Intergenerational Mobility and Assortative Mating

Lukas Althoff* Harriet Brookes Gray† Hugo Reichardt‡

November 7, 2022

1. Introduction

Throughout modern history, the US has been considered the land of opportunity. The automated linking of historical Census records has overcome many of the difficulties in empirically evaluating this claim, and a large literature documents the intergenerational mobility of Americans over the 19th and 20th century using linked census data (Abramitzky et al., 2019; Ward, 2021). However, these long-run estimates of intergenerational mobility exist almost exclusively for men, and only consider the relationship between the socioeconomic status of fathers and their sons. Incorporating women into the analysis has been notoriously hard because name changes after marriage make it challenging to trace their records from childhood to adulthood. Important progress has been made by linking historical records using information of name changes from marriage certificates that exist in some states (Craig et al., 2019), or by estimating mobility directly from survey data that asks women about their socioeconomic status and that of their parents (Jácome et al., 2021). However, availability, selection into the data, and sample sizes remain a challenge.

This paper overcomes the empirical challenge of linking women's historical census records to study the intergenerational mobility of *all* Americans from 1850 to 1940, and one of its main drivers - assortative mating. We combine two sources of data, historical census records and information from 41 million Social Security Number (SSN) applications, to trace millions of men and women over time. These 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, it also includes the

*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

maiden names of applicants' parents, massively expanding the sample, extending the coverage back in time, and increasing representativeness as it allows us to link many people who never applied for a SSN.[1] Based on this information, we follow women's historical census records from before to after marriage.

Our new data ranges from 1850 to 1940 and consists of 48 million total links, half of which are women. Our data is highly representative along all dimensions, including income, race, and sex. We link a total of 12 million women from before to after marriage, uniquely equipping us to study the long-run evolution of intergenerational mobility and assortative mating. We aim to make all of our data publicly available.

We first document the evolution of intergenerational mobility for all Americans—across men and women—from 1850 to 1940. Using proxies for household income for both parents and children, we find rank-rank elasticities of 0.3 to 0.4. Our results suggest that intergenerational mobility between 1850 and 1940 was no different from what it is today (Chetty et al., 2014). Existing estimates that include both men and women range between 0.25 and 0.4 (Jácome et al., 2021) and 0.35 (Chetty et al., 2014).

We next explore a key potential driver of intergenerational mobility, especially for women: assortative mating. We document that the levels of intergenerational mobility and assortative mating are highly correlated over time and across space. In states where marriages are more assortative—i.e., wives come from a very similar socioeconomic background as their husbands—intergenerational mobility is low. Across space, 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. The causal nexus between changes in the marriage market and changes in intergenerational mobility should be explored in further research.

We also 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 captures intergenerational mobility. When considering both parents as well as their potential complementarity, multiple elasticities become relevant, making it harder to compare intergenerational mobility across time and contexts. As a solution, we propose using the variance in children's outcomes explained by both parents' socioeconomic status (adjusted R-squared) as an alternative measure of intergenerational mobility.

We find that mothers are as predictive as fathers of their children's outcomes in

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

both income and education.

This paper contributes to a rapidly growing effort to link historical Census records. Abramitzky et al. (2019, 2020) are among the pioneers of automated record linking based on names. We build on their methodology and augment the procedure to leverage additional data sources. In particular, we combine Census records with the rich information of historical administrative data from Social Security applications. Conditioning not only a person's own but also their parents' characteristics allows us to identify more individuals with high confidence. As a result, both coverage and quality of panels is improved. Bailey et al. (2022) similarly used vital records from two states to augment the Census linking procedure. Our panel covers the entire US.

This paper also contributes to the ongoing effort of including women in long-run economic analyses. Specifically, we are part of a nascent literature that combines different sources of data to link historical Census records for women. Olivetti and Paserman (2015) constructed synthetic panels based on the first names of men and women. Our panel consists of actual high-quality links. Craig et al. (2019) used marriage certificates from Massachusetts (1850-1910) to link around 60 thousand married women in the Censuses of 1880 or 1910 to their childhood household in the Censuses of 1850 or 1880. Similarly, Bailey et al. (2022) combined a larger 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 household in the 1880-1920 Censuses. Our panel does not impose similar restrictions on the marital status or geographic location of women. Fetter et al. (2021) linked social security applicants to their fathers in the 1930 Census. Our panel leverages similar data but ranges over all Census decades available and leverages both applicants and their parents.

This paper contributes to the empirical analysis of intergenerational mobility throughout American history. For recent decades, intergenerational mobility has been well documented for both men and women using administrative data that contains 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 somewhat lower intergenerational mobility for American men born 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 patterns 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 important progress in documenting intergenerational mobility at the individual-level for both

men and women historically. They found that mobility increased from the 1910s to the 1940s birth cohorts, before plateauing compared to the 1970s cohorts. They show that Black women were a critical driver of increasing mobility over this period. Craig et al. (2019) found that mobility for women was higher than for men in the late 1800s, but this gap disappeared by 1900.

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. This hurdle has recently been addressed by developing methods to linking the historical Census records of men based on their names, age, and place of birth (Abramitzky et al., 2019; Ruggles et al., 2020). However, women are typically not covered by those methods because most of them changed their name upon marriage, making them hard to follow across historical records. In this paper, we overcome this hurdle by combining Census data with administrative records that contain the maiden names of millions of women.

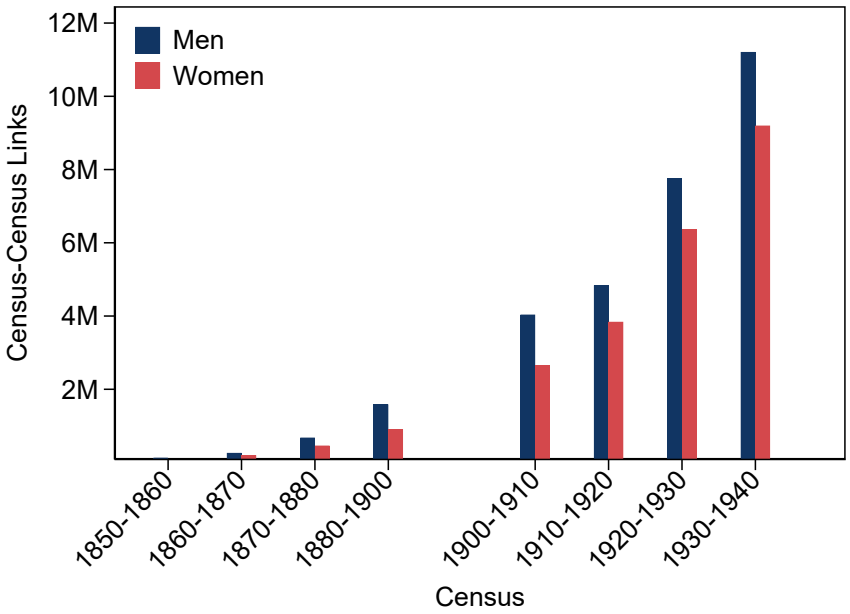
2.1 Our Panel

We successfully link 16 million Social Security applications to the Census. In addition, we are able to link 20 million parents to their Census records. Figure **??** shows our coverage of each Census decade. For earlier Census years, 1860 to 1880, we have a coverage rate of up to 5% of the Census. For 1900 and 1910, our coverage rate is between 10% and 15%, and for 1920, 1930 and 1940, we cover between 20% to 25% of the Census. Figure 1 shows the number of Census-to-Census links as a share of the Census-pair target year (i.e., the share of 1850-1860 Census links as a share of the 1860 Census).

Our panel is representative along several dimensions (see Figure 2). Most notably, 51 percent of our sample are female, almost perfectly capturing the total population in 1940. 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 (vs. 43 percent among the full population) with an average value of $3,156 in 1940-US dollars (vs. 3,216 among the full population).

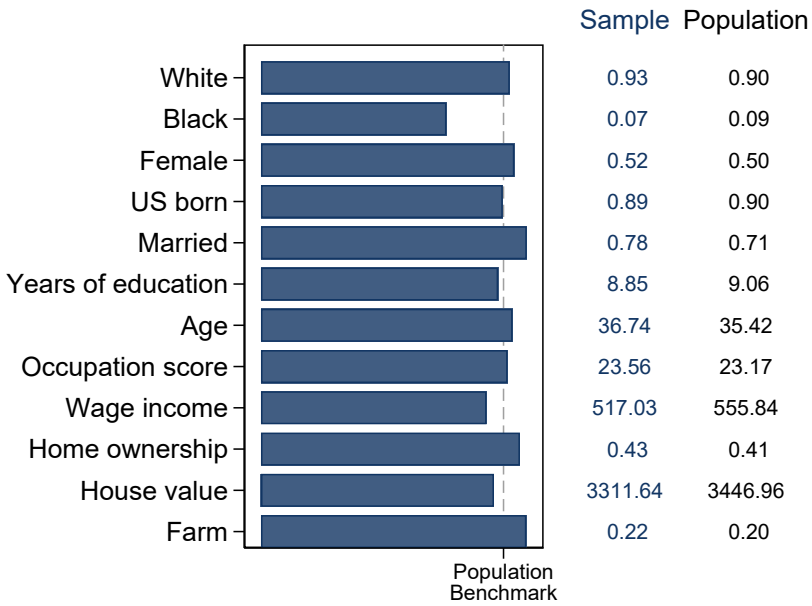
As an exceptional feature of our data, we follow up to 22 million Americans from childhood to marriage, half of whom are women (see Figure 3). This feature enables us to yield representative estimates of intergenerational mobility and study assortative mating as one of its main drivers.

FIGURE 1: Coverage of Our New Panel



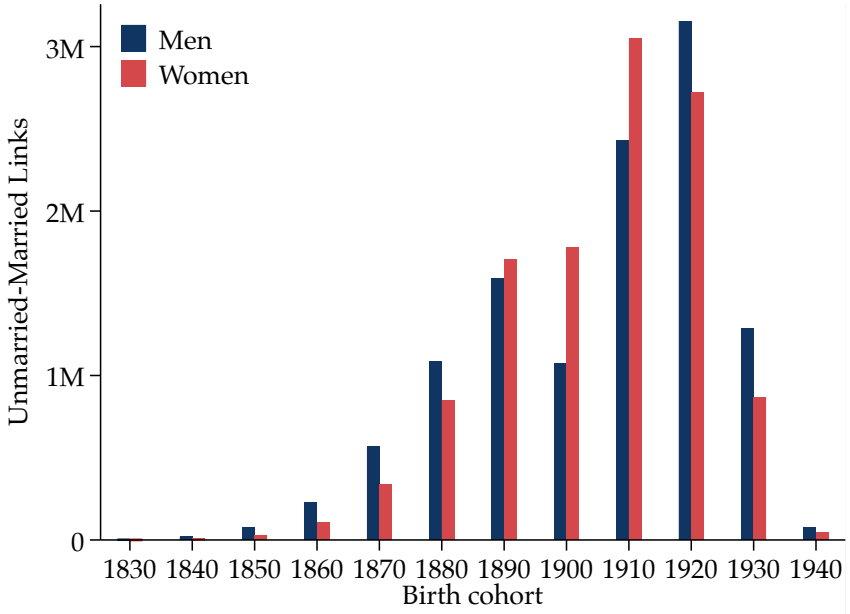
Notes: This figure shows the fraction of the full population of men and women that we successfully linked from one Census decade to the next. Our empirical analysis will also leverage links across non-adjacent Census pairs to maximize coverage.

FIGURE 2: Representativeness of Our New Panel



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.

FIGURE 3: 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 are split by birth cohort, where each year indicates a full decade (e.g., 1900 represents the birth cohorts of 1900-1909).

2.2 Data on the Maiden Names of Women

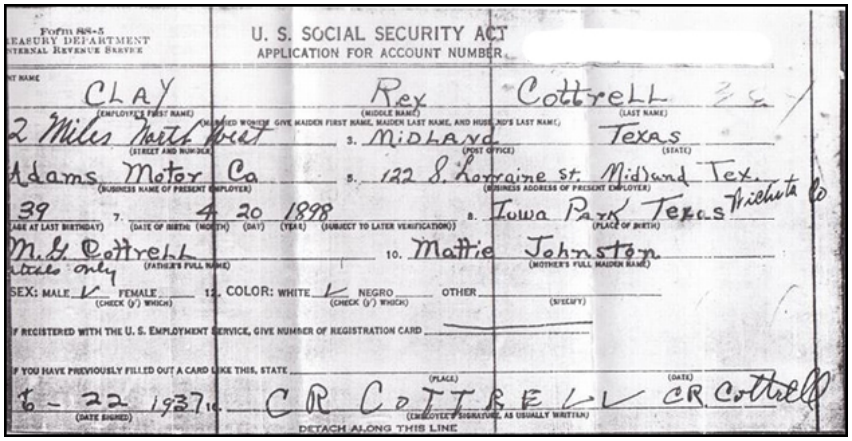
We link the decennial full-count Censuses from 1850 to 1940 to 41 million Social Security applications.[2] These applications contain the near universe of Americans who both applied for a Social Security Number (SSN) and died between 1988 and 2007. The data includes a rich set of information including each applicant's name, age, race, place of birth, and the names of their parents. Most importantly, the data includes the individual's maiden name and the maiden name of their mother. By linking those applications to the Census, we assign each person in the Census with a unique identifier (SSN), allowing us to follow them through historical records despite potential name changes upon marriage.

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

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

FIGURE 4: Social Security Application Form


Form 88-5
TREASURY DEPARTMENT
INTERNAL REVENUE SERVICE

U. S. SOCIAL SECURITY ACT
APPLICATION FOR ACCOUNT NUMBER

1. NAME: CLAY (EMPLOYEE'S FIRST NAME) Rex (MIDDLE NAME) Cottrell (LAST NAME)
2. Miles North West (STREET AND NUMBER) 3. MIDLAND (POST OFFICE) Texas (STATE)
4. Adams Motor Co (BUSINESS NAME OF PRESENT EMPLOYER) 5. 122 S. Lorraine St. Midland Tex. (BUSINESS ADDRESS OF PRESENT EMPLOYER)
6. 39 (AGE AT LAST BIRTHDAY) 7. 4 20 1898 (DATE OF BIRTH: MONTH, DAY, YEAR) (SUBJECT TO LATER VERIFICATION) 8. Iowa Park Texas (PLACE OF BIRTH) Archeta Co
9. M. G. Cottrell (FATHER'S FULL NAME) Miles only 10. Mattie Johnston (MOTHER'S FULL MAIDEN NAME)
11. SEX: MALE ✓ FEMALE (CHECK (✓) WHICH) 12. COLOR: WHITE ✓ NEGRO (CHECK (✓) WHICH) OTHER (SPECIFY)
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)
6 - 22 1937 (DATE SIGNED) 16. C R Cottrell (EMPLOYEE'S SIGNATURE, AS USUALLY WRITTEN) C.R. Cottrell (DATE)
DETACH ALONG THIS LINE

Notes: This figure shows an example 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. Source: https://www.legalgenealogist.com/2020/11/18/its-just-not-as-good/.

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 account of age misreporting 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 Assortative Mating

First, we estimate intergenerational mobility for men, women, and all Americans for individuals born between 1850 and 1900. Second, we estimate levels of assortative mating for those same cohorts and document correlations between the mobility and assortativeness across time and space.

3.1 Intergenerational Mobility

One of the more 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. The rank of the son or daughter is determined by the percentile their family income occupies during their prime ages (20-54 years) in the distribution of all family incomes in the same year. The rank of their parents is determined by the percentile their family income occupies during their own prime ages (20-54 years) in the distribution of all family incomes in the same year. If multiple observations are available for any individual's rank, we use their average as a proxy.

We estimate the following regression equation:

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

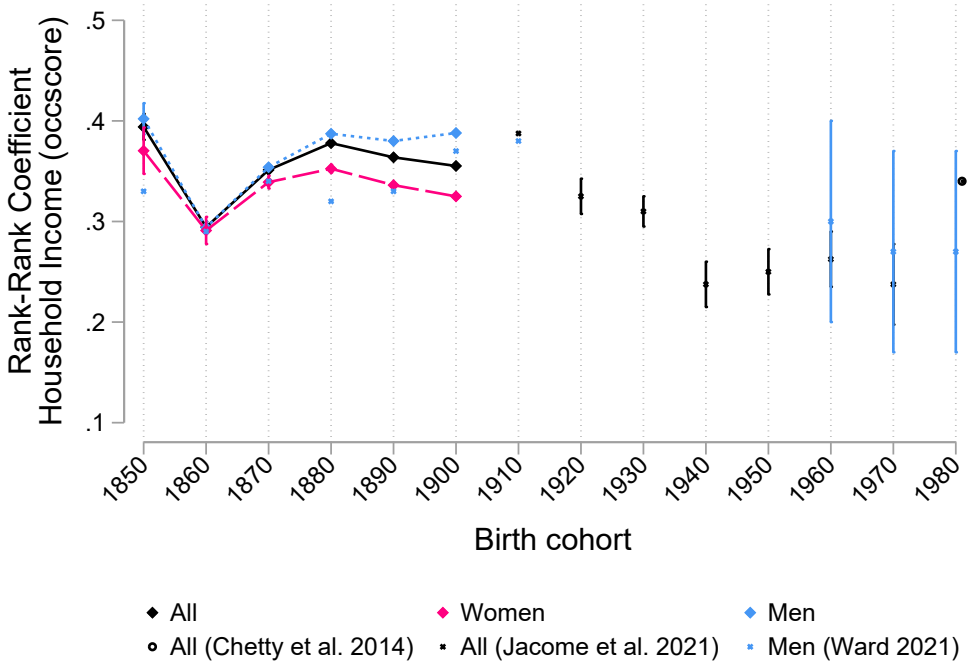
where Y_i is i's family income and $c(i)$ is i's cohort in decades. The main coefficient of interest is $\beta_{c(i)}$, capturing the rank-rank elasticity in family incomes. The lower this coefficient, the higher is cohort c's intergenerational mobility.

While women born before 1880 experienced levels of mobility similar to men, mobility between men and women quickly diverged. 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 only suggest.

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 more mobile than Black men.

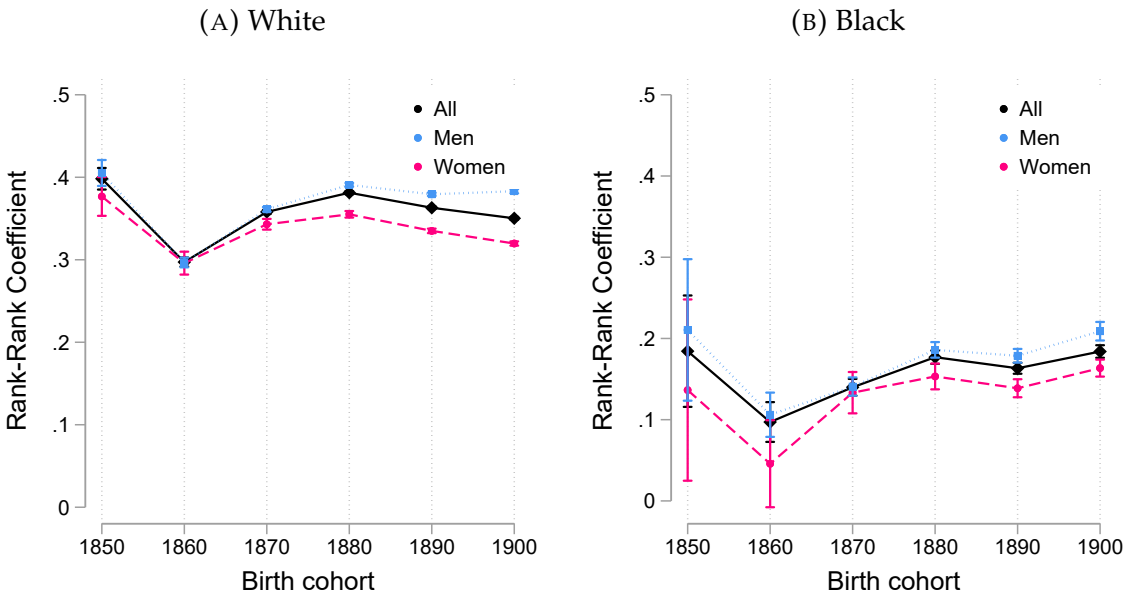
FIGURE 5: Intergenerational Elasticity of Men and Women



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.

Given that most women during the 19th and 20th Century did not participate in the labor market, we measure a woman's adult income using family income which would have been largely determined by the income of her husband. Hence, our finding that women were more mobile than their male counterparts suggests that women married men less similar in income rank to their fathers. 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.

FIGURE 6: 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.

3.2 Intergenerational Mobility & Assortative Mating

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

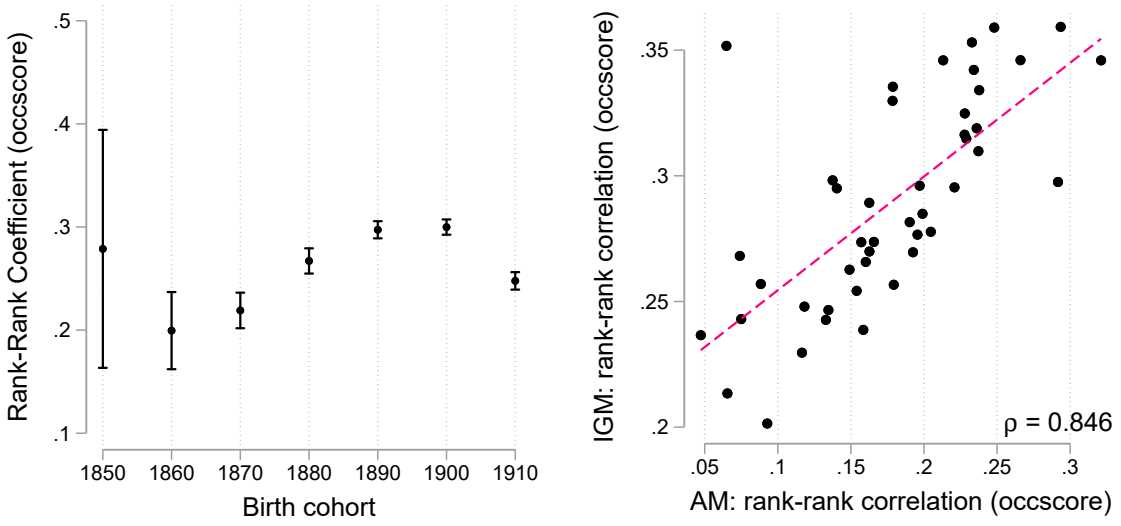
$$\text{rank}(Y_i^{\text{father}}) = \alpha_{c(i)} + \beta_{c(i)}\text{rank}(Y_i^{\text{father-in-law}}) + \varepsilon_i. \tag{2}$$

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 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 7: Assortative Mating and Intergenerational Mobility

(A) Assortative Mating (1850–1910 Cohorts) (B) AM & IGM across States (1900 Cohort)



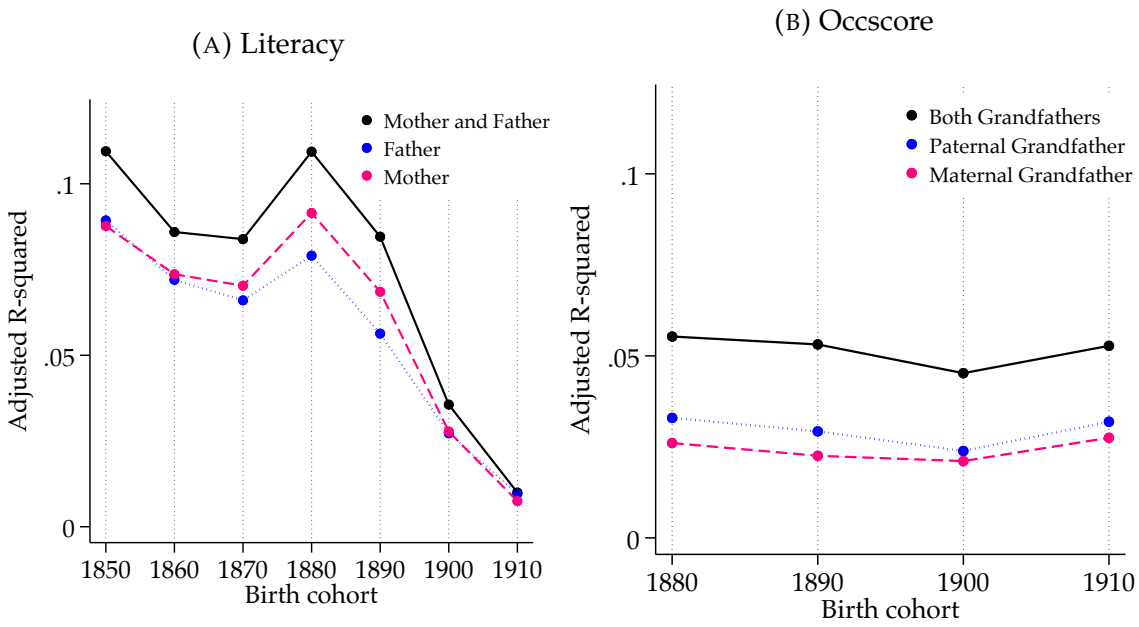
Notes: Panel (A) of this figure shows the degree of assortative mating over time as the rank-rank correlation between a person's father and their father-in-law. Panel (B) shows the correlation between assortative mating and intergenerational mobility across states for cohorts born between 1900 and 1909.

3.3 Contribution of Mothers to Intergenerational Mobility

We estimate the extent to which maternal and paternal grandfathers explain their grandchild's income rank, as shown in Figure 8. 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 9 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.

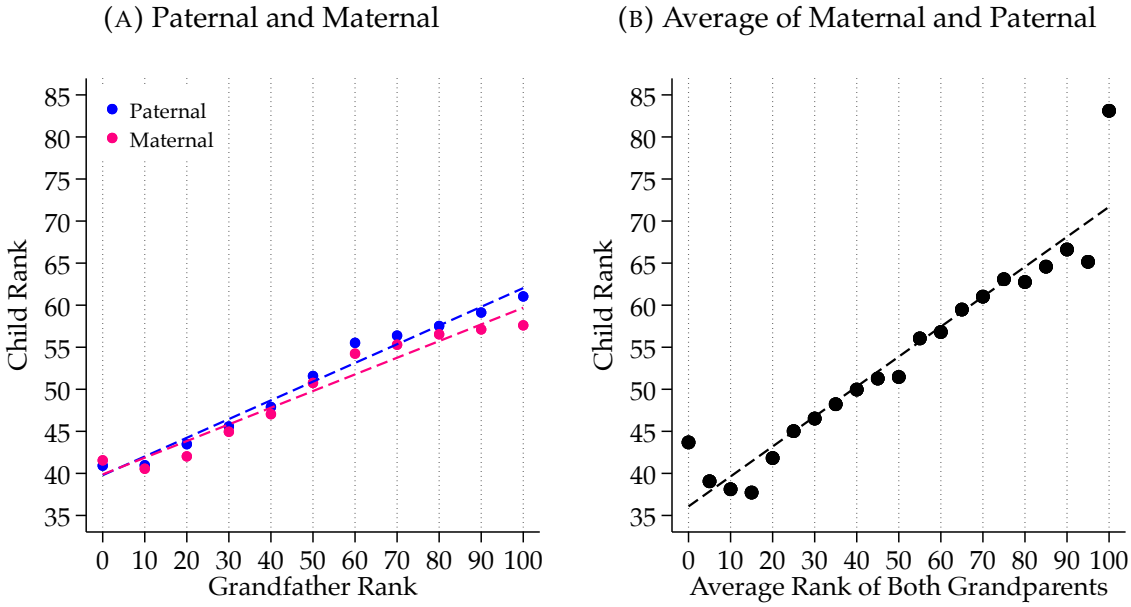
We examine the importance of both mothers and fathers in determining children's literacy outcomes. Once again, we estimate the adjusted R-squared when regression mother and father's literacy on that of their child's adult literacy. We find similar results to our income estimates. Both parents explain 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 8: 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 9: 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.

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`.

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.

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

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

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.

RUGGLES, S., S. FLOOD, R. GOEKEN, J. GROVER, E. MEYER, J. PACAS, AND M. SOBEK (2020): "IPUMS USA: Version 10.0," dataset: `https://doi.org/10.18128/D010.V10.0`.

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