FEDERAL RESERVE BANK *of* ATLANTA **WORKING PAPER SERIES**

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

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

Working Paper 2026-1a
January 2026 (Revised May 2026)

Abstract: We construct two large-scale datasets and use a rank-rank framework to describe relative and upward educational mobility by childhood county for U.S. cohorts born in 1910–19 and 1982–97. After validating our estimates against nationally representative surveys, we document increases in relative and upward educational mobility over the 20th century that mask substantial geographic convergence. Today, where children grow up matters less than it did historically. Using a state-border design, we show that increasing public spending on K-12 and postsecondary education raised upward educational mobility over the 20th century but had little effect on the persistence of educational advantage.

JEL classification: J62, I24, I28, N32

Key words: education, inequality, intergenerational mobility

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

Martha J. Bailey (marthabailey@g.ucla.edu) is a professor of economics and director of the California Center for Population Research at the University of California, Los Angeles, and a research associate at the National Bureau of Economic Research. A.R. Shariq Mohammed (ab.mohammed@northeastern.edu) is an assistant professor of economics at Northeastern University. Paul Mohnen (paul.mohnen@atl.frb.org.with) is a research economist at the Federal Reserve Bank of Atlanta.

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

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

What is less clear is how increases in educational attainment and public school spending translated into educational mobility, as measured by the extent to which parents' educational attainment shapes their children's. Theoretically, increased school spending could raise children's attainment across all backgrounds or amplify the persistence of advantage, having little effect on educational mobility or even decreasing it. On the other hand, greater school spending could level the playing field, allowing more children from disadvantaged backgrounds to transcend their parents' achievements and increase educational mobility by creating more opportunities.

This paper provides new evidence on this question by documenting the evolution of educational mobility at the local level over the 20th century and the role of public education in shaping it. To do so, we build two new large-scale datasets spanning 75 years. The first includes nearly 1.5 million children born in the U.S. between 1910 and 1919 who completed their schooling between 1930 and 1940; the second includes 2.1 million children born in the U.S. between 1982 and 1997 who completed their schooling between 2002 and 2022. The early 20th-century dataset is created by linking Social Security application records (SS-5) to children's and parents' educational attainment in the full-count 1940 Census (the first to ask about education) and the 1920 Census, which connects these outcomes to children's county of residence and local investments in public education. Our later 20th-century dataset uses the U.S. Census Bureau's links of the restricted 2000 Decennial Census, SS-5 records, and the 2001–2022

American Community Surveys (*ACS*), which allow us to connect children's and parents' educational attainment as well as children's county of residence.

We use these data to provide a novel description of both relative and upward intergenerational educational mobility *by county* at two points in the 20th century, which we measure as the relationship between parents' and children's *relative ranking* in educational attainment (Fletcher & Han, 2019). Similar to modern estimates of income mobility, we find that the conditional expectation of a child's educational attainment, given parents' attainment, is roughly linear in percentile ranks (Chetty et al., 2014a). The slope of the line characterizes *relative* educational mobility, whereas the intercept measures *absolute* upward educational mobility. This linear relationship helps smooth the lumpy education distribution, which allows us to provide a fuller characterization of educational mobility across the distribution of parent attainment and also make comparisons at the *same percentile* at two points in time.[1] The rank-rank framework also minimizes privacy concerns associated with the disclosure of small-area estimates from the Federal Statistical Restricted Data Center (FSRDC).

Because linked samples may produce biased estimates due to linking errors and unrepresentative samples, we begin the analysis by validating our estimates against those from nationally representative surveys and cross-sectional studies. Our results for educational persistence align closely with the literature's findings in both levels and trends, showing modest increases in relative educational mobility over time. In the early 20th century, a 10-percentile increase in a parent's educational rank was associated with a 4.9-percentile increase in a child's educational rank. In the late 20th century, this association had fallen to 3.9 percentile ranks, indicating less persistence and more educational mobility across generations. We add to the literature by further characterizing the expected rank of children of parents at the 25th percentile, which we call *upward* educational mobility. Children of parents at the 25th percentile

[1] Almost one third of the parents of the early 20th century had 8 years of education, whereas roughly one third of the parents of the later 20th century cohorts had exactly 12 years of education. These changes in groupings over the 20th century make comparing upward mobility at the same points in the parental attainment distribution difficult, which the rank-rank parameterization helps resolve.

of education distribution attained the 36th percentile in the early 20th century and the 38th percentile in the later 20th century, indicating a modest increase in upward mobility over our 75-year time horizon.

Using this rank-rank parameterization and large sample sizes, our main contribution is to describe the changing county-level geography of educational mobility over the 20th century and to link these changes to spending on public education. We find that national trends mask substantial differences across counties in both the early and late 20th centuries, as well as changes within counties over time. In the early 20th century, relative mobility was generally lower in the South but varied within states and regions. Upward mobility, however, was uniformly lower across the South, with almost all counties in the former slave states showing *little to no* upward educational mobility. In contrast, counties in the West, Northeast, and Industrial Midwest had very high rates of upward educational mobility. By the late 20th century, both relative and upward educational mobility increased in almost every county in the South, while declining in the Northeast, the Industrial Midwest, and the West. Put simply, while early 20th-century educational mobility depended heavily on where children grew up, place matters much less today.

Lastly, we use a state-border design to examine how changes in public education investments shaped the evolving geography of educational mobility. We find that counties located in states with relatively larger increases in per-child K-12 teachers, K-12 teacher salaries, and per-child college instructors experienced greater increases in upward educational mobility over the 20th century than neighboring counties located in states that invested less. In contrast, we find little effect on relative educational mobility. Consistent with these effects being causal, balance tests show that counties on both sides of state borders experienced similar changes in a variety of labor markets and demographic conditions over the same period. Together, these findings imply that increases in public education spending improved opportunities for all children, but that children of more educated parents maintained their educational advantage.

This paper contributes to the literature by providing the first evidence on the evolution of educational opportunity *at the county level* across the 20th century. To date, evidence on the evolution of educational mobility in the U.S. has focused on *national* changes (Couch & Dunn, 1997; Hertz et al., 2007; Hout & Janus, 2011; Torche, 2015; Fletcher & Han, 2019; Ferrie et al., 2021), largely due to the use of survey data and limited sample sizes for subnational geographies.[2] Using new large-scale, parent-child samples, we use a common measurement framework to document the changing geography of educational mobility over 75 years. This county perspective allows us to characterize how local economic conditions and educational policy shaped intergenerational educational opportunity during the human capital century.

This paper also contributes to the literature on historical educational mobility in the U.S., including work on the 19th and early 20th century (Rauscher, 2016; Feigenbaum, 2018; Althoff et al., 2025). In the closest paper to ours, Card et al. (2022) examine upward educational mobility of children ages 16–18 living with their parents in the 1940 Census (born in 1922–1924), documenting significant county differences in upward educational mobility and the causal role of educational resources on Black-White differences in upward mobility in the South. Our paper extends Card et al. (2022) forward by 75 years and also broadens the analysis by using a rank-rank measure of educational persistence to document mobility across the parental education distribution; by using a linked sample to incorporate increases in educational attainment at ages older than 16–18; by examining the effect of school resources outside the South; and by including measures of states' postsecondary educational investments.

Third, this paper contributes to a deeper understanding of the determinants of intergenerational educational mobility. A central finding for the late 20th century is that spatial differences in income mobility are driven by factors that affect children before they enter the labor market, although robust associations with measures of class size or public school expenditures per student have been lacking (Chetty et al., 2014b). This paper deepens our understanding of the pre-labor-market determinants of income mobility by demonstrating the large role of public education policy—including the number of K-12 teachers and public college instructors, as well as teacher salaries—in shaping the evolution of economic opportunity in the United States.

[2] Ferrie et al. (2021) use large, linked census data but focus on three-generation transmission rather than geographic differences.

Finally, this paper contributes to the growing literature on intergenerational occupational and income mobility in the U.S. over the 20th century (Solon, 1992; Zimmerman, 1992; Ferrie, 1996; Mazumder, 2005; Long & Ferrie, 2013; Mazumder, 2016; Bratberg et al., 2017; Feigenbaum, 2018; Bailey et al., 2020b; Connor & Storper, 2020; Song et al., 2020; Abramitzky et al., 2021; Collins & Wanamaker, 2022; Tan, 2023; Ward, 2023; Jácome et al., 2025; Davis & Mazumder, 2026). While educational mobility is a key input into occupational and income mobility (Blau & Duncan, 1967; Sewell & Hauser, 1975; Hout & DiPrete, 2006), knowledge about the interrelationships of these processes has been limited. Our estimates allow direct comparisons and reveal that many places in the U.S. are characterized by high rates of educational mobility but low rates of income mobility and vice versa. For example, among more populous counties, Sacramento, California, ranked second in the country in terms of upward educational mobility in the early 20th century but did not appear in the top-ten locations of upward occupational-income mobility (Tan, 2023). In contrast, Fairfax, Virginia, ranked third in upward income mobility in the late 20th century (Chetty et al., 2014b), but did not rank among the top ten locations for educational mobility. Our estimates of educational mobility provide an important and complementary lens through which to understand the evolution of economic opportunity in the 20th-century U.S. and motivate future work examining how and why the processes of educational, occupational, and income mobility diverge.

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

Education is a key mechanism linking parents' to children's outcomes. It is strongly related to parents' social and economic resources and also shapes children's opportunities through job preparation, credential attainment, and marital matching (Blau & Duncan, 1967; Sewell & Hauser, 1975; Hout & DiPrete, 2006). However, educational mobility is conceptually distinct from occupational and income mobility: education captures earnings and occupational *potential* rather than realized *outcomes*, and thus differs when individuals choose lower-status or lower-earning careers.

Education also alleviates several measurement challenges inherent in using income and occupation. First, most adults complete their education by their mid-20s, and the level of education remains stable over their lifetimes. This makes education less susceptible to transitory economic shocks in adulthood, which have limited studies of occupational and income mobility (Solon, 1999; Mazumder, 2005; Haider & Solon, 2006; Mazumder, 2018; Ward, 2023).

Second, education is measured for *all* adults, not just those who are employed, have an occupation, or earn wages. This is especially important for intertemporal comparisons over the 20th century, because wages and occupations are missing for a large number of Americans historically, particularly for certain subgroups. In 1940, only 14 percent of married women participated in the labor force, limiting measurement of their income and occupational mobility (Goldin, 1990; U.S. Bureau of Labor Statistics, 2023).[3] Only 17 percent of adult men were farmers or farm laborers in 1940, and much of their in-kind or farm earnings were not reported in the 1940 Census. The issue of unmeasured income also affects the 17 percent of prime-age men who were self-employed in 1940, and one quarter reported receiving over 50 dollars in non-wage income (although the amount of that income was not asked). Although these groups are not mutually exclusive, they show that a substantial part of the U.S. population in 1940 did not have an occupation or wage earnings, and many had incomplete income information. In addition, changing selection into groups with occupations and wage earnings over the 20th century limits the interpretation of intertemporal comparisons. Using educational attainment, which is measured for most Americans, allows a broad characterization of the evolution of intergenerational mobility over the 20th century.

A. Available Measures of Education

The 1940 Census was the first decennial census to ask about educational attainment, defined as the highest year of school or degree completion. Self-reported educational attainment is also available in every census year after 1940 and in every year of the *ACS*, with minimal changes to the definition over

[3] Limited information on women's own incomes or occupations limits analyses of women's intergenerational mobility to indirect methods using their husbands' or fathers' outcomes (Olivetti & Paserman, 2015; Espín-Sánchez et al., 2023; Bailey & Lin, 2025; Eriksson et al., 2026) and, by extension, limits samples to married women.

time. Since 1990, respondents with no more than a high school education were classified by their highest year of education completed, whereas those with more than a high school education were classified by their highest degree earned. Because this definitional change applies to everyone in the later 20th-century cohorts, we expect it to have little effect on educational *rankings* in our study.

Despite its advantages, several issues in measuring education affect our analysis. First, we measure parents' education in the 1940 Census at older ages than their children, raising concerns about the influence of "educational creep," the well-known tendency for educational attainment for the same birth cohort to rise over time, on our analysis (Margo, 1986). To address this concern, we later show that measuring parents' education at different ages has negligible effects on our results. Second, most children have two parents and, therefore, two measures of parental education, and theory provides little guidance about which measure to use. Our main results use the education of the most educated parent as the primary measure, which aligns with the literature and allows us to characterize education when only one parent is observed (especially for late 20th-century cohorts, when single-parent households are more common). We later show that our results are unchanged when using alternative measures of parental education. Third, one year of schooling in the early 20th century may not be directly comparable to a year of schooling in the later 20th century, in terms of curriculum, resources, or instructional quality. Because one year of education is of similar quality within a cohort, our use of educational rank will partially absorb these quality differences. Moreover, differences in educational quality will show up in our analysis to the extent that they result in greater educational attainment. However, readers should keep in mind that completed years of education may miss aspects of quality that are not reflected in attainment.

The central challenge in studying the changing geography of intergenerational educational mobility has been that parents and their adult children are not linked in large datasets when they do not co-reside. For example, parents' attainment is unobserved for about 84 percent of adults aged 21 and older in the 1940 Census and 88 percent in the 2018–2022 *ACS*. Although retrospective surveys or longitudinal data have these links, their sample sizes are too small to examine the evolution of educational mobility by county. We address this data limitation by creating two large-scale, linked census datasets for the early- and late-20th-century cohorts and by validating estimates derived from these samples using nationally representative surveys.

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

Our early 20th-century cohorts are constructed from the public Social Security application (Form SS-5) records contained in the Social Security Numerical Identification (Numident) Files released by the National Archives and Records Administration. These data cover around 40 million individuals who ever applied for a Social Security Number (SSN), died prior to 2007, and whose deaths were not state-reported—the near universe of individuals who died between 1988 and 2007 (Goldstein & Breen, 2022). Our estimates suggest that the SS-5 data contain around 50 percent of the 1920 birth cohort (Appendix Figure A1.A) and have excellent geographic coverage (Appendix Figure A1.B).[4] These data contain full name, exact birth date, sex, race, and birthplace as well as the full birth names of each Social Security applicant's parents, which is crucial for linking children to their parents.[5] Our early 20th-century cohorts restrict this sample to children born from 1910 to 1919 because they have high coverage rates in the SS-5, are likely to have at least one parent alive in the 1940 Census, and are old enough in 1940 (ages 21–30) to have completed their education.[6] We also restrict our sample to children born in the U.S. because immigrant children may have completed their schooling abroad (i.e., are not influenced by educational investments in the U.S.).

Our methodology linking these data to other sources follows the supervised machine learning approach of the LIFE-M project (https://life-m.org), which offers the benefits of at scale while limiting false positives to less than 3% (Bailey et al., 2023). We first link siblings together in the SS-5 data using parent names. Then, we link parents' *and* children's information (names, ages, and birth states) as families to the 1900–1930 Censuses. LIFE-M's family-based linking is highly accurate and increases linking rates

[4] Fetter et al. (2025) show that SS-5 records are not only representative with respect to state of birth but also state of death.

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

[6] Appendix Figure A2 shows that the educational gains past age 20 were small for individuals born in 1910–1919.

(Ruggles et al., 2025). The 1900–1930 Censuses add information on parents' year and place of birth, which we then leverage to link parents to the 1940 Census: (1) as families, (2) as couples, and (3) as individuals. Linking parents in multiple ways helps us find them in the 1940 Census, irrespective of their household structure, and also helps us identify conflicts and cull bad links from the data. We also link SS-5 children to the 1940 Census based on their full name, age (based on the exact date of birth), and birthplace information. We attempt to link girls using both their birth and married names and then reconcile the data when conflicts arise, which again helps cull bad links. (See Appendix A for more details about the linking process.)

Our final early 20th-century sample contains around 1.5 million children born in the U.S. between 1910 and 1919 with at least one parent linked to the 1940 Census and their county of residence observed in 1920. The final match rate for individuals with all information of interest is 18.4 percent of all individuals meeting our sample criteria in the SS-5 records (Appendix Table A2). Almost 50 percent of this sample are women. Because linked samples are not generally representative of the population, we use inverse propensity scores to reweight the sample to match the sex-by-race-by-cohort and education distributions in 1940, using the procedure described in Bailey et al. (2020a). These weights adjust for the fact that earlier cohorts are underrepresented in SS-5 records and the fact that Black Americans and less-educated individuals are underrepresented in the final sample because they are harder to link. Appendix Table A3 (column 5) shows that the reweighted linked samples appear well balanced with the population across age, race, and educational attainment. (See Appendix B for more details on these weights).

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

We construct an analogous linked sample for late 20th-century cohorts using high-quality Personal Identification Keys (PIKs) created by the Census Bureau and available in the restricted Federal Statistical Research Data Centers (FSRDCs).[7] PIKs allow us to combine restricted SS-5 records, the 2000

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

Census, and the 2001–2022 *ACS*. We begin by selecting a sample of children ages 3 to 18 in the 2000 Census who are living with at least one biological parent. We use their demographic information (year of birth, sex, race) from the SS-5, which tends to be more reliable than self-reported information. Then, we find these children as adults in the *ACS* after they turn 25 and observe their completed education.

The 2000 Long-Form Census contains parents' education for a 1-in-6 sample of the population. For parents not in the 2000 Long-Form Census, we locate them in the *ACS* to obtain their completed education. Our final sample contains 2.1 million children born in the U.S. between 1982 and 1997, for whom we observe the adult child's and at least one parent's educational attainment and the child's county of residence in 2000. The 2.1 million children represent about 3.5 percent of children born in the U.S. in 1982–1997, according to the 2000 Census. (See Appendix A for more details about the linking process.)

The low coverage rate is not a reflection of poor PIKing rates but rather the fact that all children, as well as parents who do not appear in the 2000 Long-Form Census, are linked to the *ACS*, which is a 1-percent sample of the population each survey year. Our additional requirement that children be at least age 25 means that younger cohorts are slightly underrepresented (since they are aged 25 or older in fewer *ACS* years). Consequently, we similarly construct inverse propensity scores to reweight the sample to match the sex-by-race-by-cohort distribution in the 2000 population. (See Appendix B for details.)

D. Benchmarking Linked Samples against Nationally Representative Surveys

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

E. Generating County-Level Estimates

Because county boundaries have changed over time, our county-level analysis restricts the sample to counties with stable boundaries over the 20th century. We define stable boundaries as those in which 95 percent of their geographic area under 1920 boundaries overlaps with their area under 2000 boundaries and vice versa, using a crosswalk by Eckert et al. (2020). We also restrict the sample to counties with at least 20 parent-child observations. Disclosing county-level estimates for the 1982–1997 cohorts from the FSRDC required following Census Bureau guidelines to protect respondents' confidentiality. Following Chetty and Friedman (2019), we added Gaussian noise to the unweighted county estimates and associated standard errors, setting the "privacy parameter" to 8. This process results in less than 5 percent of the total variation in the adjusted educational mobility estimates occurring due to noise.[8] We also present estimates using empirical Bayes methods to reduce the influence of sampling noise or disclosure-avoidance noise infusion.

II. Using Ranks to Measure the Evolution of Educational Mobility over the 20th Century

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

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

The slope coefficient, β, captures persistence in educational attainment across generations, which may reflect both low upward mobility and low downward mobility (Solon, 1999; Black & Devereux, 2011). The slope coefficient captures the product of the correlation between child and parent outcomes as well as

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

the ratio of the standard deviation of completed years of education in the child's generation relative to the standard deviation of completed years of education in the parent generation. The slope coefficient will, therefore, increase as inequality in the outcome of interest rises, even if the intergenerational correlation does not change. To account for rising inequality across generations, researchers in the income mobility literature have alternatively reported the correlation coefficient or correlation between parent and child ranks (Dahl & DeLeire, 2008; Chetty et al., 2014a).

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

$$R_i^c = \gamma + \delta R_i^p + \varepsilon_i, \tag{2}$$

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

The rank-rank approach offers some important benefits in the context of our analysis. First, this linear parameterization helps smooth the lumpy educational distribution (many individuals have the same number of years of completed education) and parameterize educational mobility at percentiles not directly observed in two time periods. In the early 20th century, more than a quarter of children reported 12 years of education, and almost one-third of their parents reported 8 years of education as their highest level of completion (see Appendix Figure A4 for parent and child educational distributions). This lumpiness means that comparisons at the same percentile of the parental educational distribution are not well defined at many points for our two cohorts, limiting intertemporal comparisons. The linear parameterization, however, allows us to predict the percentile rank of a child whose parent is at the same percentile of educational attainment in both periods. Second, the rank-rank approach allows us to make comparisons with the income mobility literature. As such, we follow Chetty et al. (2014b) and define *upward* educational mobility as the expected rank of children with parents at the 25th percentile of the parental education distribution, or $E[R_i \mid P_i = 25] = \gamma + 25\ \delta$.

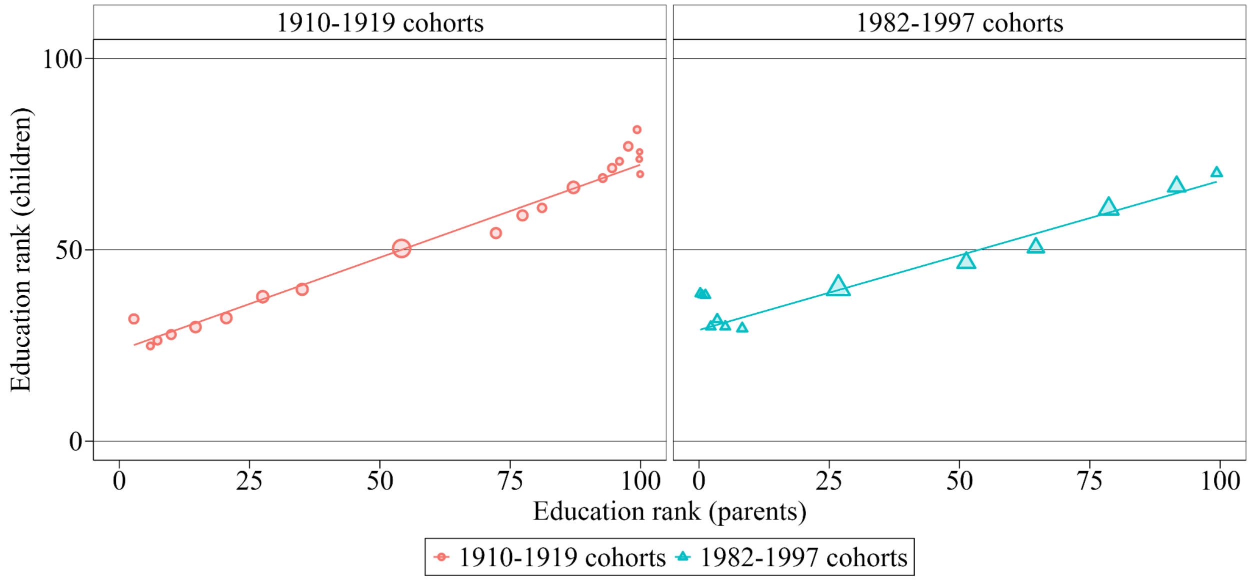
Figure 1 plots these outcomes using the percentile midpoint of each educational category, and Appendix Figure A5 presents the transition matrices for each cohort. Similar to the results for income (Chetty et al., 2014a), the expectation of a child's years of education conditional on parents' years of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

This linear relationship also holds when making different assumptions about percentile ranks; when using the lowest rank within an educational category for parents (lower bound, LB) and the highest rank for children (upper bound, UB); the UB for parents and the LB for children; as well as two other combinations of these ranks (see Appendix Figure A6). In each of these cases, the slopes of the fitted line for the midpoint definition is statistically indistinguishable from other choices (p-value for the test of the joint hypothesis that all slopes are equal is 0.54 for the 1910–1919 cohorts and 0.35 for the 1982–1997 cohorts), which implies that our estimates of changing persistence are robust to the choice of measure. Our measure of upward mobility in a given period is sensitive to our use of the midpoint, but our conclusions about *changes* in the expected rank at the 25th percentile are not (Appendix Table A4).

Table 1 presents estimates of educational mobility based on the mid-point definition derived from our rank-rank specification for the early (panel A) and later 20th-century cohorts (panel B). We augment these measures with a commonly used level-level measure of persistence and the share of children moving from the bottom 40 to the top 60 percent.[9] The latter measure provides another way to validate our results against Card et al.'s (2022) findings using an unlinked sample.

[9] For children born in 1910–1919, the 40-60 measure of upward mobility captures the likelihood of attaining at least 10 years of education conditional on parents having at most 7 years and, for children born in 1982–1997, the likelihood of attaining at least 13 years conditional on parents having at most 12 years. We choose the 40th percentile as it is one of the few thresholds that can be (approximately) observed in both the 1910–1919 and 1982–1997 birth cohorts and their parents. See Appendix Figure A4.

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



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

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

Comparable statistics from nationally-representative survey data support these findings. Although survey estimates are noisier, both educational persistence and upward educational mobility from the linked sample fall within the 95-percent confidence intervals of the survey estimates. Importantly, we fail to reject the equality of changes over time in both samples, which supports our conclusions.[10]

[10] See also Appendix Figure A7.A that shows the rank-rank slope initially declined, changed little for several decades, and declined again more recently. The path for the level-level slope is broadly similar. These estimates are consistent with Ferrie et al. (2021), who report a level-level slope of 0.37 for the 1922–1940 cohorts and 0.36 for the

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

The second test examines the sensitivity of our findings to using different measures of parents' education. The data show that using the educational attainment of the most educated parent, the least educated parent, or the mean education of the two parents (when present) has modest effects on the rank-rank slope and upward mobility at the 25th percentile (Appendix Table A6).[11] However, these different measures affect both parameters similarly in the early and later 20th century, meaning that our conclusions about changes in educational mobility over time are robust to these alternatives.

1955–1990 cohorts using linked census-survey data. Figure A7.B shows that the expected rank at the 25th percentile has seen small changes over the 20th century.

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

Table 1. Educational Mobility Estimates by Cohort and Sample

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

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

In summary, our findings are robust to a number of sensitivity checks, and we conclude that both relative mobility increased and upward educational mobility increased for American children of parents with below-median educational attainment over the 20th century. The next section uses our large sample sizes and rank-rank parameterization to measure the changing geography of educational mobility over the 20th century.

III. The Changing Geography of Relative and Upward Educational Mobility

Our rank-rank parameterization and large sample sizes allow us to provide a novel characterization of the changing geography of relative and upward educational mobility over the 20th century within a common measurement framework. To date, studies of relative educational mobility over time have used surveys (Hertz et al., 2007; Fletcher & Han, 2019) or linked datasets (Ferrie et al., 2021) to document changes in intergenerational educational mobility at the *national* level. Hilger (2015) uses 26–29-year-old children residing with their parents to describe national and state changes in educational mobility, but concerns about sample selection limit the external validity of these findings.

Figure 2 shows large geographic disparities in relative educational mobility in the early and late 20th century. Relative mobility—measured by the inverse of the rank-rank slope—tended to be lower in the South in the early 20th century and higher in the Northeast, the Midwest, and especially the West.[12] However, considerable variation in relative mobility is evident *within* states, with each state containing counties with both the lowest and the highest rates.

These varied within-state patterns are less evident for upward educational mobility. In the early 20th century, upward mobility was uniformly lower in the South, with the former slave states almost completely coinciding with the lowest expected percentile ranks (shaded in orange and red). (Only pockets of northern Missouri, western Texas, and southern Florida showed some upward educational mobility). The lightest shade of orange in these areas indicates little to no upward mobility (an expected

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

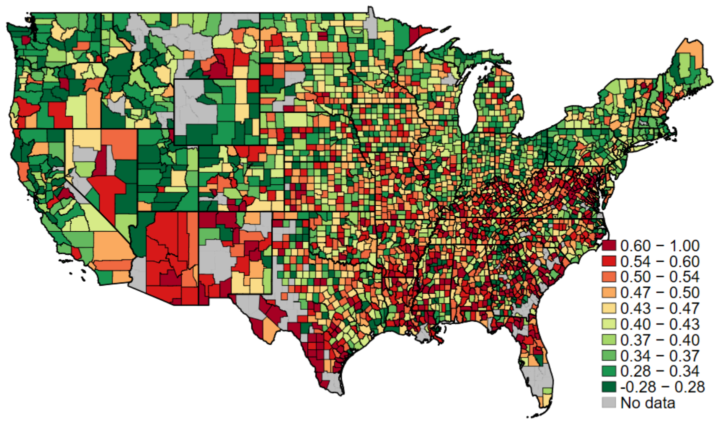
percentile rank of 25–28 for a child of a parent at the 25th percentile), and the darkest red indicates downward mobility (an expected percentile rank of 10–22 for a child of a parent at the 25th percentile). That is, outside of a handful of counties, children in former slave states saw *little to no* upward educational mobility in the early 20th century.[13] In contrast, counties in the West, Northeast, and Industrial Midwest had very high upward educational mobility rates in this period, as indicated by the light and dark green shading. Historically, opportunities for upward educational mobility depended heavily on where children grew up.

By the end of the 20th century, the geography of educational mobility had changed dramatically. So much so that the correlation between upward mobility in the two periods is only 0.25, while the corresponding correlation for relative mobility is just 0.09. Moreover, sharp regional convergence occurred in both measures. Upward mobility improved in almost every county in the South, with the expected percentile rank averaging in the mid-30s for former slave states. But this is not just a story of the South catching up. High rates of upward educational mobility in the West and the Northeast fell toward the national average. Similarly, very high rates of upward educational mobility deteriorated in the Industrial Midwest, mirroring its declining economic fortunes. Along with the South, counties in the Great Plains (Iowa, Kansas, Minnesota, Nebraska, and the Dakotas) experienced significant gains in upward educational mobility.[14] As a result, the interquartile range in upward mobility collapsed, from 15 percentiles to 6 percentiles, and cross-county variance declined by 76 percent over 75 years (Appendix Figure A8). This pattern of regional convergence also holds for the rank-rank slope and the 40-to-60 measure of upward mobility, with cross-county variances declining by 50 and 74 percent, respectively (Appendix Figures A9 and A10). A key takeaway from this analysis is that where a child grows up today matters *much less* for upward educational mobility than in the past.

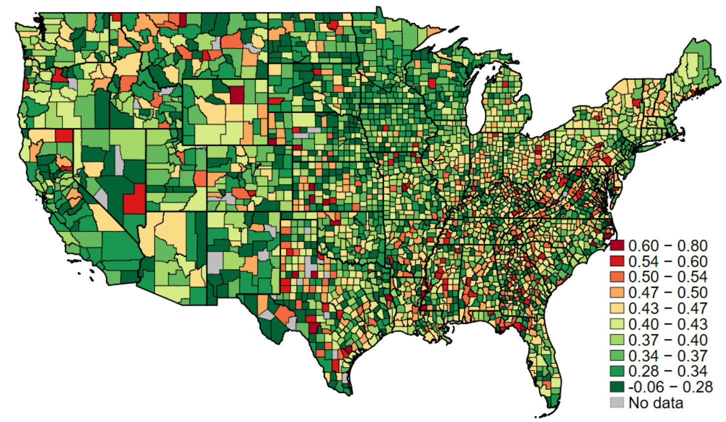
[13] The geographic pattern in upward mobility is similar to Card et al. (2022), who document county-level variation in upward mobility among teens residing with their parents in 1940. The correlation between our measure of upward mobility at the 25th percentile and theirs is 0.77. This imperfect correlation, when using the same data (albeit for slightly different cohorts), is reassuring and highlights the value of measuring completed education.
[14] Chetty et al. (2014b) also document that upward income mobility is highest in the Great Plains and lowest in the Southeast in the later 20th century.

Figure 2. The Geography of Educational Mobility by Childhood County of Residence over the 20th Century

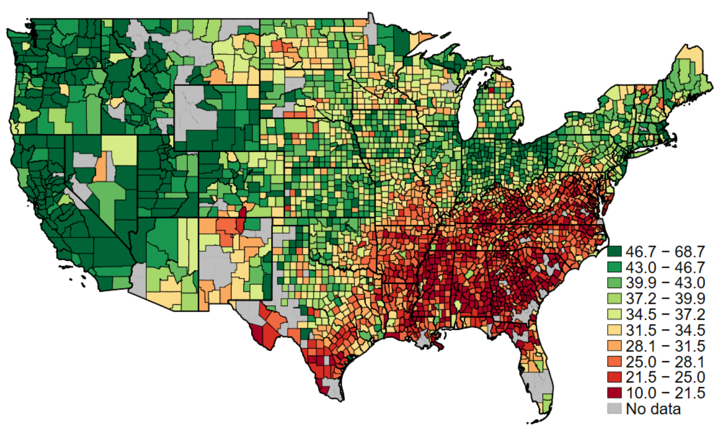
A. Rank-Rank Slope, 1910–1919 Cohorts



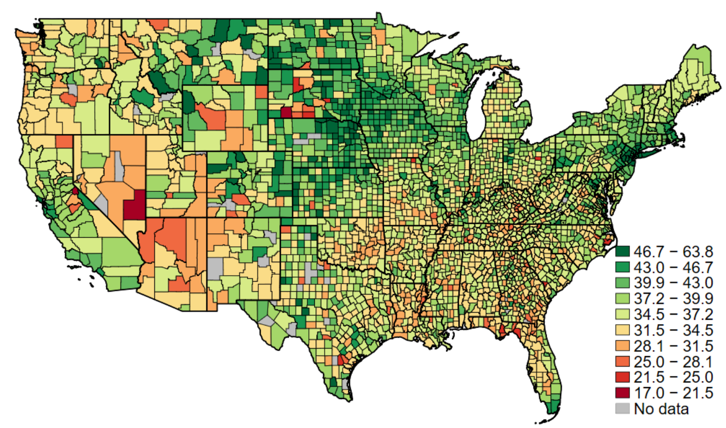
B. Rank-Rank Slope, 1982–1997 Cohorts



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



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



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

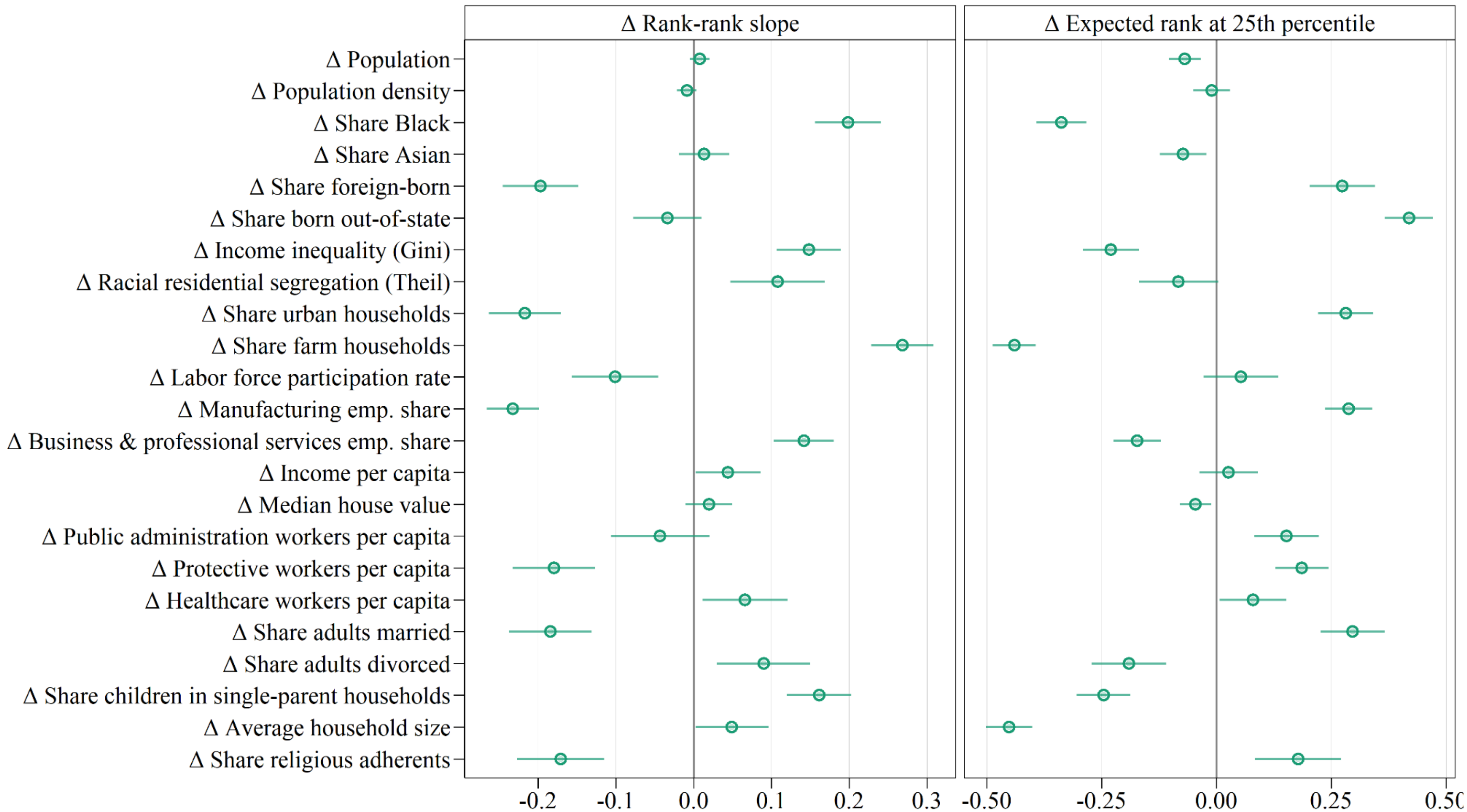
Information about the geography of intergenerational educational mobility also allows us to examine the county correlates of these changes. Following Chetty et al. (2014b), we created a rich set of county-level measures that are both available and consistently measured across the early and late 20th centuries. Figure 3 shows that greater relative and upward mobility are associated with rising economic development (e.g., share urban, manufacturing share); declining income inequality and residential segregation; a declining share of Black Americans; a rising share of immigrants; a declining share of single-parent households and falling household size; and a rising share of religious adherents. In the next section, we examine the causal role of investments in public education.

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

It is natural to think that increases in public resources for schools likely raised opportunities for children of less educated parents and in less developed areas (like the South). But this need not be the case. Rather than increasing educational mobility, rising investments in public education may have reflected higher teacher (and women's) wages and more years of school completed by all students. Moreover, larger increases in per-child educational resources may have benefited the most advantaged students (e.g., prioritizing smaller class sizes for high achievers), amplifying (rather than reducing) the persistence of parental educational advantages.

To advance the literature studying this question, we use a state-border design to help isolate the causal effect of educational spending on mobility and to minimize the influence of other changes in local characteristics (see Figure 3). This research design is motivated by the observation that around half of the cross-county variation in changing upward educational mobility over the 20th century is explained by state-level factors (i.e., state fixed effects). Said another way, otherwise very similar communities have sharply different resources for public education due to differences in a host of state educational policies (Black, 1999; Kane et al., 2006; Knight & Schiff, 2019; Card et al., 2022; Lefgren et al., 2022).

Figure 3. The Association of Changes in County Characteristics and County Educational Mobility



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

Our analysis relates changes in educational mobility in county pairs on opposite sides of a state border, Δy_{cp}, to changes in state educational resources, $\Delta \boldsymbol{Z}_{s(c)}$, within the following linear regression framework,

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

where Δy_{cp} denotes changes in educational mobility in county c in state s and part of county pair p, and δ_p is a set of county-pair fixed effects. We assign county pairs using the Census Bureau's 2010 county adjacency crosswalk and use only counties with stable boundaries and at least 20 parent-child pairs (U.S. Census Bureau, 2010). Because some counties are adjacent to more than one county, counties may appear in more than one county pair (Dube et al., 2010).

We use three relative measures of educational investments: (1) a standardized measure of the increase in the number of K-12 teachers per child aged 5 to 18; (2) a standardized measure of the increase in the average salaries of K-12 teachers (in real dollars); and (3) a standardized measure of the increase in the number of instructors at postsecondary public colleges and universities per child aged 5 to 18, which is included to capture growing state investments in higher education (Goldin & Katz, 2008).[15] The coefficients of interest, θ, document the relationship between educational mobility and relative increases in state spending on primary, secondary, and postsecondary public education.[16] (See Appendix C for data sources and variable construction details.)

These measures are an incomplete characterization of the broader set of determinants of public educational investments in both periods, but we choose them because they are well-measured and

---

[15] Over the 20th century, publicly funded postsecondary institutions have become increasingly important. In 1931, enrollment in public postsecondary institutions accounted for 50 percent of total enrollment. By 2021, this number had risen to 73 percent of total enrollment (Goldin, 2006; U.S. Department of Education, 2023a). In addition, these public institutions are more likely to serve students from more disadvantaged backgrounds. In 2022–2023, the average annual tuition and required fees at 4-year public institutions was $9,834 (minimum of in-district and in-state tuition), compared to $40,713 at 4-year private non-profit institutions and $18,241 at 4-year private for-profit institutions. The corresponding figures for 2-year institutions are $4,027, $19,517 and $16,301, respectively (U.S. Department of Education 2023c).

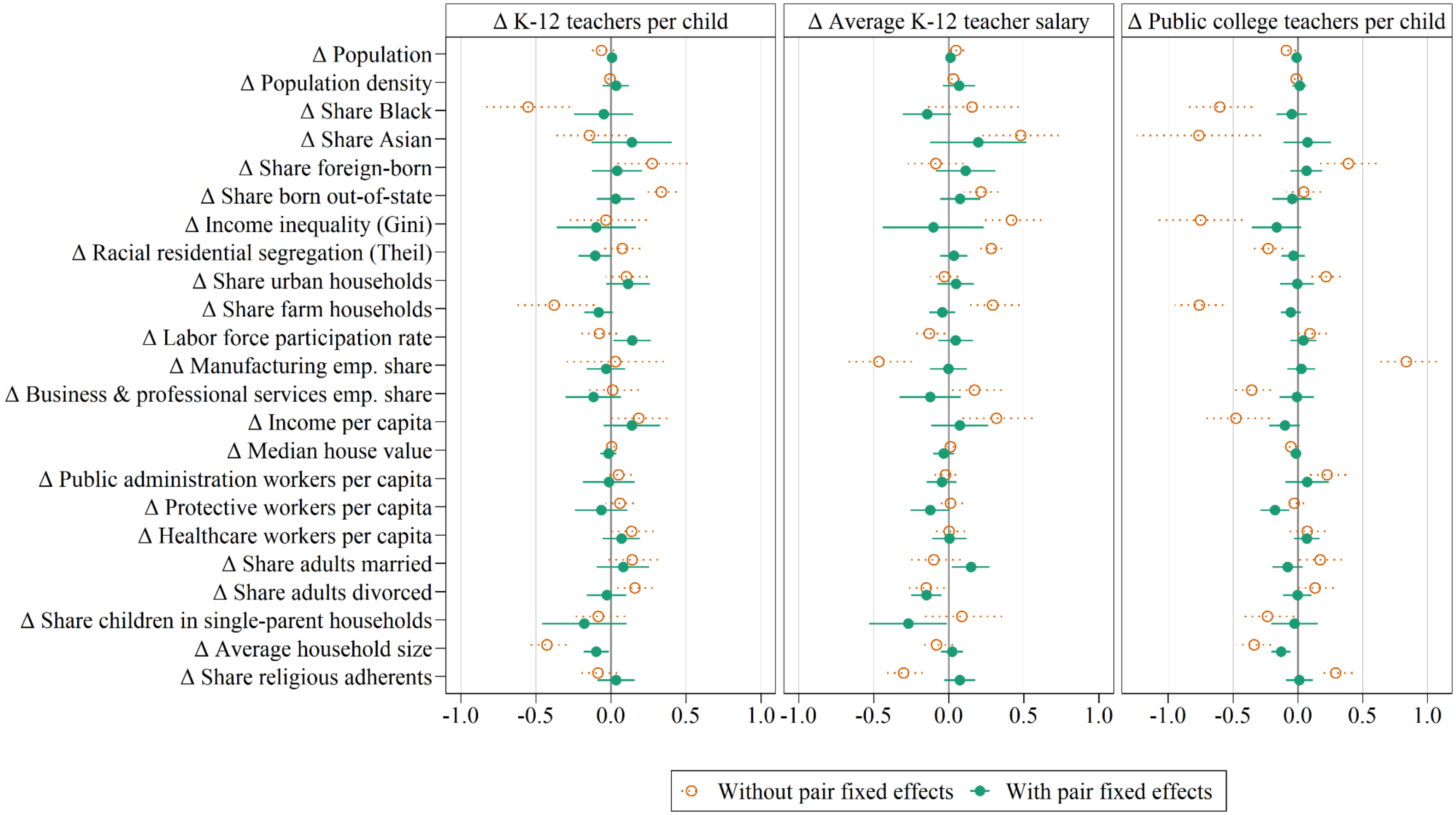
[16] Appendix Figure A11 presents maps of the counties used in the analysis and provides a visual description of the identifying variation. Note that counties can be part of multiple county pairs, because they may adjoin more than one county on the other side of the state border.

comparable across two points in the 20th century. Moreover, the K-12 measures have been widely used as indicators of public school quality (Card & Krueger, 1992, 1996; Card et al., 2022). (See Appendix Figure A12 for the correlation of these measures with educational mobility.)

A key assumption for interpreting the results as causal is that, within county pairs, the only reason for changes in relative educational resources to differ across state borders is state educational policy. That is, omitted factors affecting changes in state educational policy do not covary with changes in mobility, or $\text{cov}(\varepsilon_{cp}, \Delta\boldsymbol{Z}_{s(c)} | \delta_p)=0$. We provide evidence on this assumption by examining whether changes in observed county characteristics, ΔX_c in Figure 3, are correlated with changes in educational resources across state borders and within county pairs. Although there are many statistically and economically significant associations if we omit county-pair fixed effects, Figure 4 shows that these associations become small and generally indistinguishable from zero once we restrict comparisons to county pairs (solid markers). Out of 69 coefficients (23 characteristics × 3 education measures), only seven are statistically different from zero at the 5-percent level with county-pair fixed effects—no more than expected by chance. This balance across a wide range of changing county characteristics is reassuring and supports the validity of the research design.

We also directly test for the role of non-education state policies by assembling a database of candidate policies, including women's suffrage laws (Lott & Kenny, 1999; Miller, 2008) and state minimum wage laws in the early 20th century (mostly targeting women's wages) and later 20th century (Marchingiglio & Poyker, 2024; Federal Reserve Bank of St. Louis, 2024). Using the same specification as in (3), we find no evidence that having a state-level women's suffrage law prior to 1919 or having a higher minimum wage in either period predicts changes in greater K-12 or postsecondary public investments in education. This evidence helps assuage concerns that differences in public health investments (Miller, 2008) or labor demand drive our findings. However, we find an association between changes in K-12 education and compulsory schooling laws (Acemoglu & Angrist, 2000; Lleras-Muney, 2005; Stephens & Yang, 2014; Rauscher, 2016), consistent with states using policy bundles to promote the education of their youth. In what follows, we therefore interpret our estimates as reflecting differences

Figure 4. Balance Checks for the State-Border Design



Notes: Each dot is the coefficient from a regression of the change in a county characteristic (indicated in row label) on the change in a state public education measure (indicated in the panel title). Hollow dots are from regressions that exclude county pair fixed effects, whereas solid dots include county pair fixed effects. Here, counties are weighted by the inverse of the variance of changes in expected rank at the 25th percentile at the county level, but estimates are similar when weighting by the inverse variance of the rank-rank slope. All variables are standardized to have a mean of 0 and standard deviation of 1. 95-percent confidence intervals are based on robust standard errors, clustered at the state border level.

in state education policy *bundles* (including the effects of correlated but unmeasured state education policies, such as compulsory education laws) rather than as causal effects of individual educational measures *per se*.

Table 2 shows that these proxies for state public education resources are associated with increases in relative and upward mobility across all counties (column 1), and that these associations are not quantitatively different in the border sample (column 2). The latter finding suggests that these relationships in border counties are fairly typical of those in other locations.[17] Including county-pair fixed effects (column 3), however, largely eliminates these associations for relative mobility, whereas the positive relationship persists for upward mobility. These findings are more precise when shrinking the estimates toward the state mean to account for sampling variation and disclosure-avoidance infused noise, following the methodology of Armstrong et al. (2022) (column 4, see also Appendix Figure A13).

The magnitudes of the estimates in panel B suggest that a 1-standard-deviation larger increase in the number of per-child K-12 teachers or per-child public college instructors led to a 1 percentile-rank larger increase in upward mobility, and a 1-standard-deviation larger increase in teachers' average salaries led to a 1.8 percentile-rank larger increase in upward mobility. The finding that average teacher salaries are especially strong predictors of upward educational mobility aligns with Card et al. (2022), who document this for cohorts born around 1920 in the 1940 Census.

Overall, these three measures of public educational investments account for only 1 percent of the variation in relative mobility and roughly 10 percent of the variation in upward mobility within very similar counties across state borders. (Appendix Table A9 reveals similar findings if we measure upward mobility using the bottom 40 to top 60 percent mobility). At the state level, predicted changes in upward mobility (Table 2.B, column 3) explain around 20 percent of the variation in observed changes in upward mobility across states (Appendix Figure A14).

[17] Appendix Table A7 shows that border counties are not special in terms of observable characteristics either.

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

	(1)	(2)	(3)	(4)
A. Dependent variable: Change in county rank-rank slope				
Change in K-12 teachers per child (ages 5–18)	-0.016***	-0.014*	-0.012	-0.017*
	(0.005)	(0.008)	(0.010)	(0.009)
Change in average K-12 teacher salary	0.016***	0.013*	-0.004	-0.006
	(0.005)	(0.007)	(0.010)	(0.009)
Change in public college instructors per child (ages 5–18)	-0.018***	-0.023***	-0.009	-0.010
	(0.005)	(0.006)	(0.008)	(0.007)
R^2	0.052	0.058	0.689	0.804
Partial R^2			0.010	0.049
B. Dependent variable: Change in county expected rank of children with parents at 25th percentile				
Change in K-12 teachers per child (ages 5–18)	4.49***	4.34***	1.03*	1.18*
	(0.38)	(0.57)	(0.60)	(0.60)
Change in average K-12 teacher salary	1.33***	1.64***	1.84***	1.97***
	(0.43)	(0.43)	(0.62)	(0.61)
Change in public college instructors per child (ages 5–18)	2.45***	2.78***	0.96*	1.08**
	(0.38)	(0.43)	(0.51)	(0.51)
R^2	0.278	0.298	0.824	0.852
Partial R^2			0.080	0.121
Border sample		✓	✓	✓
County pair fixed effects			✓	✓
Bayesian shrunk estimates				✓
N	2,881	2,344	2,344	2,344

Notes: Column 1 shows the estimates from a regression of county-level rank-rank slope (panel A) or expected rank at the 25th percentile (panel B) on state public education resources, which are standardized to have a mean of 0 and a standard deviation of 1. Column 2 limits the sample to border counties (units of observation are county-by-border pairs), where counties may appear in more than one pair. Column 3 includes county-pair fixed effects as in equation 3. Column 4 replaces the county mobility measures with their Empirical Bayes analogs to reduce the influence of sampling noise or disclosure-avoidance noise infusion. The partial R^2s capture the share of the residual variation in county educational mobility explained by the state covariates after controlling for county pair fixed effects. Counties are weighted by the inverse of the variance of the relevant county mobility estimates. Robust standard errors in parentheses are clustered at the state-border level in columns 3 and 4. Stars indicate the level of statistical significance of the estimates: ***$p < 0.01$, **$p < 0.05$, *$p < 0.10$.

Altogether, these findings suggest that, while state investments in public education were an important driver of upward educational mobility over the 20th century, they had little effect on the persistence of parental educational advantage. Larger increases in state resources for public education affected mobility *similarly* for children of parents across the education distribution, and families with more education maintained their children's relative standing, likely through complementary private investments, residential sorting, and social networks (Becker & Tomes, 1986; Solon, 2004).

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

The rapid expansion of universal public education in the U.S. and growing educational attainment are well-known features of the human capital century (Goldin, 2001; Goldin & Katz, 2008). However, the lack of large-scale longitudinal data has limited our understanding of whether changing attainment and rising public educational spending have raised educational opportunities, strengthened the role of hard work and merit, and weakened the transmission of parental advantage.

This paper constructs new large-scale longitudinal datasets, allowing a novel characterization of the changing landscape of educational opportunity over 75 years and a quantification of the role of educational spending in shaping these opportunities. Similar to national survey evidence, we find that intergenerational educational mobility changed modestly at the national level over the period. However, these modest changes at the national level mask large geographic differences in the evolution of educational mobility over time. First, we find that geographic convergence in educational mobility was astounding: while educational mobility fell in the previously high-mobility areas of the West, Northeast, and Industrial Midwest, it rose in historically lower-mobility areas of the South and Great Plains. As a result, where children grow up today plays a much smaller role in their educational opportunities than it did in the early 20th century.

Second, we find that investments in public education played a meaningful role in shaping changes in educational opportunity across the 20th century. Using a state-border design that limits the influence of unobserved local economic conditions, we find that state policies that raised the number of K-12 teachers, K-12 teacher salaries, and public college instructors raised upward educational mobility but had little effect on relative mobility. Put differently, state investments affected educational outcomes *similarly* for all children, meaning that—while all children gained—children of parents with greater educational advantages maintained their relative standing. One limitation of this study is that, while educational attainment is associated with the quality of education, we only observe *years* of education in the data. Recent work by Bleemer and Quincy (2025) highlights how differences in the types and quality of college majors became stratified by parental income over the 20th century, implying differing returns to college education by social status. These findings complement our work, suggesting that differences in educational quality and college majors among children from lower socioeconomic status families may curtail gains in mobility in terms of educational attainment.

These geographic patterns in educational mobility resemble but also diverge from patterns of intergenerational income mobility, both in the early 20th century and in the later 20th century. Some resemblance is expected: through learning and opportunities for better careers, greater educational mobility should increase income mobility. Yet, educational and income mobility are not perfectly correlated in either period. For the 1910–1919 cohorts, the correlation between our estimates of upward educational mobility and county estimates of upward income or occupational mobility in Massey and Rothbaum (2021) or Tan (2023) is around 0.68.[18] For the 1982–1997 cohorts, the correlation between our county estimates of upward educational mobility and the upward income mobility estimates in Chetty et al. (2014b) is around 0.50 and, for estimates of relative educational mobility, only 0.18. Both today and in the past, many places in the U.S. are characterized by high rates of educational mobility but low rates of income mobility and vice versa. Understanding why educational and income mobility diverge and the factors that mediate this relationship remain important questions for future research.

[18] Note that our study is for a different sample than previous studies of occupational/income mobility. Tan (2023) studies White men born from 1892 to 1910 and linked across the 1910 and 1940 Censuses, assigning occupation-based income scores to fathers and sons. Massey and Rothbaum (2021) focus on men and women born from 1922 to 1940 linked between the 1940 Census and 1974–1979 tax records. They calculate fathers' income based on self-reported wage income in 1940 and children's income based on Adjusted Gross Income reported in tax data.

VI. References

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

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

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

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

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

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

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

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

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

Becker, Gary S. and Nigel Tomes. 1986. "Human Capital and the Rise and Fall of Families." *Journal of Labor Economics* 4(3, Part 2):S1–S39.

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

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

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

Bleemer, Zachary, and Sarah Quincy. 2025. "Changes in the College Mobility Pipeline Since 1900." *NBER Working Paper 33797*, National Bureau of Economic Research, Cambridge, MA.

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).

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

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

Card, David, and Alan B. Krueger. 1996. "School Resources and Student Outcomes: An Overview of the Literature and New Evidence from North and South Carolina." *Journal of Economic Perspectives* 10(4): 31–50.

Chetty, Raj and John N. Friedman. 2019. "A Practical Method to Reduce Privacy Loss When Disclosing Statistics Based on Small Samples." *AEA Papers and Proceedings* 109: 414–420.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. 2014a. “Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility.” *AER: Papers & Proceedings* 104(5): 141–147.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014b. “Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States.” *The Quarterly Journal of Economics* 129(4): 1553–1623.

Collins, William J. and Marianne H. Wanamaker. 2022. “African American Intergenerational Economic Mobility since 1880,” *American Economic Journal: Applied Economics* 14(3): 84–117.

Connor, Dylan Shane, and Michael Storper. 2020. “The changing geography of social mobility in the United States.” *Proceedings of the National Academy of Sciences* 117(48): 30309–30317.

Couch, Kenneth A. and Thomas A. Dunn. 1997. “Intergenerational Correlations in Labor Market Status: A Comparison of the United States and Germany.” *Journal of Human Resources* 32(1): 210–232.

Dahl, Molly W. and Thomas DeLeire. 2008. “The Association Between Children’s Earnings and Fathers’ Lifetime Earnings: Estimates Using Administrative Data.” Institute for Research on Poverty Discussion Paper No. 1342-08, University of Wisconsin-Madison.

Davern, Michael, Rene Bautista, Jeremy Freese, Pamela Herd, and Stephen L. Morgan. General Social Survey 1972–2024. [Machine-readable data file]. Principal Investigator, Michael Davern; Co-Principal Investigators, Rene Bautista, Jeremy Freese, Pamela Herd, and Stephen L. Morgan. Sponsored by the National Science Foundation. NORC ed. Chicago: NORC, 2024: NORC at the University of Chicago [producer and distributor]. Data accessed from the GSS Data Explorer website at gssdataexplorer.norc.org

Davis, Jonathan M.V. and Bhashkar Mazumder. 2026. “The Decline in Intergenerational Mobility after 1980.” *The Review of Economics and Statistics* 108(1): 1–15.

Dube, Arindrajit, T. William Lester, and Michael Reich. 2010. “Minimum Wage Effects Across State Borders: Estimates Using Contiguous Counties.” *The Review of Economics and Statistics* 92(4): 945–964.

Eckert, Fabian, Andrés Gvirtz, Jack Liang, and Michael Peters. 2020. “A Method to Construct Geographical Crosswalks with an Application to US Counties since 1790.” *NBER Working Paper 26770*, National Bureau of Economic Research, Cambridge, MA.

Espín-Sánchez, José-Antonio, Joseph P. Ferrie, and Christopher Vickers. 2023. “Women and the Econometrics of Family Trees,” *NBER Working Paper 31598*, National Bureau of Economic Research, Cambridge, MA.

Eriksson, Katherine, Gregory T. Niemesh, Myera Rashid, and Jacqueline Craig. 2026. “Marriage and the Intergenerational Mobility of Women: Evidence from Marriage Certificates 1850–1920.” *NBER Working Paper 34821*, National Bureau of Economic Research, Cambridge, MA.

Federal Reserve Bank of St. Louis. 2024. “Federal and State Minimum Wage Rates, Annual.” https://fred.stlouisfed.org/release/tables?eid=243906&rid=387, accessed on December 11, 2025.

Feigenbaum, James J. 2018. “Multiple Measures of Historical Intergenerational Mobility: Iowa 1915 to 1940.” *The Economic Journal* 128(612): F446–F481.

Ferrie, Joseph P. 1996. “A New Sample of Males Linked from the Public Use Microdata Sample of the 1850 U.S. Federal Census of Population to the 1860 U.S. Federal Census Manuscript Schedules.” *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 29(4): 141–156.

Ferrie, Joseph, Catherine Massey, and Jonathan Rothbaum. 2021. “Do Grandparents Matter? Multigenerational Mobility in the United States, 1940-2015.” *Journal of Labor Economics* 39(3): 597–637.

Fetter, Daniel K., Lee M. Lockwood, and Paul Mohnen. 2025. “Long-Run Intergenerational Effects of Social Security.” Working paper.

Fletcher, Jason, and Joel Han. 2019. “Intergenerational Mobility in Education: Variation in Geography and Time.” *Journal of Human Capital* 13(4): 585–634.

Goldin, Claudia. 1990. *Understanding the Gender Gap: An Economic History of American Women.* New York-Oxford: Oxford University Press.

Goldin, Claudia. 1998. “America’s Graduation from High School: The Evolution and Spread of Secondary Schooling in the Twentieth Century.” *The Journal of Economic History* 58(2): 345–374.

Goldin, Claudia. 2001. “The Human-Capital Century and American Leadership: Virtues of the Past.” *The Journal of Economic History* 61(2): 263–292.

Goldin, Claudia. 2006. “Enrollment in institutions of higher education, by sex, enrollment status, and type of institution: 1869–1995.” Table Bc523-536 in *Historical Statistics of the United States, Earliest Times to the Present: Millennial Edition*, edited by Susan B. Carter, Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright. New York: Cambridge University Press.

Goldin, Claudia, and Lawrence F. Katz. 2008. *The Race Between Education and Technology*. Cambridge: Harvard University Press.

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.

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. 2015. "The Great Escape: Intergenerational Mobility in the United States, 1930–2010." *NBER Working Paper 21217*, National Bureau of Economic Research, Cambridge, MA.

Hout, Michael and Thomas A. DiPrete. 2006. "What have we learned: RC28's contribution to knowledge about social stratification." *Research in Social Stratification and Mobility* 24(1): 1–20.

Hout, Michael, and Alexander Janus. 2011. "Educational mobility in America: 1930s–2000s." In *Whither opportunity? Rising Inequality, Schools, and Children's Life Chances*, edited by Greg J. Duncan and Richard J. Murnane, 165–185. New York: Russell Sage Foundation.

Jácome, Elisa, Ilyana Kuziemko, and Suresh Naidu. 2025. "Mobility for All: Representative Intergenerational Mobility Estimates over the 20th Century." *Journal of Political Economy* 133(1): 306–354.

Kane, Thomas J., Stephanie K. Riegg, and Douglas O. Staiger. 2006. "School Quality, Neighborhoods, and Housing Prices." *American Law and Economics Review* 8(2): 183–212.

Knight, Brian, and Nathan Schiff. 2019. "The Out-of-State Distortion." *American Economic Journal: Economic Policy* 11(1): 317–350.

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.

Lleras-Muney, Adriana. 2005. "The Relationship between Education and Adult Mortality in the United States." *Review of Economic Studies* 72(1): 189–221.

Long, Bridget Terry. 2013. "Supporting Access to Higher Education." In *Legacies of the War on Poverty*, edited by Martha J. Bailey and Sheldon Danziger, 93–120. New York: Russell Sage Foundation.

Long, Jason, and Joseph Ferrie. 2013. "Intergenerational Occupational Mobility in Great Britain and the United States Since 1850." *American Economic Review* 103(4): 1109–1137.

Lott, Jr, John R. and Lawrence W. Kenny. 1999. "Did Women's Suffrage Change the Size and Scope of Government?" *Journal of Political Economy* 107(6): 1163–1198.

Manson, Steven, Jonathan Schroeder, David Van Riper, Katherine Knowles, Tracy Kugler, Finn Roberts, and Steven Ruggles. 2024. IPUMS National Historical Geographic Information System: Version 19.0 [dataset]. Minneapolis, MN: IPUMS. http://doi.org/10.18128/D050.V19.0

Marchingiglio, Riccardo and Michael Poyker. 2024. "The Economics of Gender-Specific Minimum Wage Legislation." *Journal of Labor Economics*, https://www.journals.uchicago.edu/doi/10.1086/733493.

Margo, Robert A. 1986. "Race, Educational Attainment, and the 1940 Census." *Journal of Economic History* 46(1): 189–198.

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

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

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

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

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

Olivetti, Claudia, and M. Daniele Paserman. 2015. "In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850–1940." *American Economic Review* 105(8): 2695–2724.

Rauscher, Emily. 2016. "Does Educational Equality Increase Mobility? Exploiting Nineteenth-Century U.S. Compulsory Schooling Laws." *American Journal of Sociology* 121(6): 1697–1761.

Ruggles, Steven, Sarah Flood, Matthew Sobek, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Renae Rodgers, and Megan Schouweiler. 2024a. 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. 2024b. IPUMS Ancestry Full Count Data: Version 4.0 [dataset]. Minneapolis, MN: IPUMS. https://doi.org/10.18128/D014.V4.0

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

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

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

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

Solon, Gary. 2004. "A Model of Intergenerational Mobility Variation Over Time." In *Generational Income Mobility in North America and Europe*, edited by Miles Corak, 38–47. Cambridge, UK: Cambridge University Press.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Link to Online Appendix

Online Appendix for

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

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

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

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

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

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

Step 1: Generating Sets of Candidate Links

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

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

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

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

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

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

Step 3: Training an Algorithm to Mimic Human Linking Decisions

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

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

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

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

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

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

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

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

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

Appendix B. Details on Construction of Weights

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

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

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

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

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

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

County Demographic and Socio-economic Characteristics

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

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

We construct modern-day analogs of the above county characteristics using county-level tables from the 2000 Decennial Census published by the Census Bureau and obtained via NHGIS. One minor difference is that the share of people in specific industries (manufacturing, business and professional services, public administration) or specific occupations (protective services, healthcare) are measured using employed individuals rather than labor force participants. Because these measures cannot be computed using county-level data published by the Census Bureau, income inequality and racial residential segregation come from data made available by Opportunity Insights (https://opportunityinsights.org/data/). We also obtained the share of religious adherents in 2000 from Opportunity Insights. Appendix Table A7 shows the means of the county-level characteristics described above, for the subset of counties used in our analysis.

State Resources for Public Education

State-level data on the number of K-12 teachers, average K-12 teacher salaries and instructors at public colleges and universities for the historical period are based on data we digitized from the U.S. Office of Education's 1928–1930 *Biennial Survey of Education* (U.S. Office of Education, 1931). K-12 teachers for the school year 1929–1930 are measured using the "total number of teaching positions" (Chapter II, Table 11, column 20). Average K-12 teacher salaries for the school year 1929–1930 are measured using the "average annual salaries of teachers, supervisors, and principals" (Chapter II, Table 12, column 3). For instructors at public colleges and universities, we digitized school or state-level data on the number of professors and instructors by gender for school year 1929–1930 for all 246 publicly-controlled universities, colleges, and professional schools (Chapter IV, Table 7, columns 3–4), 134 publicly-controlled teachers colleges (Chapter V, Table 8, columns 3–4), 66 state normal schools (Chapter V, Table 14, columns 3–4), 26 city normal schools (Chapter V, Table 19, columns 6–7), and 47 county normal schools (Chapter V, Table 21, columns 3–4).

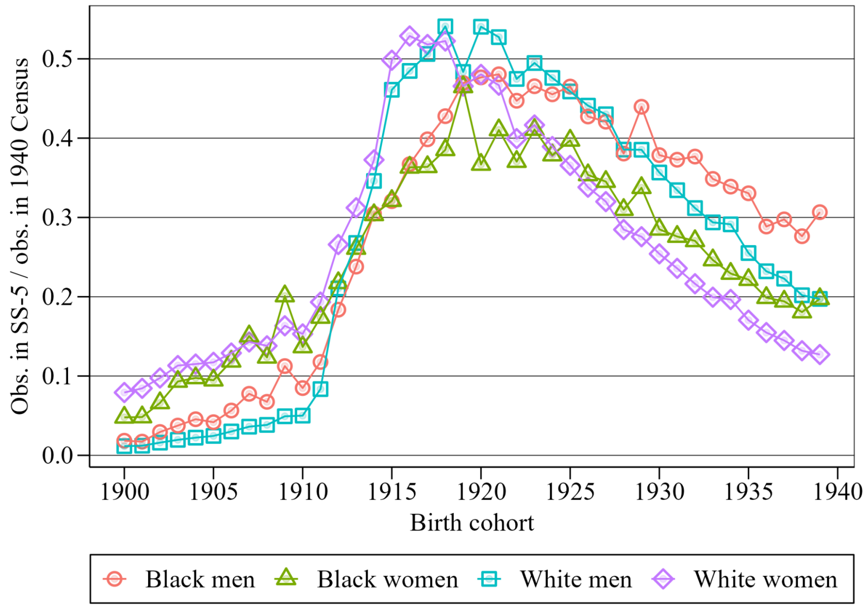
For the modern period, state-level data on K-12 teachers comes from the National Center for

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

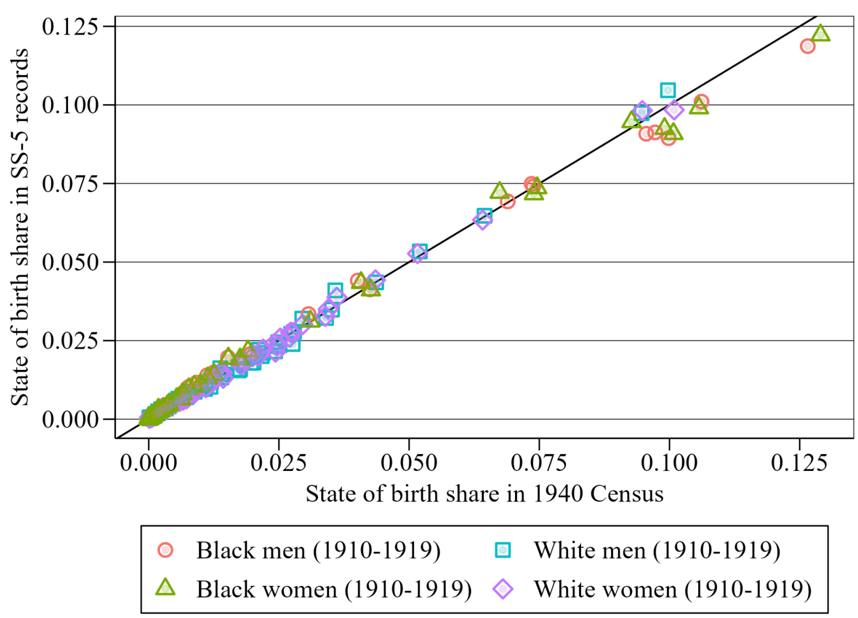
State-level K-12 teachers and public college instructors are normalized by the corresponding number of children aged 5–18 as measured in the 1930 Decennial Census for the historical period or in the 2000 Decennial Census for the modern period. Average K-12 teacher salaries are converted to 1999 dollars using the Bureau of Labor Statistics' historical "Consumer Price Index for Urban Consumers" series. Appendix Table A8 displays the resulting statistics for each state.

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

A. Cohort Coverage

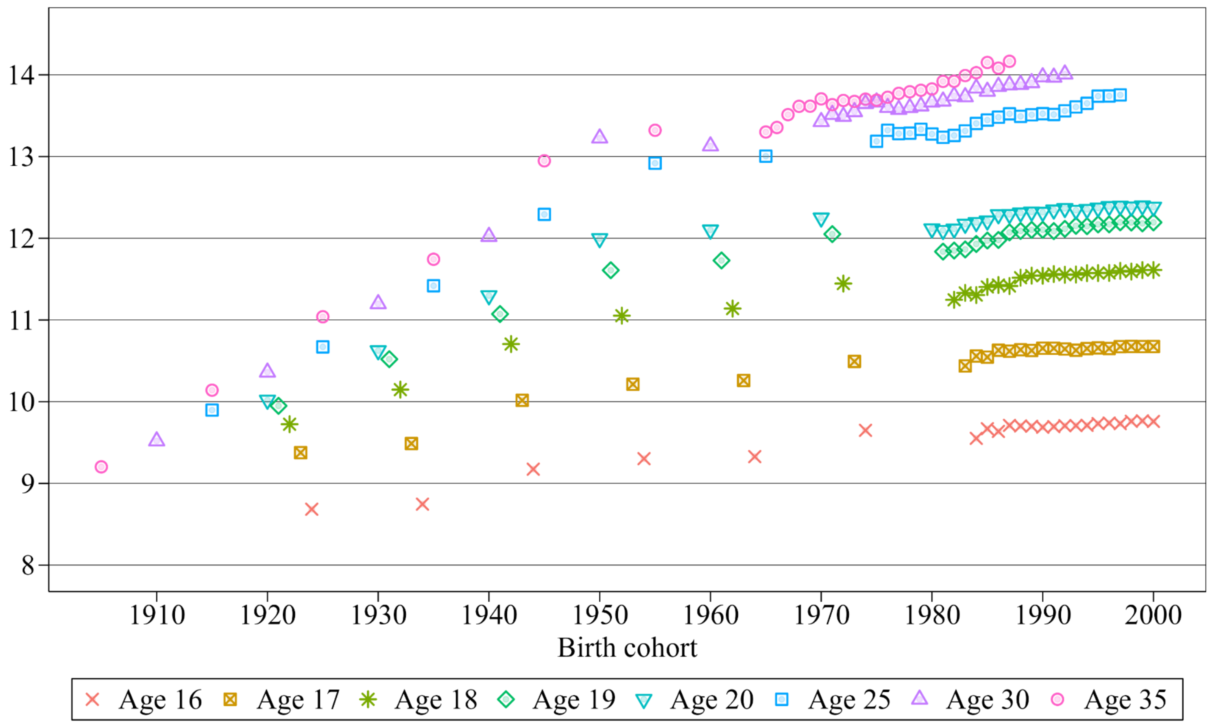


B. State of Birth Coverage



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

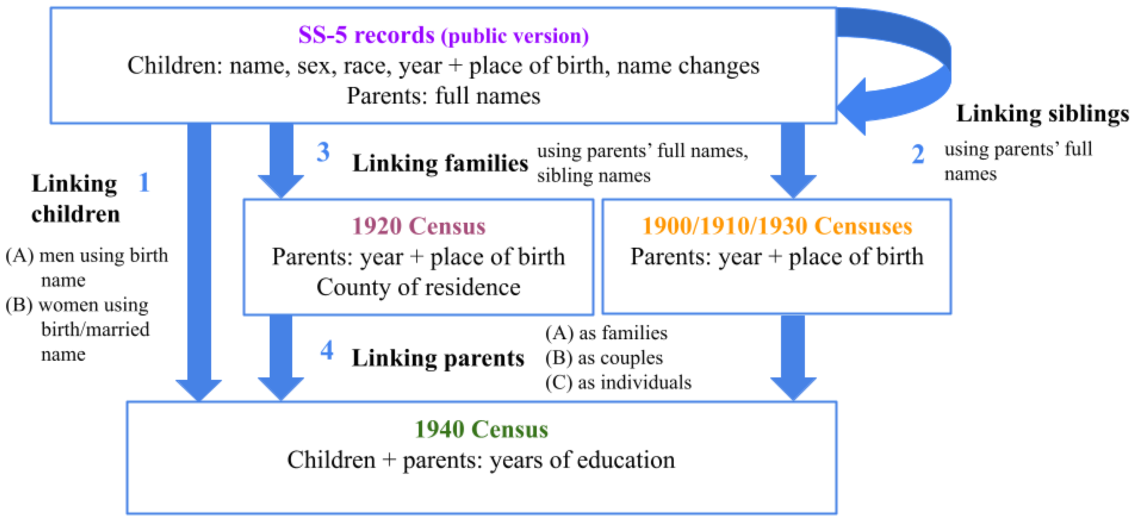
Figure A2. Average Years of Education by Age and Cohort



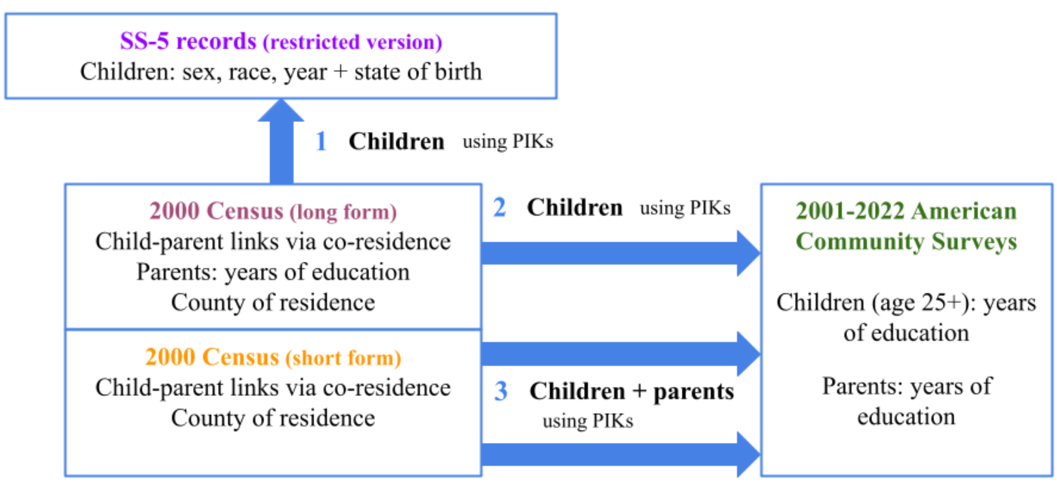
Notes: This graph plots average years of education at specified ages by cohort using data from the 1940–2000 decennial censuses and the 2001–2022 *ACS* (Ruggles et al., 2024a). Sample is restricted to individuals born in the U.S. Observations are weighted by census sampling weights.

Figure A3. Overview of Construction of Linked Samples

A. 1920-1940 Census Linked Sample

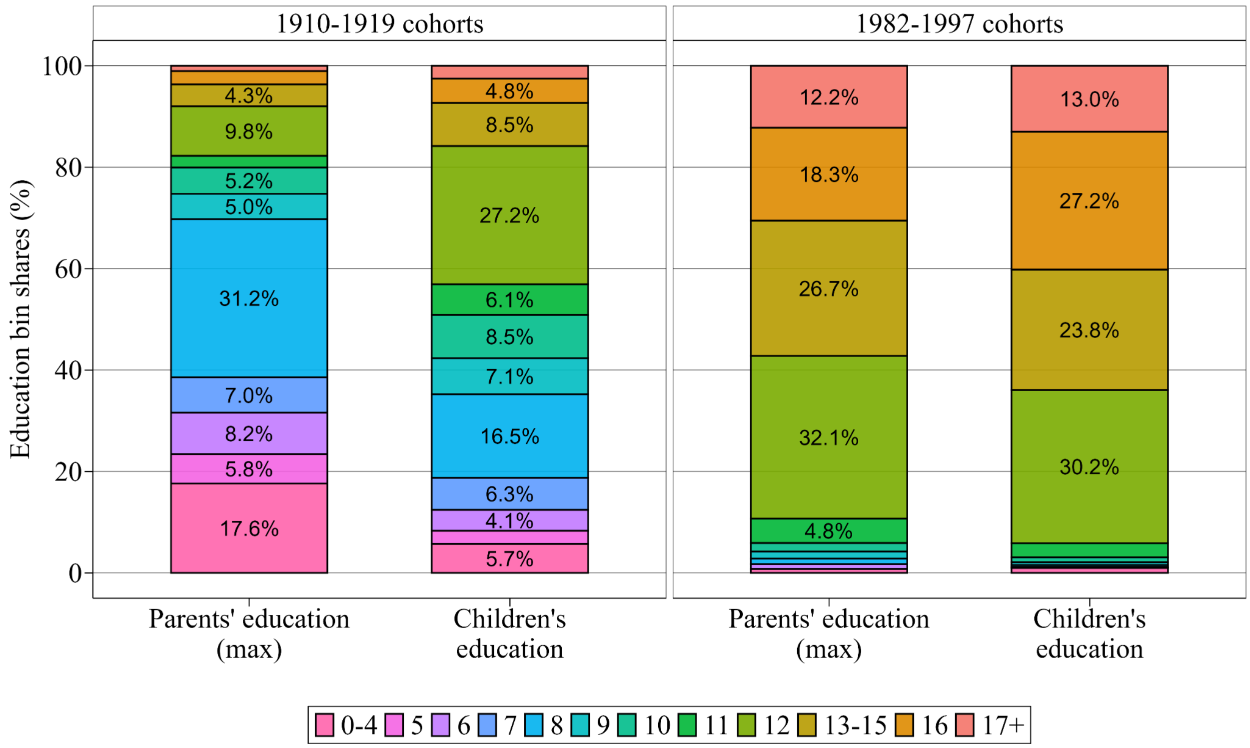


B. 2000 Census-ACS Linked Sample



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

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



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

Figure A5. National Education Transition Matrix by Cohort

A. 1910–1919 Cohorts

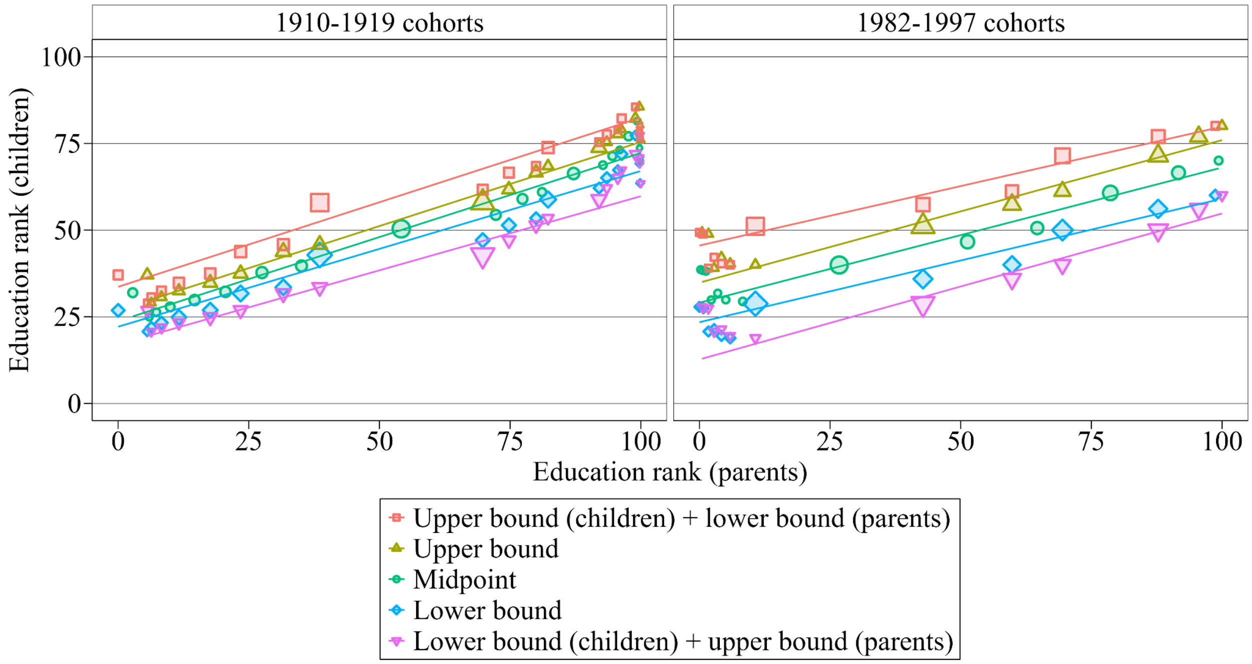
Years of education \ Max of parents' years of education	0-4	5-6	7-8	9	10	11	12	13-15	16	17+
17+	0.4%	0.4%	0.7%	1.0%	1.4%	1.6%	2.6%	4.0%	7.1%	15.2%
16	1.1%	1.2%	2.4%	3.7%	5.0%	5.6%	9.7%	13.7%	24.2%	23.0%
13-15	2.1%	2.7%	5.7%	8.7%	11.0%	12.9%	16.7%	24.7%	26.9%	25.3%
12	12.1%	16.5%	29.6%	35.4%	39.5%	40.3%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

B. 1982–1997 Cohorts

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

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

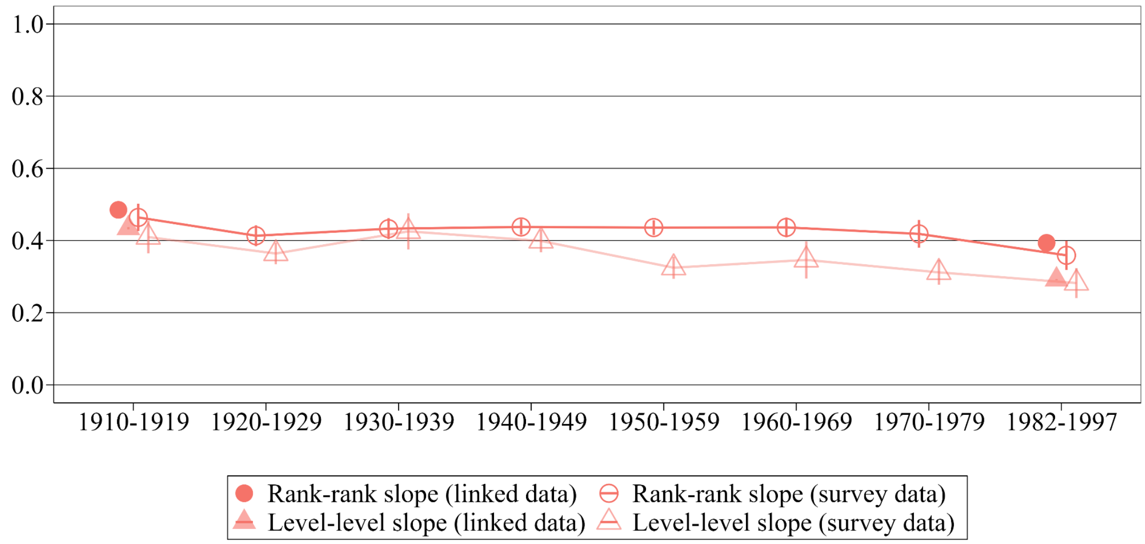
Figure A6. The Relationship Between Child and Parent Education Ranks, by Cohort: Alternative Rank Assignment Methods



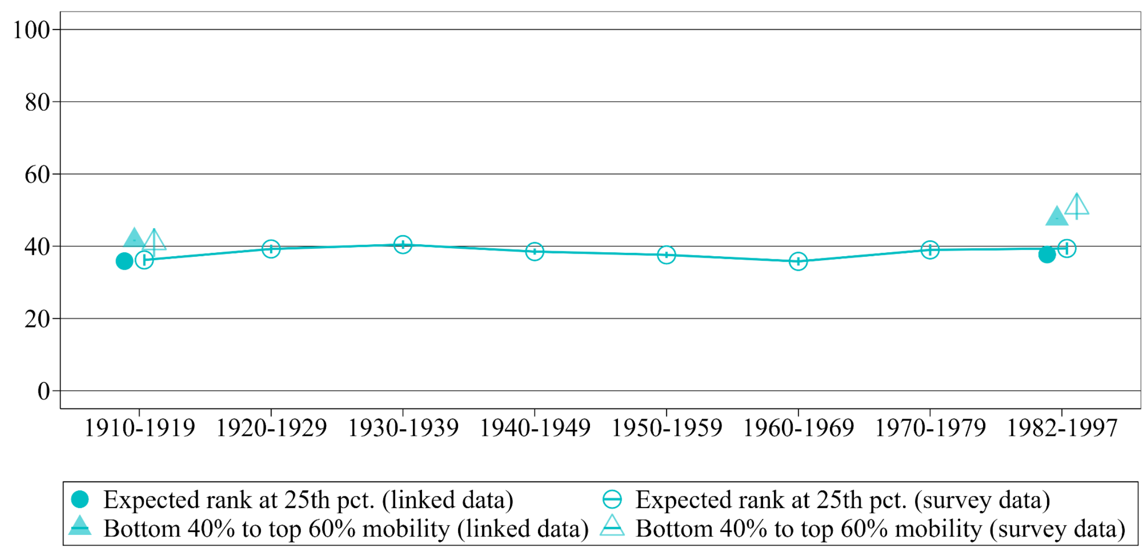
Notes: This graph plots the average education rank of children by parental education rank for different rank assignment methods, separately by cohort. Parental education is measured using the education of the most educated parent. Fitted lines are based on weighted linear regressions (see Appendix Table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

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

A. Relative Educational Mobility



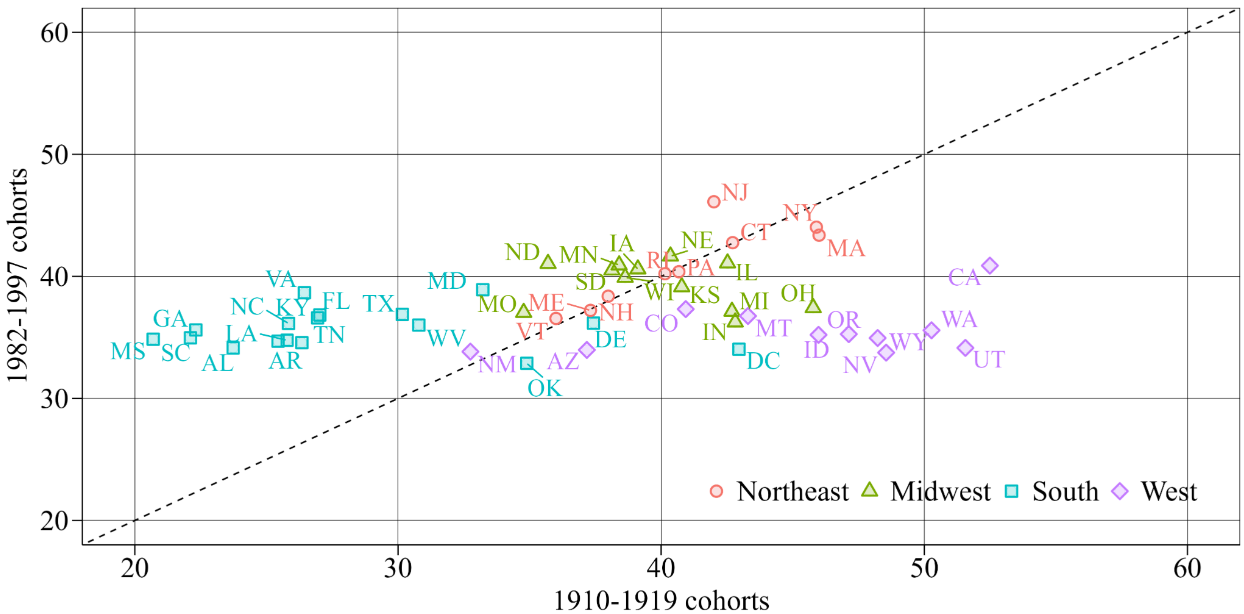
B. Upward Educational Mobility



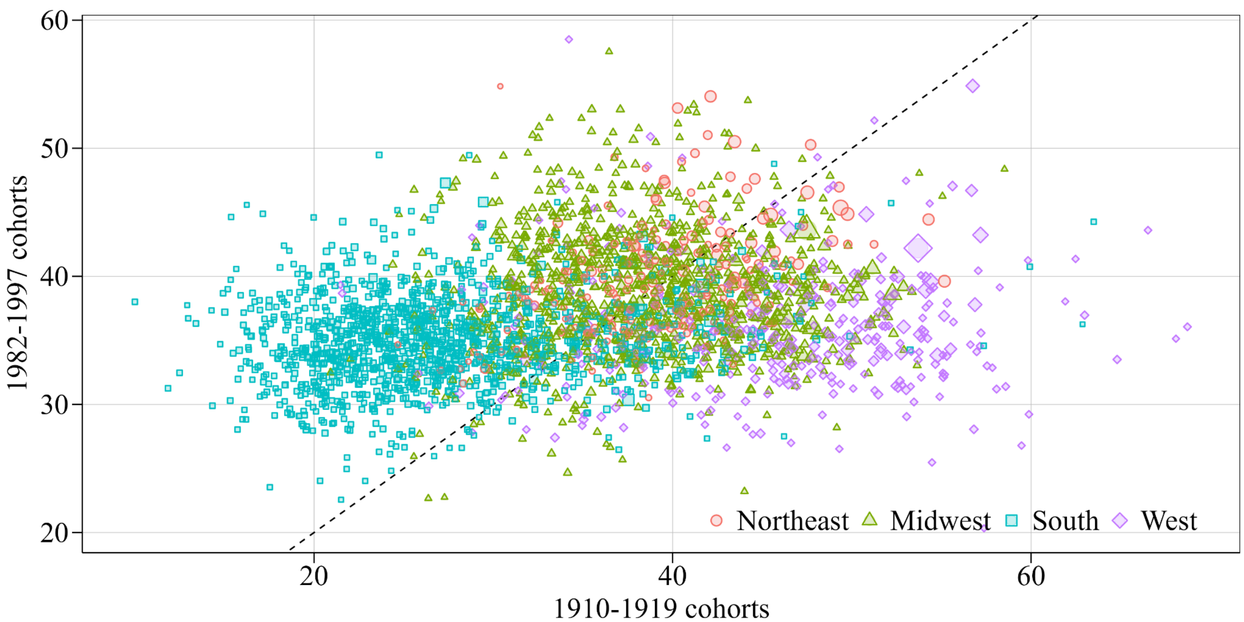
Notes: This graph plots estimates of relative and upward educational mobility by cohort and sample. Linked data for the 1910–1919 cohorts are reweighted to match the sex-by-cohort-by-race and education distribution in the 1940 population. Linked data for the 1982–1997 cohorts are reweighted to match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

Figure A8. Exp. Rank at 25th Percentile by Childhood State or County of Residence

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



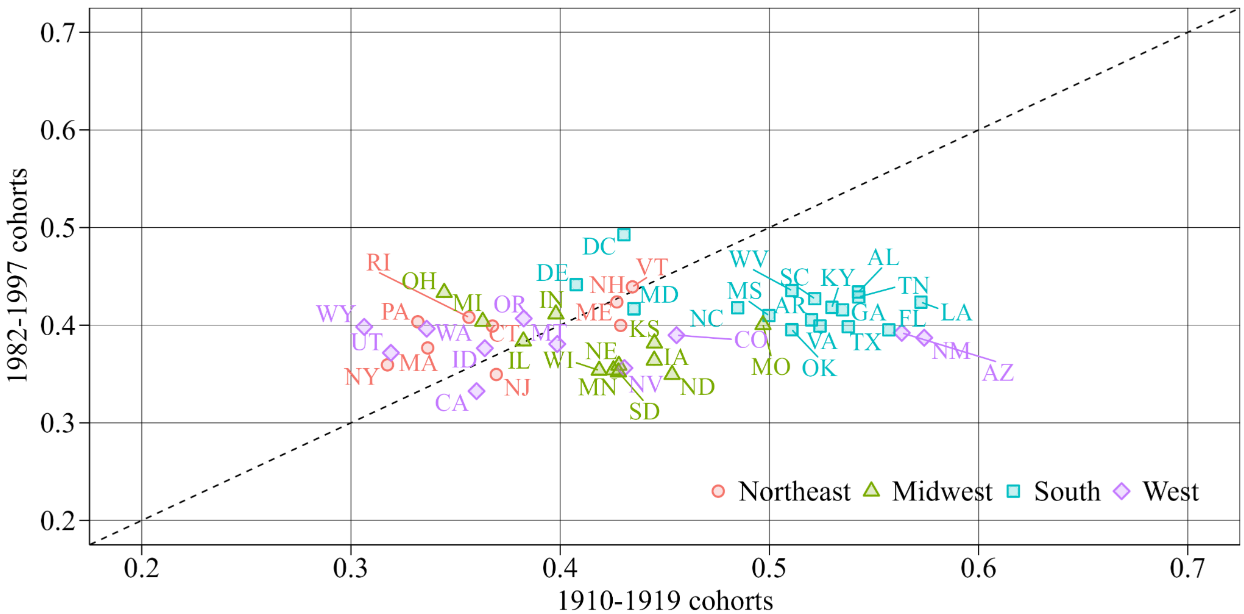
B. Expected Rank at 25th Percentile by Childhood County of Residence, 1910–19 vs. 1982–97 Cohorts



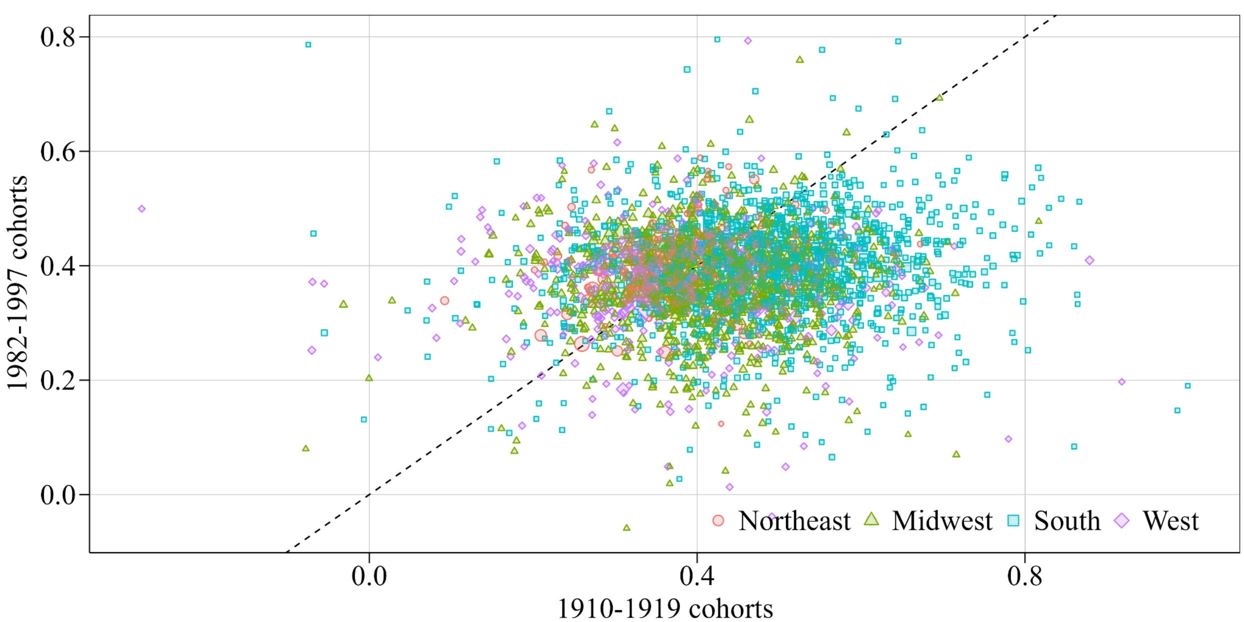
Notes: These graphs plot expected rank at the 25th percentile by 1920 childhood state or county of residence for the 1910–1919 cohorts against expected rank at the 25th percentile by 2000 childhood state or county of residence for the 1982–1997 cohorts.

Figure A9. Rank-Rank Slope by Childhood State or County of Residence

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



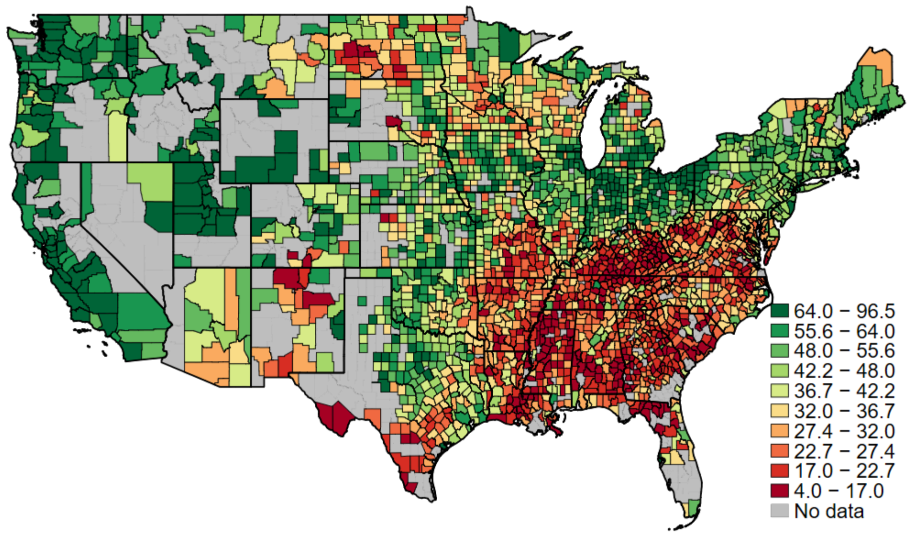
B. Rank-Rank Slope by Childhood County of Residence, 1910–19 vs. 1982–197 Cohorts



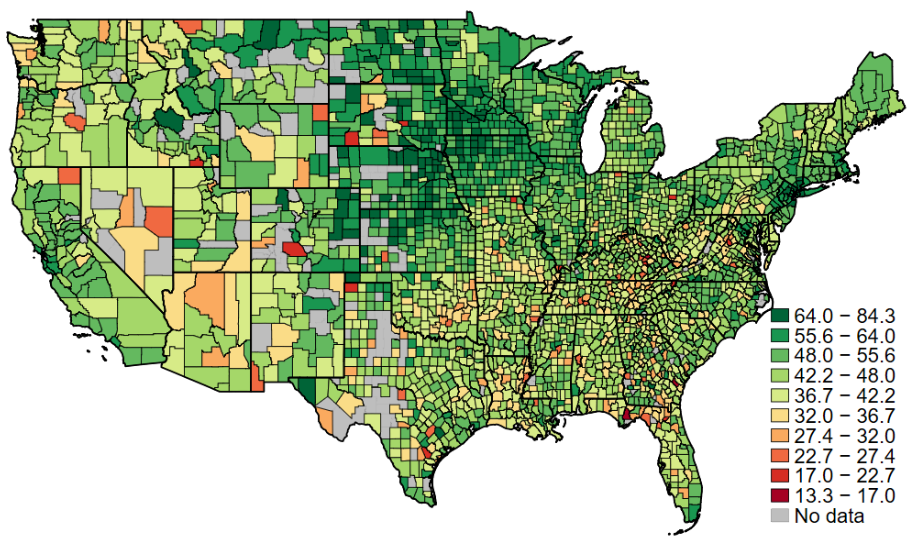
Notes: These graphs plot the rank-rank slope by 1920 childhood state or county of residence for the 1910–1919 cohorts against the rank-rank slope by 2000 childhood state or county of residence for the 1982–1997 cohorts.

Figure A10. Bottom 40 to Top 60 Percent Mobility by Childhood County of Residence over the 20th Century

A. 1910–1919 Cohorts



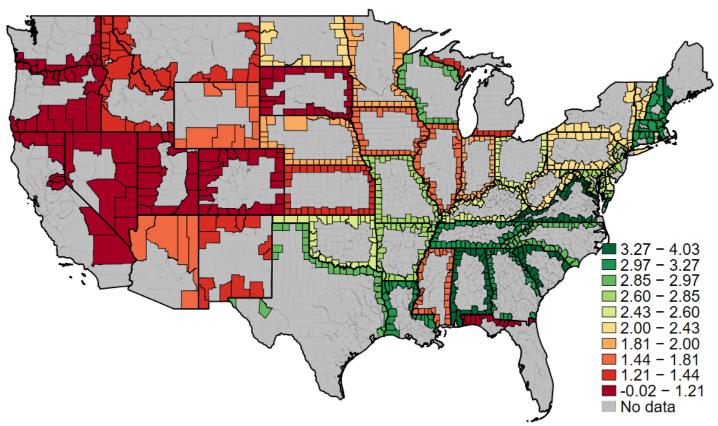
B. 1982–1997 Cohorts



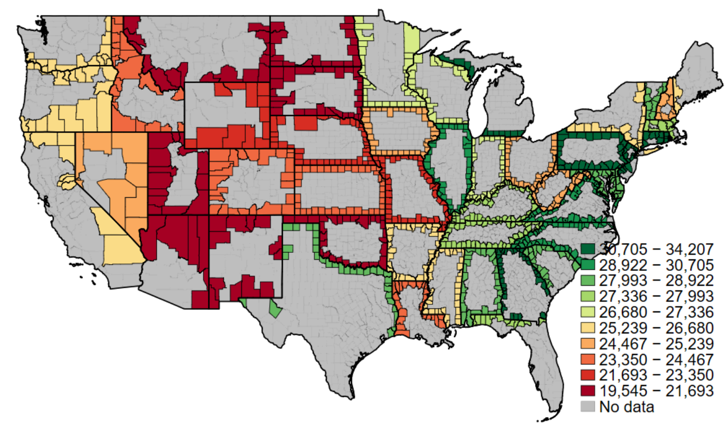
Notes: Figures plot bottom 40 to top 60 percent mobility by children's county of residence in 1920 for the 1910–1919 cohort and by county of residence in 2000 for the 1982–1997 cohort for counties with stable boundaries between 1920 and 2000 (see text for definition) and at least 20 parent-child observations.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

Figure A11. Sharp Discontinuities in Changing State Public Education Resources Across State Borders

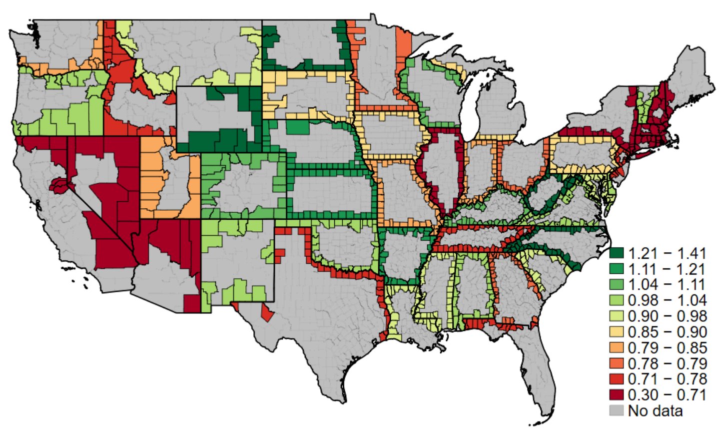
A. Change in State K-12 Teachers Per Child, 1930–2000



B. Change in State Average K-12 Teacher Salaries, 1930–2000



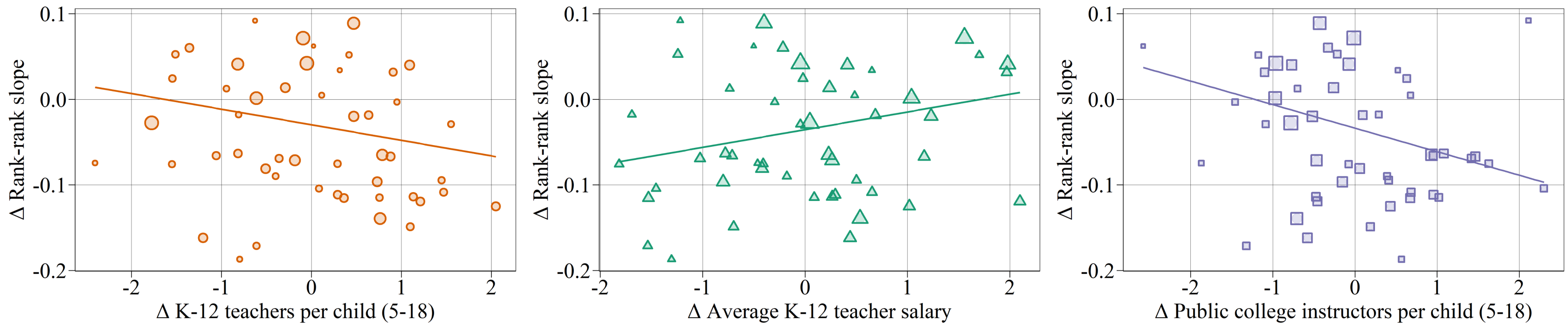
C. Change in State Public College Instructors Per Child, 1930–2000



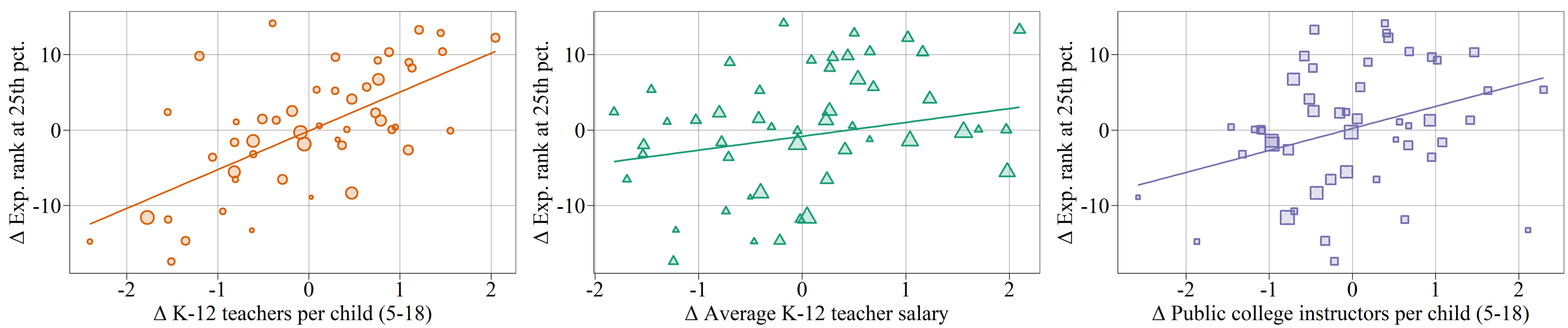
Notes: These graphs plot changes in state public education resources between 1930 and 2000 for the subset of the counties in our border regression samples. Counties are shaded according to which decile bin the associated state measure falls into.

Figure A12. The Relationship of Changes in Educational Mobility and Relative Changes in State Resources for Public Education over the 20th Century

A. Changes in Rank-Rank Slope versus Changes in K-12 Teachers, K-12 Teacher Salaries, and Public College Instructors per Child

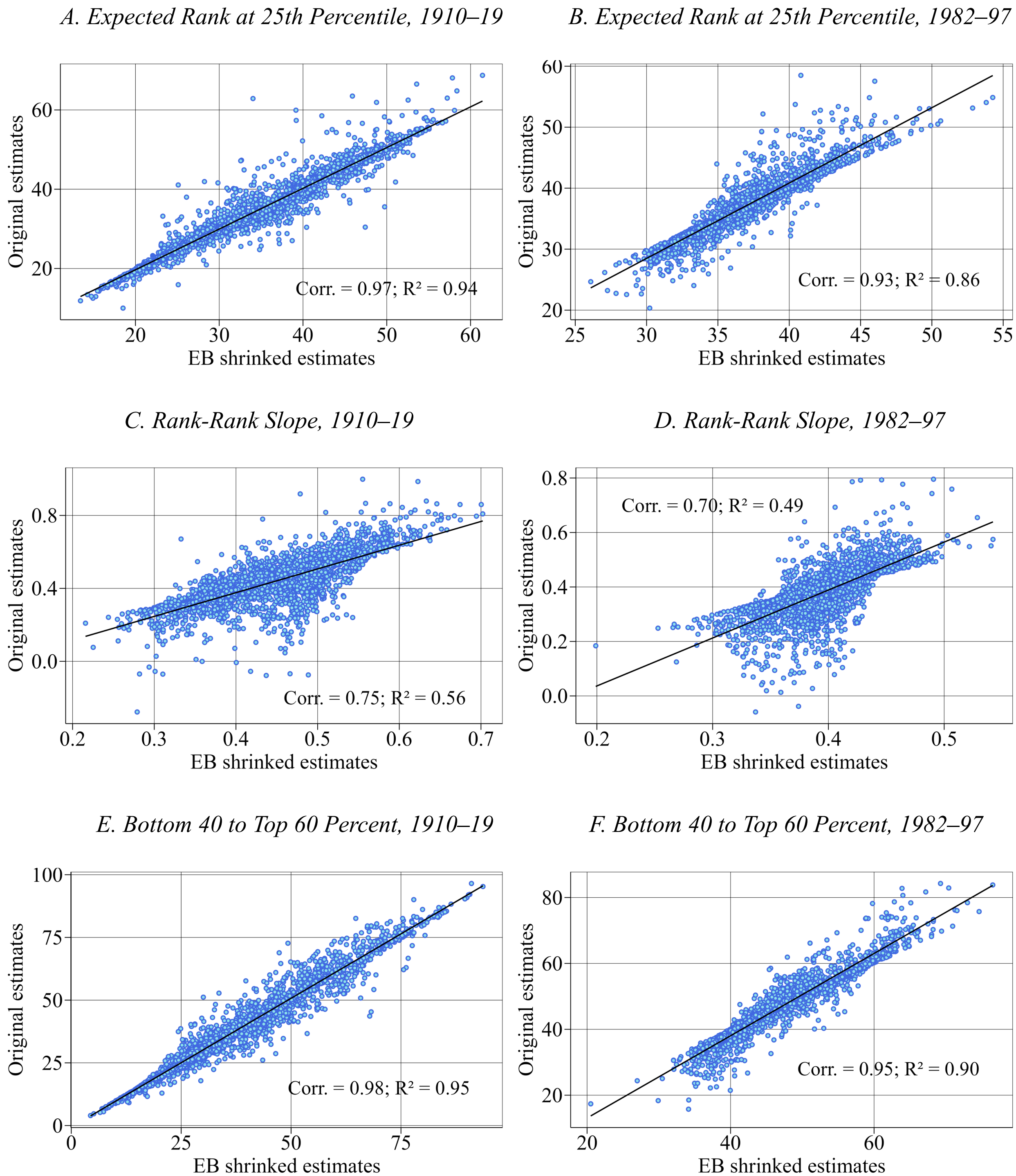


B. Changes in Upward Mobility versus Changes in K-12 Teachers, K-12 Teacher Salaries, and Public College Instructors per Child



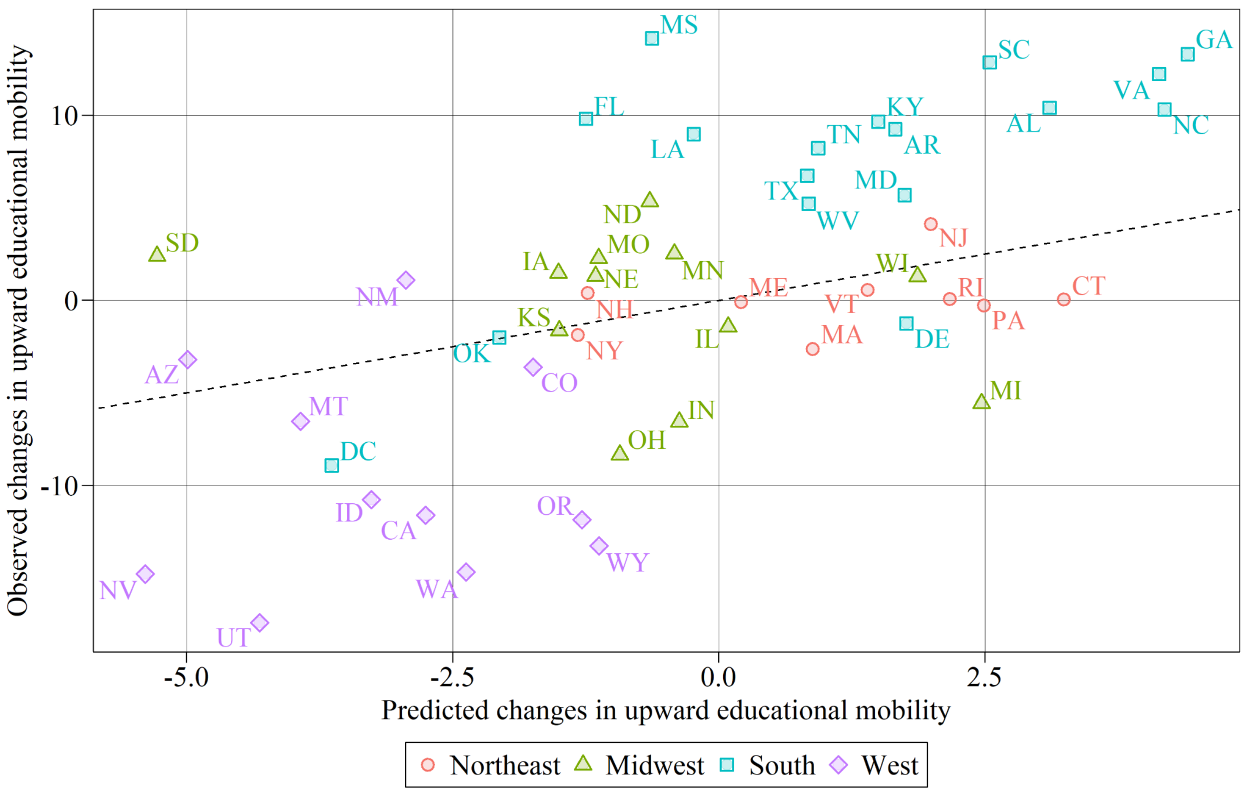
Notes: These graphs plot changes in relative or upward educational mobility by childhood state of residence between the 1910–1919 and 1982–1997 cohorts against changes in state resources for public education between 1930 and 2000. Statistics on K-12 education and public colleges and universities for 1928–1930 come from the U.S. Office of Education (1931). Analogous statistics on K-12 education for 1999–2000 and public post-secondary institutions for 2000–2001 are constructed using the Common Core of Data (CCD) and Integrated Post-secondary Education Data System (IPEDS) databases from the National Center for Education Statistics (NCES), or are directly taken from NCES tables (U.S. Department of Education, 2022). Data on the number of children aged 5–18 comes from the 1930 and 2000 Censuses. See Appendix C for details. Changes in state public education resources are standardized to have a mean of 0 and standard deviation of 1. Fitted lines are based on linear regressions where observations are weighted by the average number of children across our linked samples.

Figure A13. Original vs. Empirical Bayes Shrinked County Estimates



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

Figure A14. Observed vs. Predicted State Changes in Upward Educational Mobility over the 20th Century



Notes: This graph plots observed state changes in upward educational mobility over the 20th century on the *y*-axis against predicted state changes in upward educational mobility on the *x*-axis, which are obtained by adding the average change in upward educational mobility across states over time and the sum of observed (standardized) changes in state resources for public education multiplied by the corresponding border-design estimates from column 3 in Table 2. The dashed line indicates the 45-degree line.

Table A1. Overview of Linking Process

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

Table A1 (continued). Overview of Linking Process

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Table A7. County Characteristics by Year

	1920		2000	
	Full sample	Border sample	Full sample	Border sample
	(1)	(2)	(3)	(4)
A. Demographic Structure				
Population	35,662	42,765	92,664	87,789
Population density (people per sq mi)	137	241	223	311
Share Black	0.11	0.11	0.09	0.08
Share Asian	0.001	0.001	0.008	0.007
Share foreign-born	0.07	0.07	0.03	0.03
Share born out-of-state	0.21	0.25	0.26	0.32
Income inequality (Gini)	0.47	0.47	0.38	0.37
Racial residential segregation (Theil)	0.19	0.19	0.08	0.08
B. Economic Development				
Share urban households	0.19	0.20	0.40	0.39
Share farm households	0.52	0.50	0.05	0.05
Labor force participation rate	0.55	0.55	0.61	0.61
Manufacturing employment share	0.11	0.13	0.16	0.17
Business & professional services emp. share	0.04	0.04	0.26	0.25
Income per capita (1999 dollars)	1,981	2,067	17,487	17,410
Median house value (1999 dollars)	22,200	23,015	84,088	83,506
C. Public Services				
Public admin. workers per 1,000 people	5.8	5.8	22.5	21.6
Protective workers per 1,000 people	1.1	1.2	8.4	8.3
Healthcare workers per 1,000 people	3.4	3.4	29.7	29.8
D. Family Structure & Social Capital				
Share adults married	0.62	0.62	0.60	0.60
Share adults divorced	0.01	0.01	0.09	0.10
Share children in single-parent households	0.08	0.08	0.25	0.25
Average household size	4.4	4.4	2.5	2.5
Share religious adherents	0.43	0.42	0.53	0.54
N	2,881	1,083	2,881	1,083

Notes: This table reports the means of county characteristics in the full sample of counties and the subset of border counties used in our analysis. Some county characteristics are measured in 1930 or 1940 instead of 1920. See Appendix C for sources and variable construction details.

Table A8. State Public Education Statistics by Year

	K-12 teachers per 100 children (ages 5–18)		Average K-12 teacher salary (1999 dollars)		Public college instructors per 100 children (ages 5–18)	
	1930	2000	1930	2000	1930	2000
	(1)	(2)	(3)	(4)	(5)	(6)
United States	2.5	4.6	1,420	41,807	0.11	0.96
Alabama	1.9	5.4	7,892	36,689	0.08	1.12
Arizona	2.5	4.1	16,312	36,902	0.22	0.81
Arkansas	2.1	4.9	6,706	33,386	0.07	1.19
California	3.0	3.5	21,155	47,680	0.25	0.96
Colorado	3.5	4.7	14,478	38,163	0.23	1.34
Connecticut	2.3	5.3	18,056	51,780	0.06	0.70
Delaware	2.3	4.7	15,644	44,435	0.13	1.14
District of Columbia	2.8	5.0	22,609	47,076	0.07	0.38
Florida	2.6	3.7	8,729	36,722	0.07	0.82
Georgia	2.0	5.2	6,816	41,023	0.07	0.86
Idaho	3.3	4.6	11,957	35,547	0.15	0.88
Illinois	2.5	4.1	16,242	46,486	0.10	0.77
Indiana	2.6	4.5	14,608	41,850	0.10	0.93
Iowa	3.7	5.4	10,901	35,678	0.21	1.11
Kansas	3.7	5.2	11,549	34,981	0.21	1.35
Kentucky	1.9	4.3	8,928	36,380	0.10	1.21
Louisiana	1.9	5.0	9,377	33,109	0.07	1.00
Maine	3.0	6.5	9,387	35,561	0.14	0.78
Maryland	2.0	4.7	15,126	44,048	0.19	1.10
Massachusetts	2.4	5.6	18,683	46,580	0.05	0.76
Michigan	2.6	4.1	15,285	49,044	0.20	1.07
Minnesota	3.1	5.1	12,466	39,802	0.19	0.97
Mississippi	2.3	4.1	6,178	31,857	0.06	1.04
Missouri	2.6	5.4	12,306	35,656	0.12	0.97
Montana	4.0	5.4	12,107	32,121	0.17	1.12
Nebraska	3.8	5.7	10,732	33,237	0.17	1.39
Nevada	3.9	3.9	14,777	39,390	0.31	0.78
New Hampshire	2.5	5.6	12,495	37,734	0.21	0.77
New Jersey	2.4	5.0	21,055	52,015	0.03	0.80

Table A8 (continued). Public Education Statistics by State and Year

	K-12 teachers per 100 children (ages 5–18)		Average K-12 teacher salary (1999 dollars)		Public college instructors per 100 children (ages 5–18)	
	1930	2000	1930	2000	1930	2000
	(1)	(2)	(3)	(4)	(5)	(6)
New Mexico	2.5	3.9	11,090	32,554	0.18	1.20
New York	2.5	4.6	24,841	51,020	0.08	0.75
North Carolina	2.1	5.1	8,699	39,404	0.06	1.28
North Dakota	3.8	6.1	8,968	29,863	0.26	1.67
Ohio	2.4	5.0	16,591	41,436	0.11	0.90
Oklahoma	2.6	5.1	10,682	31,298	0.18	1.22
Oregon	2.7	3.4	16,063	42,336	0.34	1.37
Pennsylvania	2.1	4.2	16,142	48,321	0.05	0.93
Rhode Island	2.1	4.7	14,319	47,041	0.04	0.67
South Carolina	2.1	5.6	7,852	36,081	0.08	1.06
South Dakota	4.3	5.0	9,526	29,071	0.20	1.07
Tennessee	2.2	5.4	8,988	36,328	0.06	0.84
Texas	2.0	4.9	9,207	37,567	0.11	0.84
Utah	2.7	3.5	13,253	34,946	0.16	1.00
Vermont	3.0	5.3	9,596	37,758	0.22	1.26
Virginia	2.1	6.2	8,579	38,744	0.13	1.12
Washington	2.9	3.8	15,505	41,043	0.23	1.05
West Virginia	2.8	5.3	10,194	35,009	0.11	1.37
Wisconsin	2.5	5.4	13,940	41,153	0.17	1.27
Wyoming	4.8	6.4	12,346	34,127	0.19	1.56

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

Table A9. Impact of Public Education on Bottom 40 to Top 60 Percent Mobility

	Dependent Variable: Change in County Bottom 40 to Top 60 Percent Mobility			
	(1)	(2)	(3)	(4)
Change in state K-12 teachers per child (ages 5–18)	6.94*** (0.86)	6.45*** (1.22)	1.14 (1.05)	1.04 (1.04)
Change in state average K-12 teacher salary	1.48 (0.96)	1.64 (1.08)	2.79*** (1.04)	2.86*** (1.02)
Change in state public college instructors per child (ages 5–18)	6.16*** (0.80)	6.36*** (0.93)	2.60** (1.04)	2.94*** (1.06)
Border sample		✓	✓	✓
County pair fixed effects			✓	✓
Bayesian shrinked estimates				✓
R^2	0.194	0.207	0.816	0.845
Partial R^2			0.058	0.084
N	2,512	1,862	1,862	1,862

Notes: Column 1 shows the estimates from a regression of county-level bottom 40 to top 60 percent mobility on state public education resources, which are standardized to have a mean of 0 and a standard deviation of 1. Column 2 limits the sample to border counties (units of observation are county-by-border pairs), where counties may appear in more than one pair. Column 3 includes county-pair fixed effects as in equation 3. Column 4 replaces the county mobility measures with their Empirical Bayes analogs to reduce the influence of sampling noise or disclosure-avoidance noise infusion. The partial R^2's capture the share of the residual variation in county educational mobility explained by the state covariates after controlling for county pair fixed effects. Counties are weighted by the inverse of the variance of the county mobility estimates. Robust standard errors in parentheses are clustered at the state-border level in columns 3 and 4. Stars indicate the level of statistical significance of the estimates: ***$p < 0.01$, **$p < 0.05$, *$p < 0.10$.