The Missing Link
Women and Intergenerational Mobility*

Lukas Althoff† Harriet Brookes Gray‡ Hugo Reichardt§

[Most recent version here]

September 20, 2023

Abstract

The US has been considered the land of opportunity. However, representative estimates of intergenerational mobility are scarce as women's historical records are difficult to trace once their name changes upon marriage. This paper overcomes the challenge of linking women's census records using information from Social Security Number (SSN) applications. Those applications contain the maiden and married names of 60 million women—either as applicants or as applicants' mothers. Our new panel dataset covers an unprecedented number of women between 1850 and 1940 and opens a myriad of new opportunities to study women's role in the US economy. We document three important new facts. First, we construct representative estimates of intergenerational mobility from 1850 to 1940 from our new panel. We show that historically, women's intergenerational mobility tended to be higher than men's for Black and white Americans. Second, we highlight mothers' significant role in predicting their children's future outcomes, suggesting that traditional father-child comparisons overestimate intergenerational mobility. Third, we document a robust relationship between sorting in marriage markets and levels of intergenerational mobility across time and space. The more sorting there is in the marriage market, the more rigid the socioeconomic status of families across generations.

*We thank Leah Boustan, Ellora Derenoncourt, and Ilyana Kuziemko for insightful comments.
†Department of Economics, Princeton University. lalthoff@princeton.edu
‡Department of Economics, Princeton University. hbrookesgray@princeton.edu
§Department of Economics, London School of Economics. h.a.reichardt@lse.ac.uk

1. Introduction

Throughout modern history, the US has been considered the land of opportunity. Measures of intergenerational mobility—the extent to which the socioeconomic status of individuals changes from one generation to the next—reflect how access to opportunities depends on a person's parental resources. Assessing historical levels of intergenerational mobility typically requires tracing families' records over long time spans. Advances in the automated linking of historical records have overcome key difficulties to quantitatively assess intergenerational mobility of Americans over the 19th and 20th centuries (Abramitzky et al., 2019; Ward, 2021).

Virtually all long-run estimates of intergenerational mobility exclude women, leaving massive gaps in our knowledge about women's contribution to intergenerational mobility across US history. Incorporating women into the analysis has been challenging due to name changes up marriage. Recently, progress has been made in linking women's historical records by using the information of name changes from marriage certificates in some states (Craig et al., 2019) or estimating mobility directly from survey data that asks women about their socioeconomic status and that of their parents (Jácome et al., 2021). However, these data are often sparse, making it difficult to zero in on minorities or study heterogeneity across space; and some sources cover only a selected part of the population, such as married women.

In this paper, we link women's historical census records to study intergenerational mobility among *all* Americans from 1850 to 1940 and one of its main drivers—assortative mating. We combine two data sources, full-count census records and information from 41 million Social Security Number (SSN) applications, to trace millions of men and women over time. SSN applications cover the near universe of applicants who died between 1980 and 2007 and include information on applicants' names and their maiden names. Importantly, they also contain the maiden names of applicants' parents, massively expanding the sample, extending the coverage back in time, and increasing representativeness by including people who never applied for an SSN.[1]

Our new data ranges from 1850 to 1940 and consists of 36 million total links, half of which are women. Our data is highly representative along all dimensions, including income, race, and geography. We link an unprecedented 18 million women from before to after marriage, uniquely equipping us to study the long-run evolution of intergenerational mobility and assortative mating. This data will become publicly available, opening many new opportunities to study women's role in the US economy.

Based on our new panel, we document the evolution of intergenerational mobility

[1]Specifically, some occupations and industries were ineligible for Social Security immediately after the system's creation in 1935, after which the system gradually expanded to be near universal by the 1950s.

for all Americans—including men and women—between 1850 and 1940, suggesting that mobility was similar to what it is today. Based on proxies for household income for parents and children, we find rank-rank elasticities of 0.3 to 0.4. Existing estimates for more recent decades that include men and women are similar, ranging between 0.25 and 0.4 (Jácome et al., 2021; Chetty et al., 2014). Our estimates suggest that women tended to be more socially mobile than men, especially among cohorts born after 1870. Women born in the 1900s, for example, have a rank-rank elasticity in household income of 0.32—significantly lower persistence than 0.39 among men.

While father-child comparisons are the literature's standard measure of intergenerational mobility, they offer an incomplete picture of how interrelated a child's socioeconomic status is with that of their parents. In particular, unless mothers do not separately contribute to the future socioeconomic status of their children, father-child comparisons understate the persistence of socioeconomic status across generations. We extend the standard model of intergenerational mobility to flexibly allow fathers and mothers to co-determine the socioeconomic status of their children. When estimating the relationship between the socioeconomic status of a single parent and their child, one elasticity can capture intergenerational mobility. Multiple elasticities become relevant when considering both parents and their potential complementarity. As a result, it is harder to compare intergenerational mobility across time, space, and subgroups of the population.

We find that mothers as predictive of their children's outcomes as fathers. A mother's literacy status is *more* predictive of their children's literacy status, but *less* predictive of their children's income. This result is consistent with mothers influencing their children's future socioeconomic status through direct human capital transmission, whereas fathers may tend to affect their children's socioeconomic status through transmission of occupation-specific skills or employment networks. To measure of intergenerational mobility in the presence of proxies for the socioeconomic status of both parents, we propose using the variance in children's outcomes explained by both parents' socioeconomic status (adjusted R-squared).

We next explore a key potential driver of intergenerational mobility: assortative mating. Historically, women tended not to participate in the labor market, so that their economic status in adulthood was often determined by whom they married. Consistent with the importance of marriage markets, we document that intergenerational mobility and assortative mating levels are highly correlated over time and across space. In states with highly assortative marriages—i.e., where wives come from a very similar socioeconomic background as their husbands—intergenerational mobility is low. Across time, birth cohorts that tend to be more assortatively mated are less mobile, too. These results suggest that the marriage market may play a key role in shaping the economic opportunities available to men and women.

This paper contributes to our understanding of intergenerational mobility throughout

American history. For recent decades, intergenerational mobility has been well documented for men and women using administrative data containing unique identifiers. For example, Chetty et al. (2014) found that intergenerational mobility was relatively stable for people born between 1971 and 1982 (rank-rank elasticities around 0.3). Longer-run estimates typically include only men due to the difficulty of linking historical records of women in the absence of unique identifiers. For example, Abramitzky et al. (2021) documented lower intergenerational mobility for American men born around 80 years earlier, between 1880 and 1910 (rank-rank elasticities around 0.4). Based on a synthetic panel using men's and women's first names, Olivetti and Paserman (2015) document trends but not levels of intergenerational mobility. They find relatively stable mobility between 1870 and 1900, followed by a sharp decrease from 1900 to 1920 and another mild decline until 1940. Jácome et al. (2021) made substantial progress in documenting intergenerational mobility at the individual level for both men and women historically. They found that mobility increased from the 1910 to 1940 birth cohorts before plateauing compared to the 1970 cohorts. Black women's changing economic status was a critical driver of increasing mobility over this period. Craig et al. (2019) found that women's mobility was higher than men's in the late 1800s but this gap disappeared by 1900. Our sample is unprecedented in scope and representativeness in the historical context, providing new insights into the intergenerational mobility of women across time, space, and race.

We contribute to the literature on assortative mating, particularly historical patterns of assortative mating in US during the 19th and 20th centuries. Eika et al. (2019) find that marriages in the US have been increasingly assortative from the start of their sample period in 1940 until today. We document trends in assortative mating from 1850 to 1940, showing that marriages have become increasingly assortative since at least 1850, suggesting these trends are not purely a recent phenomenon. Clark and Cummins (2022) show that in England, assortative mating has been high since the 18th century. The authors show that both fathers and mothers predict the socioeconomic status of their children, suggesting that the high levels of assortative mating they observe lead to lower intergenerational mobility. We confirm the negative correlation between a generation's degree of assortative mating and its levels of intergenerational mobility across time and space in the US.

This paper also advances our understanding of women's role in the development of America's economy in the 19th and 20th century. Our sample covers the decades before the rapid rise in female labor force participation that began in the 1940s (Goldin, 1990, 1991; Goldin and Olivetti, 2013). Consistent with mothers thus likely spending more time with their children than fathers, our evidence suggests that human capital transmission was particularly strong from mothers to children. Using a schooling reform in Norway, Black et al. (2005) find that the *causal* effect of additional parental education on their children can be detected only between mothers (not fathers) and sons (not daughters).

In contrast to human capital, occupation and income seem to be more influenced by fathers: a person's earnings are better predicted by their paternal than their maternal grandfather's earnings.

This paper also contributes to a nascent effort to link women's historical census records. Abramitzky et al. (2019, 2020) are pioneers of automated record linking for men based on names. Efforts to include women are in their infancy.[2] Craig et al. (2019) use marriage certificates from Massachusetts (1850-1910) to link around 60 thousand married women in the censuses of 1880 or 1910 to their childhood households in the censuses of 1850 or 1880. Bailey et al. (2022) combine a more extensive set of vital records on births, marriages, and deaths for the states of North Carolina and Ohio, resulting in 300 thousand links of married women in the 1900-1940 censuses to their childhood households in the 1880-1920 censuses. Our panel includes 18 million married and unmarried women born between 1850 and 1940 across the entire US. Our panel is not only unprecedented in size but it is also highly representative in all dimensions. Fetter et al. (2021) also use the social security data leveraged to construct our panel, linking social security applicants to their fathers in the 1930 census. Our panel dataset ranges over all census decades available and leverages applicants and their parents. By leveraging information in addition to the census, we also augment the linking procedure developed by Abramitzky et al. (2019) to improve accuracy and coverage. Conditioning not only a person's own but also their parents' characteristics allows us to identify individuals in their childhood homes with high confidence.

2. Linking the Historical Records of Women

The main empirical challenge in studying the long-run evolution of intergenerational mobility is the lack of suitable panel data. In this section, we describe how we overcome this hurdle by combining census data with administrative records that contain the maiden names of 60 million women.

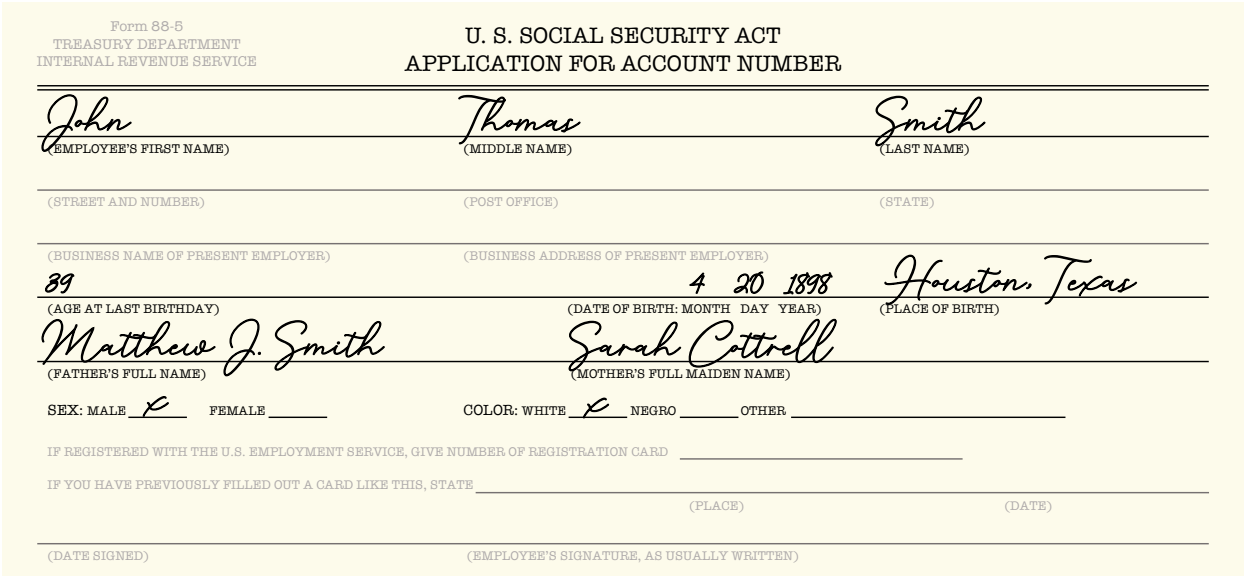
2.1 Data on the Maiden Names of Women

The key challenge to follow women's historical records is to identify them during childhood and adulthood despite name changes upon marriage. In modern administrative data, unique identifiers such as social security numbers or taxpayer identification numbers allow researchers to trace individuals' records, irrespective of potential name changes. Such identifiers do not exist in historical census data, which is the main source of individuals' socioeconomic status in early American history.

[2] Olivetti and Paserman (2015) constructed synthetic panels based on the first names of men and women.

We link the full-count censuses from 1850 to 1940 to 41 million Social Security applications to overcome the challenge of linking women's records despite name changes.[3] These applications contain the near universe of Americans who applied for a Social Security Number (SSN) and—for data privacy reasons—died before 2008. The data includes a rich set of information including each applicant's name, age, race, place of birth, and the names of their parents (see Figure 1). Most importantly, the data includes the individual's and their mother's maiden names. By linking those applications to the census, we assign each person in the census a unique identifier (SSN), allowing us to trace their historical records despite potential name changes upon marriage.

FIGURE 1: Social Security Application Form


Form 88-5
TREASURY DEPARTMENT
INTERNAL REVENUE SERVICE

U. S. SOCIAL SECURITY ACT
APPLICATION FOR ACCOUNT NUMBER

John (EMPLOYEE'S FIRST NAME) Thomas (MIDDLE NAME) Smith (LAST NAME)

(STREET AND NUMBER) (POST OFFICE) (STATE)

(BUSINESS NAME OF PRESENT EMPLOYER) (BUSINESS ADDRESS OF PRESENT EMPLOYER)

39 (AGE AT LAST BIRTHDAY) 4 20 1898 (DATE OF BIRTH: MONTH DAY YEAR) Houston, Texas (PLACE OF BIRTH)

Matthew J. Smith (FATHER'S FULL NAME) Sarah Cottrell (MOTHER'S FULL MAIDEN NAME)

SEX: MALE X FEMALE ____ COLOR: WHITE X NEGRO ____ OTHER ____

IF REGISTERED WITH THE U.S. EMPLOYMENT SERVICE, GIVE NUMBER OF REGISTRATION CARD ____

IF YOU HAVE PREVIOUSLY FILLED OUT A CARD LIKE THIS, STATE ____ (PLACE) (DATE)

(DATE SIGNED) (EMPLOYEE'S SIGNATURE, AS USUALLY WRITTEN)

Notes: This figure shows an sketch of a filled-in Social Security application form. Besides the applicants' name, address, employer, year and state of birth, and race, the application includes the father's name and the mother's maiden name.

2.2 Our Panel of 36 Million Americans

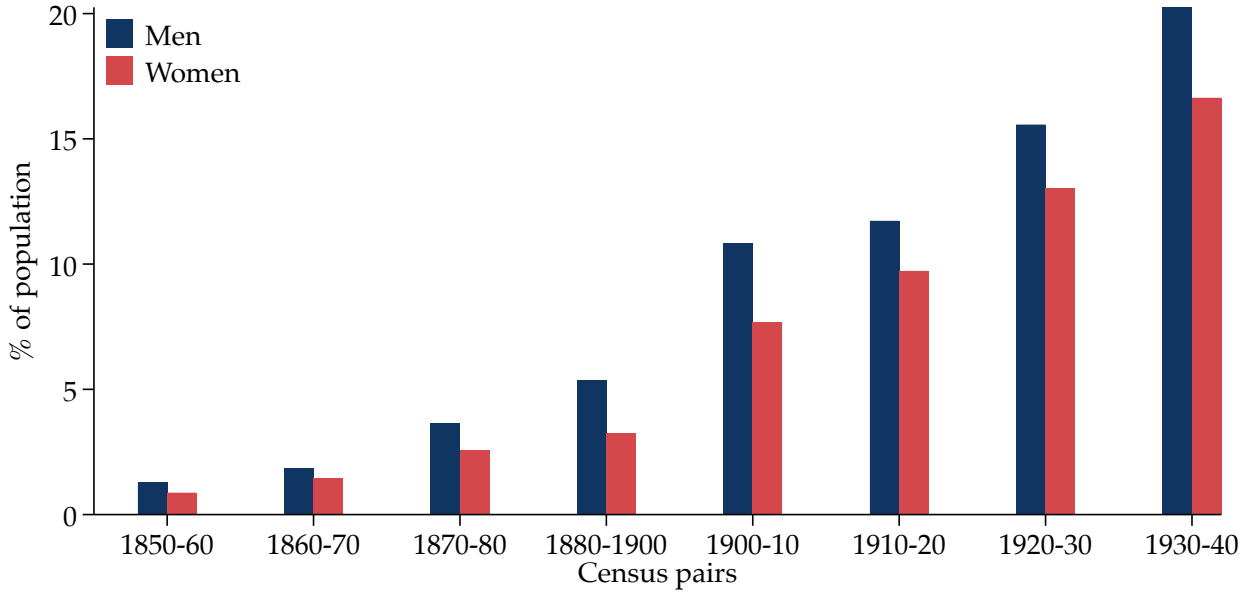
We link 36 million individuals from the Social Security records to the census—16 million applicants and 20 million parents of applicants. With 41 million applicants in the Social Security records, our implied linking rate is 40 percent. This linking rate is considerably higher than standard linking rates of around 25 percent in the literature as a result of the rich information that we leverage to find unique matches. These links assign each census record a Social Security Number, identifying individuals regardless of name changes.

In a second step, we create 112 million links between census records across time, covering large shares of the US population from 1850 to 1940 (see Figure 2).[4] Our coverage

[3] We obtained the applications from the Social Security Numerical Identification (Numident) file through the National Archives and Records Administration (NARA).

[4] For absolute numbers of links by census year and sex, see Appendix Figure A.1.

FIGURE 2: Fraction of US Population Linked in Our New Panel



Notes: This figure shows the fraction of the full population of men and women that we successfully link from one census decade to the next. Our empirical analysis also leverages links across non-adjacent census pairs, further increasing coverage.

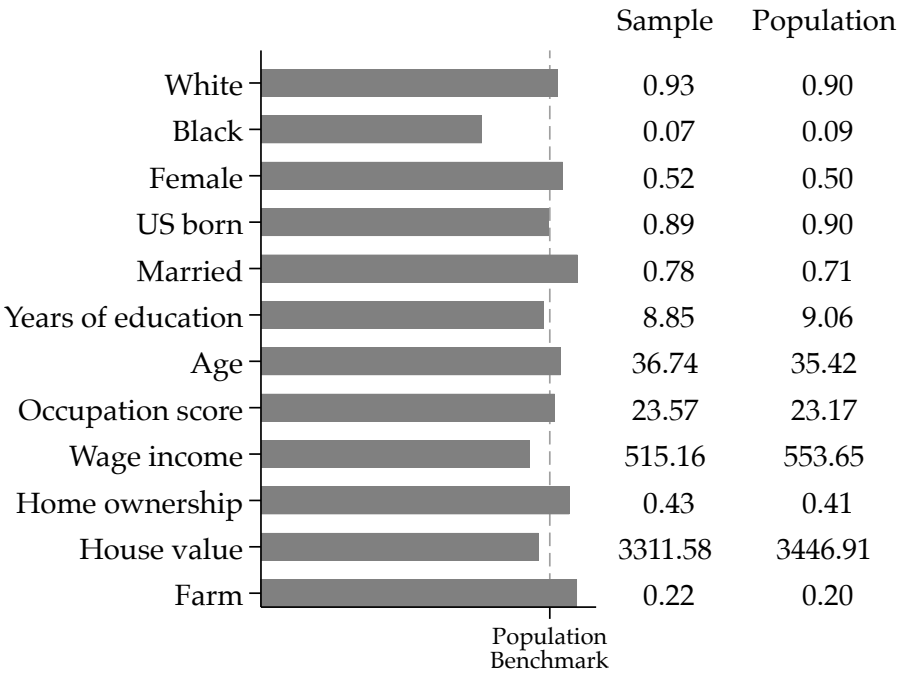
rates are slightly lower for women than for men. For the first half of the sample period, 1850 to 1900, coverage rates range from 1 to 5 percent; for the second half, coverage rates range from 7 to 20 percent.

The most innovative aspect of our panel is the large number of women whose records we trace from before and after marriage—12 million in total, covering women born between 1850 and 1940.[5] The size of this sample is unprecedented and uniquely equips us to study the intergenerational mobility of women in US history. Our sample coverage peaks for women born from the 1890s to the 1920s, where our sample contains between 1.5 and 3 million women per birth decade.

Our panel is representative along several dimensions (see Figure 3). Most notably, 52 percent of our sample are female. Our sample is also exceptionally representative in terms of race, only slightly over-covering white Americans. The income, wealth, and education of the people in our sample closely follows that of the full population. For example, 45 percent of individuals in our sample own their home, compared to 43 percent among the full population with an average value of $3,156 in 1940 compared to $3,216 among the full population. The sample is also highly representative in earlier periods.[6]

[5]See Appendix Figure A.2.
[6]See Appendix Figure A.3.

FIGURE 3: Sample Balance Prior to Weighting (1940)



Notes: This figure shows the representativeness of our panel by comparing the 1940 sample's average characteristics to those of the full population in the 1940 census. The sample is exceptional in representativeness compared to existsing panels, most notably with respect to sex and race. Because of the large sample sizes, even the smallest differences are statistically significant.

2.3 Linking Method

First, we link the Social Security applicants to each full-count census available. We adapt the method developed by Abramitzky et al. (2019) to leverage the additional information available in the administrative data to link individuals. To link women who may have changed their last names after marriage, we attempt to link a married woman's Social Security application to the census using both her married and maiden name (her father's last name). A link is established if at least a subset of information in the application matches to a single census record, with the smallest subset being the linking literature standard where we only use first name, last name, place of birth and age (Abramitzky et al., 2021). We allow for 5-year band around an individual's birth year to allow for imprecisely reporting of ages in the census. If potential links with both the maiden and married name exist in the same census decade, we do not establish either link.

Second, we link the applicant's parents to the census. If a child's application was linked to the census in the first step and the child's parents are present in their current household, we automatically establish an additional link for the parent. If the parent does not appear as an applicant in the data themselves, we assign them a synthetic SSN-like identifier. Now, the newly linked parent has a (synthetic) SSN as well as the state and year of birth and race (from the census). For mothers, we have their maiden name and their married name (father's last name).

Third, we link individuals—both applicants and their parents – that are now uniquely identified by their (for parents potentially synthetic) SSN across census decades. Linking across census records allows us to follow individuals and their socioeconomic status over time.

3. Intergenerational Mobility and Women in US History

First, we estimate intergenerational mobility for women, men, and all Americans for cohorts born between 1850 and 1900. For those estimates, we measure the socioeconomic status of a person or their parents at the household level. Our estimates suggest that while overall historical levels of intergenerational mobility were similar to today's, women tended to be more mobile than men. Second, we estimate intergenerational mobility disaggregating the socioeconomic status of parents. The results highlight mother's importance in shaping children's socioeconomic outcomes, overstating social mobility when measured in terms of father-child elasticities.

3.1 Main Estimates of Intergenerational Mobility

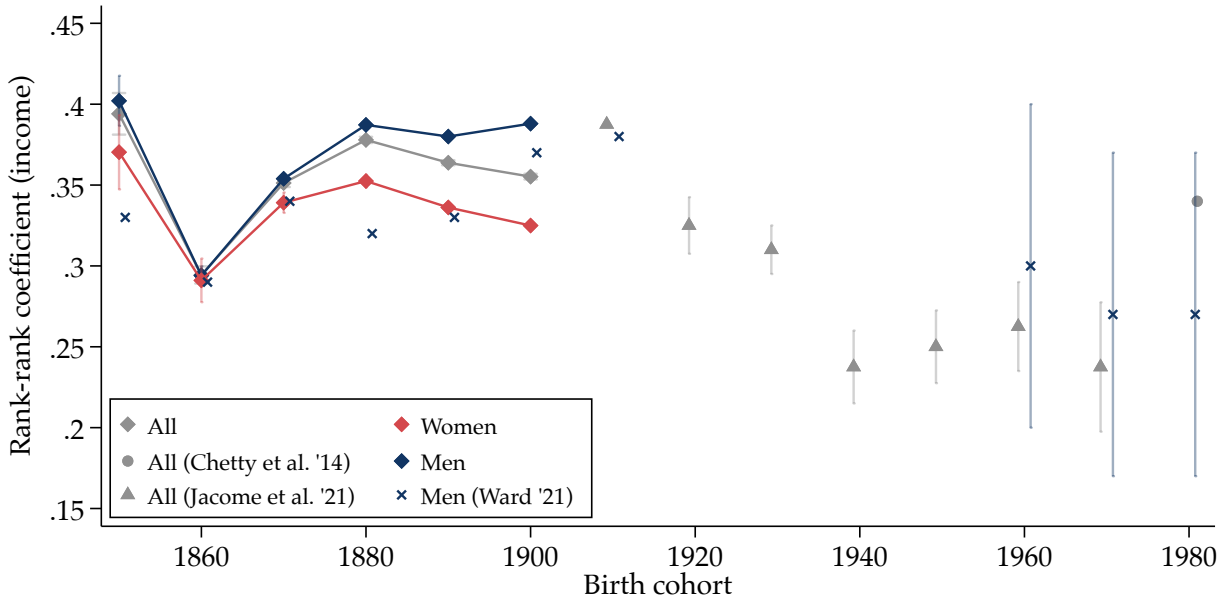
One of the common measures of intergenerational mobility is the rank-rank coefficient—the elasticity of the rank that any adult's family income occupies in the distribution with respect to the rank their parents' family income occupied within the distribution. A person's rank is determined by the percentile their family's income occupies in the overall family income distribution. If multiple observations are available for any individual's rank, we use their average as a proxy to reduce measurement error (Ward, 2021).

We estimate the following regression equation:

$$\text{rank}\left(Y_i^{\text{child}}\right) = \alpha_{c(i)} + \beta_{c(i)}\text{rank}\left(Y_i^{\text{parent}}\right) + \varepsilon_i \tag{1}$$

where Y_i is i's family income and $c(i)$ is i's cohort in decades. The main coefficient of interest is $\beta_{c(i)}$, capturing the rank-rank elasticity in family incomes. The lower this coefficient, the higher is cohort c's intergenerational mobility.

FIGURE 4: Intergenerational Immobility of Men and Women



Notes: This figure shows the rank-rank correlations between the household incomes of children and their parents. We compare our estimates for men, women, and all Americans with those in the literature of intergenerational mobility.

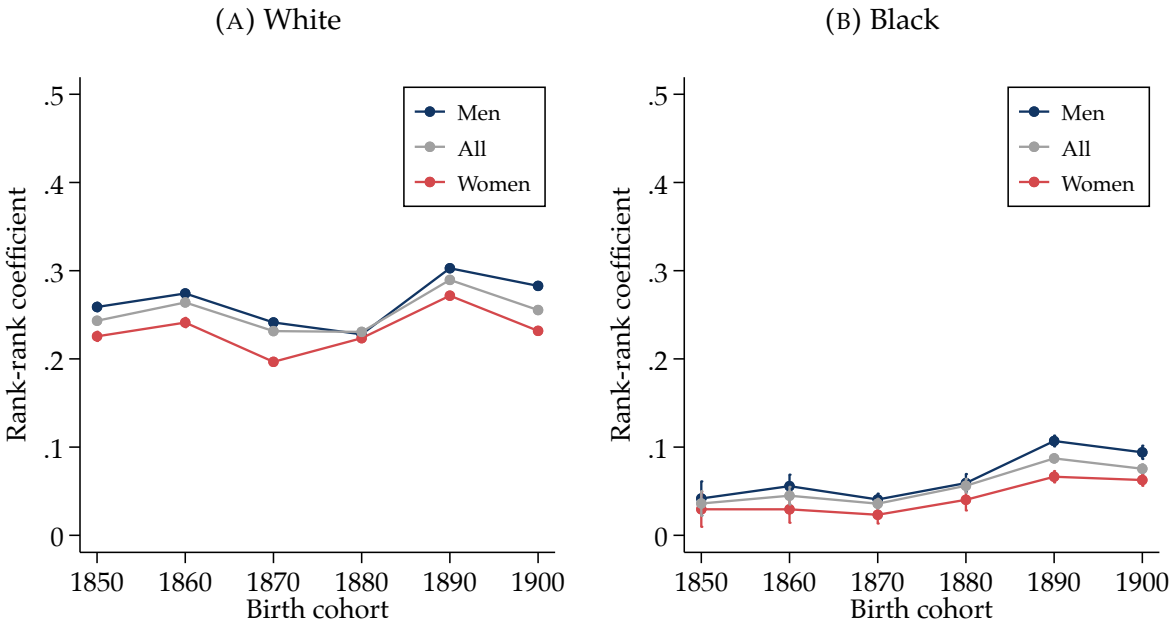
In Figure 4, we show that while women born before 1880 experienced levels of mobility similar to men, mobility between men and women began to differ significantly.[7] Women's mobility was significantly higher than that of men starting with cohorts born in the 1880s. For the cohort born between 1900 and 1909, we find rank-rank elasticities of 0.39 for men and 0.32 for women.Our estimates of the intergenerational mobility among men closely follow existing estimates (Ward, 2021). Overall, our findings suggest that society as a whole was somewhat, but not drastically, more mobile than estimates based on men suggest.

In Figure 5, we document that Black Americans were substantially more mobile than white Americans across all birth cohorts from 1850 to 1900. Because intergenerational mobility rose for white Americans and fell for Black Americans, over time their mobility rates have become more similar. Over time, the intergenerational mobility of Black and white Americans has been narrowing. Within race, women appear to be consistently more mobile then men. White men only achieved the same rate of intergenerational mobility for cohorts born around the Civil War (1861–1865), when mobility peaked for both genders. Similarly, Black women have been slightly more socially mobile than Black men.

We also document that intergenerational mobility increases over a person's lifetime

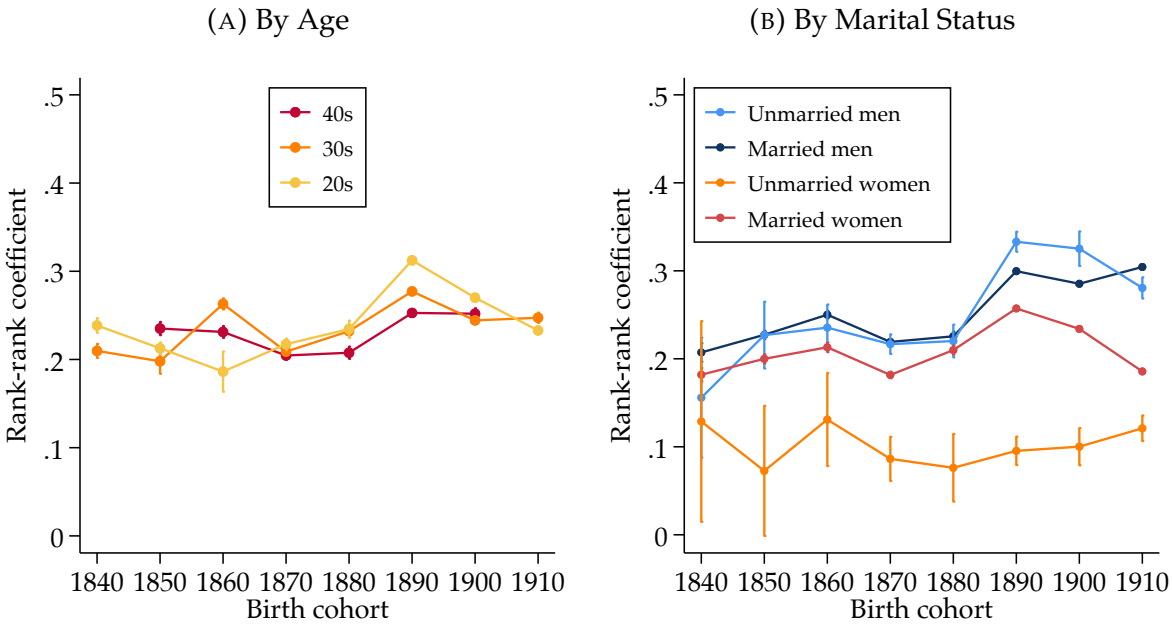
[7]Our weighted results in Figure B.4 show women to be even more mobile than their male counterparts relative to our main results. Our weighted version of the results in Figure 5, shown in Figure B.5 exhibit a similar pattern.

FIGURE 5: Intergenerational Mobility by Race



Notes: This figure shows the rank-rank coefficients of the household income from parents to children in each birth cohort. As household income, we use the sum of occupational scores of the household head and their potential spouse. We limit our sample to Americans aged 20 to 54. Panel (A) shows our estimates for white Americans; Panel (B) those for Black Americans.

FIGURE 6: Intergenerational Mobility by Age and Marital Status



Notes: This figure shows the rank-rank coefficients of the household income from parents to children in each birth cohort by age group (Panel A) and by marital status (Panel B). Only people who were *never* married between 1850 and 1940 are classified as unmarried. As household income, we use the sum of occupational scores of the household head and their potential spouse. We limit our sample to Americans aged 20 to 54.

(see Figure 6). Across birth cohorts, a persons' position in the income distribution is most similar to their parents' position in their twenties. As a person reaches their thirties and forties, their position in the income distribution becomes increasingly distinct from that of their parents'. In addition, individuals who never marry are less mobile than those who marry. The higher mobility of married individuals suggests that marriage may provide a means to achieving a socioeconomic status different from one's parents.

Given that most women during the 19th and 20th Century did not participate in the labor market, their family income was largely determined by their husband's income. Hence, whatever level of intergenerational mobility men attained, women attained additional mobility through marriage. Meanwhile, men entered occupations where they achieved a more similar income rank to that of their fathers. Hence, our results for the mobility of women are potentially reflective of marriage patterns. To this end, we explore the role assortative mating—the correlation between spouses' socioeconomic background—as a key driver of intergenerational mobility for women before WWII. In a society with perfect assortative mating, where the rich marry the rich and the poor marry the poor, there is likely little room for women to deviate from the socioeconomic status of their parents. However, we find evidence that women do deviate from the socioeconomic status of their fathers, requiring further exploration into the mechanism of assortative mating.

3.2 Additional Estimates: Mobility in Literacy

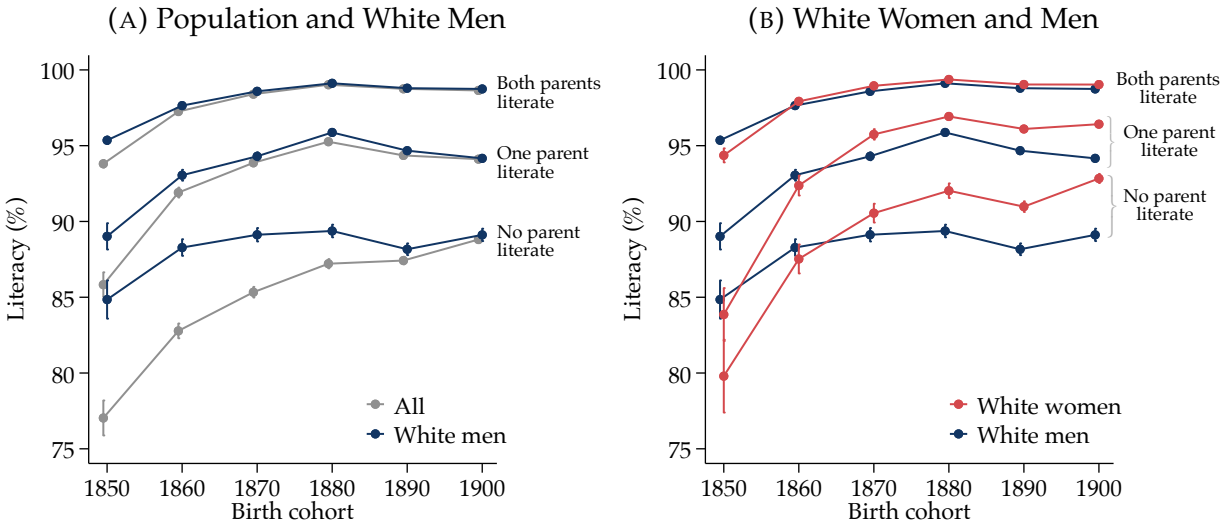
Literacy is a key outcome to study the intergenerational mobility of women for three reasons. First, is was a key indicator of human capital until the early 20th century. Second, it is one of the few outcomes that exist in all historical census records. Third, it is an outcome that contains valuable information for both men and women. In contrast, individual-level measures of income are ill-suited to study parent-child dependencies because before the second half of the 20th century, most women did not participate in the labor market. In this section, we therefore focus on literacy to measure intergenerational mobility at the individual (vs. family) level. This measure comparability between men and women allows us to estimate the separate impact that mother and fathers literacy have on the literacy of their child. We estimate the parental transmission of literacy using the following equation:

$$\text{Lit}_i = a_{c(i)} + \beta^{\text{m}}_{c(i)}\text{Lit}_i^{\text{mother}} + \beta^{\text{f}}_{c(i)}\text{Lit}_i^{\text{father}} + \beta^{\text{i}}_{c(i)}\text{Lit}_i^{\text{mother}} \times \text{Lit}_i^{\text{father}} + \epsilon_i$$

where Lit_i is person i's literacy status, resulting in an estimate of how much a person's literacy depends on the literacy of their mother ($\beta^{\text{m}}_{c(i)}$), father ($\beta^{\text{f}}_{c(i)}$), and the interaction of both ($\beta^{\text{i}}_{c(i)}$). We break up our analysis by race and sex.

Whether or not a person is literate depends greatly on their parents' literacy through-

FIGURE 7: Intergenerational Mobility in Literacy



Notes: This figure shows the average literacy rate of prime-age (20-54) Americans and prime-age white men by the literacy status of their parents. Estimates reflect the sum of $a_{c(i)}$ and $\beta^{m}_{c(i)}$, $\beta^{f}_{c(i)}$, or $\beta^{i}_{c(i)}$ respectively. Lower dispersion corresponds to higher intergenerational mobility. Shaded areas are robust confidence bands.

out our sample period (birth cohorts of 1850s to 1900s). Figure 7 shows our estimates separately for white men and the entire population. Both series suggest that whether a person is literate or not has become less dependent on their parents' literacy. This finding is consistent with rapidly rising literacy rates across the population during this era. However, focusing on white men understates the degree to which intergenerational mobility has increased.

For people born in the 1850s who had two literate parents, literacy rates were 17 percentage points higher than those of individuals with no literate parent. For the 1900s cohort, this gap had shrunk to 9 percentage points. For white men, the gap shrank by far less, from 15 to 13 percentage points. This result is consistent with the expansion of literacy to groups that have historically attained lower education than white men.

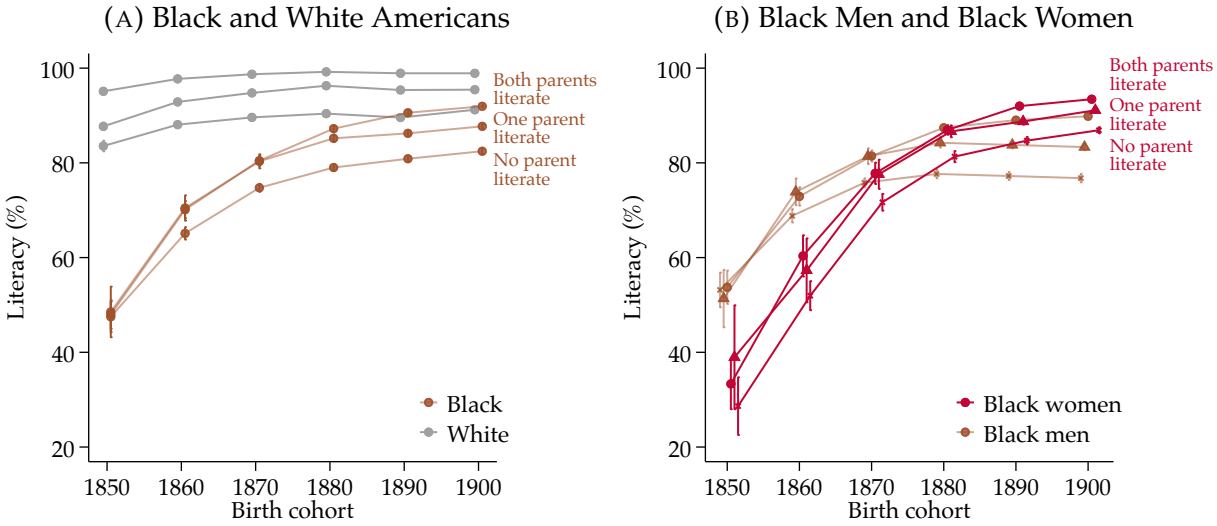
In contrast to white men, white women experienced a much faster increase in mobility (see Panel B of Figure 7). While white women born in the 1850s with one or no literate parents were 5 percentage points less likely to be literate than white men within the same parent group, women born in the following decades caught up and surpassed their male counterparts. White women born in the 1900s with one (no) literate parent were 2 (6) percentage points more likely to be literate than their male counterparts.

Mobility patterns in literacy were vastly different for Black Americans (see Figure 8). Black men with one or no literate parent had almost the same literacy rate as Black men with both parents literate up until the 1880s birth cohort. after which there was a slight divergence. Black men with no literate parents faced a larger gap.

Black women experienced a faster increase in mobility relative to Black men. In con-

trast to Black men born in the 1850s, Black women with one literate parent were more mobile than women with two or no literate parents. However, this switched for Black women born in the 1860s. In a similar pattern to Black men, the literacy rates of Black women with one and two literate parents converged up until the 1880s birth cohort, after which they diverged. From the 1880s birth cohort onward, Black women became more likely to be literate than their male counterparts.

FIGURE 8: Intergenerational Mobility in Literacy by Race



Notes: This figure shows the average literacy rate of prime-age (20-54) Black Americans by the literacy status of their parents. Panel (A) contrasts Black Americans with white Americans. Panel (B) contrasts Black men with Black women. Estimates reflect the sum of $a_{c(i)}$ and $\beta^{m}_{c(i)}$, $\beta^{f}_{c(i)}$, or $\beta^{i}_{c(i)}$ respectively. Lower dispersion corresponds to higher intergenerational mobility. Shaded areas are robust confidence bands.

Compared to white Americans, Black Americans experienced significant increases in mobility in terms of literacy rates through the 1850s to 1900s birth cohorts. Figure 8 shows our estimates for the entire Black and white population respectively. Black individuals born in 1850 with two literate parents were 45 percentage points less likely to be literate than their white counterparts. By the 1900s birth cohort, Black individuals with two literate parents were only 5 percentage points less likely to be literate than white men with two literate parents. While the gap between white and Black individuals within the same parent group closed substantially over time, Black individuals born in the 1900s were still less mobile.

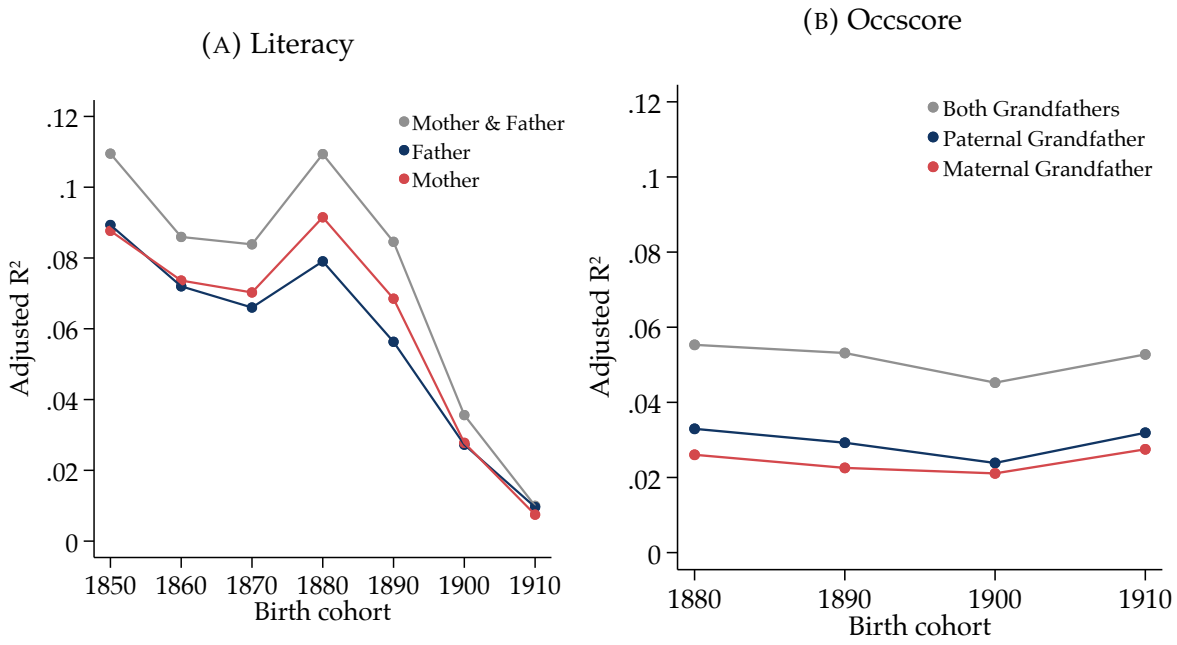
3.3 Mothers' Contribution to Intergenerational Mobility

We estimate the extent to which maternal and paternal grandfathers explain their grandchild's income rank, as shown in Figure 9. We compare the adjusted R-squared in our rank-rank elasticities when including paternal grandfather, maternal grandfather or both grandfathers. We find that including both grandfathers explains a child's income rank

better than if we only included one line. Importantly, when looking at paternal and maternal contribution respectively, we find that the maternal line explains as much as the paternal line. Figure 10 shows the elasticity of a child's income rank with respect to the income ranks of maternal and paternal grandfathers as well as their average. The gradient of the relationship between the average rank of both grandparents is steeper than it is using only one parent. Taken together, these results indicate the importance of accounting for pooled family resources in understanding intergenerational mobility.

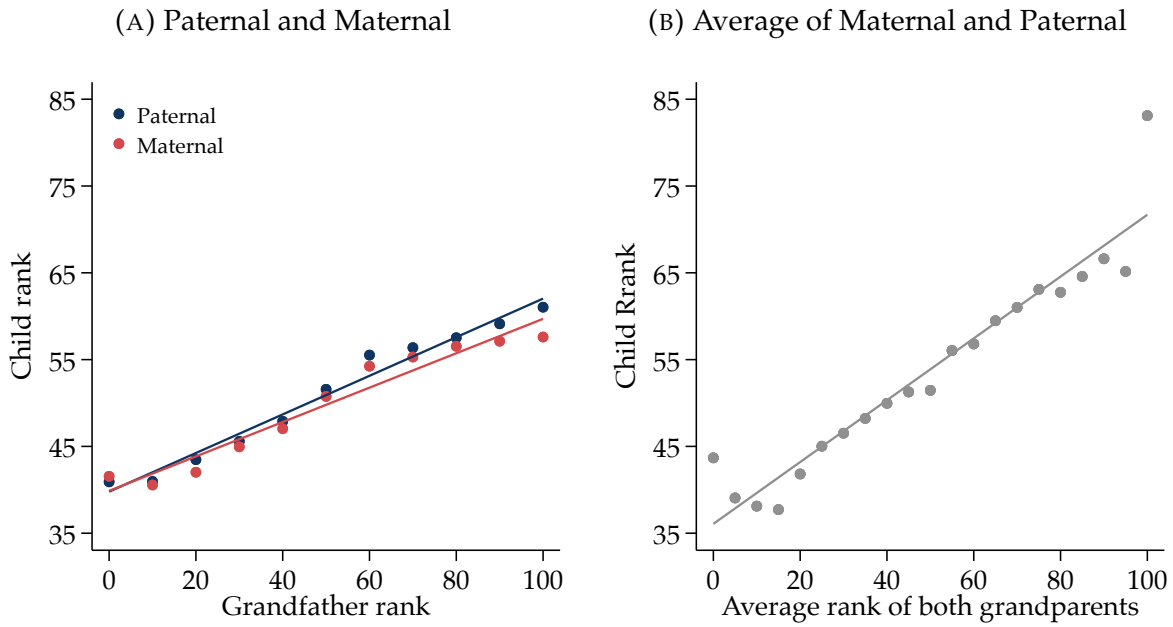
We examine the importance of both mothers and fathers in determining children's literacy outcomes. Once again, we estimate the adjusted R-squared when regression mother and father's literacy on that of their child's adult literacy. We find similar results to our income estimates. Both parents explain a child's literacy outcomes more than one parent, and mother's explain just as much as fathers, and in some cases contribute to their child's literacy to a larger degree.

FIGURE 9: Contribution of Mothers to Intergenerational Mobility



Notes: Panel (A) of this figure shows the intergenerational mobility in literacy based on the literacy of a person's mother, father, or both parents. Panel (B) shows the intergenerational mobility in occupational scores. The child's outcome is their household-level occupational income scores. Because mothers tended not to participate in the labor market, we use the maternal and paternal grandparents' occupational income scores as parent outcomes.

FIGURE 10: Paternal and Maternal Contributions to Intergenerational Mobility



Notes: Panel (A) of this figure shows the average occupational income rank (at the household level) of a child depending on their maternal and paternal grandfathers' occupational income rank. Panel (B) shows the same but averages over the ranks of both grandparents. The steeper slope suggests that a person's maternal and paternal socioeconomic background explain their own outcomes far better than only considering one side of their ancestral line.

3.4 Intergenerational Mobility & Assortative Mating

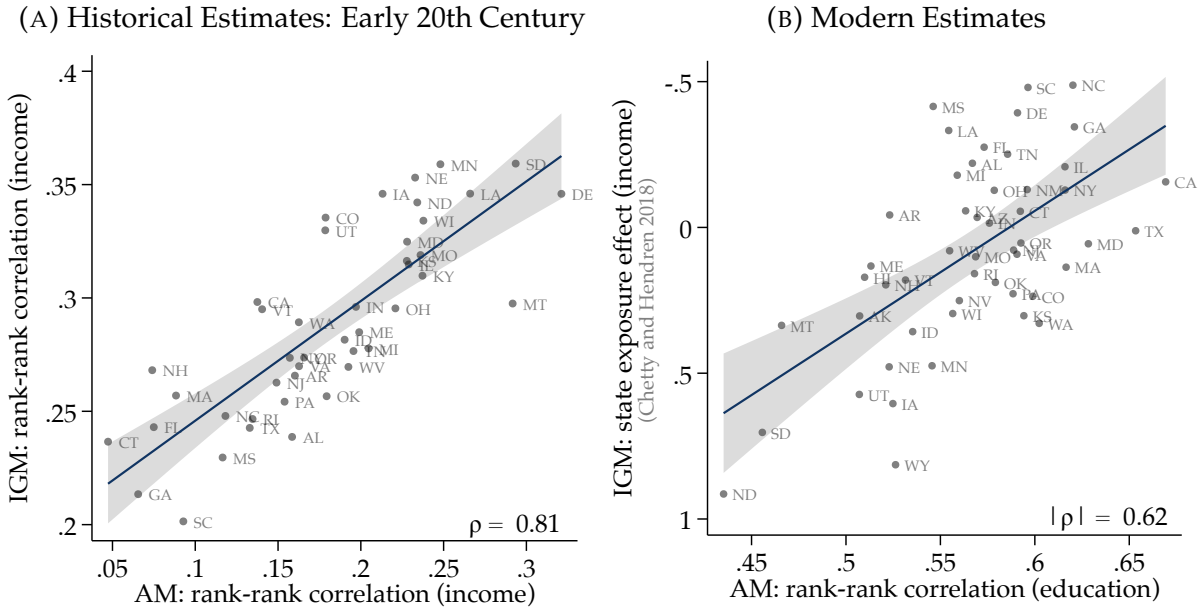
We measure assortative mating as the rank-rank elasticity between a person's father and their father-in-law:

$$\text{rank}\left(Y_i^{\text{father}}\right) = \alpha_{c(i)} + \beta_{c(i)}\text{rank}\left(Y_i^{\text{father-in-law}}\right) + \varepsilon_i. \tag{2}$$

The main coefficient of interest is $\beta_{c(i)}$, capturing the rank-rank elasticity in the incomes of fathers and fathers-in-law. The higher this coefficient, the higher is cohort c's degree of assortative mating.

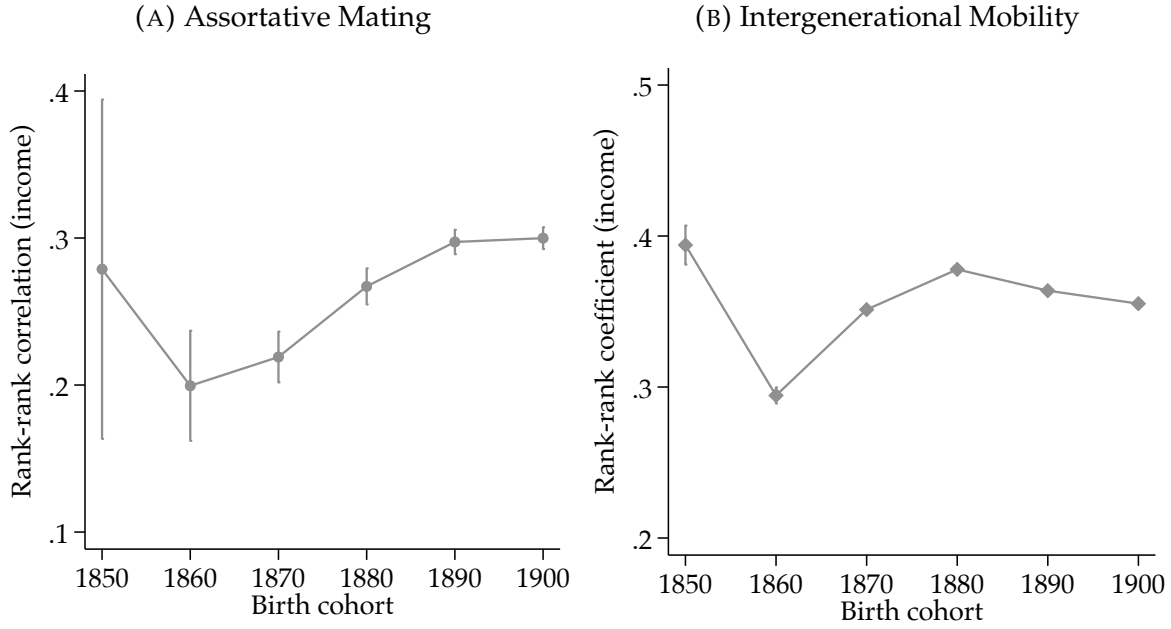
We document a strong empirical link between low intergenerational mobility and high levels of assortative mating—both across time and space. First, for birth cohorts where assortative mating was stronger, intergenerational mobility was lower. Second, for a given cohort, individuals born in state with strong assortative mating experience lower intergenerational mobility. Indeed, over 80 percent of the state variation in mobility can be accounted for by state differences in assortative mating.

FIGURE 11: Low Intergenerational Mobility in States with High Assortative Mating



Notes: This figure relates a cohorts' intergenerational mobility (IGM) to the assortativeness of its marriages. Panel (A) shows estimates for the 1900s birth cohort from our new panel; IGM is the parent-child rank-rank correlation in household income (occscore); AM is the father-father-in-law rank-rank correlation in household income (occscore). Panel (B) shows estimates for the 1980s birth cohort; IGM is measured as the causal effect of exposure to places for children with parents at the 25th percentile of the household income distribution, aggregating county-level estimates from Chetty and Hendren (2018) to the state-level; AM is measured as the rank-rank correlation in years of education following Eika et al. (2019).

FIGURE 12: Assortative Mating and Intergenerational Mobility



Notes: Panel (A) of this figure shows the degree of assortative mating over time as the rank-rank correlation between the income of a person's parents and their parents-in-law. Panel (B) shows the degree of intergenerational mobility over time as the rank-rank correlation between the income of a person's parents and their own household income in adulthood.

4. Conclusion

This paper provides representative long-run estimates of intergenerational mobility for Americans born between the 1850s and the 1910s. Our results show that intergenerational mobility has historically been rather similar to mobility in more recent periods. This finding contradicts historical anecdotes that suggest that the 19th century was one of exceptionally high mobility. For white Americans, mobility peaked for people born around the turn of the 20th century (1880-1900s) and those born in the decades after WWII (1940s-1970s). For Black Americans, this peak was achieved by those who were born into and subsequently freed from slavery. Across the decades we study, Black Americans were only somewhat more mobile than white Americans.

We find that both white and Black women were more mobile than their male counterparts for birth cohorts after 1870. The differences between the intergenerational mobility estimates of men and women, measured in terms of family income, suggests that women chose spouses or careers that resulted in their income rank being less similar than that of their fathers, relative to men. We explore the channel of assortative mating to determine the degree to which marriage patterns relate to intergnerational mobility. We find that lower assortative mating is correlated with high mobility. Lastly, we document new evidence on the maternal contribution to intergenerational mobility. We find that both parents play a role in the mobility of their children, and that mothers play as much of a role as fathers.

In addition to further assessing the channel of assortative mating in intergenerational mobility, we plan to link individuals from the 1930 and 1940 censuses to administrative death records that allow us to extend our results to 2000 at the neighborhood-level. Neighborhood-level outcomes include proxies for education, income, and wealth which we use as additional outcomes to measure intergenerational mobility.

REFERENCES

ABRAMITZKY, R., L. BOUSTAN, E. JACOME, AND S. PEREZ (2021): "Intergenerational Mobility of Immigrants in the United States over Two Centuries," *American Economic Review*, 111, 580–608.

ABRAMITZKY, R., L. BOUSTAN, AND M. RASHID (2020): "Census Linking Project: Version 1.0," dataset: `https://censuslinkingproject.org`.

ABRAMITZKY, R., L. P. BOUSTAN, K. ERIKSSON, J. J. FEIGENBAUM, AND S. PÉREZ (2019): "Automated Linking of Historical Data," Working Paper 25825, National Bureau of Economic Research.

BAILEY, M. J., P. Z. LIN, S. MOHAMMED, P. MOHNEN, J. MURRAY, M. ZHANG, AND A. PRETTYMAN (2022): "LIFE-M: The Longitudinal, Intergenerational Family Electronic Micro-Database," dataset: `https://doi.org/10.3886/E155186V2`.

BLACK, S. E., P. J. DEVEREUX, AND K. G. SALVANES (2005): "Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital," *American Economic Review*, 95, 437–449.

CHETTY, R. AND N. HENDREN (2018): "The impacts of neighborhoods on intergenerational mobility II: County-level estimates," *The Quarterly Journal of Economics*, 133, 1163–1228.

CHETTY, R., N. HENDREN, P. KLINE, E. SAEZ, AND N. TURNER (2014): "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility," *American Economic Review Papers and Proceedings*, 104, 141–147.

CLARK, G. AND N. CUMMINS (2022): "Assortative mating and the Industrial Revolution: England, 1754-2021," Economic History Working Papers 114608, London School of Economics and Political Science, Department of Economic History.

CRAIG, J., K. A. ERIKSSON, AND G. T. NIEMESH (2019): "Marriage and the Intergenerational Mobility of Women: Evidence from Marriage Certificates 1850-1910," .

EIKA, L., M. MOGSTAD, AND B. ZAFAR (2019): "Educational Assortative Mating and Household Income Inequality," *Journal of Political Economy*, 127, 2795–2835.

FETTER, D. K., L. M. LOCKWOOD, AND P. MOHNEN (2021): "Long-Run Intergenerational Effects of Social Security," Tech. rep.

GOLDIN, C. (1990): *Understanding the gender gap: An economic history of American women*, gold90-1, National Bureau of Economic Research.

——— (1991): "The Role of World War II in the Rise of Women's Employment," *American Economic Review*, September, 741–756.

GOLDIN, C. AND C. OLIVETTI (2013): "Shocking Labor Supply: A Reassessment of the Role of World War II on Women's Labor Supply," *American Economic Review Papers and Proceedings*, 103, 257–262.

JÁCOME, E., I. KUZIEMKO, AND S. NAIDU (2021): "Mobility for All: Representative Intergenerational Mobility Estimates over the 20th Century," Working Paper 29289, National Bureau of Economic Research.

OLIVETTI, C. AND M. D. PASERMAN (2015): *American Economic Review*, 105, 2695–2724.

WARD, Z. (2021): "Intergenerational Mobility in American History: Accounting for Race and Measurement Error," Working Paper 29256, National Bureau of Economic Research.

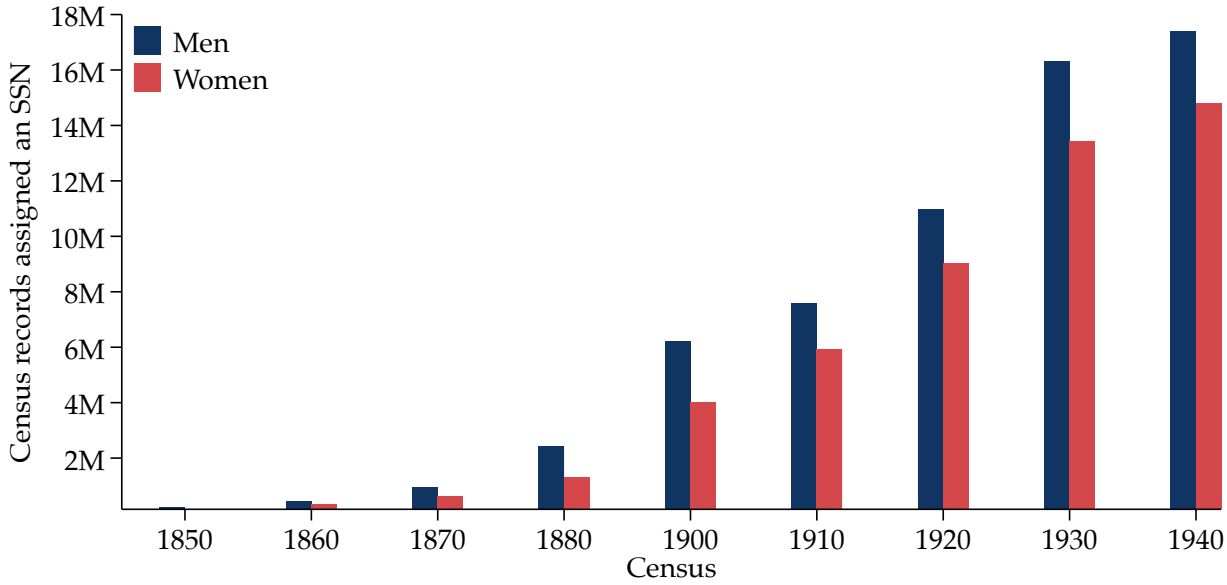
APPENDIX

A Data Appendix 21

B Weighted results 23

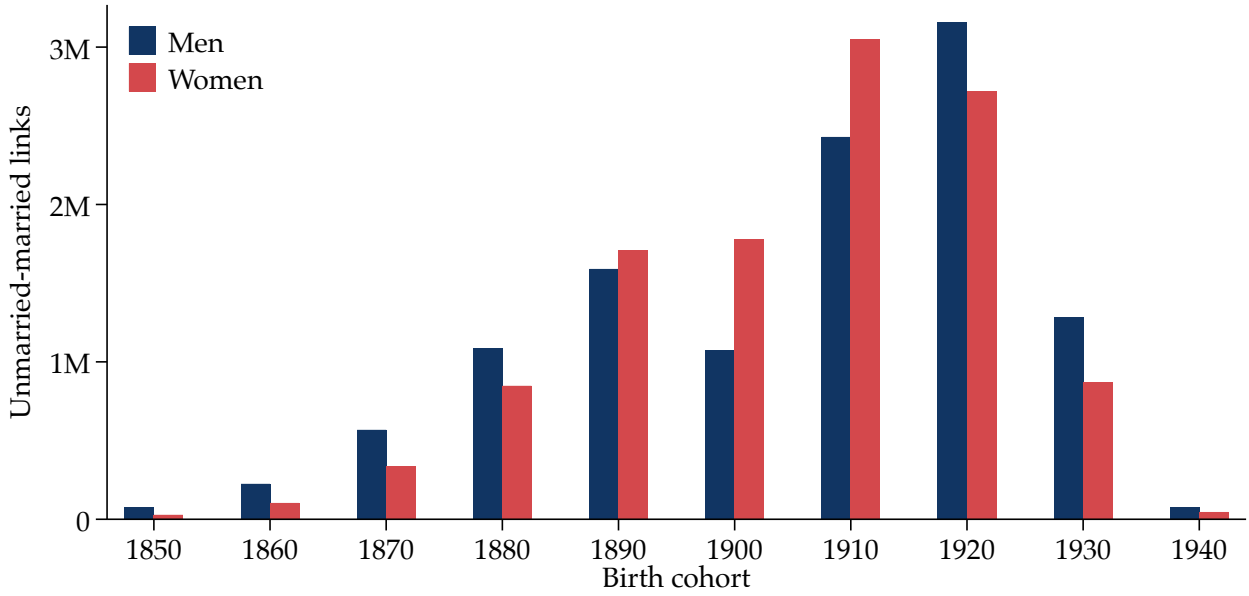
A. Data Appendix

FIGURE A.1: Number of Census-Social Security Links



Notes: This figure shows the number of census records that we have assigned a Social Security Number (SSN) by sex and census year. In total, we link 112,096,305 census records to an SSN—62,496,789 for men and 49,599,516 for women.

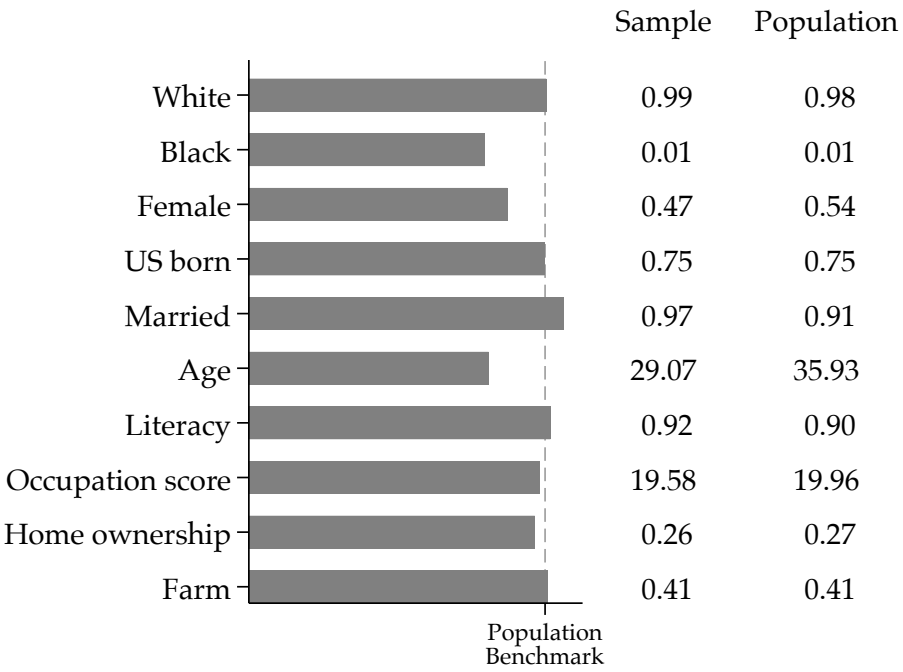
FIGURE A.2: Number of Unmarried to Married Links



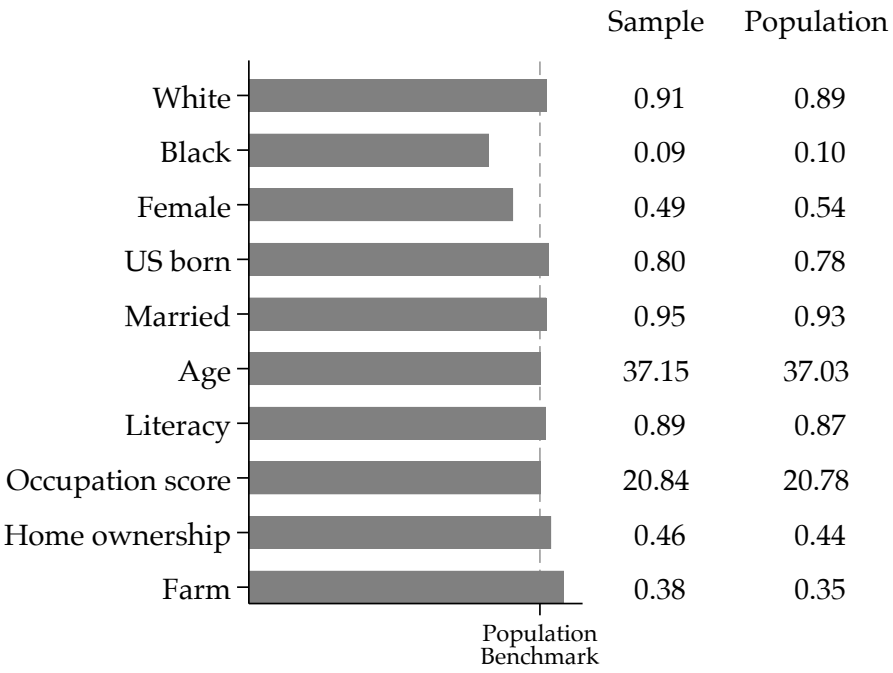
Notes: This figure shows the number of men and women we link from their childhood home's census record to their census record after marriage. The totals—11,588,403 men and 11,490,587 women—are split by birth cohort (where 1900, for instance, represents individuals born between 1900 and 1909 inclusively).

FIGURE A.3: Sample Balance

(A) 1860



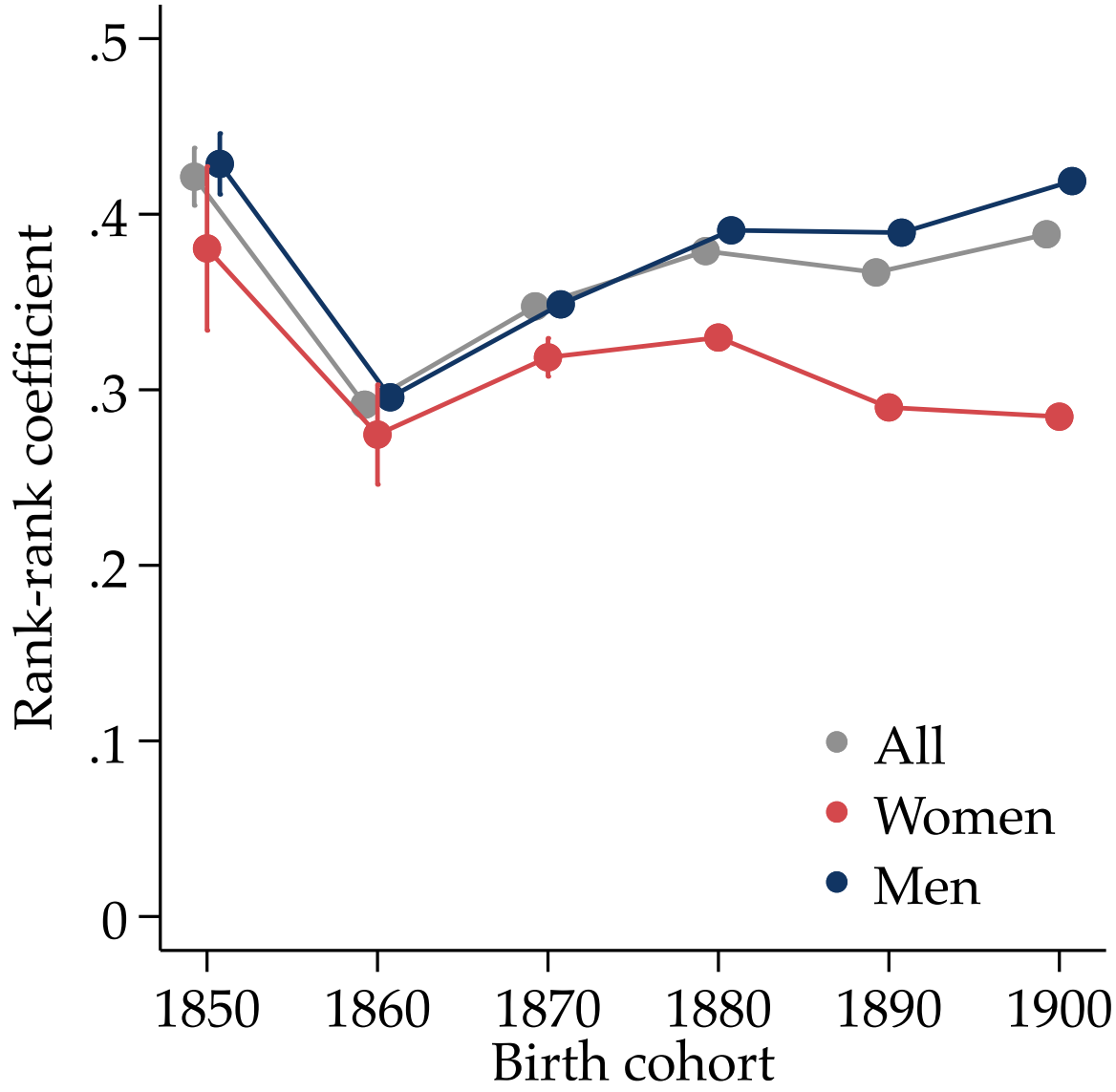
(B) 1900



Notes: This figure shows the representativeness of our panel by comparing the 1860 and 1900 sample's average characteristics to those of the full population in the 1860 and 1900 census respectively. The sample is exceptional in representativeness compared to existsing panels, most notably with respect to sex and race. Because of the large sample sizes, even the smallest differences are statistically significant.

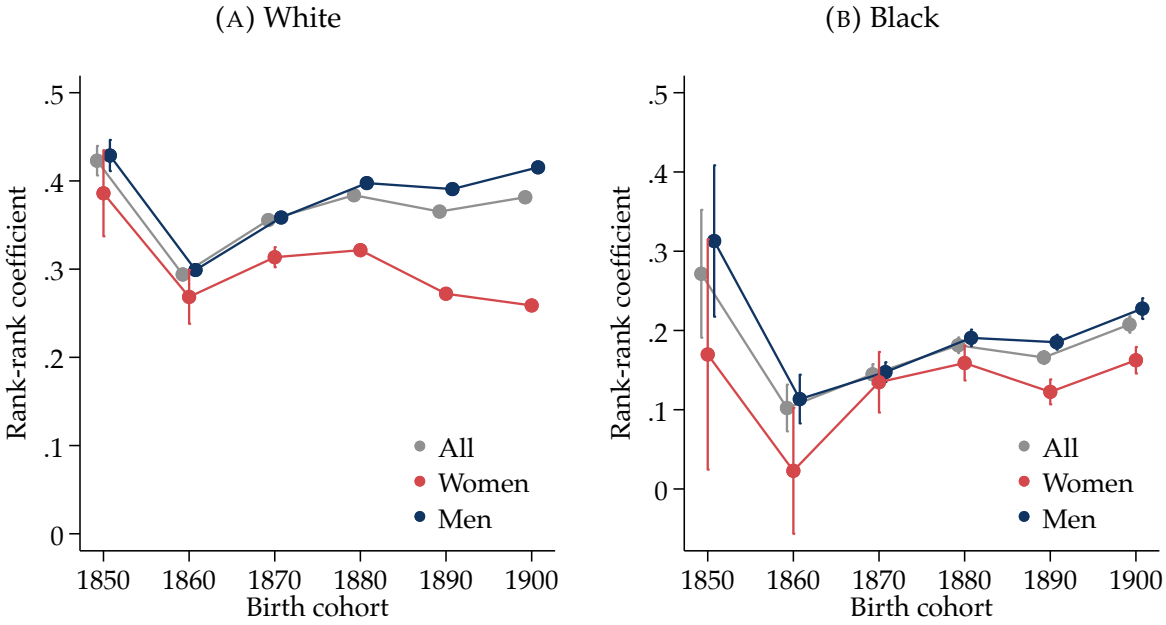
B. WEIGHTED RESULTS

FIGURE B.4: Intergenerational Mobility, weighted



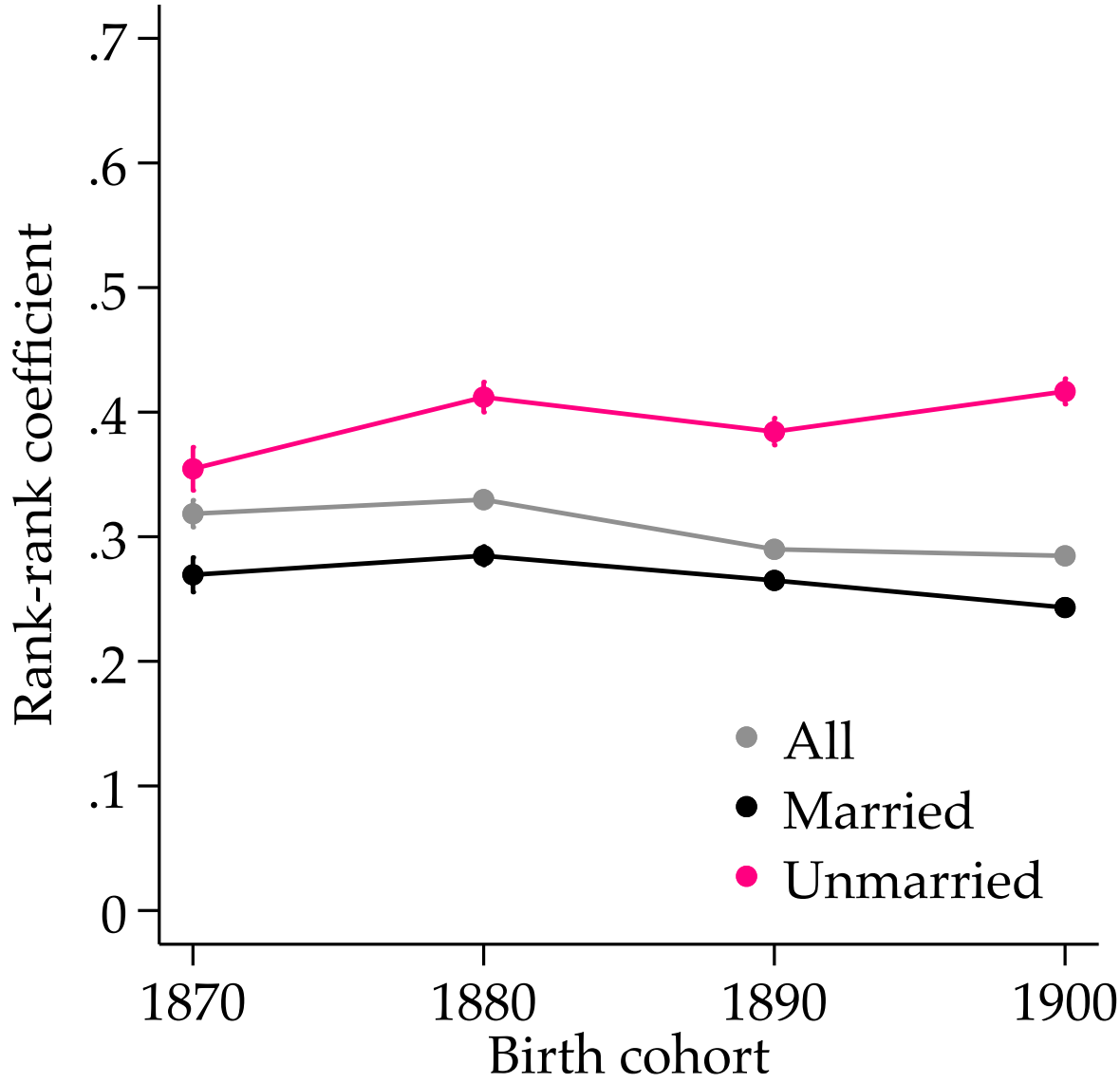
Notes: This figure shows the rank-rank correlations between the household incomes of children and their parents. These results are the weighted equivalent of those presented in Figure 4. Results are weighted to match the adult child population's characteristics on race, age, occupation category (white collar, farmer, semi-skilled, unskilled), region of residence, urban location, farming status, ownership of residence and marriage. These inverse propensity weights are created separately for each child's adult year of observation.

FIGURE B.5: Intergenerational Mobility by Race, weighted



Notes: This figure shows the rank-rank coefficients of the household income from parents to children in each birth cohort. As household income, we use the sum of occupational scores of the household head and their potential spouse. We limit our sample to Americans aged 20 to 54. Panel (A) shows our estimates for white Americans; Panel (B) those for Black Americans. These results are the weighted equivalent of those presented in Figure 4. Results are weighted to match the adult child population's characteristics on race, age, occupation category (white collar, farmer, semi-skilled, unskilled), region of residence, urban location, farming status, ownership of residence and marriage. These inverse propensity weights are created separately for each child's adult year of observation.

FIGURE B.6: Intergenerational Mobility by Marital Status, weighted



Notes: This figure shows the rank-rank coefficients of the household income from parents to children in each birth by marital status. Only people who were *never* married between 1850 and 1940 are classified as unmarried. As household income, we use the sum of occupational scores of the household head and their potential spouse. We limit our sample to Americans aged 20 to 54. These results are the weighted equivalent of those presented in Figure 6 (Panel B). Results are weighted to match the adult child population's characteristics on race, age, occupation category (white collar, farmer, semi-skilled, unskilled), region of residence, urban location, farming status, ownership of residence and marriage. These inverse propensity weights are created separately for each child's adult year of observation.