Mobility for All: Representative Intergenerational Mobility Estimates over the 20th Century

Elisa Jácome, Ilyana Kuziemko, and Suresh Naidu*

Preliminary and incomplete. Comments welcome.

Abstract

We present what is to the best of our knowledge the first long-run estimates of intergenerational relative mobility for samples that are representative of the full U.S.-born population. We develop a simple mobility measure that allows easy inclusion of non-whites and women for the 1910s to 1970s birth cohorts. We show a robust decline in both the intergenerational-elasticity and rank-rank persistence measures between the 1910s and 1940s birth cohorts. Both measures tend to drift up afterwards, so we find that persistence measures mirror the u-shaped trends in inequality over this period. Decomposing the IGE into within- and between-group components, we show that absolute convergence of incomes by race explain a large share of the decline in intergenerational mobility.

*We thank Ahna Pearson for excellent research assistance. Leah Boustan, James Feigenbaum, and Nathan Hendren have provided invaluable feedback at various stages of this project. We thank seminar participants at Berkeley, NBER, NYU Wagner, Princeton, USC, and Wharton.

1 Introduction

Intergenerational relative mobility—how tied an individual's place in the income distribution is to his parents' place in the income distribution while he was growing up—has long been an object of interest, especially in the United States. A high level of mobility is viewed as an important part of Americans' identity as a nation: that it is a "land of opportunity" even for those who grew up poor. In political philosophy, intergenerational relative mobility is widely viewed as a proxy for equality of opportunity (Roemer and Trannoy, 2015) and the overall fairness of a society.

The rise in inequality since the 1970s and 1980s has further increased interest in intergenerational mobility. The rise in inequality may be less troubling if it is accompanied by rising mobility as well. Cross-country variation, however, suggests that, at least in modern data, high levels of inequality correlate with greater income persistence between children and parents (Corak, 2013). Direct evidence on the correlation in inequality and mobility across *time* within the US has been more limited. As Song *et al.* (2020) write: "evidence of long-term trends in intergenerational mobility is largely absent" (p. 251).

The main contribution of this paper is simple: to our knowledge, we present the first long-run estimates of historical intergenerational relative mobility for a *representative* sample of U.S.-born individuals. In particular, we show mobility estimates for children born in the 1910s through the 1970s.[1] As Table 1 shows, a handful of papers have made important contributions to our understanding of long-run trends in intergenerational relative mobility. However, for data reasons, they include only subsets (and typically *advantaged* subsets) of the population. Song *et al.* (2020) shows mobility of occupational prestige from 1830 to 1980, but only for white men. Using a very clever synthetic-panel strategy based on the status information conveyed by first names, Olivetti and Paserman (2015) can compare occupational mobility between fathers and sons to that of fathers and sons-in-law, but only for white men and married white women. Collins and Wanamaker (2017) and Ward (2020) include Black Americans, but only men. Papers on intergenerational relative mobility that include women and non-whites present results only for more modern periods (see, e.g., Chetty *et al.*, 2014b, Solon, 1992, Chetty *et al.*, 2018, or Mazumder, 2018) or short snapshots of time (see, e.g., Card *et al.*, 2018). For example, in Chetty *et al.*, 2018 a key assumption underlying the calculation of steady-state racial inequality gaps is that race-specific intergenerational elasticities (IGE) are constant over time.

The goal of including women and non-whites motivates our methodology. Instead

[1]Note that we do not examine intergenerational *absolute* mobility, which captures the probability that a child's income as an adult surpasses her parents' income (in real dollars) while she was a child. For recent work on intergenerational absolute mobility, see Chetty *et al.* (2017) and Berman (2018).

of relating occupational status of one generation to the next (which complicates looking at women, as few worked in the historical period), we introduce a family-income-to-income-score mobility concept. Essentially, we relate the self-reported family income (adjusted for inflation) of the adult child to an *income score* meant to predict her family income while she was growing up, based on characteristics such as her race, her father's occupation, her region, and, when available, her father's education. Family income is readily reported in many datasets, and is a question that male as well as female respondents can answer. Moreover, it will naturally pick up income gaps by race.

We locate (to the best of our knowledge) all surveys that ask individuals their current family income and their father's occupation while they were growing up, ranging from well-known surveys like the General Social Survey to more obscure ones like "Americans View Their Mental Health" and "The National Survey of Black Americans." All of these surveys also ask race and the region (at the very least, South versus elsewhere) where the respondent was born or grew up. Our baseline estimates use the 1940 Census to predict family income for fathers by $Race \times Occupation \times South$ cells (though we show robustness to many variations in predicted family income, including using father education whenever available). We then relate self-reported family income of the adult child to her predicted income while growing up, estimating both intergenerational elasticity and rank-rank mobility relationships.

Our main finding is that both IGE and rank-rank mobility measures fell (meaning that mobility rose) between the 1910s and 1940s birth cohorts. Between the 1940s and 1970s birth cohorts, these measures trended upward again, especially between the 1940s and 1960s cohorts. Overall, we find a u-shape with respect to these persistence measures, with the through at the 1940s birth cohorts. Interestingly, the u-shape tracks well measures of inequality in the family income of the adult children in our sample. This result is the within-country-over-time variant of the "Great Gatsby Curve" relating modern-day mobility estimates to inequality using cross-country variation. Of course, just as with the cross-country correlation, we make no claims to causality between the adult children's inequality measures and their mobility.

As noted, our paper is motivated by the absence of historical mobility measures that are representative of the full U.S.-born population. In the second part of the paper, we focus on subgroups (mostly the four subgroups defined by Black/white race and sex) and in particular how movements of these subgroups contributed to (or slowed) the increase in mobility from the 1910s to 1940s cohorts. Like inequality measures and unlike means, the full-population IGE (or rank-rank) slope is not a weighted average of subgroup IGE slopes. In particular, the between-subgroup differences in parental income play a major role. We make this point more precise in the paper, but the basic logic is shown in Figure 1. In this case, the full-population IGE is greater than the IGE measure from *either* subgroup, because group B comes from such an extremely disadvantaged part of the parental-income distribution and remains disadvantaged in adulthood.

In our subgroup analysis, we show several key changes in the mapping of parental income to own adult income among our key subgroups and how they relate to the overall change in full-population mobility. From the 1910s–1920s to the 1940s–1950s birth cohorts, there is substantial catch-up for Black Americans, with their entire regression line shifting upward by roughly thirty log points (or seven rank percentiles). Whites also enjoy income growth in real terms, and their IGE and rank-rank slopes become flatter (meaning that, *within the white population*, parental income matters less in predicting own adult family income). Our decomposition shows that the Black-white catch-up in levels of income accounts for roughly half of the rise in mobility and the flattening of the white IGE slope accounts for the remainder. This result highlights the independent contribution of reducing inequality between endogamous subgroups for reducing intergenerational persistence.

Of particular interest are Black women, both because there has been almost no mobility work that has included them and because, of the four main subgroups we examine in detail, they are the worst off in terms of adult family income over most of our period (and differently from more recent data e.g. Chetty *et al.* (2018)). We conclude our paper by showing that accounting for other race and gender groups besides white men, particularly Black women, increases the IGE consistently over the 20th century, and significantly so in the early 20th century.

The remainder of the paper is organized as follows. In the next section we describe the various datasets we use. In Section 3, we describe our methodology, in particular family-income-to-income-score mobility concept. In this section we describe in detail how we calculate income scores to approximate parental income. Section 4 presents our results for the full, representative population. Sections 5 presents a decomposition of the IGE and decomposes the rise in mobility into differential mobility by race and gender. Section 6 concludes.

2 Data

In this section, we briefly describe the datasets that we use in this paper and share summary statistics. Far greater detail can be found in Appendix B.

2.1 Datasets and sampling rules

We have located to the best of our knowledge all surveys that ask family income, father's occupation with sufficient detail, race, and region of the country where the respondent was born or grew up (at least to the level of South versus other regions). We end up using 14 different surveys, and details on all of them are provided in Appendix B. Most readers will be familiar with some (e.g., the General Social Survey or the American National Elections Survey), but others are not as well known (e.g., the National Black Election Study or Americans View their Mental Health).

In some cases, the data we use are in fact panel datasets that follow individuals and families over time (e.g., the National Longitudinal Study of Mature Women and Older Men) and have often been used to measure mobility for more modern periods. To remain consistent within our methodology, however, we do not use the *panel* components of these datasets. In the first wave, these panel datasets often ask the adult respondent questions about their own childhood, and it is this linkage that we use to estimate the respondents' family income in childhood.

Following the IGE literature, we restrict attention to native-born men and women in the 30–50 age range in order to ensure that we are measuring lifecycle earnings as closely as possible. While some recent papers have not limited themselves to ages close to forty, in all cases we limit ourselves to this age range. Papers that use occupation scores of the adult child may well have less worry about life-cycle bias as occupation may be more stable across a career, but as we are using self-reported family income, we take care not to stray too far from prime-age years.

In many cases, the data collection for these surveys was explicitly meant to be representative and provide survey weights to correct deviations due to sampling error. In those cases, we use the provided sampling weights. Of course, some of these surveys target one sex or one race (e.g., National Black Election Survey) so are clearly not representative of the full U.S.-born population. In the early cohorts, we also have a substantially lower share of women in our data than in the general population. For this reason, we will always re-weight the dataset so that each cohort has weighted shares for white women, white men, Black women and Black men of 0.44, 0.44, 0.06 and 0.06, respectively. In the appendix, we show that our main results barely change under other weighting schemes, including not weighting at all.

2.2 Summary statistics

The first panel of Table 2 show summary statistics of the fathers of the respondents in our main dataset, separately by decade of birth. We do not weight at all so that readers can get a sense of the raw data.

The decline of agriculture as a dominant occupation for fathers is readily apparent for children in the 1910–1950 birth cohorts, falling from over one-third to less than one-tenth. As noted, we do not have father's education in every survey, but the table shares summary statistics from those surveys that do include father's education. In our earliest birth cohorts, the fathers in our data are born in the last few decades of the nineteenth century and thus grew up before the high school movement, which is reflected in their low levels of secondary education. Less than twenty percent of the fathers of our 1910s and 1920s birth cohorts graduated from high school. College graduation was a rarity for these fathers and as late as the 1960s birth cohorts less than one in six of respondents have fathers who completed college.

Summary statistics for the adult children (i.e., the survey respondents) appear in the second panel of the table. The age of respondents is relatively similar and always close to forty, as we would expect from our 30–50 age restrictions. In contrast to past historical work which focused on whites (white men, in fact), our samples have coverage of Black individuals very close to their population shares. Past work that has applied linkage techniques to the Black population in the Census have also tended to result in samples somewhat smaller than the population shares (Collins and Wanamaker, 2017; Ward, 2020).

A number of trends among the children in our data merit comment. The rise of educational attainment from the 1910 to the 1950 birth cohorts is striking. High school attainment increases from one-half to 90 percent; and college graduation rates nearly triple, from ten to twenty-eight percent. The increase in education from one generation to the next is massive as well: for the 1910s to 1930s birth cohorts, the likelihood they graduate from high school is triple that of their fathers.

Another marked trend for the adult children in our data is the decline in veteran status (which the table reports only for men in surveys that asked about veteran status). While over seventy percent of men in our 1920 cohort report military service, by the 1950s cohort military service has become quite rare. Finally, another noticeable trend is union membership: while it holds steady in the mid- to high-twenties for our early cohorts, it begins a steady decline with the 1950s cohort.

Table 3 separates our data (unweighted, as in the previous table) by time period, race and sex and compares it to the relevant population in the Census. As before, we see that in all periods and separately for men and women, our data are very close to representative on race (roughly ten percent of the sample). In fact, one of the only variables on which there are small discrepancies between our raw survey data and the Census data is education in the earliest birth cohorts. For example, whites in our data have a high-school completion rate about ten percentage points higher than their Census counterparts (the differences are positive but slightly smaller for Black individuals). This difference is smaller for all groups in later birth cohorts.

Otherwise, our raw survey data is remarkably similar to the Census in terms of age, the share living or originating from the South (an especially important variable for Black respondents), and marital patterns. This table emphasizes the fact that, practically speaking, it is simply easier to gather representative data in a single cross-section (as our surveys do) than to maintain a representative sample over many years via panel data (whether the connections over time are created by Census linking or explicitly following the same person longitudinally as in the Panel Study of Income Dynamics [PSID] or National Longitudinal Surveys [NLS]).

3 Methodology

In this section, we introduce a new mobility concept that we can calculate for all U.S.-born Americans from twentieth-century birth cohorts. We relate the father's *income score* to the adult child's *family income*. Our income score predicts father's income using occupation, race, and Southern region (at minimum) and in some cases education as well. So, importantly, we do not assign the same income scores to white and Black respondents. Our use of the adult child's family income instead of the child's occupation in adulthood allows us to incorporate women as well, given that many women did not work in the historical period but can still report a measure family income to survey enumerators.

3.1 Calculating father's income scores

IPUMS provides 1950-based occupational income scores that go back as far as 1850, which calculate the median total income of the people (pooling men and women) in each occupation in 1950. We modify the standard IPUMS *occscore* variable in a number of ways.

First, not all our surveys give father's occupation categories that are as detailed as those in the Census. Across all of our surveys, we can harmonize occupations into 28 categories (corresponding to the categories in the ANES). We thus sort Census occupations into these 28 bins.

Second, we limit the Census samples to men between the ages of 30 to 50 who are living with a biological child less than 18 years old. This sample restriction should better proxy family income of *fathers* with a given occupation, which is the population of interest when we try to predict income during the respondent's childhood.

Third, we use the 1940 instead of 1950 Census (though we show robustness to using 1950, as well as robustness to many other modifications of the family income score).

The 1940 Census better captures the pre-Great-Compression wage and income distribution, a point made by Collins and Wanamaker (2017) and Ward (2020). Moreover, the full-count sample of the 1940 Census is available, whereas only the one-percent sample of the 1950 is currently public and only sample-line respondents are asked about income, which helps reduce noise when estimating the median income of smaller cells.

Finally, unlike the standard Census *occscore* variable, we take the median household income of this sample by occupation, race (Black versus white) and region (South versus elsewhere). Given widespread discrimination and occupational segregation, using occupational scores computed from pooled Black and white populations likely mismeasures occupational incomes. In order to focus on Black-white differences in absolute and relative mobility, we thus generate occupation scores separately by race.

One feature of historical measurement of occupational incomes is that farm income is notoriously difficult to impute, as it is both highly volatile (being subject to weather and price shocks) as well as difficult to measure (as comprehensive measurement of agricultural costs is difficult). Moreover, the 1940 Census income variable excludes income from self-employment, which includes most farmers. We therefore follow the approach of Collins and Wanamaker (2017) to calculate the income of farmers in 1940, using the income of farm laborers in 1940 as well as the ratio of farmer-to-farm-laborer income in the 1960 Census to impute the income of 1940 farmers. We similarly adjust the income of self-employed businessmen in 1940 using a similar approach. More detail on these adjustments is available in Appendix B.

Figure 2 compares the standard Census 1950 occscore variable (on the y-axis) to our income score (on the x-axis), all in 1950 dollars. Not surprising, income scores of Black individuals are almost always to the left of whites', and in particular Black Southern income scores (gray diamonds) are to the left of Black income scores from other regions (pink circles). A Southern income-score penalty exists for white fathers as well, with the blue solid triangles (white Southerners) typically to the left of the green solid circles (white non-Southerners).

While we show robustness to several modifications of this income score later in the paper, this measure serves as our baseline income score, as we can calculate it for all of our sample. We briefly foreshadow some of the adjustments we make to predicted childhood income here, and defer details to Section 4.6. A natural question is the validity of our 1940 occupation scores for earlier cohorts, especially for farmers (Feigenbaum 2018). We use earlier data sources to measure farm income (namely, the 1900 Census of Agriculture to construct race-by-region farmer income) as well as non-farm income (namely, the 1901 Cost of Living Survey). For surveys in which father's educational attainment is available, we also use this information to predict parental income during childhood.

3.2 Comparison to past proxies for parental income

Data limitations have long plagued the study of mobility in the United States, and our approach is no exception. We briefly review the main approaches in the literature, highlighting their advantages and disadvantages to better put our approach and results in context.

Papers using historical data

An advantage of studying older cohorts is that the Census provides de-identified data for those in the 1940 Census and earlier Censuses (and will de-identify the 1950 Census next year). Recent papers have using linking algorithms to find the same individual across Censuses based on their name, year of birth and place of birth. Ferrie (1996) was an early and important contribution to this literature.

However, this approach is not without complications and limitations. First, there is an active literature on the correct linking methodology and the preferred tolerance for rates of falsely matching and missing true matches (see, e.g., Abramitzky *et al.*, 2019 and Bailey *et al.*, 2020). Matching methodologies are still in flux and best practices will likely evolve as machine-learning techniques improve. Second, at least with the available current technologies, the linked population is not representative of the full population. Most obviously, the linked sample is not representative by sex, as women often change their names upon marriage and thus a representative group of women cannot generally be linked. To date, all published mobility papers using Census linking drop all women. Even beyond gender, certain types of names are very hard to link with precision. For example, there are too many John Smiths born in New York State in any given year to know with confidence that men with those characteristics in two different Census years are in fact the same person; conversely, long, foreign names are often changed, preventing matches.

An important example of a group which proves challenging for Census linking are Black Americans. For example, an important paper that does include Black Americans is Ward (2020), but his linked sample is only two-percent Black before those observations are up-weighted. Collins and Wanamaker (2017) are able to find reliable adult matches for three and five percent of Black children in the 1880 and 1900 Census, respectively. Moreover, Black Americans, and particularly Black men, are systematically under-counted in Censuses.[2] In short, linking historical Censuses proves quite difficult beyond white men.

As income is not available until 1940, most mobility work using Census data focuses

[2] O'Hare (2019) calculates that the net under-count rate for the Black population has gone from 8.4% in 1940 to 2.5% in 2010.

on occupational status of the father (as we do).[3] But a single observation of father income has noise from two sources, as Ward (2020) recently highlights. First, fathers change occupations from year to year, especially when occupations are measured at the three-digit level that is often used in this literature. While the attenuation bias from occupation churn is likely smaller than that from year-to-year changes in family income, as occupation tends to be more stable, it could still be substantial. Ward (2020) shows that papers that measure mobility using a single year of father's income substantially over-estimate mobility due to attenuation bias. Second, Census-takers appear to record occupation with substantial error, at least in the historical period. As Ward (2020) details, in a case when a re-census was required in St. Louis in 1880, one-third of occupations were reported differently only five months later, despite the reference date for the occupation being unchanged.

Given the challenges of linking, researchers have turned to creative solutions. As already noted, Olivetti and Paserman (2015) use a synthetic panel of first names, which allows them to examine (married) women as well as men (though they only include whites). To the extent that children stay in their parents' households as an adult, then household surveys like the Census allow researchers to observe both child and parents *without* needing to link, an insight Card *et al.* (2018) and Hilger (2015) have used to study intergenerational mobility with respect to education.

Our approach in many ways circumvents the challenges associated with linking, though it introduces others. As our data come from simple cross-sections, they tend to be representative, as it is much easier for a survey to find a representative cross-section in a given year than to maintain representativeness as they follow a sample across time. Indeed, as we already discussed, the percent of Black respondents in our data is very close to that in the full U.S. population, even for our earliest cohorts.

We do not observe fathers for a single year, but rather observe them in the recollections of their children during their own prime-age adult years. In that sense, we do not face the problem that Census researchers face of having the bad luck of observing the father in a particularly unrepresentative year in terms of his occupation. It seems natural to assume that the adult child would remember the occupation her father mostly did, so the retrospective nature of our data likely aids in identifying the main occupation of the father.

A challenge for our method is that there is no way to rigorously assess the accuracy of the adult children in our sample in their recollection of their father's occupation (though we show some partial validations in Section 3.3). Our method is also not useful for measuring intergenerational income persistence for cohorts born before 1910—the

[3]But going forward, it is worth noting that even for children growing up in a period where the Census records family income, it can only provide a *single snapshot* of income, and thus future work using linked Census data will suffer from attenuation bias.

types of surveys we use only became common beginning in the 1950s and thus cohorts before 1910 cannot be studied without violating our age range of observing adult children between ages thirty and fifty.

Papers using more modern data

For those interested in studying more modern cohorts, two data sources are especially useful. First, some dataset have been collected with the express purpose of measuring intergenerational mobility, such as the PSID and the NLS datasets. Second, IRS data allow linking of a small number of cohorts (those born around 1980).

The PSID has many advantages for modeling intergenerational mobility (e.g., numerous observations of father's income to limit attention bias), but long panels suffer from attrition bias, where richer, more educated, and whiter populations stay in the sample for longer periods of time. Indeed, as we show in Appendix Table 1, individuals for whom we observe ten years of childhood household income have fathers who are slightly whiter and much more educated than the general population of fathers. Schoeni and Wiemers (2015) show that the patterns of attitrion by parent and child income results in biased estimates of intergenerational mobility.

Chetty *et al.* (2014b) pioneered the use of administrative data to study U.S. mobility. As most individuals must file taxes each year, it is typically possible to observe several years of parental income. Because administrative data typically include Social Security or other personal identification numbers, linking these data is much easier than linking historical Census data, and the resulting linked sample is a much larger share of the full population and more representative. Though even with these administrative data, roughly seven percent of children cannot be linked to parents for various reasons.

As the IRS data have only been available since the 1990s, intergenerational mobility measures to date can be confidently estimated with these data for only a few cohorts (those born around 1980 and shortly thereafter). For a child born in 1980, a researcher can see several years of parental data, though only for her adolescent years. And then (assuming a few years' lag for data processing) researchers can find that child in her late thirties to observe adult outcomes. For children born a few years after 1980, researchers have a better look at parental income during her childhood, instead of just her adolescence, but then must observe her in her mid- to early thirties. Thus, given the relatively limited time that these data cover, it is difficult to estimate mobility measures for even a single cohort when both the adult child and the parent is around age forty. Quite obviously, these data cannot be used to understand long-run changes in intergenerational relative mobility.

Relative to IRS data, our sample sizes are small, preventing us from breaking the data into neighborhoods or single percentiles as in Chetty *et al.* (2014a), Chetty and Hendren (2018a) and Chetty and Hendren (2018b). Our use of father's occupation (which we use in conjunction with race and region, and in some cases education, to transform into an income scores) assigns to each child the median household income of fathers in a certain occupation-race-region cell, and thus misses any within-occupation variance of father's income.

3.3 Validating our income score

We perform a number of exercises that we hope will bolster readers' confidence in the accuracy of our income score, while acknowledging we cannot fully validate the measure.

First, roughly speaking, brothers and sisters grow up in the same families in the US, so adult men and women should report similar occupations for their fathers. Strictly speaking, small differences might arise between the average income-score of men and women. If parents have sex-based stopping rules when making fertility decisions, then boys and girls might wind up growing up in systematically different families (as in Asher *et al.* (2018), using data from India), but evidence for sex-based fertility patterns in the US is much weaker. Second, even if boys and girls grow up in identical families in terms of parental income, small differences might arise because men have higher mortality rates and thus selection into surviving into prime age could have gender differences (especially for our oldest cohorts, men are less likely to live until age 50, a point to which we return later).

These small potential discrepancies notwithstanding, we would be suspicious of any parental income estimate that gives significantly different estimates for men and women. We thus regress the log of estimated parental income on a female dummy, separately for each of our birth decades. We repeat this analysis with ranked parental income as the outcome as well and report the results in Appendix Table 2. The coefficient on the female dummy is always close to zero, has no consistent sign, and of the 14 regressions, it is significant at the ten-percent level only twice (in the 1960s) and never at the five-percent level.[4]

We also validate our occupation measure by comparing the mix of (coarsened) occupations our respondents report their father having to that of actual fathers in the Census in the years the respondents would be children (Appendix Table 4). As noted earlier, Ward (2020) warns that Census-takers made errors in recording the occupation variable, but we would still be worried if our respondents' recollection of their fathers'

[4]Appendix Table 3 shows the top five occupations reported by male and female respondents in each birth cohort. In all birth cohorts, at least four—if not all five—of the top occupations coincide between male and female respondents.

occupation differed dramatically from fathers' occupations in the Census in the years the respondents were growing up.

Finally, Appendix Figure 1 shows the Gini coefficient based on our predicted parental income measure, separately by respondents' birth decade. We cannot compare these Gini coefficients to those from, say, the 1940 or 1950 Census, as our measure will pick up none of the inequality coming from within-cell variation. Nonetheless, the measure does capture the known decline in inequality from 1930s to 1950s.[5]

3.4 Specifications

We estimate variants of the following two specifications, both standard in the mobility literature. We begin with the classic log-log specification estimated in Becker and Tomes (1979):

$$\log(y_{ic}) = \beta^c \log(y_{ic}^p) + f(\text{age}_i) + \delta_y + \epsilon_i. \tag{1}$$

In this equation, β^c is an estimate of the IGE for cohort c. We control for a quadratic in the age at which we observe the adult child (though recall we already restrict sampled to be observed at ages 30–50, which should limit life-cycle effects). We also include survey-year fixed effects in all specifications.

Next, we following Chetty *et al.* (2014a) and calculate ranks for fathers and children. The rank of the father is the percentile (based on the income score described in the previous subsection) within all fathers having a child in cohort c. Similarly, the rank of the child is the rank of family income among all children born in cohort c. The mapping of child's rank to parent's rank (the copula of the joint distribution) tends to be linear and can handle zeros, which may be missed in the (logarithmic) IGE specification. Chetty *et al.* (2014a) focus on this specification:

$$\text{Rank}_{ic} = \gamma^c \text{Rank}_{ic}^p + \delta_y + \epsilon_i. \tag{2}$$

In this estimation, γ^c is an estimate of the rank-rank correlation for cohort c.

In the final panel of Table 2, we show these income and rank measures, where incomes are all in 1950 dollars. There is only minor top- and bottom-coding of the adult children's family income in each birth decade. Real family income of the children grows robustly over the 1910–1940 birth cohorts, consistent with strong post-war economic growth. Fathers income *score* grows more slowly, as by construction is can only represent occupational upgrading across time (as it is based on the 1940 income

[5]In this graph, inequality is high and declining slightly from the 1910s to the 1930s. Given limited income data from this period, it is difficult to compare our measure to any "ground truth" from the Census or other sources.

distribution, though we revisit this assumption in robustness checks).

An obvious and important point is that with sufficient granularity in the income measures, the average rank for fathers and children must both be fifty. Reassuringly, we see an average close to fifty across all of our birth decades. Missing information for parental income is a challenge even for modern, administrative data, and so we include the share of children who have missing information for father's income, which rises especially after 1950. We do not include these observations in most of the analysis, but later we show robustness to adding them back into the sample with various assumptions about missing fathers' incomes.

4 Results for representative samples

The main finding we describe in this section is a rise in intergenerational relative mobility between the 1910 and 1940 cohorts. In the next section we try to understand this trend by splitting up the sample by gender and race, but in this section we merely aim to establish the robustness of this main result.

4.1 Main results

The first series of Figure 3 shows the IGE for survey respondents over time, pooling across surveys and applying our baseline population-adjusted weights. We show the IGE separately by decade of birth. Between the 1910s and 1940s birth cohorts, the IGE falls markedly, from roughly 0.59 to 0.35. We then see an increase in this measure in subsequent birth cohorts, so that the IGE appears to take on a u-shape over time.

The second series shows the results from the rank-rank specification, which mirrors very closely those of the IGE in terms of trends. As is typically found in other papers, our rank-rank coefficients are lower in magnitude than our IGEs: it begins the sample period just under 0.40 and declines to a low of just over 0.23 for the 1940 birth cohort. Like the IGE, it also rises after 1940, though in a somewhat less pronounced manner.

For several reasons, we focus on the *decline* in the IGE and rank-rank measures that occurs from the 1910s to the mid-century birth cohorts, instead of the subsequent rise thereafter. First, as we noted in our discussion of Table 2, the share of data with missing information about fathers increases over time, so trends toward the latter part of our sample period might reflect sample selection.[6] Second, beginning with the 1960s birth cohorts, modern panel data such as the PSID and later linked administrative IRS become available, so we feel our relative contribution to understanding mobility patterns in the modern period is smaller.

[6]However, to the extent that the estimates in this later period suffer from measurement error, then if anything, the u-shape over this entire time period will be even more marked.

Figure 4 shows the decline in intergenerational persistence between the 1910–1919 cohorts and the 1940–1949 birth cohorts as bin-scatters figures. The first panel shows the change in the IGE relationship: a large shift upward (reflecting real income growth relative to parents across the parental income distribution) as well as a significant flattening of the slope (because the upward shift is especially large among individuals growing up with less family income).

The second panel of Figure 4 shows that the decline in the rank-rank is also large and precisely estimated. Given that by construction there can never be an overall increase in rank (its average must always be 50) we see only a flattening of the slope. While caution is warranted in terms of comparing the levels of our rank-rank estimates (which use parental income scores) to those from modern administrative data (which use actual income data averaged over several years from the parents), we use the modern estimates as rough benchmarks. The rank-rank slope we find for the 1910s–1920s cohorts is roughly equal to the modern US, whereas the slopes we find for mid-century cohorts is close to Canada's.

Finally, we plot a third series in Figure 3, the Gini coefficient based on the self-reported family income for the respondents in our surveys (that is, the adult children in the parent-child pair). Quite strikingly, the inequality and persistence measures move in tandem over these birth cohorts. As we noted in the Introduction, much of the support for the hypothesis that inequality and mobility are inversely related (the so-called Great Gatsby relationship) comes from cross-country data. Our results in Figure 3 provide some evidence for the Great Gatsby curve using variation over time within the US. To date, such evidence has been lacking. For example, Chetty *et al.* (2014b) finds no decrease in rank-rank income mobility over a short, more recent period (the 1971 to 1986 cohorts), despite the increase in inequality over this period.

Interestingly, Song *et al.* (2020) find a similar effect to our results for cohorts born at mid-century. They note that while their long-run results suggest general stability among birth cohorts in the twentieth century, a potential exception is the "drop in the intergenerational correlation estimated from pooled social surveys for the 1950 cohort (born between 1946 and 1955), namely the early baby-boom generation." While they do not emphasize it as much as the 1946–1955 decline, their persistence correlations are also lower for the 1936–1945 and the 1956–1965 cohorts, which we also find.[7] Song *et al.* (2020) write that "we consider the deviation of the 1950 birth cohort best interpreted as suggestive. Proper interpretation of this deviation awaits future research with further evidence." We view our results as adding yet another piece of evidence in support of a temporary but significant increase in mobility for mid-century U.S. birth cohorts.

[7] As Song *et al.* (2020) note, other papers finds hints of such a result as well. Using data from the GSS, Hout (1988) find rising intergenerational mobility of occupational status from the early 1970s to the mid-1980s, which would correspond to some of our most mobile birth cohorts.

Similarly, while Ward (2020) has a gap from 1920–1960 in their long-run time-series of male mobility, they find that mobility is much lower in 1920 than in 1960, consistent with our results for representative samples.

In summary, we have so far provided evidence of a significant decline in IGE and rank-rank persistence measures between the 1910s and 1940s birth cohorts. Importantly, these results reflect samples that are representative of the full U.S.-born population, including women and non-whites. In the subsequent subsections, we attempt to show robustness of our results to what we consider the most central concerns.

4.2 Corroborating evidence from respondent's education

On average, more educated individuals have higher earnings and family income, with the return to education varying over time. Thus, it would be somewhat surprising if the predictive power of parental income on children's education did not fall given that its predictive power over children's family income did.

To examine this idea, we estimate variants of equations (1) and (2) where in both cases we put the adult child's self-reported years of schooling as the outcome variable (available in all of our datasets). Figure 5 shows the results from both of these estimations, as usual, by birth decade. The relationship between father's income and respondent educational attainment declines sharply between the 1910 and 1950 birth cohorts. Appendix Figure 2 illustrates these changes using bin-scatter figures, and highlights that this weakening relationship is largely driven by the rapid increase in respondents' high school completion in the bottom half of the income distribution.

4.3 Adjusting for father's education

As noted earlier, we proxy the child's family income while she was growing up by taking the median of 1940 Census household income for all fathers with the same occupation, race and region (South versus elsewhere). Our hope is that this measure can proxy for the respondent's long-run income during childhood. The extent to which it does *not* pick up idiosyncratic, mean-zero variation in family income from year to year is in fact a strength, as it reduces attenuation bias. But if it misses systematic variation in family income not picked up by *occupation* $\times$ *race* $\times$ *South*, then it will lead to bias. Moreover, it is not *a priori* obvious the sign or trend over time of the bias.

As noted earlier, for more than half of our surveys, respondents were also asked about their fathers' education. We can thus augment the fathers' income scores by predicting income by *occupation* $\times$ *race* $\times$ *South* $\times$ *education category*. Father's education is one of the most important reasons why family income could systematically deviate from our *occupation* $\times$ *race* $\times$ *South*-based income score. Indeed, adding information about education significantly increases the power of our income scores to predict 1940 family income (the R-squared rises from 20.6 to 24.3).

The first panel of Appendix Figure 3 compares the IGE with the original *occupation*$\times$ *race*$\times$*South* income score to those with these augmented scores, both using the same sample of respondents who are asked fathers' education. The two series are nearly identical in both levels and trends: in particular, both show the marked decline between the 1910s and 1940s birth cohorts. The second panel shows that the decline in the rank-rank measure is also unchanged by augmenting parental income score with father's education. Thus, when we significantly improve our income scores with an important predictor, the trends in mobility remain unchanged, providing some reassurance that systematic, unobserved within *occupation*$\times$*race*$\times$*South* cell variation in income is not driving our results.

4.4 Adjustments to farm income

Past work (e.g., Song *et al.*, 2020, Collins and Wanamaker, 2017, Ward, 2020) on historical mobility has paid special attention to farmers, both because they are such a large part of the population in the nineteenth and early twentieth centuries and because their income is hard to model. For example, it is well known that the relative position of farmers declined from the late nineteenth to mid-twentieth century. Using data from the 1950 U.S. Census and the 1915 Iowa Census, Feigenbaum (2018) finds that 1950 IPUMS occscores are a good predictor of actual income in 1915, with the important exception of farmers. Farmers experienced a substantial fall in relative income between 1910 and 1950 (for example, in 1915 Iowa, famers made median income, but in the full 1950 U.S. population, farmers were at the bottom of the income distribution).

As noted earlier, our baseline income scores acknowledge the difficulty in using income data from the 1940 Census to measure farm income, and thus follow the approach in Collins and Wanamaker (2017) to adjust farmers' incomes. In our baseline measure, for example, white respondents born in the 1910s–1920s cohorts outside of the U.S. South and who have farmer fathers are estimated to be growing up around the 30th percentile of the income distribution.

Nevertheless to test for the robustness of the main result to this baseline adjustment for farmer income, we begin by using an alternative source of data to calculate farmer income. Specifically, we follow the approach in Goldenweiser (1916) and Abramitzky *et al.* (2012) and use the 1900 Census of Agriculture to calculate farmers' net earnings. In our calculations, we allow for variation at the *race*$\times$*South* level and take into account the share of each group that is not farm owners (i.e., part owners, or cash or share tenants). Using this data source tends to increase farmer fathers' rank in the income distribution for the earliest cohorts (e.g., by roughly five rank percentiles for white farmers outside of the South). Appendix Figure 4 shows that our main result of a marked decline in persistence remains when we use this alternative data source.

As an alternative approach, we simply drop farmers to ensure that our mobility patterns are not being entirely driven by this population, a population for which it is hard to estimate childhood income. Again, the conclusion that mobility increased substantially between the 1910s and 1940s birth cohorts persists.

4.5 Alternatives to household income in the 1940 Census

We also check the robustness of our main result to alternative ways of constructing income scores. We begin by using father's income—as opposed to household income—for fathers with a given occupation, race, and Southern residence. Appendix Figure 5 shows that the IGE, and especially the rank-rank estimates, hardly change and the decline in persistence remains.

We then consider using an alternative data source, not only for farmers, but for all occupations. We combine our estimates from the 1900 Census of Agriculture with average earnings from the 1901 Cost of Living Survey and again adjust for variation at the $race \times South$ level (Preston and Haines 1991). The third series in this figure shows that despite using completely different data sources—measured forty years apart in time—for both agricultural and non-agricultural occupations, the increase in mobility remains marked between the 1910s and 1940s cohorts. These alternative income scores also allow us to account for changes in the relative status of certain occupations over the first half of the twentieth century. For example, in the earliest cohorts, children whose fathers were semi-skilled operative and kindred workers have ranked childhood income around the 30–35th percentile of the income distribution with this alternative measure, compared to the 40–50th percentile using our baseline measure.

Finally, we combine our data sources, so that fathers are assigned income scores using the most contemporaneous data sources possible. In other words, the 1910 and 1940 cohorts are given income scores that use the 1900- and 1940-based income scores, respectively. The middle cohorts are given a weighted average of the two, and the later cohorts are assigned income scores using later Censuses than the 1940 one (i.e., the 1950, 1960, and 1970 Censuses). Again, the general patterns for both the IGE and rank-rank estimates remain unchanged.

4.6 Other robustness checks

Table 2 shows that the information needed to calculated predicted childhood income is not always available. This situation arises almost always because the respondent does not report father's occupation (e.g., because she doesn't remember, chooses not to report it, or she grew up without her father). In Appendix Figure 5 we also show robustness to a particular extreme assumption about this group: that their household had zero income, or in other words, that their family had the lowest possible rank for predicted childhood income.

Moreover, as discussed in Section 3.1, while we believe that our revisions to the IPUMS *occscore* methodology are valuable and appropriate, we show in Appendix Figure 5 robustness to merely using the *occscore* variable instead of our income score.

Finally, Appendix Figure 6 shows the robustness of the main result to alternative weighting schemes: namely, using the provided survey weights without any additional adjustments for population shares and using no weights at all. In all of these checks, we continue to find a marked increase in mobility between the 1910 and 1940 cohorts.

5 Decomposing the rise in mobility

In this section, we show how to decompose the overall IGE or rank-rank relationship into factors related to subgroups. We then use the results from the previous section to show how much changes in mobility or income among subgroups, particularly race and gender, explain the overall decline in IGE and rank-rank coefficients that we found in the previous section.

5.1 Decomposing the IGE and rank-rank slopes

Consider any partition of the full sample, emitting subgroups $g \in G$ with subgroup g's share of the total sample given by p_g. Further, let β_g^{IGE} be equal to β from estimating equation the IGE equation $y_i^c = \alpha + \beta y_i^p + e_i$ (where, as usual, y^c and y^p are the adult child's and the parent's log income, respectively) on the subgroup g.

From the OLS formula and the law of total covariance, the whole-population IGE is given by:

$$\begin{aligned}
\beta^{\text{IGE}} =& \frac{Cov(y^c, y^p)}{Var(y^p)} \\
=& \frac{1}{Var(y^p)}\Big(E_g[Cov(y^c, y^p)] + Cov(E[y^c \mid g], E[y^p \mid g])\Big) \\
=& \underbrace{\sum_{g \in G} p_g \frac{Var(y^p|g)}{Var(y^p)} \beta_g^{\text{IGE}}}_{\text{Weighted average of subgroup slopes}} + \underbrace{\frac{Cov(E[y^c|g], E[y^p|g])}{Var(y^p)}}_{\text{Between-group covariance of subgroup averages}},
\end{aligned} \tag{3}$$

where E_g denotes the expectation over groups g.

A slight modification gives a similar expression for the whole-population rank-rank slope. In this case, β_g^{RR} is the beta from $r_i^c = \alpha + \beta r_i^p + e_i$ (where, as usual, r^c and r^p are the within-cohort rank of the child and the parent) on the subgroup g. The same application of the law of total covariance gives:

$$\begin{aligned}\beta^{RR} =& \sum_{g\in G} p_g \frac{Var(r^p|g)}{Var(r^p)}\beta_g^{RR} + \frac{Cov(E[r^c|g], E[r^p|g])}{Var(r^p)} \\ =& 12 * \Big(\sum_g p_g Var(r^p|g)\beta_g^{RR} + \sum_g p_g \mathbf{E}[r^p|g]E[r^c|g] - 0.25\Big)\end{aligned} \tag{4}$$

To ease intuition and to focus on the key application for our paper, we rewrite the IGE decomposition for two groups, W and B:

$$\begin{aligned}\beta^{\mathrm{IGE}} =& p_w \frac{Var(y^p|W)}{Var(y^p)}\beta_W^{\mathrm{IGE}} + (1-p_W)\frac{Var(y^p \mid B)}{Var(y^p)}\beta_B^{\mathrm{IGE}} \\ &+ \frac{p_w \mathbf{E}[y^p \mid W]\cdot \mathbf{E}[y^c \mid W] + (1-p_W)\mathbf{E}[y^p \mid B]\cdot \mathbf{E}[y^c \mid B] - \mathbf{E}[y^p]\mathbf{E}[y^c]}{Var(y^p)}.\end{aligned} \tag{5}$$

The decomposition makes clear the important role of between-group differences in parental income. To see this point, assume for the moment that W and B are two distinct subgroups, but are drawn independently *from the same distribution* of parental income y^p. Thus, there exists no between-group variation in y^p and $Var(y^p) = Var(y^p \mid W) = Var(y^p \mid B)$ and $\mathbf{E}[y^p] = \mathbf{E}[y^p \mid W] = \mathbf{E}[y^p \mid B]$. In this special case of no between-group differences in parental income, $\beta^{\mathrm{IGE}} = p_w\beta_W^{\mathrm{IGE}} + (1-p_W)\beta_B^{\mathrm{IGE}}$, or in other words, the full-population IGE is the average of the two subgroup IGE slopes weighted by the subgroup share of the total population. While it is clear that racial samples are drawn from different distributions of parental income, it is also clear that gender subgroups are drawn from the same distribution of parental income: men and women grow up in the same households as children.

5.2 Mobility by race and sex

As we noted in the Introduction and shown above, an important reason to examine the mobility of representative samples is that relating mobility measures of subgroups to the full-cohort mobility is complicated. In this subsection, we show the mappings of father to adult children's incomes separately by the race and gender of the respondent. We then use the decomposition in equation 3 to show how changes in relative mobility and relative income of these groups contributed to (or stymied) the overall rise in mobility from the 1910s to the 1940s birth cohorts. Given the discussion above, we expect that the between-group component will prove important for a decomposition along racial subgroups, but it must be the case that if changes in mobility by gender are important, it would have to be because within-gender mobility changes a lot for one gender relative to the other (given that men and women have virtually identical population shares and distributions of parental income).

5.2.1 Main results by race

We begin by showing Black and white mobility for the earlier, less-mobile 1910–1920s cohorts compared to the more mobile 1940s–1950s cohorts, in Figure 6 (IGE in subfigure (*a*) and rank-rank relationships in (*b*)). Perhaps the most striking aspect of the graph is how little overlap there is in the support of the Black versus white mobility graphs: Black fathers' income overlaps only modestly with white father's income. In the rank-rank figure, almost no whites grew up in the bottom ten percent of predicted family income and few Black respondents grew up with income above the 30th percentile, so the overlap of the two groups happens almost entirely between the tenth and thirtieth percentiles of parental income. One advantage of our "small data" is that the vast differences between how Black and white adult children grew up is readily apparent: with "big data" one can capture the tiny number of Black children who grew up in rich families and thus extend the regression lines over the entire 0–100 domain of parental income rank. But even today prime-age Black adults are vastly under-represented in the upper parts of the parental income distribution while growing up.[8]

Another striking result is the significant catch up among Black individuals between these two periods. In the IGE graph, the entire Black regression line shifts upward by about thirty log points, whereas these is a much more modest upward shift for whites. The rank-rank graph shows a similar pattern, as we would expect. Whereas a child in the earlier cohorts born to a Black father at the tenth percentile (which is a very typical percentile for Black children in this era) would be predicted to have an adult family income around the 25th percentile (compared to the 40th percentile for a similarly situated white child from this era). But for mid-century cohorts, Black children born at the tenth percentile (still a very typical place for Black children) are predicted to appear at the 35th percentile as adults (thereby halving the gap with their white counterparts from around 15 to 7 percentile ranks).

While we have so far focused on Black-white catchup, the regression lines explaining white mobility also change over time. In both the IGE and the rank-rank estimates, the slopes flatten modestly. As the large majority group, the flattening of the mobility slopes among white individuals will have an important effect on the overall full-cohort IGE, as is clear from the decomposition above.

[8] The tiny share of Black children in the upper ranks of parental income distribution even in modern data can be seen in the Appendix figures of Chetty *et al.* (2018).

5.2.2 Main results by gender

As discussed in the introduction, a major motivation for our mobility concept is that it enables us to look at the intergenerational mobility including women. In this subsection we look separately at patterns of mobility by gender. Figure 7 shows IGE and rank-rank estimates separately by gender. For both measures and for all birth decades, persistence measures for women are greater than or equal to those for men. But they do not exhibit any differences over time.

Examining Figure 7 though the lens of equation 5, with the groups being men and women highlights that including women increases the overall IGE. This is because women and men come from the same distributions of parent income (growing up together in the same households), and have generally equal population shares. So, the overall IGE is the simple average of the male and female IGE, and since the female IGE is higher than the male IGE, this pulls down the overall level of mobility. But since the pattern of IGE changes for women closely tracks that of men, it would be difficult to explain the changing patterns of overall mobility with factors that are gender-specific (even though women are generally poorer than men throughout our sample period, this does not differentially covary with parent income).

As noted in Section 2, some of our datasets include only women (e.g., the National Longitudinal Study of Mature Women) or only men (the Occupational Change in a Generation datsets), so a possible concern is that the differences in mobility are coming entirely from the fact that the male and female estimates are in some cases coming from different datasets. In Appendix Figure 7 we show robustness to limiting to datasets that include both men and women. For the IGE, we still see that women's persistence measures are in all birth decades greater than those of men; for the rank-rank, the only violation is the 1930s.

Using a different methodology and only considering married women, Olivetti and Paserman (2015) find that women are less mobile than men in the mid-1800s but slightly more mobile by the 1920s. Note that our sample periods do not overlap, as even our oldest cohorts—those born in the 1910s—are not observed until the 1950s. Importantly, unlike Olivetti and Paserman (2015), we can include women who are unmarried—that is, who are never married, widowed, or divorced.

5.3 Main results by race and sex

Figures 8 and 9 further breaks down the results in Figure 6 by sex as well. Figure 8, shows that, among men, Black individuals have closed almost the entire mobility gap with whites by mid-century. Of course, as the graph also makes clear, Black men still grow up in poorer households, so their adult income is still much lower than whites.

But by mid-century, essentially the same regression line explains Black and white male IGE mobility. Similarly, for the rank-rank results, men born between the tenth and thirtieth percentiles (the only part of the parental income distribution where there is significant overlap between Black and white respondents) are mapped to relatively similar places in the adult income distribution, regardless of race (roughly the 35th to 45th percentiles). Importantly, if closure of the mobility gap had held for future cohorts, the Black-white family income gap for men would have closed within a few generations.

That Black women outperform Black men (either in absolute terms or in comparison to gender gaps among whites) is a robust finding using more modern data. An important recent contribution to this literature is Chetty *et al.* (2018). For many important outcomes, there is no racial mobility gap at all among women (i.e., while there is still an absolute racial gap among women in, say, college-going, it is nearly entirely explained by the fact that Black women are born to poorer families; but Black and white women from the same point in the parental-income distribution get mapped to nearly the same outcomes in adult life). An important contribution of our focus on representative cohorts is that we can examine whether this gender gap among Black individuals has always held or is a newer phenomenon.

As noted, our results by sex and race suggest that the better outcomes of Black women relative to men did not hold historically. While Black women and men grow up in the same families, the Black women we observe as adults are substantially poorer (while white women are also poorer than white men, the gap is much smaller). In contrast to more modern cohorts, marriage rates for the cohorts we study are high for both Black and white respondents (typically around 80–90 percent are ever married), which deepens the puzzle of how Black men and women have such different outcome as adults.

One key similarity, however, is significant convergence between Black and white women from the 1910s–1920s cohorts to the 1940s–1950s cohorts, with the regression line depicting Black women's IGE shifting upward a similar magnitude to that for Black men. However, because Black women born in 1910s–1920s are so poorly off as adults (by far the poorest of the four groups we study), the Black-white gap among women in the 1940s–1950s birth cohorts is still quite pronounced.

In Figure 10, we show the patterns of IGE cumulating our four subgroups beginning with white men, then adding white women, Black men, and finally Black women. There is little gender difference in IGE for whites, but there is a larger difference for Black individuals. Adding the two margins of race and gender dramatically increases the IGE.

Appendix Figure 8 suggests a proximate cause of this relative improvement for

Black women: the improving mortality of Black men. Widowhood is a major risk for Black women in our sample, even though we require respondents to be observed between ages 30 and 50, still in their prime-age years. In the early cohorts, roughly ten percent of black women in our sample report being widows. The relationship with father's income is noisy, but suggests if anything a positive relationship of widowhood and father's status. For our mid-century cohorts, the average share of black women who report being widows has fallen to less than five percent. In this case, the gradient with father's income is the more expected negative sign, though it is also noisy. White women in the early cohorts were not complete strangers to widowhood either, with just under five percent self-reporting as widows. Interestingly, there is very little gradient with respect to father's predicted income. For the mid-century birth cohorts, this share has fallen to nearly zero.[9]

5.4 Decomposing the decline in overall mobility by race and sex

The results above suggest that racial catch-up played an important role in declining mobility, with a limited role for gender differences. Returning to Figure 6 with the decomposition in mind allows us to assess the effects of the various movements in the by-race IGE mappings. Figure 6 depicts a number of different changes over time, some of which will increase mobility (the level increase for Black respondents, the slope decrease for white and Black individuals) and some of which will reduce mobility (the level increase for whites). The decomposition can quantify the various contributions.

We begin by considering the changes in levels—that is, a positive shift for both Black and white Americans, though a larger increase for Black individuals, resulting in significant catch-up. Figure 11 shows that if Black individuals had experienced the same real income growth as white individuals (without changing the slopes for either group), then 60 percent of the IGE and 45.5 percent of the rank-rank decline would not have been realized. Thus, Black respondents' catch-up to whites in levels explains a large share of the total decline in persistence, despite them only being a small share of the population.

We can also ask what share of the total decline in persistence is explained by the flattening of the white mobility slope. Figure 11 shows that the change in the white slope accounts for much of the remaining change in mobility between the 1910–1920 and 1940–1950 cohorts, with very similar effects across the two measures (57% in the case of rank-rank, 56% in the case of IGE).

[9] In Appendix Figure 9, we show that these high rates of widowhood for Black women are also found in the Census.

6 Conclusion

We provide, to the best of our knowledge, the first evidence on long-run intergenerational relative mobility trends for representative samples of the U.S.-born population. We find a robust decline in IGE and rank-rank persistence measures from the 1910s to the 1940s birth cohorts. For cohorts born after the 1940s, we find that persistence drifts back upward. These persistence measures track the u-shape in inequality measures of the family income of the adult children in our samples. We thus provide some of the first evidence that the "Great Gatsby Curve" holds using within-country variation across time, instead of cross-sectional variation across countries.

Besides presenting mobility estimates for samples that are representative of the full U.S.-born population, we also show formally how the movement of subgroups (e.g., by sex or by race) contributes to the change in the full-population mobility measures. Relative changes for small groups can have disproportionately large effects so long as they are drawn from extreme parts of the parental-income distribution. We show that Black respondents catching up to whites in levels explained 60% of the rise in mobility, despite being only 10–12% of the population.

References

Abramitzky, R., Boustan, L. P. and Eriksson, K. (2012). Europe's tired, poor, huddled masses: Self-selection and economic outcomes in the age of mass migration. *American Economic Review*, **102** (5), 1832–56.

—, —, —, Feigenbaum, J. J. and Pérez, S. (2019). *Automated linking of historical data*. Tech. rep., National Bureau of Economic Research.

Asher, S., Novosad, P. and Rafkin, C. (2018). Intergenerational mobility in india: Estimates from new methods and administrative data. *World Bank Working Paper*.

Bailey, M. J., Cole, C., Henderson, M. and Massey, C. (2020). How well do automated linking methods perform? lessons from us historical data. *Journal of Economic Literature*, **58** (4), 997–1044.

Becker, G. and Tomes, N. (1979). An equilibrium theory of the distribution of income and integenerational mobility. *JPE*, **87** (6), 1153–1189.

Berman, Y. (2018). The long run evolution of absolute intergenerational mobility. *Available at SSRN 3269831*.

Bowles, S. (1972). Schooling and inequality from generation to generation. *Journal of Political Economy*, **80** (3, Part 2), S219–S251.

Card, D., Domnisoru, C. and Taylor, L. (2018). *The intergenerational transmission of human capital: Evidence from the golden age of upward mobility*. Tech. rep., National Bureau of Economic Research.

Chetty, R., Grusky, D., Hell, M., Hendren, N., Manduca, R. and Narang, J. (2017). The fading american dream: Trends in absolute income mobility since 1940. *Science*, **356** (6336), 398–406.

— and Hendren, N. (2018a). The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects. *The Quarterly Journal of Economics*, **133** (3), 1107–1162.

— and — (2018b). The impacts of neighborhoods on intergenerational mobility ii: County-level estimates. *The Quarterly Journal of Economics*, **133** (3), 1163–1228.

—, —, Jones, M. R. and Porter, S. R. (2018). *Race and economic opportunity in the United States: An intergenerational perspective*. Tech. rep., National Bureau of Economic Research.

—, —, — and — (2020). Race and economic opportunity in the united states: An intergenerational perspective. *The Quarterly Journal of Economics*, **135** (2), 711–783.

—, —, Kline, P. and Saez, E. (2014a). Where is the land of opportunity? the geography of intergenerational mobility in the united states. *The Quarterly Journal of Economics*, **129** (4), 1553–1623.

—, —, —, — and Turner, N. (2014b). Is the united states still a land of opportunity?

recent trends in intergenerational mobility. *American Economic Review*, **104** (5), 141–47.

Collins, W. J. and Wanamaker, M. H. (2017). *African American Intergenerational Economic Mobility Since 1880*. Tech. rep., National Bureau of Economic Research.

Corak, M. (2013). Income inequality, equality of opportunity, and intergenerational mobility. *Journal of Economic Perspectives*, **27** (3), 79–102.

Feigenbaum, J. J. (2015). Intergenerational mobility during the great depression.

— (2018). Multiple measures of historical intergenerational mobility: Iowa 1915 to 1940. *The Economic Journal*, **128** (612), F446–F481.

Ferrie, J. P. (1996). A new sample of males linked from the public use microdata sample of the 1850 us federal census of population to the 1860 us federal census manuscript schedules. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, **29** (4), 141–156.

Goldenweiser, E. A. (1916). The farmer's income. *The American Economic Review*, **6** (1), 42–48.

Hilger, N. G. (2015). *The great escape: Intergenerational mobility in the united states since 1940*. Tech. rep., National Bureau of Economic Research.

Hout, M. (1988). More universalism, less structural mobility: The american occupational structure in the 1980s. *American Journal of sociology*, **93** (6), 1358–1400.

Long, J. and Ferrie, J. (2013). Intergenerational occupational mobility in great britain and the united states since 1850. *American Economic Review*, **103** (4), 1109–37.

Mazumder, B. (2015). Estimating the intergenerational elasticity and rank association in the us: Overcoming the current limitations of tax data.

— (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.

Merriam, W. R. (1902). *Census Reports: Twelfth Census of the United States, Taken in the Year 1900*, vol. 2. Washington, United States Census Office.

O'Hare, W. P. (2019). Census coverage of the black population. In *Differential Undercounts in the US Census*, Springer, pp. 83–91.

Olivetti, C. and Paserman, M. D. (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.

Preston, S. H. and Haines, M. R. (1991). Appendix a, assigning income and unemployment estimates to individuals in the national sample of the 1900 united states census. In *Fatal Years: Child Mortality in Late Nineteenth-Century America*, Princeton University Press, pp. 211–220.

Roemer, J. E. and Trannoy, A. (2015). Equality of opportunity. In *Handbook of income distribution*, vol. 2, Elsevier, pp. 217–300.

Ruggles, S., Flood, S., Foster, S., Goeken, R., Pacas, J., Schouweiler, M. and Sobek, M. (2021). IPUMS USA: Version 11.0 [dataset].

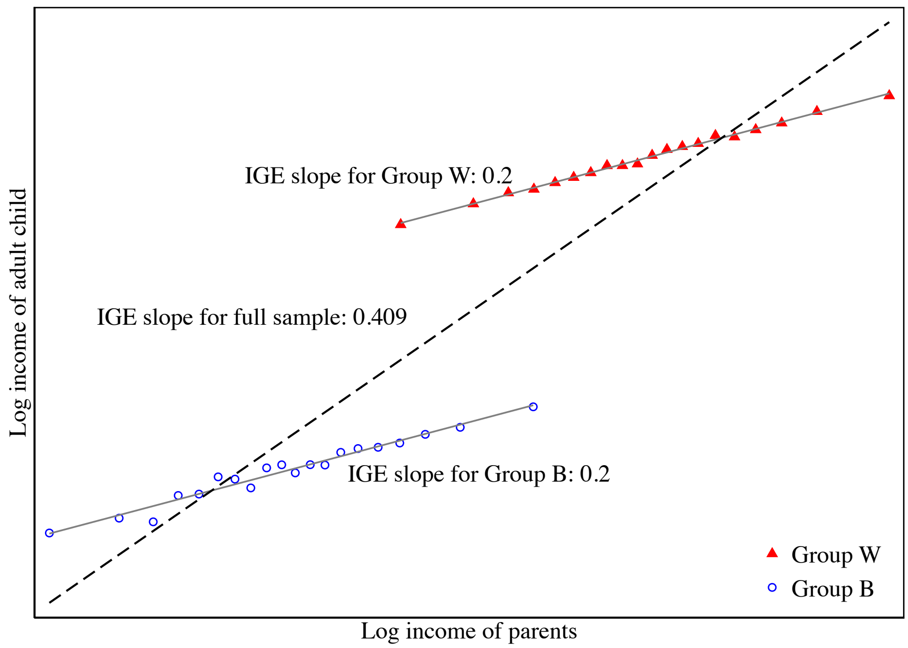
Schoeni, R. F. and Wiemers, E. E. (2015). The implications of selective attrition for estimates of intergenerational elasticity of family income. *The Journal of Economic Inequality*, **13** (3), 351–372.

Solon, G. (1992). Intergenerational income mobility in the united states. *The American Economic Review*, pp. 393–408.

Song, X., Massey, C. G., Rolf, K. A., Ferrie, J. P., Rothbaum, J. L. and Xie, Y. (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.

Ward, Z. (2020). Intergenerational mobility in american history: Accounting for race and measurement error.

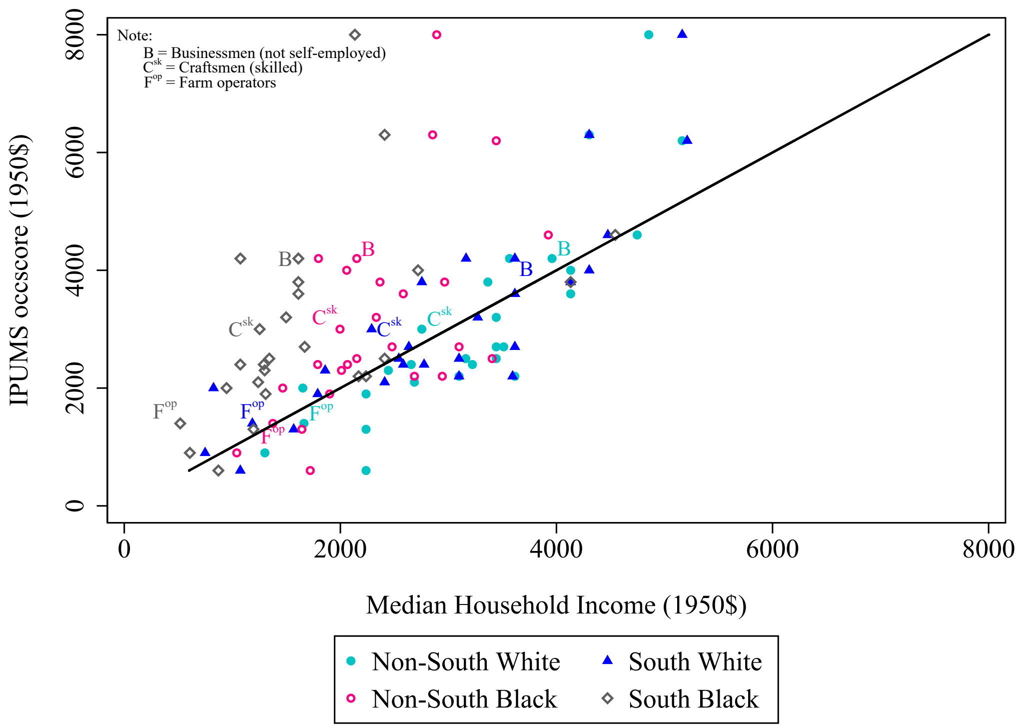
Figure 1: Illustration that IGE slope is not a weighted average of sub-sample IGE slopes



Sources: Data generated by the authors for the sake of illustration.

Notes: In this case, group W is the large majority (90% of the population) and B is the minority (10% of the population).

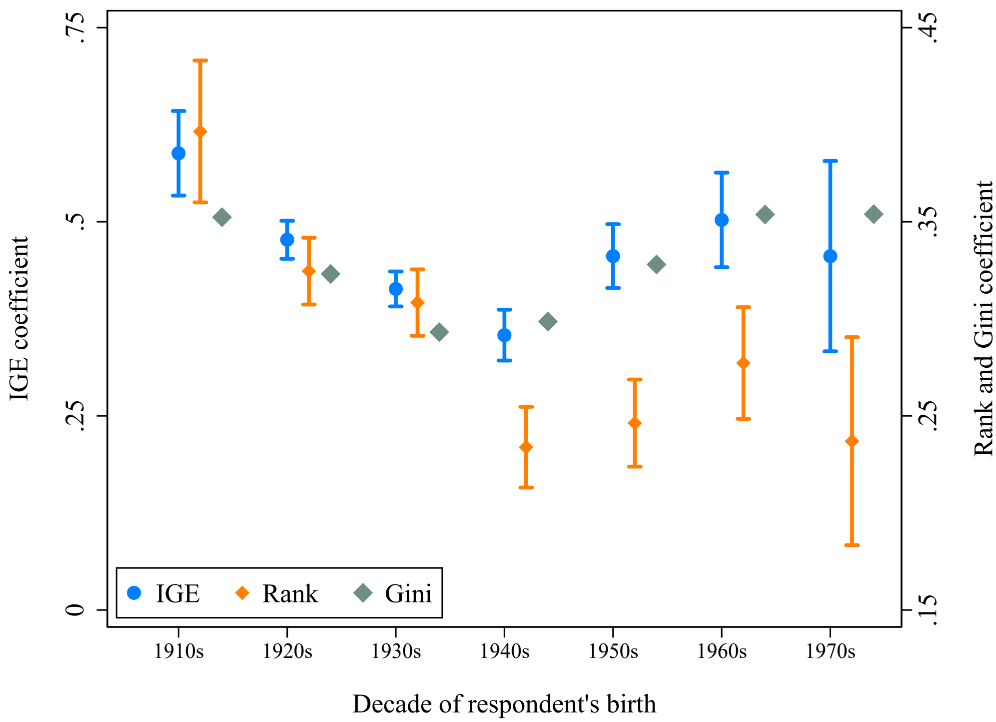
Figure 2: Comparing our income scores to IPUMS *occscore* variable



Sources: 1940 Census.

Notes: We sample all men between the ages of 30–50 living with at least one biological child under the age of 18. The y-axis plots the *occscore* variable generated by IPUMS. The x-axis plots our income score (which is calculated specifically for this subsample and predicts income as a function of occupation, race, and Southern residence). We include a 45-degree line to aid comparisons. We highlight with labels a few salient occupations.

Figure 3: IGE and rank-rank measures by birth decade

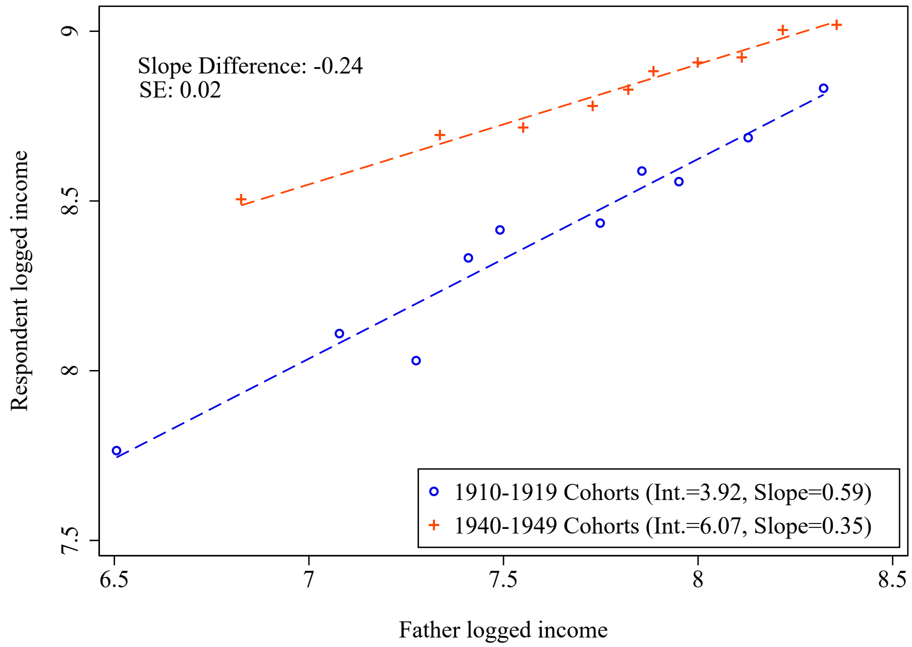


Sources: This figure combines 14 different surveys, which are described in Section 2 and in further detail in Appendix B.

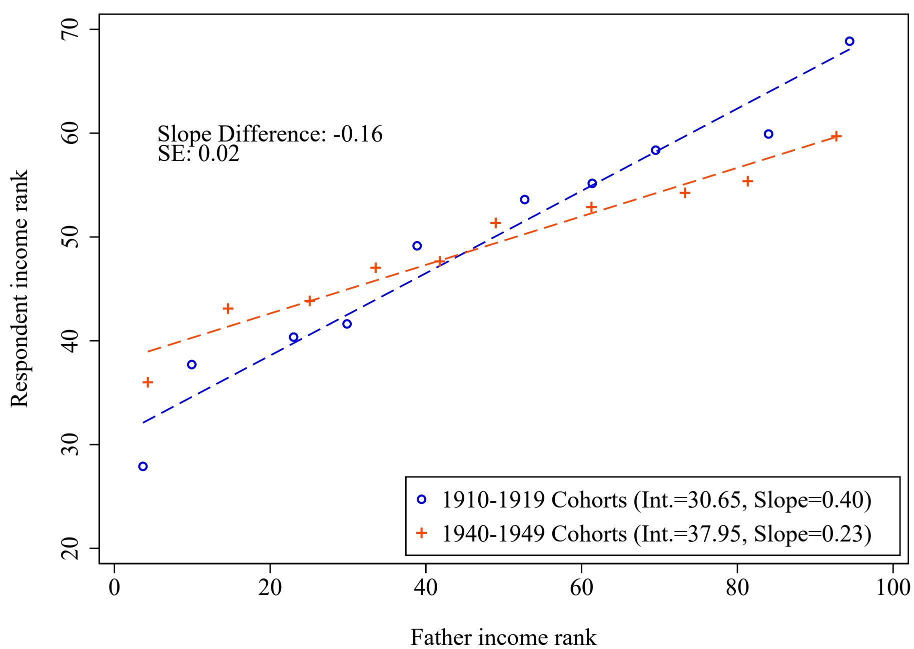
Notes: The IGE and rank-rank are based on the baseline sample of respondents age 30–50. The Gini coefficient uses the self-reported family income of the *children* in our samples. We use sample weights where provided and further re-weight each birth cohort (i.e., decade) so that they have representative *race* × *sex* shares.

Figure 4: Bin-scatter depictions of the decline in intergenerational persistence

(a) Intergenerational elasticities



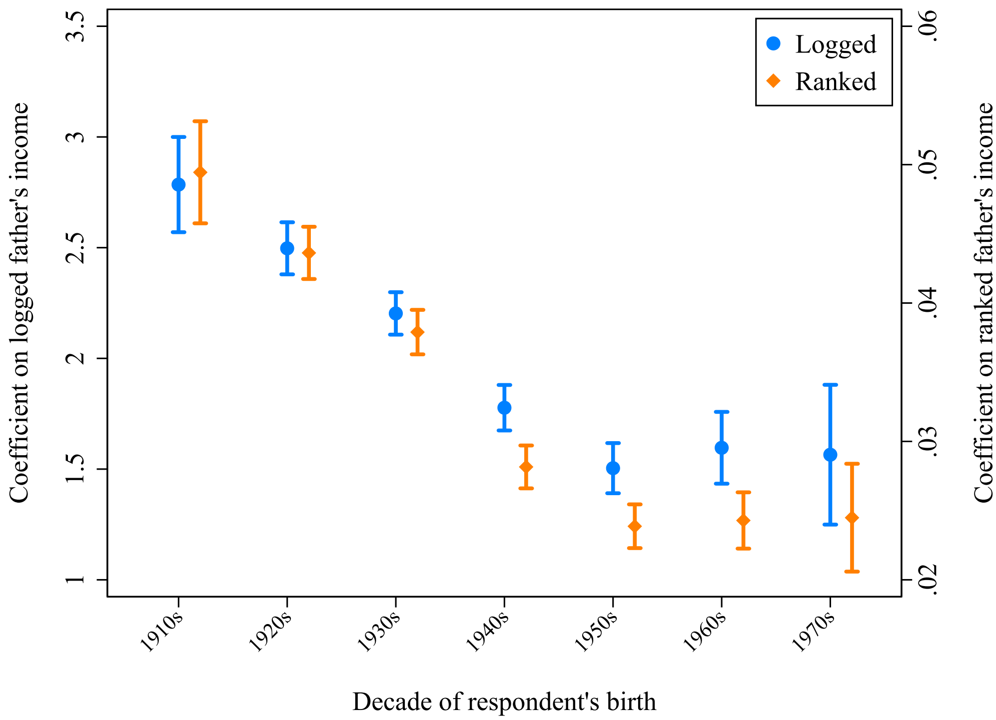
(b) Rank-rank relationships



Sources: Data come from 14 different surveys, described in Section 2 and in further detail in Appendix B.

Notes: The IGE and rank-rank are based on the baseline sample of respondents age 30–50. We use sample weights where provided and further re-weight each birth cohort (i.e., decade) so that they have representative $race \times sex$ shares.

Figure 5: Weakening relationship between respondent's educational attainment and father's income, by birth cohort

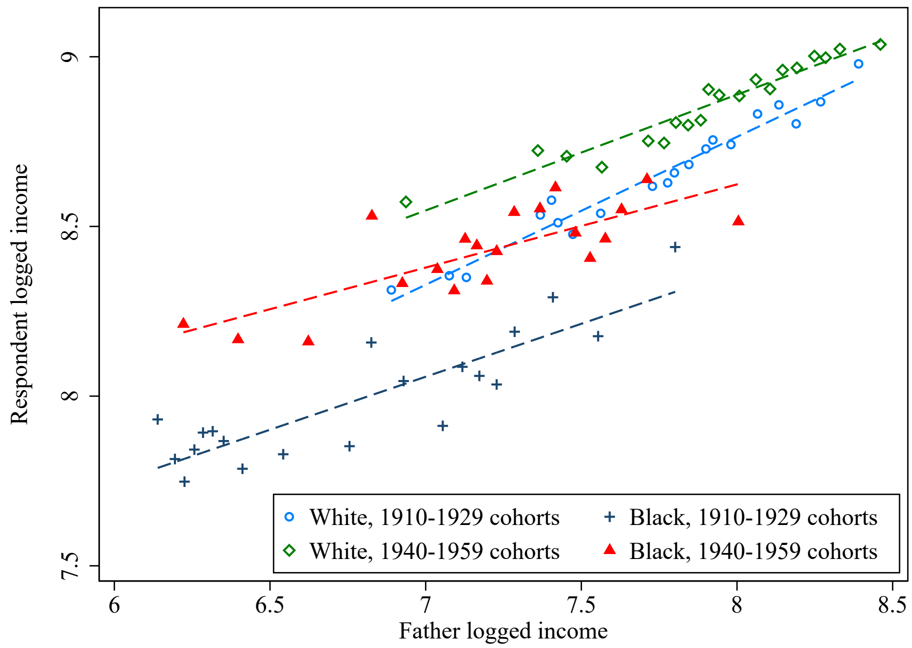


Sources: This figure combines 14 different surveys, which are described in Section 2 and in further detail in Appendix B.

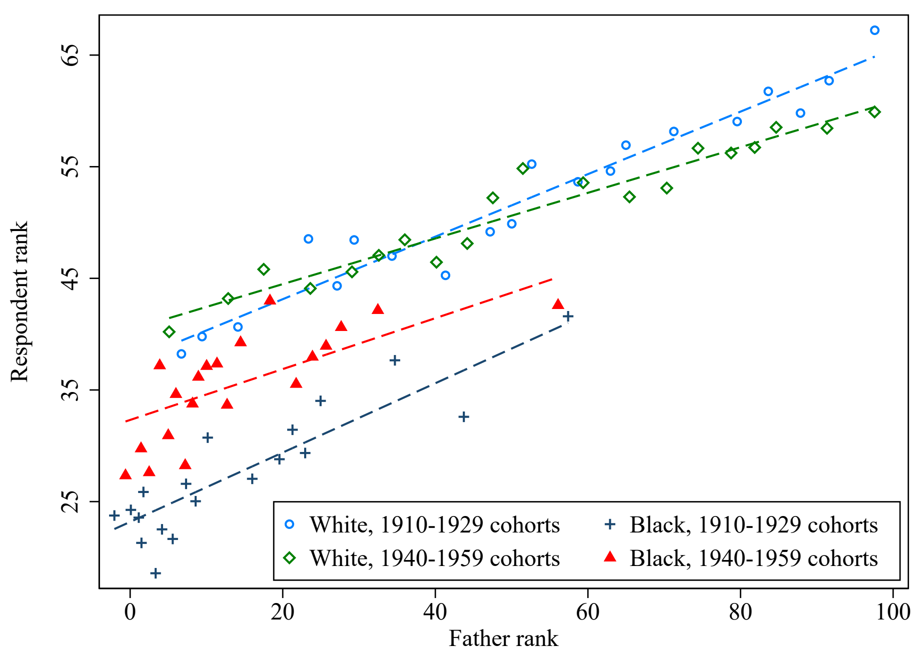
Notes: The estimates are based on the baseline sample of respondents age 30–50. We use sample weights where provided and further re-weight each birth cohort (i.e., decade) so that they have representative $race \times sex$ shares. We use a respondent's years of schooling as the dependent variable and regress it on logged or ranked father's income score, similar to equations (1) and (2).

Figure 6: Mobility by race, 1910s–1920s versus 1940s–1950s

(a) Intergenerational elasticities



(b) Rank-rank relationships

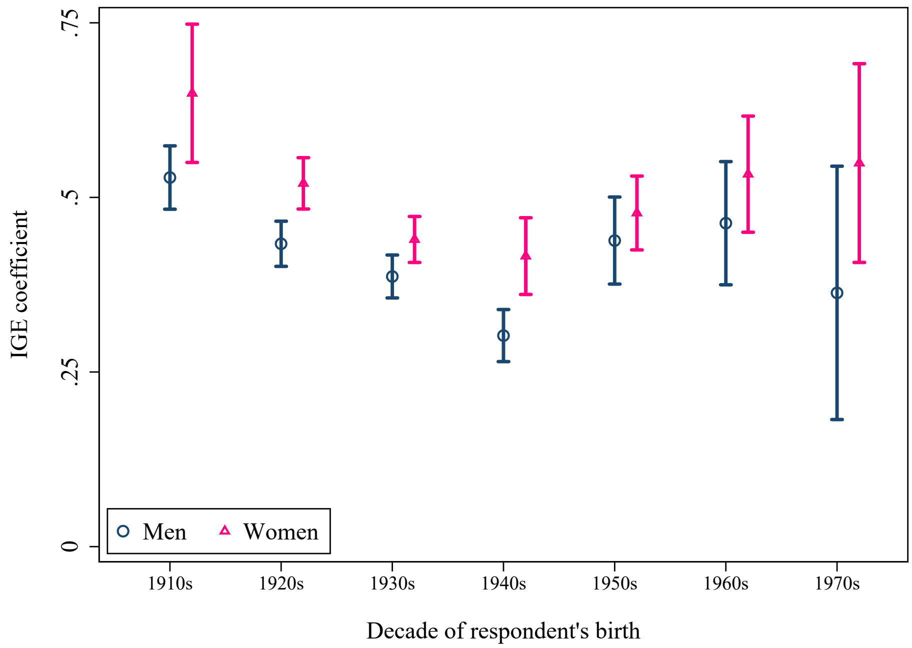


Sources: This figure combines 14 different surveys, which are described in Section 2 and in further detail in Appendix B.

Notes: The IGE and rank-rank are based on the baseline sample of respondents age 30–50. We use sample weights where provided and further re-weight each birth cohort (i.e., decade) so that they have representative *race* × *sex* shares.

Figure 7: IGE and rank-rank measures by birth decade, by sex

(a) Intergenerational elasticity



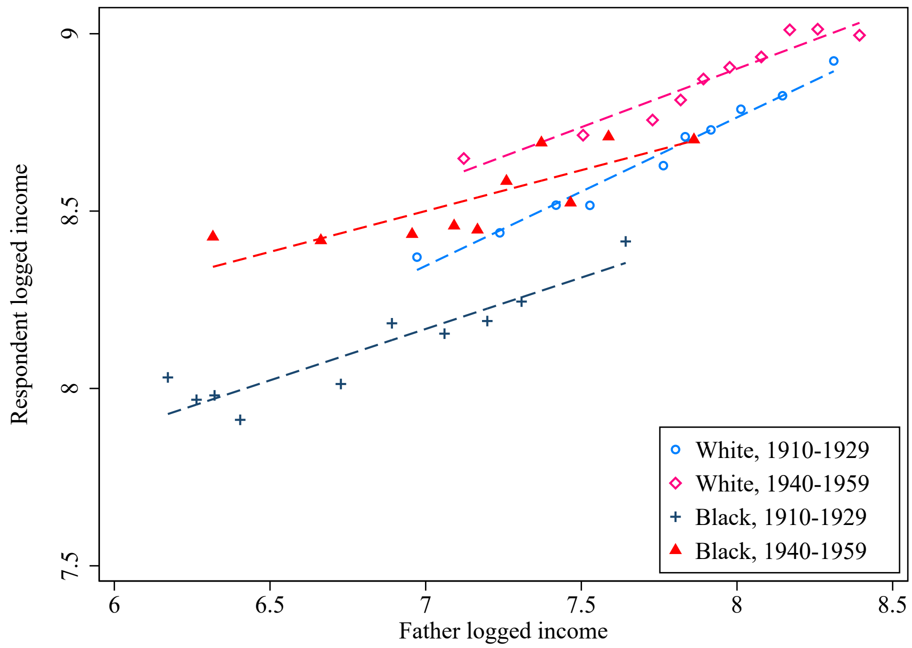
(b) Rank-rank coefficient

Sources: This figure combines 19 different surveys, which are described in Section 2 and in further detail in Appendix B.

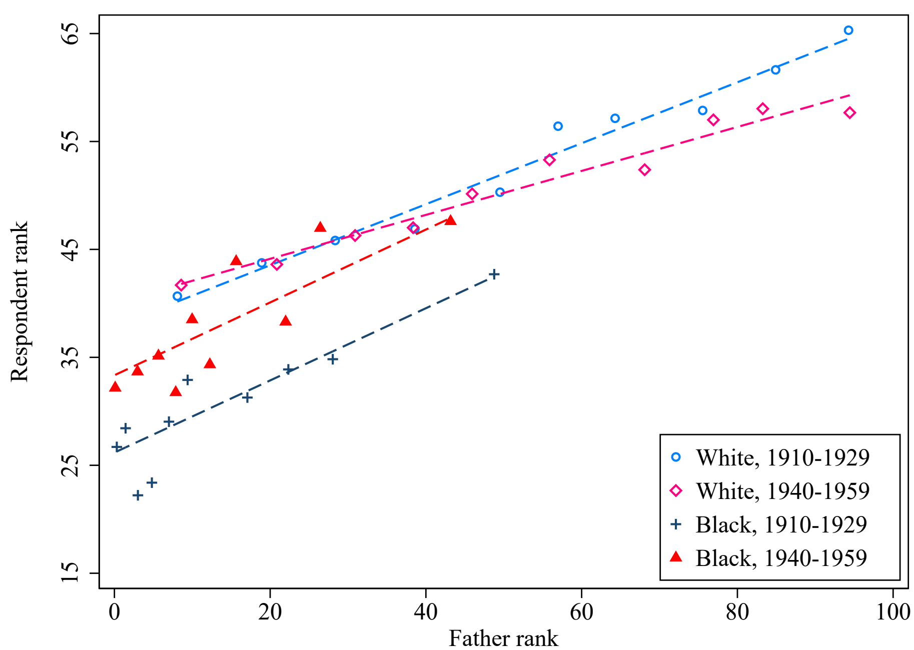
Notes: The IGE and rank-rank are based on the same samples of respondents age 30-50. We use sample weights where provided and further weight each birth cohort so that they have representative *race* $\times$ *sex* shares.

Figure 8: Mobility by race for men, 1910s–1920s versus 1940s-1950s

(a) Intergenerational elasticities



(b) Rank-rank relationships

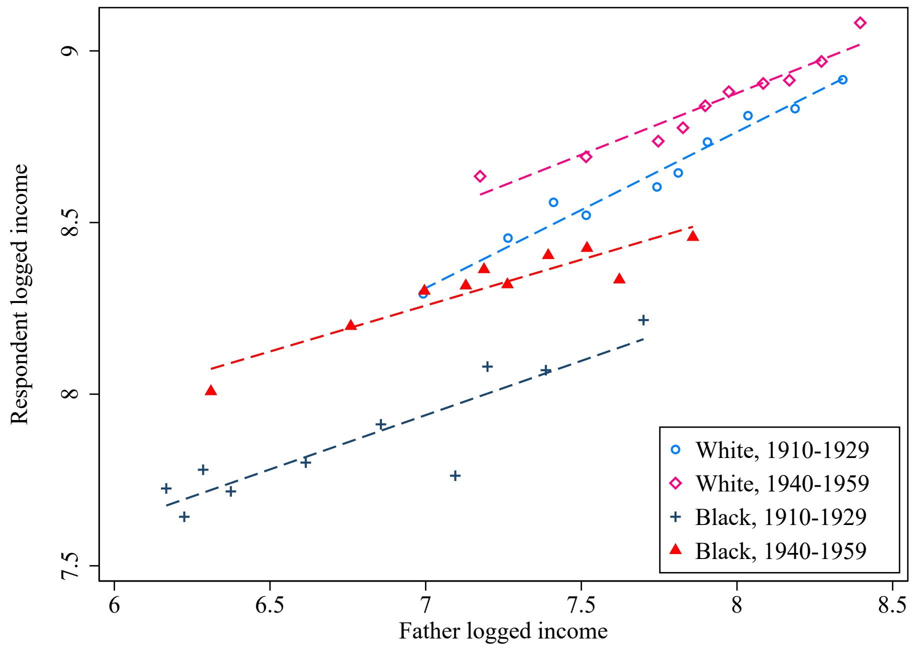


Sources: This figure combines 14 different surveys, which are described in Section 2 and in further detail in Appendix B.

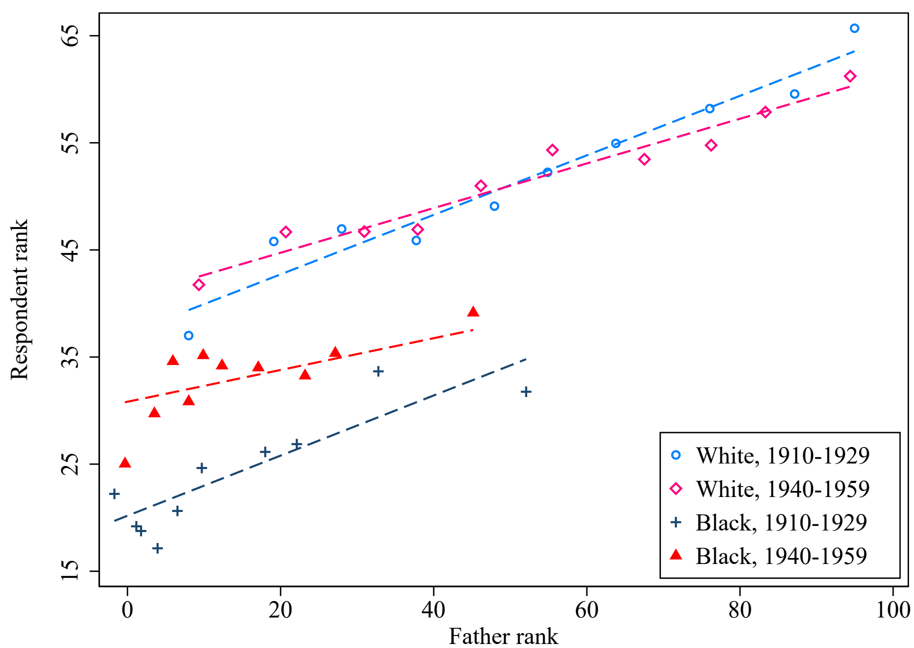
Notes: The IGE and rank-rank are based on the baseline sample of respondents age 30–50. We use sample weights where provided and further re-weight each birth cohort (i.e., decade) so that they have representative *race* × *sex* shares.

Figure 9: Mobility by race for women, 1910s–1920s versus 1940s-1950s

(a) Intergenerational elasticities



(b) Rank-rank relationships

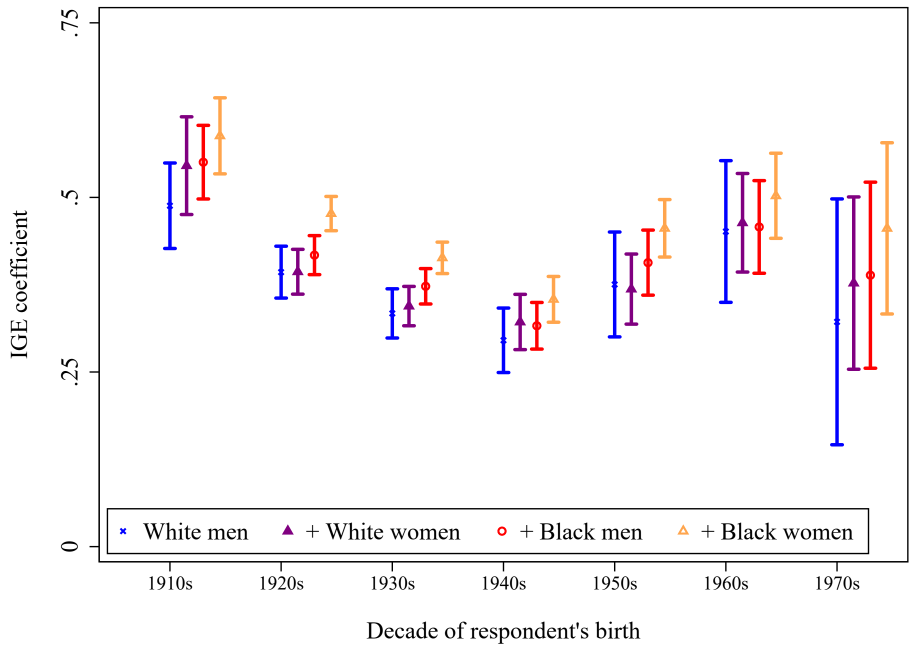


Sources: This figure combines 14 different surveys, which are described in Section 2 and in further detail in Appendix B.

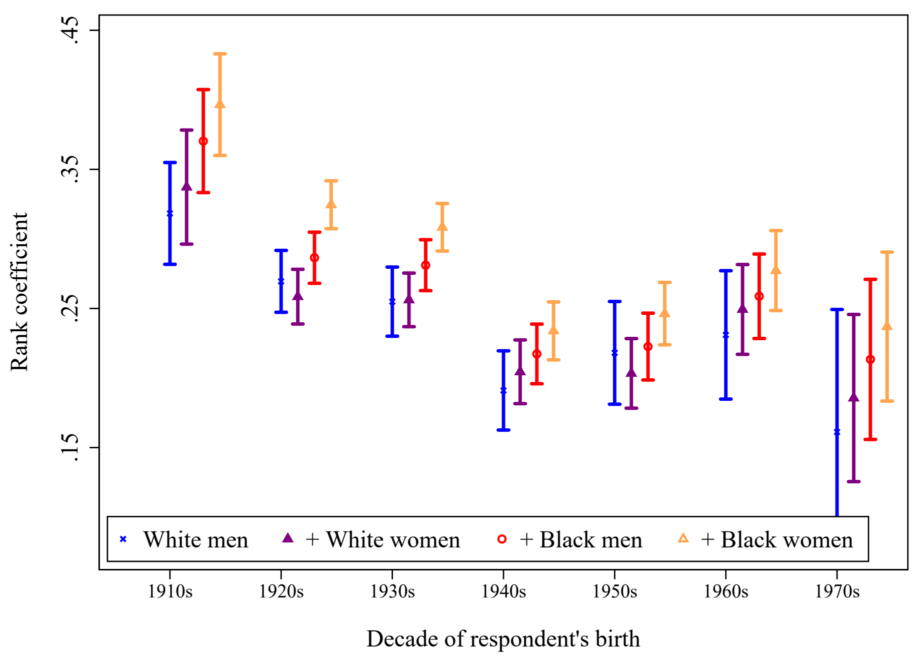
Notes: The IGE and rank-rank are based on the baseline sample of respondents age 30–50. We use sample weights where provided and further re-weight each birth cohort (i.e., decade) so that they have representative *race* $\times$ *sex* shares.

Figure 10: Mobility patterns over the 20th century including under-represented groups

(a) Intergenerational elasticities



(b) Rank-rank relationships

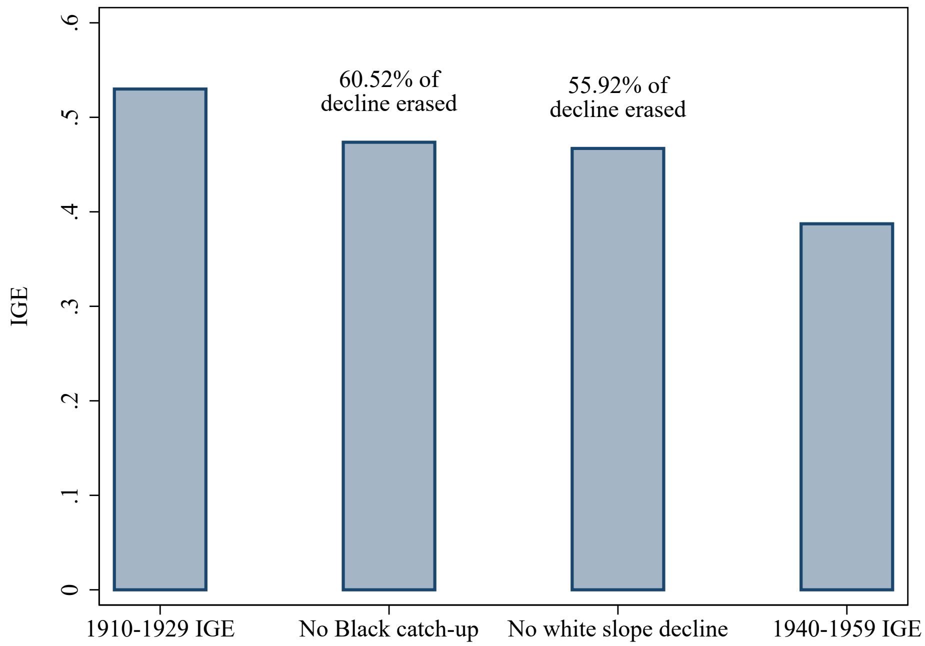


Sources: This figure combines 14 different surveys, which are described in Section 2 and in further detail in Appendix B.

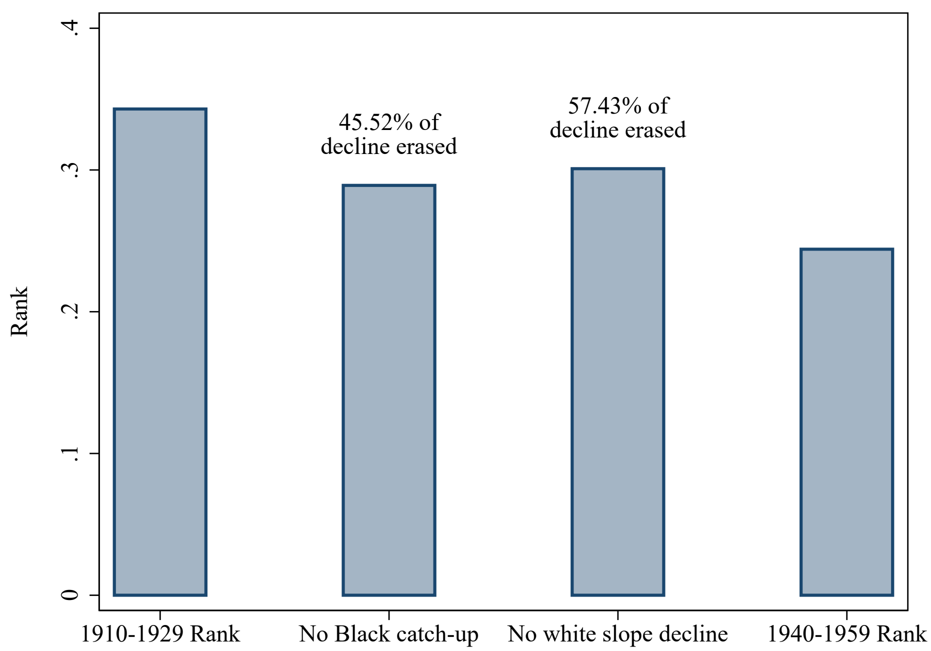
Notes: The IGE and rank-rank are based on the baseline sample of respondents age 30–50. We use sample weights where provided and further re-weight each birth cohort (i.e., decade) so that they have representative $race \times sex$ shares.

Figure 11: Decomposing the fall in mobility from the 1910s-1920s to 1930s-1940s

(a) Decomposition of IGE



(b) Decomposition of Rank-Rank



Sources: This figure combines 14 different surveys, which are described in Section 2 and in further detail in Appendix B.

Notes: This figure shows the contribution of difference components of the decomposition in Section 5 to the change in intergenerational mobility for cohorts born in 1910–1929 to those born in 1940–1959. It shows the contributions of reductions in white-only IGE and the contribution of the between group convergence in levels.

Table 1: Select Review of Intergenerational Mobility Papers

Paper	Cohorts	Income/status proxy: Parent(s)	Income/status proxy: Child	Links	Sample
Ward (2020)	1850–1910	Occ.×Race×Place	Occ.	Match	All ♂
Collins and Wanamaker (2017)	1880–1970	Occ.×Race×Place	Occ.	Match & Retr.	All ♂
Song *et al.* (2020)	1830–1980	Occ.	Occ.	Match & Retr.	White ♂
Long and Ferrie (2013)	1840, 1930	Occ.	Occ.	Match & Retr.	White ♂
Olivetti and Paserman (2015)	1840–1910	Occ.	Occ.	Syn. panel	White ♂ & married ♀
Feigenbaum (2018)	1900	Inc.	Inc.	Match	Iowa ♂
Feigenbaum (2015)	1900–1910	Inc.	Inc.	Match	Urban ♂
Card *et al.* (2018)	1920	Edu.	Edu.	Same household	Repres.
Bowles (1972)	1930	Inc.	Inc.	Retrosp.	CPS ♂
Mazumder (2015)	1950–1970	Inc.	Inc.	Panel data	Repres.
Chetty *et al.* (2014a)	1980–1982	Inc.	Inc.	Claim dep.	Repres.
Chetty *et al.* (2020)	1978–1983	Inc.	Inc.	Claim dep.	Repres.
Our paper	1910–1970	Occ.×Race×South	Inc.	Retrospective	Repres.

Notes: Since many papers do not explicitly consider birth cohorts, the "cohorts" column refers to the birth decade(s) that most of the sample comes from, given the age restrictions used in the paper. In the "Links" column, "match" refers to matching across datasets (e.g., Census matching by name, age and state of birth); "Syn. panel" refers to matching based on characteristics but not individual identity; "Claim dep." refers to matching by whether the parent ever claims the child as a dependent to the IRS; "Retrosp." refers to adult children being asked retrospectively about the characteristics of their parents (e.g., occupation and education).

Table 2: Summary statistics, by birth decade

	1910s	1920s	1930s	1940s	1950s	1960s	1970s
Father demographics:							
Foreign-born	0.22	0.17	0.11	0.05	0.04	0.03	0.06
High school educated	0.17	0.19	0.26	0.45	0.60	0.70	0.80
College educated	0.03	0.04	0.05	0.09	0.16	0.21	0.27
Farming occupation	0.37	0.28	0.23	0.14	0.08	0.04	0.03
Respondent demographics:							
Female	0.12	0.32	0.44	0.42	0.57	0.51	0.57
Age	45.88	41.57	36.96	37.82	37.06	38.80	36.65
Black	0.11	0.13	0.15	0.13	0.14	0.12	0.13
High school educated	0.50	0.61	0.71	0.86	0.91	0.92	0.92
College educated	0.10	0.14	0.16	0.28	0.29	0.33	0.42
Moved regions	0.21	0.22	0.22	0.24	0.22	0.21	0.22
Union member	0.17	0.28	0.27	0.25	0.18	0.14	0.12
Veteran	—	0.76	0.54	0.31	0.11	0.09	0.07
Father income:							
Income score (1950$)	2,190	2,248	2,313	2,582	2,755	2,872	2,901
Missing income	0.12	0.13	0.14	0.20	0.21	0.33	0.32
Father income rank	46.38	46.48	46.15	46.70	46.82	47.40	47.57
Respondent income:							
Family income (1950$)	5,475	6,810	7,295	7,755	7,471	8,301	7,830
Missing income	0.15	0.10	0.06	0.07	0.09	0.09	0.07
Bottom coded	0.09	0.05	0.03	0.03	0.05	0.04	0.04
Top coded	0.07	0.07	0.08	0.18	0.16	0.10	0.08
Family income rank	49.47	48.68	47.33	46.33	45.82	46.84	45.91
Observations	5,307	13,896	12,915	10,395	8,483	4,637	1,664

Notes: This figure combines 14 different surveys, which are described in Section 2 and in further detail in Appendix B. All of the shares in this table are based on the baseline sample of respondents age 30–50 and are unweighted.

Table 3: Summary statistics, comparison to Census

	1910–1929		1930–1949		1950–1969	
	Census	Survey	Census	Survey	Census	Survey
Panel A: White Men						
Share of Men	0.90	0.91	0.90	0.90	0.87	0.88
Age	39.51	43.28	38.69	36.83	40.59	37.63
High school graduate	0.51	0.61	0.81	0.80	0.92	0.92
College graduate	0.12	0.15	0.27	0.26	0.30	0.33
Southern born/grew up	0.30	0.29	0.31	0.31	0.28	0.27
Resides in the South	0.28	0.28	0.33	0.31	0.34	0.31
Married	0.87	0.90	0.81	0.84	0.68	0.64
Widowed	0.01	0.01	0.00	0.01	0.00	0.01
Family income, 1950$	6,124	6,758	7,712	8,040	8,519	8,356
Respondent rank	52.57	51.52	53.20	52.20	52.20	52.64
Observations	195,091	12,856	214,612	11,984	297,783	5,208
Panel B: Black Men						
Share of Men	0.10	0.09	0.10	0.10	0.13	0.12
Age	39.41	44.52	38.54	37.18	40.13	37.32
High school graduate	0.21	0.28	0.62	0.59	0.85	0.85
College graduate	0.03	0.04	0.10	0.11	0.13	0.16
Southern born/grew up	0.86	0.83	0.77	0.73	0.60	0.60
Resides in the South	0.54	0.54	0.51	0.56	0.57	0.59
Married	0.75	0.81	0.63	0.69	0.50	0.51
Widowed	0.02	0.02	0.01	0.02	0.01	0.02
Family income, 1950$	3,817	4,321	5,738	5,980	6,318	6,019
Respondent rank	27.59	31.70	39.19	37.23	38.72	38.53
Observations	21,002	1,301	24,293	1,362	38,206	712
Panel C: White Women						
Share of Women	0.89	0.79	0.88	0.81	0.86	0.85
Age	39.50	41.00	38.74	38.17	40.64	37.83
High school graduate	0.55	0.66	0.81	0.82	0.94	0.93
College graduate	0.07	0.09	0.17	0.19	0.30	0.32
Southern born/grew up	0.30	0.31	0.31	0.31	0.28	0.27
Resides in the South	0.28	0.30	0.32	0.31	0.34	0.31
Married	0.86	0.86	0.79	0.78	0.70	0.65
Widowed	0.03	0.03	0.02	0.02	0.01	0.01
Family income, 1950$	6,033	6,869	7,527	7,615	8,469	7,933
Respondent rank	51.06	51.72	51.45	50.60	51.75	50.67
Observations	201,503	3,981	217,061	8,033	302,610	6,138
Panel D: Black Women						
Share of Women	0.11	0.21	0.12	0.19	0.14	0.15
Age	39.27	40.90	38.70	37.17	40.18	37.21
High school graduate	0.25	0.32	0.63	0.58	0.88	0.86
College graduate	0.04	0.05	0.09	0.11	0.17	0.19
Southern born/grew up	0.86	0.84	0.77	0.72	0.61	0.60
Resides in the South	0.55	0.60	0.51	0.56	0.58	0.60
Married	0.66	0.64	0.50	0.53	0.40	0.37
Widowed	0.08	0.09	0.06	0.06	0.03	0.03
Family income, 1950$	3,560	3,597	4,962	4,743	5,706	5,057
Respondent rank	23.72	23.82	32.87	28.95	34.65	32.61
Observations	24,081	1,065	29,808	1,931	45,166	1,062

Notes: All of the survey shares are based on the baseline sample of respondents age 30–50 and are unweighted. We use the 1960, 1980, and 2000 Census from IPUMS and keep individuals born in the same years as survey respondents.

Table 4: IGE and rank coefficient, by birth cohort

(a) Intergenerational elasticity

	(1) 1910	(2) 1920	(3) 1930	(4) 1940	(5) 1950	(6) 1960	(7) 1970
IGE coefficient	0.588 [0.028]	0.477 [0.013]	0.413 [0.011]	0.354 [0.017]	0.456 [0.021]	0.502 [0.031]	0.456 [0.063]
Observations	5,307	13,896	12,915	10,395	8,483	4,637	1,664

(b) Rank-rank coefficient

	(1) 1910	(2) 1920	(3) 1930	(4) 1940	(5) 1950	(6) 1960	(7) 1970
Rank coefficient	0.397 [0.019]	0.325 [0.009]	0.308 [0.009]	0.234 [0.011]	0.246 [0.011]	0.277 [0.015]	0.237 [0.027]
Observations	5,307	13,896	12,915	10,395	8,483	4,637	1,664

NOTES: The IGE and rank-rank estimates—calculated using equations (1) and (2), respectively—are based on the baseline sample of respondents age 30–50. We use sample weights where provided and further re-weight each birth cohort (i.e., decade) so that they have representative *race* × *sex* shares.

Mobility for All: Appendix Materials

A Additional figures and tables referenced in the text **44**

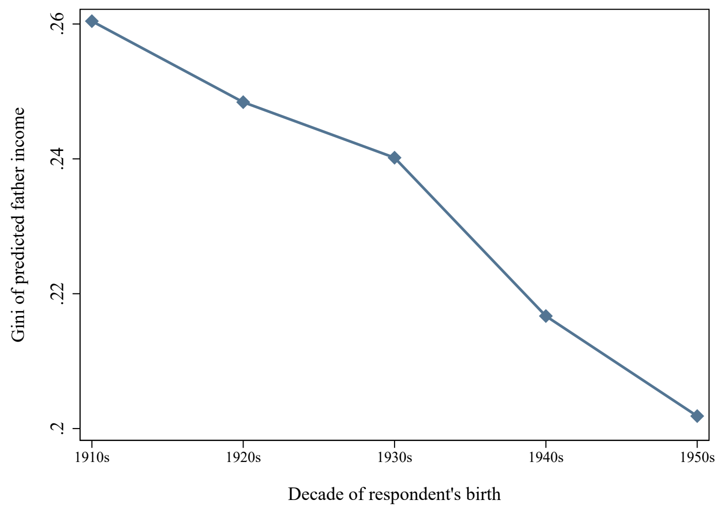
B Additional Detail on Data Sources **57**

List of Appendix Figures

1 Declining Gini coefficient of predicted father's income, by birth cohort . 44
2 Bin-scatter depictions of the weakening relationship between respondent education and father's income 45
3 Mobility by birth decade, adjusting father income score for education . . 46
4 Mobility by birth decade, various adjustments for farmers 47
5 Mobility by birth decade, various income score adjustments 48
6 Mobility by birth decade, robustness to weights 49
7 Mobility measures by birth decade, by sex (restricted to common surveys) 50
8 Widowhood by race for women, 1910s–1920s versus 1940s–1950s 51
9 The share of Black women who are widows declines over time 52

A Additional figures and tables referenced in the text

Appendix Figure 1: Declining Gini coefficient of predicted father's income, by birth cohort

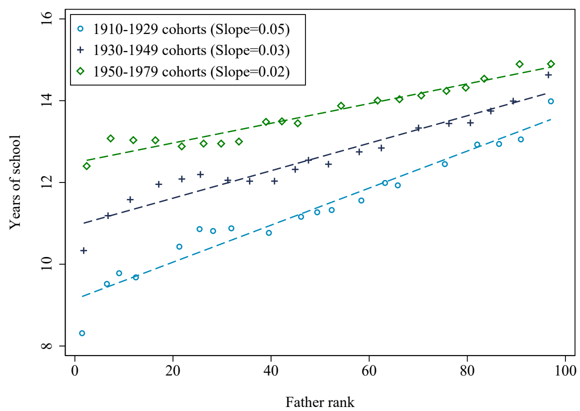


Sources: This figure combines 14 different surveys, which are described in Section 2 and in further detail in Appendix B.

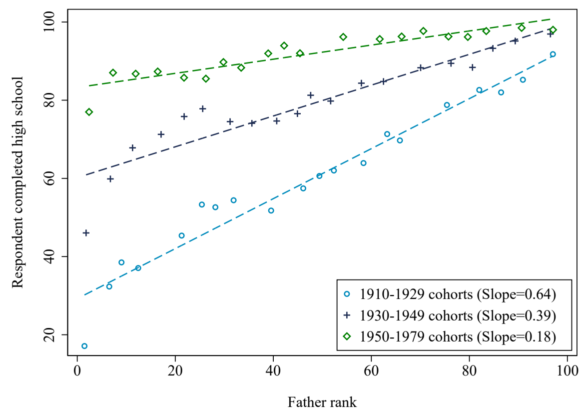
Notes: This figure plots the Gini coefficient of fathers' income scores, separately for each birth cohort in the sample.

Appendix Figure 2: Bin-scatter depictions of the weakening relationship between respondent education and father's income

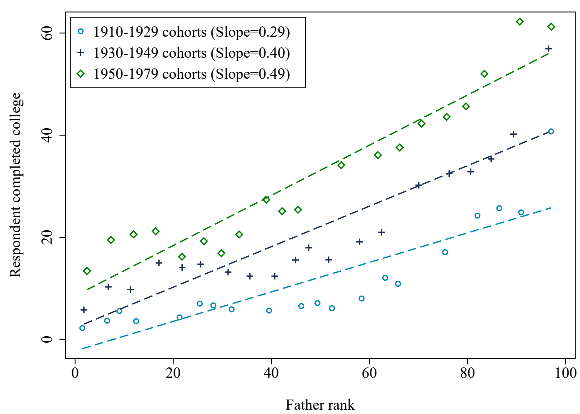
(a) Years of schooling



(b) High school completion



(c) College completion

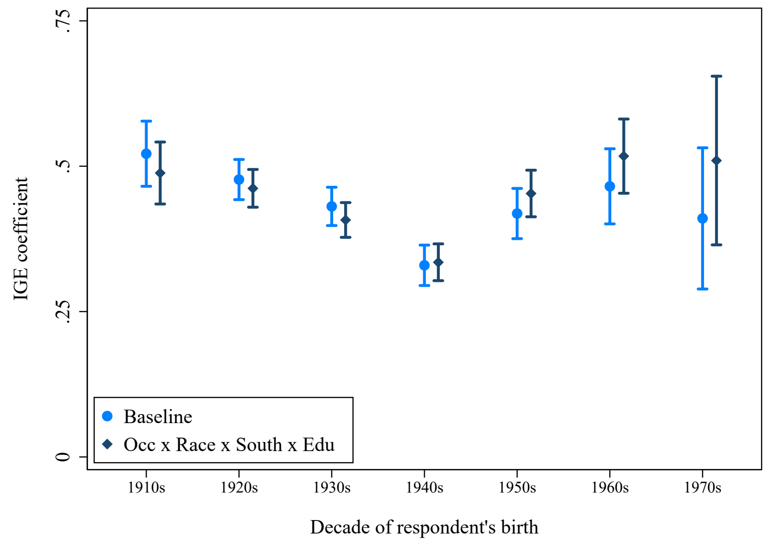


Sources: Data come from 14 different surveys, described in Section 2 and in further detail in Appendix B.

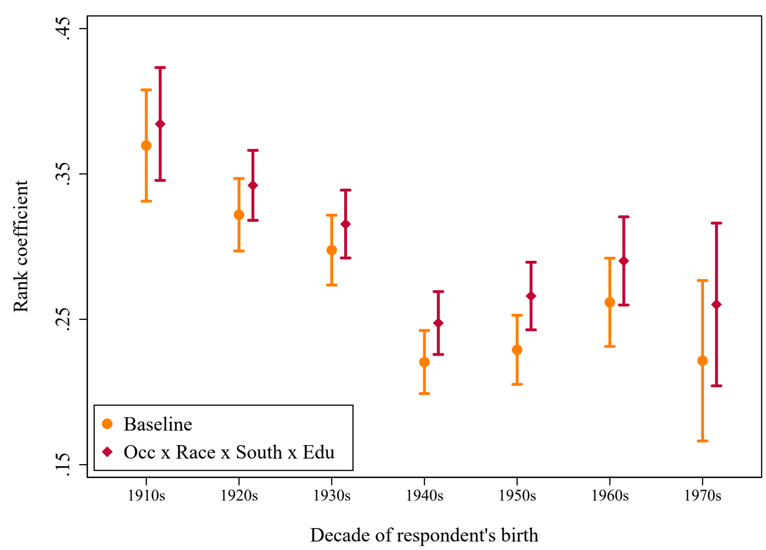
Notes: The estimates are based on the baseline sample of respondents age 30–50. We use sample weights where provided and further re-weight each birth cohort (i.e., decade) so that they have representative *race* × *sex* shares. Further details on the construction of education variables are available in Appendix B.

Appendix Figure 3: Mobility by birth decade, adjusting father income score for education

(a) Intergenerational elasticity



(b) Rank-rank coefficient

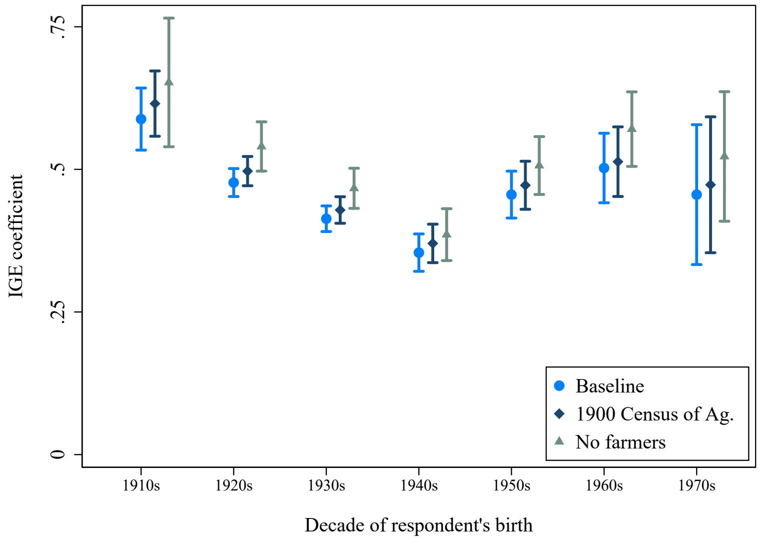


Sources: This figure combines 14 different surveys, which are described in Section 2 and in further detail in Appendix B.

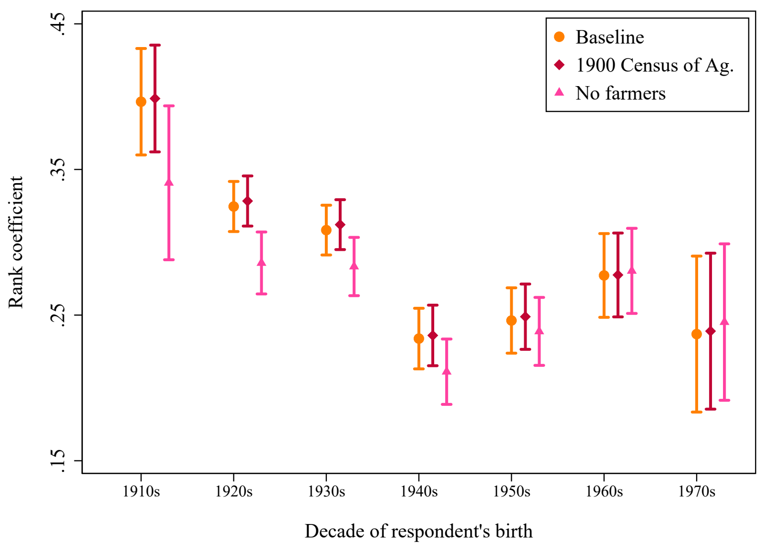
Notes: The IGE and rank-rank are based on the baseline sample of respondents aged 30–50 who provided information on their fathers' education. We use sample weights where provided and further re-weight each birth cohort (i.e., decade) in this sample so that they have representative *race* × *sex* shares.

Appendix Figure 4: Mobility by birth decade, various adjustments for farmers

(a) Intergenerational elasticity



(b) Rank-rank coefficient



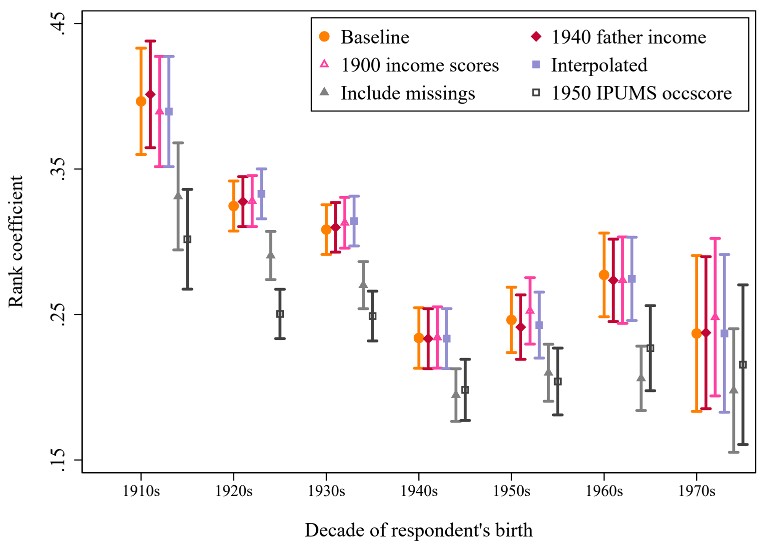
Sources: This figure combines 14 different surveys, which are described in Section 2 and in further detail in Appendix B.

Notes: The first series is the baseline series and is presented for the sake of comparison. The second series uses the 1900 Census of Agriculture to estimate income for fathers who are farmers. In both the first and second series, the IGE and rank-rank are based on the baseline sample of respondents aged 30–50. We use sample weights where provided and further re-weight each birth cohort (i.e., decade) in this sample so that they have representative *race* × *sex* shares. The third series drops all respondents whose fathers work in agricultural occupations; the remaining respondents are re-ranked in this subsample and weights are constructed so that each birth cohort in this subsample has representative *race* × *sex* shares.

Appendix Figure 5: Mobility by birth decade, various income score adjustments

(a) Intergenerational elasticity

(b) Rank-rank coefficient

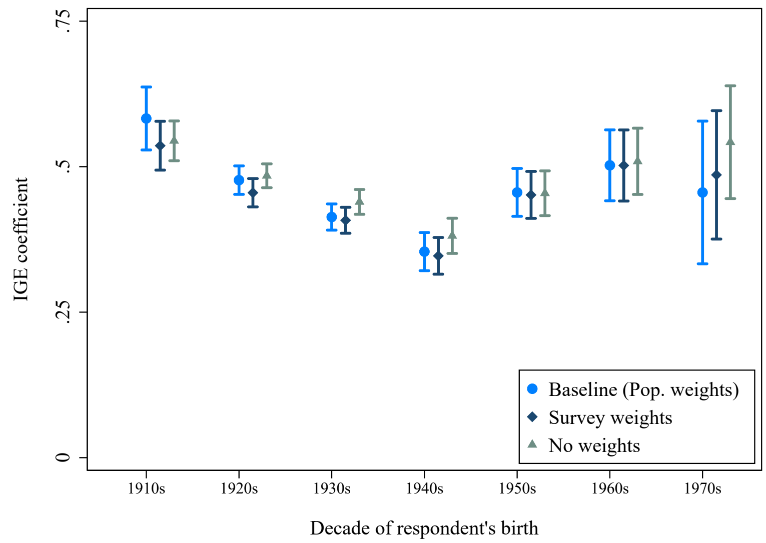


Sources: This figure combines 14 different surveys, which are described in Section 2 and in further detail in Appendix B.

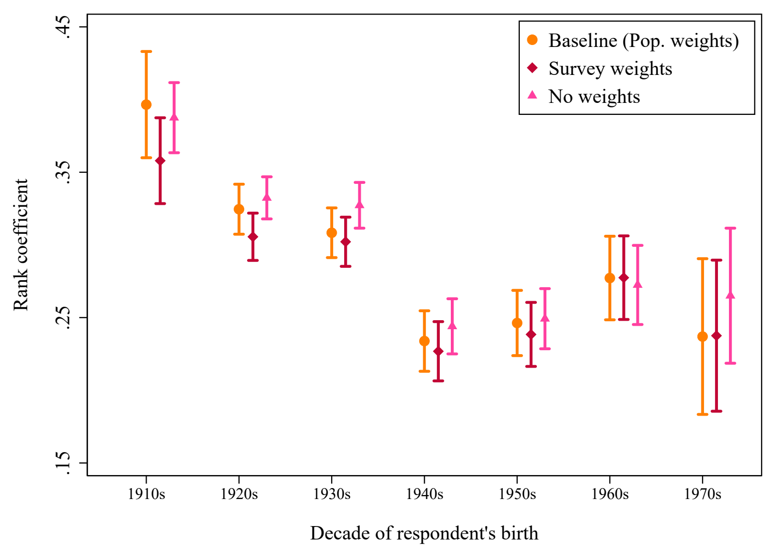
Notes: With the exception of the series that includes missing father income, all estimates are based on the baseline sample of respondents aged 30–50. We use sample weights where provided and further re-weight each birth cohort in this sample so that they have representative $race \times sex$ shares. "1940 father income" refers to using median (personal) income for fathers with a certain occupation, race, and Southern residence. "1900 income scores" refers to using the 1900 Census of Agriculture and the 1901 Cost of Living Survey to construct income scores. "Interpolated" means using the most contemporaneous dataset possible for each cohort to predict father income. "1950 IPUMS occscore" refers to using the *occscore* variable from IPUMS. "Include missings" in the second panel refers to giving respondents with unavailable father occupation an income of zero prior to ranking. This last series thus includes more respondents than in our baseline sample, so we re-weight each birth cohort in this lager sample so that they have representative $race \times sex$ shares.

Appendix Figure 6: Mobility by birth decade, robustness to weights

(a) Intergenerational elasticity



(b) Rank-rank coefficient

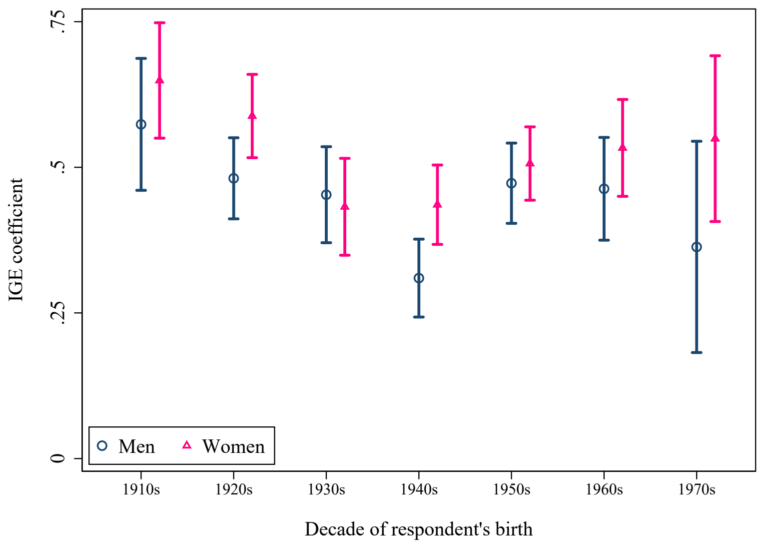


Sources: This figure combines 14 different surveys, which are described in Section 2 and in further detail in Appendix B.

Notes: The IGE and rank-rank are based on the baseline sample of respondents aged 30–50 who provided information on their fathers' education. In the first series, we re-weight survey weights so that each birth cohort has representative *race* × *sex* shares. The second series simply uses the provided survey weights (or a weight of one when no survey weight is available). The estimates from the third series are unweighted.

Appendix Figure 7: Mobility measures by birth decade, by sex (restricted to common surveys)

(a) Intergenerational elasticity



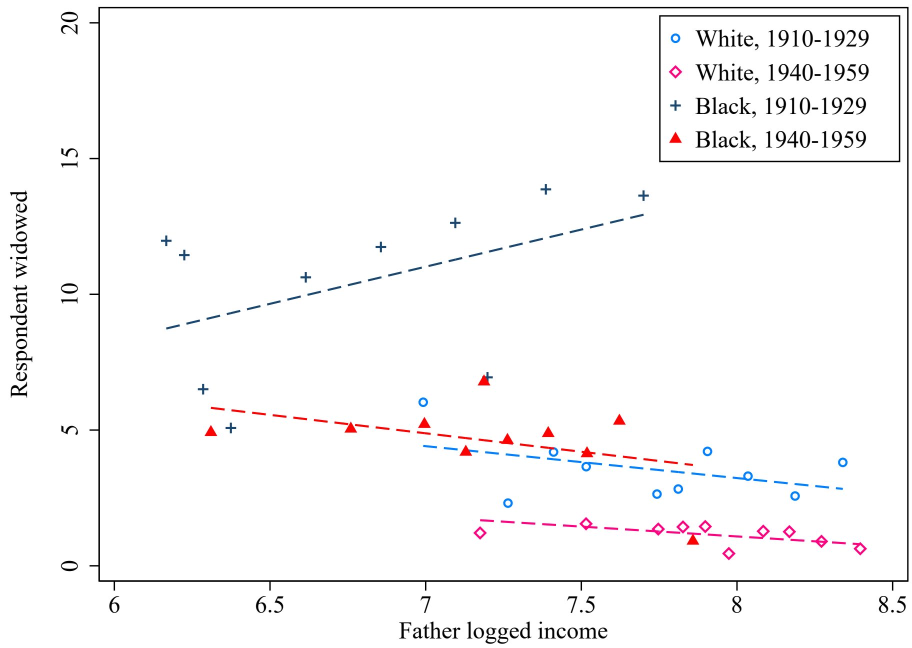
(b) Rank-rank coefficient

Sources: This figure combines 19 different surveys, which are described in Section 2 and in further detail in Appendix B.

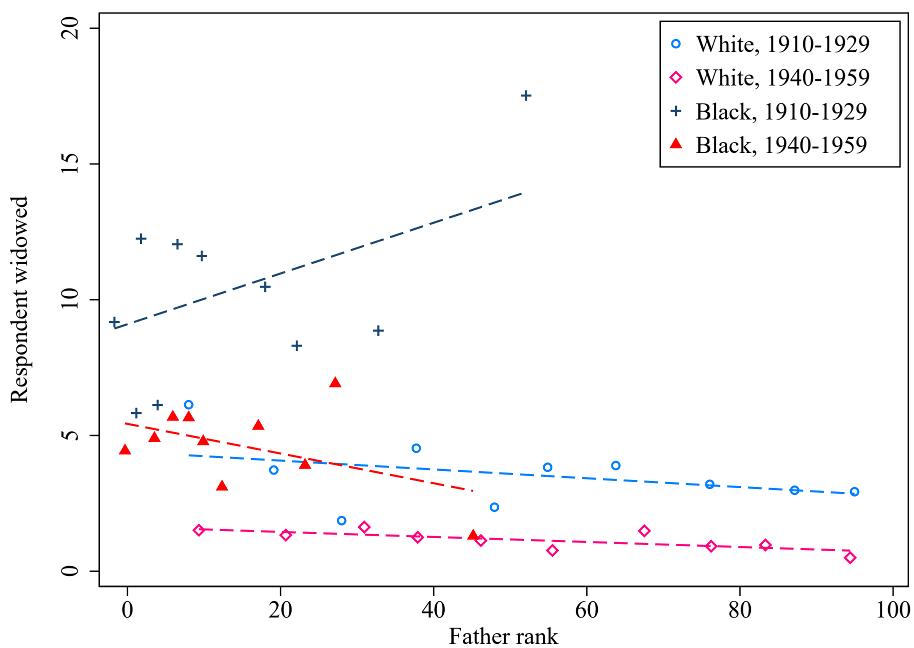
Notes: This figure is identical to Figure 7 except that in this figure, we use only samples that income both men and women. The IGE and rank-rank are based on the same samples of respondents age 30-50. We use sample weights where provided and further weight each birth cohort so that they have representative $race \times sex$ shares.

Appendix Figure 8: Widowhood by race for women, 1910s–1920s versus 1940s–1950s

(a) Widowhood by parental-income estimate



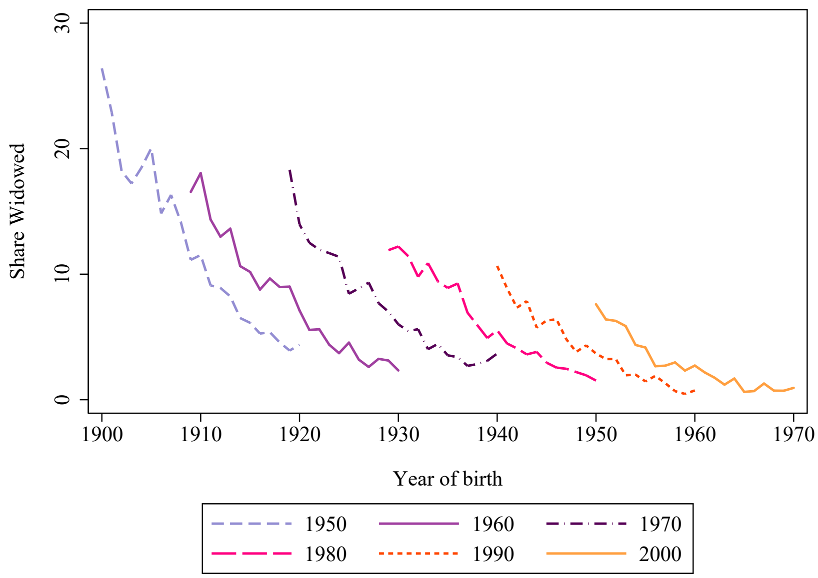
(b) Widowhood by parental-income rank



Sources: This figure combines 14 different surveys, which are described in Section 2 and in further detail in Appendix B.

Notes: These estimates are based on the baseline sample of respondents age 30–50. We use sample weights where provided and further re-weight each birth cohort (i.e., decade) so that they have representative $race \times sex$ shares.

Appendix Figure 9: The share of Black women who are widows declines over time



Sources: 1950–2000 Census data from IPUMS.

Notes: The sample used is Black women ages 30–50 born in the United States. Each line uses a different Census to calculate the share of individuals born in a specific year that are widowed.

Appendix Table 1: Summary Statistics in Panel Study of Income Dynamics

	1968 Men	All Fathers	All Sons	+1 Year of Son Income	Father's Income: 1 year	5 years	10 years
Son characteristics:							
Age	—	—	—	38.09	37.98	37.93	37.94
Black	0.09	0.11	0.10	0.10	0.09	0.09	0.08
HS educated	0.56	0.76	0.86	0.92	0.92	0.93	0.94
College edu.	0.16	0.23	0.29	0.34	0.35	0.34	0.37
Family income	—	—	—	83,354	86,073	88,568	93,252
Father characteristics:							
HS educated	—	—	—	—	0.80	0.81	0.85
College edu.	—	—	—	—	0.28	0.29	0.32
Farm occupation	—	—	—	—	0.05	0.04	0.04
Observations	1,756	4,296	3,261	3,238	2,825	2,403	1,839

Notes: This table uses the Panel Study of Income Dynamics dataset from 1968 through 2015. The first column considers all men ages 30–50 in the 1968 wave of the survey. The second and third columns consider all men identified as fathers and sons, respectively, using the Family Identification Mapping System (FIMS). For both columns 2 and 3, we use the earliest observation in the age range 30–50 for which the individual has a non-zero weight. The fourth column only considers sons with at least one year of available income between ages 30–50. Columns 5, 6, and 7 then restrict the sample in column 4 to those individuals with one, five, or ten years of data on father's occupation during ages 30–50.

Appendix Table 2: Differences in Income Scores, by Respondent Sex and Birth Cohort

(a) Logged Father's Income

	(1) 1910	(2) 1920	(3) 1930	(4) 1940	(5) 1950	(6) 1960	(7) 1970
Female	-0.006	0.023	-0.025	0.004	-0.006	-0.023*	0.001
	[0.034]	[0.022]	[0.020]	[0.013]	[0.011]	[0.013]	[0.023]
Observations	5,307	13,896	12,915	10,395	8,483	4,637	1,664

(b) Ranked Father's Income

	(1) 1910	(2) 1920	(3) 1930	(4) 1940	(5) 1950	(6) 1960	(7) 1970
Female	-0.314	1.412	-1.828	0.646	-0.406	-1.502*	-0.828
	[1.838]	[1.268]	[1.132]	[0.808]	[0.758]	[0.899]	[1.589]
Observations	5,307	13,896	12,915	10,395	8,483	4,637	1,664

NOTES: This table uses our baseline sample ages 30–50 to regress logged and ranked father's income on an indicator variable for whether a respondent is female. Age and age squared controls are included in the first panel, and survey-year fixed effects are included in both panels.

Appendix Table 3: Top Five Occupations Reported by Male and Female Respondents, by Birth Cohort

Birth Cohort	Male Respondents	Share of male sample	Female Respondents	Share of female sample
1910s	1. Farm operator	0.33	1. Farm operator	0.36
	2. Craftsman (skilled)	0.15	2. Craftsman (skilled)	0.16
	3. Craftsman (semi-skilled)	0.15	3. Craftsman (semi-skilled)	0.12
	4. Unskilled laborer (non-farm)	0.07	4. Unskilled laborer (non-farm)	0.08
	5. Businessman (self-employed)	0.05	5. Businessman (self-employed)	0.07
1920s	1. Farm operator	0.24	1. Farm operator	0.25
	2. Craftsman (skilled)	0.17	2. Craftsman (semi-skilled)	0.19
	3. Craftsman (semi-skilled)	0.17	3. Craftsman (skilled)	0.14
	4. Unskilled laborer (non-farm)	0.07	4. Businessman (not self-employed)	0.10
	5. Businessman (self-employed)	0.06	5. Unskilled laborer (non-farm)	0.07
1930s	1. Farm operator	0.19	1. Farm operator	0.19
	2. Craftsman (skilled)	0.18	2. Craftsman (semi-skilled)	0.19
	3. Craftsman (semi-skilled)	0.18	3. Craftsman (skilled)	0.17
	4. Unskilled laborer (non-farm)	0.07	4. Businessman (not self-employed)	0.11
	5. Businessman (self-employed)	0.06	5. Unskilled laborer (non-farm)	0.07
1940s	1. Craftsman (skilled)	0.20	1. Craftsman (skilled)	0.19
	2. Craftsman (semi-skilled)	0.17	2. Craftsman (semi-skilled)	0.18
	3. Farm operator	0.11	3. Businessman (not self-employed)	0.12
	4. Businessman (not self-employed)	0.11	4. Farm operator	0.09
	5. Unskilled laborer (non-farm)	0.06	5. Unskilled laborer (non-farm)	0.07
1950s	1. Craftsman (skilled)	0.20	1. Craftsman (skilled)	0.19
	2. Craftsman (semi-skilled)	0.16	2. Craftsman (semi-skilled)	0.17
	3. Businessman (not self-employed)	0.13	3. Businessman (not self-employed)	0.12
	4. Farm operator	0.06	4. Unskilled laborer (non-farm)	0.06
	5. Unskilled laborer (non-farm)	0.06	5. Farm operator	0.06
1960s	1. Craftsman (skilled)	0.21	1. Craftsman (skilled)	0.20
	2. Craftsman (semi-skilled)	0.16	2. Craftsman (semi-skilled)	0.16
	3. Businessman (not self-employed)	0.13	3. Businessman (not self-employed)	0.13
	4. Unskilled laborer (non-farm)	0.05	4. Unskilled laborer (non-farm)	0.05
	5. Protective service officer	0.05	5. Protective service officer	0.05
1970s	1. Craftsman (skilled)	0.16	1. Craftsman (skilled)	0.20
	2. Craftsman (semi-skilled)	0.13	2. Craftsman (semi-skilled)	0.15
	3. Businessman (not self-employed)	0.13	3. Businessman (not self-employed)	0.14
	4. Businessman (self-employed)	0.08	4. Protective service officer	0.05
	5. Protective service officer	0.08	5. Unskilled laborer (non-farm)	0.05

Notes: All shares are weighted using population-adjusted survey weights. The fourth and fifth most common father occupations were reported the same number of times by both male and female respondents born in the 1970s cohort; the tied occupations are ranked alphabetically in the table.

Appendix Table 4: Occupations of Census Fathers and Survey Respondents' Fathers, by Birth Cohort

	1910–1919		1920–1929		1930–1939		1940–1949		1950–1959		1960–1969	
	Census (1930)	Survey	Census (1940)	Survey	Census (1950)	Survey	Census (1960)	Survey	Census (1970)	Survey	Census (1980)	Survey
Coarsened Occupations												
Accountants and auditors	0.52	0.53	0.62	0.49	0.88	0.58	0.99	0.69	1.11	1.08	1.05	0.87
Clergymen	0.41	0.65	0.41	0.70	0.40	0.61	0.44	0.64	0.54	0.91	0.57	0.61
Public-school teachers	0.48	0.67	0.83	0.50	0.85	0.57	1.14	0.85	2.06	1.30	2.83	1.98
Dentists	0.24	0.29	0.27	0.16	0.21	0.23	0.21	0.15	0.23	0.30	0.28	0.23
Physicians and surgeons	0.40	0.65	0.49	0.27	0.59	0.41	0.62	0.50	0.70	0.71	0.79	0.62
Engineers	0.72	1.07	0.70	0.86	1.56	1.05	2.58	2.23	3.61	3.22	2.96	4.02
Lawyers and judges	0.45	0.29	0.56	0.41	0.51	0.47	0.51	0.52	0.67	0.64	0.92	0.91
Social and welfare workers	0.03	0.02	0.04	0.04	0.08	0.09	0.12	0.06	0.19	0.12	0.29	0.22
Nurses (trained or student)	0.01	0.01	0.02	0.08	0.01	0.06	0.03	0.08	0.15	—	0.24	0.05
Other professional and technical	0.74	0.43	1.02	0.67	1.61	0.98	2.42	1.85	4.60	3.37	4.38	3.92
Semi-professional	0.88	0.88	0.91	0.64	1.49	1.16	2.35	1.81	3.08	2.19	3.55	2.38
Businessmen (self-employed)	6.35	6.29	4.73	4.15	6.52	3.61	4.29	3.32	3.30	3.65	3.73	4.37
Businessmen (not self-employed)	5.24	4.58	5.50	7.20	6.18	7.96	8.09	11.63	9.39	12.56	12.31	13.09
Bookkeeper	0.38	0.18	0.41	0.35	0.30	0.38	0.25	0.22	0.44	0.18	0.20	0.19
Stenographers	0.14	0.29	0.12	0.28	0.16	0.22	0.12	0.13	0.17	0.11	0.08	0.08
Other clerical workers	3.41	1.76	4.27	2.99	4.83	3.05	5.28	3.96	5.01	3.70	5.08	3.42
Sales: higher-status	1.41	1.27	1.01	1.05	1.11	0.98	1.52	1.34	2.08	1.52	2.01	1.97
Sales: inside sales	4.33	1.92	6.96	2.23	4.85	2.70	5.09	3.48	4.81	3.63	4.06	3.68
Sales: lower-status	0.19	0.37	0.13	0.18	0.05	0.19	0.05	0.06	0.08	0.08	0.10	0.06
Foremen	2.14	1.72	1.89	2.15	2.62	3.03	3.30	3.58	3.92	4.04	4.55	3.76
Craftsmen (skilled)	17.17	15.48	15.58	15.76	18.16	17.55	19.03	19.81	19.01	19.90	17.13	20.35
Craftsmen (semi-skilled)	15.07	13.44	17.65	17.81	20.41	18.43	20.46	17.42	18.90	16.57	16.97	15.95
Protective service officers	1.32	1.22	1.45	2.06	2.35	2.12	3.72	3.07	4.19	4.17	4.46	5.05
Private household workers	0.09	0.15	0.25	1.18	0.04	0.77	0.03	0.20	0.03	0.01	0.01	—
Other service workers	2.44	2.01	2.96	3.05	2.54	3.35	2.41	3.25	3.25	2.82	3.08	2.92
Farm laborers	3.37	1.89	4.24	2.98	2.13	3.49	1.37	2.85	1.03	1.55	0.79	0.93
Unskilled non-farm laborers	10.95	7.56	12.19	7.47	6.27	7.18	5.47	6.17	4.47	5.80	4.44	5.23
Farm operators	20.18	34.40	14.79	24.28	10.73	18.79	5.01	10.14	2.55	5.86	1.97	3.14

Notes: Census shares are weighted using provided weights. For survey estimates, we re-weight survey weights so that each birth cohort has representative *race* × *sex* shares. All Censuses and surveys use the same sample: Black and white fathers aged 30–50.

B Additional Detail on Data Sources

B.1 Income Scores for Fathers

B.1.1 Coarsened Occupations

Across all surveys, we harmonize occupations into 28 categories, corresponding to the main occupations in the American National Election Survey. The ANES occupation we use are:

- Accountants and auditors
- Clergymen
- Teachers
- Dentists
- Physicians and surgeons
- Engineers
- Lawyers and judges
- Social and welfare workers
- Nurses
- Other professional and technical occupations
- Semi-professional occupations
- Self-employed businessmen, managers, and officials
- Businessmen, managers, and officials
- Bookkeepers
- Stenographers, typists, and secretaries
- Other clerical workers
- Higher-status sales workers in "outside" sales
- Inside sales workers (e.g., salesmen, clerks)
- Lower-status sales workers in "outside" sales (e.g., peddlers, newsboys)
- Foremen
- Skilled craftsmen and kindred workers
- Semi-skilled operatives and kindred workers
- Protective service workers
- Private household workers
- Other service workers
- Farm laborers
- Non-farm laborers
- Farm operators

B.1.2 Census-based Income Scores

Because we want our baseline income scores to approximate the income of the fathers' generation, we restrict the decennial Census to individuals who resemble the survey respondents' fathers (Ruggles *et al.*, 2021). In particular, we restrict the sample to men who are between the ages of 30 and 50, whose race was recorded as either white or Black, and who had a child younger than 18 present in the household. For the 1950 Census, we also restrict the sample to men who were sample-line individuals (i.e., who were asked questions about income). We then use a crosswalk that maps Census occupations into our 28 coarsened occupations.

Next, we calculate the average income in each occupation for individuals with certain characteristics. In particular, we take averages by (1) occupation × race × South, and (2) occupation × race × South × education.[10] As described in the main text, the 1940-based occupation × race × South variation serves as our baseline approach for approximating father income. The 1940 income variable (i.e., wage and salary income) excludes income from self-employment, including income from farming. We thus implement two notable changes to our baseline income score following the approach in Collins and Wanamaker (2017). First, we use fathers ages 30 to 50 in the 1960 Census to calculate the ratio of farmer income to farm laborer income by race and Southern residence. We then use farm laborers' income in 1940 as well as these ratios to impute the 1940 income of farmers.[11] Second, we adjust the income of self-employed non-farm workers using a similar approach: we consider fathers ages 30 to 50 in the 1960 Census and compute ratios of mean earnings for self-employed workers relative to wage-and-salary workers. When then impute the earnings of self-employed non-farm workers in 1940, by race and Southern residence.[12]

In the robustness checks, we calculate analogous income scores—at the occupation × race × South level—using the 1950, 1960, and 1970 Censuses. In all of these variations, we calculate averages of the *inctot* (i.e., total personal income) variable.[13] Finally, we also calculate the averages of the 1950*occscore* variable—which reflects the median total income of all persons with that particular occupation in that Census—for the 28 coarsened occupations (with no additional variation at the race or region level). To make sure these measures are comparable throughout the analysis, all income scores are reported in 1950 dollars.

B.1.3 Alternative Income Scores

In the robustness checks of the paper, we consider a number of alternative income scores. The first variation we consider is one that uses alternative data sources (i.e., data not from the decennial Censuses). In particular, for non-farmers, we use information on average earnings by occupation from the 1901 Cost of Living Survey (Preston and Haines 1991) and collapse this information to our coarsened occupations. We use fathers ages 30–50 in the 1940 full-count Census to adjust these income values by race

[10] Because there are no accountants in the 1940 Census, accountants are assigned the same income as professionals with the same other characteristics. For education variations, we use five levels of education: less than 8th grade, 8th grade, some high school, completed high school, and at least some college.

[11] We also follow Collins and Wanamaker (2017) and adjust farmer and farm laborer income measures upward to reflect the value of in-kind income throughout these calculations.

[12] When constructing income scores at the occupation × race × South × education, we allow the ratios in the 1960 Census to also vary along all of these dimensions. If there were fewer than 20 individuals in the 1960 Census with those characteristics, we use the income of individuals with the same recorded race, occupation, and region to compute ratios.

[13] The 1950 Census only asked the income question to a small subset of the population. As such, for the 1950 occupation × race × South variation, if there were no respondents in the 1950 Census with a particular occupation, Southern residence, and recorded race, we impute the income value using the income of similar individuals (i.e., same race, same residence, similar occupation).

and Southern residence. For fathers who are farmers, we assign an income value using the 1900 Census of Agriculture. In particular, we use information on farm output and expenses from Merriam (1902) and follow the approach in Goldenweiser (1916) and Abramitzky *et al.* (2012) to calculate farmers' income by race and Southern residence. We then adjust these values by the share of farmers in that race and region that were owners (giving non-owners 50% of farm income).

A second variation we consider is one in which we assign fathers an income score using the data source that is closest in time to when the respondent grew up. In particular, we assign the 1910 cohort the income scores using the 1900 Census of Agriculture and the 1901 Cost of Living Survey, and the 1940 cohort the income scores constructed using the 1940 Census. The 1920 and 1930 cohorts are assigned a weighted average of these two data sources. Finally, the 1950, 1960, and 1970 cohorts are assigned income scores constructed using their corresponding Census.

Two final variations we consider are those that only vary the income of fathers in farming. The first uses our baseline income scores, but inflates the income of white fathers who were farmers in the decades prior to 1950. Specifically, we inflate father farming income by 30%, 20%, 10%, and 5% for white respondents born in the 1910s, 1920s, 1930s, and 1940s cohorts, respectively. The second approach we take is excluding all respondents whose fathers were either farmers or farm laborers.

B.2 Harmonizing Survey Datasets

We typically include a survey in the analysis if it meets three conditions: First, it must survey adult individuals born in the 20th century. Second, it must ask survey respondents about their household income. And third, it must ask respondents about their fathers' occupation while they were growing up, and the available occupation codes must be able to be collapsed to our coarsened occupations. The surveys that meet these conditions then usually also include other useful information, including demographic characteristics of the respondent (e.g., age, country of birth, education, occupation) as well as of the father (e.g., education).

In the end, we have fourteen harmonized surveys:

- American National Election Studies (ANES),1956-1970
- Americans View Their Mental Health (AVTMH), 1957 & 1976
- General Social Survey (GSS), 1972-2018
- National Fertility Survey (NFS), 1970
- NLS Mature Women (NLSMW), 1968
- NLS Older Men (NLSOM), 1966
- NLSY79, 2002 [2]
- NLS Young Men (NLSYM), 1981 [2]
- NLS Young Women (NLSYW), 1988 [2]
- National Survey of Black Americans (NSBA), 1979-1980
- National Survey of Families and Households (NSFH), 1987-1988
- Occupational Changes in a Generation (OCG), 1962 & 1973

We restrict the sample to native-born respondents aged 30–50. We also include respondents in this age range for whom we do not know where they were born. We exclude foreign-born respondents because we cannot know with certainty whether they grew up in or outside of the United States. Because we assign U.S.-based income scores to the father of each respondent and because the average income for the same occupation can differ across countries, we refrain from assigning income scores to the fathers of these respondents and thus do not include them in the analysis.

Once we identify and clean these surveys, we pool them together for the analysis. An individual is in our baseline sample if he/she has an available family income, recorded race, region of birthplace/childhood (South vs. non-South), and father's occupation. Together, these four components allow us to measure the respondent's income level and compare it to his/her father's income score.

B.2.1 Respondent Family Income

In all of our harmonized surveys, respondents are asked about their family income in that year. Some surveys provide the information in categories, while others provide exact numerical values. To be consistent in our coding, we rely on the bin structure of the surveys and assign respondents the midpoint of that category.[14]

For surveys that report exact values, we replicate the bin structure for assigning respondents a family income value. In particular, we first find a survey that took place around the same time period and use that survey's bin structure as a template. We then assign individuals the midpoint of their corresponding bin.[15] Ultimately, we want to observe a *roughly* equal proportion of respondents in each bin. When the outlined procedure does not yield this result, we consider alternative bin structures (in other cleaned surveys) until we find a bin structure that results in the desired distribution.

Finally, for consistency, we ensure that each survey has roughly 10–12 bins for respondent family income. For surveys that have significantly more bins, we combine bins and assign respondents the midpoint of the new category (while simultaneously ensuring that each bin has roughly the same share of respondents).

B.2.2 Assigning Income Scores to Survey Respondents' Fathers

We obtain father occupation from the respondent, who typically reports his/her father's occupation when the respondent was around 15 or 16 years old. As previously mentioned, we harmonize father occupations into 28 coarsened categories. To do so,

[2]Note that these surveys are repeated cross-sections. We select one cross-section to clean by first observing the median age in the earliest cross-section of the survey. We then calculate the year in which the median age of respondents would be around 40. If the survey was not conducted in this year, we take the nearest survey year. We want the typical respondent to be around 40 years of age in order to minimize lifecycle bias and to preserve an average age of 40 in all of our cleaned surveys.

[14]The exception to this step is that for individuals who make the least (i.e., whose income falls in the bottom bin), we assign them 0.75 $\times$ the upper boundry of the category. For respondents who make the most (i.e., whose income falls in the top bin), we assign them 1.25 $\times$ the lower boundry of the category.

[15]For instance, because NSFH interviews took place in 1987 and 1988, we use the 1988 bins from the GSS as a template for the bin structure of respondent family income for NSFH respondents.

we construct crosswalks between the 1950 Census occupations and our coarsened occupations, as well as analogous crosswalks for the 1960, 1970, 1980, and 2010 Census occupations. If the occupations in a survey did not match the Census occupations, then we created survey-specific crosswalks between the available occupation codes and our coarsened occupations.

Once we finish coarsening occupations, we merge our father income scores by father occupation, race, and whether the *respondent* grew up in the South. While our surveys provide father occupation, they do not report information on his race. Therefore, we proxy father race with respondent race.

Our surveys do not report the state or region in which the respondent's father worked when the respondent was growing up. We can, however, observe the region in which the respondent was born or grew up. We thus use respondent residence in childhood/adolescence to proxy for father residence. Whenever we have information on both birthplace as well as childhood region, we use the latter to proxy for father residence.

B.2.3 Educational Attainment

Our constructed measures of educational attainment always reflect years of schooling *completed*. In some surveys, respondent and father education are binned (i.e., "less than grade school," "grade school," "less than high school," etc.), while in other surveys they are categorical (i.e., 0-20+ years of schooling). To harmonize across surveys, we create two education variables.

The first binned variable assigns consecutive, ascending values as follows:

(0) no education (0 years)
(1) less than grade school (1-7 years)
(2) grade school (8 years)
(3) less than high school (9-11 years)
(4) high school (12 years)
(5) some college (13-15 years)
(6) college+ (16+ years)

In contrast, the second binned variable assigns *years of schooling* in the following manner:

(0) no education
(6) less than grade school
(8) grade school
(10) less than high school
(12) high school
(14) some college
(16) college+

We create these two variables for the respondent and for the respondent's father. Whenever available, we make similar variables for the respondents' mothers. Finally, we create indicator variables denoting high school and college completion for the respondent, for the father, and for the mother if possible.

B.2.4 Weighting Scheme

We construct two types of weights for our analysis, a centered weight and population adjusted weights.

We begin by taking the provided weight in each survey and dividing it by its mean so that the weight has an average of 1. For surveys that consist of repeated cross-sections (i.e., the ANES and GSS), we re-center the weight in each survey year. If a survey does not have a weight, we create a weight with all values set to 1. We then combine these re-centered weights into one variable: the centered weight.

The main weight we use in the analysis builds on the centered weight, but adjusts it further for population characteristics. In particular, because some of our surveys are not representative on race or sex, certain cohorts in the pooled dataset will not be nationally representative. We therefore adjust the centered weights so that the share of white men, white women, Black men, and Black women in each cohort (i.e., decade) is 44, 44, 6, and 6 percent, respectively.

Throughout the analysis, we sometimes narrow the sample to certain respondents (e.g., individuals whose fathers are not in agricultural occupations, individuals with available information on father's education). For these secondary samples, we also adjust the centered weight so that the share of white men, white women, Black men, and Black women in each cohort of that sub-sample is 44, 44, 6, and 6 percent, respectively.

B.2.5 Ranking Respondents and Fathers

In addition to logging respondent family income and fathers' income scores, we also rank respondents and their fathers. In particular, we rank respondents relative to other survey respondents born in the same birth year. Similarly, we rank fathers relative to all other fathers with children born in the same year. Notably, we rank respondents and their fathers on the condition that we have a minimum of 100 observations in a given birth year for the relevant sample. Our baseline analysis sample ends up including individuals born in every year between 1911 and 1979. Finally, in our baseline approach, we use the population adjusted weights when creating ranks.

Whenever we consider secondary samples of individuals, we typically re-rank so that individuals are compared against the other individuals in that sub-sample and we use the population adjusted weights from that sub-sample.