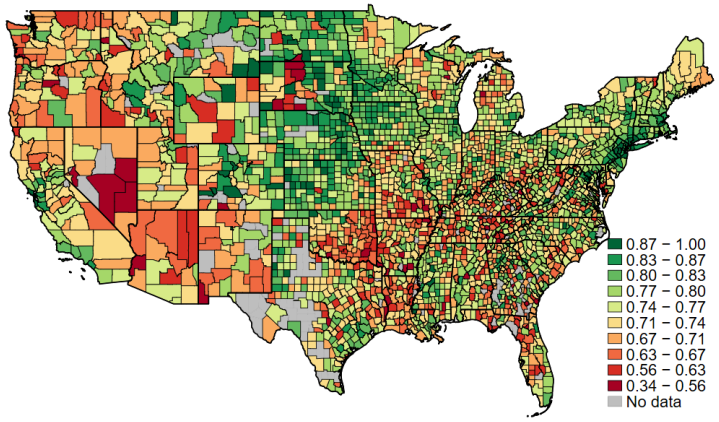
Notes: This graph plots upward educational mobility by 1920 childhood state/county of residence for the 1910–1919 cohorts against upward educational mobility by 2000 childhood state/county of residence for the 1982–1997 cohorts. In panel B, counties are weighted by the average number of children in the bottom 40% across both cohorts in the relevant samples.

Figure A8. Educational Persistence at the Top by Childhood County of Residence

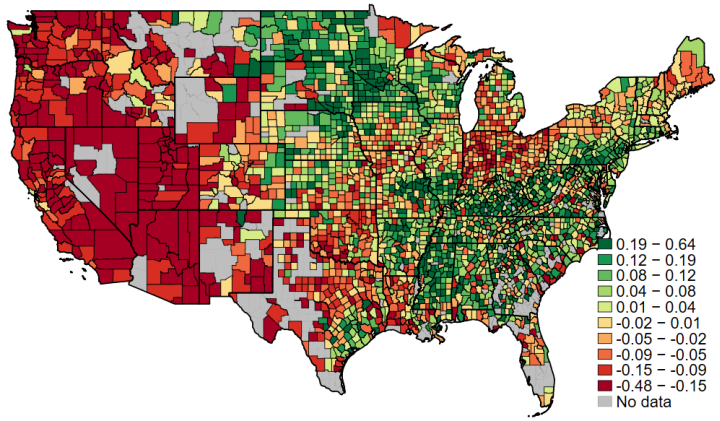
Panel A. 1910–1919 Cohorts



Panel B. 1982–1997 Cohorts



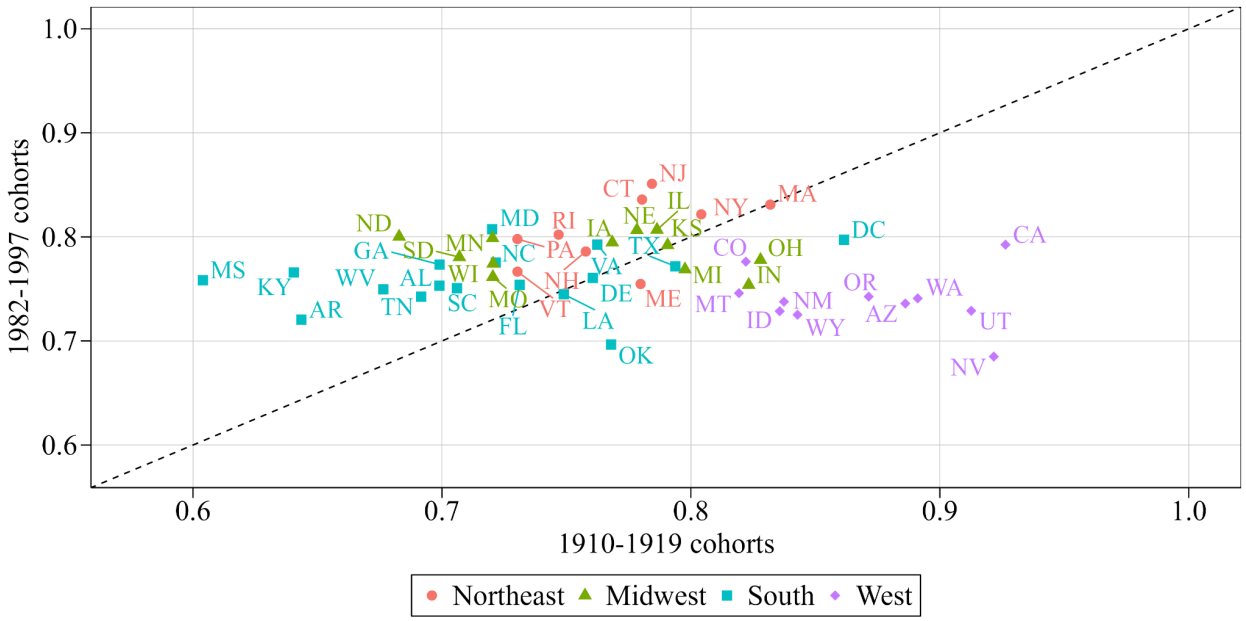
Panel C. Change Between 1910–1919 and 1982–1997 Cohorts



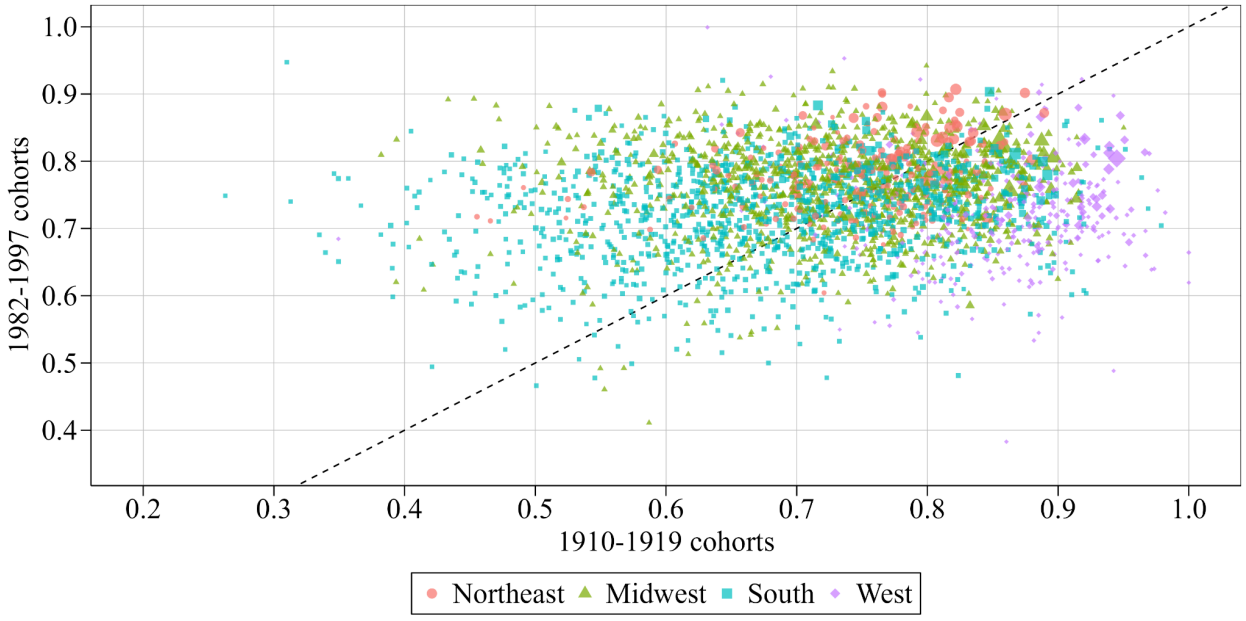
Notes: These graphs plot educational persistence at the top by childhood county of residence in 1920 or 2000, depending on the cohort. Sample restricted to counties with at least 20 observations and "stable" boundaries between 1920 and 2000 (see text). In panels A and C, counties are shaded according to which decile bin they fall in. In panel B, counties are shaded using the decile bin cutoffs from panel A for comparability.

Figure A9. Educational Persistence at the Top by Childhood State/County of Residence

Panel A. Educ. Persistence at the Top by Childhood State of Residence, 1910–19 vs. 1982–97 Cohorts



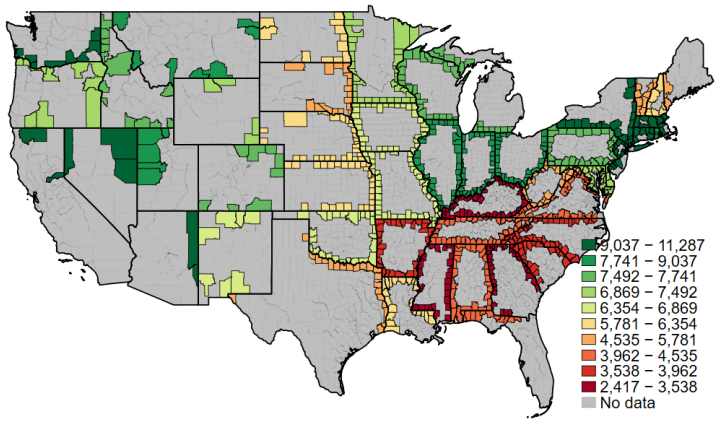
Panel B. Educ. Persistence at the Top by Childhood County of Residence, 1910–19 vs. 1982–97 Cohorts



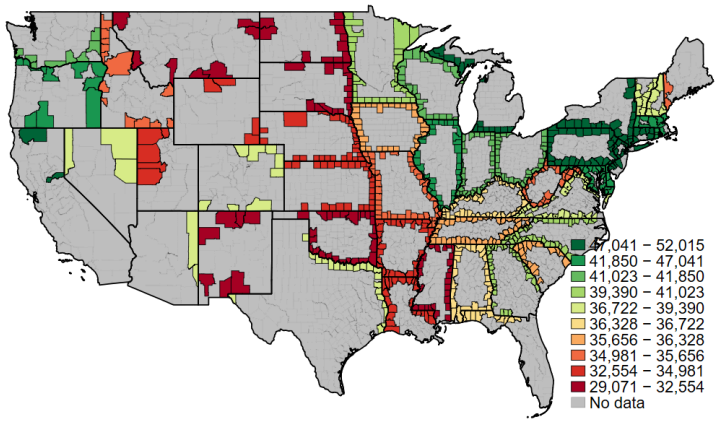
Notes: This graph plots educational persistence at the top by 1920 childhood state/county of residence for the 1910–1919 cohorts against educational persistence at the top by 2000 childhood state/county of residence for the 1982–1997 cohorts. In panel B, counties are weighted by the average number of children in the top 60% across both cohorts in the relevant samples.

Figure A10. State Border Maps: Average K-12 Teacher Salary

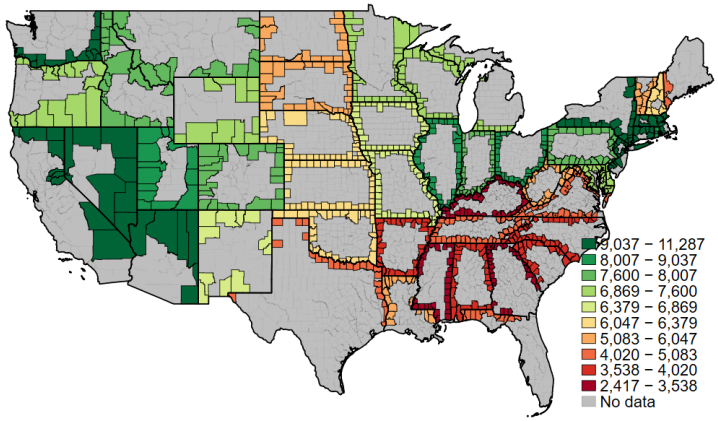
Panel A. Upward Educational Mobility, 1910–1919 Cohorts



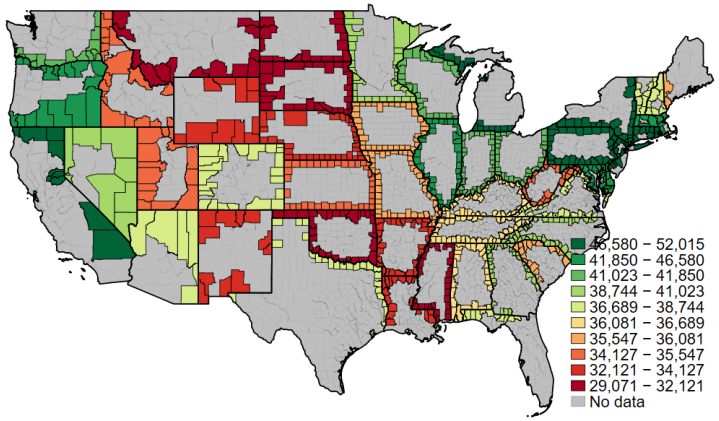
Panel B. Upward Educational Mobility, 1982–1997 Cohorts



Panel C. Educational Persistence at the Top, 1910–1919 Cohorts

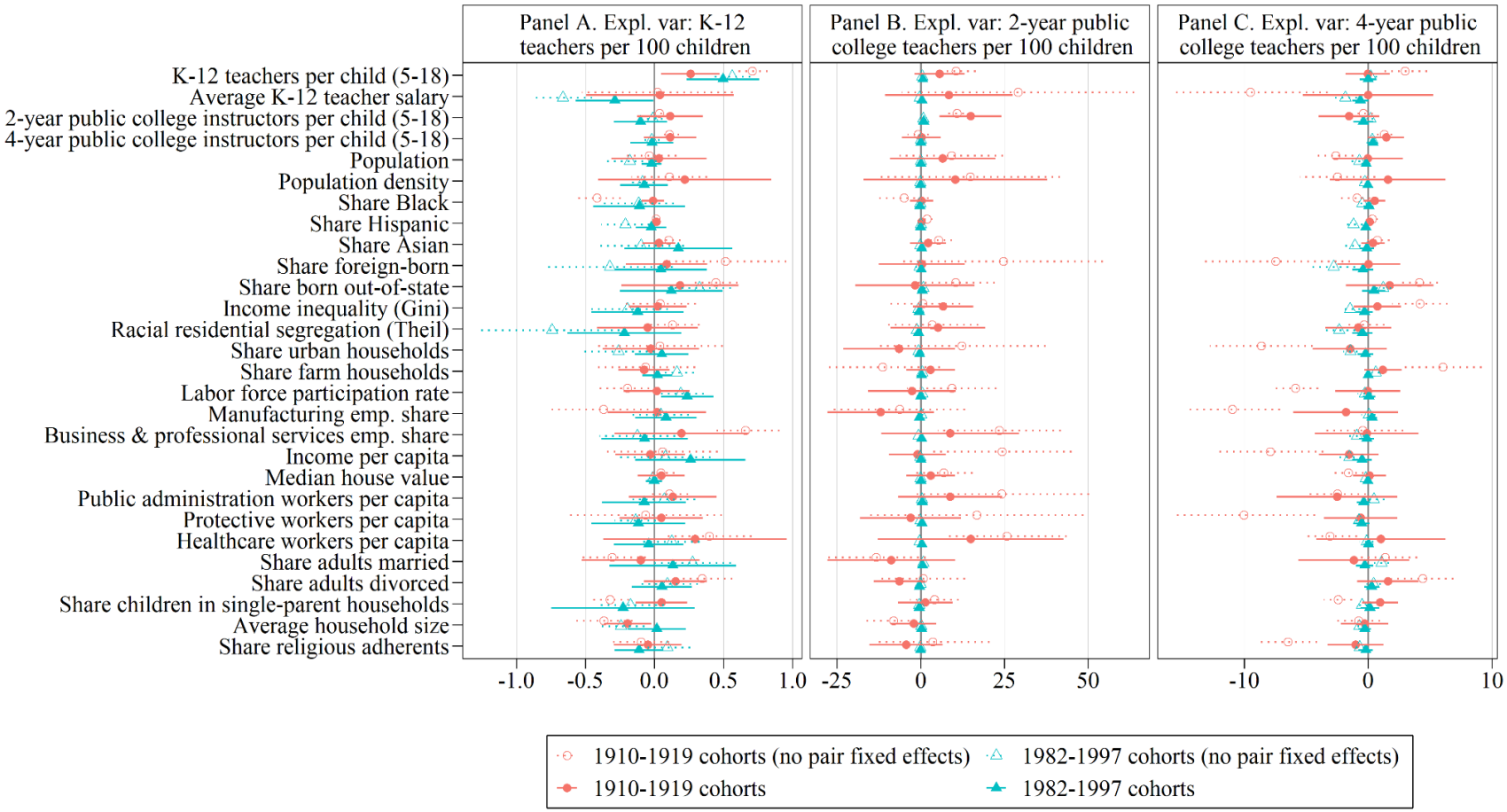


Panel D. Educational Persistence at the Top, 1982–1997 Cohorts



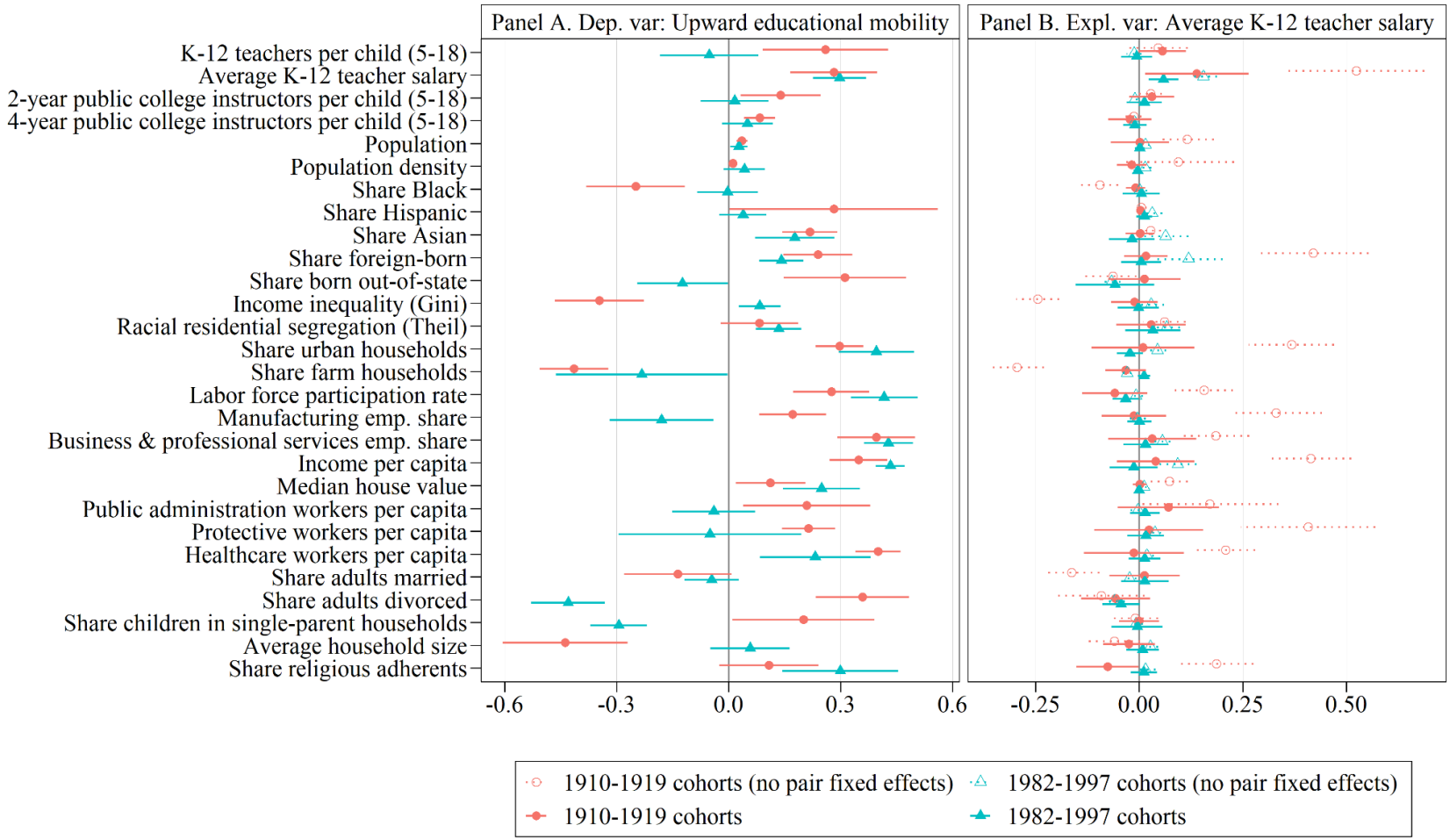
Notes: These graphs plot the counties in the state border regression sample (counties on opposite sides of state borders with at least 20 observations in both periods and "stable" boundaries between 1920 and 2000). Counties shaded according to decile associated with corresponding state average K-12 teacher salary (in 1999 dollars).

Figure A11. Balance Checks for K-12 Teachers and 2-Year and 4-Year Public College Instructors Per 100 Children



Notes: Each dot/triangle is the coefficient from a variant of equation (1) where upward educational mobility is replaced with a county characteristic (indicated in row label) as the dependent variable and the key state-level explanatory variable is indicated in the panel title (either excluding or including county pair fixed effects). In all regressions, counties are weighted by the number of children in the bottom 40% in the relevant sample. All county characteristics are normalized to have a weighted mean of 0 and standard deviation of 1 within each cohort. 95% confidence intervals based on robust standard errors, clustered at the state border level.

Figure A12. Corr. between Educ. Persist. at the Top and County Characteristics and Teacher Salaries Balance Checks



Notes: In panel A, each dot/triangle is the (weighted) correlation coefficient between educational persistence at the top and a county characteristic (indicated in row label), separately by cohort. In panel B, each dot/triangle is the coefficient from a variant of equation (1) where educational persistence at the top is replaced with a county characteristic (indicated in row label) as the dependent variable and the key explanatory variable is the state-level average K-12 teacher salary (either excluding or including county pair fixed effects). In all regressions, counties are weighted by the number of children in the top 60% in the relevant sample. All county characteristics are normalized to have a weighted mean of 0 and standard deviation of 1 within each cohort. 95% confidence intervals based on robust standard errors (clustered at the state border level in panel B).

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% 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% 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% 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% 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% 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 (cont.)

	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% hold-out sample)	
(F) Linking parents in SS-5 records to the 1940 Census as couples (with age and birthplace information via 97% 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 year of birth, 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% hold-out sample)	
(G) Linking parents in SS-5 records to the 1940 Census as individuals (with age and birthplace information via 97% 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 Linked Sample, 1910–19 Cohorts

	All	Men	Women
	(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	0.17
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 in our 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) shown in the bottom two rows.

Table A3. Educational Persistence Estimates by Cohort, Sex, and Sample

	All	Men	Women
	(1)	(2)	(3)
Panel A: 1910–1919 cohorts			
Pooled surveys	0.399	0.411	0.388
	(0.017)	(0.029)	(0.018)
N	2,851	1,096	1,755
1920-1940 Census linked sample	0.394	0.418	0.370
	(0.001)	(0.002)	(0.002)
N	1,457,003	733,288	723,715
Panel B: 1983–1994 cohorts			
Pooled surveys	0.294	0.289	0.302
	(0.025)	(0.040)	(0.030)
N	2,485	1,091	1,394
2000 Census-ACS linked sample	0.287	0.286	0.289
	(0.001)	(0.002)	(0.002)
N	1,297,000	667,000	630,000

Notes: This table shows educational persistence estimates (the slope coefficient from a regression of own years of education on parental years of education, a constant and a quadratic in age), for different cohorts, gender groups, and samples. Observations in the pooled GSS/NSFH surveys are weighted by normalized sampling weights. Observations in the 1920-1940 Census and 2000 Census-ACS linked samples are weighted by inverse propensity scores that re-weight the sample to match the sex-by-cohort-by-race distribution in the 1940 and 2000 population, respectively. Robust standard errors in parentheses. The sample sizes for the 2000 Census-ACS linked sample have been rounded in line with disclosure guidelines.

Table A4. Upward Educational Mobility Estimates by Cohort, Sex, Race/Ethnicity, and Sample

	All	Men					Women				
		White	Black	Hispanic	Asian	Native	White	Black	Hispanic	Asian	Native
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: 1910–1919 cohorts											
Pooled surveys	0.472 (0.019)	—	—	—	—	—	—	—	—	—	—
N	843	—	—	—	—	—	—	—	—	—	—
1920-1940 Census linked sample	0.446 (0.001)	0.490 (0.002)	0.211 (0.005)	0.221 (0.018)	0.463 (0.028)	0.307 (0.038)	0.505 (0.001)	0.321 (0.005)	0.249 (0.020)	0.564 (0.031)	0.335 (0.043)
N	424,281	194,325	12,342	597	371	174	202,192	12,187	490	273	129
Panel B: 1982–1997 cohorts											
Pooled surveys	0.509 (0.018)	—	—	—	—	—	—	—	—	—	—
N	1,228	—	—	—	—	—	—	—	—	—	—
2000 Census-ACS linked sample	0.476 (0.001)	0.402 (0.001)	0.306 (0.003)	0.410 (0.002)	0.673 (0.005)	0.227 (0.006)	0.573 (0.001)	0.490 (0.003)	0.560 (0.003)	0.780 (0.005)	0.365 (0.007)
N	815,000	297,000	36,500	61,500	13,000	6,400	279,000	38,500	59,000	11,000	6,400

Notes: This table shows upward educational mobility estimates for different cohorts, sex-by-race groups, and samples. Observations in the pooled GSS/NSFH surveys are weighted by normalized sampling weights. Observations in the 1920-1940 Census and 2000 Census-ACS linked samples are weighted by inverse propensity scores that re-weight the sample to match the sex-by-cohort-by-race distribution in the 1940 and 2000 population, respectively. Robust standard errors in parentheses. The sample sizes for the 2000 Census-ACS linked sample have been rounded in line with disclosure guidelines.

Table A5. Educational Persistence Estimates by Cohort, Sex, Race/Ethnicity, and Sample

	All	Men					Women				
		White	Black	Hispanic	Asian	Native	White	Black	Hispanic	Asian	Native
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: 1910–1919 cohorts											
Pooled surveys	0.825 (0.009)	—	—	—	—	—	—	—	—	—	—
N	2,008	—	—	—	—	—	—	—	—	—	—
1920-1940 Census linked sample	0.777 (0.001)	0.771 (0.001)	0.569 (0.008)	0.543 (0.039)	0.781 (0.016)	0.668 (0.042)	0.801 (0.001)	0.651 (0.007)	0.710 (0.031)	0.832 (0.017)	0.702 (0.038)
N	1,032,722	516,911	6,124	194	757	152	499,962	6,456	229	539	156
Panel B: 1982–1997 cohorts											
Pooled surveys	0.774 (0.014)	—	—	—	—	—	—	—	—	—	—
N	1,773	—	—	—	—	—	—	—	—	—	—
2000 Census-ACS linked sample	0.769 (0.001)	0.739 (0.001)	0.584 (0.003)	0.658 (0.003)	0.857 (0.003)	0.457 (0.008)	0.839 (0.001)	0.762 (0.003)	0.770 (0.003)	0.909 (0.002)	0.598 (0.008)
N	1,321,000	556,000	43,000	37,500	29,000	5,600	524,000	43,000	35,000	25,500	5,550

Notes: This table shows educational persistence at the top estimates for different cohorts, sex-by-race groups, and samples. Observations in the pooled GSS/NSFH surveys are weighted by normalized sampling weights. Observations in the 1920-1940 Census and 2000 Census-ACS linked samples are weighted by inverse propensity scores that re-weight the sample to match the sex-by-cohort-by-race distribution in the 1940 and 2000 population, respectively. Robust standard errors in parentheses. The sample sizes for the 2000 Census-ACS linked sample have been rounded in line with disclosure guidelines.

Table A6. Public Education Statistics by State and Year

	K-12 teachers per 100 children (ages 5–18)		Average K-12 teacher salary (1999 dollars)		2-year public college instructors per 100 children (ages 5–18)		4-year public college instructors per 100 children (ages 5–18)	
	1920	2000	1920	2000	1930	2000	1930	2000
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
United States	2.3	4.6	7,194	41,666	0.02	0.59	0.11	0.77
Alabama	1.5	5.4	4,020	36,689	0.04	0.53	0.05	0.99
Arizona	2.1	4.1	10,623	36,902	0.02	0.55	0.28	0.56
Arkansas	1.7	4.9	3,962	33,386	0.02	0.46	0.05	0.99
California	2.6	3.5	10,565	47,680	0.12	0.78	0.3	0.68
Colorado	3	4.7	7,716	38,163	0.01	0.44	0.25	1.26
Connecticut	2.4	5.3	9,336	51,780	0.04	0.42	0.03	0.6
Delaware	2	4.7	7,043	44,435	0	0.63	0.14	0.81
D. C.	2.5	5	11,287	47,076	0.06	0	0.02	0.82
Florida	2.3	3.7	4,302	36,722	0.01	0.82	0.09	0.44
Georgia	1.6	5.2	3,538	41,023	0.01	0.51	0.06	0.66
Idaho	3.1	4.6	7,741	35,547	0.05	0.31	0.11	0.78
Illinois	2.2	4.1	8,978	46,486	0.02	0.62	0.1	0.54
Indiana	2.3	4.5	8,007	41,850	0	0.3	0.11	0.95
Iowa	4.3	5.4	6,869	35,678	0.02	0.57	0.19	0.82
Kansas	3.4	5.2	6,321	34,981	0.02	0.79	0.2	1.04
Kentucky	1.8	4.3	3,430	36,380	0.01	0.63	0.11	1.03
Louisiana	1.5	5	6,005	33,109	0.01	0.27	0.06	0.93
Maine	3.6	6.5	5,008	35,561	0.06	0.13	0.09	0.86
Maryland	1.7	4.7	7,492	44,048	0.02	0.71	0.19	0.98
Massachusetts	2	5.6	10,482	46,580	0.02	0.46	0.03	0.61
Michigan	2.6	4.1	7,566	49,044	0.02	0.51	0.25	0.99
Minnesota	3	5.1	7,326	39,802	0.02	0.57	0.19	0.7
Mississippi	1.9	4.1	2,417	31,857	0.01	0.53	0.05	0.71
Missouri	2.3	5.4	6,620	35,656	0.02	0.45	0.1	0.83
Montana	4.9	5.4	7,957	32,121	0.05	0.27	0.13	1.16
Nebraska	4	5.7	6,354	33,237	0	0.65	0.18	1.24
Nevada	4.2	3.9	9,659	39,390	0	0.54	0.38	0.72

Notes: This table reports public education statistics by state and year for the continental United States.

Table A6. Public Education Statistics by State and Year (cont.)

	K-12 teachers per 100 children (ages 5–18)		Average K-12 teacher salary (1999 dollars)		2-year public college instructors per 100 children (ages 5–18)		4-year public college instructors per 100 children (ages 5–18)	
	1920	2000	1920	2000	1930	2000	1930	2000
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
New Hampshire	2.8	5.6	6,304	37,734	0	0.62	0.23	0.54
New Jersey	2.1	5	10,648	52,015	0.02	0.44	0.03	0.72
New Mexico	2.4	3.9	6,669	32,554	0.03	0.81	0.18	0.88
New York	2.4	4.6	10,432	51,020	0.03	0.46	0.06	0.71
North Carolina	1.9	5.1	3,854	39,404	0.01	1.27	0.06	0.67
North Dakota	4.2	6.1	6,047	29,863	0.03	0.41	0.24	1.49
Ohio	2.4	5	9,037	41,436	0	0.49	0.13	0.76
Oklahoma	2.3	5.1	6,379	31,298	0.02	0.52	0.18	0.93
Oregon	4	3.4	7,226	42,336	0.06	0.94	0.34	1.1
Pennsylvania	1.8	4.2	7,641	48,321	0	0.4	0.05	0.87
Rhode Island	1.9	4.7	8,887	47,041	0	0.14	0.05	0.54
South Carolina	1.6	5.6	3,854	36,081	0	0.56	0.08	0.83
South Dakota	4.1	5	5,781	29,071	0	0.19	0.22	1.11
Tennessee	1.8	5.4	4,103	36,328	0	0.36	0.06	0.69
Texas	1.9	4.9	5,083	37,567	0.02	0.57	0.1	0.6
Utah	2.7	3.5	8,239	34,946	0	0.33	0.18	1.02
Vermont	3.2	5.3	5,540	37,758	0.01	0.42	0.21	1.18
Virginia	1.9	6.2	4,535	38,744	0	0.57	0.13	1.02
Washington	3	3.8	10,208	41,043	0.06	0.74	0.21	0.73
West Virginia	2.4	5.3	5,307	35,009	0.03	0.15	0.11	1.53
Wisconsin	2.3	5.4	7,600	41,153	0.01	1.03	0.18	0.85
Wyoming	4.4	6.4	7,218	34,127	0	1.26	0.24	0.86
Standard deviation (SD)	0.89	0.75	2,261	5,983	0.02	0.26	0.09	0.24
Coefficient of variation (SD/mean)	0.34	0.15	0.32	0.15	1.13	0.48	0.61	0.28

Notes: This table reports public education statistics by state and year for the continental United States.

Table A7. County Characteristics by Year

	1920		2000	
	Mean	*N*	Mean	*N*
	(1)	(2)	(3)	(4)
Panel A: Public education				
K-12 teachers per 100 children (5–18)	2.7	2,805	5.5	2,800
Average K-12 teacher salary (1999 dollars)	11,153	2,804	41,772	2,799
2-year public college instructors per 100 children (5–18)	0.02	2,805	0.47	2,805
4-year public college instructors per 100 children (5–18)	0.10	2,805	0.50	2,805
Panel B: Demographic structure				
Population	36,469	2,805	94,887	2,805
Population density (people per sq mi)	140	2,805	229	2,805
Share Black	0.11	2,805	0.08	2,805
Share Hispanic	0.01	2,805	0.05	2,805
Share Asian	0.001	2,805	0.008	2,805
Share foreign-born	0.07	2,805	0.03	2,805
Share born out-of-state	0.21	2,805	0.26	2,805
Income inequality (Gini)	0.47	2,804	0.38	2,805
Racial residential segregation (Theil)	0.19	2,805	0.08	2,805
Panel C: Economic development				
Share urban households	0.2	2,805	0.4	2,805
Share farm households	0.51	2,805	0.05	2,805
Labor force participation rate	0.55	2,805	0.61	2,805
Manufacturing employment share	0.11	2,804	0.17	2,805
Business & professional services emp. share	0.04	2,804	0.26	2,805
Income per capita (1999 dollars)	1,996	2,804	17,559	2,805
Median house value (1999 dollars)	22,387	2,804	84,832	2,805
Panel D: Public services				
Public admin. workers per 1,000 people	5.8	2,805	22.3	2,805
Protective workers per 1,000 people	1.1	2,805	8.4	2,805
Healthcare workers per 1,000 people	3.4	2,805	29.8	2,805
Panel E: Family structure & social capital				
Share adults married	0.62	2,805	0.6	2,805
Share adults divorced	0.01	2,805	0.1	2,805
Share children in single-parent households	0.08	2,805	0.25	2,805
Average household size	4.4	2,805	2.5	2,805
Share religious adherents	0.43	2,802	0.53	2,805

Notes: This table reports unweighted means of county characteristics and the number of counties with non-missing data. Sample restricted to subset of counties in our analysis. See Appendix B for details on variable construction.

Table A8. The Changing Relationship Between Educational Persistence at the Top and Resources for Public Education

	Dependent variable: Educational persistence at the top							
	1910–1919 cohorts				1982–1997 cohorts			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
K-12 teachers per 100 children (ages 5–18)	-0.0002 (0.004)	-0.007 (0.008)	0.006 (0.012)	-0.005 (0.009)	0.020*** (0.004)	0.035*** (0.007)	0.011 (0.009)	-0.004 (0.004)
Average K-12 teacher salary (1,000s of 1999 dollars)	0.020*** (0.002)	0.019*** (0.002)	0.008** (0.003)	0.006*** (0.002)	0.004*** (0.0004)	0.006*** (0.001)	0.002 (0.001)	0.0002 (0.001)
2-year public college instructors per 100 children (ages 5–18)	0.470*** (0.133)	0.621** (0.244)	-0.152 (0.485)	-0.188 (0.333)	0.005 (0.009)	0.003 (0.018)	0.005 (0.023)	0.003 (0.011)
4-year public college instructors per 100 children (ages 5–18)	0.243*** (0.048)	0.326*** (0.064)	0.043 (0.170)	0.148 (0.096)	-0.014 (0.011)	0.022 (0.025)	0.007 (0.030)	0.014 (0.011)
County pair fixed effects			✓	✓			✓	✓
County controls				✓				✓
R^2	0.263	0.230	0.815	0.888	0.171	0.255	0.810	0.924
Partial R^2 (educ. vars.)	–	–	0.031	0.029	–	–	0.019	0.007
p-value (educ. vars.)	0.000	0.000	0.103	0.072	0.000	0.000	0.596	0.668
N	2,737 (counties)	2,200 (county × pairs)			2,737 (counties)	2,200 (county × pairs)		

Notes: This table shows regressions of county educational persistence at the top on state-level public education resources, separately by cohort. In columns 1 and 5, the sample consists of all counties in the U.S. with stable boundaries and at least 20 children in the top 60% in the relevant sample. In all other columns, the sample consists of border counties-by-pairs. The regressions in columns 3 and 7 include county pair fixed effects, and the regressions in columns 4 and 8 additionally control for the non-education county controls listed in Appendix Table A7 (Panels B–E). In all regressions, counties are weighted by the corresponding number of children in the top 60% in the relevant sample. Robust standard errors in parentheses (clustered at the state border level in columns 2–4 and 5–8). *** 1%, ** 5%, * 10% significance.