NBER WORKING PAPER SERIES

FAMILY TREES AND FALLING APPLES:
HISTORICAL INTERGENERATIONAL MOBILITY ESTIMATES FOR WOMEN AND MEN

Kasey Buckles
Joseph Price
Zachary Ward
Haley E.B. Wilbert

Working Paper 31918
http://www.nber.org/papers/w31918

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2023, Revised November 2024

This work is supported by grants from the Russell Sage Foundation (G-1063) and the National Science Foundation (SES-2049762). We are grateful to excellent research assistants at the University of Notre Dame and the BYU Record Linking Lab for their work on this project, including Brianna Felegi, Adrian Haws, Andrew Lugo, and Nate McGhie. We have benefited from feedback at many seminars and conferences, and from conversations with Martha Bailey, Greg Clark, Bill Evans, Greg Niemesh, Claudia Olivetti, Seth Sanders, Lowell Taylor, and Jan Stuhler. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Family Trees and Falling Apples: Historical Intergenerational Mobility Estimates for Women and Men
Kasey Buckles, Joseph Price, Zachary Ward, and Haley E.B. Wilbert
NBER Working Paper No. 31918
November 2023, Revised November 2024
JEL No. I30, J0, N0, N01

ABSTRACT

Efforts to document long-term trends in socioeconomic mobility in the US have been hindered by the lack of large, representative datasets linking parents to adult children, a challenge especially acute for women due to surname changes. We use a new dataset, the Census Tree, which incorporates genealogy data to link tens of millions of fathers to their sons and daughters in historical US censuses. We find that relative mobility was remarkably similar across sons and daughters in the 1835 to 1915 birth cohorts. Additionally, assortative mating was much stronger than previously estimated, as daughters married husbands with very similar backgrounds.

Kasey Buckles
University of Notre Dame
Department of Economics
3052 Jenkins-Nanovic Halls
Notre Dame, IN 46556
and NBER
kbuckles@nd.edu

Joseph Price
Department of Economics
Brigham Young University
2129 WVB
Provo, UT 84602
and NBER
joseph_price@byu.edu

Zachary Ward
Department of Economics
Baylor University
One Bear Place
Waco, TX 76798
and NBER
zach.a.ward@gmail.com

Haley E.B. Wilbert
University of Notre Dame
Department of Economics
3052 Jenkins-Nanovic Halls
Notre Dame, IN 46556
hwilbert@nd.edu

A Information and data access for the Census Tree. is available at https://censustree.org

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. Unfortunately, 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.

There are two particular challenges when attempting to measure the intergenerational transmission of socioeconomic status (SES) for women. The first is that because many women change their surnames when they marry, it can be difficult to link information about a woman's childhood with her adult outcomes. In modern datasets, these links can sometimes be made using unique identifiers constructed from social security numbers (Chetty et al. 2014, Wagner and Layne 2014, Ferrie et al. 2021), but these are not available in historical data. As a result, most previous efforts to create census links on a large scale have omitted women entirely (e.g., Abramitzky et al. 2022, Feigenbaum 2016). To overcome this, we use data from a new linked dataset, the Census Tree (Buckles et al. 2024, Price et al. 2021). Many of the census-to-census links in the Census Tree are created by users of one of the world's largest online genealogy platforms, FamilySearch.org. These users have private information including maiden names that allows them to create links between a daughter's childhood and adult homes that would be impossible for a trained research assistant or a machine-learning algorithm to make. The Census Tree is unprecedented in its size, containing over 700 million census-to-census links—nearly half of which are for women.

The second challenge to estimating historical mobility for women is that most women did not report an occupation in datasets like the census (and in many cases were not asked), making it difficult to use this to proxy their economic rank in the same way that prior work has done for men. Those who did report occupations likely represent a selected group; for instance, Black women had much higher labor force participation rates than White women over the past 150 years (Goldin 1977, Boustan and Collins 2014), and single women were more likely to report an occupation than married women. To address this challenge, we introduce a novel approach. Instead of relying on one's own occupation to measure SES, we use the occupations of one's nearby neighbors to create a measure of "neighborhood status." This method is possible because census enumerators walked door to door to record households, such that households listed sequentially were close by (Logan and Parman 2017). Importantly, this measure can be created for all individuals, regardless of gender and whether they reported an occupation. It also allows for variation in status within occupations, and better reflects the combined resources of the household than relying on the occupation of the household head alone.

With these innovations, we present the most comprehensive picture of historical intergenerational mobility in the United States to date. We estimate the intergenerational transmission (IGT) of socioeconomic status for men and women born between 1835 and 1915 by measuring the strength of the relationship between the child's neighborhood status and their father's economic status, as proxied with occupation, race, and location (similar to Abramitzky et al. 2021, Collins and Wanamaker 2022, and Ward 2023). The Census Tree's extensive set of links includes at least two observations for the father for the majority of the children in our sample, which allows us to instrument the father's economic status with a second observation to account for measurement error when using a single snapshot of the father (Modalsli and Vosters 2024, Ward 2023). Our final estimation sample contains 21.0 million sons and 14.4 million daughters.

Our key finding is that the level and trend of intergenerational mobility were strikingly similar for women and men across the entire period. Our IV estimates of the transmission of status range from 0.58-0.69 for women and from 0.55-0.67 for men. For both genders, intergenerational mobility increased—or IGT decreased—across the 19th to early 20th century birth cohorts, suggesting a gradual improvement in relative mobility over time. The similarity of mobility estimates for men and women holds across a variety of specifications. For instance, while our preferred approach uses an IV method to account for measurement error, we observe similar patterns for male and female mobility when using OLS. The consistency remains when we use educational attainment as a proxy for the child's status (only available in the 1940 Census), and when we drop fathers who are farmers. Additionally, we observe similar transmission levels across gender when estimating intergenerational transmission by father's state of residence, by race, or by ethnicity. Overall, the results consistently show that fathers' status was almost equally predictive of sons' and daughters' outcomes.

The similarity of mobility estimates across sons and daughters is surprising given the literature's emphasis on parental investment and childhood environment as the primary drivers of mobility patterns (e.g., Becker et al. 2018, Chetty and Hendren 2018). In the past, a daughter's status was usually determined by that of her husband, but husbands were (hopefully) not raised in the same household as the daughters and thus did not benefit from the same parental investments. Given this, one would expect imperfect assortative mating, or "regression to the mean in marriage" (Becker and Tomes 1985), to weaken the link between a daughter's status and that of her parents. Our finding that transmission rates are the same—or even slightly higher—between fathers and daughters as between fathers and sons suggests that assortative mating was very high throughout American history, with daughters marrying men who had similar childhood conditions to their own. Put another way, it appears that assortative mating is so strong that it is as if women are marrying their brothers. Yet, at the same time, most historical estimates of assortative mating in the United States (Bailey and Lin 2024, Espín-Sanchez et al. 2023, Eriksson et al. 2023, Olivetti et al. forthcoming) are too low to be consistent with the similarity between our father-son and father-daughter mobility estimates, raising a puzzle over the true extent of assortative mating in American history. Fortunately, with our data, we can directly estimate marital sorting based on the status of the wife's and husband's fathers' economic status.

We create new estimates of assortative mating based on the strength of association between a wife's father's status and her father-in-law's status. We estimate that assortative mating was very strong, with associations ranging from 0.64 to 0.76 for children born between the 1840 and 1910 birth cohorts. These associations decreased over time, similar to our transmission estimates between father and child. Note that the associations between father and father-in-law may imply an even higher association between a wife and husband's latent status as the fathers are one generation removed from the wife and husband, but since we do not observe the wife's latent status, we cannot estimate this association directly. We find higher estimates of assortative mating than others in the literature for two reasons: (1) we account for measurement error using an IV strategy and (2) we use a measure of economic status that allows for differences within occupation by race and region. Without these two changes, we find much lower rates of assortment. However, a high rate of assortment helps to understand the similarity in our mobility estimates for married men and women (Chadwick and Solon 2002) and is consistent with others who estimate high degrees of assortative mating when accounting for latent status or measurement error in other contexts (Althoff et al. 2024, Clark and Cummins 2022; Collado, Ortuño-Ortín, and Stuhler 2023; Curtis 2021).

We conclude our paper by comparing our findings to other work on historical mobility. As most prior work uses a measure of the child's status that is based on their own occupation (or their husband's, in the case of married women), we reproduce our results using this more traditional measure. We show that while our use of the neighborhood status measure yields lower point estimates of the intergenerational transmission coefficients, the overall patterns are the same. Additionally, when we replicate the pseudo-linking method proposed by Olivetti and Paserman (2015) but maintain the same population (Black and White) and measures (children's neighborhood status), male and female mobility estimates remain similar, though the pseudo-linking approach estimates less mobility than our direct links. Finally, we combine our estimates of mobility by gender for the 1835-1915 birth cohorts with those of Jácome, Kuziemko, and Naidu (forthcoming), who produce similar estimates for the 1910-1970 birth cohorts. This allows us to create the longest-ever time series of historical mobility estimates by gender for the United States, spanning 130 years. Throughout this entire period, mobility follows similar patterns for men and women, and transmission to daughters is at least as strong as transmission to sons for all cohorts.

II. Data: The Census Tree

Our analysis relies on data from the Census Tree, which is the largest existing set of linked records 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. (2024), but we provide a summary here (see also Price et al. 2021).

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. Relevant for our purposes, this information includes 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.

To create the Census Tree, Buckles et al. (2024) build on the Family Tree links in two ways. First, a subset of these links is extracted 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.[2] 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."[3] 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.[4]

The full Census Tree has over 391 million links for men and over 314 million links for women. For men, the number of links between adjacent censuses is between 3.4 and 4.7 times larger than the CLP, and match rates are over 80% for the censuses in the 20th century. For women, match rates are over 75% between adjacent censuses in the 20th century, and well over 50% in the 19th century. These high match rates are critical to our empirical strategy for two reasons. First, they allow us to create tens of millions of intergenerational links between fathers and their adult sons and daughters. Second, we are able to find most fathers in multiple censuses, which helps address issues with measurement error in the father's socioeconomic status. [5] In the next section, we elaborate on the advantages of the Census Tree for our estimation strategy and provide summary statistics for our estimation sample.

III. Estimating the Intergenerational Transmission of Status

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

---

[2] 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.

[3] If a link is identified between censuses A and B, and censuses B and C, then the link between A and C is implied.

[4] The Census Tree dataset (Price et al. 2023) is available at censustree.org, along with the code used to create it, and the models and training data for the machine learning methods.

[5] Another important characteristic of linked data is precision, or the number of matches that are correct. The Family Tree, CLP, MLP, and Census Tree all have precision above 90%.

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

where the socioeconomic status of the child is predicted by the status of the parent. This measures how much parental status intergenerationally transmits to the child. 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 simple, but there is a large literature on how to accurately estimate β_1 (Solon 2002, Mazumder 2005, Chetty et al. 2014, Nybom and Stuhler 2016). In our case, three issues are especially salient: measurement error in the father's status, the lack of an SES measure for those who do not have an occupation (especially women), and the possibility that our linked sample is not representative of the population. We address each of these in turn.

A. Measurement Error

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). This could be due to fathers being 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 these problems 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.

Instead, we use an instrumental variables strategy to address measurement error, in which one observation of the father's status is instrumented with a second taken from another census. The first stage is given in Equation (2) and the reduced form in Equation (3):

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

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

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 and first stage coefficients (i.e., $\hat{\beta}_{IV} = \frac{\hat{\delta}_1}{\hat{\pi}_1}$). 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. A key concern about this assumption is that transitory error is persistent over time such that error in one observation is correlated with error in the second (Mazumder 2005). However, modern-day studies show such errors only persist for about three years (Modalsli and Vosters 2024). Since our second observation is taken from a different census (at least 10 and at most 20 years away), it is unlikely that transitory errors are correlated across over such a long span, though we cannot verify this assumption with historical data.

This method requires linking multiple censuses to get a second observation for the father—or that the father can be "doubly-linked." In datasets with low linking rates, this requirement can result in a significant loss of observations and intensify any bias from having a nonrandom sample (Ward 2023), as multiple links compound the problem of linkage failures. To illustrate, if linking rates across censuses are independent, then a 30% rate for one link reduces to a 9% rate for two links. But if the linking rate increases to 70%, then the doubly-linked rate would be 49%, or over five times greater.[6] Thus, the Census Tree's high match rates—especially between adjacent censuses for men—mean that we are able to maintain large, representative samples. We elaborate on the construction and representativeness of our doubly-linked samples below.

B. Measuring Socioeconomic Status

The next issue is what measure to use for the father's and child's status. Educational attainment and income are preferred, 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 birth cohort a percentile rank based on its human capital level, measured by literacy in the 1850-1930 Censuses and by years of education in the 1940 Census. Because percentile rankings are made within birth cohorts, this measure captures how an occupation's status changes over time. Ward (2023) creates the "adjusted Song score" that further accounts for differences within occupations by race and region. This approach 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).

[6] 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.

However, a significant drawback of any occupation-based measure is that it is not observed for anyone without an occupation—including most women during this time period. Prior attempts to estimate mobility for women have addressed this by using a woman's husband's occupation (Bailey and Lin 2024, Eriksson et al. 2023, Olivetti and Paserman 2015) in place of her own. But this method may not fully capture a married women's status if she has sources of status that are independent of and not perfectly correlated with her husband's occupation. In households where both partners work, we may want a household measure of status that combines the two, but historical datasets often underreport married women's work (Chiswick and Robinson 2021, Goldin 1990). This issue is particularly evident in the 19th century; for instance, women were not even asked for their occupations in the 1850 Census. Finally, using spousal occupations means that single women are excluded entirely from mobility estimates—a population that accounts for between 25 and 28% of women over our sample period.

To address these problems, we create a new measure of an individual's status that is based on the average occupational status of their nearest geographic neighbors. We take advantage of the fact that the historical U.S. censuses were typically collected by a workforce of enumerators going door-to-door, recording responses by hand on a sheet of paper (Logan and Parman 2017). As a result, those households on the same census sheet were almost always one's immediate neighbors, and those on adjacent sheets also lived nearby. Households are assigned a serial number that orders them across sheets (Hacker et al. 2021).

For each son and daughter in our data (i.e., the dependent variable y_{child}), we create a measure of socioeconomic status that is the average occupation-based status of the household head for the serial numbers before and after their own household. In the results shown here, we average the ten prior and ten subsequent household heads, for twenty total; our results are robust to using more or fewer adjacent households to construct this average. We require all of the averaged observations to be in the same town, using the geographic identifiers from the Census Place Project (Berkes, et al. 2022).[7] We call this measure "neighborhood adjusted Song score" or "neighborhood-SES," which we abbreviate as n-SES. Because this issue is more important for women than men, we can still use the occupation-based measure (adjusted Song score) for the father ($y_{parent,t}$), and thus our IGT estimates are the transmission of father's status on child's n-SES. [8]

[7] For non-relatives living in the house (e.g., domestic servants), we use their own adjusted Song score to avoid confusing the status of the neighborhood they work in with their own status.

[8] We use the adjusted Song score as the measure of the father's status to make estimates more comparable with prior

We show that n-SES effectively captures economic status. First, we compare an individual's n-SES to their adjusted Song score in Table 1, for a sample of 25-55 year-old male household heads with occupations. There is a high and consistent correlation across all censuses, ranging between 0.635 and 0.693. Second, in Table 2, we compare both n-SES and adjusted Song to non-occupational measures of status that are available in certain censuses (real property, personal property, education). We find that the correlation of n-SES with these other measures is similar to the correlation of the adjusted Song score with these same measures. In the 19th century, adjusted Song score has a higher correlation with wealth proxies than the n-SES measure, but this flips in the 20th century for home values. The adjusted Song score and n-SES are similarly correlated with income and years of education in 1940. Importantly, we find that the education correlation is similar for both women and men, which strengthens confidence that our neighborhood proxy reflects socioeconomic status similarly across genders. These relationships in the data give us confidence when using this measure to address the problem of missing occupations.

In Figure 1, we show that we can include many more people in our sample when using n-SES than when using measures like the occscore or the adjusted Song score that are based on an individual's occupation. In each panel, we plot the ratio of the estimation sample size when using n-SES as the measure of the child's status to that when using their own occupation (or their spouse's in the case of married women), by gender and birth cohort.[9] For the married sample in Panel A, using n-SES increases the potential sample size by 1 to 8% for men and by 2 to 9% for women. This highlights the fact that the n-SES measure does not only solve the problem of unobserved status for women—it allows for the inclusion of others who are unemployed, not in the labor force (including the disabled, retirees, and those in school), or with difficult-to-classify occupation strings.[10]

In Panel B, there is a slightly larger increase in the sample size for single men, ranging from 7 to 9%, as this group is more likely to be unemployed. The big differences, however, appear for single women. For the 1840 birth cohort, the sample of single women increases by nearly 3.5 times when using the n-SES measure. This ratio falls monotonically as women's labor force participation

work. Below we explore the sensitivity of our results to using n-SES to measure status for both generations, and to using adjusted Song for both.

[9] 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.

[10] IPUMS-USA codes individuals with difficult to classify occupations as "Not Yet Classified," which includes 1.4-2.1 million people between the 1900 and 1930 Censuses.

increases, allowing us to observe an occupation for more single women.[11] Still, for the 1910 birth cohort, using n-SES increases the sample of single women by 34%. The decline in the ratio over time highlights the key advantage of the n-SES measure: because it can be constructed for everyone, our estimates will not be affected by changing selection into marriage or labor force participation.

C. Representativeness of 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 we can construct an adjusted Song score for the father. We then identify those children for whom we can 1) link to an observation with an observed n-SES 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.[12] 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.[13] This constitutes the "doubly-linked" sample described above.

A key group for whom we cannot use this method to create intergenerational links is the children of enslaved fathers, as they were not enumerated in the 1850 and 1860 censuses. For this group, we append a random sample of southern-born Black adults in the 1870 and 1880 Censuses to our linked dataset. We can calculate the n-SES for this group using those censuses, and then impute their father's "occupation" to be enslaved and assign it the lowest status measure in the distribution (effectively 0 on the 0-100 scale). This allows us to capture the impact of emancipation, 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.

---

[11] Between the 1860 and 1940 censuses, the fraction of women reporting an own occupation increases from 33% to 64%. In the 1850 census, women were not given the opportunity to report an occupation.

[12] 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.

[13] 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 status 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).

In Table 3, we show how our sample is affected by the requirements that we have both an adult observation for the child and a second occupational observation for their father. 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 and construct their n-SES. 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 49.2% 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 42.5% of the sample, or 37.5 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. In total, the Census Tree estimation sample is 5.2 times larger for men, with over 20 million observations. When including the 14.4 million doubly-linked observations we can construct for women, the Census Tree has 8.7 times more father-child pairs than the sample using exact-conservative links from the Census Linking Project.

The descriptive statistics for our doubly-linked estimation sample are shown in Table 4. There are three key things to note about the (adult) children. First, our sample is 59% male, on average. The over-representation of men is expected given that the supervised and unsupervised linking methods included in the Census Tree are usually unable to link women when their surnames change. Nevertheless, the user-made links from the Family Tree still allow us to include over 14 million links for women. Second, Black men and women are under-represented in our sample, as 10-15% of the national population was Black during this time period. But here too, the Census Tree's sample sizes mean that we are able to include 1.4 million Black Americans in our doubly-linked sample. Third, the majority of children in the sample are married (68-78%), but we still retain unmarried individuals. For the fathers, we see that a high and declining fraction are farmers; we explore the sensitivity of our results to excluding this population in the next section.

Given that there is nonrandom selection into our doubly-linked sample, we follow Bailey et al. (2020) and reweight the linked data to match population characteristics based on the adult child's observable characteristics throughout our analysis. 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, and use the predicted probabilities to create inverse probability weights for each individual.[14, 15] In Table 5 we replicate Table 4 using the Bailey weighting procedure and show that—as expected—the weighted sample matches known features of the US population closely.[16] Prior work has shown that estimates weighted this way are highly robust to differences in sample size and sample construction (Abramitzky et al. 2024).

IV. Intergenerational Mobility Results

A. Baseline Estimates

Figure 3 shows our main estimates of intergenerational mobility for the 1840-1910 birth cohorts, for daughters and sons.[17] Starting with the IV estimates based on Equations (2) and (3), we see that persistence was high but falling over this period. To interpret these coefficients, just over two-thirds of the initial rank gap across two fathers persisted for their children in the 1840 birth cohort (where again, the child's SES is the average of their neighbor's adjusted Song scores, and the father's is his own). Over time, this link weakened, with the coefficients dropping from 0.67 to 0.55 for sons and from 0.69 to 0.58 for daughters in the 1910 birth cohort. Our preferred estimates use an IV strategy to correct for measurement error in the father's status, but we also plot OLS estimates in Figure 3, keeping the sample the same. The OLS estimates show a similar pattern over time, but the coefficients are 20-30% smaller due to attenuation bias caused by measurement error. Ward (2023) also finds a similar pattern of high but declining persistence (or low but improving mobility) in a men-only sample using mostly occupation-based scores.[18] We will directly compare mobility estimates from the neighborhood and occupational-based measures later.

---

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

[15] 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.

[16] For example, the fraction Black in the U.S. fell from 15.7% in the 1850 census to 10.7% in the 1910 census. In our weighted sample, the percent Black falls from 13.3% for the 1850 birth cohort to 9.4% for the 1910 cohort. Our weighted sample is also just over 50% male. Note that we cannot directly compare our doubly-linked sample to population characteristics, as the characteristics for a particular birth cohort are constructed using multiple censuses in which the cohort is observed as an adult.

[17] We calculate standard errors for our coefficients, but do not include confidence intervals on this and most other figures because they are too small to be represented visually.

[18] The finding of low but improving relative mobility since the 19th Century contrasts with studies reporting high and

A striking finding in Figure 3 is that the estimates of transmission are very similar for daughters and sons. The IV coefficient for daughters is only 0.02 to 0.04 higher than for sons.[19] The similarity across daughters and sons also holds when using OLS. The similarity in parent-child transmission estimates across genders is consistent with post-World War II findings from Chadwick and Solon (2002), who use the Panel Study of Income Dynamics, and Chetty et al (2020), who use modern-day income tax data. But as discussed above, it is not obvious that transmission to sons and daughters would be so similar historically, as this was an era when the overwhelming majority of women did not work outside the home and so their socioeconomic status depended on their husband's. The fact that they are suggests strong assortative mating—a possibility we explore in the next section.

When comparing estimates for sons and daughters, it is important to consider possible differences in sample composition. It is the case that our links for married daughters mainly come from the user-generated Family Tree links, as traditional matching methods fail to link most women when their surnames change with marriage. If these Family Tree links are non-representative in ways not corrected by weighting on observable characteristics, then our estimates for daughters (mostly from the Family Tree) could differ from those for sons (who have relatively more matches from machine learning) due to this difference in link composition. To investigate this, we reproduce the IGT estimates in Figure 3 using only Family Tree links in Appendix Figure A2. First, we find that estimates *for sons* are nearly identical whether using the Census Tree or the Family Tree (comparing IV estimates in Figure 3 and Appendix Figure A2), suggesting that genealogical links and non-genealogical links yield similar IGT estimates. This is consistent with Buckles et al. (2024), who further show that intergenerational estimates for men in the Census Tree are similar to estimates using the Census Linking Project (Abramitzky et al. 2022), which does not use genealogical links. Second, Appendix Figure A2 shows that IGT estimates for daughters and sons are also similar in the Family Tree, indicating that the difference in link composition across genders does not drive results.

Farming was the most important occupation in our sample, yet the closeness of estimates between sons and daughters is not driven by farming. We show in Appendix Figure A3 that if one drops farmers—who made up 38-52% of fathers—from the sample, father-son and father-daughter

declining mobility (e.g., Long and Ferrie 2013, Song et al. 2020). As Ward (2023) explains in detail, the difference is due to (1) accounting for measurement error in the father's status, (2) including Black individuals in the sample, and (3) using a status measure that captures within-occupation inequality across regions in the US.

[19] In Figure A1 we plot the binscatter relationship between the father's status and the child's; here again, estimates for daughters and sons are nearly identical.

transmission estimates remain closely aligned, though the level of the IGT estimates falls. A lower transmission rate when dropping farmers has been found by others (Olivetti and Paserman 2015, Ward 2023) as it partially reflects that farming was a relatively persistent occupation (Haws, Just, and Price 2024). But despite the change in levels, we continue to find the same overall trends and strong similarity in the strength of transmission to daughters and sons.

Finally, the regression coefficients represented in Figure 3 are estimated using the Bailey et al. (2020) weighting procedure. In Panel B of Appendix Figure A3, we show that omitting these weights leads to slightly higher IGT estimates. Nevertheless, our key findings hold: parent-child transmission is high but decreasing throughout the period, and the transmission coefficients for daughters and sons are very similar.

B. Results by Race

We find that the similarity in mobility estimates across gender also holds within race. Figure 4 splits the linked sample by Black and White fathers and plots IV estimates of the child's n-SES rank on the father's occupational rank. First, we observe that when limiting the sample to the children of White fathers, our IGT estimates are generally lower, with the exception of the 1840 and 1850 birth cohorts—a result we return to below. A lower IGT for White families is consistent with prior work that finds that estimates based on White-only samples overstate the degree of mobility (understate persistence) in U.S. history, as they do not capture barriers to economic advancement for Black Americans (Jácome et al. forthcoming, Ward 2023). Despite the change in levels, the trends are similar to those in Figure 3, and we continue to find that IGT for daughters and sons was similar, with slightly higher estimates for daughters.

We also find that the transmission between father and child was similar for Black boys and girls. This result is somewhat surprising given higher female labor force participation rates for Black women than for White women throughout American history, which may have helped Black women to further break the tie with the father. However, it is consistent with historical work from Jácome et al (forthcoming) that mobility rates were similar across gender for Black children born between 1910 and 1940.

We estimate very high transmission rates for Black families before the 1880 birth cohorts, which implies that gaps across Black fathers *widened* for their children, such that there was no convergence toward the Black mean.[20] While having a slope above one is not possible for the whole

[20] Ward (2023, Figure A13) also found a high persistence estimate for Black families, so this result is not unique to the Census Tree sample. We estimate a sharp decline between the 1870 and 1890 birth cohorts, which is also found by

sample, it is within the Black population because 95% of Black fathers were below the 15th percentile of the distribution. That is, while gaps widened among Black families (largely due to diverging levels of status for Black agricultural workers across the South and non-South), these gaps were widening *within* the low end of the status distribution. If one ignores these regional differences within occupation and uses the *unadjusted* Song score for the father's score, the within-Black estimate is less than 0.40 across birth cohorts (see Figure A4).[21] While the level and trend of Black mobility are sensitive to the measure of economic status, the key takeaway is that we continue to find intergenerational mobility estimates that are similar for sons and daughters, even within racial groups.

In Figure 5, we show how intergenerational mobility estimates change when going from a sample traditionally used in economic history (White men) to the full population of Black and White men and women. Following Jácome, Kuziemko, and Naidu (forthcoming), we start with a While male sample and then progressively add White women, Black men, and then finally Black women. As expected, given the results above, adding White women to the White male sample has minimal impact on estimates. Rather, the significant changes occur when Black men and women are added. Overall, the population-level mobility estimates indicate that intergenerational transmission decreased from 0.68 for the 1840 birth cohort to 0.58 for the 1910 birth cohort. Interestingly, adding Black men and women born in the 1840 and 1850s cohorts, most of whom were emancipated, does not substantially increase estimates. This contrasts with results when using the adjusted Song score, where adding Black families causes a large increase in IGT estimates (Ward 2023). The difference arises because emancipation did not improve Black workers' occupational scores (as many remained in low ranked jobs), but it did improve their neighborhood ranks, as some Black families lived near White neighbors (Logan and Parman 2017). Overall, similar to Jácome, Kuziemko, and Naidu (forthcoming), we find that samples that do not include Black women will understate persistence following emancipation.

C. Results by Ethnicity

Gender differences in mobility might be larger across descendants from different source countries due to differences in gender norms across ethnicities. Indeed, cultural assimilation appears

---

Althoff et al. (2024) when using an alternative sample and approach to uncover latent human capital ranks.

[21] Another issue is that our status measure does not distinguish between farm owners and tenants, as this is not consistently observable in the population schedules. For estimates of within-Black mobility in the late 19th century that address this issue, see Collins et al. (2024), where they link population schedules to the agricultural censuses to observe farm values and acreage.

incomplete for the children of immigrants, as their views on gender norms, their fertility rates, and their rates of female labor force participation remain correlated with their ancestor's source country (Alesina et al. 2013, Blau et al. 2013, Fernández and Fogli 2009, Guinnane et al. 2006). In countries with high female labor force participation, daughters might have been less tied to their father's status, as they could earn wages and save to move to better neighborhoods. Using our data, we can estimate gender differences in mobility by the father's country of birth, examining whether there are starker differences between father-daughter and father-son estimates than in the broader US population. We do so by pooling all birth cohorts and estimating our preferred specification where we use the IV method to estimate the association between the father's economic status and the child's neighborhood status for 20 major European sources, as well as Mexico and Canada.

We find that mobility estimates were also remarkably similar across gender by father's source country, revealing a correlation of 0.942 between father-son and father-daughter estimates (see Figure 6). While this finding does not rule out the importance of gender norms for female outcomes during this period, it suggests that such differences did not drive a large wedge between the rate of social mobility across sons and daughters. Despite this gender similarity, there was significant variation in transmission estimates across countries, ranging from around 0.23 for children of Norwegian-born fathers to nearly 0.5 for children from Mexican-born fathers. For comparison, the figure also includes estimates for US-born White and Black children, showing that transmission was lower among the descendants of immigrants than for US-born groups, consistent with the findings in Abramitzky et al. (2021) that the children of immigrants experience greater relative mobility than the children of the US-born.

D. Results by Childhood State

We also test whether mobility estimates by gender show similar patterns across place within the United States. To do this, we focus on the father's state of residence, and limit the sample to the White population to ensure that gender similarities across states are not driven by the racial composition of the state. We find a strong correlation between father-son and father-daughter transmission estimates at the state level (see Figure 7). The correlation across state/gender mobility estimates is 0.97, with most states showing slightly stronger transmission between father-daughter than father-son (except in a handful of Western states). Similar to the earlier estimates by father's country of birth, we find that the national-level estimates mask wide variation in estimates by state. Southern states had stronger parent-child transmission estimates, while Northeastern and Midwestern states had much lower estimates. These spatial patterns are qualitatively similar to other studies that estimate relative mobility across space using linked historical data in the early 20th century (Connor and Storper 2020, Tan 2023).

We can compare our transmission estimates to others in the literature, and show that they are similar in level to those from Bailey and Lin (2024) for Ohio and Eriksson et al. (2023) for Massachusetts. For example, in Figure A5 we show that for the 1870 Ohio birth cohort, our father-daughter estimate is 0.30, which is much closer to the 0.36 estimate reported in Bailey and Lin (2024) than our national level estimate of 0.64. Over time, we find a slight decline in the Ohio IGT from 0.30 for the 1870 cohort to 0.26 for the 1900 cohort, which is found by Bailey and Lin (2024), though they estimate a steeper decline. In Massachusetts, we estimate lower IGT estimates than in Ohio, where we estimate that the father-daughter transmission declined from 0.29 for the 1850 birth cohort to 0.19 for the 1900 cohort. These estimates are similar to Eriksson et al. (2023, Table 2), albeit higher, who report 0.22 for the 1850 cohort and 0.18 for the later 1890 cohort.

V. The Role of Marital Status and Assortative Mating

So far, we have pooled married and single children when estimating intergenerational mobility by gender. But the channels for the transmission of socioeconomic status might depend on marital status—especially for daughters during a period of low female labor force participation. For married women, status depends on their choice of partner, and transmission is stronger when assortative mating is higher. Single women are more likely to work, which may operate as an engine of mobility for them. However, many unmarried daughters in the period lived with their fathers, which would generate a strong correlation in their status—especially when using our neighborhood-based measure of status, as the daughter would be assigned the status of their *father's* neighbors.[22]

We first plot transmission estimates for single sons and daughters in Panel A of Figure 8; again, we note that the latter are excluded from historical mobility estimates that rely on marriage records (Eriksson et al. 2023) or the husband's occupation (Olivetti and Paserman 2015) to measure status. Here, we see that the results that pooled single and married children masked a slightly larger gap in mobility estimates between single sons and daughters—transmission to single daughters is stronger. Panel B of Figure 8 shows that the strikingly similar estimates for men and women in the

[22] While this makes the transmission of status from fathers to live-in sons and daughters somewhat mechanical, we argue that this is exactly what our IGT measure is meant to capture. The status of a single child living with their father will naturally be very closely linked to his status. Put another way, the child's well-being is closely linked to the level of resources that the father's occupation provides.

full sample are driven by the similarity across married men and women. This fact suggests that women married men with similar underlying status—that is, assortative mating was very high (Curtis 2021; Clark and Cummins 2022).

To explore this, 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's adjusted Song score to instrument for the first, as we did with the IGT estimates above.

Figure 9 shows the trend in assortative mating estimates based on the husband's birth cohort. 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. The trend is similar to the trend for intergenerational mobility, where the association between the father and father-in-law falls from 0.75 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 much larger than most others in the recent literature (Bailey and Lin 2024, Espín-Sánchez et al. 2023, Eriksson et al. 2023, Olivetti et al. forthcoming). 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 ranges between 0.04 and 0.42 over the same period (see Appendix Table A1). The difference in estimates could be due to sample construction, but that does not appear to be the primary reason; rather, it is because 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 OLS with the fathers' occscore as the measure of status, we estimate that assortative mating is between 0.12 and 0.25, which is similar to other estimates (see Figure A6). Our higher estimates when using IV are

[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, $E[father]=father+motherAM$, where AM captures 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.

consistent with others who address measurement error in the United Kingdom (Clark and Cummins 2022) and Canada (Curtis 2021).

One caveat is that we use the parent's generation to estimate assortment. An advantage of this approach is that status is predetermined from the wife and husband's childhood and therefore is not biased by an endogenous cause or effect of marriage on the wife or husband's occupation (Almås et al. 2023). Yet because we are using the prior generation, it may be that assortment is even higher between husband and wife directly than indirectly between father and father-in-law. Indeed, our assortment estimates are lower in magnitude than those in Althoff et al. (2024), who use a semiparametric latent variable model and find that the correlation of latent human capital ranks between husband and wife is about 0.77-0.88 between 1870 and 1910 birth cohorts.[24] Another issue is that 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 between father and father-in-law for the population as a whole (Bailey and Lin 2024; Eriksson, Lake, and Niemesh 2022). However, our estimates are much higher than those listed above that also use the parents' status to measure assortative mating (Eriksson et al. 2023; Olivetti et al. forthcoming).[25]

VI. Comparisons to Prior Work

A. Using Adjusted Song Score to Measure the Child's Status

For our main intergenerational mobility estimates, we use a novel measure of socioeconomic status—the neighborhood-SES—which removes the potential for bias due to selection into labor force participation and marriage, and better captures the full set of resources available to the household. However, because our measure is new, it is important to ask how our findings compare to those using conventional occupational-based measures for *both* the parent and the child (Song et al. 2020, Ward 2023).

[24] These numbers come from taking the square root of R-square estimates in Appendix Figure A6 of Althoff et al. (2024), which uncovers the correlation.

[25] Our high estimate of assortment at the national level also reflects geographic sorting. When one's marriage market mostly consists of others who grew up in the same area, and there is significant variation in status across places, we will observe higher levels of assortative mating at higher levels of geography. Indeed, when we add controls for the childhood town of residence, we find that our estimates of assortative mating decrease.

Beginning with the estimates for sons in Figure 10 Panel A, we see that our IV estimates are higher in magnitude when using the son's adjusted Song score as his measure of SES, especially for earlier cohorts.[26] The father-son transmission coefficient falls from 0.82 for 1840 birth cohorts to 0.63 for the 1910 birth cohort. These high estimates align with previous work by Ward (2023), who uses the same score for father-son estimates. The differences in estimates across n-SES and own adjusted-Song measures could be due to changing sample composition, as the former includes all sons who can be doubly-linked to a father, while the latter includes the subset of those with a reported adult occupation. However, when we restrict the n-SES to those with a reported occupation, we find a similar trend as our main estimates, ruling out that sample composition explains the difference. Instead, the difference is likely due to the fact that our n-SES measure is a fundamentally different measure of socioeconomic status than an individual-based measure. By taking the average status of one's neighbors, our measure captures the full set of resources available to the household that affect status, such as wealth (or lack thereof) or the earnings of multiple household members. It also allows for more variation in status within the occupation cells defined by the Song score.[27] A household in which the wife also works for pay may be able to afford to move to a higher-status neighborhood as a result of her income; this will be captured by the n-SES measure even if the husband's occupation does not change. The results in Figure 10 suggest that, using this concept of status, sons' fortunes appear to be slightly less tied to their fathers.

Still, the key finding of our empirical analysis holds: we see similar trends in the IGT estimates across cohorts for daughters and sons, with slightly stronger transmission to daughters. When using the adjusted Song score to measure the child's status, the difference in estimates between sons and daughters is larger, ranging from 0.02 to 0.10 (compared to 0.02 to 0.04 when

[26] We calculate a son's adjusted Song score using his occupation. For daughters, we follow the literature and use her husband's occupation, thus actually estimating father/son-in-law mobility (Olivetti and Paserman 2015). For single children, we use their own adjusted Song, and drop those who do not report an occupation. When creating the adjusted Song score, we stratify by gender as well as by race and region. 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 prior work.

[27] For example, consider two White male contemporaries: one is a teacher at a wealthy boarding school in New York City, while the other teaches at a poor school in the city's slums. The adjusted Song score would assign these two men the same status, as they have the same occupation, same race and cohort, and live in the same region of the country. However, if the former lives among members of the professional class, while the latter lives among tradespeople and blue-collar workers, the neighborhood measure will appropriately assign him higher status. Similarly, a young person with generational wealth may report a low status occupation for themselves (e.g. artist), but if they continue to live among neighbors with high-status occupations, their n-SES will better reflect their true position in society.

using n-SES).[28] Nevertheless the similarity of estimates between father-son and father-daughter, during a time when most women did not work, is consistent with a high rate of assortative mating.

As an additional check on the sensitivity of our results to the use of the neighborhood measure, Appendix Table A2 shows that using 1940 years of education (ranked by birth cohort to align with the father's score) instead of neighborhood status as the dependent variable still yields similar estimates for father-son (0.47) and father-daughter (0.44) transmission. Ideally, we would use education across all years since it is observable for sons and daughters regardless of labor force participation, but unfortunately education is only first available in 1940.[29]

B. Prior Estimates of Father-Daughter Transmission of SES

The magnitudes of our transmission estimates are higher than the most prominent 19th and early 20th Century father-daughters estimates from the literature in Olivetti and Paserman (OP) (2015). OP use the same 1850-1940 censuses to estimate mobility, but instead of having direct links across censuses, they use pseudo-links based on one's first name. They find transmission estimates between 0.3 and 0.5, with higher transmission to sons-in-law than for sons for some earlier cohorts, but the reverse for cohorts born in the 1900s. One reason our transmission estimates are higher is that this pseudo-linking methodology can be subject to measurement error (Santavirta and Stuhler 2023). OP pseudo-link one-percent samples of the US Censuses—importantly, not the full-count Censuses—which introduces measurement error because the children in the first sample are mostly not the adults in the second sample, and thus the wrong fathers' occupational scores are used to proxy for the child. This issue can be addressed by using full-count data instead of samples. Other reasons why our point estimates differ are that we include Black families in our estimates and use a different measure of socioeconomic status.

In Figure 11 we replicate the OP method using the full-count data, using the same name-based strategy to approximate the father's status and including Black families. We find higher transmission estimates for both men and women (0.65-0.84) than the directly-linked estimates.[30] We

[28] If we instead use n-SES to measure both the child's status *and the father's*, our IV estimates are higher in levels—ranging from 0.77 to 0.87 for women and between 0.74 and 0.82 for men, with gaps between 0.00 and 0.08 (Figure A3, Panel C).

[29] Another variable available for men and women regardless of labor force participation is longevity, which Black et al. (2024) analyze and also document similar parent-son and parent-daughter correlations.

[30] We also pseudo-link on first name, sex, race and birthplace combinations, not just first name and sex. Figure 11 plots the regression results of $y_{child} = a_0 + a_1\bar{y}_{father} + \varepsilon_i$ where y_{child} is the neighborhood score observed in a cross-sectional full-count census, and $\bar{y}_{father}$ is the average adjusted Song score of fathers of children with the same first name, birthplace and race combinations observed in cross-sectional census 20-40 years prior.

also continue to find that mobility was increasing throughout the period (as in Figure 3), unlike OP who find that mobility decreased from 1870 to 1920. The results in Figure 11 suggest that differences in the levels and trends between our estimates and OP are not due to their use of the name-based method, but rather to differences in the samples and measure of socioeconomic status. But our main finding is stubbornly persistent: even with the name-based method that does not rely on linking, we find that male and female transmission estimates are similar in level and trend, with the female transmission estimates being slightly higher (by 0.01 to 0.06). We prefer the directly linked estimates since names may provide additional information content that biases mobility estimates (Santavirta and Stuhler 2023).

C. Combining Estimates to Construct a 130-Year Time Series

Using the Census Tree links from 1850-1940, we have constructed a time series of mobility estimates by gender for the 1840 to 1910 birth cohorts. Meanwhile, Jácome, Kuziemko, and Naidu (JKN) (forthcoming) use data from a collection of surveys to produce estimates of the intergenerational elasticity of income by gender for birth cohorts from 1910 to 1970—picking up right where we leave off. While the data and exact measures of status are different, the two sets of estimates have some important similarities. First, both use a measure of the parent's status that is constructed from census data on the father's race, region, and occupation. Second, both use a measure of the *child's* outcome that is meant to capture household status—neighborhood SES in our case, and log family income in JKN. Third, both use weights to make the estimates more representative of the population.

Figure 12 shows the time series for the 1840 to 1970 birth cohorts when combining the estimates from these two sources. We make two observations. First, while there is a discontinuity when switching between the two time series, it seems that the increase in mobility that we observe for the 1850-1910 birth cohorts continued through 1940, at which point mobility began to decrease (or the IGT began to increase). Second, transmission to daughters is stronger than transmission to sons for all birth cohorts across this 130-year span, with economically small gaps across genders for the pre-1910 cohorts, and much larger gaps for post-1910 cohorts based on estimates from JKN.

The difference in gender mobility gaps over time appears to be due to how our proxies capture Black-White differences in child status. JKN show that the larger father-daughter coefficient is due to larger Black-White gaps in family income for girls than for sons. Because Black women had much lower family income than other race/gender groups, they had low rates of upward mobility, which causes higher transmission estimates for daughters. In terms of decomposing the IGT estimate into within-race and between-race effects, there is a larger between-race effect for daughters than for sons. But when JKN measure within-race mobility, they find small differences across gender, which agrees with other modern work by Chadwick and Solon (2002) and Chetty et al. (2020). Since our measure of childhood status is based on the neighborhood, we measure similar Black-White gaps in neighborhood status for sons and daughters, as they live in the same areas.[31] As a result, our estimates of the gender gap in transmission are smaller than those in JKN.

VII. Conclusion

In this paper, we present the most comprehensive estimates of historical intergenerational mobility 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. Importantly, because the Census Tree includes women as well as men, we can 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.

Our main finding is that intergenerational mobility for women is nearly identical to that for men. That is, the father's economic status predicts a daughter's status just as well as it does a son's. We show that this result is robust across a number of different samples and specifications and that it is not driven by any bias arising from using genealogical data to generate some of the links. This result is surprising, because for our historical context, most daughters' status was determined by that of her husband—someone who was not raised in the same childhood household, and thus did not receive the same human capital investments as the daughter.

One explanation for this result is that assortative mating was historically very high in the United States. We show this directly by estimating a strong association between the economic status of a father and the father-in-law. This finding has significant implications for the history of American inequality because if the rich married the rich and the poor married the poor, then it could help to explain cross-sectional inequality and the trend in intergenerational mobility (Fernandez and Rogerson 2001; Eika, Mogstad, and Zafar 2019; Ermisch, Francesconi, and Siedler 2006; Collado, Ortuño-Ortín, and Stuhler 2023, Keller and Shiue 2023).

[31] In our data, the average rank for neighborhood status for White boys is 54.1 and 22.2 for Black boys, resulting in a Black-White male gap of 31.9 ranks. The average rank for White daughters is 53.2 and 22.0 for Black daughters, leaving a similar Black-White female gap of 32.2 ranks.

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, Harriet Brookes Gray, Katherine Eriksson, Santiago Pérez, and Myera Rahid. "Finding 'John Smith': Using Extra Information for Historical Record Linkage." Working Paper. (2024).

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 (2021): 580-608.

Alesina, Alberto, Paola Giuliano, and Nathan Nunn. "On the origins of gender roles: Women and the plough." *The Quarterly Journal of Economics* 128, no. 2 (2013): 469-530.

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, 2024. https://lukasalthoff.github.io/pdf/igm_mothers.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 US." In *The Economic History of American Inequality: New Evidence and Perspectives*. National Bureau of Economic Research, 2024.

Becker, Gary and Nigel Tomes. "Human Capital and the Rise and Fall of Families." *Journal of Labor Economics* 4, no. 3 Part 2 (1986): S1-S39.

Becker, Gary S., Scott Duke Kominers, Kevin M. Murphy, and Jörg L. Spenkuch. "A theory of intergenerational mobility." *Journal of Political Economy* 126, no. S1 (2018): S7-S25.

Berkes, Enrico, Ezra Karger, and Peter Nencka. "The Census Place Project: A Method for Geolocating Unstructured Place Names." *Explorations in Economic History* 87 (September 20, 2022): 101477. https://doi.org/10.1016/j.eeh.2022.101477.

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

Black, Sandra E., Neil Duzett, Adriana Lleras-Muney, Nolan G. Pope, and Joseph Price. *Intergenerational Correlations in Longevity*. No. w31034. National Bureau of Economic Research, 2024.

Blau, Francine D., Lawrence M. Kahn, Albert Yung-Hsu Liu, and Kerry L. Papps. "The transmission of women's fertility, human capital, and work orientation across immigrant generations." *Journal of Population Economics* 26 (2013): 405-435.

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.

Buckles, Kasey, Adrian Haws, Joseph Price, and Haley Wilbert. "Breakthroughs in Historical Record Linking Using Genealogy Data: The Census Tree Project." 2024. 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.

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.

Chetty, Raj, and Nathaniel Hendren. "The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects." *The Quarterly Journal of Economics* 133, no. 3 (2018): 1107-1162.

Chiswick, Barry R., and RaeAnn Halenda Robinson. "Women at work in the United States since 1860: An analysis of unreported family workers." *Explorations in Economic History* 82 (2021): 101406.

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.

Collins, William J., Nicholas Holtkamp, and Marianne H. Wanamaker. "Black Americans' Landholdings and Economic Mobility after Emancipation: Evidence from the Census of Agriculture and Linked Records." *The Journal of Economic History*, 2024, 1–34. https://doi.org/10.1017/S0022050724000299

Connor, Dylan Shane, and Michael Storper. "The Changing Geography of Social Mobility in the United States." *Proceedings of the National Academy of Sciences* 117, no. 48 (November 16, 2020): 30309–17. https://doi.org/10.1073/pnas.2010222117.

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.

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.

Espín-Sánchez, José-Antonio, Joseph P. Ferrie, and Christopher Vickers. *Women and the Econometrics of Family Trees*. No. w31598. National Bureau of Economic Research, 2023.

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).

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

Fernández, Raquel, and Alessandra Fogli. "Culture: An empirical investigation of beliefs, work, and fertility." *American Economic Journal: Macroeconomics* 1, no. 1 (2009): 146-177.

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.

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

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.

Guinnane, Timothy W., Carolyn M. Moehling, and Cormac Ó. Gráda. "The fertility of the Irish in the United States in 1910." *Explorations in Economic History* 43, no. 3 (2006): 465-485.

Hacker, J. David, Jonas Helgertz, Matt A. Nelson, and Evan Roberts. "The influence of kin proximity on the reproductive success of American couples, 1900–1910." *Demography* 58, no. 6 (2021): 2337-2364.

Haws, Adrian, David R. Just, and Joseph Price. "Who (Actually) Gets the Farm? Intergenerational Farm Succession in the United States." *American Journal of Agricultural Economics* (2024). https://doi.org/10.1111/ajae.12493

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.

Jácome, Elisa, Ilyana Kuziemko, and Suresh Naidu. "Mobility for All: Representative Intergenerational Mobility Estimates over the 20th Century." *The Journal of Political Economy,* forthcoming. https://doi.org/10.1086/732527.

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.

Keller, Wolfgang, and Carol H. Shiue. *Intergenerational Mobility of Daughters and Marital Sorting: New Evidence from Imperial China*. No. w31695. National Bureau of Economic Research, 2023.

Logan, Trevon D., and John M. Parman. "Segregation and Homeownership in the Early Twentieth Century." *American Economic Review* 107, no. 5 (May 1, 2017): 410–14. https://doi.org/10.1257/aer.p20171081.

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.

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.

Modalsli, Jørgen, and Kelly Vosters. “Spillover Bias in Multigenerational Income Regressions.” *Journal of Human Resources* 59, no. 3 (February 4, 2022): 743–76. https://doi.org/10.3368/jhr.0520-10919r2.

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. "The Emergence of the Socioeconomic Gradient in Women's Marriage Outcomes." *The Review of Economics and Statistics*, (forthcoming): 1–29. https://doi.org/10.1162/rest_a_01481.

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.

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.

Tan, Hui Ren. "A Different Land of Opportunity: The Geography of Intergenerational Mobility in the Early Twentieth-Century United States." *Journal of Labor Economics* 41, no. 1 (January 1, 2023): 77–102. https://doi.org/10.1086/718595.

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.

Wagner, Deborah, and Mary Layne. *The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' (CARRA) Record Linkage Software.* CARRA Working Paper Series #2014-01. U.S. Census Bureau, July 1, 2014.

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.

Table 1: Correlations between own and neighborhood adjusted Song score (n-SES)

Census Year	Correlation	Observations
1870	0.689	4,957,040
1880	0.693	6,554,934
1900	0.664	9,592,202
1910	0.671	12,592,360
1920	0.669	15,189,684
1930	0.654	18,270,156
1940	0.635	20,570,452

Notes: Table shows the correlation between the individual's adjusted Song score and their neighborhood adjusted Song score (n-SES), for male heads of household with at least one child, age 25-55. See text for details on the construction of the n-SES. We include the 1870-1940 censuses as these are the census years in which we observe the son or daughter's adult outcome.

Table 2: Correlations between own and neighborhood adjusted Song score and available measures of socioeconomic status

		Correlation with:		
Census Year		Own adjusted Song	Neighborhood adjusted Song	Observations
1870				
	Property Wealth	0.282	0.188	4,956,835
	Total Wealth	0.379	0.245	4,956,680
1930				
	House Value	0.319	0.426	7,637,454
1940				
	House Value	0.417	0.498	7,687,861
	Income	0.511	0.460	14,487,922
	Education	0.489	0.455	20,157,858

Notes: Table shows the correlation between the individual's adjusted or neighborhood Song score (n-SES) and measures of socioeconomic status available in the 1870, 1930, and 1940 censuses. The sample includes male heads of household age 25-55. See text for details on the construction of the n-SES.

Table 3: 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 n-SES	Percent Retained	Also has a Second Adjusted Song for the Father	Percent Retained
1850	6,978,173	3,103,526	44.47%	2,326,855	33.34%
1860	8,578,939	3,611,853	42.10%	3,205,347	37.36%
1870	12,245,005	4,803,220	39.23%	4,261,983	34.81%
1880	16,437,894	9,452,357	57.50%	7,832,669	47.65%
1900	20,903,258	12,332,454	59.00%	10,792,402	51.63%
1910	23,060,478	10,115,169	43.86%	9,061,812	39.30%
Total	88,203,747	37,750,699	49.23%	37,481,068	42.49%

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 a neighborhood SES measure (n-SES) when the child is between the ages of 25 and 55. 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 4: Descriptive statistics of the linked intergenerational sample (unweighted)

Birth Cohort	1840	1850	1860	1870	1880	1890	1900	1910
Panel A: Child Characteristics								
Black	2.9	1.7	2.2	4.7	4.7	4.6	4.4	4.0
Age	37.4	38.6	44.3	39.9	39.1	38.8	37.6	30.0
	(4.7)	(9.1)	(6.2)	(5.9)	(4.7)	(4.8)	(4.8)	(3.0)
Male	58.2	59.0	60.5	59.6	58.0	58.1	58.4	61.0
Neighborhood SES	53.6	53.0	54.7	54.0	53.6	53.1	52.7	51.7
	(20.7)	(20.6)	(19.8)	(20.5)	(20.4)	(19.8)	(19.1)	(18.8)
Married	77.8	74.4	77.7	75.5	75.6	78.1	75.8	67.7
Panel B: Father Characteristics								
Age	40.9	39.4	40.0	40.0	37.0	41.3	38.1	37.8
	(6.7)	(7.1)	(7.2)	(7.2)	(7.1)	(6.5)	(7.1)	(7.1)
Adjusted Song Score	55.5	54.9	55.0	53.3	52.3	50.2	49.8	50.0
	(25.0)	(24.7)	(24.3)	(25.3)	(25.5)	(25.9)	(25.9)	(25.9)
White Collar	8.0	9.4	10.4	11.3	11.8	13.4	15.0	17.0
Farmer	60.4	56.6	53.1	51.7	50.8	47.7	43.1	38.2
Unskilled	14.7	17.1	21.5	23.6	24.5	24.2	26.6	28.0
Skilled	14.1	15.4	14.8	13.2	12.8	14.5	15.1	16.5
N	1,320,179	2,243,961	3,233,320	4,573,210	3,235,326	5,918,861	7,833,114	7,055,545

Notes: This table shows the descriptive statistics of the sample, with standard deviations in parentheses. Each column reports statistics by the child's birth cohort, rounded to the nearest decade. 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.

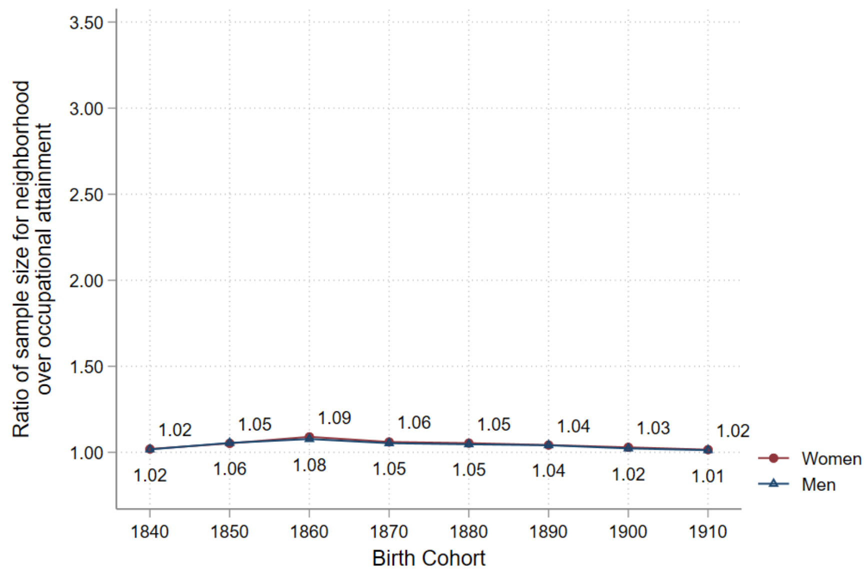
Table 5: Descriptive statistics of the linked intergenerational sample (weighted)

Birth Cohort	1840	1850	1860	1870	1880	1890	1900	1910
Panel A: Child Characteristics								
Black	14.6	13.3	8.9	9.9	11.3	9.7	9.5	9.4
Age	36.9	36.7	44.5	39.3	35.8	41.2	37.6	29.9
	(5.1)	(9.8)	(7.3)	(7.0)	(5.7)	(7.4)	(5.2)	(3.1)
Male	52.2	53.2	55.1	53.2	49.9	53.0	49.3	51.1
Neighborhood SES	49.8	49.7	52.8	52.3	51.9	51.5	51.2	49.9
	(22.9)	(22.6)	(21.4)	(21.7)	(22.0)	(20.5)	(20.0)	(19.8)
Married	79.8	73.0	75.5	73.3	68.8	76.0	77.3	69.5
Panel B: Father Characteristics								
Age	41.0	39.6	40.6	40.0	36.9	41.5	38.0	37.8
	(6.9)	(7.3)	(7.1)	(7.2)	(7.2)	(6.5)	(7.1)	(7.1)
Adjusted Song Score	49.1	49.1	51.8	50.8	49.2	47.8	47.3	47.3
	(28.6)	(28.2)	(26.4)	(26.8)	(27.3)	(27.2)	(27.0)	(27.1)
White Collar	7.2	8.7	10.0	11.0	11.6	13.3	14.6	16.5
Farmer	51.8	48.4	49.4	48.9	47.0	44.4	42.3	37.9
Unskilled	15.1	18.8	26.1	26.7	28.3	26.8	27.9	29.1
Skilled	12.8	14.6	14.3	13.2	13.0	15.2	15.0	16.3
N	1,320,179	2,243,961	3,233,320	4,573,210	3,235,326	5,918,861	7,833,114	7,055,545

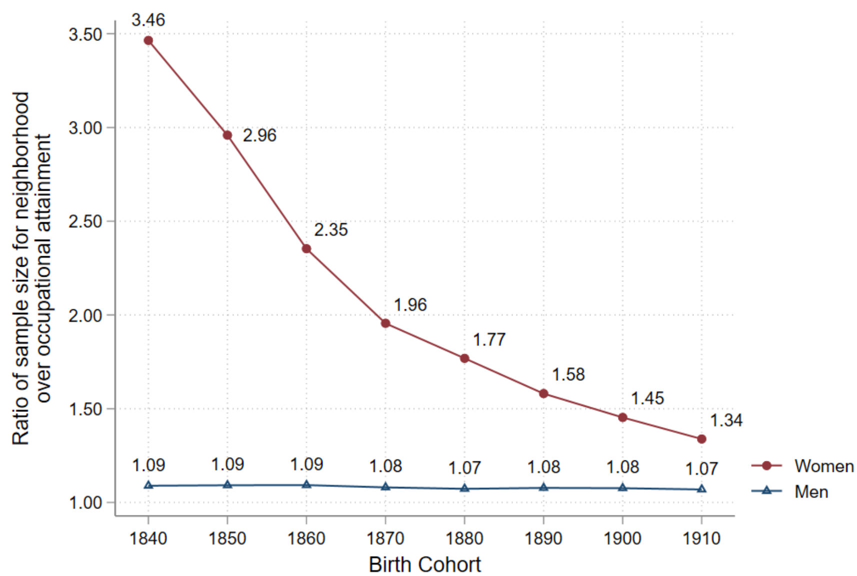
Notes: This table shows the descriptive statistics of the sample, with standard deviations in parentheses. Each column reports statistics by the child's birth cohort, rounded to the nearest decade. The table is weighted to be representative on observables following the process of Bailey et al. (2020). 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.

Figure 1: Neighborhood-SES is observable for more people than adjusted Song score, especially single women.

Panel A: Married

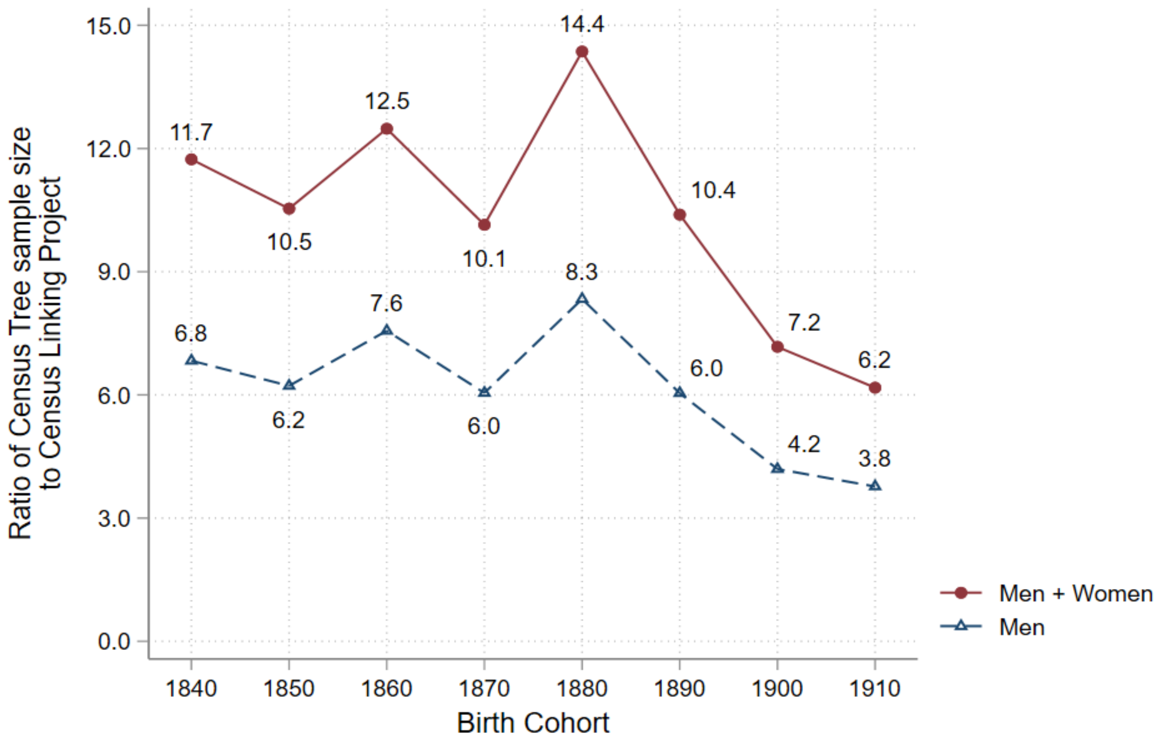


Panel B: Single



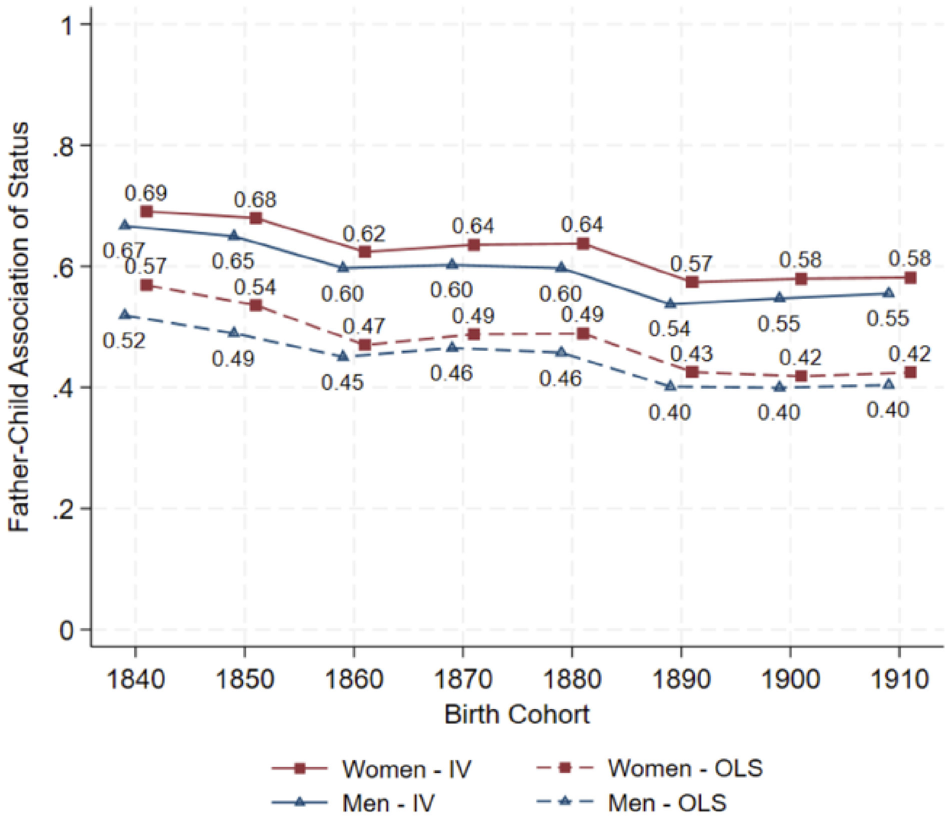
Notes: These figures show the ratio of sample sizes when using neighborhood status (n-SES) as a proxy for a child's status, compared to using occupational status with the adjusted Song score. For example, in the 1840 birth cohort, there are 3.46 times more single daughters for whom we can measure n-SES than occupational status (Panel B). The ratio is lower for married women in Panel A because we use their husband's occupation to proxy for their status, and most husbands reported an occupation.

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 5.2 times more male links in the Census Tree intergenerational dataset, and 8.7 time more male and female links.

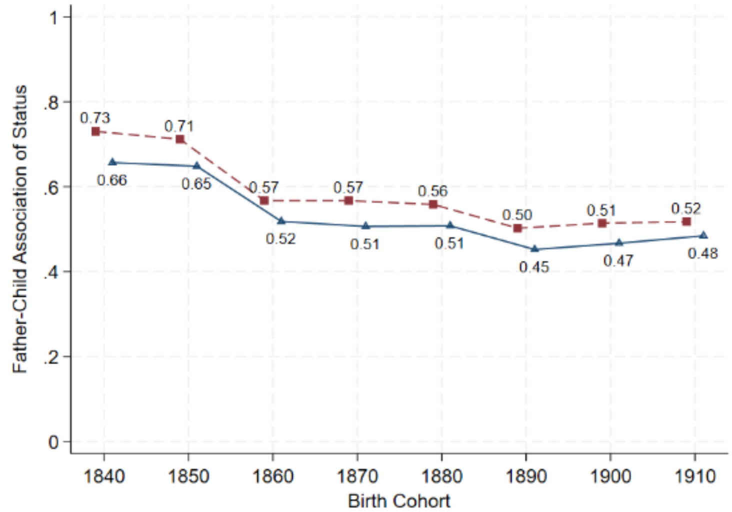
Figure 3: Estimates of intergenerational mobility for 1840-1910 birth cohorts, by gender.



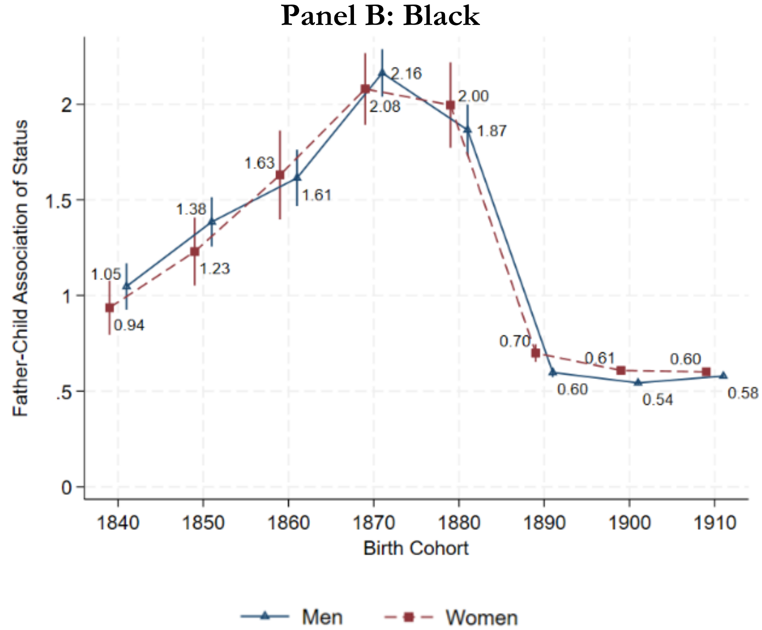
Notes: Figures show coefficients from a regression of the child's status (as measured by n-SES) on the father's (measured using Adjusted song). There are 21.0 million father-son links and 14.4 million father-daughter links in the figure. The IV estimates use a second observation for the father's status as an instrument for the first. All estimates are weighted to be representative of the population following the procedure outlined by Bailey et al. (2020).

Figure 4: Estimates of intergenerational mobility for 1840-1910 birth cohorts, by race.

Panel A: White

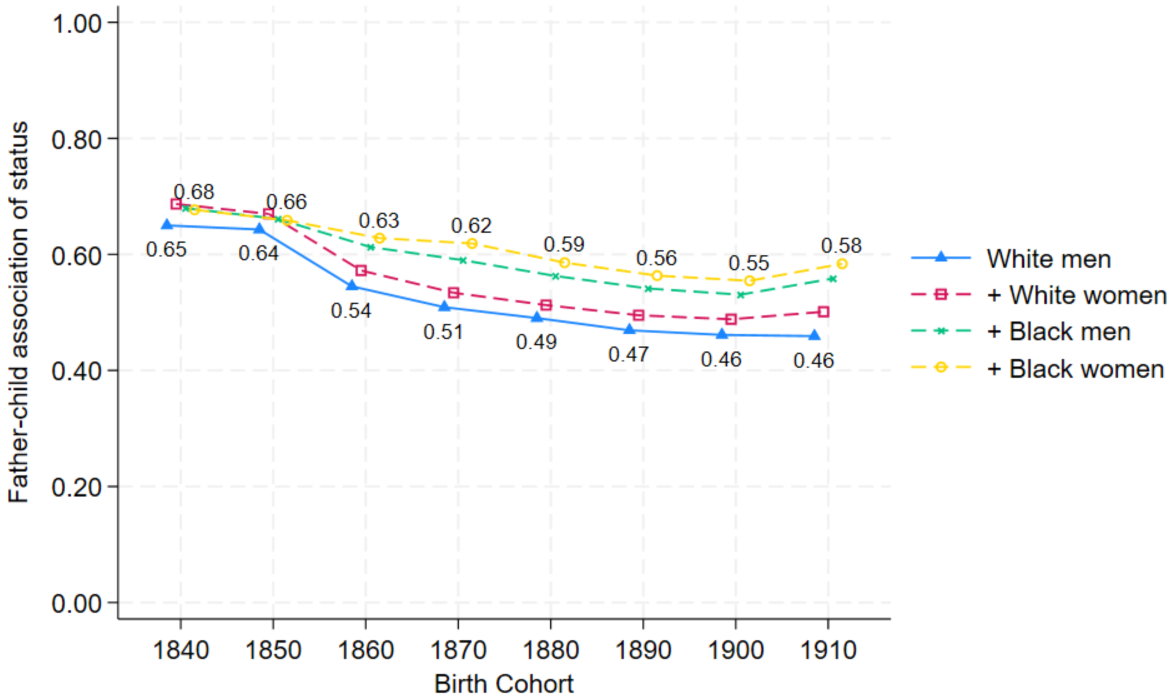


Panel B: Black



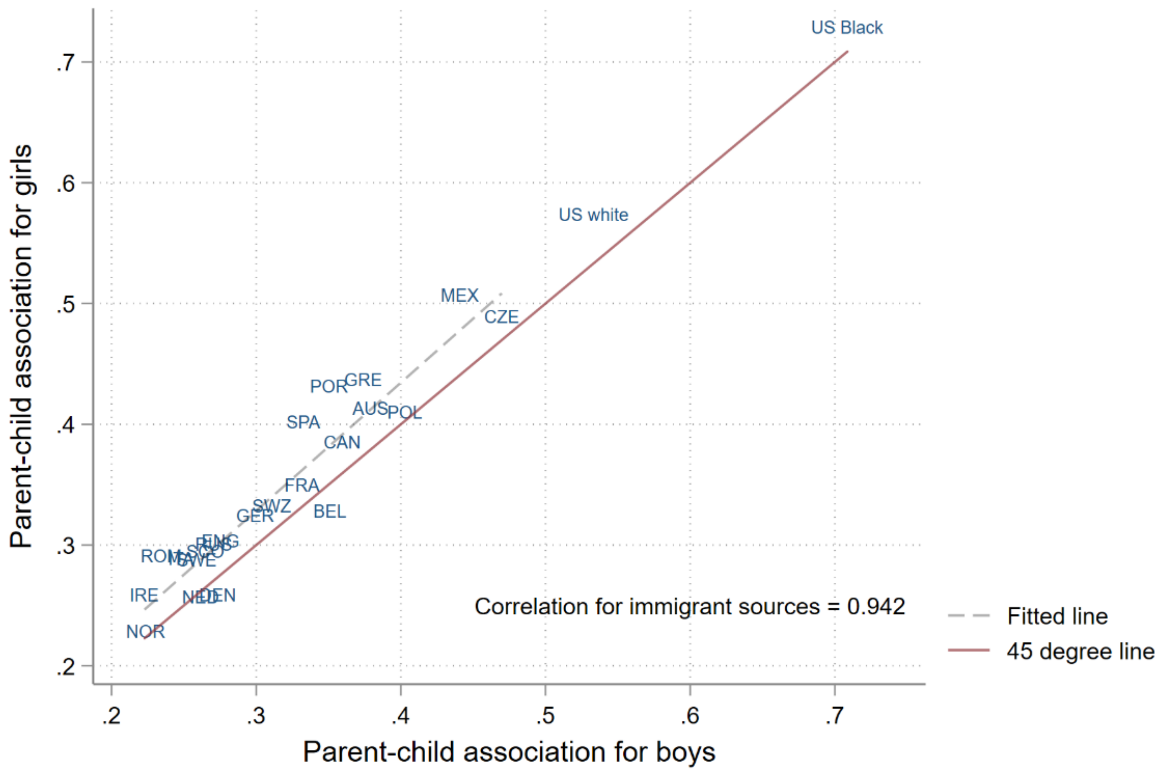
Notes: All estimates are from an IV regression of the child's neighborhood SES on the father's adjusted Song score, using a father's second observation as an instrument for the first. Estimates are weighted to be representative of the population following the procedure outlined by Bailey et al. (2020).

Figure 5: Including women in the sample raises estimates of intergenerational transmission.



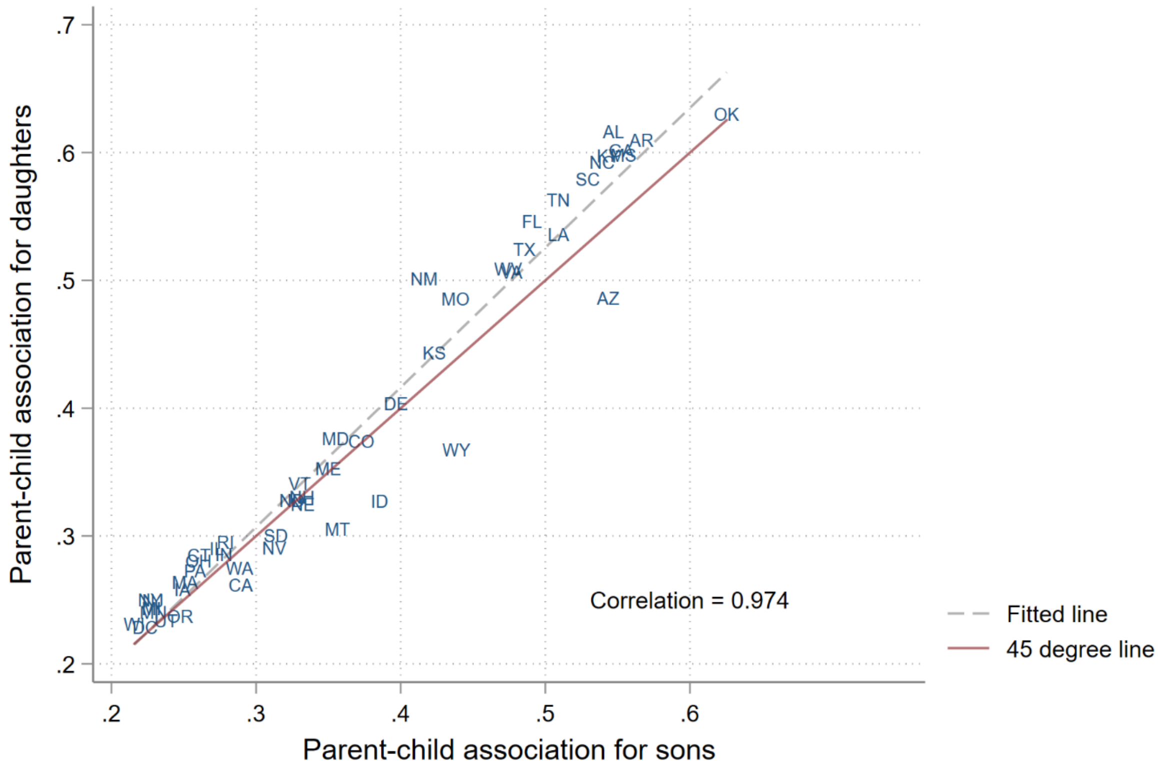
Notes: This figure shows how mobility estimates change when adding more groups to the sample. The bottom line is a sample of only White men; the line above pools White women into the sample; the line above that adds Black men into the sample; and the top line finally adds Black women to get population-level estimates. All estimates are from an IV regression of the child's neighborhood SES on the father's adjusted Song score, using a father's second observation as an instrument for the first. Estimates are weighted to be representative of the population following the procedure outlined by Bailey et al. (2020).

Figure 6: Mobility estimates are similar across gender by father's country of birth



Notes: All estimates are from an IV regression of the child's neighborhood SES on the father's adjusted Song score, using a father's second observation as an instrument for the first. Estimates are weighted to be representative of the population following the procedure outlined by Bailey et al. (2020). Estimates are produced separately by father's country of birth. The correlation and fitted line is shown only for foreign-born sources. We include US-born White and US-born Black estimates for reference. For these estimates, we pool all birth cohorts together.

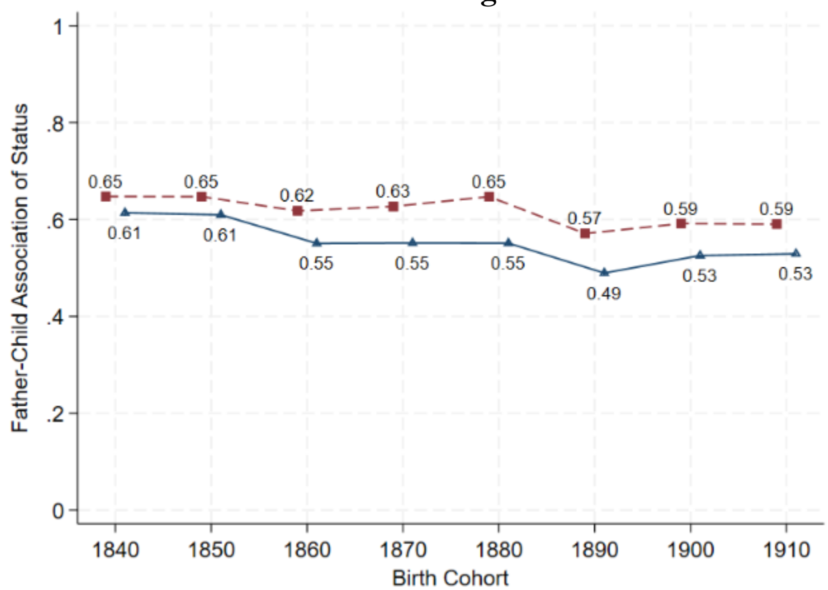
Figure 7: State-level estimates of mobility are similar across daughters and sons



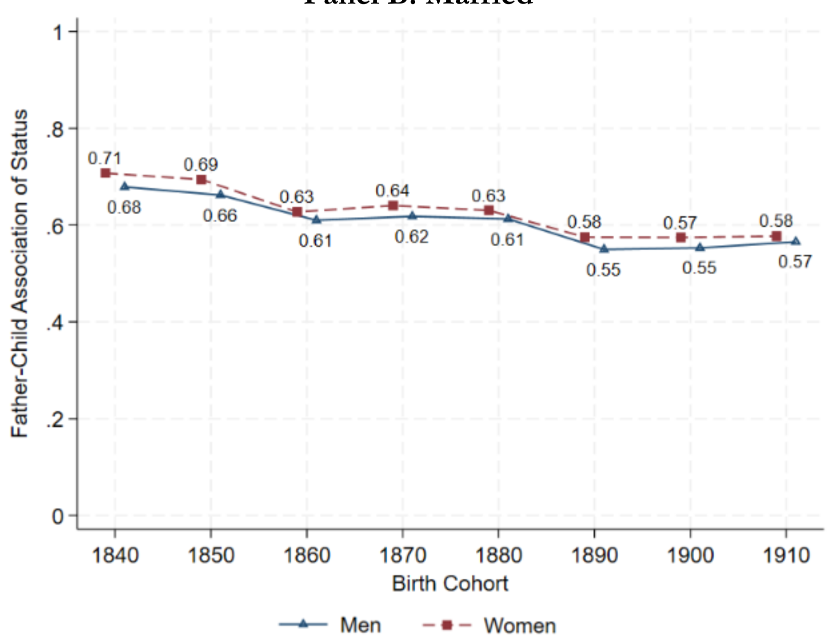
Notes: All estimates are from an IV regression of the child's neighborhood SES on the father's adjusted Song score, using a father's second observation as an instrument for the first. The sample is limited to White sons and daughters, and then weighted to be representative of the population following the procedure outlined by Bailey et al. (2020). Estimates are produced separately by childhood state of residence. Estimates above the 45-degree line imply that transmission is greater between father-daughter than father-son. For these estimates, we pool all birth cohorts together.

Figure 8: Estimates of intergenerational mobility for 1840-1910 birth cohorts, by marital status.

Panel A: Single

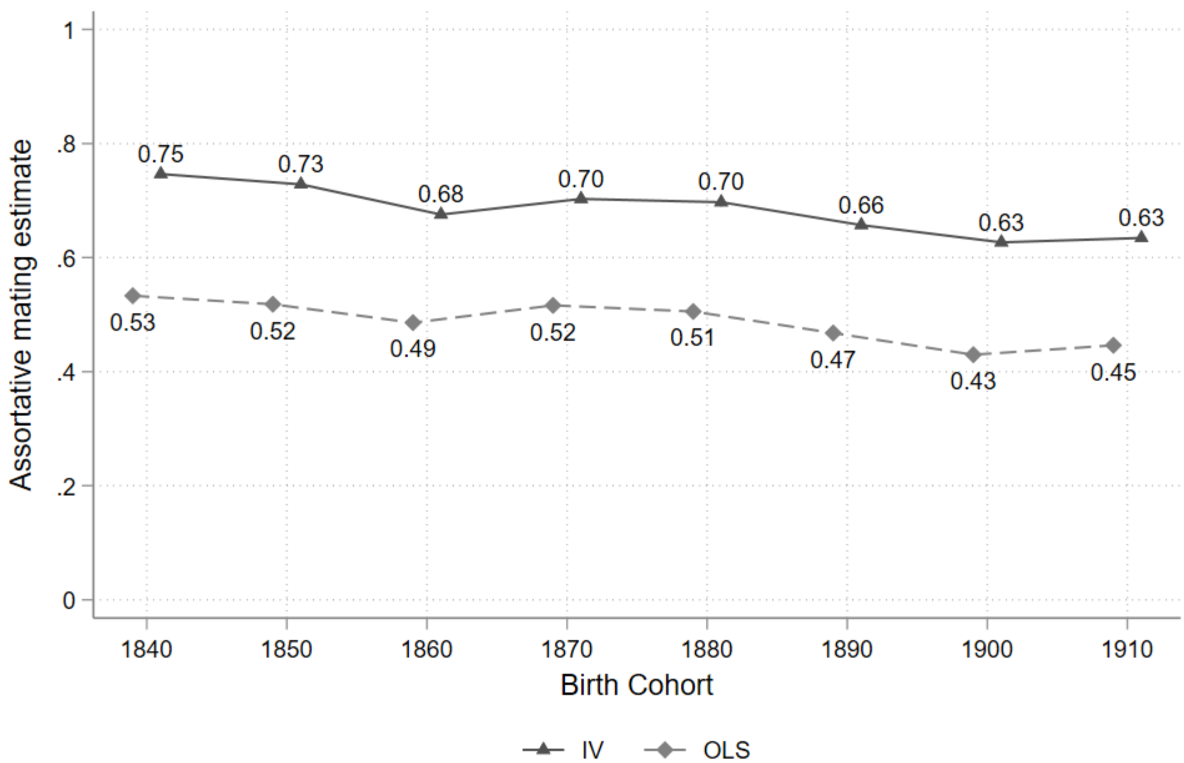


Panel B: Married



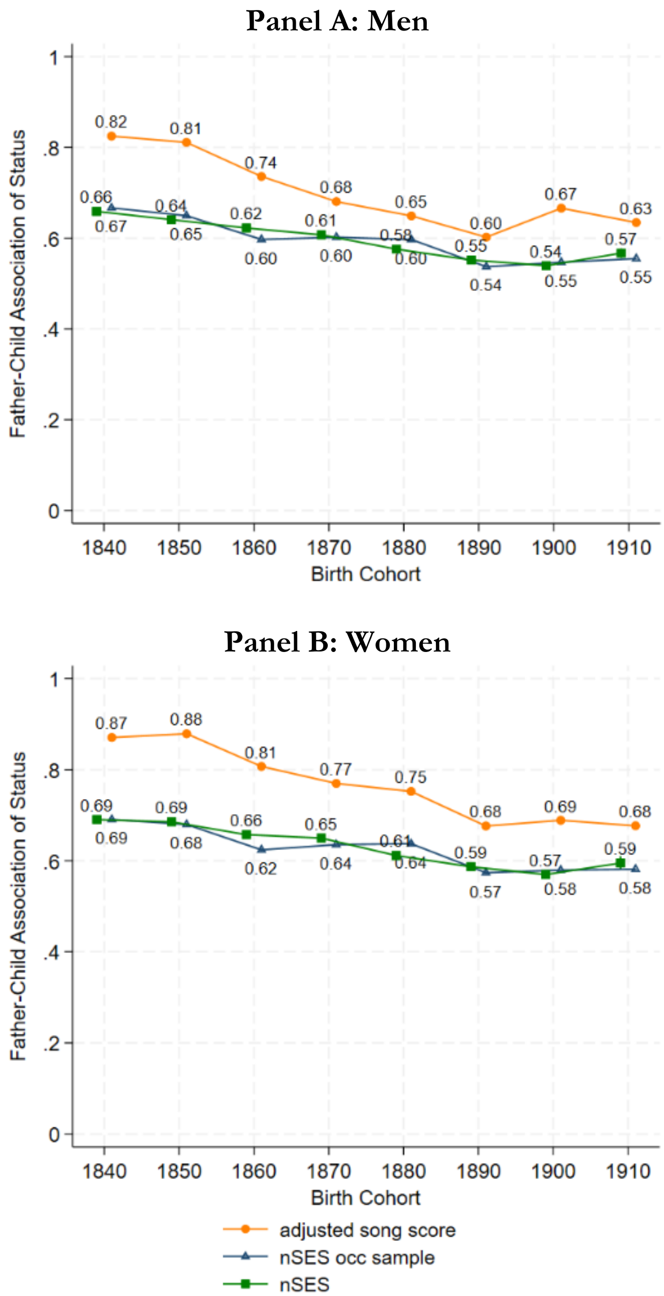
Notes: All estimates are from an IV regression of the child's neighborhood SES on the father's adjusted Song score, using a father's second observation as an instrument for the first. Estimates are weighted to be representative of the population following the procedure outlined by Bailey et al. (2020).

Figure 9: Assortative mating was strong.



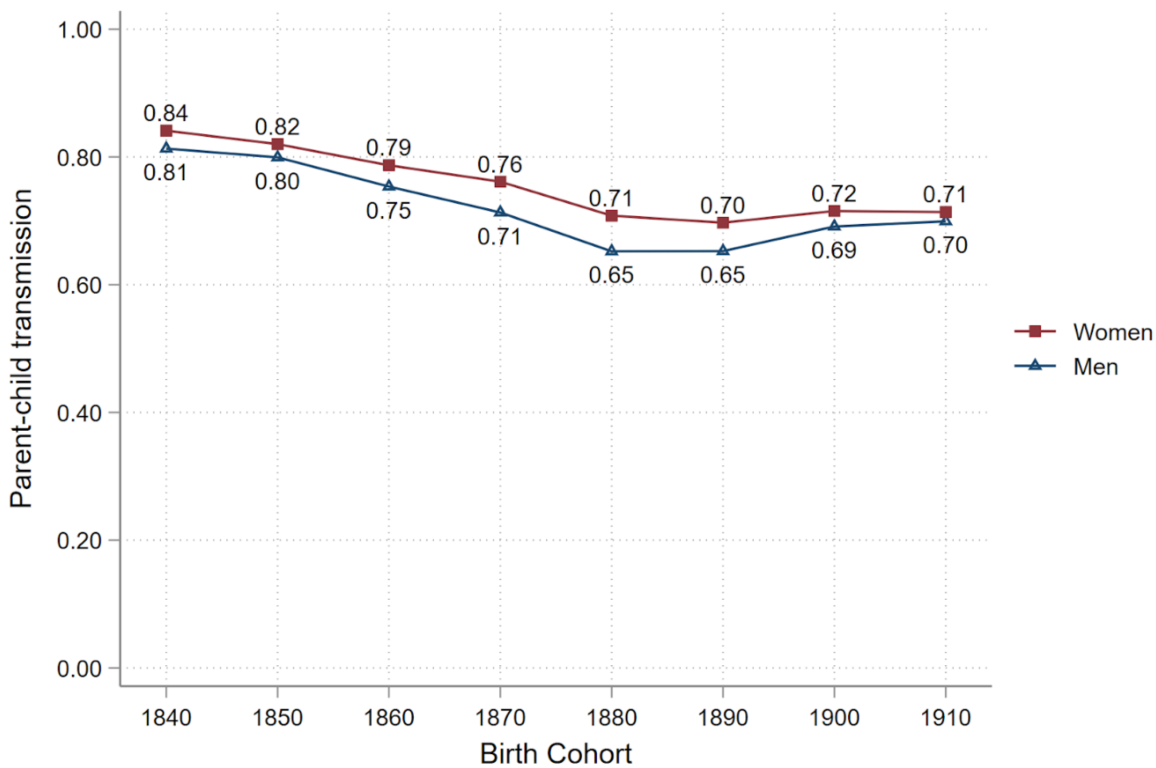
Notes: Coefficients are from regressions of the husband's father's status on the wife's father's status, for married couples where both are observed. The measure of status is the adjusted Song score. The IV regressions instrument the primary observation for the wife's father's status with a second observation. 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).

Figure 10: Comparison of intergenerational mobility estimates when using neighborhood SES (n-SES) and adjusted Song score as measure of child's status.



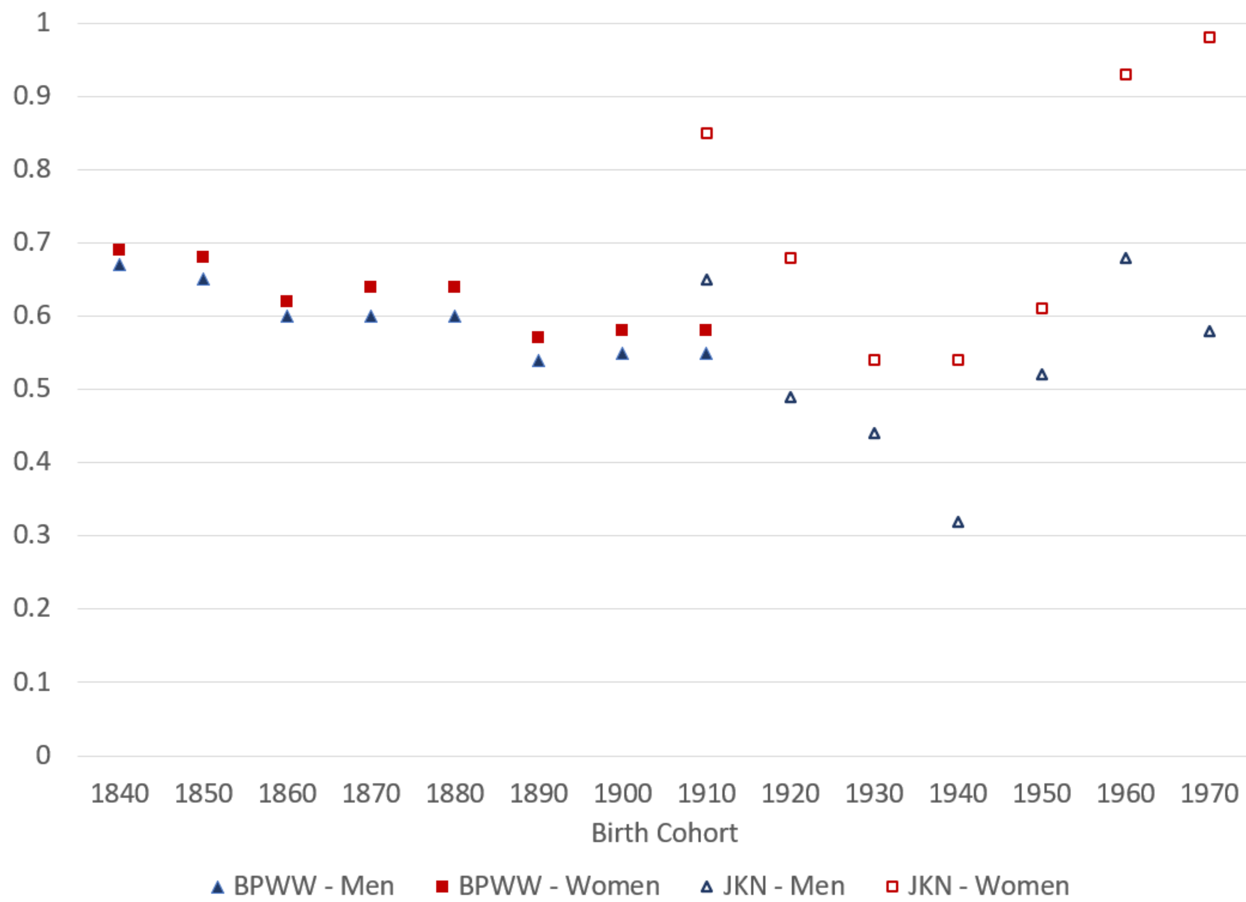
Notes: All estimates are from an IV regression of the child's SES on the father's adjusted Song score, using a father's second observation as an instrument for the first. Estimates are weighted to be representative of the population following the procedure outlined by Bailey et al. (2020). When using the adjusted Song as the measure of the child's SES, we calculated the score for sons using their own occupation, and for daughters using her husband's if she is married and her own if she is single. We drop those who do not report an occupation. The green line shows estimates using the n-SES as the measure of status, but limiting the sample to those for whom we can construct an adjusted Song score using the procedure described.

Figure 11: Comparison to estimates using a name-based strategy



Notes: Coefficients are from a regression of the child's neighborhood status on the average of fathers' adjusted Song Scores. The average is taken from fathers of the children with the same first name, sex, and birthplace of the child 20-40 years prior. This approach follows Olivetti and Paserman (2015) and uses only cross-sectional full-count censuses rather than directly linked data. The total number of observations ranges from 8,417,961 in the 1840 cohort and 39,982,198 in the 1890 cohort.

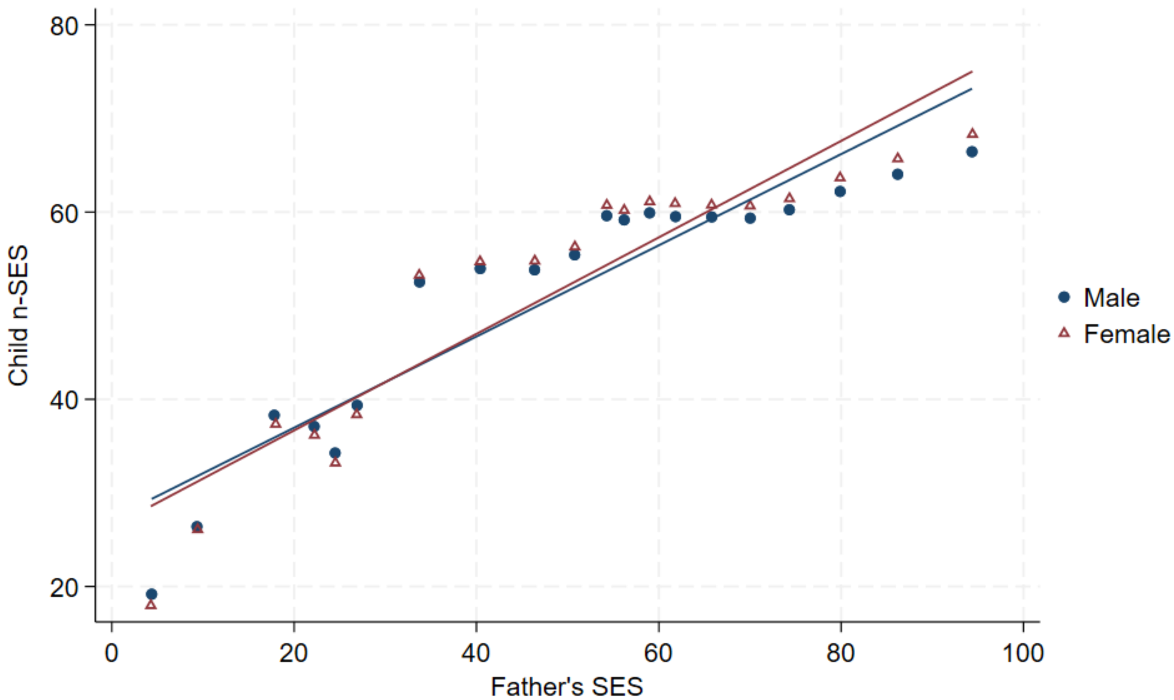
Figure 12: Intergenerational mobility for 1840-1970 birth cohorts, combining our estimates with those in Jácome, Kuziemko, and Naidu (forthcoming)



Notes: Estimates from the current paper (BPWW, solid markers) are the same as in Figure 3. The estimates from Jácome, Kuziemko, and Naidu (JKN, open markers) are taken from their Table A.6. JKN regress the log of the child's household income as an adult on a measure of the parents' household income when the child was approximately age 10, constructed using the father's occupation, race, and region.

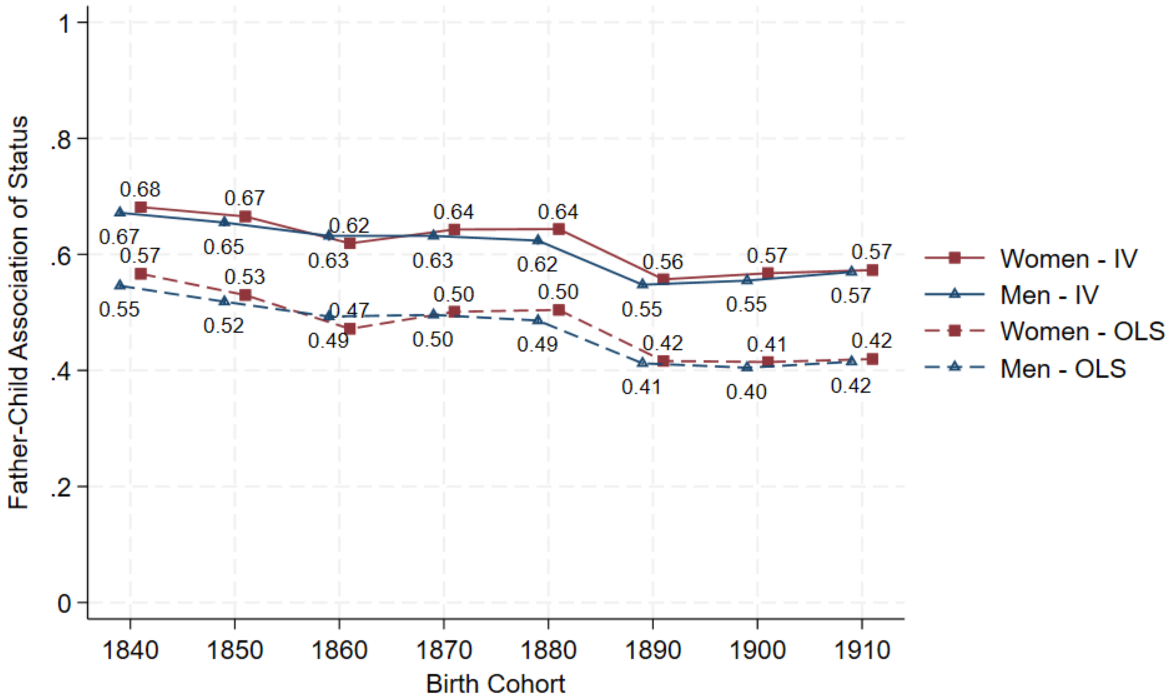
Online Appendix

Figure A1: Binscatter plot for mobility estimates by gender



Notes: This figure shows the binscatter relationship between the child's n-SES and the father's adjusted Song Score. Estimates are weighted to be representative of the population following the procedure outlined by Bailey et al. (2020).

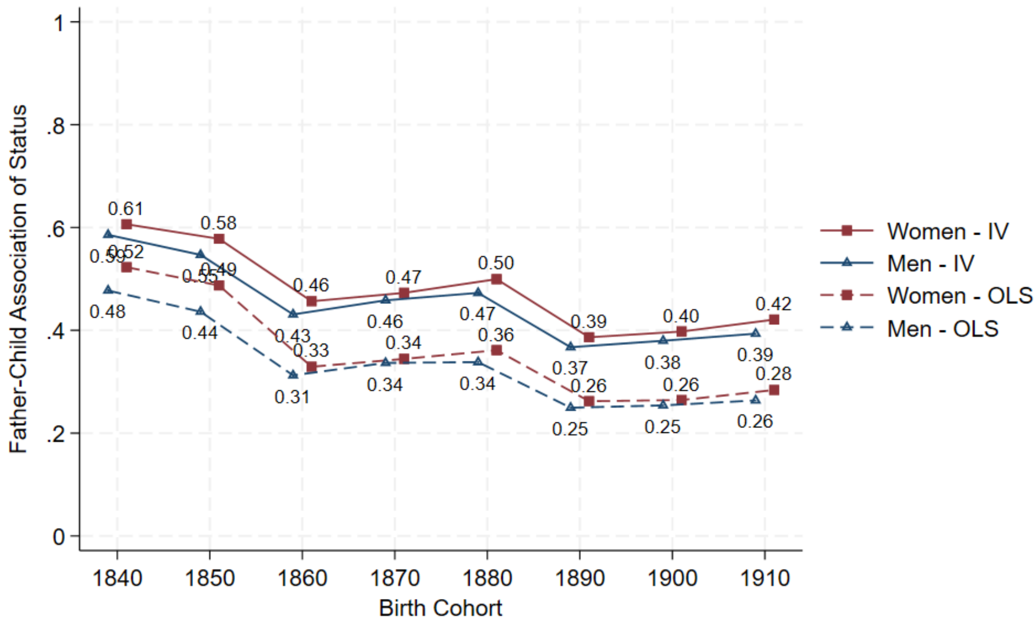
Figure A2: Replicating main IGT estimates using only Family Tree links



Notes: This figure reproduces the mobility estimates in Figure 3 using only links from the Family Tree (that is, made by FamilySearch users).

Figure A3: Additional specifications for main results

Panel A: Drop farmers



Panel B: Unweighted

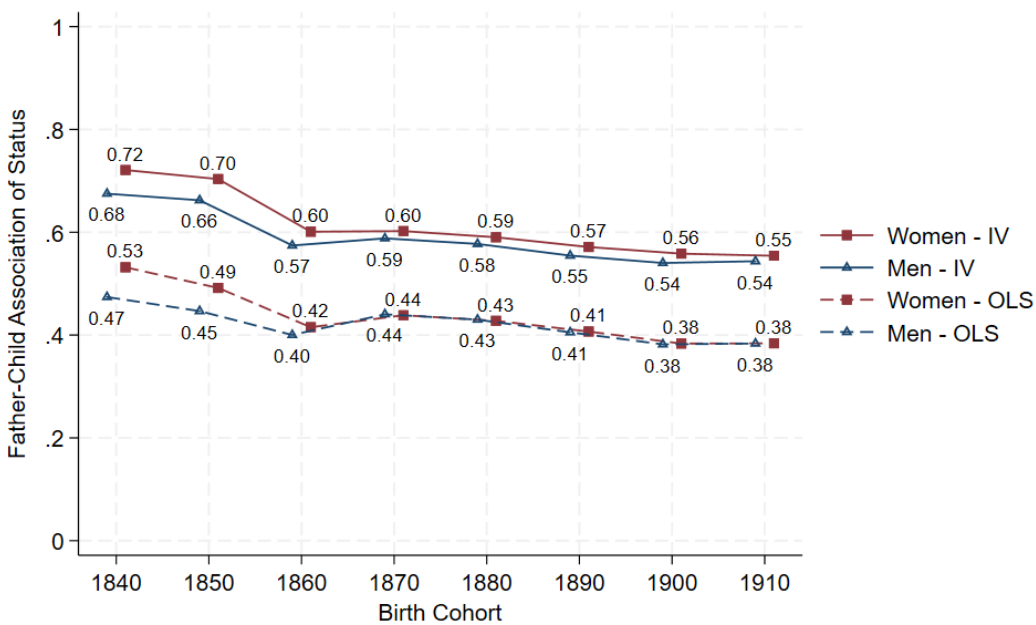
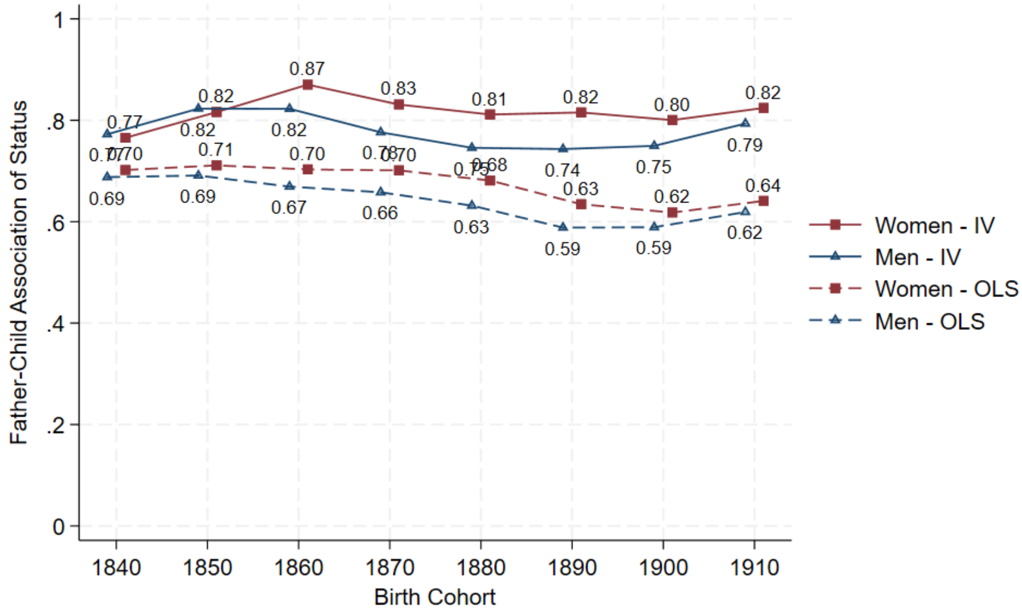


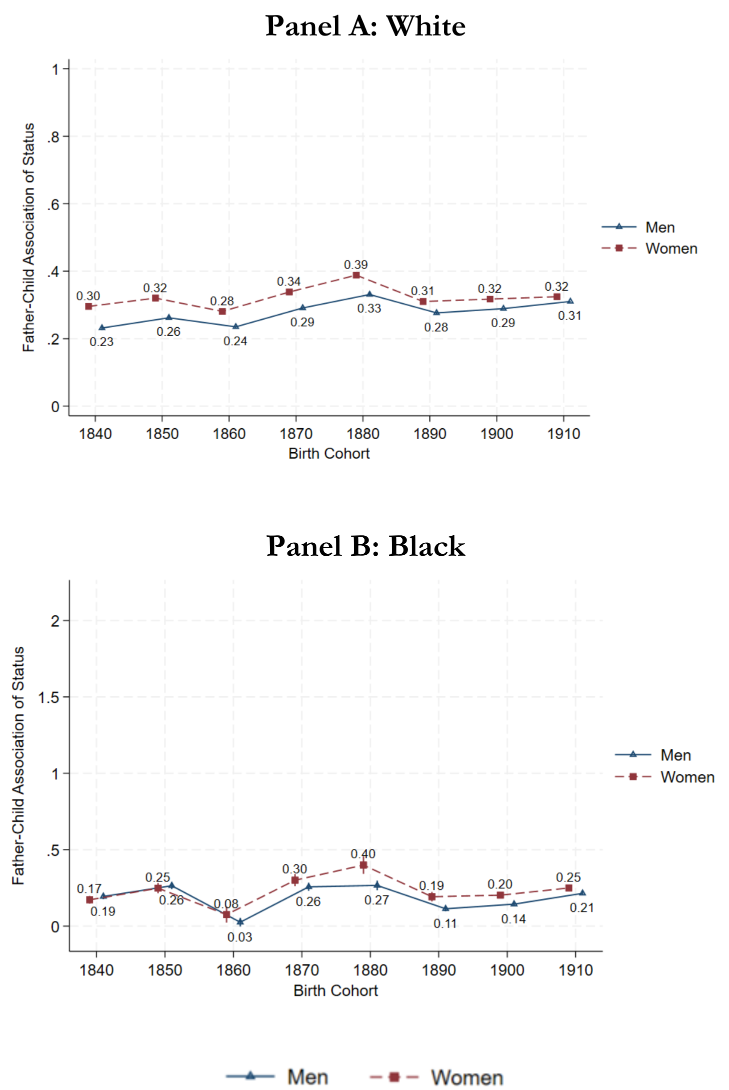
Figure A3: Additional specifications for main results (continued)

Panel C: Using father's n-SES as his measure of status



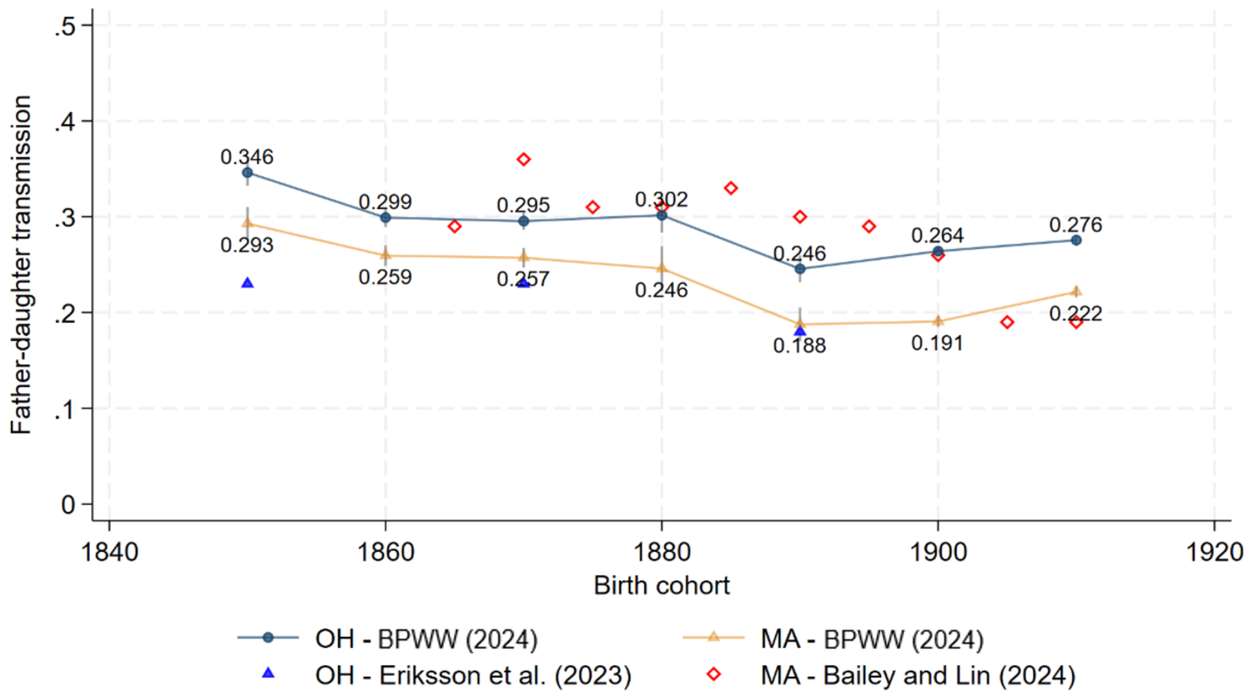
Notes: Figure A3 reproduces the estimates in Figure 3 but after dropping fathers who are farmers (Panel A), without weights (Panel B), and when using the father's n-SES as his measure of status but keeping the sample the same (Panel C).

Figure A4: Estimates of intergenerational mobility for 1840-1910 birth cohorts, by race, using occupation-only score



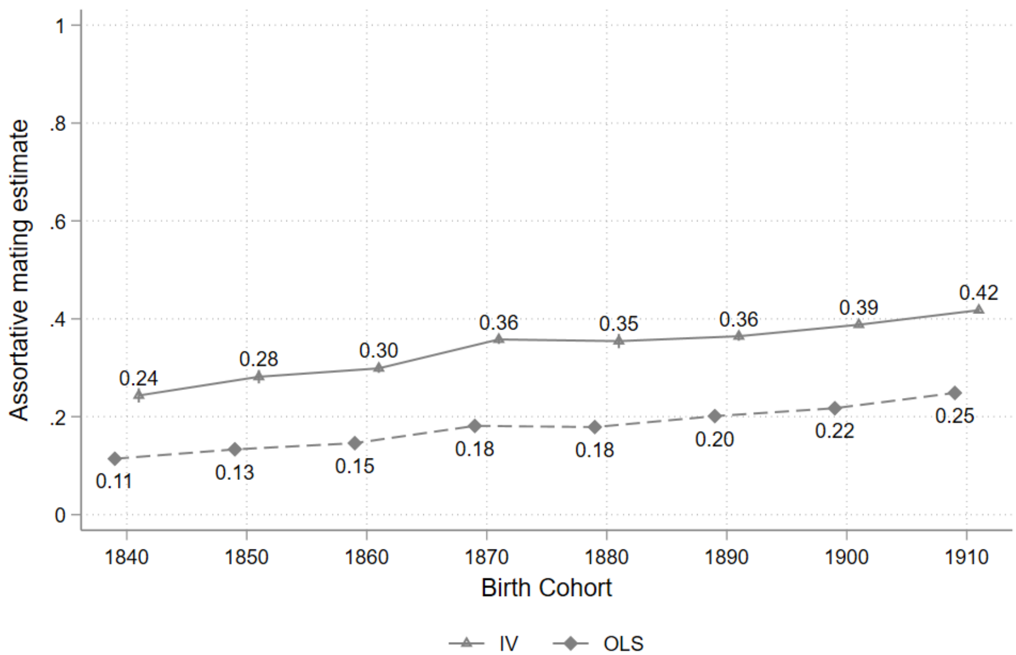
Notes: All estimates are from an IV regression of the child's neighborhood SES on the father's Song score, using a father's second observation as an instrument for the first. Estimates are weighted to be representative of the population following the procedure outlined by Bailey et al. (2020).

Figure A5: Comparing our IGT estimates for Massachusetts and Ohio to Eriksson et al. (2023) and Bailey and Lin (2024)



Notes: The BPWW (2024) estimates are from this paper, for an IV regression of the child's neighborhood SES on the father's Song score, using a father's second observation as an instrument for the first. Estimates are weighted to be representative of the population following the procedure outlined by Bailey et al. (2020). We limit the sample to those living in Ohio or Massachusetts in childhood. We compare to rank-rank estimates from Eriksson et al. (2023) and Bailey and Lin (2024), which are taken from Figure 7 of Bailey and Lin (2024).

Figure A6: Assortative mating estimates when using fathers' occscore



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 7, but instead of using the fathers' adjusted Song scores as their measure of status, we use the IPUMS variable *occscore*.

Table A1: Other estimates of assortative mating in the literature

Paper	Estimate	Status measure	Estimation	Population
BPWW (2024)	0.63-0.73	Occ, race, region	IV	White and Black
BPWW (2024)	0.43-0.52	Occ, race, region	OLS	White and Black
BPWW (2024)	0.25-0.42	Occ	IV	White and Black
BPWW (2024)	0.12-0.25	Occ	OLS	White and Black
Althoff et al. (2024)	0.78-0.88	Human capital rank	Semiparametric latent variable method	White and Black
Bailey and Lin (2024)	0.25-0.35	Occ	OLS	Ohio marriages
Espín-Sanchez et al. (2023)	0.416	Occ. with farmer adjustment	Non-linear IV	White and Black
Eriksson et al. (2023)	0.20-0.29	Occ	OLS	White MA marriages
Olivetti et al. (forthcoming)	0.04-0.09	Occ	First-name method	White

Notes: BPWW (2024) is the current paper. The estimates are taken from Figure A7 in Althoff et al. (2024)., Figure D.1 in Bailey and Lin (2024), Figure 4 in Eriksson et al. (2023), and Figure 2 in Olivetti et al. (forthcoming). The column "Status measure" lists the characteristics that are used to construct the measure of socioeconomic status, where "Occ" refers to occupation.

Table A2: Gender differences in mobility are also small when using educational attainment as the dependent variable

	I	II	III	IV	V	VI
Female x Father's Adj. Song Score	-0.0278	0.00361	-0.0160	-0.0363	0.00167	-0.0515
	(0.000933)	(0.000961)	(0.00715)	(0.00131)	(0.00151)	(0.0122)
Father's Adjusted Song Score	0.345	0.266	0.571	0.473	0.405	0.816
	(0.000549)	(0.000617)	(0.00414)	(0.000779)	(0.000974)	(0.00730)
Female	4.917	2.872	8.219	5.142	2.846	8.209
	(0.0567)	(0.0590)	(0.134)	(0.0723)	(0.0847)	(0.164)
Constant	43.79	48.88	31.59	37.82	41.81	29.44
	(0.0326)	(0.0377)	(0.0622)	(0.0418)	(0.0545)	(0.0807)
Estimation method	OLS	OLS	OLS	IV	IV	IV
Population	All	White	Black	All	White	Black
Observations	13,771,139	13,216,844	554,295	13,771,111	13,216,819	554,292
R-squared	0.146	0.093	0.101	0.126	0.070	0.089

Notes: This table uses linked data where the child is observed in the 1940 Census to estimate differences in educational mobility between sons and daughters. The dependent variable is the percentile rank of years of education by birth cohort (rounded to the nearest decade, similar to the adjusted Song score). The first row tests whether the father's influence on educational outcomes differs between sons and daughters. Columns I and IV use the full sample, combining Black and White children, while Columns II and V are limited to White children, and Columns III and VI are limited to Black children. Columns IV-VI also use an instrumental variables approach, where the father's second adjusted Song score is used as an instrument for his first score. We also instrument for the interaction between gender (female) and the father's adjusted Song score using the interaction between gender and the father's second adjusted Song score.