Family Trees and Falling Apples: Historical Intergenerational Mobility Estimates for Women and Men[1]

November 2023

Kasey Buckles
University of Notre Dame,
NBER, and IZA

Joseph Price
Brigham Young University,
NBER, and IZA

Zachary Ward
Baylor University and NBER

Haley Wilbert
University of Notre Dame

Abstract

Efforts to document long-term trends in socioeconomic mobility in the United States have been hindered by the lack of large, representative datasets that include information linking parents to their adult children. This problem has been especially acute for women, who are more difficult to link because their surnames often change between childhood and adulthood. In this paper, we use a new dataset, the Census Tree, that overcomes these issues by building on information from an online genealogy platform. Users of the platform have private information that allows them to create links among the 1850 to 1940 decennial censuses; the Census Tree combines these links with others obtained using machine learning and traditional linking methods to produce a dataset with hundreds of millions of census-to-census links, nearly half of which are for women. With these data, we produce estimates of the intergenerational transmission of socioeconomic status from fathers to their sons and daughters. We find that for married men and women, the patterns of mobility over this period are remarkably similar. Single women, however, are less mobile than their male counterparts. We also present new estimates that show that assortative mating was much stronger than previously estimated for the US.

[1] This work is supported by grants from the Russell Sage Foundation (G-1063) and the National Science Foundation (SES-2049762). We are grateful to the excellent research assistants at the University of Notre Dame and the BYU Record Linking Lab for their work on this project, including Brianna Felegi, Nate McGhie, and Adrian Haws. We have benefited from feedback at many seminars and conferences, and from conversations with Martha Bailey, Greg Clark, Bill Evans, Greg Niemesh, and Claudia Olivetti.

I. Introduction

It may be "self-evident that all men are created equal," but it is also self-evident that children are born into unequal families. This lottery of birth – whether children are born into rich or poor families – has long-run implications for a child's income, education, and wealth (Black and Devereux 2011, Chetty et al. 2014, Corak 2013). There is a second feature of this lottery—whether one is born male or female—that also affects one's place in society. Women, conspicuously absent from Jefferson's Declaration, may have had relatively fewer avenues for overcoming the circumstances of their birth as their labor market opportunities and rights were more limited in the past. However, these conditions also mean that women's status was often tied to their husband's, which may be *less* correlated with her own family background.[2] Due to limitations in historical data, however, most estimates of intergenerational mobility over the last two centuries are based on male-only samples. As a result, the stakes of the birth lottery for women, and how they have changed throughout American history, remain unclear.

The exclusion of women is not the only way that historical data are unrepresentative of the population. Historical mobility estimates are often based on datasets that track individuals across censuses, but only a select group can be successfully linked from childhood to adulthood (Bailey et al. 2020). In the absence of a unique administrative identifier such as a registration number, one must have stable and unique characteristics to be successfully linked, such as first name, surname, birthplace, and birth year. The lack of a stable surname for most women explains why they have been excluded from nearly all linked samples. But those with common names, like "John Smith," are also excluded because it is unclear which John Smith is the correct one. This common name problem contributes to low linking rates across censuses, such that linked samples capture only 10 to 30 percent of the underlying population (Abramitzky et al. 2021a). If this 10 to 30 percent is a select group, then a researcher may draw incorrect inferences about the level and trend of intergenerational mobility.

In this paper, we present the most comprehensive picture of historical intergenerational mobility in the United States to date, making three key contributions that lead to novel findings. First, we use a new linked dataset, the Census Tree, that is well beyond the current frontier in terms of the number, quality, and representativeness of the links (Buckles et al. 2023, Price et al.

[2] The evidence on the relative strength of parent-daughter and parent-son associations is mixed. Modern-day evidence suggests that parent-daughter associations are weaker, (Chadwick and Solon 2002; Chetty et al. 2020), but studies using earlier cohorts sometimes find the opposite (Eriksson et al., 2023; Jácome, Kuziemko, and Naidu 2021).

2021). The advantage of the Census Tree is that it uses information provided by FamilySearch.org, one of the largest web-based genealogy platforms in the world. FamilySearch users have private information (e.g., sibling names, maiden names) that allows them to find and link records in a way that would not be possible for a trained research assistant or machine learning algorithm. The FamilySearch data is exceptionally rich: it contains 158 million male links across two censuses, an amount that is greater than the number of conservative links in the Census Linking Project (Abramitzky et al. 2022). The Census Tree data combine these links from FamilySearch users with others made by standard linking methods—including a machine learning approach that uses the FamilySearch links as training data—to yield a dataset with 391 million links for men among the 1850 to 1940 censuses. Ultimately, while most recent papers are based on samples that contain between 10 and 30% of the possible links among censuses (Abramitzky et al. 2021b, Collins and Wanamaker 2022), the Census Tree contains between 68 and 75% of these links.[3] This substantial increase in the number of linked individuals translates into a much larger analysis sample, allowing us to address concerns about selection bias and measurement error.

The Census Tree also enables our second contribution: the estimation of intergenerational mobility patterns for women. The access to records and private information that FamilySearch users have (especially maiden names) allows them to create links for women at rates nearly as high as those for men—there are 153 million user-made links for women in the data. When combined with links using other methods in the Census Tree, we obtain 314 million female links. This set of census-to-census links for women is truly unprecedented; previous efforts to create census links at a large scale have omitted women entirely (e.g., Abramitzky et al. 2022, Feigenbaum 2016). Others have used marriage records to link women from childhood to adulthood (Bailey and Lin 2022, Eriksson et al. 2023, Withrow 2021); however, marriage records are typically held at the state-level and therefore do not allow researchers to estimate mobility for the whole population. The Census Tree links for women are highly representative of the population, just as they are for men. We use these links to build female mobility estimates for the 1835 to 1915 birth cohorts, and we compare the strength of the transmission of status for daughters to that for sons, by marital status. Thus,

[3] These estimates refer to the average match rate among all 36 census-to-census pairs between 1850 and 1940, and indicate upper- and lower-bounds, depending on the approach used to construct the denominators for the match rates. See Buckles et al. (2023) for more on how match rates are calculated.

we can create mobility estimates for birth cohorts before survey data that start with the 1910s birth cohorts (Jácome, Kuziemko, and Naidu 2021).

Third, the information on women in the Census Tree allows us to generate new estimates of assortative mating between wives and husbands for the United States. We are able to compare the socioeconomic status of the couple's fathers, where again the very large sample size allows us to address selection issues and measurement error. We use these estimates to understand the role of assortative mating in generating economic mobility for married men and women. These estimates contribute to a small but growing literature on assortative mating in the U.S., which has relied on marriage records (Bailey and Lin 2022, Eriksson et al. 2023), indirect links across censuses based on first names (Olivetti et al. 2022), or social security applications (Althoff, Gray, and Reichardt 2023).

With these advantages, we produce gender-specific estimates of the intergenerational transmission (IGT) of socioeconomic status for cohorts born between 1835 and 1915, leading to several new findings. We estimate mobility in a variety of ways, including traditional estimates where the son's occupational status is regressed on a single snapshot of the father's status for a sample of white men (e.g., Song et al. 2020). We also create updated estimates following Ward (2023) where we add Black men to the sample, instrument the father's status with a second observation to account for measurement error, and use improved measures of economic status that account for within-occupation differences by race and region. This analysis leads to our first result: once both Census Tree and Census Linking Project samples are weighted to be representative of the population, they yield very similar estimates of mobility for men, no matter the method of estimation. These results bolster confidence in the extensive research that relies on Census Linking Project links and on the Bailey et al. (2020) weighting procedure.

Having produced revised estimates of mobility for men using the Census Tree data, we then turn to estimates that include women. For married women, we follow the literature in using her husband's occupation to construct her measure of status. Doing so generates our second finding: the strength of the transmission of socioeconomic status to married daughters and married sons is remarkably similar. Intergenerational transmission estimates for married women are within 3 to 5 percent of those for married men. Because transmission estimates do not weaken much for married daughters, this result suggests a high level of assortative mating—a topic we return to in later analysis. The trend is exactly the same as well, where we find decreasing transmission (or increasing mobility) between 1840 and 1910 birth cohorts.

For single men and women, we use their own occupation, and obtain a third finding: single daughters' fortunes appear to be more tied to that of their fathers than their married counterparts, while the opposite is true for sons. As a result, there is a large level difference in our estimates of the intergenerational transmission of status for single men and women, where father-daughter transmission is stronger than father-son transmission. This difference is partly due to racial differences in labor force participation. Black women were more likely to report an occupation than white women, and thus the single women sample has a higher proportion of Black individuals, which in turn lowers mobility estimates. The importance of Black women for mobility estimates is also stressed by Jácome, Kuziemko, and Naidu (2021); our results align in that we find that overall parent-child persistence estimates are stronger when including Black women. However, we also show that for a white sample, transmission was stronger for single women than single men for most birth cohorts.

Finally, we create new estimates of assortative mating based on the association between a wife's father's status and the status of her father-in-law. Our fourth finding is that assortative mating in this period was stronger than previous work suggests. The father and father-in-law association ranged between 0.64 and 0.76 for the 1840 to 1910 birth cohorts; these estimates are more than twice as high as other recent estimates for the same period, which vary between 0.15 and 0.35 (Althoff, Gray, and Reichardt 2023, Bailey and Lin 2022, Eriksson et al. 2023, Olivetti et al. 2022). We show that there are two explanations for our higher estimates: 1) we account for measurement error using the IV approach, and 2) we account for racial and regional differences in status within occupation. We further provide additional evidence that assortative mating was strong: we estimate with 1940 education data that a husband's education was associated with *both* the wife's education *and* the father-in-law's status at the same time. A father-in-law effect could be due to couples matching on parental status in addition to spousal status, or due to measurement error attenuating the spousal correlation in education (Ferrie, Massey, and Rothbaum 2021). Either way, it shows that the standard approach to estimating assortative mating via spousal education understates the true degree of assortment (e.g., Mare 1991; Eika, Mogstad, and Zafar 2019), a point that is also made by Collado, Ortuño-Ortín, and Stuhler (2023). This strong rate of assortment helps to understand the similarity in our mobility estimates for married men and women (Curtis 2021, Clark and Cummins 2022).

II. Intergenerational Mobility Estimates: The Frontier

There has been a tremendous amount of scholarship on the IGT of socioeconomic status, spanning many decades and countries—indeed, there is now a literature of literature reviews on the topic (Ganzeboom, Treiman, and Ultee 1991; Solon 1992; Black and Devereux 2011; Torche 2015; Mazumder 2018). In this section, rather than attempting a comprehensive survey of recent papers, we identify three that collectively represent movement along the frontier and highlight the value of our contributions.

First, Feigenbaum (2016) introduced the potential for using machine learning methods to dramatically reduce the cost of high-quality record linking, and applied this method in his paper estimating the IGT of income and education for a cohort of sons in the 1915 Iowa State Census (Feigenbaum 2018). Feigenbaum takes advantage of the fact that this state census recorded education and income—outcomes not available in the full national census until 1940. To facilitate the matching of childhood observations (with information on the father's outcomes) in the 1915 Iowa census and the cohort's adult outcomes in the 1940 national census, he first creates by hand a training dataset that has many examples of both true and false links. He then uses this training data to create a machine learning algorithm that quickly and cheaply makes additional matches. Feigenbaum's paper demonstrates that the combination of supervised (which require training data) and unsupervised (which do not) linking methods can produce higher match rates than those obtained by previous linking methods—he can match 59% of the sons in the 1915 Iowa census to their adult observations. However, because the Iowa census was a unique event, he is not able to provide similar estimates for other states, or for other time periods in Iowa. Furthermore, the sample excludes women entirely, and 99% of the linked sample is white. As we describe in Section IV, the Census Tree dataset also combines both supervised and unsupervised methods to obtain the links, but using the complete U.S. censuses from 1850 to 1940.

Another innovation comes from Olivetti and Paserman (2015), who take up the challenge of producing IGT estimates for women. Rather than try to solve the problem of linking women's childhood and adulthood records, they instead impute the parent's socioeconomic status, using the information contained in the child's first name. For example, they assign each "Luke" the average status of all parents with a child named Luke around the time of birth, with the insight being that the strategy works for "Susan" as well. They are therefore able to produce estimates of the IGT of socioeconomic status (as measured by occupation) for both men and women in the 1870-1940 censuses. However, it is unclear how much measurement error is introduced by assigning childhood status based on first name. We will be able to produce actual childhood-to-adulthood links for millions of women, thus eliminating this source of measurement error.

Finally, Ward (2023) uses links created by the Census Linking Project (Abramitzky et al. 2022) to link men in the 1870-1940 U.S. Censuses to their fathers, making three critical adjustments to the traditional approach to estimating the IGT of socioeconomic status. First, he includes Black men, and shows that not doing so causes one to severely overestimate the amount of socioeconomic mobility in the U.S. during this time period. Second, he takes advantage of the fact that for many fathers, occupation is observed in multiple censuses. This allows him to implement averaging and instrumental variables strategies that reduce measurement error in the father's status. This adjustment also leads to the conclusion that mobility was lower in the past than previously thought. Third, he refines the measure of socioeconomic status, incorporating information about not only occupation, but also region, race, and cohort. We adopt each of these adjustments in our paper, and our much larger sample allows us to include many more fathers with multiple observations; our estimation sample is over five times larger than that used in Ward (2023). And again, we can produce intergenerational mobility estimates for women, which Ward is unable to do.[4]

III. Estimating the Intergenerational Transmission of Status

The trend in relative mobility is estimated with the simple descriptive regression:

$$y_{child} = \beta_0 + \beta_1 y_{parent} + \varepsilon_{child} \tag{1}$$

where the socioeconomic status of the child is predicted by the status of the parent. If the coefficient β_1 is positive and large, then a parent's status "matters" (noncausally) for the child's outcome and indicates low relative mobility. If β_1 is close to zero, it reflects high mobility where parental background does not matter. The regression is deceptively simple, but there is a large literature on how to accurately measure β_1 (Solon 2002, Mazumder 2005, Chetty et al. 2014, Nybom and Stuhler 2016). Since our paper aims to measure historical mobility for the whole population, we will discuss how problems with measurement, particularly in historical data, may bias estimates.

[4] Jácome, Kuziemko, and Naidu (2021) aggregate many surveys to produce estimates of the intergenerational transmission of socioeconomic status for the 1910-1970 birth cohorts, which is beyond the period in our analysis. Althoff, Gray, and Richardt (2023) use social security records to construct census links for men and women in the 1850-1900 birth cohorts; they obtain match rates of 2% (1%) for men (women) in the earliest census pair, and of 20% (16%) for the latest. They use these links to produce OLS estimates of intergenerational mobility in household income.

The first measurement issue is the representativeness of the sample, which biases estimates in an unclear direction. Historical data on parental and child outcomes are created by linking children from their childhood home to their adult outcomes in a later census. However, since linking is non-random, the sample may mismeasure mobility for the underlying population (Bailey et al. 2020). To address this issue, researchers weight the linked sample to reflect population characteristics. But there may still be concern about unobservable selection into the linked sample, especially since linked samples often capture less than 20 percent of the population. Besides nonrandom linkage, historical data may fail to capture mobility for the entire US population because specific demographic groups are excluded from the data. For example, the enslaved and their descendants are often not in linked historical data, which then biases estimates of β_1 downward (toward greater mobility) because this population started low and ended low in the economic distribution (Ward 2023). Another group excluded from linked samples is women. Excluding women may bias estimates of β_1 upward or downward, depending on whether parental status is more predictive of child status for sons or daughters (Chadwick and Solon 2002, Jácome, Kuziemko, and Naidu 2021).

The second measurement issue is that linked data contain false links. Researchers link individuals across censuses based on a limited set of characteristics, such as name, age, and birthplace. Because the set is limited, an individual (e.g. "John Smith") may be linked to someone else with a similar name, age, and birthplace combination. This issue biases estimates of β_1 downward and falsely implies greater mobility (Bailey et al. 2020). To address this problem, researchers have resorted to more conservative linking algorithms that drop individuals with the same name, birthplace, and age combinations. However, there is a tradeoff here: dropping common names improves linking accurately, but also leads to more non-representative samples. Ideally, a researcher would have a linked sample that is both highly accurate and covers a large portion of the US population – like the dataset we use in this paper.

Third, a long literature argues that a single snapshot of the father inadequately captures his permanent economic status (Solon 1992, Mazumder 2005). This issue also applies to occupation-based status measures used in historical studies (Ward 2023). One reason for measurement error is that fathers may be placed into the wrong occupation code, which is a problem that also exists in modern-day surveys (Kambourov and Manovskii 2008, Moscarini and Thomsson 2007, Vom Lehn, Ellsworth, and Kroff 2022). One solution to this problem is to average more than five observations of a father's occupational status (after transforming occupation codes into a unidimensional scale, such as an income score), but this is not possible in a historical context where censuses are taken ten years apart. Another is to use an instrumental variables strategy to address measurement error, where a parent's status in one year is instrumented with a second observation.[5] This method requires linking multiple censuses to get multiple father observations, which can result in a significant loss of observations and intensify any bias from having a nonrandom sample (Ward 2023).

The data requirements for addressing all of these measurement issues are extensive. It necessitates a linked sample that has the following properties: (1) a high linking rate, (2) high accuracy, (3) population representativeness, and (4) multiple observations for the father. The next section will elaborate on how we construct this sample.

IV. Data and Measurement

A. The Census Tree Linked Dataset

Our analysis relies on data from the Census Tree, which is the largest existing set of record links for people in the 1850-1940 U.S. Censuses. Detailed descriptions of the Census Tree and the methodology used to create it are available in Buckles et al. (2023), but we provide a summary here.

The foundation of the Census Tree is data from FamilySearch.org, one of the world's largest internet genealogy platforms. The platform works like a wiki; when users have an ancestor in common, they can link their family trees, and any of the 12 million+ users can add information to a profile for a deceased person. This information includes birth dates, death dates, family histories, and—relevant for our purposes—digitized records including the historical U.S. censuses. When census records from two different years have been attached to a single profile, this creates an observable census-to-census link, and the set of these user-made links constitutes a dataset called the "Family Tree." Critically, the users typically have private information (e.g., sibling names, maiden names) that allows them to find and link records in a way that would not be possible for a trained research assistant or machine learning algorithm. The Family Tree dataset alone contains over 158 million links for men and 153 million links for women among the 1850-1940 censuses.

[5] The IV estimate takes the form:

$$y_{parent,t} = \pi_0 + \pi_1 y_{parent,s} + u_{parent,t} \tag{2}$$

$$y_{child} = \delta_0 + \delta_1 y_{parent,s} + v_{child} \tag{3}$$

Where $y_{parent,t}$ is the first parent observation and $y_{parent,s}$ is the second parent observation. The IV estimate is a ratio of the reduced form in Equation (3) and the first stage in Equation (2) (i.e., $\hat{\beta}_{IV} = \frac{\hat{\delta}_{IV}}{\hat{\pi}_1}$). For the instrumental variables estimate to be a consistent estimate of the true β_1, the assumption is that error in the first observation is uncorrelated with error in the second observation.

To create the Census Tree, Buckles et al. (2023) build on the Family Tree links in two key ways (see also Price et al. 2021). First, a subset of these links is extracted as a "truth set" to be used as training data for a machine learning algorithm to make new links. The algorithm uses features of the data that are based on characteristics commonly used in unsupervised linking strategies (e.g. names, birth years, birthplace), but is also able to incorporate less stable information like place of residence.[6] Second, matches from the Census Linking Project (Abramitzky et al. 2022), the IPUMS Multigenerational Longitudinal Panel (Helgertz et al. 2023), and from FamilySearch's algorithms are added, along with "implied links."[7] The links are filtered for quality, and conflicts among the different sources are resolved by comparing the number of other matches on the same census sheet. The combined set of Family Tree links and new links made using these approaches constitutes the Census Tree (Price et al. 2023).[8]

Figure 1 demonstrates the incredible advances made by the Family Tree and Census Tree. In it, we show the match rates between adjacent censuses by gender, for these two datasets and for the Census Linking Project (CLP) data.[9] We use the exact-conservative matches from the CLP which require first name, last name, and birthplace to be unique within a five-year age ban, as these are the links used to produce the main estimates in Ward (2023). Match rates are constructed as the fraction of people who are over age 10 in the latter census that we can find in the previous census, adjusting for emigration and under-enumeration.

There are four key takeaways from the figure. First, for men, the Family Tree data alone contain more links than the CLP data (exact-conservative), and the Census Tree attains match rates that are 3.4 to 4.7 times larger than the CLP, reaching rates over 80% for the censuses in the 20th century.[10] These high match rates correspond to 126 million additional links for men between adjacent censuses. Second, for the user-made links in the Family Tree, match rates are similar for men and women. This is consistent with family members having interest in and information about family members of both genders. Third, the Census Tree is over twice as large as the Family Tree

[6] Note that the algorithm does not require that a person live in the same place in the two censuses, but it can assign a higher probability that the records are a match when they are geographically closer to one another. Price et al. (2021) document the gains to precision and recall when geographic information is included.
[7] If a link is identified between censuses A and B, and censuses B and C, then the link between A and C is an implied link.
[8] The Census Tree dataset is available at censustree.org, along with the code used to create it, and the models and training data for the machine learning methods.
[9] See Buckles et al. (2023) for match rates and sample sizes among non-adjacent censuses in the Census Tree.
[10] Another important characteristic of linked data is precision, or the number of matches that are correct. All of the methods shown in Figure 1 have rates of precision above 90%, and Price et al. (2021) show that the Census Tree does not sacrifice precision to attain its higher match rates.

for both men and women; this is due to the new links made by the additional supervised and unsupervised linking strategies described above. Fourth, the additional steps made to create the Census Tree links add more links for men than for women; this primarily reflects the fact that the training data cannot "teach" the algorithm to match maiden and married names. Nevertheless, we link over three-fourths of linkable women between adjacent censuses in the twentieth century. For the 1800s, match rates for both genders are lower, but the Census Tree still contains well over half of the possible links.

When combining all census-to-census pairs, the Census Tree contains 391 million links among censuses for men, and 314 million links for women. The remarkable coverage of the Census Tree helps to address concerns about selection into the sample. Buckles et al. (2023) compare the linkable population to the Census Tree and Family Tree samples and show that the samples are representative of the population along observable dimensions. Moreover, the samples are large enough to support re-weighting to make the samples match extremely closely.

B. Measuring Socioeconomic Status

Educational attainment and income are preferred measures of socioeconomic status, but unfortunately, these are not available in US Census data before 1940. As a result, historical IGT estimates typically use a measure based on occupation. One example is the "occscore" measure available from IPUMS, which assigns each occupation its median total income among those with the same occupation in 1950. Song et al. (2020) instead assign each occupation in a given birth cohort a percentile rank based on the occupation's human capital level, as captured by literacy for the 1850-1930 Censuses and years of education in the 1940 Census. Because percentile rankings are made within birth cohorts, this measure captures changes in the status of an occupation over time. Ward (2023) creates an "adjusted Song score" that further stratifies the data allowing for within-occupation differences by race and region, which follows recent evidence that occupation-only scores miss key inequalities that are important for measuring intergenerational mobility (Saavedra and Twinam 2020, Collins and Wanamaker 2022). For this paper, we additionally stratify by gender when creating the adjusted Song score.[11]

A limitation of this method is that variation within occupation, race, place, cohort, and gender is unobserved, which may attenuate estimates of income mobility. For example, two white

[11] In practice, our results are nearly identical when we do and do not stratify by gender, and this choice does not explain any differences between our results and previous work.

male Pennsylvanian farmers of the same age will be assigned the same status, even if one is more successful. For this reason, estimates of historical mobility should not be directly compared to income mobility estimates based on modern-day tax records (Chetty et al. 2014).

A key group of fathers for whom we cannot use this method to create an adjusted Song score are enslaved men, as they were not enumerated in the 1850 and 1860 censuses. For these men, we impute their "occupation" to be enslaved and then assign it the lowest status measure in the distribution (effectively 0 on the 0-100 scale).[12] This allows us to capture the impact of emancipation on mobility, which has been estimated to increase absolute mobility since children had improved outcomes relative to parents, but did little to improve relative mobility since Black boys started at the bottom of the distribution, and ended there as adult men (Ward 2023). Using the Census Tree data, we can determine whether including women in the sample changes this current understanding of mobility.

C. The Estimation Sample

To construct our sample for intergenerational mobility estimates, we first identify all children between the ages of 0 and 14 (inclusive) in the 1850 to 1910 censuses for whom an adjusted Song score for the father can be constructed. We then identify those for whom we can 1) link to an observation with an adjusted Song score (or husband's score) when they are between the ages of 25 and 55, and 2) find at least one additional observation for the father's adjusted Song score, which will be used as the instrument in our IV approach. We take a second father observation up to 20 years away from the "primary" observation when he is with the child.[13] We refer to this as our "doubly-linked" sample. When we observe multiple adult observations for the son or daughter, we use the observation closest to age 40. When we observe more than one supplemental observation for the father, we use the one from the closest census to our focus observation.[14]

Since our empirical strategy requires a sample that is doubly linked rather than singly linked, the advantage of the Census Tree's higher linking rate becomes even more apparent. This is because

---

[12] Specifically, we append a random sample of adult southern-born Black men and women from the 1870 and 1880 Censuses to the linked sample, and impute the father's adjusted Song score as described.

[13] This differs from Ward (2023) who takes a second father observation up to ten years away. Mobility estimates are not qualitatively changed when expanding the window for fathers up to twenty years, but it does allow for larger sample sizes, which potentially reduces nonrandom selection into the sample.

[14] For example, if the child is observed in the 1900 census, we use the father's adjusted Song score in 1900 as the primary observation. The child's adult adjusted Song score would be observed in the 1920, 1930, or 1940 census, where we select the observation closest to age 40. The father's second observation can come from, in order of selection, the 1910, 1920, or 1880 census (the 1890 census would be used before the 1920 census had it not been destroyed).

multiple links compound the problem of linkage failures. To simplify, if linking rates across censuses are independent, then a 30 percent rate for one link reduces to a 9 percent rate for two links. But if the linking rate increases to 70 percent, then the double-link rate would be 49 percent – over five times greater. In practice, linkage failures are not independent since factors that make it more difficult to make a link (e.g., a common name) may lead to failure for both links.

We cannot use the same backward-looking method used to construct the census-specific match rates in Figure 1 to construct match rates for the estimation sample, because a child's adult observation could come from multiple different censuses. However, in Table 1 we report the number of observations lost by each of the restrictions described above. The first column shows the primary sample—the number of children aged 0-14 in each census between 1850 and 1910 for whom an Adjusted song score can be calculated for the father. In column two, we report the number for whom we can find the child as an adult with an adjusted Song score (or for their husband in the case of married women). This singly-linked sample is what is most commonly used in the literature; with the Census Tree we are able to create a single link for 43 percent of the base sample. When we further restrict this sample to those for whom we can observe a second adjusted Song score for the father (the doubly-linked sample), we still retain 37 percent of the sample, or 32.5 million total observations.

We can compare the size of the resulting estimation sample to that constructed by Ward (2023) using the Census Linking Project links. Figure 2 shows the ratio of the sample sizes in the two data sets by birth cohort.[15] In total, the Census Tree estimation sample is over five times larger for men, with over 20 million observations. When including the 11.5 million doubly-linked observations we can construct for women, the Census Tree has eight times more father-child pairs than the sample using exact-conservative links from the Census Linking Project.

To address any remaining nonrandom selection into the sample, we follow Bailey et al. (2020) and reweight the linked data to match population characteristics based on the adult child's observable characteristics. To do so, we pool the linked sample with the full-count census data for each year we observe the child as an adult (1870-1940). We then use a probit model to predict whether individuals are in the linked sample based on demographics, migration, broad occupation categories, and geography.[16] We then use the predicted probabilities to create inverse probability

[15] In all results organized by birth cohort, we group birth years into ten-year bands around the central year. For example, the 1900 birth cohort includes births between 1895 and 1904.

[16] The probit model predicts based on (1) gender, (2) whether one is ever married, (3) age (indicator variables based on

weights for each individual. In the following results, we always use weighted data unless otherwise mentioned. Sometimes we compare Census Tree mobility estimates to those based on Family Tree links or the Census Linking Project links. When doing so, we use the same process to create custom weights for the other datasets, with the exception of pooling with women or using gender to predict linkage for the Census Linking Project.

The descriptive statistics for our main linked sample are shown in Table 2. There are three key things to note. First, the fraction male in the sample ranges from 51 to 59 percent, depending on the birth cohort. The fraction male is not 50 percent because to be included in the sample (and intergenerational regression), one needs to either report an occupation or have a husband with an occupation. Because single women have a lower labor force participation rate than single men, the overall sample is slightly more male. Second, our sample includes millions of observations for Black men and women. Third, the overwhelming majority of the sample is married (73-87%), but we still retain unmarried individuals. Later, we will explore heterogeneity in mobility by marital status because marriage might be an important channel for upward mobility.

V. Intergenerational Mobility Results

A. Reproducing Estimates for Men Using the Census Tree

First, we investigate whether the Census Tree, with its significantly higher linking rate, alters our understanding of male mobility throughout American history. We focus on men initially to enable comparisons with prior estimates, which have primarily relied on male samples. Later, we will add women to the sample to separately ask how gender influences mobility.

Before estimating mobility using the Census Tree data, we begin by using links from the Census Linking Project to recreate various mobility estimates based on Equation (1).[17] First, we use the traditional approach, as implemented by Song et al. (2020), where the occupational score of the son is regressed on the occupational score of the father. These estimates are the bottom line plotted in Figure 3, Panel A. However, as highlighted by Ward (2023), father-son associations become stronger after making three improvements to the data. First, we account for within-

decadal age) (4) region of residence (5) whether one lived in a different state than birth (6) whether on holds a white-collar job, farmer, skilled, unskilled or no job (7) whether someone lived in an urban area. An indicator for Black is also included and interacted with the previous variables, except for moving across states. After weights are implemented, we do a final adjustment for years when the enslaved sample is included to ensure the Black share of sample is correct for the 1870 and 1880 Censuses.

[17] Figure A1 plots the binscatter relationship between the father's status and the child's status and shows that it is approximately linear.

occupational differences in status by race and region; second, we add Black men to the sample; and third, we instrument the first father observation with a second one to account for measurement error. These three adjustments are incrementally added in the figure. Making these adjustments changes the traditional narrative of the trend in relative mobility throughout American history. Initially, there is a "high but decreasing mobility" story, as indicated by the father's status being weakly predictive of the sons in the earliest 1840 birth cohort (0.24) but becoming more predictive over time for the 1910 cohort (0.37). In other words, America used to be a high-mobility society, but that disappeared over time (Long and Ferrie 2013, Song et al. 2020). However, after making the adjustments, the narrative changes to "low but increasing mobility." Now as shown by the top blue line, the father's status is strongly predictive of the son's status for the 1840 cohort (0.83) but becomes less predictive over time (0.65 for the 1910 cohort).

Our interest is to check whether these mobility estimates change when using improved links from the Census Tree. Figure 3, Panel B plots the Census Tree estimates and shows that mobility estimates are very similar across linking methods—despite the sample size increasing by five times for the Census Tree. For our preferred estimates, the Census Tree data estimates fall from 0.80 for the 1840 cohort to 0.63 for the 1910 cohort, which is similar to the 0.83 to 0.65 fall with the Census Linking Project data. There are wider differences for estimates when changing the status measure to an occupation-only score, where the Census Tree has smaller persistence estimates. However, both linking methods show a "high and decreasing mobility" level and trend. Overall, it appears that increased linking rates in the Census Tree do not substantially change mobility estimates for men.

These results assuage concerns about the representativeness of our doubly-linked sample and demonstrate the effectiveness of the Bailey et al. (2020) weighting procedure. When we do not weight the data (Figure A2), the preferred IGT estimates for the Census Tree and Census Linking Project both drop, indicating higher mobility. The reason for this is that there is a lower share of Black men in the unweighted data than in the weighted data because Black men are more difficult to successfully link. Weighting corrects for this by increasing the Black share of the sample, which in turn raises IGT estimates. To understand why, recall that the overall IGT estimate is a weighted average of within-race IGT estimates and the between-race persistence of the Black-white gap (Hertz 2008). The weights for each of the within-race and between-race components are determined by their respective shares of the overall variation, where the between-race share is about one-third (Ward 2023). Because between-race persistence estimates are high throughout American history (around 0.90-0.95 (Margo 2016)), adding Black men to the sample increases the IGT toward this 0.90-0.95 estimate (Jácome, Kuziemko, and Naidu 2021; Ward 2023). If one uses a white-only sample, then the unweighted estimates are more similar to the weighted estimates, suggesting that the major way weighting influences mobility estimates is by adjusting for the racial composition of the sample.

B. Genealogy-Based Links Produce Reliable Mobility Estimates

Besides linking at a higher rate, the other big advantage of the Census Tree data is that we can include women. But before we estimate mobility for women, we first ask whether the sample of women in the Census Tree, which relies more heavily on the genealogical links in the Family Tree, yields biased estimates. This could occur because FamilySearch users and their links may not be representative of the population in a way that cannot be addressed by weighting. To check on the representativeness of genealogical links, we exploit the fact that FamilySearch users also link men. If male mobility estimates from the Family Tree alone yield similar mobility estimates to those obtained using the full Census Tree, we can be more confident that female mobility estimates are reliable.

Figure 4 compares male mobility estimates for the genealogically linked Family Tree to those obtained using the broader Census Tree, where all results are weighted to be representative of the population. While the point estimates differ slightly by linking method, crucially, the trends are nearly identical. Overall, the Family Tree estimates are about 3-7 percent higher than the Census Tree estimates, suggesting lower mobility. The slightly higher IGT estimates for the Family Tree are likely because it contains fewer false links, where false links reduce the father-son association (Bailey et al. 2020). We are therefore confident that our estimates for women will be informative for the entire population, even though links from childhood to adulthood are disproportionately based on the Family Tree. Given this result, we proceed with our estimates of mobility by gender.

C. Mobility Estimates for Women

When presenting our main results by gender, we also separate the sample by marital status. We do this because socioeconomic status is measured differently for married and single women—for the former we use the husband's occupation to construct our measure, while for single women we use their own occupation. Furthermore, we are interested in exploring whether IGT was greater for married or single women. In effect, the estimates will allow us to determine which is more strongly tied to a woman's father's status: her husband's if she is married, or her own if she is single and working? An important consideration is that selection into marriage and labor force participation is likely changing over this period, and that the selection patterns differ by race (Goldin 1977, Elliot et al. 2012). We discuss this further below.

The first result with the new linked data is that IGT estimates for married women and married men are strikingly similar both in terms of the level and trend. We show this in Figure 5, Panel A, where we estimate father-child associations for married women using their husband's socioeconomic status as a proxy for their own status (Olivetti and Paserman 2015). The married male transmission estimates fall from 0.81 for the 1840 birth cohort to 0.66 for the 1910 birth cohort, while the married female (or, father/son-in-law) estimates fall from 0.85 to 0.66. The similarity in the levels and trends is not a result of the fact that observations for women are disproportionately from the Family Tree—if one uses only the Family Tree links to estimate *both* married male and female mobility, the same pattern holds (Figure A3).

The fact that estimates are so similar between father and son and father and son-in-law is far from trivial. The son and son-in-law were (hopefully) raised in different households; therefore, any unobserved skill, preferences, or cultural values that are directly transmitted to the son are not directly transmitted to the son-in-law in the same way. Since the parent-child association does not weaken much via marriage, it suggests that women matched with men with similar underlying status – that is, assortative mating was very high (Curtis 2021; Clark and Cummins 2022). One reason for high assortment could be matching on unobservable or latent status (Collado, Ortuño-Ortín, and Stuhler 2023), but another possibility is that high assortment is driven by geography. If where one grows up shapes both adult outcomes as well as the marriage market, then the father and son association will be similar to the father and son-in-law association (Olivetti et al. 2022).[18]

The magnitudes of the female estimates are much higher than the most prominent national estimates from the literature in Olivetti and Paserman (OP) (2015). OP use the same 1850-1940 censuses to estimate mobility based on pseudo links from first names. One reason our transmission estimates are higher is that this pseudo-linking methodology can be subject to measurement error that attenuates estimates (Santavirta and Stuhler 2023). OP pseudo-link across one-percent samples of the US Censuses (rather than full-count Censuses), which introduces measurement error because

[18] We find evidence that geography is an important explanation for assortative mating patterns. When measuring assortative mating via the association between the father's and father-in-law's status, the strength of this association decreases after adding controls for geography, such as childhood town of residence. We leave a full exploration of assortative mating patterns for future work.

the children in the first one-percent sample may be different from the adults in the second one-percent sample. A second reason why our point estimates differ is that OP do not pseudo-link Black sons or daughters, in part due to limitations in observing the population before emancipation. A third reason is that OP do not account for within-occupation differences by race and region. Ward (2023, Figure 6B) addresses these three issues and updates the OP estimates by using full-count data, including the Black population, and updating the measure of socioeconomic status while retaining the name-based method. Female parent-child associations are estimated to trend from 0.91 for the 1840 cohort to 0.85 for the 1910 cohort – estimates that are more similar in level and trend to our directly linked estimates shown in Figure 5A. Therefore, the name-based approach and directly linked approach appear to yield similar levels and trends, though the directly linked estimates do show a steeper fall in the parent-child association (or rise in mobility) than the name-based estimates. We prefer the directly linked estimates since names may provide additional information content that biases mobility estimates (Santavirta and Stuhler 2023).

We can also estimate mobility for single women who report an occupation (57% of all single women in the 1870-1940 censuses). Single women are not always included in historical mobility estimates, in part because others have used marriage records (Eriksson et al. 2023) or rely on the husband's occupation to measure status (Olivetti and Paserman 2015). We show mobility estimates for single women in Figure 5, Panel B. Rather than a strong similarity in estimates between men and women, as was found for the married group, there is a large difference in mobility between single women and men. Overall, we estimate a stronger rate of transmission between fathers and single daughters than between fathers and single sons. For example, the single-daughter estimate is 0.71 for the 1910 cohort, which is 25 percent higher than the single-son estimate of 0.57.

A key reason why we estimate higher persistence for single women is because the single female sample has a higher share Black (15%) than the single male sample (10%).[19] This is because Black women had a higher labor force participation rate than white women, a pattern long recognized by economic historians (Goldin 1977, Boustan and Collins 2014). In our data, depending on the year, the fraction of single Black women reporting an occupation was sometimes twice that of single white women (see Table A1). The high proportion of Black women increases estimates of

[19] Consistent with this result, weighting for representativeness matters more for single female estimates than for the other subgroups. Without the weights, single female estimates are 10-25 percent lower than the weighted estimates for the post-emancipation birth cohorts (see Figure A4 for the Census Tree and A5 for the Family Tree). Once again, this result is because the weights increase the Black share of the sample to correct for the lower linking rate of Black women, and increasing the Black share increases IGT estimates.

IGT. This is not because IGT is stronger for single Black women—though that is the case. Rather, it is because greater weight is placed on the between-race persistence of the Black-white gap (Hertz 2008; Jácome, Kuziemko, and Naidu 2021).

If one instead uses the white sample to estimate mobility for single women (Figure 6, Panel B), the gap between single men and women is smaller and disappears by the 1890 birth cohort. While race helps to explain the difference in the estimates between single men and single women, it cannot explain the difference between married men and married women (Figure 6, Panel A). This is because the married estimates rely on the husband's status, and there were no strong differences in labor force participation across Black and white men.

Figure 7 shows more clearly the importance of including Black women in the sample when estimating relative mobility. We run the same process as before in Figure 3, where we show how "traditional" estimates of mobility (i.e., a white-only sample, an occupation-only status measure, no correction for measurement error) change when improving the data to include Black women, to account for within-occupation differences by race/region, and to correct for measurement error when using an IV strategy. Panel A shows the results for married women, which follows a similar pattern we observed before for males in Figure 3. However, a clear difference appears in Panel B: adding Black women increases single-female estimates by about 70-100 percent, whereas it only increases married-female estimates by 23-36 percent in Panel A.

Given the disparities in mobility between single men and single women, overall mobility estimates for the population may change when adding women to the sample. In Figure 8, we start with a white male sample—the one that has been traditionally used in economic history research—and then progressively add white women, Black men, and then finally Black women.[20] Our results for the 1840-1910 cohorts in Figure 7 show that adding white women to the white male sample does slightly increase persistence rates, but not by much. Rather, the larger changes occur when adding Black men to the sample, and then finally Black women. The overall mobility estimates suggest that the IGT decreased from 0.82 for the 1840 birth cohort and changed to 0.66 for the 1910 birth cohort.[21] Similar to Jácome, Kuziemko, and Naidu (2021), we find that samples that do not include Black women will understate persistence.

[20] See Jácome, Kuziemko, and Naidu (2021) for the same process for birth cohorts between 1910-1990.
[21] The finding that adding women to the sample increases IGT estimates (or lowers mobility) holds when using the occupation-only measure of status (See Figure A7). However, once again, the greater change to mobility estimates occurs not when adding women, but when including Black Americans in the sample.

The importance of Black women may be surprising given modern-day evidence that Black-white mobility gaps in individual income are *non-existent* for women, but much larger for men (Chetty et al. 2020). We can use the data to estimate the historical antecedent of these mobility gaps, like Jácome, Kuziemko, and Naidu (2021) and Collins and Wanamaker (2022), but we can extend the data back to the 19th century for both men and women. To estimate mobility gaps, we regress the child's score on the father's score separately by gender and race and plot the gaps for those raised at the 10th percentile (Figure A6).[22] In contrast to today's lack of mobility gap for women, we find a strong one in the past. We also find similar magnitudes for men and women: both groups fell behind their white counterparts by about 12-19 points. This is suggestive evidence that barriers to Black female mobility have fallen since the early 20th century, a point that is also made by Jácome, Kuziemko, and Naidu (2021).

VI. Assortative Mating

The level and trend of intergenerational mobility estimates presented thus far may strongly depend on the amount of assortative mating between men and women. If the rich marry the rich and the poor marry the poor, then cross-sectional inequality will increase and these differences may persist to the next generation (Fernandez and Rogerson 2001; Eika, Mogstad, and Zafar 2019; Ermisch, Francesconi, and Siedler 2006; Collado, Ortuño-Ortín, and Stuhler 2023). As discussed before, we have already found indirect evidence of high assortment since the father-son and father/son-in-law IGT estimates are very similar (Curtis 2021, Clark and Cummins 2022). But this indirect evidence can be improved upon using direct measures available in our dataset.

In this section, we directly estimate assortative mating based on the association between the wife's father and the husband's father's status. Specifically, we regress the husband's father's status on the wife's father's (his father-in-law), using OLS:

$$y_{father} = \gamma_0 + \gamma_1 y_{father-in-law} + \varepsilon_{father} \tag{4}$$

We also produce IV estimates using a second measurement of the father-in-law to instrument for the first.

A benefit of using the parent's generation to estimate assortment is that status is predetermined from the wife and husband's childhood and therefore is not biased by an endogenous

[22] While Chetty et al. (2014, 2020) often focus on mobility gaps at the 25th percentile, we focus on the 10th percentile since the bulk of the Black population was below the 25th percentile in the 19th and early 20th centuries.

cause or effect of marriage on the wife or husband's occupation (Almås et al. 2023). However, we can only estimate assortative mating for the sample of couples where we can observe the father for both. This primarily misses intermarriage with migrants whose fathers are unobserved in the source country. Because intermarriage with the foreign-born tends to reduce assortative mating, we likely overstate assortment for the population as a whole (Bailey and Lin 2022; Eriksson, Lake, and Niemesh 2022).

We estimate a very strong association between the father and father-in-law's status, which suggests a strong degree of assortment in American history. Figure 8 shows the trend in assortative mating estimates based on husband's birth cohort. The trend is similar to the trend for intergenerational mobility, where the association between the father and father-in-law falls from 0.73 for the 1840 birth cohort to 0.63 for the 1910 birth cohort. The fact that both intergenerational transmission and assortative mating fall at the same time hints that falling assortment may be an important mechanism for increasing mobility between 1850 and 1940.[23]

Our assortative mating estimates are at least double others in the recent literature (Althoff, Gray, and Reichardt 2023, Bailey and Lin 2022, Eriksson et al. 2023, Olivetti et al. 2022). Using either marriage records from specific states, social security applications, or pseudo-links across generations, others estimate that the association between the husband's father and father-in-law range between 0.04 and 0.30 over the same period (see Table A2). The difference in estimates could be due to sample construction, but that does not appear to be the primary reason; rather, it is that we account for measurement error using an IV method and adjust for within-occupation differences in status. If we instead follow the approach of others and use an occupational-based measure of status and OLS, we estimate assortative mating is between 0.12 and 0.25 (plotted in Figure A8), which is similar to other estimates. Our higher estimates when using IV are consistent with others who address measurement error in the United Kingdom (Clark and Cummins 2022) and Canada (Curtis 2021).

We measure assortment based on the parents' status, which may be different from assortment based on the wife and husband's status. This is because we do not observe an independent measure of occupation-based status for many wives, and even if we did, we would be

[23] We do not include the mother's status in our main intergenerational equation because we do not have an independent observation of the mother's status. Therefore, the father-son estimates captures both the father "effect" as well as the omitted product of the mother's "effect" and assortative mating. That is, $E[\hat{\beta}_{father}] = \beta_{father} + \beta_{mother}\delta_{AM}$, where δ_{AM} capture assortative mating via a regression of the mother's status on the father's status. If the mother and father effects are constant, then reduced assortative mating mechanically weakens the father-son estimate.

concerned that occupations are endogenously determined in marriage. Prior work has addressed these issues by using educational attainment in the 1940 census as the measure of status (e.g., Bailey and Lin 2022). We can also use this measure to examine the extent to which children match on their own status and that of their fathers.

We do this by going a step further than the standard assortative mating regression and estimate whether the husband's status has an *additional* association with the father-in-law's status on top of the wife's in 1940.[24] Before estimating this augmented regression, we first recreate the typical assortative mating estimate where we regress the wife's education (percentile ranked) on the husband's. This regression yields an estimate of 0.54 (see Table A3). If we additionally control for the wife's father's adjusted Song score, then we find that it is predictive of the husband's education level with a coefficient of 0.12. If we instrument the father-in-law's status with a second measure, then the father-in-law coefficient increases to 0.19. Ultimately the sum of the spouse's coefficient and father-in-law's coefficient is 0.67, which is higher than the original 0.54. These results could imply that the husband matched on both the wife's status and the father-in-law's status. Alternatively, it could indicate that the wife's education is measured with error and the father-in-law's status (instrumented) provides more information about the true status of the wife (Ferrie, Massey, and Rothbaum 2021). Either way, the results indicate that a standard assortative mating regression between wife and husband will understate the true level of assortment.[25]

VII. Conclusion

In this paper, we present the most comprehensive estimates of historical intergenerational mobility and assortative mating in the United States to date. Our estimates rely on a new linked dataset, the Census Tree, that is well beyond the current frontier in terms of the number, quality, and representativeness of the links. For men, these features of the Census Tree allow us to address concerns about the representativeness of mobility estimates (Bailey et al. 2020) and bias due to measurement error (Solon 1992) in prior work. More importantly, because the Census Tree

---

[24] This regression is similar in spirit to a multigenerational regression that checks whether there is additional association with the grandparent's status after conditioning on the parent's (e.g. Braun and Stuhler 2018; Charles, Hurst, & Killewald 2013). However, like how a statistically significant "grandparent effect" could be a true causal effect or a mirage due to measurement error (Solon 2018), a "father-in-law" effect could also be real or due to measurement error in the spouse's education.

[25] Collado, Ortuño-Ortín, and Stuhler (2023) make a similar point that assortative mating is stronger than observed in direct correlations due to additional mating on unobservable characteristics. They estimate that assortative mating between husband and wife is 0.76 when using modern-day data from Sweden.

includes women as well as men, we are able to produce estimates of the transmission of status to daughters as well as to sons. The links for women also allow us to generate new estimates of assortative mating that are highly representative of the population and correct for measurement error in the parent's status.

We generate several new findings. First, we conclude that the larger linked samples in the Census Tree produce estimates for men that are very similar to those from the Census Linking Project, once the samples are weighted to be representative of the population. Second, we find that intergenerational mobility for married women is nearly identical to that for married men—a non-trivial result that is consistent with a high level of assortative mating. Third, we find that single women have much lower mobility than their married counterparts, while the opposite is true for men, leading to large gender gaps in mobility for single people. We show that this is due to high rates of labor force participation and low mobility for single Black women. Fourth, once we use our doubly-linked sample to produce IV estimates of the strength of assortative mating that reduce bias due to measurement error, it appears that assortative mating was much stronger than prior work suggests.

Our paper contributes to a large literature that has attempted to measure historical socioeconomic mobility in the United States. The topic has received so much attention because the careful measurement of mobility is critical for social sciences research on how policies, institutions, and practices interact to make it easier or harder for a person to overcome the circumstances of their birth. Work on these questions has been hindered by disagreement about the facts due to measurement error and the omission of large segments of the population—issues we are able to address using the Census Tree data. In doing so, we find heterogeneity in mobility by race, gender, and marital status that must be taken into account in future work that seeks to understand not only *whether and when* the United States has been a land of opportunity, but *for whom*.

References

Abramitzky, Ran, Leah Boustan, Katherine Eriksson, James Feigenbaum, and Santiago Pérez. "Automated Linking of Historical Data." *Journal of Economic Literature* 59 (2021a): 865–918. https://doi.org/10.1257/jel.20201599.

Abramitzky, Ran; Leah Boustan, Katherine Eriksson, Myera Rashid, Santiago Pérez. "Census Linking Project: 1850-1940." Harvard Dataverse. (2022). https://doi.org/10.7910/DVN/XUXYSR.

Abramitzky, Ran, Leah Boustan, Elisa Jácome, and Santiago Pérez. "Intergenerational Mobility of Immigrants in the United States over Two Centuries." *American Economic Review* 111, no. 2 (2021b): 580-608.

Almås, Ingvild, Andreas Kotsadam, Espen R. Moen, and Knut Røed. "The Economics of Hypergamy." *Journal of Human Resources* 58, no. 1 (2023): 260-281.

Althoff, Lukas, Harriet Brookes Gray, and Hugo Reichardt. "The Missing Link: Women and Intergenerational Mobility." Accessed September 20, 2023. https://hugoreichardt.github.io/pdf/igm_women.pdf

Bailey, Martha J., Connor Cole, Morgan Henderson, and Catherine Massey. "How Well Do Automated Linking Methods Perform? Lessons from US Historical Data." *Journal of Economic Literature* 58 (4) (2020): 997–1044. https://doi.org/10.1257/jel.20191526.

Bailey, Martha J. and Peter Z. Lin. "Marital Matching and Women's Intergenerational Mobility in the Late 19th and Early 20th Century U.S." (2022). Working paper. https://drive.google.com/file/d/1qQjE9j3CW2I6YR9GqtEOXxyNSLcvY9sr/view

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

Boustan, Leah Platt, and William J. Collins. "The Origin and Persistence of Black-White Differences in Women's Labor Force Participation." In *Human Capital in History: The American Record*, pp. 205-240. University of Chicago Press, 2014.

Braun, Sebastian Till, and Jan Stuhler. "The Transmission of Inequality across Multiple Generations: Testing Recent Theories with Evidence from Germany." *The Economic Journal* 128, no. 609 (2018): 576-611.

Buckles, Kasey, Adrian Haws, Joseph Price, and Haley Wilbert. "Breakthroughs in Historical Record Linking Using Genealogy Data: The Census Tree Project." NBER Working Paper no. 31671, 2023. https://doi.org/10.3386/w31671.

Chadwick, Laura, and Gary Solon. "Intergenerational Income Mobility Among Daughters." *American Economic Review* 92 no. 1 (2002): 335–44. https://doi.org/10.1257/000282802760015766.

Charles, Kerwin K., Eric Hurst, and Alexandra Killewald. "Marital Sorting and Parental Wealth." *Demography,* 50 no. 1 (2013): pp. 51-70.

Chetty, Raj, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter. "Race and Economic Opportunity in the United States: An Intergenerational Perspective." *The Quarterly Journal of Economics* 135, no. 2 (2020): 711-783.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility." *American Economic Review* 104 no. 5 (2014): 141–47. https://doi.org/10.1257/aer.104.5.141.

Clark, Greg and Neil Cummins. Assortative Mating and the Industrial Revolution: England, 1754-2021. (2022).

Collado, M. Dolores, Ignacio Ortuño-Ortín, and Jan Stuhler. "Estimating Intergenerational and Assortative Processes in Extended Family Data." *The Review of Economic Studies* 90, no. 3 (2023): 1195-1227.

Collins, William J., and Marianne H. Wanamaker. "African American Intergenerational Economic Mobility since 1880." *American Economic Journal: Applied Economics* 14 no. 3 (2022): 84–117. https://doi.org/10.1257/app.20170656.

Corak, Miles. "Income Inequality, Equality of Opportunity, and Intergenerational Mobility." *Journal of Economic Perspectives* 27 no. 3 (2013): 79–102. https://doi.org/10.1257/jep.27.3.79.

Curtis, M., *The her in inheritance: Matching and mobility in two million Quebec marriages, 1800-1970*. Working Paper, Université Libre de Bruxelles. (2021).

Eika, Lasse, Magne Mogstad, and Basit Zafar. "Educational Assortative Mating and Household Income Inequality." *Journal of Political Economy* 127, no. 6 (2019): 2795-2835.

Elliott, Diana B., Kristy Krivickas, Matthew W. Brault, and Rose M. Kreider. Historical Marriage Trends from 1890-2010: A Focus on Race Differences. In *Annual Meeting at the Population of America. San Francisco, CA*, pp. 012-12. (2012).

Eriksson, Katherine, Gregory T. Niemesh, Myera Rashid, and Jacqueline Craig. "Marriage and the Intergenerational Mobility of Women: Evidence from Marriage Certificates 1850-1920." Accessed October 2023. https://conference.nber.org/conf_papers/f191084.pdf

Eriksson, Katherine, Addison Lake, and Gregory T. Niemesh. "Immigration Shocks and Marriage Market Sorting." In *AEA Papers and Proceedings*, vol. 112 (2022): pp. 214-218.

Ermisch, John, Marco Francesconi, and Thomas Siedler. "Intergenerational Mobility and Marital Sorting." *The Economic Journal* 116, no. 513 (2006): 659-679.

Feigenbaum, J., A Machine Learning Approach to Census Record Linking (2016).

Feigenbaum, J.J., "Multiple Measures of Historical Intergenerational Mobility: Iowa 1915 to 1940." *The Economic Journal*, *128*(612) (2018): pp.F446-F481.

Fernández, Raquel, and Richard Rogerson. "Sorting and Long-Run Inequality." *The Quarterly Journal of Economics* 116, no. 4 (2001): 1305-1341.

Ferrie, Joseph, Catherine Massey, and Jonathan Rothbaum. "Do Grandparents Matter? Multigenerational Mobility in the United States, 1940–2015." *Journal of Labor Economics* 39, no. 3 (2021): 597-637.

Ganzeboom, Harry B. G., Donald J. Treiman, and Wout C. Ultee. 1991. "Comparative Intergenerational Stratification Research: Three Generations and Beyond." *Annual Review of Sociology* 17 (1): 277–302. https://doi.org/10.1146/annurev.so.17.080191.001425.

Goldin, Claudia. "Female Labor Force Participation: The Origin of Black and White Differences, 1870 and 1880. *The Journal of Economic History*, *37* no. 1 (1977): pp.87-108.

Helgertz, Jonas, Steven Ruggles, John Robert Warren, Catherine A. Fitch, David Hacker, Matt Nelson, Joseph Price, Evan Roberts, and Matthew Sobek. "IPUMS Multigenerational Longitudianal Panel: Version 1.1 [Dataset]." (2023). Minneapolis, MN: IPUMS. https://doi.org/10.18128/D016.V1.1.

Hertz, Tom. "A Group-Specific Measure of Intergenerational Persistence." *Economics Letters* 100, no. 3 (2008): 415-417.

Jácome, Elisa, Ilyana Kuziemko, and Suresh Naidu. "Mobility for All: Representative Intergenerational Mobility Estimates over the 20th Century." NBER Working Paper no. 29282, 2021. https://doi.org/10.3386/w29289.

Kambourov, G. and Manovskii, I. "Rising Occupational and Industry Mobility in the United States: 1968–97. *International Economic Review*, *49* no. 1 (2008): pp.41-79.

Long, Jason, and Joseph Ferrie. "Intergenerational Occupational Mobility in Great Britain and the United States Since 1850." *American Economic Review* 103, no. 4 (2013): 1109-1137.

Margo, Robert A. "Obama, Katrina, and the Persistence of Racial Inequality." *The Journal of Economic History* 76, no. 2 (2016): 301-341.

Mare, Robert D. "Five Decades of Dducational Assortative Mating." *American Sociological Review,* (1991): 15-32.

Mazumder, Bhashkar. "Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data. *Review of Economics and Statistics*, *87,* no. 2 (2005): pp.235-255.

Mazumder, Bhashkar. "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 no. 1 (2018): 213–34.

Moscarini, Giuseppe and Kaj Thomsson. "Occupational and Job Mobility in the US." *The Scandinavian Journal of Economics*, *109* no. 4, (2007): pp.807-836.

Nybom, Martin, and Jan Stuhler. "Heterogeneous Jncome Profiles and Lifecycle Bias in Intergenerational Mobility Estimation." *Journal of Human Resources* 51, no. 1 (2016): 239-268.

Olivetti, Claudia, and M. Daniele Paserman. "In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1940." *American Economic Review* 105, no. 8 (2015): 2695–2724. https://doi.org/10.1257/aer.20130821.

Olivetti, Claudia, M. Daniele Paserman, Laura Salisbury, and E. Anna Weber. *Who married,(to) whom, and where? Trends in marriage in the United States, 1850-1940*. 2022. http://www.claudiaolivetti.website/uploads/6/8/0/0/68001935/opsw_nov2022.pdf

Price, Joseph, Kasey Buckles, Adrian Haws, and Haley Wilbert. The Census Tree, 1850-1940. Ann Arbor, MI:

Inter-university Consortium for Political and Social Research [distributor]. (2023). https://doi.org/10.3886/E193225V1.

Price, Joseph, Kasey Buckles, Jacob Van Leeuwen, and Isaac Riley. "Combining Family History and Machine Learning to Link Historical Records: The Census Tree Data Set." *Explorations in Economic History* 80 (2021): 101391. https://doi.org/10.1016/j.eeh.2021.101391.

Saavedra, Martin, and Tate Twinam. "A Machine Learning Approach to Improving Occupational Income Scores." *Explorations in Economic History,* 75 (2020): 101304. https://doi.org/10.1016/j.eeh.2019.101304.

Santavirta, Torsten, and Jan Stuhler. "*Name-Based Estimators of Intergenerational Mobility.*" Forthcoming at *The Economic Journal.* (2023). https://janstuhler.files.wordpress.com/2023/01/ej_revised_manuscript.pdf

Solon, Gary. "Intergenerational Income Mobility in the United States." *The American Economic Review* 82, no. 3 (1992): 393–408.

———. "Cross-Country Differences in Intergenerational Earnings Mobility." *Journal of Economic Perspectives*, 16 no. 3 (2002): 59–66. https://doi.org/10.1257/089533002760278712.

———. "What Do We Know So Far about Multigenerational Mobility?." *The Economic Journal* 128, no. 612 (2018): F340-F352.

Song, Xi, Catherine G. Massey, Karen A. Rolf, Joseph P. Ferrie, Jonathan L. Rothbaum, and Yu Xie. "Long-Term Decline in Intergenerational Mobility in the United States since the 1850s." *Proceedings of the National Academy of Sciences* 117, no. 1 (2020): 251–58. https://doi.org/10.1073/pnas.1905094116.

Torche, Florencia. "Analyses of Intergenerational Mobility: An Interdisciplinary Review." *The ANNALS of the American Academy of Political and Social Science* 657, no. 1 (2015): 37–62. https://doi.org/10.1177/0002716214547476.

Vom Lehn, Christian, Cache Ellsworth, and Zachary Kroff. "Reconciling Occupational Mobility in the Current Population Survey. *Journal of Labor Economics* 40, no. 4 (2022): pp.1005-1051.

Ward, Zachary. "Intergenerational Mobility in American History: Accounting for Race and Measurement Error." Forthcoming, *American Economic Review* (2023). https://drive.google.com/file/d/1drVpb0NRqWPB6Zt6RP6TfMLQ7UavhJVy/view.

Withrow, Jennifer. "Three Essays on Labor and Marriage Markets: Farm Crisis and Rural-to-Urban Migration in the United States, 1920-1940." (2021).

Table 1: Observations in the primary, singly-linked, and doubly-linked Census Tree samples

		Single Linking		Double Linking	
Census Year	# Children with an Adjusted Song for the Father	Also has Adult Adjusted Song for Self or Husband	Percent Retained	Also has a Second Adjusted Song for the Father	Percent Retained
1850	6,978,173	2,755,518	39.49%	2,054,868	29.45%
1860	8,578,939	3,169,721	36.95%	2,809,186	32.75%
1870	12,245,005	4,200,813	34.31%	3,721,330	30.39%
1880	16,437,894	8,219,165	50.00%	6,799,859	41.37%
1900	20,903,258	10,716,158	51.27%	9,371,161	44.83%
1910	23,060,478	8,689,324	37.68%	7,782,712	33.75%
Total	88,203,747	37,750,699	42.80%	32,539,116	36.89%

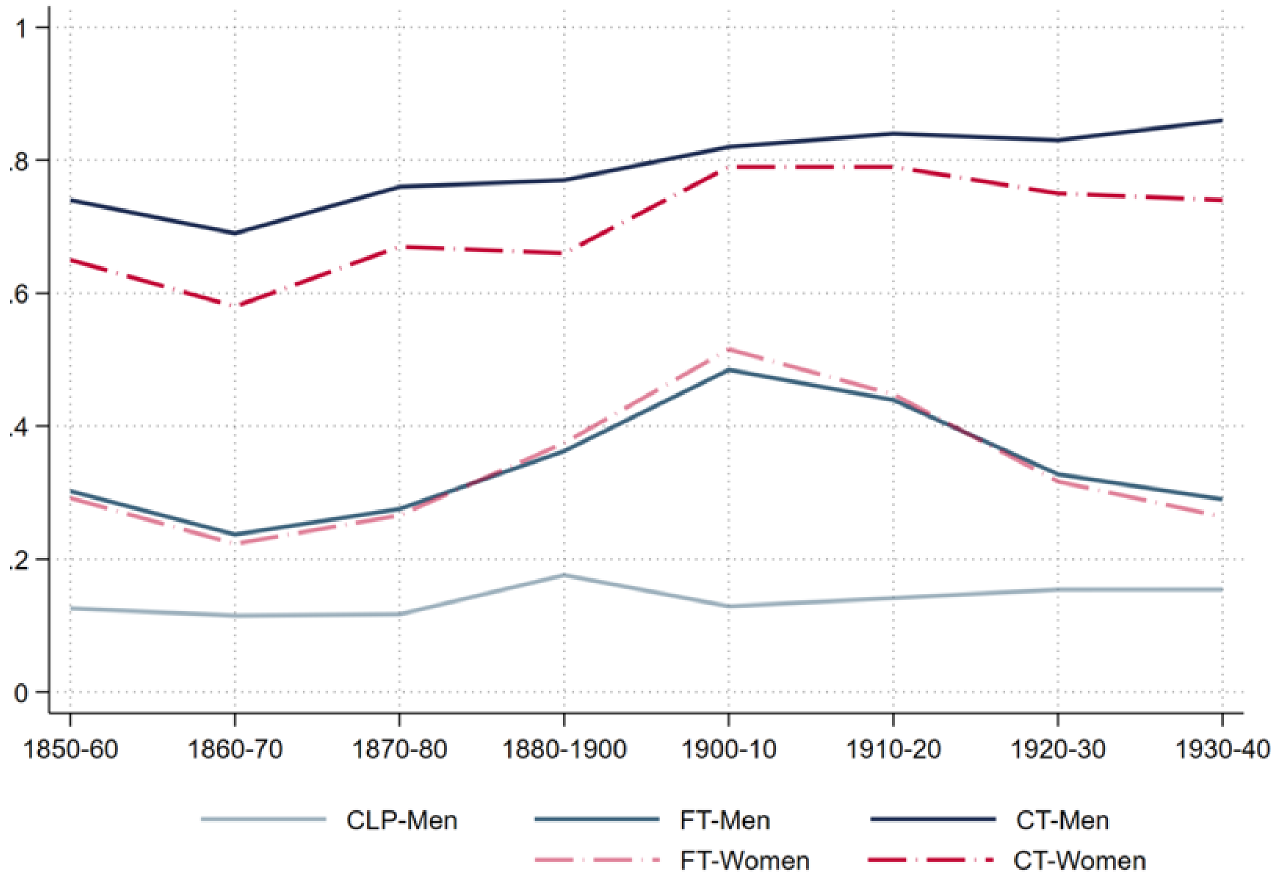
Notes: Column one lists the number of children age 0-14 in each census year for whom an adjusted Song score can be created for the father (the primary observation). In column two, we keep those for whom we are able to link an adult adjusted Song score when the child is between the ages of 25 and 55; for married women, we use their husband's score (the singly-linked sample). The third column is the percent of the primary sample that can be singly-linked. In the fourth column, we keep those from the singly-linked same that can also be linked to a second observation with an adjusted Song score for the father (the doubly-linked sample). The final column reports the percent of the primary sample that can be doubly-linked.

Table 2: Descriptive statistics of the linked intergenerational sample

Birth Cohort	1840	1850	1860	1870	1880	1890	1900	1910
Panel A: Child Characteristics								
Black	10.9	13.0	9.4	10.4	11.8	9.7	9.3	9.0
	(31.1)	(33.6)	(29.1)	(30.5)	(32.2)	(29.7)	(29.1)	(28.6)
Age	37.0	36.6	44.6	39.5	36.1	41.1	37.7	30.0
	(5.0)	(9.7)	(7.2)	(6.8)	(5.6)	(7.2)	(5.2)	(3.1)
Men	56.5	57.5	59.0	56.5	53.4	55.6	51.1	52.8
	(49.5)	(49.4)	(49.1)	(49.5)	(49.8)	(49.6)	(49.9)	(49.9)
Adjusted Song Score	52.7	50.3	52.6	52.6	53.1	53.0	51.8	50.1
	(28.6)	(29.1)	(28.5)	(28.7)	(29.0)	(28.0)	(27.7)	(27.3)
Married	86.6	79.8	81.1	78.6	74.4	80.5	80.8	73.0
	(34.0)	(40.0)	(39.1)	(40.9)	(43.6)	(39.6)	(39.3)	(44.3)
White Collar	16.2	16.2	20.7	23.1	25.9	29.1	32.1	31.2
	(36.8)	(36.8)	(40.5)	(42.1)	(43.8)	(45.4)	(46.6)	(46.3)
Farmer	42.4	38.1	35.8	29.9	25.4	19.9	15.7	11.6
	(49.4)	(48.5)	(47.9)	(45.7)	(43.5)	(39.9)	(36.4)	(32.0)
Unskilled	28.8	33.6	29.7	32.2	33.7	33.5	35.7	43.5
	(45.3)	(47.2)	(45.7)	(46.7)	(47.2)	(47.2)	(47.9)	(49.5)
Skilled	12.4	11.9	13.6	14.7	14.8	17.3	16.4	13.5
	(33.0)	(32.4)	(34.3)	(35.4)	(35.5)	(37.8)	(37.0)	(34.2)
Panel B: Father Characteristics								
Age	41.0	39.5	40.5	40.0	36.8	41.4	38.0	37.8
	(6.8)	(7.3)	(7.1)	(7.2)	(7.2)	(6.5)	(7.1)	(7.1)
Adjusted Song Score	51.0	48.9	51.1	50.2	48.6	47.6	47.4	47.5
	(27.5)	(27.9)	(26.3)	(26.7)	(27.2)	(27.1)	(26.9)	(27.0)
White Collar	8.0	9.0	9.3	10.3	10.8	13.0	14.5	16.5
	(27.2)	(28.7)	(29.0)	(30.4)	(31.0)	(33.6)	(35.2)	(37.1)
Farmer	59.6	54.0	50.3	49.8	47.8	45.0	42.4	37.9
	(49.0)	(49.8)	(49.9)	(49.9)	(49.9)	(49.7)	(49.4)	(48.5)
Unskilled	17.4	20.8	26.3	26.8	28.5	26.8	27.8	29.1
	(37.9)	(40.6)	(44.0)	(44.3)	(45.1)	(44.2)	(44.8)	(45.4)
Skilled	14.7	16.0	14.0	12.9	12.8	15.1	15.0	16.4
	(35.5)	(36.6)	(34.7)	(33.6)	(33.4)	(35.8)	(35.7)	(37.0)
N	1,185,094	1,960,058	2,824,783	4,107,865	2,933,468	5,468,975	7,347,848	6,683,435

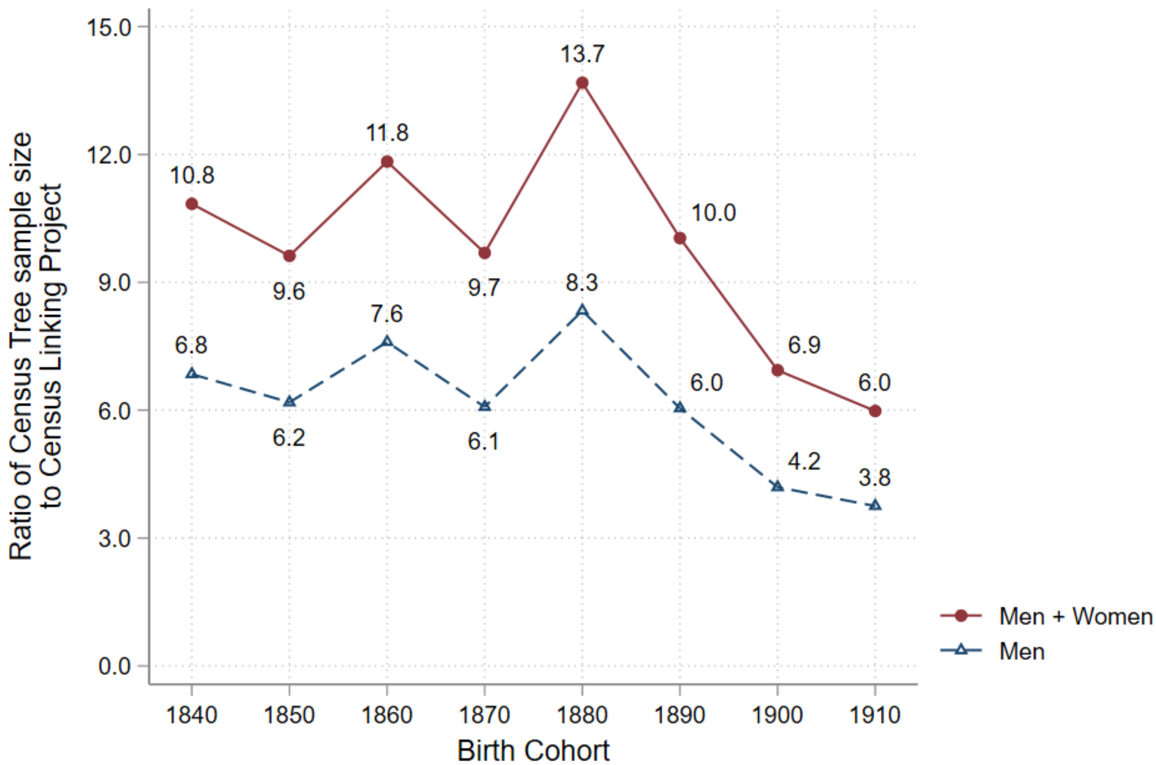
Notes: This table shows the descriptive statistics of the sample. Each column reports statistics by the child's birth cohort, rounded to the nearest decade. The table is weighted to be representative on observables following the process of Bailey et al. (2020). Occupation categories are based on occ1950 codes from IPUMS, where white collar jobs include occ1950 codes that start with 0, 2, 3 or 4. Farmers are codes that start with 1. Unskilled are codes that start with 6, 7, 8 or 9 (excluding non-occupation codes above 970). Skilled occupations are codes that start with 5. If a daughter is married, the occupation and Adjusted Song score is that of her husband's. If a daughter is single, then it is her own occupation.

Figure 1: Comparing match rates between adjacent censuses in the Census Linking Project, Family Tree, and Census Tree samples



Notes: Figure shows match rates obtained between adjacent censuses by the Census Linking Project (exact conservative matches), the Family Tree, and the Census Tree, by gender. The Census Linking Project only includes men. The 1890 census was destroyed by a fire, so we include the 1880-1900 census pair. Match rates are constructed following the procedure described in Price et al. (2021).

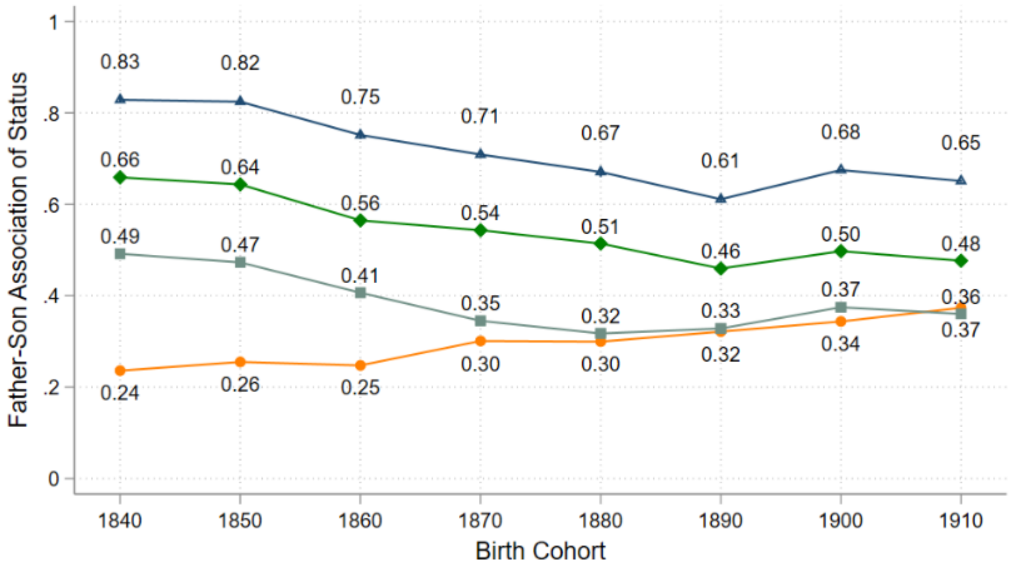
Figure 2: The Census Tree yields larger sample sizes compared to the Census Linking Project when double linking.



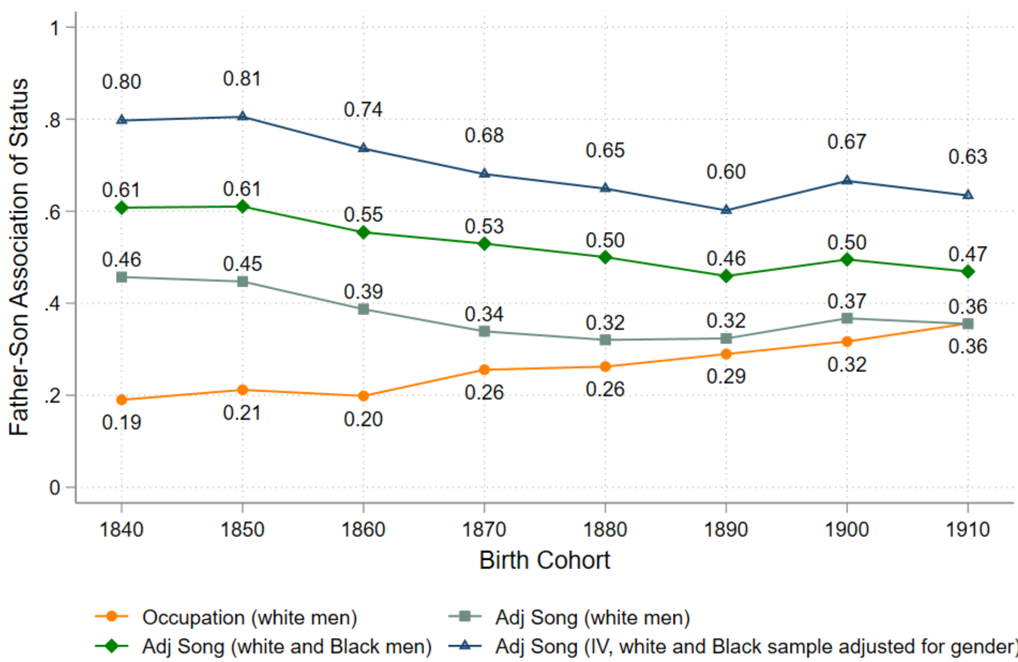
Notes: This figure shows the ratio of sample sizes when using Census Tree links or Census Linking Project links (exact-conservative) to create an intergenerational dataset. On average, there are five times more male links in the Census Tree intergenerational dataset, and eight time more male and female links.

Figure 3: Estimates of intergenerational mobility for men in 1840-1910 birth cohorts

Panel A: Replication of Ward (2023), using Census Linking Project data

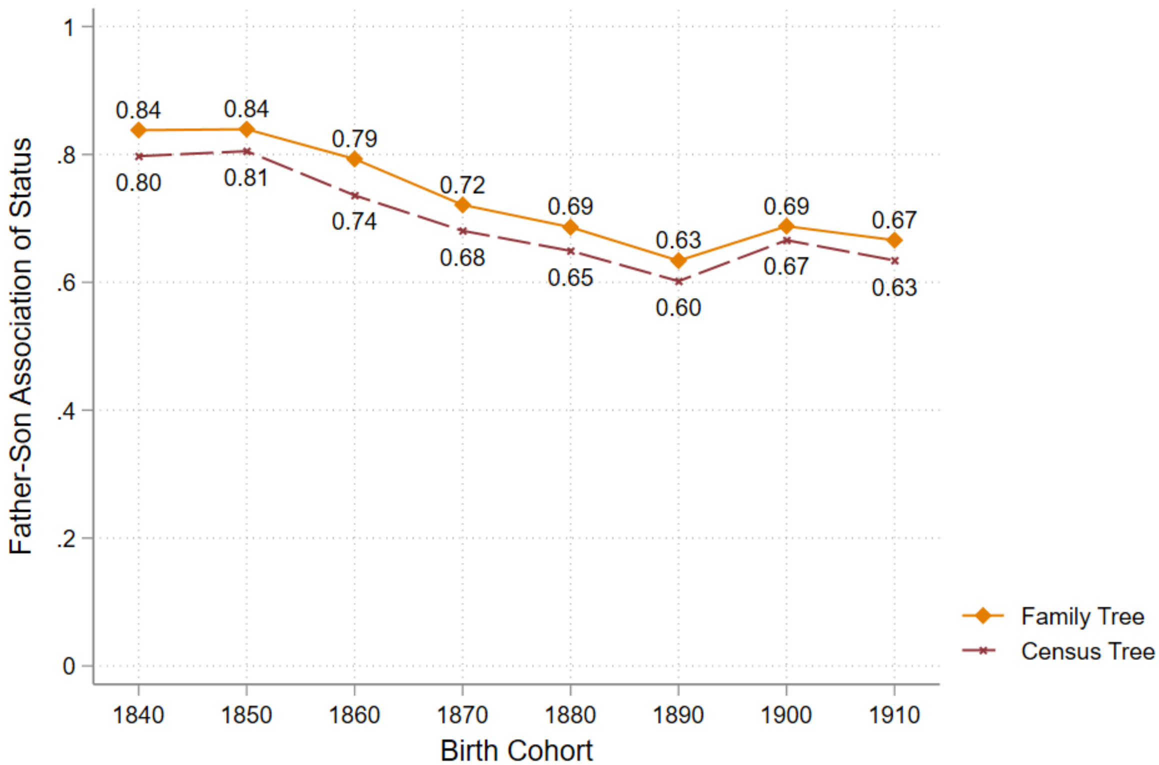


Panel B: Replication using Census Tree data



Notes: Figures show coefficients from a regression of the son's status on the father's. Panel A uses links from the Census Linking Project, while Panel B uses links from the Census Tree. The top panel has 3,915,677 father-son links in the sample (3,787,802 white and 127,875 Black), while the bottom panel has 20,073,579 (19,039,041 white and 1,034,538 Black). The bottom line "Occupation (white males)" uses the occupation score from Song et al. (2020) for a sample of white males. The next line uses a score that allows for within-occupation differences by race and region (Ward 2023). The third line adds Black males to the sample. The final line instruments the first father observation with a second. All estimates are weighted to be representative of the population following the procedure outlined by Bailey et al. (2020).

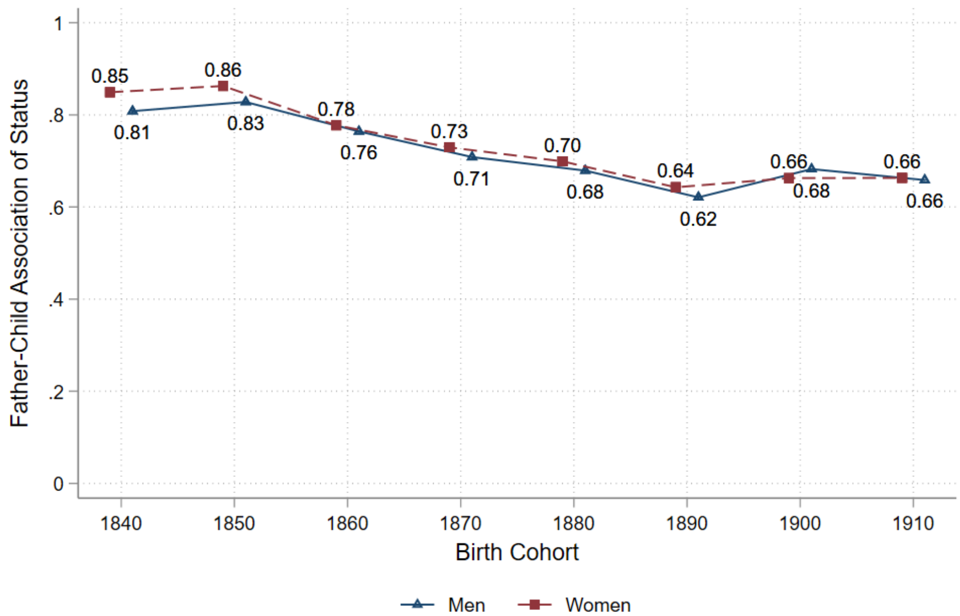
Figure 4: Comparing male mobility estimates from the Family Tree and Census Tree datasets



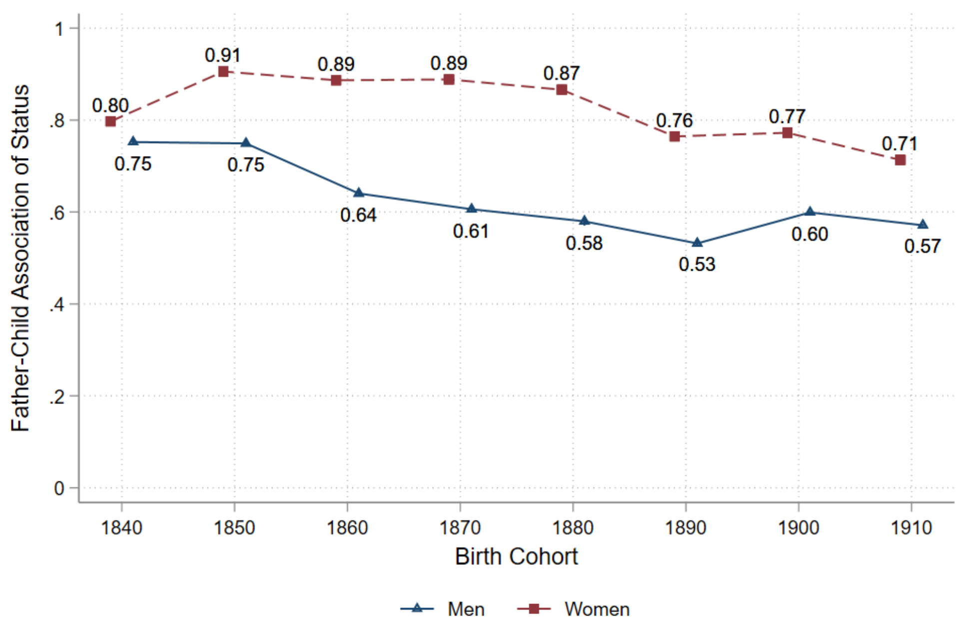
Notes: The Census Tree series reproduces the topline series from Figure 2, Panel B. The Family Tree series (represented by diamonds) shows estimates using only links from the Family Tree. All estimates are weighted to be representative of the population following the procedure outlined by Bailey et al. (2020). The total number of observations in the Census Tree and Family Tree series are 19,074,136 and 8,781,622, respectively.

Figure 5: Intergenerational mobility estimates for men and women

Panel A: Married



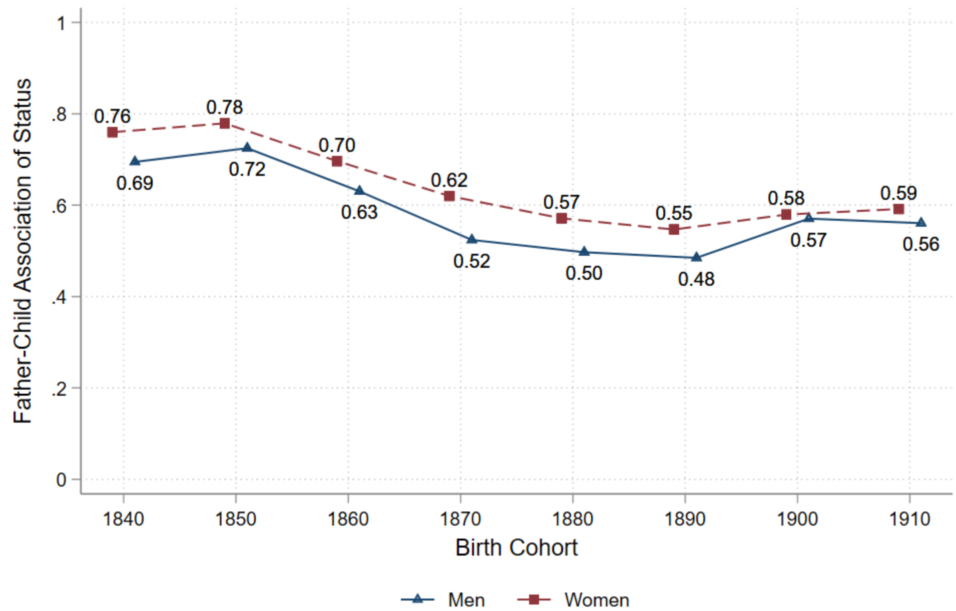
Panel B: Single



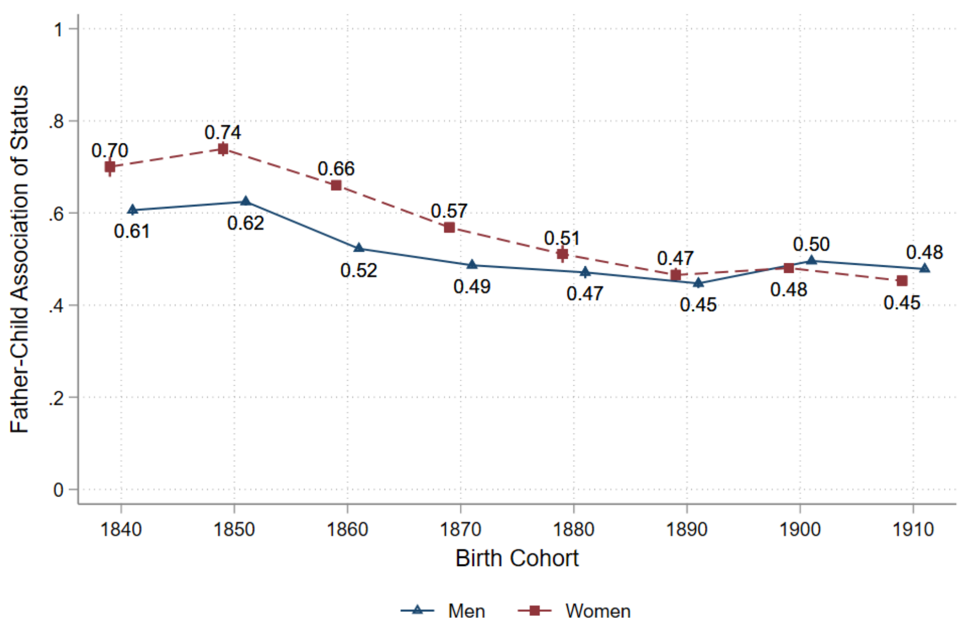
Notes: Panel A shows mobility estimates for married people, from an IV regression of the son's (or son-in-law's, in the case of women) adjusted Song score on the father's, using a second father's observation as an instrument for the first. Panel B shows the same for single people, but the woman's own occupation is used as the dependent variable. Estimates are weighted to be representative of the population following the procedure outlined by Bailey et al. (2020).

Figure 6: Intergenerational mobility estimates for white men and women

Panel A: Married



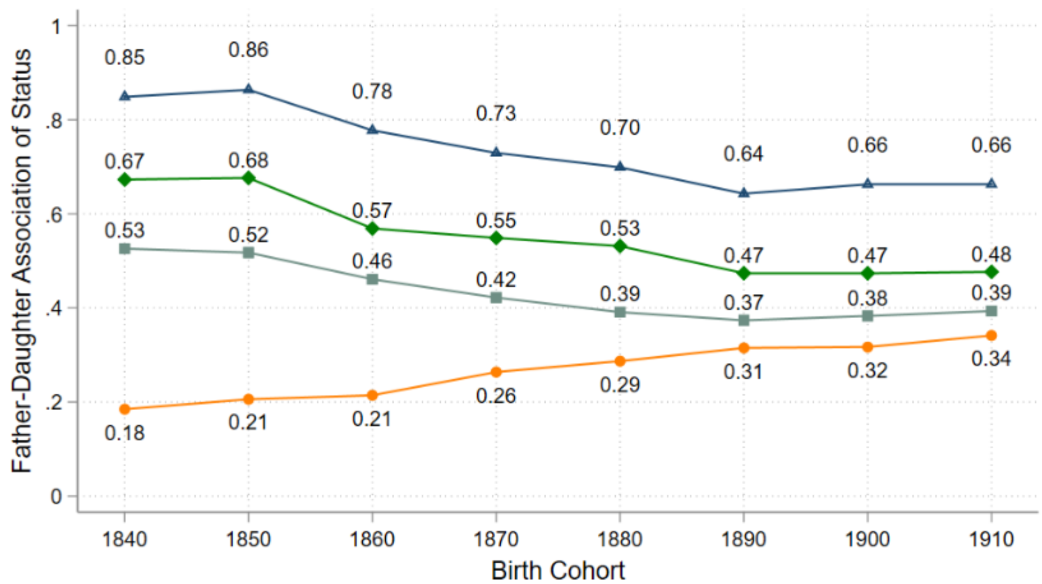
Panel B: Single



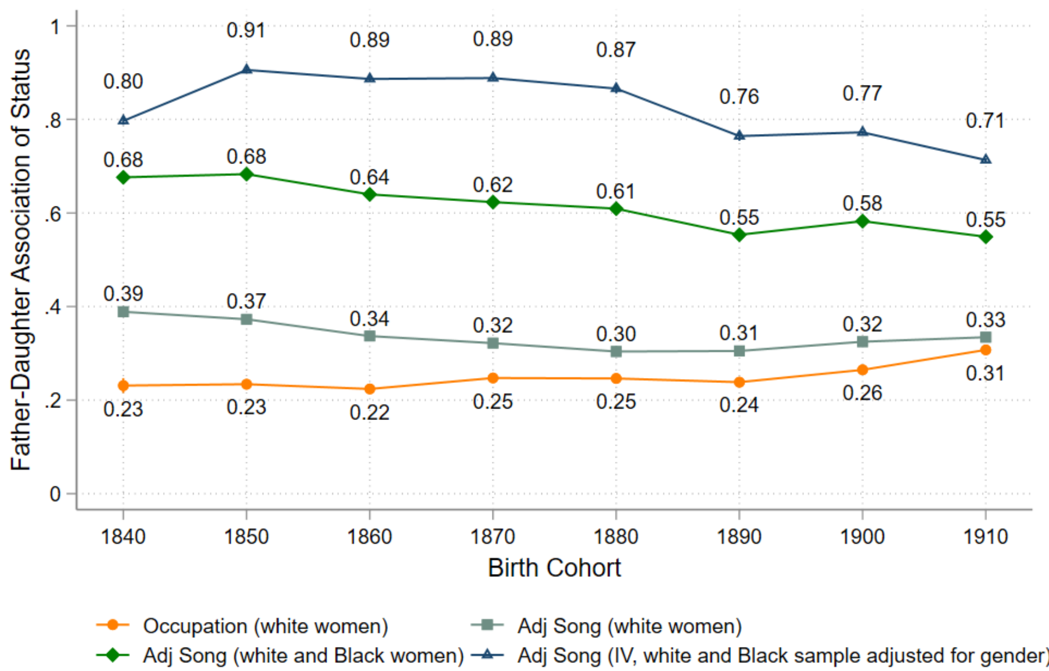
Notes: Panel A shows mobility estimates for married people, from an IV regression of the son's (or son-in-law's, in the case of women) adjusted Song score on the father's, using a second father's observation as an instrument for the first. Panel B shows the same for single people, but the woman's own occupation is used as the dependent variable. The estimates for the figure are the same as in Figure 5, but this figure limits the sample to the white population. Estimates are weighted to be representative of the population following the procedure outlined by Bailey et al. (2020).

Figure 7: Estimates of intergenerational mobility for women in 1840-1910 birth cohorts

Panel A: Married Women

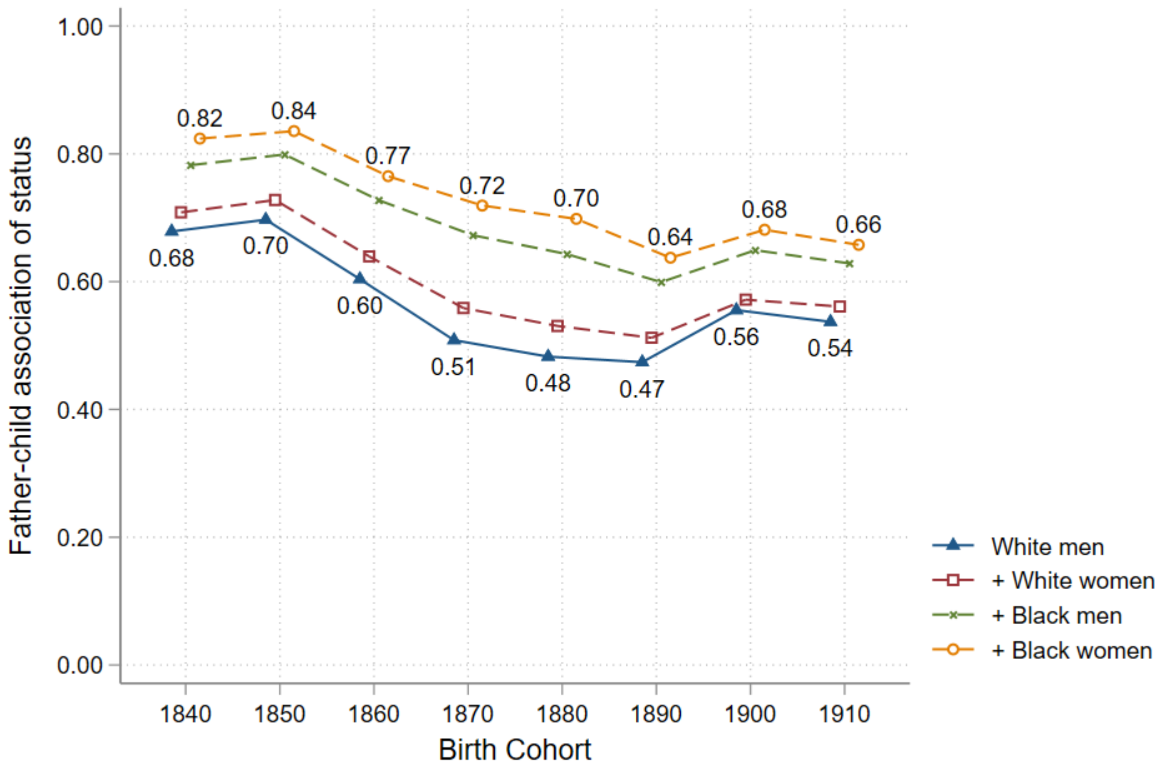


Panel B: Single Women



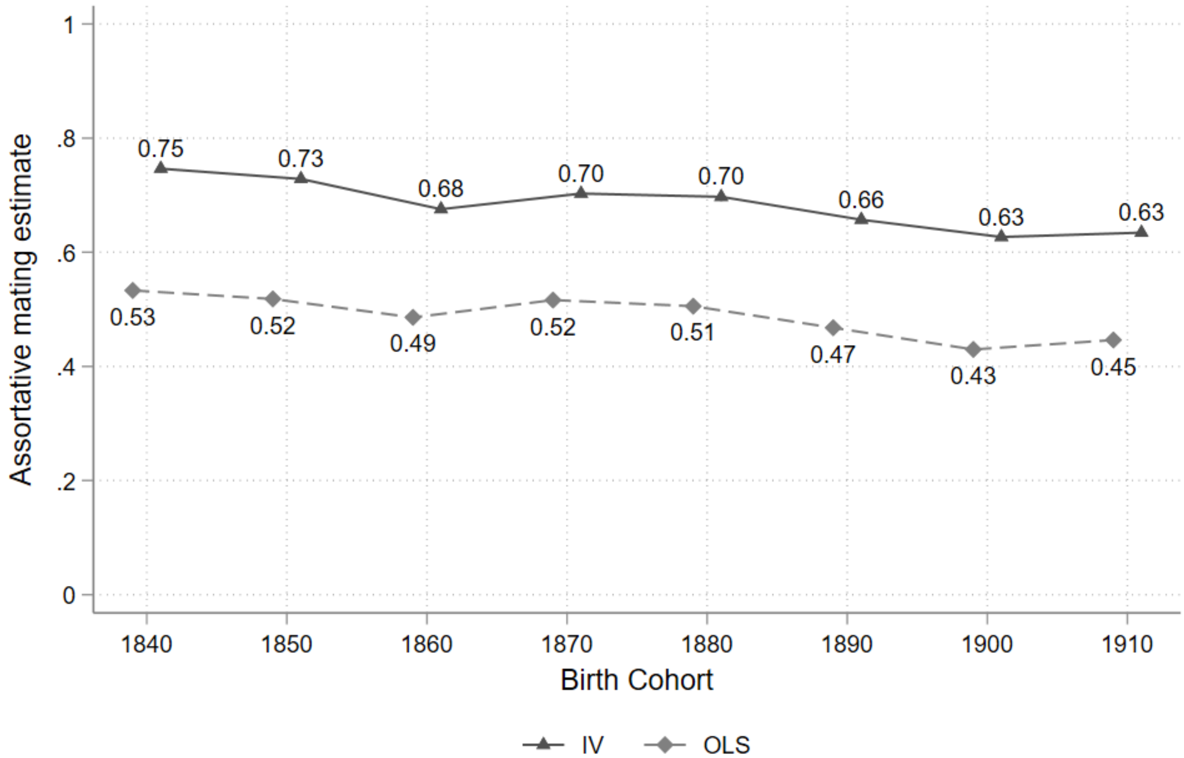
Notes: Figures show coefficients from a regression of the daughter's status on the father's. Panel A uses links for married women, while Panel B uses links for single women. The bottom line "Occupation (white females)" uses the occupation score from Song et al. (2020) for a sample of white males. The next line uses a score that allows for within-occupation differences by race and region (Ward 2023). The third line adds Black females to the sample. The final line instruments the first father observation with a second. All estimates are weighted to be representative of the population following the procedure outlined by Bailey et al. (2020).

Figure 8: Including women in the sample raises estimates of intergenerational transmission.



Notes: This figure shows how mobility estimates change when adding more groups to the sample. The bottom line is a sample of only white men; the line above that pools white women into the sample; the line above that adds Black men into the sample; and the top line finally adds Black women to get population-level estimates. The point estimate is the coefficient on the father's status level after instrumenting one father observation with a second. The child's own status is the dependent variable for single people and married men, the husband's status is the dependent variable for married women. Estimates are weighted to be representative of the population following the procedure outlined by Bailey et al. (2020).

Figure 9: Assortative mating was strong.



Notes: Coefficients are from regressions of the husband's father's status on the wife's father's status, for married couples where both are observed. The measure of status is the adjusted Song score. The IV regressions instrument the primary observation for the wife's father's status with second observation. There are 3,245,572 total observations in the regressions. Estimates are weighted to be representative of the population following the procedure outlined by Bailey et al. (2020).

Online Appendix

Table A1. Single women reporting an occupation in the 1870-1940 censuses, by race

Year	Total	White Women	Black Women
1870	33.13	28.47	57.49
	(1,796,921)	(1,508,154)	(288,767)
1880	43.08	38.45	69.53
	(2,212,261)	(1,882,458)	(329,803)
1900	49.08	45.30	74.53
	(3,809,920)	(3,316,614)	(493,306)
1910	57.83	54.26	82.17
	(4,590,541)	(4,003,920)	(586,621)
1920	59.64	57.32	76.78
	(5,275,881)	(4,647,240)	(628,641)
1930	62.60	60.55	75.43
	(6,233,578)	(5,373,863)	(859,715)
1940	64.06	63.42	67.96
	(7,499,546)	(6,445,519)	(1,054,027)

Notes: Sample is restricted to women age 25-55 who are not currently married. Each column gives the percent of women in the sample for whom an own adjusted Song score can be constructed. The number of observations is reported in parentheses.

Table A2. Other estimates of assortative mating in the literature

Paper	Estimate	Status measure	Estimation	Population
Buckles et al. (2023)	0.63-0.73	Occ, race, region	IV	White and Black
Buckles et al. (2023)	0.43-0.52	Occ, race, region	OLS	White and Black
Buckles et al. (2023)	0.25-0.42	Occ	IV	White and Black
Buckles et al. (2023)	0.12-0.25	Occ	OLS	White and Black
Althoff et al. (2023)	0.20-0.30	Occ	OLS	Social security apps linked to census
Bailey and Lin (2022)	0.25-0.35	Occ	OLS	Ohio marriages
Eriksson et al. (2023)	0.20-0.29	Occ	OLS	White MA marriages
Olivetti et al. (2022)	0.04-0.09	Occ	First-name method	White, first-name method

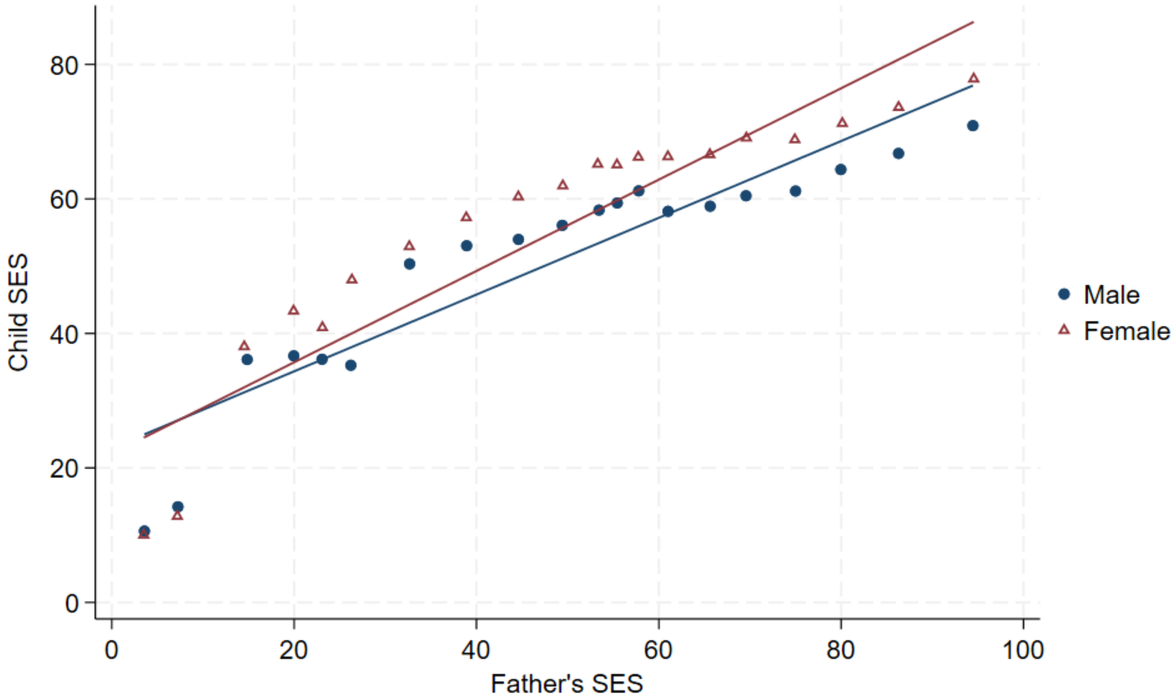
Notes: Buckles et al. (2023) is the current paper. The estimates are taken from Figure 12 in Althoff et al. (2023), Figure E.1 in Bailey and Lin (2022), Figure 4 in Eriksson et al. (2023), and Figure 2 in Olivetti et al. (2022).

Table A3. Assortative mating estimates based on 1940 husband's education

	(1)	(2)	(3)
Wife's years of education percentile rank	0.546	0.505	0.482
(0-100)	(0.00111)	(0.00113)	(0.000648)
Wife's Father's adj. Song score		0.119	0.187
		(0.000996)	(0.000892)
Estimation type	OLS	OLS	IV
Observations	1,674,154	1,674,154	1,674,154
R-squared	0.349	0.366	0.360

Notes: The sample is the linked intergenerational sample where education is observed for both the wife and husband in 1940. The dependent variable is the percentile rank of the husband's years of education. We percentile rank to make the range of the education variable and the father and father-in-law's status measure the same. Estimation type "IV" in column (3) uses an instrumental variables strategy where the wife's father's status is instrumented with a second observation. This table shows that husband's education was simultaneously associated with the wife's education and her father's education level.

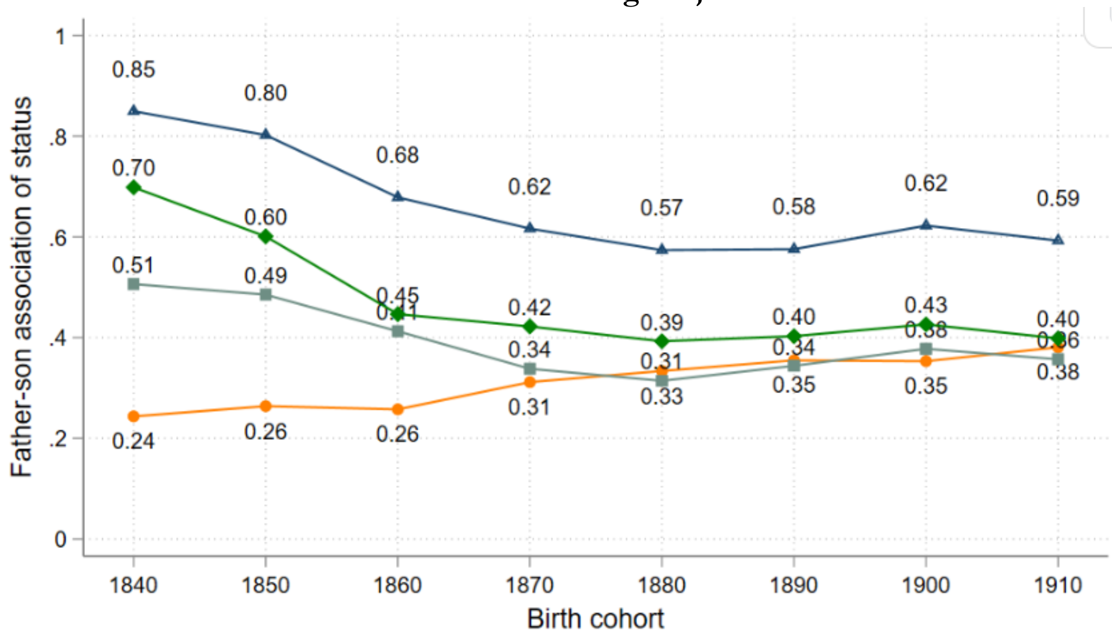
Figure A1. Binscatter plot for mobility estimates by gender



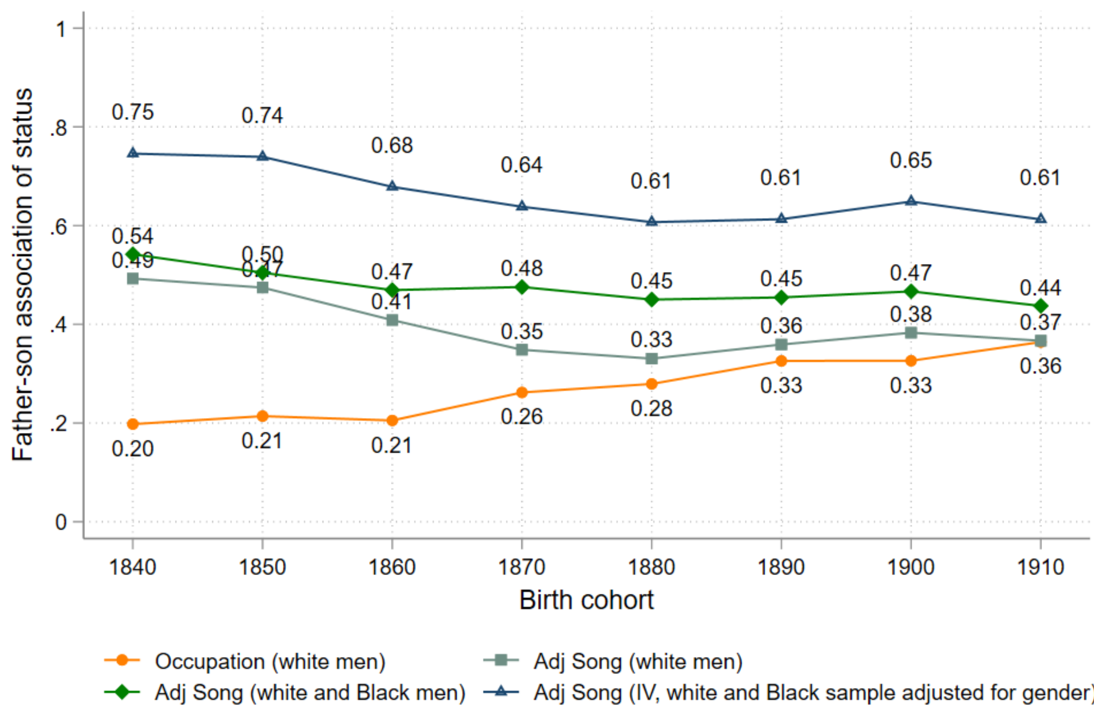
Notes: This figure shows the binscatter relationship between the child's adjusted Song score and the father's adjusted Song Score. Female's score is their own if they are single, but their husbands if they are married. Estimates are weighted by the custom sample weights.

Figure A2: Unweighted estimates of intergenerational mobility for men

Panel A: Census Linking Project data

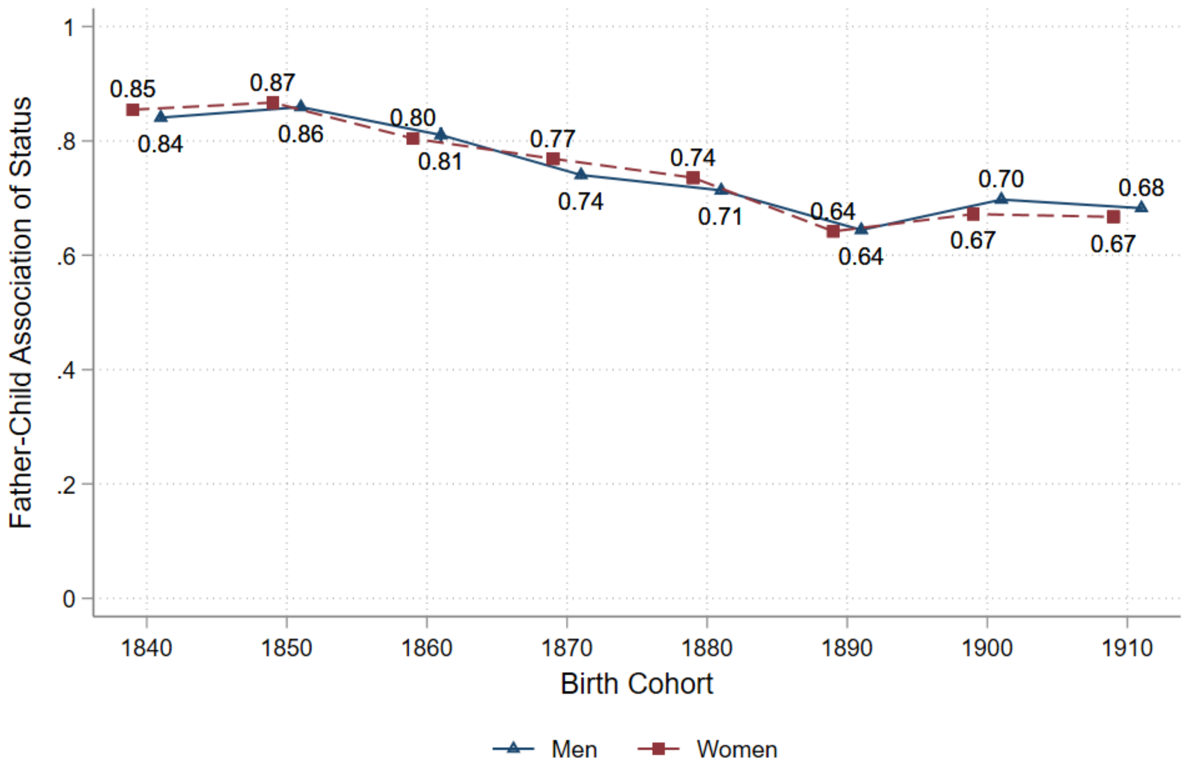


Panel B: Replication using Census Tree data



Notes: Figures show coefficients from a regression of the son's status on the father's. Panel A uses links from the Census Linking Project, while Panel B uses links from the Census Tree. This figure recreates estimates from Figure 3, but does not use weighted data.

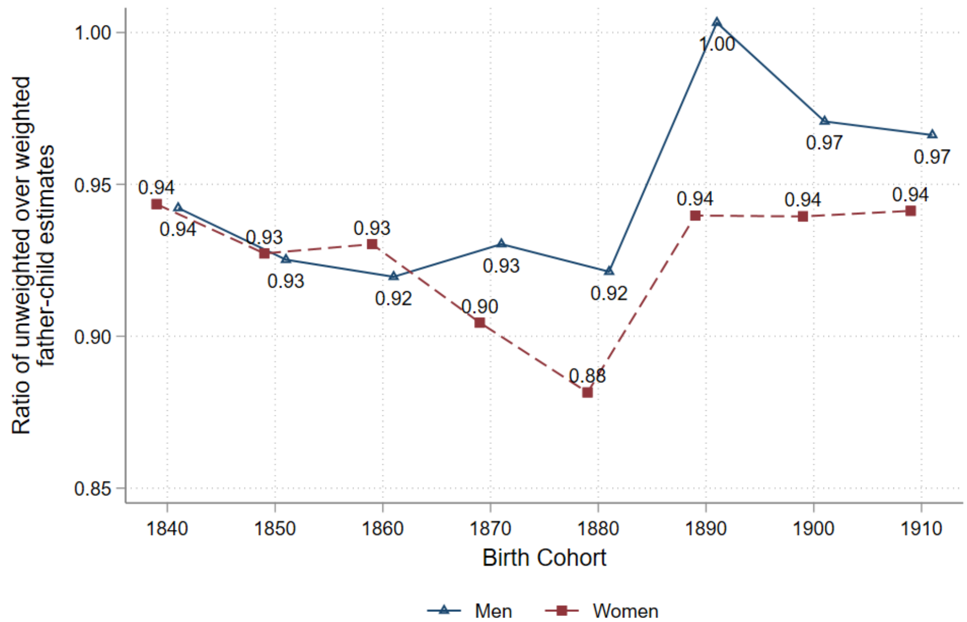
Figure A3: Intergenerational mobility estimates for married men and women, Family Tree links



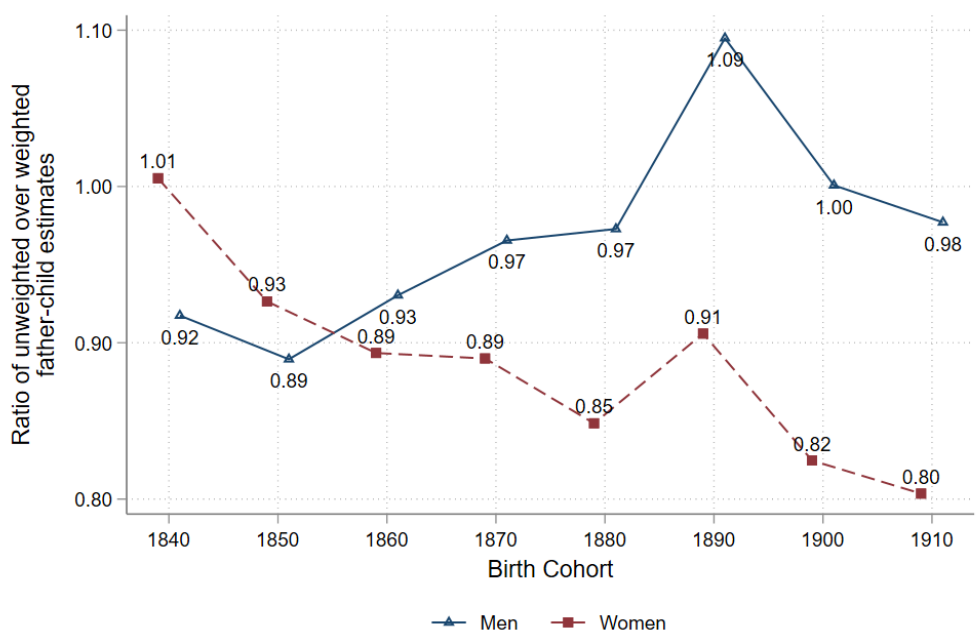
Notes: This figure shows mobility estimates for married people, from an IV regression of the son's (or son-in-law's, in the case of women) adjusted Song score on the father's, using a second father's observation as an instrument for the first. It is a replication of Figure 5A, but when using links from the Family Tree instead of the Census Tree.

Figure A4: Unweighted intergenerational estimates by marital status, Census Tree

Panel A. Married



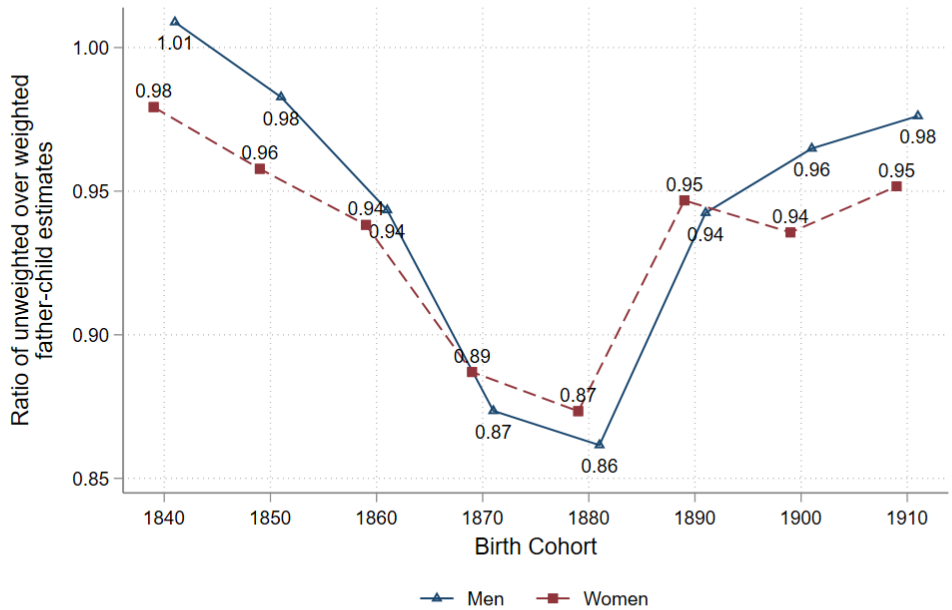
Panel B. Single



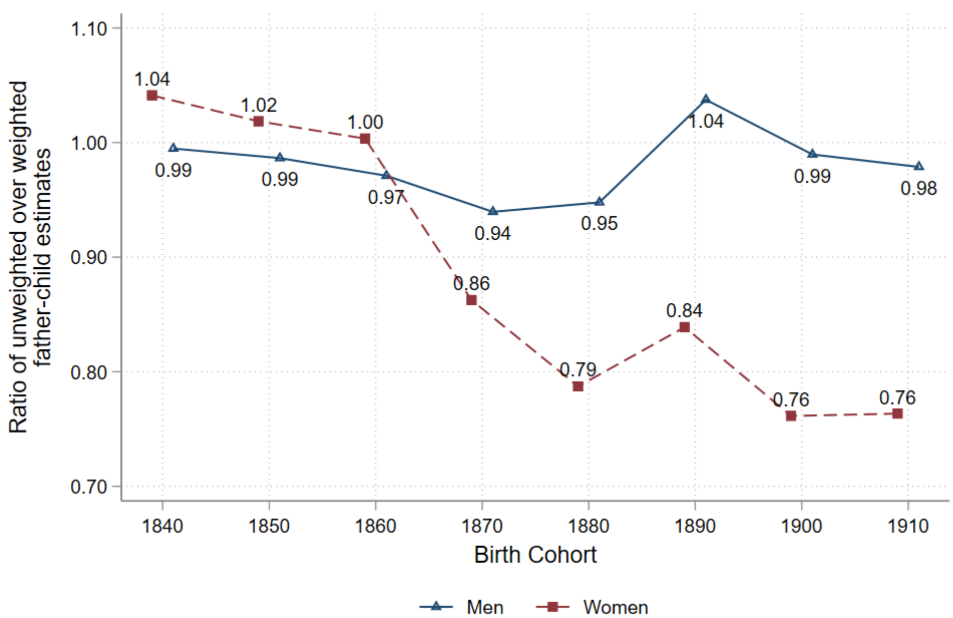
Notes: This figure shows the ratio of intergenerational transmission estimates when using an unweighted sample over the weighted sample. The links are based on the Census Tree. The weighted point estimates are shown in Figure 5. This figure shows that unweighted transmission estimates are mostly lower than weighted estimates.

Figure A5: Unweighted intergenerational estimates by marital status, Family Tree

Panel A. Married



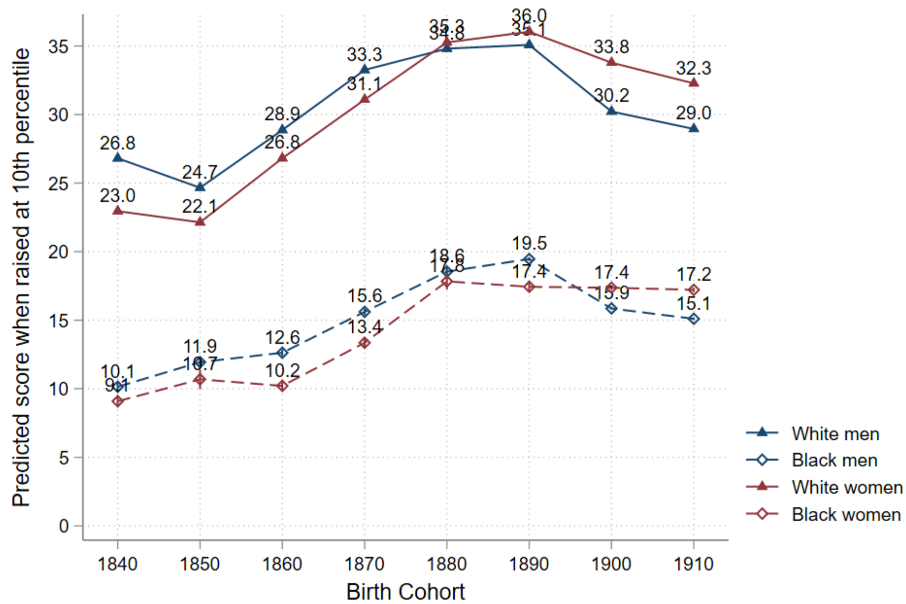
Panel B. Single



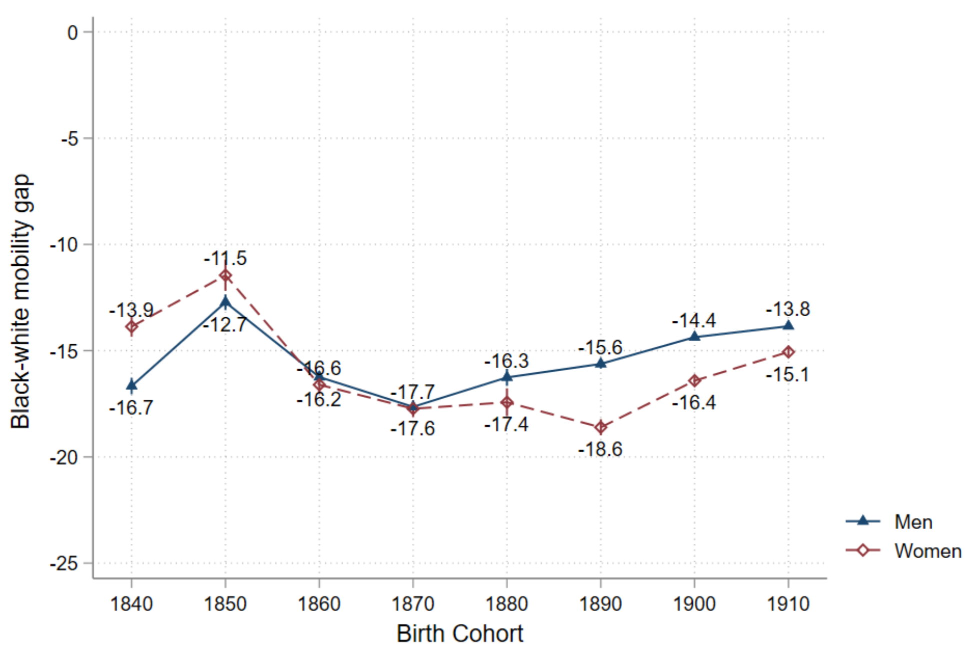
Notes: This figure shows the ratio of intergenerational transmission estimates when using an unweighted sample over the weighted sample. The links are based on the Family Tree. The weighted point estimates are shown in Figure A3.

Figure A6: Black-white mobility gaps at the 10th percentile, by gender

Panel A: Predicted levels for children raised at the 10th percentile

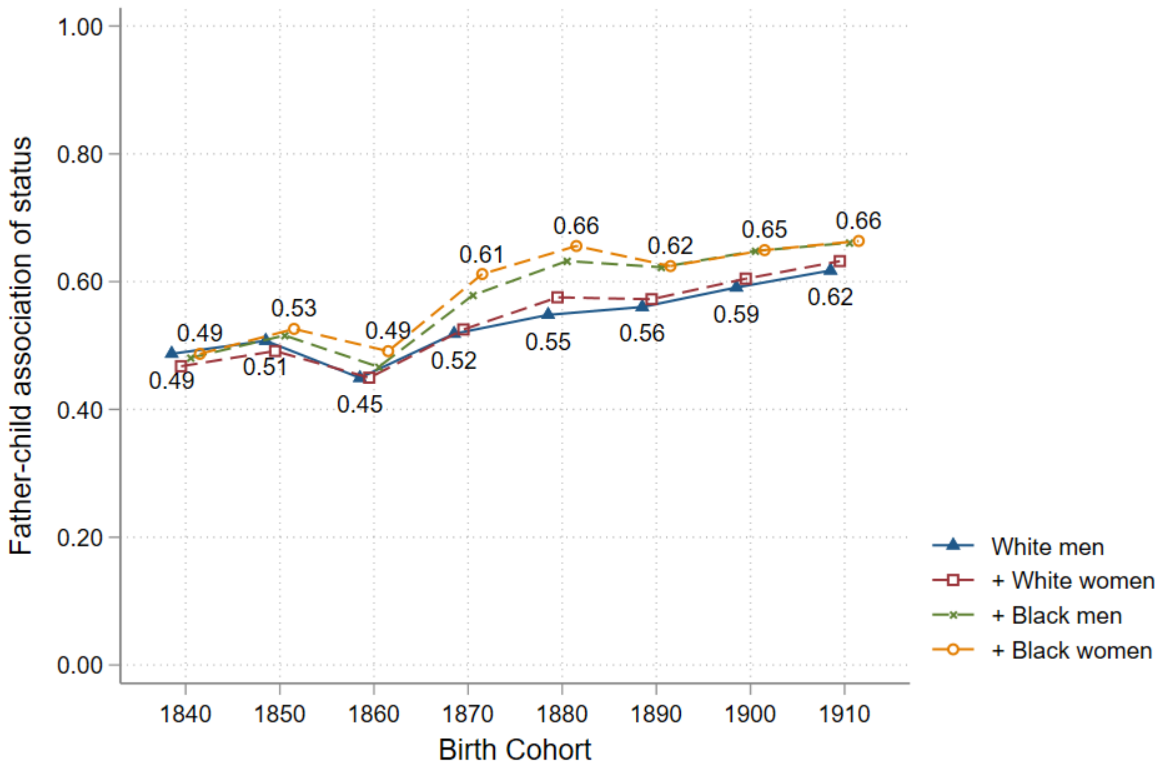


Panel B: Difference in mobility gaps by gender



Notes: Panel A shows the results of a regression of the child's adjusted Song score on their father's, and plots the predicted outcomes for a child raised at the 10th percentile. This regression is run separately by race and gender. Panel B plots the difference between Black and white's outcomes shown in Panel A.

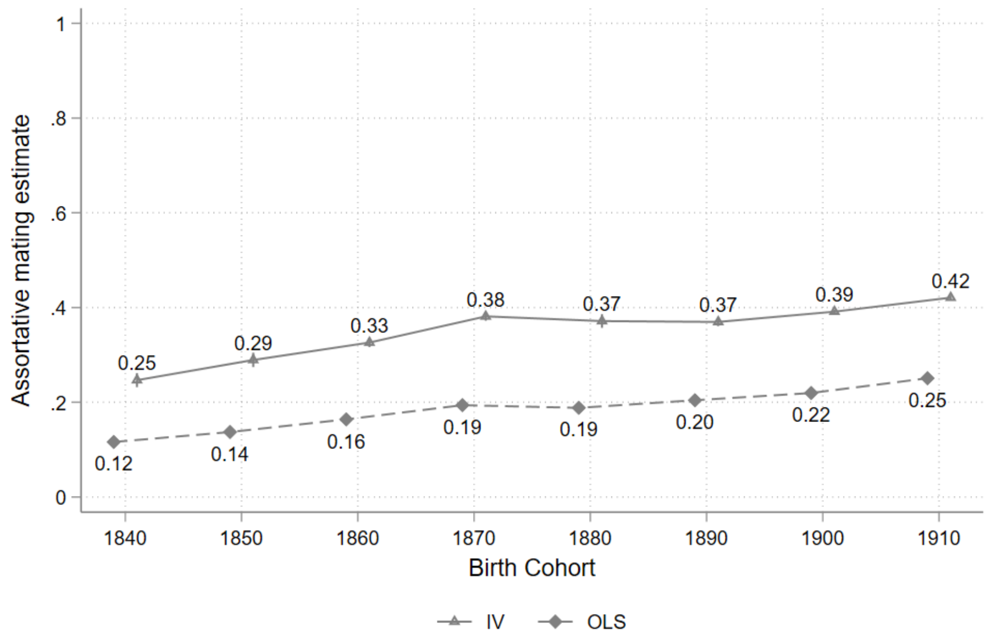
Figure A7: Including women in the sample slightly raises estimates of intergenerational transmission (occupation-only score)



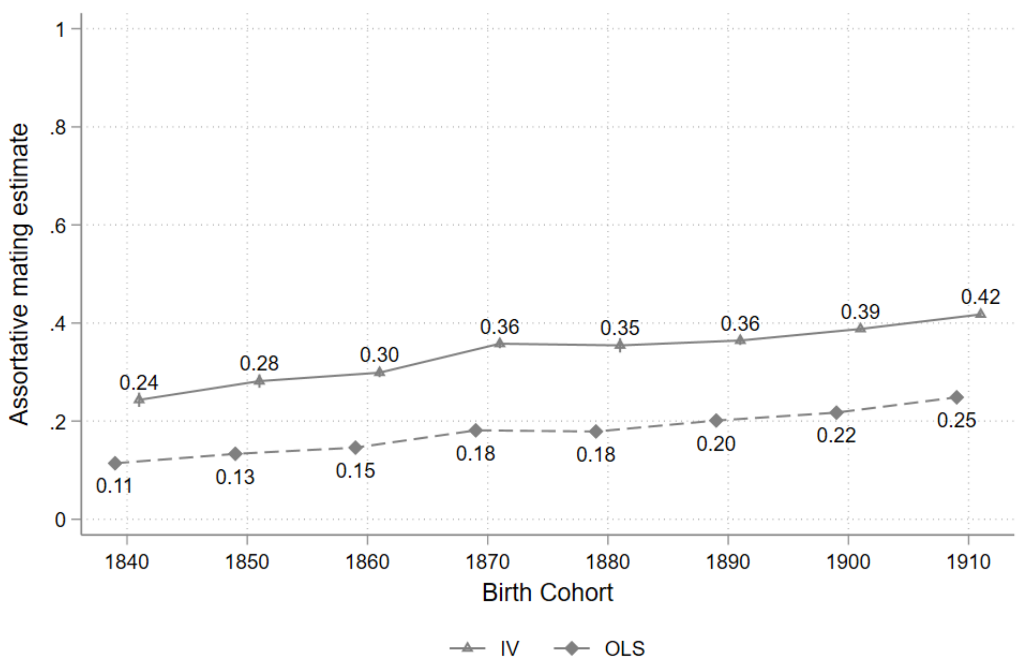
Notes: This figure is analogous to Figure 8, but with the occupation-only Song score (Song et al. 2020) as the measure of socioeconomic status.

Figure A8. Assortative mating estimates when using occupation-only status

Panel A. Black and white



Panel B. White



Notes: These figures estimate assortative mating by regressing the husband's father's status on the wife's fathers. The method is same as in Figure 8, but instead of using an adjusted status measure that allows for within-occupation differences by race and region ("adjusted Song Score"), we use an occupation-only measure of status ("Song score") (Song et al. 2020).