NBER WORKING PAPER SERIES

MARRIAGE AND THE INTERGENERATIONAL MOBILITY OF WOMEN: EVIDENCE FROM MARRIAGE CERTIFICATES 1850-1920

Katherine Eriksson
Gregory Niemesh
Myera Rashid
Jacqueline Craig

Working Paper 34821
http://www.nber.org/papers/w34821

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2026

Reva Balakrishnan, Abi James, Simran Kalkat, and Daniel Scates provided excellent research assistance. We have benefited from helpful comments from Leah Boustan, Gregory Clark, Matthew Curtis, José Antonio Espín-Sánchez, Walker Hanlon, Bhashkar Mazumder, Joel Mokyr, Charles Moul, and Melissa Thomasson. We also thank seminar and conference participants at Stanford University, Northwestern University, Yale University, University of California - Berkeley, University of Michigan, the University of Pittsburgh, Louisiana State University, Southern Denmark University, Oxford University, Warwick University, Bristol University, The U.S. Census Bureau seminar series, the 2023 NBER Economics of Mobility Fall Meeting, the 2023 Western Economic Association Meetings, the 2022 Economic History Association Meetings, the 2019 Cliometrics Society Meetings, and Statistics Norway's Workshop on Intergenerational Mobility, Gender and Family Formation in the Long Run. Rashid acknowledges generous funding from the Northwestern Center for Economic History and the Robert Eisner Memorial Fellowship. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Marriage and the Intergenerational Mobility of Women: Evidence from Marriage Certificates 1850-1920
Katherine Eriksson, Gregory Niemesh, Myera Rashid, and Jacqueline Craig
NBER Working Paper No. 34821
February 2026
JEL No. J12, J62, N31

ABSTRACT

We document that women’s economic mobility improved nearly a century before married women gained broad labor market opportunities. Using Massachusetts marriage registers linked to U.S. censuses (1850–1920), we create new father–child links for women to estimate intergenerational mobility and assortative mating, overcoming a key historical linkage barrier. Estimates from a structural marriage market model suggest assortative mating fell 61% from 1850–1870 to 1900–1920. Counterfactuals imply women’s mobility would have been far lower absent the decline in assortative mating. Had late cohorts faced early cohort sorting, the rank–rank slope between a woman’s father and husband would have been 2.5 times higher.

Katherine Eriksson
University of California, Davis
Department of Economics
and NBER
kaeriksson@ucdavis.edu

Gregory Niemesh
Miami University
Department of Economics
and NBER
niemesgt@miamioh.edu

Myera Rashid
Northwestern University
Department of Economics
mrashid@u.northwestern.edu

Jacqueline Craig
Miami University of Ohio
jackiekcraig22@gmail.com

1 Introduction

The "American Dream" is commonly understood as the idea that opportunity in the United States is not bound by familial origins but by ability and effort. For many women during the nineteenth and early twentieth centuries, economic mobility was primarily determined through the marriage market. With nearly 85 percent of white women married and limited documented participation in the labor market after marriage, the choice of spouse became the central mechanism for women's economic advancement, or its constraint (Goldin, 1983). Consequently, assortative mating, the extent to which individuals marry within their socioeconomic group, shaped both women's adult economic standing and the intergenerational transmission of status.

This paper asks how assortative mating changed in the United States during the nineteenth and early twentieth centuries, and what those changes implied for women's intergenerational mobility. We present novel estimates of women's economic persistence over this period, and focus on marriage as one channel through which familial advantage was transmitted to the next generation. Because we measure a married woman's economic standing using her husband's status, marital sorting has direct implications for observed mobility. Stronger sorting makes daughters' adult household status more closely reflect their family background, whereas weaker sorting weakens that relationship even if underlying intergenerational transmission is unchanged.

To answer these questions, we construct a new set of intergenerational links for women from Massachusetts marriage registers (1850-1914) directly linked to U.S. census records. Because the registers report women's surnames at birth, we can link women across censuses and across generations despite surname changes at marriage. This new dataset enables us to observe father-daughter and father-son pairs across a 70-year period to track changes in intergenerational mobility. We also provide novel evidence of how assortative mating shapes women's mobility trends. Using the structural framework developed by Espín-Sánchez, Ferrie and Vickers (2023), we recover the unobserved assortative mating parameter from empirical correlations between the economic statuses of the husband, the husband's father, and the wife's father. This approach separates changes in assortative mating from other determinants of intergenerational persistence and allows us to quantify the contribution of assortative mating to women's measured mobility.

Our analysis begins by estimating relative rank-rank mobility for four 20-year marriage cohorts spanning 1850 to 1920. In the earliest cohort (1850-1870), persistence is high, with a rank-rank slope of about 0.37, and men and women exhibit similar levels of mobility. Mobility rises for both genders in later cohorts, but the gains are larger for men. By 1900-1920, the rank-rank slope falls from 0.38 to 0.21 for men (a 45 percent decline) and from 0.37 to 0.30 for women (an 18 percent decline), yielding a sizable gap between genders by the final cohort. These differences arise because men's mobility is estimated from father-son links, whereas women's mobility is estimated from father-son-in-law links. Thus, women's measured persistence reflects marriage market sorting as well as intergenerational transmission.

To understand the cause of these gender-specific trends, we turn to the structural model of assortative mating that uncovers the underlying correlation in economic status between spouses.

This approach is useful in historical settings where married women's economic status is rarely observed directly. Using a framework based on Espín-Sánchez, Ferrie and Vickers (2023), we identify assortative mating by connecting three empirical correlations: between a husband and his father, between a husband and his father-in-law, and between the fathers of spouses. Together, these correlations pin down a structural parameter for assortative mating, capturing the direct relationship between spouses.

Our results suggest high initial levels of assortative mating with significant declines by the final cohort (1900-1920). While the correlation in economic status between spouses' fathers increases by 31 percent from the 1850–1870 cohort to the 1900–1920 cohort, the structural assortative mating parameter between spouses declines by approximately 61 percent over the same period. This divergence suggests that the increasing correlation between the fathers' statuses reflects changes in the inheritance of economic status rather than stronger spousal matching. By the final cohort, assortative mating between spouses had declined significantly, indicating that marrying a partner of similar socioeconomic status became less prevalent in the late 19th century.

These changes in the strength of assortative mating played a pivotal role in shaping women's intergenerational mobility during the period. Counterfactual analyses using the assortative mating parameter reveal that stronger assortative mating significantly increased persistence in women's economic status. For example, if the 1900-1920 cohort had faced the high level of assortative mating observed in the 1850–1870 cohort, the rank-rank slope between a woman's father's and her husband's status would have been 2.5 times higher than actually observed. Conversely, weaker assortative mating, as observed in the 1900–1920 cohort, would have substantially reduced the transmission of economic status for earlier cohorts.

Finally, to understand why assortative mating declined over the late nineteenth and early twentieth centuries, we decompose the estimated structural parameter across major demographic groups. The approach separates the role of changing population composition, differences in economic status across groups, and shifts in intergenerational associations within groups. Results suggest that immigration-related changes account for most of the decline in assortative mating, through both the growing share of children with immigrant parents and the narrowing of economic differences between immigrant and native-born families. In contrast, differences by internal migration or rural-urban background do little to explain the aggregate trend, since all these groups experienced similar declines over the period. These findings help us understand the demographic channels through which assortative mating weakened and help explain why women's mobility improved over time.

We contribute to the broader literature on intergenerational mobility, marriage, and economic history in several ways. First, we provide novel estimates of women's intergenerational mobility in the nineteenth and early twentieth centuries using a newly created set of direct intergenerational links for women. Our evidence complements work by Olivetti and Paserman (2015), who use given-name pseudo-links, Jácome, Kuziemko and Naidu (2025) for the mid-twentieth century, Bailey and Lin (2024) for Ohio, Buckles et al. (2025) using genealogical tree links, and Althoff, Gray and

Reichardt (2024) using SSA Numident-based links. Our linking strategy relies on women's surnames at birth recorded in marriage registers, which provides an alternative way to follow women across historical censuses and may mitigate selection concerns that potentially arise in genealogy-based linked data.

Second, we estimate a structural model that links assortative mating to women's intergenerational mobility, addressing the challenge of limited data on women's economic status. Assortative mating typically refers to the correlation in status (e.g., income, education) between spouses, but historical data rarely allow for such direct estimates. Assortative mating is often proxied by the correlation in economic status between spouses' fathers (Buckles et al., 2023; Bailey and Lin, 2024), and prior work using this proxy finds rising sorting over a period similar to ours. The distinction is important: the correlation between fathers need not track the underlying correlation between spouses. In theory, these two correlations can move in opposite directions. Consistent with this distinction, we find that the correlation between spouses' fathers increases over time even as the structural assortative mating parameter between spouses declines.

Finally, our results contribute to the broader literature on long-run trends in intergenerational mobility (Ferrie, 2005; Parman, 2011; Long and Ferrie, 2013; Feigenbaum, 2018; Song et al., 2020; Chetty et al., 2017; Aaronson and Mazumder, 2008; Davis and Mazumder, forthcoming), reinforcing the finding that mobility has not monotonically fallen over time in the United States (Ward, 2023). We also find higher persistence for both men and women than much of the existing historical evidence. A likely explanation is that we follow Ward (2023) to address measurement error in fathers' status by instrumenting with a second observation from another census.

The paper proceeds as follows. Section 2 introduces the theoretical framework linking assortative mating to mobility outcomes. Section 3 describes the data sources and record-linking methodology. Section 4 presents the empirical results, highlighting key trends in mobility. We use the theoretical framework to uncover trends in assortative mating in Section 5, and explore its contribution to women's intergenerational mobility. Section 6 discusses the external validity of our findings. Finally, Section 7 concludes with a discussion of the implications of our findings for understanding the historical dynamics of economic mobility and gender.

2 Theoretical Framework

In this section, we present a simple theoretical framework to examine the interaction between marriage and intergenerational mobility, building on the model developed in Espín-Sánchez, Ferrie and Vickers (2023). This framework connects the underlying structural parameters of inheritance and assortative mating to the empirical relationships estimated from a family tree data-generating process. Moreover, the model allows us to perform counterfactual analyses to explore how women's mobility would have changed under higher or lower levels of assortative mating across cohorts.

The analysis relies on two-generation data where only male status outcomes are observed, specifically the occupational income scores of the husband, the husband's father, and the wife's father (husband's father-in-law), denoted as X_i^h, X_i^f, and X_i^{fl}, respectively. This setup provides three empirical relationships that can be estimated using the data.

$$b_f^h = \mathbb{E}\big[X_i^h X_i^f\big] \tag{1}$$

$$b_{fl}^h = \mathbb{E}\big[X_i^h X_i^{fl}\big] \tag{2}$$

$$b_{fl}^f = \mathbb{E}[X_i^f X_i^{fl}] \tag{3}$$

The first two equations capture the correlations between the husband's status and his father's status, as well as between the husband's status and his wife's father's status. These empirical relationships are valuable in their own right and are interpreted as measures of intergenerational mobility for the husband and the wife, respectively, consistent with prior studies (Olivetti and Paserman, 2015; Jácome, Kuziemko and Naidu, 2025; Ward, 2023). This interpretation is grounded in the historical context, where the husband's occupation contributed the majority of household economic resources. In our sample of married white couples from the nineteenth and early twentieth centuries, fewer than 6% of married white women participated in out-of-home market work (Author's calculations using Ruggles et al. (2024)).[1] These empirical relationships capture how individuals' economic position in the national distribution shifts relative to the household standing they experienced as children. The third equation captures the correlation between the husband and wife's fathers' statuses.

The intergenerational transfer of income (or status) is captured through gender-specific inheritance equations:

$$X_i^h = \beta_f X_i^f + \beta_m X_i^m + e_i^h \tag{4}$$

$$X_i^w = \beta_{fl} X_i^{fl} + \beta_{ml} X_i^{ml} + e_i^w \tag{5}$$

Here, h and w represent the husband and wife, respectively, while f and m refer to the father and mother of the husband, and fl and ml indicate the father and mother of the wife (or the husband's father-in-law and mother-in-law). The β coefficients capture the causal structural inheritance parameters that reflect the transmission of status from parent to child through mechanisms such as wealth transfers, investments in human capital, and genetics. The error terms e_i^h and e_i^w represent factors affecting the status of the husband and wife that are uncorrelated with parental status, constituting the "un-inherited" component of income or status. Equations 4 and 5 highlight that inheritance, in its most flexible form, can vary by gender and by the interaction of the child's and parent's genders (Espín-Sánchez, Gil-Guirado and Vickers, 2022).

Men and women sort into marriages based on their income or status, a process captured by the structural assortative mating parameter. This parameter, ρ, represents the correlation between the

[1] While an ideal measure of household economic mobility might include consumption or wealth, capturing the wife's contributions to household resources through in-home market and non-market work, such data are unavailable for this period.

statuses of spouses and is formally defined as:

$$\rho = \mathbb{E}\big[X_i^h X_i^w\big] \tag{6}$$

In our context, while the husband's status (X_i^h) is observable, the wife's status (X_i^w) is not directly observed. As a result, ρ must be inferred indirectly using observable correlations between other status measures within the dataset.

The structural equations can be linked to the empirical relationships by multiplying both sides of equation 4 by X_i^f and equation 5 by X_i^{fl}.[2] Taking expectations and imposing exclusion restrictions on the error terms yields the following equations:

$$b_f^h = \beta_f + \rho\beta_m \tag{7}$$

$$b_{fl}^h = b_{fl}^f\big(\beta_f + \beta_m\big) \tag{8}$$

This system comprises two equations with three unknowns: β_f, β_m, and ρ. To close the system, a third equation for the assortative mating parameter ρ is obtained by multiplying the left-hand sides of equations 4 and 5, taking expectations, and deriving the correlation between the statuses of spouses (ρ). Applying the same operation to the right-hand sides and taking expectations gives:

$$\rho = \mathbb{E}\big[X_i^h X_i^w\big] = b_{fl}^f\big(\beta_f + \beta_m\big)^2 \tag{9}$$

This framework establishes a clear link between the structural inheritance parameters and the observed empirical relationships. The system of equations 7 through 9 defines three unknown structural parameters: inheritance from the father (β_f), inheritance from the mother (β_m), and assortative mating (ρ). By solving these equations, we express the structural parameters as functions of the three empirical relationships observed in the data, enabling us to estimate their values. Status variables in the structural model are normalized and interpreted as correlations. For comparability with the existing mobility literature, we estimate empirical relationships in Section 4 using non-normalized measures, and then use normalized measures when analyzing assortative mating in the structural model in Section 5. Normalization does not meaningfully affect the results.

We use the results from this model in two key ways. First, the estimated empirical associations ($b_f^h, b_{fl}^h, b_{fl}^f$) enable us to identify the structural assortative mating parameter (ρ) and analyze how it evolves over time. This is crucial because the correlation between the fathers of spouses does not provide an unbiased measure of assortative mating. As shown in equation 9, ρ depends on

[2] We follow Proposition 1 from Espín-Sánchez, Ferrie and Vickers (2023) to obtain identification of the structural parameters β_f, β_m, and ρ. The assumptions of this proposition are two-fold. First, all possible correlations between the unmarried parents are equal (i.e. $\mathbb{E}\big[X_i^f X_i^{fl}\big] = \mathbb{E}\big[X_i^f X_i^{ml}\big] = \mathbb{E}\big[X_i^{fl} X_i^m\big] = \mathbb{E}\big[X_i^m X_i^{ml}\big]\big]$). Second, a number of exclusion restrictions are placed on the error terms to make equations 4 and 5 well specified. $\mathbb{E}\big[X_i^f \varepsilon_i^h\big] = 0$ implies that there is no interaction effect of the parent statuses, which could be driven by a marriage match surplus. $\mathbb{E}\big[X_i^{fl} \varepsilon_i^h\big] = 0$ implies there is no direct effect of the wife's father on the husband's status that does not work through the status of the wife. Finally, the strength of associative matching is constant across the two generations used in estimation.

both b^f_{fl} and the two structural inheritance terms. To frame this issue differently, an increase in b^f_{fl} could result from (1) an increase in assortative mating (ρ) or (2) a decrease in the sum of the inheritance terms $(\beta_f + \beta_m)$. However, the inheritance terms themselves are functions of the other two empirical correlations, b^h_f and b^h_{fl}. Consequently, it is possible for b^f_{fl} to rise even if assortative mating (ρ) declines, depending on the changes in these correlations.

Second, we use the estimates of the three structural parameters to conduct counterfactual analyses of women's intergenerational mobility. After solving for the inheritance and sorting parameters for each cohort, we substitute the estimate of ρ from one cohort into structural equations 7-9 for the other cohort while leaving that cohort's estimated inheritance terms unchanged. This allows us to assess how women's mobility in one cohort would have changed if assortative mating had remained at the level observed in the other cohort.

3 Data

We construct a novel two-generation family tree dataset that links both men and women to their childhood households for the nineteenth and early twentieth centuries. Using this dataset, we estimate three empirical correlations in economic status: between the husband and his father, between the husband and his wife's father, and between the fathers of the husband and wife. A key innovation in our approach is the use of birth surnames recorded in marriage registers to link adult married women to their fathers in their childhood households in historical censuses. This record linkage involves two distinct matching procedures.

First, couples identified in the marriage index are linked to a post-marriage decennial census to observe their economic status as adults. Second, the sample of successfully matched married couples is linked to their respective childhood households in a pre-marriage census to observe their fathers' economic status. This second linkage is conducted separately for husbands and wives. Economic status for both fathers and adult children is measured using reported census occupations, which are assigned an associated occupational income or wealth score, as described in detail below.

Our dataset includes four cohorts of father-son and father-daughter pairs linked across two censuses: 1850–1870, 1860–1880, 1880–1900, and 1900–1920. This section outlines the details of the linking procedures and their performance, followed by a discussion of how we measure economic status.

3.1 Linking Daughters and Sons to Fathers

We begin with digitized marriage registers for Massachusetts covering all registered marriages from 1850-1915 (FamilySearch, 2016). For both spouses, the marriage index lists their full names, birth years, and parents' full names. We construct four marriage cohorts to link to census records: 1850-70, 1860-80, 1880-1900, and 1900-1920. The sample includes marriages where at least one spouse was 20 years old or younger during their childhood census, increasing the likelihood of observing them in their childhood home.[3]

To measure adult status, we match each cohort to a subsequent post-marriage census using names and years of birth. For instance, marriages from 1850-1869 are linked to the 1870 census, while those from 1860-1879 are linked to the 1880 census, and so on. Couples marrying between 1860 and 1869 are matched to two post-marriage censuses (1870 and 1880) and are part of both the 1850-70 cohort and 1860-1880 cohort. Successfully matched couples are further linked to the childhood census conducted 20 years prior, with each spouse matched separately using names, year and state of birth, and parental names. Father's status is then measured from the childhood census.[4]

The marriage records provide uncommonly rich detail on individuals, their spouses, and their parents. We take advantage of this extra information by using a probabilistic matching method to improve the quality and number of linked records.[5] For the first match of married couples to a post-marriage census, we compare surname, two given names, and two birth years across records. Balancing small differences across these dimensions is key to selecting the best link. The second match, from the post-marriage census to the pre-marriage census, is more complex. Here, we match using four strings – child's given name and surname, father's given name, and mother's given name – along with the child's year and state of birth, to reconstruct familial and intergenerational connections.

The supervised machine learning record linkage algorithm developed by Feigenbaum (2016) allows for the effective use of detailed family structure data in the marriage registers. This approach replicates the careful, time-intensive manual linking performed by humans and scales it to otherwise cost-prohibitive large datasets. The process begins by randomly sampling couples from the marriage certificates and manually constructing true links using trained research assistants who make decision based on the information provided in the marriage registers and decennial census.[6] A logistic regression model is then trained on a subset of these true links. Parameters are optimized to adjudicate between matches with similar likelihood scores and to establish a minimum predicted likelihood threshold for a match, balancing false positives against the number of true links identified. These parameters are validated against a separate set of manually linked records not used in model training. The validated model is then applied to predict matches for the remaining marriage

[3]FamilySearch.org digitized the microfilmed marriage indices and kindly provided us access (FamilySearch, 2016). They did not digitize place of birth, which is commonly used as a blocking variable in record linkage algorithms. Complete count decennial censuses (1850-1920) were accessed through the NBER and are provided by Ancestry.com and IPUMS (Ruggles et al., 2024). The population schedules of the 1890 Decennial Census were destroyed in a 1921 fire, preventing us from creating 1870–1890 or 1890–1910 links.

[4]An alternative approach would have been linking in the opposite direction—from the marriage register to the childhood census and then to the adult census. However, tests revealed that this approach resulted in significantly fewer high-quality matches.

[5]This contrasts with much of the historical intergenerational mobility literature, which has primarily relied on automated matching procedures using exact or unique matches across two census waves (Long and Ferrie, 2013; Abramitzky, Boustan and Eriksson, 2012; Collins and Wanamaker, 2014, 2015, 2022; Abramitzky et al., 2020; Ward, 2023).

[6]The authors themselves and a trained set of undergraduate research assistants from Northwestern and UC Davis helped create the training data. Random samples of hand linked data made by research assistants were verified by the authors.

certificate sample, where true matches were not manually coded.

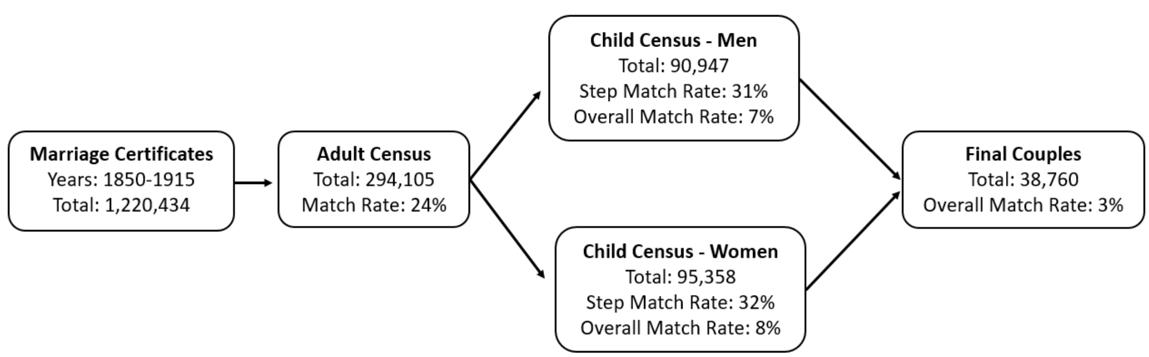
The process is repeated for the post-marriage to pre-marriage census links. Humans make links to a childhood census for a randomly sampled subset of the successful links from the marriage record to post-marriage adult census match. Then, a logistic regression prediction model is trained, parameters selected and cross-validated, and the model scaled up to the full set of potential matches. In this second step, the linking process benefits from the inclusion of parents' names listed in the marriage register, which the prediction model incorporates. This additional information, unavailable in standard census-to-census links, improves the quality of the intergenerational links by reducing false positives.

The results of this linking methodology are summarized in Figure 1. We successfully match 294,105 of the 1,220,434 couples in the marriage registers meeting our sample restrictions, achieving a 24 percent match rate in the first step. This match rate is comparable to those typically achieved in census-to-census linking, especially for censuses in the nineteenth century where quality of transcription is worse (Abramitzky et al., 2021*b*). For these matched couples, we then link each spouse individually to their childhood households in a pre-marriage census. This second step results in a sample of 90,947 father-son pairs and 95,358 father-daughter pairs, with match rates of 31 and 32 percent, respectively. Finally, we find a second observation for the father's occupation in a surrounding census with a match rate of 82 and 80 percent. The overall match rates across all three steps are 6 percent for both men and women. This rate represents links made after two steps of successful matching, resulting in a lower match rate than would be achieved in a single step. To conduct the marital sorting analysis, we restrict the sample to couples where *both* the husband and wife from a single marriage registration are successfully linked to their fathers. This stringent requirement yields 38,760 couples, representing a 3 percent overall match rate, similar to that found in the literature on assortative mating where both spouses must be matched back to their fathers Espín-Sánchez, Ferrie and Vickers (2023).

Our work introduces an innovative approach by linking both spouses to their fathers, utilizing marriage records to facilitate this connection. This approach, while novel, results in lower match rates compared to the historical literature on father-son mobility, which typically requires linking only two data sources rather than three. In our analysis of status associations between fathers (i.e. father/father-in-law), both spouses must be independently linked to their fathers. This requirement inherently limits the analysis primarily to *native-born* couples, as it is challenging to link an immigrant spouse to their father unless the father immigrated with the family when the spouse was a child. Consequently, observations involving immigrant spouses are underrepresented. A related implication is that the reported match rate for double-matched couples is artificially deflated. The denominator includes all marriage registrations, even those with immigrant spouses who could not realistically be linked to a U.S. census because their fathers never migrated. By contrast, the analytically relevant population consists only of native-born couples and couples with an immigrant spouse who arrived with her father, i.e. those who could plausibly appear in the census. The marriage register does not contain information that would allow us to identify which cases fall into this linkable subset, which limits our ability to adjust the match rate accordingly.

Mobility estimates are generally biased when a father's economic status is measured with error (Solon, 1992). Using a single observation of the father's status, even if based on a proxy like occupation, introduces measurement error due to transitory shocks or linking errors, such as linking to the wrong father (Mazumder and Acosta, 2015; Bailey et al., 2020). To address this, we follow Ward (2023) by instrumenting one observation of the father's status with a second observation to mitigate bias from measurement error. Rather than creating new links to identify a second observation, we rely on pre-existing links established by other researchers, specifically those from the Census Tree Project (Price et al., 2021; Buckles et al., 2025). The search for a second observation prioritizes proximity, beginning with the nearest census, 10 years forward then back, before expanding to a 20-year window in both directions. This approach yields a second observation for approximately 80% of fathers in our linked sample.[7] Table 1 further reports match rates and number of observations for each of the four cohorts.

Figure 1: Illustration of double match procedure and corresponding match rates



Notes: This figure depicts a summary of the steps of our double-match procedure to link men and women from the marriage registers to their adult and childhood censuses. Each step reports the number of observations in that step as well as the overall match rates which use the sample of marriage certificates. In the second step of linking from adult to childhood census, we additionally report step match rates which use the sample of 294,105 men and women that were successfully linked in the first step as denominator.

3.2 Performance of the Linked Sample

We evaluate the performance of our linked sample in three ways: by testing the record linkage algorithm on a cross-validation sample of "true" links, comparing our links to an independently created multi-generational linked dataset, and assessing how representative our sample is across key characteristics relative to the population.

[7]Using the MLP links (Helgertz et al., 2023), we find slightly fewer second observations (70% coverage). However, mobility estimates are virtually identical to those obtained using the Census Tree Project links.

Table 1: Summary of sample sizes and match rates across the two linkage steps by cohort

	Pool of marriages	Matches: Marriage to adult census		Matches: Adult census to child census		Matches: Additional Dad Ob			Both matched to child census
	Obs	Obs	Step match rate	Obs	Step match rate	Obs	Step match rate	Total match rate	Obs
1850-70	217,858	39,553	0.18	M: 13,143	0.33	10,196	0.78	0.05	6,330
				W: 14,979	0.38	11,506	0.77	0.05	
1860-80	250,598	56,745	0.23	M: 19,214	0.34	17,375	0.90	0.07	7,887
				W: 19,359	0.34	17,181	0.89	0.07	
1880-00	372,213	91,235	0.25	M: 27,743	0.30	24,104	0.87	0.06	11,034
				W: 27,134	0.30	22,670	0.84	0.06	
1900-20	379,765	106,572	0.28	M: 30,847	0.29	22,488	0.73	0.06	13,509
				W: 33,866	0.32	24,680	0.73	0.06	
Total	1,220,434	294,105	0.24	M: 90,947	0.31	74,163	0.82	0.06	38,760
				W: 95,358	0.32	76,037	0.80	0.06	

Notes: This table reports sample sizes and match rates at each step of the linkage process for four marriage cohorts: 1850–1870, 1860–1880, 1880–1900, and 1900–1920. The sample begins with all marriages registered in Massachusetts during each cohort period. Each couple is first linked to a post-marriage adult census, and each spouse is then linked to a pre-marriage childhood census. To construct the instrument for fathers' economic status, we obtain a second observation of the father's occupation from a census surrounding the childhood census. The total match rate is the product of the three step-specific match rates.

In training the model, we optimize two hyper-parameters: one sets the minimum predicted likelihood for a match, and the other determines the minimum difference in likelihood between the top two potential matches to call the highest likelihood a match. These hyper-parameters are chosen to maximize the True Positive Rate (TPR) while maintaining a the share of matches that are false positives below 10% in the training data, which aligns with that reported in census-to-census linkage studies of men, as suggested by Abramitzky et al. (2021*b*) and Bailey et al. (2020). The TPR measures the share of true links correctly identified by the algorithm.

The algorithm's share of false positives in the hold-out validation sample align with existing methodologies and are near the target rate, while the proportion of true links correctly identified remains high. In the marriage register-to-adult census linkage, 11% of matches in the full sample are false positives, while the algorithm captures 77% of true matches. The adult census-to-childhood census linkage performs even better, with an 8% of matches coded as fals positives for both men and women and a TPR of 93%. The improvement in the second linkage step likely reflects the availability of more detailed family information. In contrast to standard census-to-census algorithms, which rely on surname, given name, state of birth, and birth year, our second step also incorporates the given names of the parents. By comparison, the marriage register-to-adult census match relies only on surname, given names, and birth years for both spouses. Appendix Table B2 reports a full set of cross-validation results by match step, marriage cohort, and for men and women.

The target false positive share of 10% for our algorithm was chosen to align with rates commonly used in the literature, but this choice is not fixed. Bailey et al. (2020) demonstrates that higher shares of false positive in the linked sample can attenuate mobility estimates. However, lowering the false positive rate involves a trade-off: it reduces the sample size and raises concerns about the sample's representativeness. In robustness checks, we reduce the target share of false positives to 5% and find that the mobility estimates remain highly consistent with those from the base sample using a 10% target rate.

A potential concern is that record-linkage error may differ systematically by gender. When linking couples from marriage registers to their adult census records, an incorrect match can still preserve the correct husband–father connection because men typically retain the same surname across censuses. In contrast, an incorrectly linked couple almost always implies an incorrect match between the wife (husband's status) and her own father, since her surname changes at marriage. As a result, the husband–wife's-father correlation could contain more measurement error than the husband–own-father correlation. However, our linkage procedure substantially mitigates this concern. For both men and women, the algorithm requires agreement on the given names of both parents in addition to other identifying characteristics, which produces links with a very low share of false positives and high overall accuracy. This design leaves less room for gender-asymmetric linkage error than standard census-to-census linking methods. Additionally, we re-estimate men's and women's mobility using increasingly conservative linkage thresholds that successively lower the target false-positive rate, thereby reducing potential measurement error. As shown in the robustness checks, the estimated gender gaps and cohort trends remain essentially unchanged across these samples. Taken together, these checks suggest that any difference in measurement error between men and women is small and unlikely to drive our findings. Moreover, for measurement error to account for the observed decline in persistence over time, the rate of mislinking would have to increase sharply across cohorts, a pattern inconsistent with our observed linkage quality and robustness checks using stricter false-positive thresholds.

Nonrandom selection into the linked sample is a common concern (Bailey et al., 2020; Abramitzky et al., 2021*b*). Appendix Figure A1 plots the representativeness of our matched sample relative to the population of married men and women who are born in MA and recorded as living in MA during the later census decade.[8] The matched sample is representative of the population in terms of literacy, being US-born and average occupational status (IPUMS *occscore* variable). However, matched observations are less likely to be from farm households, a consistent difference across cohorts and for both men and women. The matched sample also contains fewer individuals with immigrant parents, which is expected given the challenges of linking to immigrant families discussed earlier. Finally, the matched sample is slightly more likely to include individuals with children in the 1900–1920 cohort.[9]

When differences between the population and the linked sample arise, we address this by applying inverse propensity-score weighting (IPW), as recommended by Pérez (2017), Zimran (2019), Bailey et al. (2020), and Abramitzky et al. (2021*b*). These weights adjust the linked sample to reflect the population of adults born in Massachusetts. Specifically, we reweight on age, literacy, occupational score bins, farming households, living in group quarters and having children.

One key contribution of our linked sample is its ability to track women, particularly married women, across censuses. Here, we compare them to genealogical links created through the FamilySearch Family Tree, which also tracks women across the marital transition (Buckles et al., 2025). The Family Tree's manual links are created by the public searching for their ancestors, professional genealogists, and trained research assistants at the Family History Lab at Brigham Young University. These genealogical links can trace women across censuses by using private family knowledge unavailable to researchers relying solely on the information recorded in the decennial censuses. While our approach is more limited in scope and the number of matches, it offers three key advantages over genealogical methods. First, it relies solely on publicly available information from marriage records to link women. Second, it can be easily replicated and applied to cases where such genealogical links are unavailable. Third, it reduces biases associated with a sample of individuals whose descendants were uniquely able to link them across censuses.

To compare the sample of linked women and men from our approach to genealogical links, we

[8]We choose this as the population to compare to because there is no equivalent comparison group to being married in Massachusets in the censuses. This also allows for the construction of appropriate weights using census data. The couples in the marriage registers were all married in MA as adults and 72% of those in our linked sample were born in MA. We also restrict the sample of our links to be compared accordingly to MA-born and residing in MA as adults. Later, this choice of population will also allow for a direct comparison to genealogical links.

[9]The 1860–1880 cohort in our sample is less likely to be urban than the population, a discrepancy partly due to the IPUMS coding of the *urban* variable for Massachusetts. The 1880 census classifies observations as "urban" at a much lower rate than adjacent censuses or other states due to Massachusetts' unique town-based boundaries. Interested readers can refer to the IPUMS documentation for the *urban* variable.

first evaluate the accuracy of our links. Appendix Table A1 compares samples of linked men and women created using our method (with varying false positive rates used to choose cutoff parameters) to genealogical links from the FamilySearch Family Tree. In the absence of ground truth data, we report accordance rates, which measure the share of matches that both methods link to the same record in the later census, conditional on attempting to link the same individual from a base year (Abramitzky et al., 2025). For both methods, we focus on individuals born in Massachusetts and residing there as adults. Our method achieves high accordance rates for both women (98–99%) and men (97–99%) across different false positive thresholds.

In Appendix Figure A1 we compare the representativeness of the Family tree links an our own links to that of the population of men and women born and living in Massachusetts during the later census decade. Compared to genealogical links, our method is more representative of the population on two attributes (parental nativity and the presence of children) but less representative on one (farming households). Abramitzky et al. (2025) also report lower representativeness of immigrant groups in genealogical data. For all other attributes, our method performs similarly to genealogical links on average in terms of representativeness.

3.3 Measuring Economic Status

A key difference between modern and historical mobility studies is the lack of income data in historical U.S. censuses. The U.S. Decennial Census did not collect income information before 1940 and only occasionally recorded household wealth (e.g., 1860 and 1870). Consequently, historical mobility studies rely on proxies for income, typically derived from reported occupations (Ferrie, 2005; Long and Ferrie, 2007, 2013; Olivetti and Paserman, 2015; Feigenbaum, 2018; Collins and Wanamaker, 2014; Abramitzky et al., 2021*a*; Ward, 2023). Our preferred approach constructs an occupational wealth score based on the total property value reported by heads of households in the 1870 census (Ferrie, 1999; Collins and Zimran, 2019; Ager, Boustan and Eriksson, 2021; Collins and Zimran, 2023). This measure draws on detailed data to predict wealth by occupation, region, race, and immigrant status. Moreover, wealth serves as a key indicator of economic security, representing the resources available at any given time. Wealth can improve educational and occupational opportunities for children, provide capital to start a business, and act as a safety net during emergencies (Hamilton and Darity, 2017). Because this measure does not capture within-occupation changes after 1870, we supplement it with robustness checks using three alternative status proxies that produce similar results.

We calculate the mean total property wealth (real and personal) for all white males aged 18 to 65, segmented by occupation, region, and immigrant status. If fewer than 50 observations are available for an occupation-region-immigrant cell, we use the mean wealth for the occupation by region, and if necessary, the national occupation mean. To reflect the evolving occupational structure of the American economy in the 19th and early 20th centuries, we place each observation's wealth score into the cohort-specific national wealth distribution and calculate its percentile rank. For example, fathers of the 1850–1870 cohort are ranked in the 1850 distribution, while their sons are ranked in the 1870 distribution.

This method implies that a son remaining in the same occupation as his father does not necessarily retain the same rank; he may experience upward or downward mobility. Similarly, transitioning to an occupation with a higher wealth score than the father's does not always result in an improved rank, as shifts in the overall occupational distribution could keep the son's rank stable or even lower it. Like all occupational scores, our wealth score cannot capture within-occupation variation in status. In addition, because each occupation–region–immigrant cell is assigned a fixed 1870 dollar value, the measure cannot reflect changes in the relative wealth of those cells between 1850 and 1910 (e.g., if machinists' average wealth rises relative to farmers' after 1870). This limitation applies to all occupational scores derived from a single cross-section and is distinct from the aggregate compositional shifts addressed by our re-ranking step.

We also demonstrate that our results are consistent across several alternative proxies for income or economic status. Additional appendix tables present mobility and marital sorting estimates using the IPUMS *occscore* variable, which is based on the 1950 income distribution; income scores from a 1901 Cost of Living report and the 1900 Census of Agriculture for farmers; and a literacy-based occupational score derived from Song et al. (2020) and Ward (2023).

4 Empirical Status Correlations among Fathers, Husbands, and Fathers-in-law

This section presents the core empirical relationships that underpin our analysis of assortative mating and intergenerational mobility. We estimate correlations of status between a husband and his father, between a husband and his wife's father, and between the fathers of husbands and wives. These three relationships summarize how economic status is associated across generations and across family lineages, providing the empirical basis for the structural model of assortative mating developed in the section 2. We describe the estimation approach, report the resulting correlations and their evolution across cohorts, and evaluate the robustness of these empirical patterns to alternative measures and sample definitions.

4.1 Intergenerational Mobility

We measure intergenerational relative mobility using rank-rank regressions, which regress an adult's economic status rank on their father's rank while allowing the intercept and slope to differ by gender:[10]

$$Adult\ Rank = \alpha + \beta_0 Woman + \beta_1 Rank\ Father + \beta_2 Woman \times Rank\ Father + \epsilon \tag{10}$$

[10]Intergenerational elasticity (IGE) estimates are provided in the appendix. The log-log IGE estimate assumes a linear relationship between a father's log economic status and the child's log economic status, but this assumption breaks down in the tails of the income distribution in modern data (Corak and Heisz, 1999; Chetty et al., 2014*b*). Following Dahl and DeLeire (2008); Chetty et al. (2014*a*); Mazumder (2014); Collins and Wanamaker (2022), we estimate a rank-rank regression that imposes a linear relationship.

The intercept term α captures the absolute rank mobility of men born to fathers at the bottom of the wealth distribution, while β_0 measures the differential absolute mobility for women. Our primary coefficients of interest are β_1, which measures the relative rank mobility for men, and β_2, which measures the difference in mobility for women relative to men. This regression is estimated separately for each of the four cohorts. To address life-cycle bias, all estimates include a quartic in the father's age (relative to age 40) and a quartic in the adult husband's age when economic status is measured (relative to age 40).[11]

Mobility estimates will be biased due to the extent that the status of the father is measured with error (Solon, 1992). A single observation of father's economic status, even if based on a proxy like occupation, would cause measurement error in the presence of transitory shocks or errors in the data from linking to the incorrect father (Mazumder and Acosta, 2015; Bailey et al., 2020). We follow Ward (2023) in instrumenting one father observation with another to eliminate the bias posed by measurement error. Section 3 provides details on the procedure used to find a second father observation. All results report the 2SLS estimates of the β's in equation 10.

Biased estimates of mobility can also be caused by unrepresentative samples, which occurs when the linking process is more likely to find a successful match based on characteristics of the observation (Solon, 1992; Bailey et al., 2020). Bailey et al. (2020) and Abramitzky et al. (2021*b*) propose methods to evaluate the representativeness of linked samples. When unrepresentativeness is an issue, reweighting the sample to reflect the original population may provide a solution. The main results limit the sample to married adults born in Massachusetts, which also provides the population with which we create inverse probability weights.

Turning now to the results, we begin by examining the evolution of mobility over time, focusing on key trends in the rank-rank estimates for men and women across cohorts. Our findings suggest an overall increase in mobility for both men and women during the sample period.

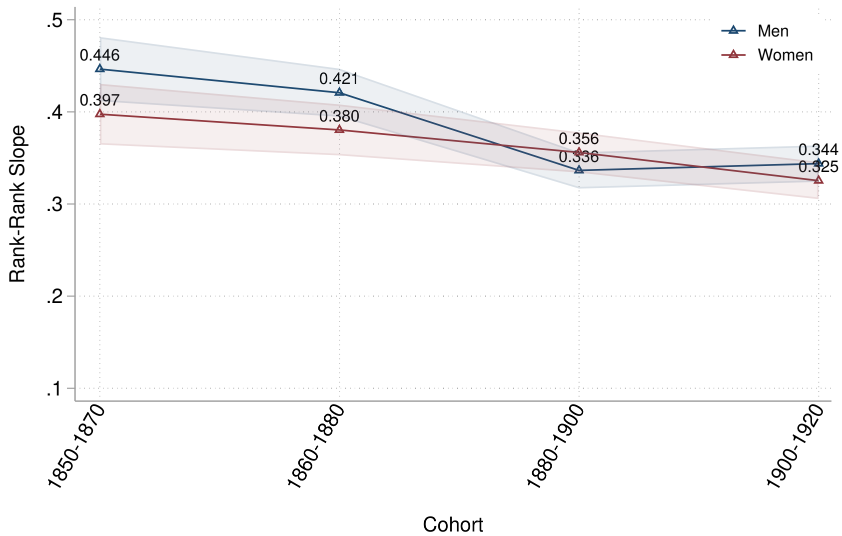
Figure 2 plots the mobility estimates for men ($b^h_f = \beta_1$) and for women ($b^h_{fl} = \beta_1 + \beta_2$). Panel (a) presents estimates for the full sample of all linked observations for couples listed in the Massachusetts marriage registers. For men, the rank-rank parameter declines by 23 percent, from 0.446 in the 1850–1870 cohort to 0.344 in the 1900–1920 cohort. In the earlier cohorts (1850–1870 and 1860–1880), women exhibited greater mobility than men, with statistically significant differences in the rank-rank parameter across genders. However, this gap diminished over time, with the estimates converging by the 1900–1920 cohort. While both men and women experienced declines in the rank-rank parameter, the decline for women (18 percent) was less pronounced than for men. Appendix Table A2 presents the full coefficient estimates from equation 10 with standard errors to provide inference on the gender differences.

The full sample includes in-migrants to Massachusetts as well as observations born in the state, making it difficult to isolate the causes of mobility changes over time. To better address these

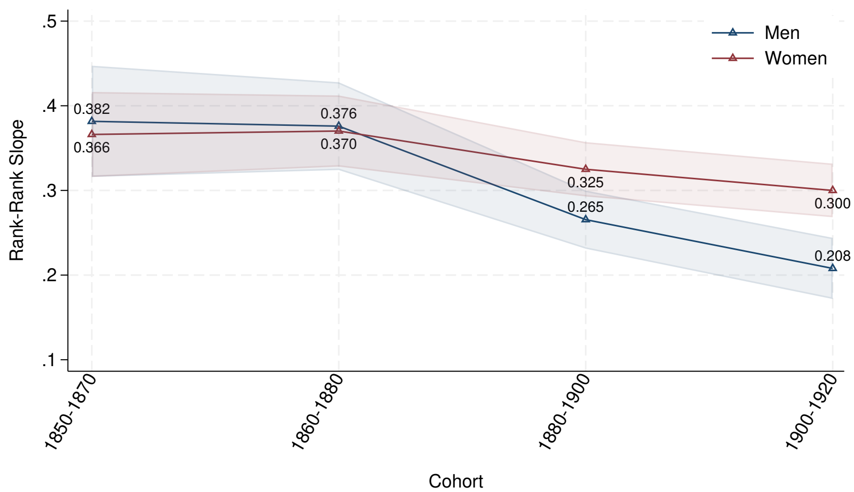
[11] Life-cycle bias arises because fathers and adult children may not be observed at the same age (Grawe, 2006). Haider and Solon (2006) show that life-cycle bias is minimized when income is measured at age 40. Since our data record occupation at a single point in time, we include controls to capture the life-cycle pattern of economic status (Aaronson and Mazumder, 2008; Lee and Solon, 2009).

Figure 2: Intergenerational Mobility Estimates: Rank-Rank Slopes

(a) Full Sample



(b) Massachusetts Born



Notes: Each entry represents an estimate of the rank-rank parameter from separate regressions of own wealth score rank on father's wealth score rank by cohort, using a second father observation as an instrument for father's status taken from the Census Tree Project links. The mobility estimate for men corresponds to β_1, while the estimate for women is $(\beta_1 + \beta_2)$ from equation 10. Shaded areas indicate 95 percent confidence intervals. All regressions include control variables: a quartic in father's age and a quartic in husband's age, both measured at the time economic status is observed. Panel (a) presents results for the full sample, while Panel (b) restricts the analysis to individuals born in Massachusetts. For the Massachusetts-born sample, observations are weighted to adjust for differential selection into the linked sample. Observations are not weighted in the full-sample results in Panel (a).
Sources: 1850-1920 Decennial Census data from Ruggles et al. (2024). Marriage certificates from *FamilySearch.org*. Census Tree Project links from (Price et al., 2021; Buckles et al., 2025)

dynamics, we focus much of the subsequent analysis on a Massachusetts-born sample. Mobility trends in this sample, presented in Panel (b), are broadly similar to those in the full sample, with a few notable differences. First, there is a downward level shift in the rank-rank parameters, indicating higher mobility within the Massachusetts-born sample. For men, the rank-rank parameter for the first cohort is 0.382 compared to 0.446 in the full sample, a difference of 14 percent.

Second, unlike in the full sample, the rank-rank parameters for women are statistically indistinguishable from those for men in the first two cohorts. Across cohorts, mobility increases for both genders in the Massachusetts-born sample, but the improvement begins only with the 1880–1900 cohort. The increase in mobility for men is more pronounced than for women, leading to a divergence in rank-rank parameters by the final two cohorts. Specifically, men's mobility increases by 46 percent, with the rank-rank parameter decreasing from 0.382 to 0.208. Women also experience a substantial improvement in mobility, with the rank-rank parameter declining by 18 percent, from 0.366 to 0.300.[12]

The mobility trends depicted in Figure 2 remain consistent across different measures of economic status. Appendix Table A3 presents rank-rank mobility slopes calculated using several alternative occupational scores. Our base results rely on an occupational wealth score derived from wealth data recorded in the 1870 decennial census. This measure, calculated for each occupation, census division, and immigrant/non-immigrant group, was chosen for its suitability over a long historical period, its ability to provide detailed demographic distinctions, and its effectiveness in capturing the significant wealth gap between immigrants and the native-born (within occupations, while flexibly controlling for age).

Similar mobility trends are observed across three alternative measures: a pooled (non-immigrant-specific) 1870 wealth score, an occupational income score constructed from the 1901 Cost of Living Survey (Preston and Haines, 1991; Abramitzky et al., 2021*a*), and a human capital measure based on occupational literacy rates (Song et al., 2020; Ward, 2023). For both men and women, persistence is higher in the first two cohorts and declines in the 1880–1900 and 1900–1920 cohorts, mirroring the patterns seen with the base measure.

These alternative measures, however, handle immigrant–native gaps differently than our base measure, often producing slightly higher rank-rank parameters (lower mobility). Children of immigrant fathers experienced faster upward mobility than children of native-born fathers, and immigrant fathers constitute a significant and growing share of our linked sample over time. Because the pooled status scores used in the alternative measures place immigrant fathers at higher initial ranks than our base measure does, they mechanically reduce the measured distance that their children travel in rank space. As a result, these measures mute the differentially high upward mobility of children of immigrant fathers and yield estimates that imply somewhat higher persistence than our

[12]Rank-rank slopes offer a convenient global summary of intergenerational mobility, but they can mask important variations in mobility at different points along the distribution (Deutscher and Mazumder, 2023). Appendix C uses binscatter plots and transition matrices to examine how intergenerational mobility varies across the wealth distribution, highlighting cohort and gender differences in upward and downward mobility that are obscured by average rank-rank slopes.

baseline specification.

Finally, we observe anomalous results when using a measure of occupational status that falls well outside our sample period. A commonly used historical measure, the IPUMS *occscore* variable, is constructed from the median income in occupational cells based on the national income distribution in the 1950 decennial census. Using this measure results in a steady increase in persistence across all cohorts, likely due to its temporal misalignment. The measure is separated by roughly 100 years from the earliest father cohort in our sample, reflecting income distributions shaped by the "Great Compression" of the mid-20th century with scope for dramatic changes in occupational rankings due to industrialization and the decline of agriculture. For these reasons, we focus on our base measure, which yields results consistent with alternative measures that are more temporally aligned with our study period.

Mobility trends are robust to different specifications and sample limitations, as shown in Appendix Table A4. During periods of structural transformation, such as the shift from agriculture to industry, mobility estimates may be influenced by how farmers and farm income are treated (Xie and Killewald, 2013). To address this, Panel (A) excludes all observations where either the father or the child reports their occupation as farmer. Life-cycle bias, though partially addressed through quartic age controls for both fathers and husbands, may still affect our estimates. Panel (B) restricts the sample to individuals whose economic status was measured between ages 30 and 50, a range where measurement error is less likely to occur (Mazumder, 2005; Haider and Solon, 2006). Finally, Panel (C) reports estimates using the log-log IGE mobility parameter. Across all these alternative specifications, we find substantively similar mobility trends to those in the base specification: relatively high persistence in the first two cohorts, followed by a sustained increase in mobility.

Our results are robust to using more conservative linkage parameters that yield lower false positive rates. Appendix Table A5 reports husband–father and husband–father-in-law correlations from samples constructed under progressively stricter thresholds, down to a 5% share of false positives. We adopt the 10% threshold as our baseline because it produces mobility estimates nearly identical to those from stricter cutoffs while retaining a larger, more representative sample given the trade-off between false positives and sample size.

4.2 A Proxy of Marital Sorting: Father and Father-in-law Correlations

We estimate the association in economic status between the fathers of spouses using a rank-rank regression, similar to the approach used to estimate mobility. Equation 11 specifies this relationship, regressing the occupational wealth score rank of the husband's father (Y_i^f) on that of the wife's father (Y_i^{fl}). This empirical specification corresponds to b_{fl}^f in the theoretical framework.

$$Y_i^f = \alpha + b_{fl}^f Y_i^{fl} + v_i \tag{11}$$

To address potential measurement error, the occupational wealth score of the wife's father (Y_i^{fl}) is instrumented using a second observation of his status. Each regression includes quartic terms for the ages of both the husband's and wife's fathers at the time economic status is measured. Observations are weighted to adjust for differential selection into the linked sample.

The estimates for $\widehat{b^f_{fl}}$ capture one dimension of sorting in the marriage market based on socioeconomic background, the correlation between the socioeconomic status of spouses' parents. An interesting empirical relationship in its own right, it is distinct from the economic concept of assortative mating between spouses themselves. Figure 3 presents our initial findings on the Massachusetts marriage market during the nineteenth and early twentieth centuries. As expected, we observe a strong positive association between the socioeconomic status of husbands' and wives' fathers across all periods. Notably, the evidence indicates an upward trend, with a 31 percent increase from the 1850–1870 cohort to the 1900–1920 cohort. This trend aligns closely with the timing of observed changes in mobility, beginning with low levels in the first two cohorts and increasing significantly between the 1860–1880 and 1900–1920 cohorts. The pattern remains robust across alternative specifications, including regressions of the father-in-law's status on the father's status, the use of alternative measures of economic status, and restrictions based on age or farmer status (see Appendix Table A6).

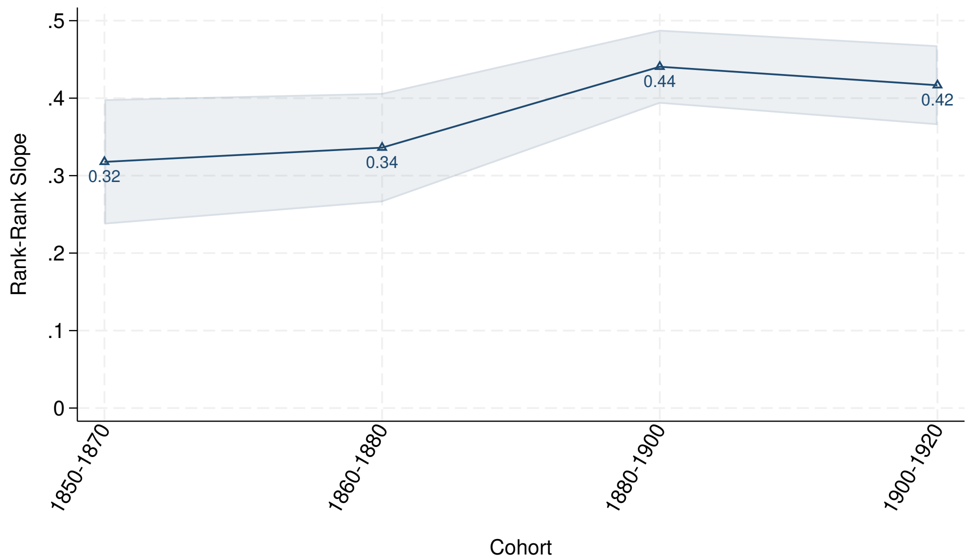
These estimates are somewhat smaller than those for more recent periods in the United States. For example, Charles, Hurst and Killewald (2013) report a parental wealth correlation of 0.4, and Gihleb and Lang (2020) find a correlation of around 0.6 for spouses' education. In historical U.S. contexts, our findings align with studies showing increasing correlations in the status of spouses' fathers. Olivetti et al. (forthcoming) report correlations in the occupational status of spouses' fathers ranging from 0.008 to 0.03 in the Northeast—10 to 30 times smaller than our estimates. The authors note that their pseudo-linking methodology likely biases these estimates downward due to the use of imperfect proxies for parental status. After adjusting for this bias, they find national correlations between 0.1 and 0.3, which are more consistent with our results.

Buckles et al. (2023) provide national estimates for the father/father-in-law correlation using family tree links and the adjusted Song score for occupational status. Their estimates are higher than ours, ranging from 0.63 to 0.75, and are relatively stable with small declines over time. These discrepancies may arise from two factors. First, national-level correlations are typically higher than regional ones due to regional inequality and the prevalence of intra-regional marriages. Second, the adjusted Song score, as constructed by Ward (2023), does not account for differences in fathers' immigrant status, which we found to be significant in Section 4 for mobility estimates.

5 The Marriage Market and Assortative Mating

The preceding analysis of intergenerational mobility reveals intriguing patterns that cannot be fully understood without considering the role of the marriage market. Our structural analysis of the marriage market formally connects these empirical relationships to the underlying assortative-

Figure 3: A Proxy of Marital Sorting - Father on Father-in-law Estimates



Notes: Each entry reports the estimated rank-rank parameter from separate regressions of the husband's father's wealth score rank on the wife's father's wealth score rank, stratified by cohort. A second observation of the wife's father's wealth score is used as an instrument. These estimates correspond to $\widehat{b^f_{fl}}$ in equation 11. Shaded areas represent 95% confidence intervals. All regressions control for a quartic in age for both fathers, measured at the time their economic status is observed. Observations are weighted to account for differential selection into the linked sample.
Sources: 1850-1920 Decennial Census data from Ruggles et al. (2024). Marriage certificates from *FamilySearch.org*.

matching parameter. During this period, the economic status of white married women was often shaped by their choice of partner, as their participation in the out-of-home labor market was limited (Goldin, 1983). It is therefore crucial to examine how patterns of assortative mating—the tendency for individuals to marry those with similar economic status—affected mobility trends over time. This section applies the econometric framework introduced in Section 2 to estimate the extent of assortative mating in the marriage market and assess its impact on observed trends in intergenerational mobility.

Assortative mating is unobserved during this time period, because married women often lack recorded occupations and status to correlate with their husbands. To uncover an underlying structural assortative mating parameter, we use the theoretical framework described in Section 2 and based on Espín-Sánchez, Ferrie and Vickers (2023). The framework connects assortative mating to three key empirical correlations observed in two-generation linked data: (1) between a husband and his father, (2) between a husband and his father-in-law, and (3) between the fathers of the spouses. We use these estimates from the previous section to calculate the assortative mating parameter and explore its trend over time. Finally, we use the structural parameters to conduct counterfactual analyses, investigating how different levels of assortative mating would have influenced women's intergenerational mobility.

Our main analysis focuses on a linked sample of Massachusetts-born individuals where both the husband and wife are matched to their fathers. This approach offers significant flexibility in modeling inheritance and assortative mating parameters, using realistic assumptions on inheritance parameters. However, it reduces the sample size and limits the inclusion of observations with immigrant spouses. We demonstrate that the trends observed in our results are robust to concerns about the sample's representativeness.

The structural parameters for assortative mating (ρ) and the inheritable portion of parental status (β_f for fathers and β_m for mothers) can be recovered using the equations outlined in Section 2:

$$\begin{aligned} \widehat{b_f^h} &= \beta_f + \rho\beta_m \\ \widehat{b_{fl}^h} &= \widehat{b_{fl}^f}(\beta_f + \beta_m) \\ \rho &= \widehat{b_{fl}^f}(\beta_f + \beta_m)^2 \end{aligned} \tag{12}$$

where the terms with hats represent estimates for son and daughter mobility and the correlation between spouses' fathers. Appendix Table A7 prints the full set of empirical correlations used in this exercise. We provide the 95 percent confidence interval from bootstrapping the regression for 1,000 repetitions to perform inference on the structural parameters.

Figure 4a displays the base estimates for the assortative mating parameter, ρ, shown in black. The results indicate strong assortative mating between spouses in the first two cohorts, followed by a significant decline in the last two cohorts. Specifically, assortative mating decreased by approximately 61 percent from the 1850–1870 cohort to the 1900–1920 cohort. These findings suggest that marrying a spouse of similar socioeconomic status became less important over the course of the 19th century, at least within the scope of our sample.

The results for assortative mating highlight the limitations of interpreting the correlation between the husbands' and wives' fathers' statuses as a direct measure of assortative mating between spouses, as traditionally understood by economists. While we observe increases in the correlation between the fathers (b_{fl}^f), we simultaneously find decreases in assortative mating between spouses (ρ). This apparent contradiction can be resolved by examining the system of equations in 12.

The correlation between the fathers (b_{fl}^f) can increase for reasons unrelated to an increase in assortative mating (ρ). As shown in the final equation in 12, ρ can decline even as (b_{fl}^f) increases if the sum of the inheritance parameters ($\beta_f + \beta_m$) also declines. This is precisely what occurs in our data. Furthermore, the second equation in 12 demonstrates that the ratio of (b_{fl}^f) and (b_{fl}^h) determines the sum of the inheritance parameters. In our data, b_{fl}^f rises and ρ falls, leaving the sum of the inheritance terms as the only component that can adjust downward to reconcile these patterns.

The downward trend observed in our base results is also evident in the larger sample where only one spouse is required to be linked to their father, addressing concerns about the representativeness of the smaller, double-linked sample. Espín-Sánchez, Ferrie and Vickers (2023) demonstrate that the assortative mating parameter ρ becomes the product of the mobility estimates for sons and daughters ($b^h_f \times b^h_{fl}$) when mothers are assumed to have no influence on their child's status.[13] This assumption implies that the inheritance parameter for mothers is zero ($\beta_m = 0$), an assumption often implicitly made by researchers. Conversely, our base results allow both parents to contribute to status, with potentially different impacts.

Figure 4a shows the assortative mating parameter under this restrictive model applied to the double-linked sample (gray line, labeled "Product"). While substantially smaller than the base results, it exhibits a similar pattern: higher assortative mating in the first two cohorts, followed by a 62 percent decline by the 1900–1920 cohort. The light blue line ("Product – single-linked") plots the same product of mobility estimates for the larger, single-linked sample. This line shows comparable levels and patterns of assortative mating, reinforcing the robustness of the results across samples.

5.1 The Contribution of Assortative Mating to Women's Intergenerational Mobility

Given that women's adult economic status during this period was heavily influenced by their choice of spouse, changes in the strength of assortative mating play a critical role in determining women's intergenerational mobility. In this section, we calculate counterfactual mobility rates under varying levels of assortative mating observed across cohorts from Figure 4a. Specifically, we substitute the value of ρ from cohort A into the structural equations 12 for cohort B and solve for b^h_{fl}, holding the inheritance terms constant. This approach yields a counterfactual estimate of women's intergenerational mobility in cohort B, assuming they faced the strength of assortative mating present in cohort A.

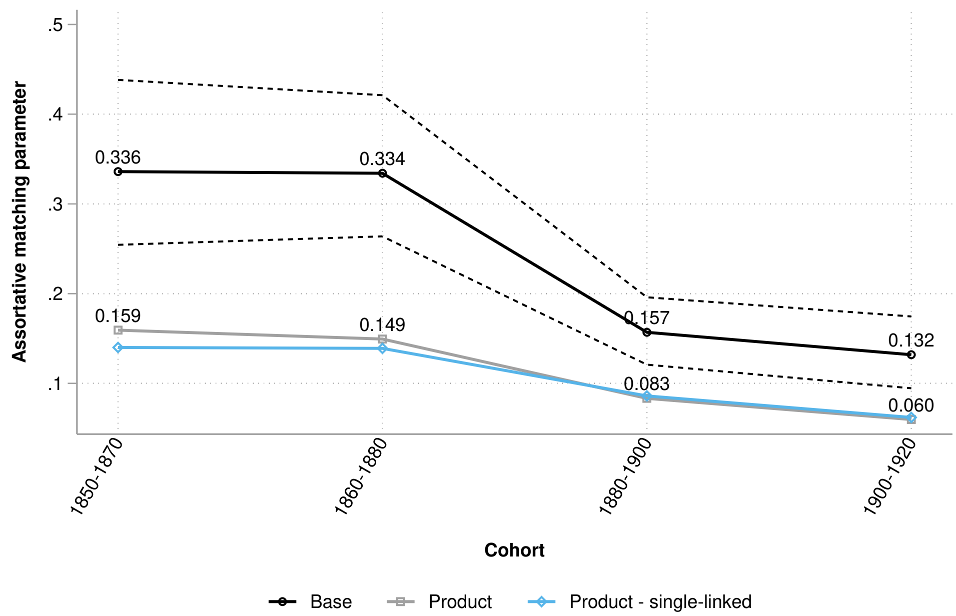
The actual estimates of women's intergenerational mobility are shown in black in Figure 4b. We provide two counterfactual scenarios for women's mobility: one based on the highest level of assortative mating, observed in cohort 1 (1850–1870), plotted in gray, and the other based on the lowest level, observed in cohort 4 (1900–1920), plotted in blue. The results clearly demonstrate that stronger assortative mating leads to higher intergenerational persistence for women. For instance, if the 1880–1900 cohort had faced the high level of assortative mating present in cohort 1, their counterfactual persistence would have been 114 percent higher (0.548 vs. 0.256). Similarly, for the 1900–1920 cohort, counterfactual persistence under cohort 1's assortative mating would have been 154 percent higher (0.592 vs. 0.233). These findings suggest that, without the reductions in assortative mating between the 1860–1880 and 1880–1900 cohorts, not only would women have lost the gains made in intergenerational mobility, but they would have experienced substantial increases in intergenerational persistence.

Conversely, if the strength of assortative mating in the first two cohorts had been as low as that observed in cohort 4 (1900–1920), a woman's father's economic status would have played a much smaller role in determining her own adult economic outcomes. For example, the rank-rank

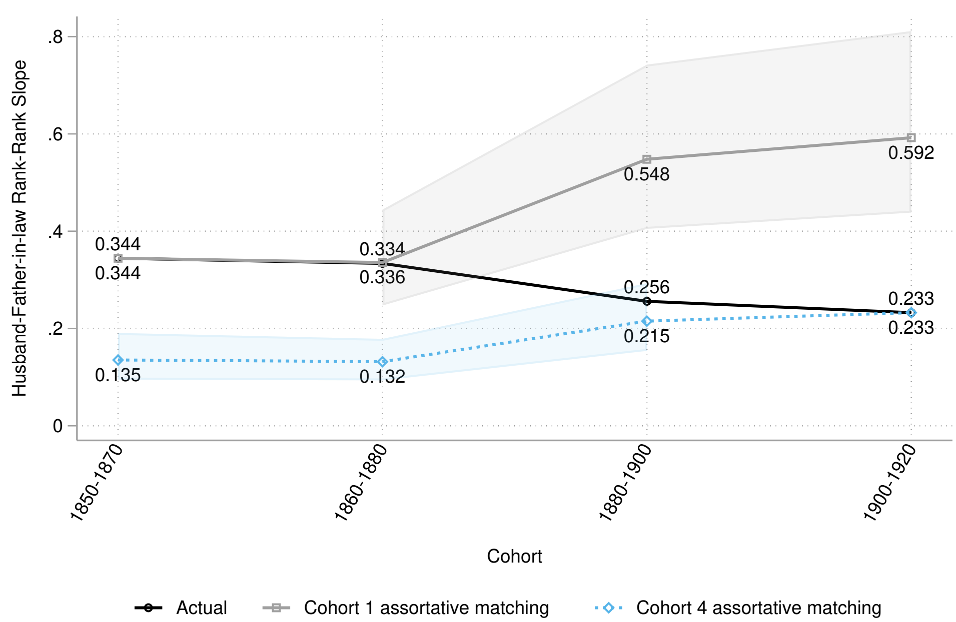
[13] This corresponds to Proposition 2 in Espín-Sánchez, Ferrie and Vickers (2023).

Figure 4: assortative mating Parameter Estimates and Mobility Counterfactual

(a) Structural assortative mating Parameter Estimates (ρ)



(b) Counterfactual Women's Mobility by Strength of assortative mating



Notes: Panel A presents estimates of the structural assortative mating parameter, ρ. The base results (black) are derived from solving the system of equations in Section 2, with 95% confidence intervals shown as dashed lines, based on 1,000 bootstrap iterations. The blue and gray lines represent assortative mating assuming zero maternal inheritance, calculated as the product of mobility terms, with the blue line for samples linking at least one spouse to their father and the gray line for samples linking both spouses to their fathers (confidence intervals omitted for clarity). Panel B shows counterfactual analyses using the base estimates. Actual women's mobility ($\beta_1 + \beta_2$) appears in black, while the gray line shows counterfactual mobility assuming the 1850–1870 cohort's assortative mating level, and the blue line assumes the 1900–1920 level. Shaded areas plot 95% confidence intervals, except where counterfactuals use the cohort's actual ρ value. *Sources:* 1850-1920 Decennial Census data from Ruggles et al. (2024). Marriage certificates from *FamilySearch.org*.

slope between a woman's husband's and her father's status in cohort 1 would have been 61 percent lower (0.135 vs. 0.344). This highlights the significant impact of declining assortative mating on weakening the intergenerational transmission of status for women.

5.2 Why Did Assortative Mating Fall?

A large decline in assortative mating between the 1850–1880 and 1880–1920 marriage cohorts raises a natural follow–up question: *which underlying demographic and behavioral forces drove that aggregate shift?* We answer this in two stages. First, we document differences in assortative mating across key demographic subgroups – parent nativity, internal migration, and urban residence. Second, we adapt a law-of-total-covariance decomposition to our IV setting, isolating each subgroup's contribution to the structural assortative mating parameter so we can see which cohort-level changes matter most for the aggregate decline. Because the drop is concentrated between 1850–1880 and 1880–1920, we pool the four cohorts into these two wider groups to gain power.

We focus on immigration, internal migration, and urbanization as explanatory factors because these demographic forces have been identified as central drivers of both intergenerational mobility and social stratification in the United States between 1850 and 1920. Ferrie (2005) and Long and Ferrie (2013) argue that high levels of internal migration prior to 1880 enabled high intergenerational mobility, as families relocated across regions with large wage and opportunity differentials (Salisbury, 2014; Ward, 2020). Migration rates declined after 1880 coinciding with diminishing locational arbitrage (Kim, 1995, 1998). Although the literature establishes internal migration as an important determinant of mobility, its theoretical effect on assortative mating is ambiguous, so we treat it as an empirical question. Urbanization, too, transformed economic and marriage markets. As Americans moved from rural areas into cities, the transition out of farming initially facilitated upward mobility (Song et al., 2020), but by the early 20th century, most of that transition had already occurred. Within urban labor markets, however, opportunities became more stratified, and recent research shows urbanization had a "first-order effect" on marriage market sorting (Olivetti et al., forthcoming). Immigration may also have reshaped the mobility landscape. Between 1850 and 1913, over 30 million immigrants entered the U.S., many starting in low-wage jobs. The children of these immigrants might move up rapidly, however, some work finds that second-generation mobility was limited (Abramitzky, Boustan and Eriksson, 2014; Ward, 2020). Olivetti and Paserman (2015) report a negative relationship between immigration rates and overall mobility. Our own results, by contrast, suggest that second-generation immigrants were more mobile than the children of native-born parents within our sample (See Appendix Figure A2), making immigration a particularly relevant dimension for understanding cohort-level changes in assortative mating.

5.2.1 Assortative Mating By Demographic Subgroup

First, we document how assortative mating evolved during the late nineteenth century across three demographic dimensions: parent nativity, internal migration, and rural-versus-urban childhood residence. Figure 5 plots estimates of ρ for each group, calculated from the wife's vantage point (e.g., the "native-born parents" line restricts the sample to wives with native-born parents, regardless of the husband's origins). Replicating the exercise from the husband's perspective yields virtually identical patterns (see Appendix Figure A3).

The nativity split shows the starkest contrast: assortative mating is initially much stronger for daughters of native-born parents than for daughters of immigrants, but the gap closes by 1880–1920 as native-born sorting falls and immigrant sorting rises sharply. Meanwhile, internal migrants versus stayers and rural versus urban children display no meaningful cross-group differences. All start with high assortative mating that declines through 1880–1920. Although the aggregate assortative mating parameter is not a simple weighted average of subgroup slopes, the parent nativity-specific results suggest that differential changes for these subgroups across cohorts may play a meaningful role in explaining the overall decline. In contrast, the internal migration and urbanization patterns closely follow the aggregate trend in Figure 4a, suggesting that they may offer less explanatory power.

5.2.2 Decomposition

We formalize each subgroup's contribution to the decline by decomposing the aggregate structural assortative mating parameter

$$\rho = \frac{\widehat{b^h_{fl}}^{\,2}}{\widehat{b^f_{fl}}}, \tag{13}$$

where $\widehat{b^h_{fl}}$ is the husband–father IV slope (rank of husband on rank of wife's father) and $\widehat{b^f_{fl}}$ is the father–father IV slope (rank of husband's father on rank of wife's father). Throughout, subscripts h and f refer to husband and husband's father, respectively; fl indicates the wife's father.

Two features distinguish our decomposition from existing OLS-based approaches (e.g.(Hertz, 2008; Jácome, Kuziemko and Naidu, 2025)). First, we decompose aggregate IV estimates instead of OLS as in the previous literature. Because our slopes come from a two–stage least-squares design, the law-of-total-covariance expansion introduces first-stage covariance terms that do not appear under the OLS weighting. The second difference stems from the non-linearity of ρ. Equation 13 is a non-linear function of moments; consequently ρ cannot be expressed as a simple weighted average of subgroup-specific ρ's. We therefore decompose the numerator and denominator separately and then reassemble them.

We begin by expanding the IV estimator for the husband–father slope. The decomposition for the father-father slope proceeds similarly, except with appropriate subscripts. For brevity, we show the two-group case ($A, B \in G$). Let $g \in G$ index mutually exclusive demographic groups (e.g. immigrant parents vs. native-born parents). Denote by p_g a group's population share, by $\beta^h_{fl,g}$ and $\beta^f_{fl,g}$ its within-group IV slopes, by $\sigma^g_{y_f Z}$ the group-specific first-stage covariance, and by $\mu^g_{y_h}$, $\mu_{y^g_{fl}}$, and μ^g_Z the relevant mean ranks. The decomposition of the aggregate husband-father IV slope can be written as:

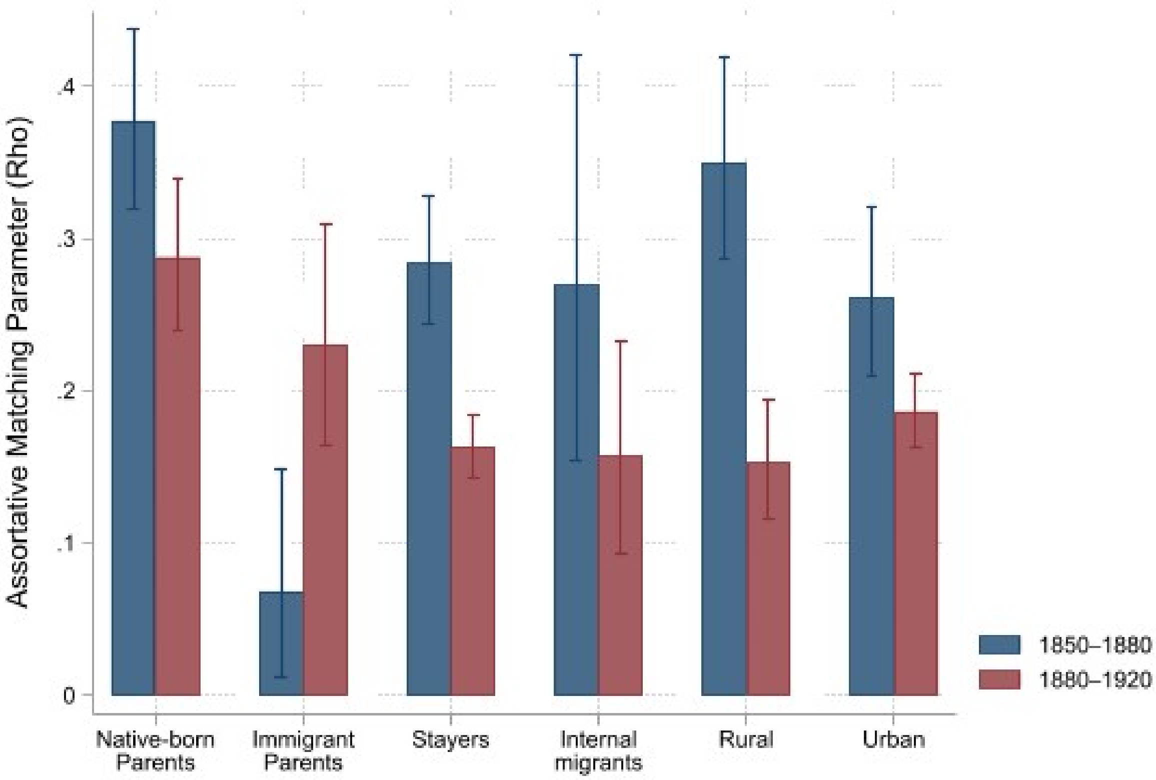


Figure 5: Estimated assortative mating by demographic group

Notes: This figure plots the estimated assortative mating parameter (ρ) by demographic subgroup, calculated from the wife's perspective. Each bar represents the cohort-specific ρ for marriages formed by women in the indicated group: daughters of native-born or immigrant parents, daughters who remained in Massachusetts (stayers) or moved elsewhere within U.S. (internal migrants), and daughters raised in rural or urban towns. Blue bars correspond to the 1850–1880 cohorts and red bars to the 1880–1920 cohorts; vertical lines denote 95% confidence intervals. Estimates are obtained from the structural model described in Section 2, using the same specification as the aggregate results in Figure 4a.

$$\beta^{\mathrm{IV}} = b^h_{fl} = \frac{\mathrm{Cov}(y^h, Z)}{\mathrm{Cov}(y^{fl}, Z)} = \frac{p_A\, \beta^h_{fl,A}\, \sigma^A_{y^{fl}Z} \;+\; p_B\, \beta^h_{fl,B}\, \sigma^B_{y^{fl}Z} \;+\; p_A p_B\, (\mu^A_{y^h} - \mu^B_{y^h})(\mu^A_Z - \mu^B_Z)}{p_A\, \sigma^A_{y^{fl}Z} \;+\; p_B\, \sigma^B_{y^{fl}Z} \;+\; p_A p_B\, (\mu^A_{y^{fl}} - \mu^B_{y^{fl}})(\mu^A_Z - \mu^B_Z)}, \quad (14)$$

The literature on intergenerational mobility uses expressions like the one above to decompose aggregate changes in OLS estimates into the underlying shifts in population composition, group-specific mobility, and the convergence of economic status across groups (see, for example, Jácome, Kuziemko and Naidu (2025) for a decomposition by race and gender for the United States in the twentieth century). Following this, we express the change in our IV estimate of women's mobility, (b^h_{fl}), as the combined effect of the three factors mentioned above in the OLS decomposition, with an additional factor that enters because our slopes are estimated by IV rather than OLS - $\sigma_{y^{fl}Z}$, which captures the strength of the instrument in the first-stage.

Our ultimate aim, however, is to explain changes in the assortative mating parameter ρ. We apply the same logic to both IV slopes that define ρ: $\widehat{b^h_{fl}}$ in the numerator and $\widehat{b^f_{fl}}$ in the denominator. Expanding each slope with the four factors above and then substituting them back into ρ yields a complete decomposition that shows exactly how population shares, within-group slopes, instrument strength, and status gaps combine to drive the aggregate decline. The decomposition also provides a natural set of counterfactuals: holding one factor at its early-cohort level while allowing the others to evolve to their late-cohort values quantifies its contribution to the overall decline. The full decomposition of the aggregate ρ can be written as:

$$\rho = \frac{\widehat{b^h_{fl}}^{\,2}}{\widehat{b^f_{fl}}} = \frac{\left(\dfrac{p_A\, \beta^h_{fl,A}\, \sigma^A_{y^{fl}Z} + p_B\, \beta^h_{fl,B}\, \sigma^B_{y^{fl}Z} + p_A p_B\, (\mu^A_{y^h} - \mu^B_{y^h})(\mu^A_Z - \mu^B_Z)}{p_A\, \sigma^A_{y^{fl}Z} + p_B\, \sigma^B_{y^{fl}Z} + p_A p_B\, (\mu^A_{y^{fl}} - \mu^B_{y^{fl}})(\mu^A_Z - \mu^B_Z)}\right)^2}{\left(\dfrac{p_A\, \beta^f_{fl,A}\, \sigma^A_{y^{fl}Z} + p_B\, \beta^f_{fl,B}\, \sigma^B_{y^{fl}Z} + p_A p_B\, (\mu^A_{y^f} - \mu^B_{y^f})(\mu^A_Z - \mu^B_Z)}{p_A\, \sigma^A_{y^{fl}Z} + p_B\, \sigma^B_{y^{fl}Z} + p_A p_B\, (\mu^A_{y^{fl}} - \mu^B_{y^{fl}})(\mu^A_Z - \mu^B_Z)}\right)} \quad (15)$$

Non-linearity of ρ complicates the interpretation of counterfactuals relative to a single-slope setting, because each factor affects both the numerator and denominator of ρ. Population shares illustrate the idea: we replace p_B and p_A in the late-cohort expression with the early-cohort levels (7% immigrant parents vs 30%) to gauge the role of changing composition. This higher share of daughters with immigrant parents simultaneously changes the association between a husband and his wife's father and the association between the two fathers. Thus, when we hold a factor at its 1850–1880 value in the 1880–1920 formula, we substitute it consistently in both parts of the decomposition. Interpretation of counterfactuals on the group slopes requires even more care, and is less intuitive. As shown in Section 2, the empirical slopes (b^h_{fl} and b^f_{fl}) are linked through a structural model, so we cannot hold one constant while letting the other evolve as observed; we instead hold the entire covariance structure at its 1850–1880 configuration.

Figure 6 summarizes the results of our counterfactual exercises. The first two bars show the actual assortative mating parameters for 1850–1880 and 1880–1920, with standard error bars to mark the 5th and 95th percentiles of the bootstrap distribution (1,000 reps). Each subsequent

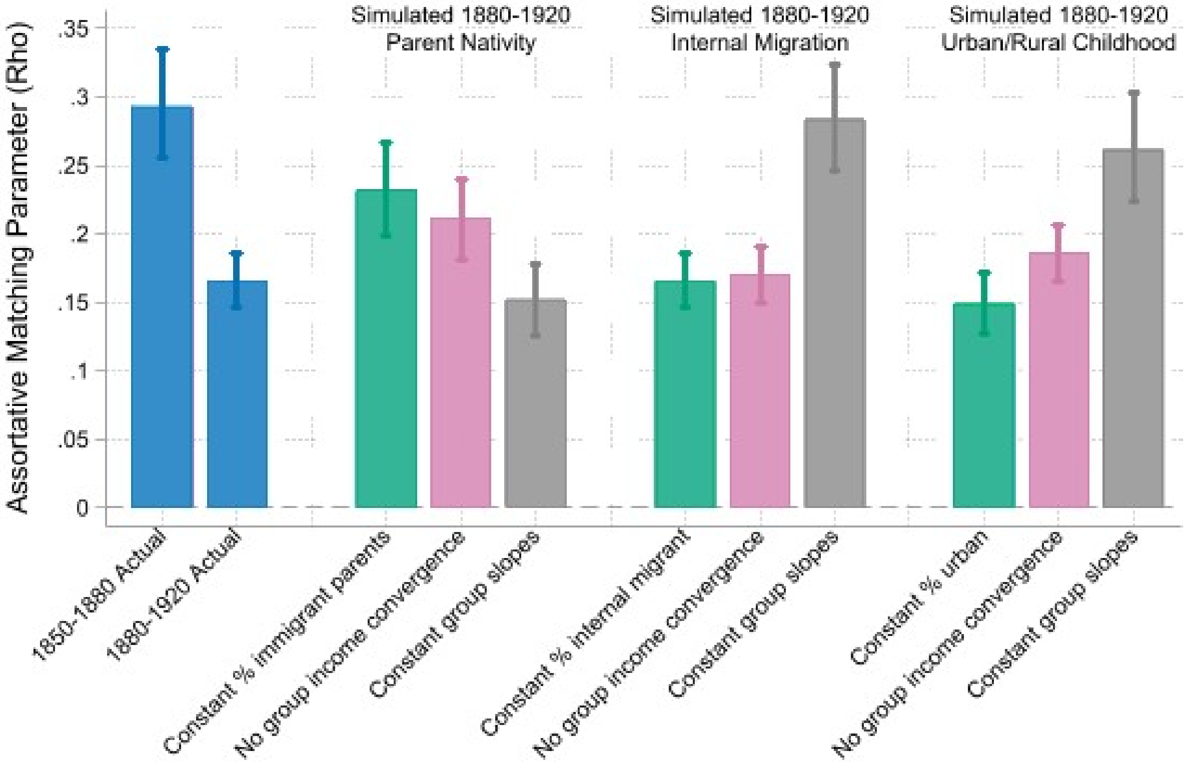


Figure 6: Subgroup contribution to decline in assortative mating

Notes: The figure summarizes counterfactual decompositions of the assortative mating parameter $\rho = \widehat{b^h_{fl}}^2 / \widehat{b^f_{fl}}$ across the 1850–1880 and 1880–1920 marriage cohorts. The first two bars plot the observed ρ for each cohort, with 5th and 95th percentile bootstrap confidence intervals (1,000 reps). Subsequent bars show simulated 1880–1920 values obtained by holding one factor in the IV decomposition fixed at its 1850–1880 level: subgroup population shares (green), mean intergenerational rank gaps (pink), and the full within-group covariance structure—including both IV slopes and first-stage covariance terms (gray). All remaining components are allowed to evolve to their observed 1880–1920 values. Counterfactuals are shown for subgroups defined by parent nativity, internal migration status, and rural versus urban childhood residence. Bars lying closer to the 1850–1880 benchmark indicate a larger explanatory role for that factor in the aggregate decline in assortative mating.

bar holds one factor at its early-cohort value while letting others evolve to their true 1880-1920 values. The nearer a simulated 1880–1920 bar sits to the 1850–1880 benchmark, the more that factor potentially explains of the decline.

Our results suggest that immigration-related factors account for the largest share of the decline in assortative mating between cohorts. The first set of counterfactuals decomposes the aggregate assortative mating parameter by parent nativity. Holding the share of immigrant parents constant (green bar) explains 52% of the decline. The next largest contributor is the wealth score gap counterfactual, which accounts for 36% of the decline (pink bar).[14] Holding subgroup slopes constant (gray bar) yields a counterfactual assortative mating level for the 1880–1920 cohort that is only

[14]In this exercise, we hold constant the intergenerational rank gaps across all parent types: $(\mu^A_{y^h} - \mu^B_{y^h}),: (\mu^A_{yf} - \mu^B_{yf}),: (\mu^A_{yfl} - \mu^B_{yfl}),:$ and $: (\mu^A_{yZ} - \mu^B_{yZ})$.

slightly below the observed value. This suggests that changes in the slope terms contributed little to the overall decline.

The remaining two sets of counterfactual results indicate that urbanization had a minimal effect on the decline in assortative mating, while internal migration had none. Holding constant the share urban and the share internal migrant (green bars) produces negligible changes in the simulated 1880–1920 value. Fixing the mean status gaps across groups (pink bars) explains 16% of the decline for urban-rural residence, and 3% of the decline for migrants versus stayers. Holding subgroup slopes constant fully explains the aggregate decline (gray bars). For example, when both migrant and stayer slopes are fixed at 1850–1880 levels, the simulated bar matches the early-cohort value. However, this offers little insight into the underlying mechanism causing the decline in aggregate. The slope counterfactuals, while mechanically large, do not imply that internal migration or urbanization are distinctive drivers of the decline in ρ. As Figure 5 shows, both groups began with similar assortative mating levels and declined in tandem, tracking the aggregate trend. In short, the slope counterfactuals for internal migration and urban residence confirm that the forces driving the aggregate decline also operated within each group, without revealing any group-specific mechanisms.

The subgroup patterns above (by parent nativity, migration, and urban residence) suggest both compositional forces and changing choice sets; next, we speculate on plausible mechanisms and how they map into our framework. In mid-century New England, marriages were strongly constrained by parental networks, locality, and women's legal dependence, yielding matches concentrated within narrow social circles; as women's legal and economic autonomy expanded under Married Women's Property Acts and related reforms, and as urbanization thickened marriage markets, parental background became less decisive in spouse choice. In terms of our structure, this shift from parent-brokered to couple-selected matches reduces sorting on parental status, lowering assortative mating—even if father–father-in-law correlations move differently.

Classic assortative-mating models predict stronger positive sorting when complementarities are high, but when market integration relaxes search constraints and raises the feasibility of "out-of-network" matches, sorting on parental background can weaken even if sorting on other traits (e.g., tastes, personality) strengthens (Kalmijn, 1994). Thicker urban marriage markets reduce reliance on parental networks. Framed this way, industrial-era urbanization offers a possible explanation for declines in assortative mating despite stable or rising father–father correlations driven by inheritance. Evidence on marital distance during industrialization supports this mechanism. Marital distance is the shortest genealogical path connecting spouses through pre-existing family and in-law ties prior to their marriage. Ghosh, Hwang and Squires (2025) finds that the sharpest rise in distance occurred for the 1870–1915 birth cohorts, with roughly three-quarters of the effect operating through migration and changes in local marriage pools and about one-quarter through preference shifts within a location, i.e., couples crossing prior social boundaries more often.

6 External Validity of Marriage-market Mobility Estimates

Our analysis shows that changing patterns of assortative mating within the Massachusetts marriage market contributed to early gains in women's economic mobility. Because these results are based on linked marriage records from Massachusetts, their generalizability depends on the historical and geographic scope of the data. In this section, we assess the extent to which our findings apply more broadly by comparing our estimates to those from other sources and examining how differences in geography, linkage methods, and marriage behavior influence the observed relationship between marriage and mobility. We focus on the father–child status correlations, which are more commonly reported in the literature. Appendix Table A8 summarizes related studies on women's intergenerational mobility and their methodological choices along these dimensions.

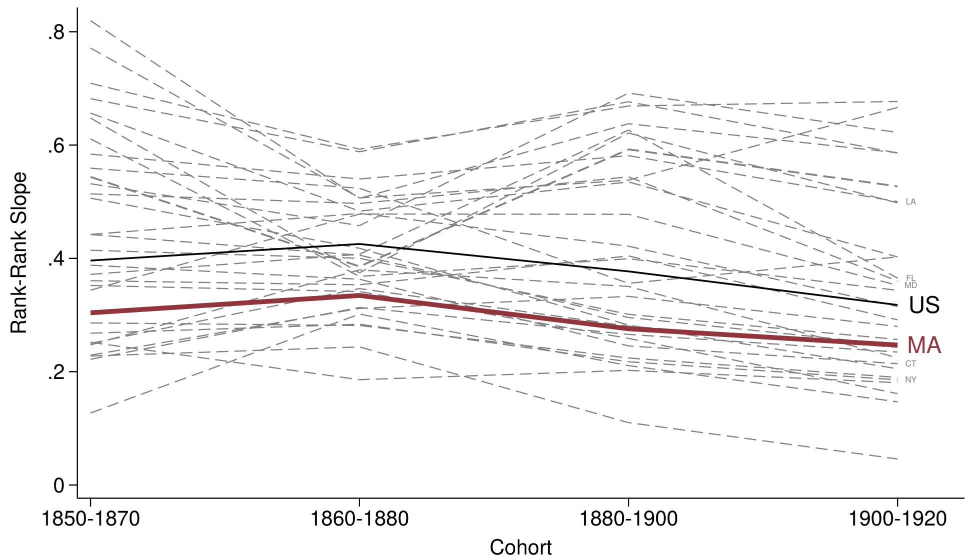
Regional variation in marriage timing, singlehood, and occupational structure can influence both the opportunity for assortative mating and the measured degree of intergenerational persistence. Mobility patterns vary across geographic subunits and differ from national-level estimates. Massachusetts in the nineteenth and early twentieth centuries was relatively mobile. Additionally, large regional differences in average occupational wealth scores—higher in the Northeast—can lead to higher persistence estimates at the national level compared to individual states.[15] Together, these factors explain why our estimates of persistence in Massachusetts are lower than the national estimates for men reported by Ward (2023), though both show a similar decline in persistence over the late 19th and early 20th centuries.

To show that Massachusetts-born men experienced different levels and changes in mobility compared to the rest of the nation, we use a standard record-linkage methodology and our constructed wealth scores. Figure 7 presents 2SLS rank-rank slopes by state of birth and for the nation as a whole, based on linked native-born white men from the Census Linking Project (Abramitzky et al., 2020). This analysis includes both married and single men. Persistence estimates for Massachusetts-born men rank in the middle to lower third among states in each cohort, though their mobility trends over time closely align with national patterns.

While alternative linkage methods provide valuable benchmarks, our marriage-based links directly capture the channel through which assortative mating operates, making them particularly relevant for understanding how marriage influenced women's economic outcomes. The choice of linking methodology can influence mobility estimates by introducing different biases into the record linkage process. Our use of marriage certificates to link married women across generations yields a representative sample of married couples, with match rates and false positive rates comparable to other methods. For example, the LIFE-M project uses a combination of birth, death, and marriage records from Ohio to create intergenerational links for women born in the late 19th and early 20th centuries (Bailey et al., 2022). Bailey and Lin (2024) use the LIFE-M data to estimate women's

[15] The national mobility estimate can exceed those for most individual states. The full population parameter is a weighted average of the individual subgroup slope parameters where the weights are the share of the population in the group *and* the ratio of the subgroup variance in incomes to the variance of incomes in the population, *plus* a term that captures level differences in income between the subgroups (Hertz, 2008; Jácome, Kuziemko and Naidu, 2025).

Figure 7: Intergenerational Mobility of Massachusetts-born Men Relative to the Nation



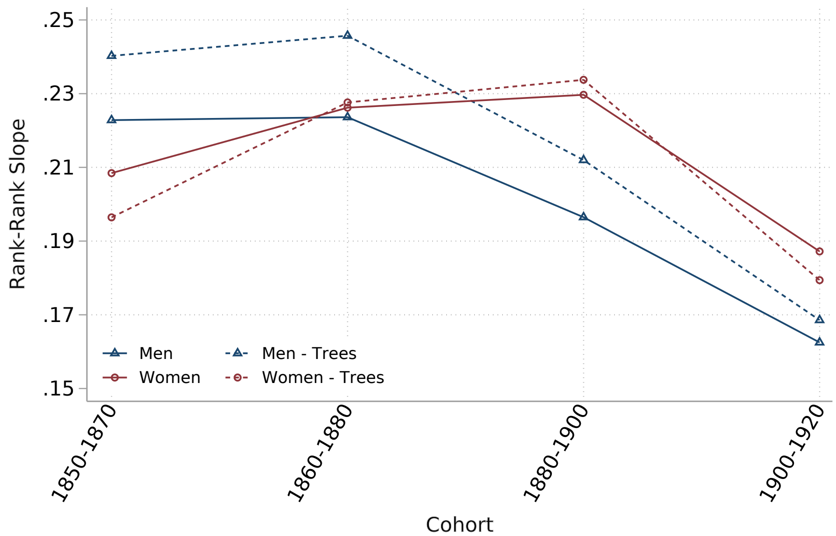
Notes: This figure plots rank-rank slopes of son's wealth score rank on father's wealth score rank for Massachusetts-born men in red and for the entire US in solid black. Gray lines plot the rank-rank slope for men born in other states. The father's occupational score is instrumented using a second observation of his economic status, and include controls for a quartic in the father's age and a quartic in the adult child's age, both measured at the time their economic status is observed. Observations are weighted to adjust for differential selection into the linked sample. Record links were created using the Census Linking Project and the conservative ABE-method with exact name matching.
Sources: Abramitzky et al. (2020), Buckles et al. (2023), and Ruggles et al. (2024).

intergenerational mobility in Ohio and report rank-rank relative mobility levels similar to ours for overlapping cohorts, along with a comparable decline in persistence across later birth cohorts. Specifically, the rank-rank slope for women born in 1870 was 0.35, decreasing to 0.29 for the 1890 birth cohort.

The only nationally representative dataset with intergenerational links for women is Buckles et al. (2023), which uses genealogist-constructed family trees. However, this approach may be subject to selection bias, as it relies on observations with descendants likely to engage in genealogy. A key advantage of our study is that it begins with a comprehensive population of men and women married in Massachusetts, avoiding this potential bias. The similarity of our estimates to those derived from the family tree data in Buckles et al. (2023) helps alleviate concerns about selection bias or the representativeness of Family Tree dataset.

When restricted to comparable samples, mobility levels and trends are largely consistent between genealogist-linked data and our double-match procedure using marriage registers. To ensure comparability, we limit the data to individuals born in Massachusetts, residing in Massachusetts as adults, and recorded as married in the census. We apply our main wealth score ranks to both generations across all cohorts. Figure 8 shows rank-rank slopes for men and women in our linked dataset and the family tree dataset under these restrictions. Both samples reveal an inverted U-shaped pattern of persistence for women, with minimal differences in levels. For men, the family tree data indicates slightly less mobility than our linked sample in every cohort, but the overall pattern is similar: no change in mobility from the first to second cohorts, followed by a steady decline in persistence.

Figure 8: Rank-rank slopes of linked data versus family tree data



Notes: This figure plots rank-rank parameters of son's (husband's) wealth score rank on father's (father-in-law's) wealth score rank men in blue and women in red. Solid lines show mobility estimates from linked data using marriage certificates to create intergenerational links, whereas dashed lines show mobility estimates using genealogical family tree data from Buckles et al. (2023). Samples are limited to observations born in Massachusetts, residing in Massachusetts as adult, and married. Economic status for father and son is measured using the wealth score ranks discussed in the Data section. *Sources:* 1850-1920 Decennial Census data from Ruggles et al. (2024). Marriage certificates from *FamilySearch.org*. Family tree links from various years of the Census Tree Project Buckles et al. (2023).

Because our main analysis is restricted to married couples, understanding who married and how marriage correlated with socioeconomic status is crucial for interpreting our results on assortative mating and mobility. The decision to marry, and the timing of that decision, carried significant economic implications for women. Men and women in the United States followed the "European Marriage Pattern" of late marriage and high rates of singlehood as early as the 1850s (Hajnal, 1965; Fitch and Ruggles, 2000). Nationally, marriage rates by age 50 were high by modern standards, ranging from 87% to 90% for women and 85% to 92% for men. However, Massachusetts stood out with significantly lower marriage rates compared to the national average, and women in the state were notably more likely to remain single than men. Massachusetts consistently ranked first or second among states for the highest singlehood rates (Appendix Figure A4). While the nation saw a gradual decline in marriage rates, the decrease for Massachusetts-born women was three to four times larger. Furthermore, a sharp drop in marriage rates for Massachusetts cohorts began in the mid-1890s, decades earlier than the national trend, which started in the 1930s (Haines, 1996).

Marriage patterns varied by social class and gender, with differing links to socioeconomic status (SES). In the 19th century, marriage rates for women were negatively correlated with SES, while rates for men increased with SES (Olivetti et al., forthcoming). Women with greater resources outside of marriage, hence more options, were less likely to marry, whereas men with more resources were better able to marry and establish households. In Western societies, including the U.S., men were often required to meet a resource threshold to marry and form a household (Ruggles, 2016).[16]

To provide context for how marriage-market participation affects measured persistence, we compare our main findings for married couples to those for single men and women. This comparison offers insights for researchers interested in understanding the implications of excluding singles, while our primary focus remains on the role marriage plays in the economic mobility of women. Our linked data based on a sample of registered marriages does not allow us to formally model the decision to marry. However, we can estimate an intergenerational rank-rank slope for women who never married. These women retained their surname from childhood, allowing us to create intergenerational links using the ABE conservative exact-name method to estimate their mobility.

Interpreting the occupational status correlation between an adult daughter and her father presents challenges for two key reasons. First, while single women's labor force participation (LFP) was substantially higher than that of married white women, it remained much lower than men's. In our sample of single white Massachusetts-born women, LFP increased from 40% in the 1850–1870 cohort to 82% in the 1900–1920 cohort. Even among single women, many are excluded from intergenerational mobility estimates due to nonparticipation in the labor force or not having a recorded occupation in the census.

Second, a single woman's decision to work was strongly influenced by her father's economic status. Daughters of higher-ranked men were significantly less likely to participate in the labor force. Panel A of Table 2 shows that each one-unit increase in a father's wealth score rank was associated with a 0.5 to 0.6 percentage point lower likelihood of LFP across the first three cohorts—a substantial economic effect. For example, moving a father from the 25th to the 75th percentile of the sample wealth rank distribution corresponded to a reduction in LFP of 40% to 60% of the mean LFP. However, as labor force participation surpassed 80% by the 1900–1920 cohort, the impact of a father's rank on LFP diminished to half of what was observed in earlier cohorts.

Mobility estimates for single women differ significantly from those for married women in our sample, because "mobility" is defined differently for these two groups. For single women, mobility is based on their own occupational status, whereas for married women, it is tied to their husband's occupational status. Panel B of Table 2 reports very low persistence for single women in the first two cohorts (0.099 and 0.153), with somewhat higher persistence in the later cohorts (0.23), though still lower than the persistence observed for married women (0.30). During this period,

[16] Farmers and proprietors (self-employed) had higher marriage rates than wage workers, with high-wage workers marrying more often than low-wage workers (Ruggles, 2016). Farmers tended to marry later, waiting to inherit or acquire a farm, as wives' unpaid labor was essential for its operation. Similarly, self-employed proprietors relied on their wives' labor to run shops. Oppenheimer (1988) explains variation in marriage timing through a search-theoretic model linked to men's economic opportunities.

women faced limited opportunities in the labor market, and their employment was concentrated in far fewer occupations compared to men. This occupational segregation likely contributed to the lower correlation between daughters' and fathers' occupational statuses relative to men's mobility estimates.

Single men born in Massachusetts exhibit mobility levels that are relatively similar to their married counterparts. To estimate this, we use links created by the Census Linking Project, restricting the sample to never-married white men born in Massachusetts, aged 20–40 in the adult census. Panel C of Table 2 indicates that while single men did not follow the same mobility trends as married men, their levels of persistence were comparable. For example, single men in the first cohort had a rank-rank slope with their fathers of 0.28, compared to 0.38 for married men. However, as persistence declined for married men over time, single men exhibited slightly higher persistence in the later cohorts, with a slope of 0.24 versus 0.21 for married men in the final cohort. Estimating the full population mobility parameter would be a complicated weighted average of the married and single samples.[17]

Taken together, these comparisons indicate that the patterns we document for Massachusetts marriages, declining assortative mating and rising women's mobility, are broadly consistent with other historical evidence. While variation in geography, linkage strategy, and marital composition affect estimated levels of persistence, the direction and timing of changes align closely across datasets. These findings suggest that shifts in marriage-market sorting observed in Massachusetts reflect a broader process of changing marital behavior and economic opportunity in the early twentieth-century United States.

7 Conclusion

This study presents new evidence for women's intergenerational economic mobility in the 19th and early 20th century United States, highlighting the role of marital sorting in shaping these patterns. Using a novel dataset constructed from Massachusetts marriage registers linked to census records, we find that men and women initially had similar levels of mobility in the 1850-70 cohort. As men's mobility improved quicker than women's, a gender gap opened by the early 20th century. Our analysis reveals that shifts in assortative mating were critical to these trends. Initially, strong assortative mating constrained women's mobility, but as assortative mating declined, women experienced modest gains in mobility.

While this study provides valuable insights into women's intergenerational mobility, several limitations warrant consideration. First, our analysis focuses on Massachusetts, a state with unique economic and social characteristics, which may limit the generalizability of our findings to other regions. However, Massachusetts' early industrialization and rich historical records provide a compelling context for studying mobility during this transformative period. Second, our use of occupa-

[17] The full population parameter is a weighted average of subgroup slope parameters, weighted by population shares, the ratio of subgroup income variance to overall income variance, and a term capturing income-level differences across subgroups (Hertz, 2008; Jácome, Kuziemko and Naidu, 2025).

Table 2: Mobility for Single Men and Women

	1850-1870	1860-1880	1880-1900	1900-1920
	Panel A: Single women's labor force participation			
Father Rank	-0.0054	-0.0057	-0.0056	-0.0037
	(0.0003)	(0.0003)	(0.0003)	(0.0002)
Observations	13,599	18,239	30,778	37,654
Mean LFP Rate	0.399	0.494	0.597	0.816
$\beta \times$ std. dev. Dad's Rank	-0.15	-0.16	-0.15	-0.09
$\beta \times 75^{th} - 25^{th}$ pctile of Dad's Rank	-0.26	-0.26	-0.26	-0.12
	Panel B: Rank-rank mobility of single women			
Father Rank	0.099	0.153	0.231	0.225
	(0.018)	(0.014)	(0.015)	(0.011)
Observations	4,964	8,321	17,309	28,842
	Panel C: Rank-rank mobility of single men			
Father Rank	0.276	0.314	0.276	0.238
	(0.018)	(0.018)	(0.014)	(0.012)
Observations	7,596	8,846	14,655	19,756

Notes: This table uses samples of never married 20-40 year old Massachusetts-born men and women linked to their father in a census 20 years prior. Panel A regresses an indicator for labor force participation, and panel B regresses the adult woman's own occupational score rank, on father's occupational score rank. In all regressions, the father's occupational score is instrumented using a second observation of his economic status taken from the Census Tree Project links, and include controls for a quartic in the father's age and a quartic in the adult child's age, both measured at the time their economic status is observed. Observations are weighted to adjust for differential selection into the linked sample. Heteroskedasticity-robust standard errors reported in parentheses. Panels A and B uses links created with the ABE conservative exact name matching procedure (Abramitzky, Boustan and Eriksson, 2012). Panel C uses links created by the Census Linking Project, again using the ABE conservative exact name method (Abramitzky et al., 2020).
Sources: Census Linking Project Links from Abramitzky et al. (2020). 1850-1920 Decennial Census data from Ruggles et al. (2024). Census Tree Project links from (Price et al., 2021; Buckles et al., 2025)

tional wealth scores as proxies for economic status do not fully capture women's contributions to household resources, particularly non-market work. These considerations suggest opportunities for future research to expand and deepen our understanding of these dynamics in broader contexts.

Understanding the link between changes in assortative mating and broader economic or social transformations is a key avenue for future research. A national dataset could offer the geographic variation needed to isolate potential drivers of these changes, such as industrialization, urbanization, and expanded educational opportunities. Examining the interaction between local labor markets and marriage decisions—especially the influence of economic shocks or industrial growth on assortative mating—presents another promising area of study. Additionally, investigating these patterns across demographic subgroups, such as immigrant versus native-born populations or different ethnic groups, could reveal how social and economic forces uniquely affected mobility. Finally, analyzing the role of social networks and geographic proximity in marriage decisions could shed light on how local factors influenced the persistence or reduction of economic inequality across generations. These approaches highlight the potential for a deeper and more nuanced understanding of intergenerational mobility and marital sorting on a national scale.

8 Bibliography

Aaronson, Daniel, and Bhashkar Mazumder. 2008. "Intergenerational Economic Mobility in the United States, 1940 to 2000." *Journal of Human Resources*, 43(1): 139–172. Publisher: University of Wisconsin Press Section: Articles.

Abramitzky, Ran, Leah Boustan, Elisa Jacome, and Santiago Perez. 2021*a*. "Intergenerational Mobility of Immigrants in the United States over Two Centuries." *American Economic Review*, 111(2): 580–608.

Abramitzky, Ran, Leah Boustan, Katherine Eriksson, James Feigenbaum, and Santiago Pérez. 2021*b*. "Automated Linking of Historical Data." *Journal of Economic Literature*, 59(3): 865–918.

Abramitzky, Ran, Leah Boustan, Katherine Eriksson, Santiago Pérez, and Myera Rashid. 2020. "Census Linking Project: Version 2.0 [dataset]." *https://censuslinkingproject.org.*

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2012. "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration." *American Economic Review*, 102(5): 1832–56.

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2014. "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration." *Journal of Political Economy*, 122(3): 467–506.

Abramitzky, Ran, Leah Platt Boustan, Harriet M. Brookes Gray, Katherine Eriksson, Santiago Pérez, Hannah M. Postel, Myera Rashid, and Noah Simon. 2025. "Finding John Smith: Using Extra Information for Historical Record Linkage." National Bureau of Economic Research Working Paper 33999.

Ager, Philipp, Leah Boustan, and Katherine Eriksson. 2021. "The Intergenerational Effects of a Large Wealth Shock: White Southerners after the Civil War." *American Economic Review*, 111(11): 3767–94.

Althoff, Lukas, Harriet Brookes Gray, and Hugo Reichardt. 2024. "The Missing Link: Women and Intergenerational Mobility." Working Paper.

Bailey, Martha J., and Peter Z. Lin. 2024. "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.* University of Chicago Press.

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

Bailey, Martha J., Peter Z. Lin, A.R. Shariq Mohammed, Paul Mohnen, Jared Murray, Mengying Zhang, and Alexa Prettyman. 2022. "LIFE-M: The Longitudinal, Intergenerational Family Electronic Micro-Database." Distributed on 2022-12-21.

Buckles, Kasey, Adrian Haws, Joseph Price, and Haley E.B. Wilbert. 2025. "Breakthroughs in historical record linking using genealogy data: The Census Tree project." *Explorations in Economic History*, 98: 101717.

Buckles, Kasey, Joseph Price, Zachary Ward, and Haley E.B. Wilbert. 2023. "Family Trees and Falling Apples: Historical Intergenerational Mobility Estimates for Women and Men." National Bureau of Economic Research Working Paper 31918.

Charles, Kerwin, Erik Hurst, and Alexandra Killewald. 2013. "Marital Sorting and Parental Wealth." *Demography*, 50(1): 51–70.

Chetty, Raj, David Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, and Jimmy Narang. 2017. "The fading American dream: Trends in absolute income mobility since 1940." *Science*, 356(6336): 398–406.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014*a*. "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." *Quarterly Journal of Economics*, 129(4): 1553–623.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. 2014*b*. "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility." *American Economic Review*, 104(5): 141–147.

Collins, William J., and Ariell Zimran. 2019. "The economic assimilation of Irish Famine migrants to the United States." *Explorations in Economic History*, 74: 101302.

Collins, William J., and Ariell Zimran. 2023. "Working Their Way Up? US Immigrants' Changing Labor Market Assimilation in the Age of Mass Migration." *American Economic Journal: Applied Economics*, 15(3): 238–269.

Collins, William J., and Marianne H. Wanamaker. 2014. "Selection and Economic Gains in the Great Migration of African Americans: New Evidence from Linked Census Data." *American Economic Journal: Applied Economics*, 6(1): 220–252.

Collins, William J., and Marianne H. Wanamaker. 2015. "The Great Migration in Black and White: New Evidence on the Selection and Sorting of Southern Migrants." *The Journal of Economic History*, 75(04): 947–992.

Collins, William J., and Marianne H. Wanamaker. 2022. "African American Intergenerational Economic Mobility since 1880." *American Economic Journal: Applied Economics*, 14(3): 84–117.

Corak, Miles. 2020. "The Canadian Geography of Intergenerational Income Mobility." *The Economic Journal*, 130(631): 2134–2174.

Corak, Miles, and Andrew Heisz. 1999. "The Intergenerational Earnings and Income Mobility of Canadian Men: Evidence from Longitudinal Income Tax Data." *Journal of Human Resources*, 34(3): 504–533.

Dahl, Molly, and Thomas DeLeire. 2008. "The association between children's earnings and fathers' lifetime earnings: estimates using administrative data." Institute for Research on Poverty Discussion Paper No. 1342-08. http://irp.wisc.edu/publications/dps/pdfs/dp134208.pdf.

Davis, Jonathan M.V., and Bhashkar Mazumder. forthcoming. "The Decline in Intergenerational Mobility after 1980." *The Review of Economics and Statistics*, 1–47.

Deutscher, Nathan, and Bhashkar Mazumder. 2023. "Measuring Intergenerational Income Mobility: A Synthesis of Approaches." *Journal of Economic Literature*, 61(3): 988–1036.

Espín-Sánchez, José-Antonio, Joseph P Ferrie, and Christopher Vickers. 2023. "Women and the Econometrics of Family Trees." National Bureau of Economic Research Working Paper 31598.

Espín-Sánchez, José-Antonio, Salvador Gil-Guirado, and Chris Vickers. 2022. "La "Doña" è Mobile: The Role of Women in Social Mobility in a Pre-Modern Economy." *The Journal of Economic History*, 82(1): 1–41.

FamilySearch. 2016. "Massachusetts Marriages, 1695-1910." *Database: http://FamilySearch.org*, Index based upon data collected by the Genealogical Society of Utah, Salt Lake City.

Feigenbaum, James. 2016. "A Machine Learning Approach to Census Record Linking." Working Paper.

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

Ferrie, Joseph P. 1999. *Yankeys Now: Immigrants in the Antebellum U.S. 1840-1860. NBER Series on Long-term Factors in Economic Development*, New York - Oxford:Oxford University Press.

Ferrie, Joseph P. 2005. "History Lessons: The End of American Exceptionalism? Mobility in the United States Since 1850." *Journal of Economic Perspectives*, 19(3): 199–215.

Fitch, Catherine A, and Steven Ruggles. 2000. "Historical Trends in Marriage Formation, United States." In *Ties that Bind: Perspectives on Marriage and Cohabitation.* , ed. Linda Waite, Christine Bachrach, Michelle Hindin, Elizabeth Thomson, and Arland Thornton, 59–88. Aldine de Gruyter.

Ghosh, Arkadev, Sam Hwang, and Munir Squires. 2025. "The Melting Pot: Industrialization and Integration in America." Working Paper.

Gihleb, Rania, and Kevin Lang. 2020. "Educational Homogamy and Assortative Mating Have Not Increased." In *Change at Home, in the Labor Market, and On the Job*. Vol. 48 of *Research in Labor Economics*, , ed. Solomon W. Polachek and Konstantinos Tatsiramos, 1–26. Emerald Publishing Limited.

Goldin, Claudia. 1983. "The Changing Economic Role of Women: A Quantitative Approach." *The Journal of Interdisciplinary History*, 13(4): 707–733.

Grawe, Nathan D. 2006. "Lifecycle bias in estimates of intergenerational earnings persistence." *Labour Economics*, 13(5): 551–570.

Haider, Steven, and Gary Solon. 2006. "Life-Cycle Variation in the Association between Current and Lifetime Earnings." *The American Economic Review*, 96(4): 1308–1320.

Haines, Michael R. 1996. "Long-term marriage patterns in the United States from colonial times to the present." *The History of the Family*, 1(1): 15–39.

Hajnal, John. 1965. "European Marriage Patterns in Perspective." In *Population in History: Essays in Historical Demography.* , ed. David Glass and D.E.C. Eversley, 101–143. Chicago:Aldine.

Hamilton, Darrick, and William A Darity. 2017. "The Political Economy of Education, Financial Literacy, and the Racial Wealth Gap." *Federal Reserve Bank of St. Louis Review*, 99(1): 59–76.

Helgertz, Jonas, Steven Ruggles, John Robert Warren, Catherine A. Fitch, J. David Hacker, Matt A. Nelson, Joseph P. Price, Evan Roberts, and Matthew Sobek. 2023. "IPUMS Multigenerational Longitudinal Panel: Version 1.1 [dataset]."

Hertz, Tom. 2008. "A group-specific measure of intergenerational persistence." *Economics Letters*, 100(3): 415–417.

Jácome, Elisa, Ilyana Kuziemko, and Suresh Naidu. 2025. "Mobility for All: Representative Intergenerational Mobility Estimates over the Twentieth Century." *Journal of Political Economy*, 133(1): 306–354.

Kalmijn, Matthijs. 1994. "Assortative Mating by Cultural and Economic Occupational Status." *American Journal of Sociology*, 100(2): 422–452.

Kim, Sukkoo. 1995. "Expansion of Markets and the Geographic Distribution of Economic Activities: The Trends in U. S. Regional Manufacturing Structure, 1860-1987." *The Quarterly Journal of Economics*, 110(4): 881–908.

Kim, Sukkoo. 1998. "Economic Integration and Convergence: U.S. Regions, 1840-1987." *The Journal of Economic History*, 58(3): 659–683.

Lee, Chul-In, and Gary Solon. 2009. "Trends in Intergenerational Income Mobility." *The Review of Economics and Statistics*, 91(4): 766–772.

Long, Jason, and Joseph Ferrie. 2007. "The Path to Convergence: Intergenerational Occupational Mobility in Britain and the US in Three Eras." *Economic Journal*, 117(519): 61–71.

Long, Jason, and Joseph Ferrie. 2013. "Intergenerational Occupational Mobility in Great Britain and the United States since 1850." *American Economic Review*, 103(4): 1109–37.

Mazumder, Bhashkar. 2005. "Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data." *The Review of Economics and Statistics*, 87(2): 235–255.

Mazumder, Bhashkar. 2014. "Black–White Differences in Intergenerational Economic Mobility in the U.S." *Economic Perspectives*, 38(Q I): 1–18.

Mazumder, Bhashkar, and Miguel Acosta. 2015. "Using Occupation to Measure Intergenerational Mobility." *The ANNALS of the American Academy of Political and Social Science*, 657(1): 174–193.

Olivetti, Claudia, and M. Daniele Paserman. 2015. "In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1940." *American Economic Review*, 105(8): 2695–2724.

Olivetti, Claudia, M. Daniele Paserman, Laura Salisbury, and E. Anna Weber. forthcoming. "The Emergence of the Socioeconomic Gradient in Women's Marriage Outcomes." *The Review of Economics and Statistics*, 1–29.

Oppenheimer, Valerie Kincade. 1988. "A Theory of Marriage Timing." *American Journal of Sociology*, 94(3): 563–591. Publisher: University of Chicago Press.

Parman, John. 2011. "American Mobility and the Expansion of Public Education." *The Journal of Economic History*, 71(1): 105–132.

Pérez, Santiago. 2017. "The (South) American dream: Mobility and economic outcomes of first- and second-generation immigrants in nineteenth-century Argentina." *The Journal of Economic History*, 77(4): 971–1006.

Preston, Samuel H, and Michael R Haines. 1991. *Fatal Years: Child Mortality in Late Nineteenth-Century America.* Princeton University Press.

Price, Joseph, Kasey Buckles, Jacob Van Leeuwen, and Isaac Riley. 2021. "Combining family history and machine learning to link historical records: The Census Tree data set." *Explorations in Economic History*, 80(C): S0014498321000024.

Ruggles, Steven. 2016. "Marriage, Family Systems, and Economic Opportunity in the USA Since 1850." In *Gender and Couple Relationships*. Vol. 6, , ed. Susan M. McHale, Valarie King, Jennifer Van Hook and Alan Booth, 3–41. Cham:Springer International Publishing. Series Title: National Symposium on Family Issues.

Ruggles, Steven, Matt A. Nelson, Matthew Sobek, Catherine A. Fitch, Ronald Goeken, J. David Hacker, Evan Roberts, and J. Robert Warren. 2024. "IPUMS Ancestry Full Count Data: Version 4.0 [dataset]."

Salisbury, Laura. 2014. "Selective migration, wages, and occupational mobility in nineteenth century America." *Explorations in Economic History*, 53(C): 40–63.

Solon, Gary. 1992. "Intergenerational Income Mobility in the United States." *American Economic Review*, 82(3): 393–408.

Song, Xi, Catherine G. Massey, Karen A. Rolf, Joseph P. Ferrie, Jonathan L. Rothbaum, and Yu Xie. 2020. "Long-term decline in intergenerational mobility in the United States since the 1850s." *Proceedings of the National Academy of Sciences*, 117(1): 251–258. Publisher: Proceedings of the National Academy of Sciences.

Ward, Zachary. 2020. "The Not-So-Hot Melting Pot: The Persistence of Outcomes for Descendants of the Age of Mass Migration." *American Economic Journal: Applied Economics*, 12(4): 73–102.

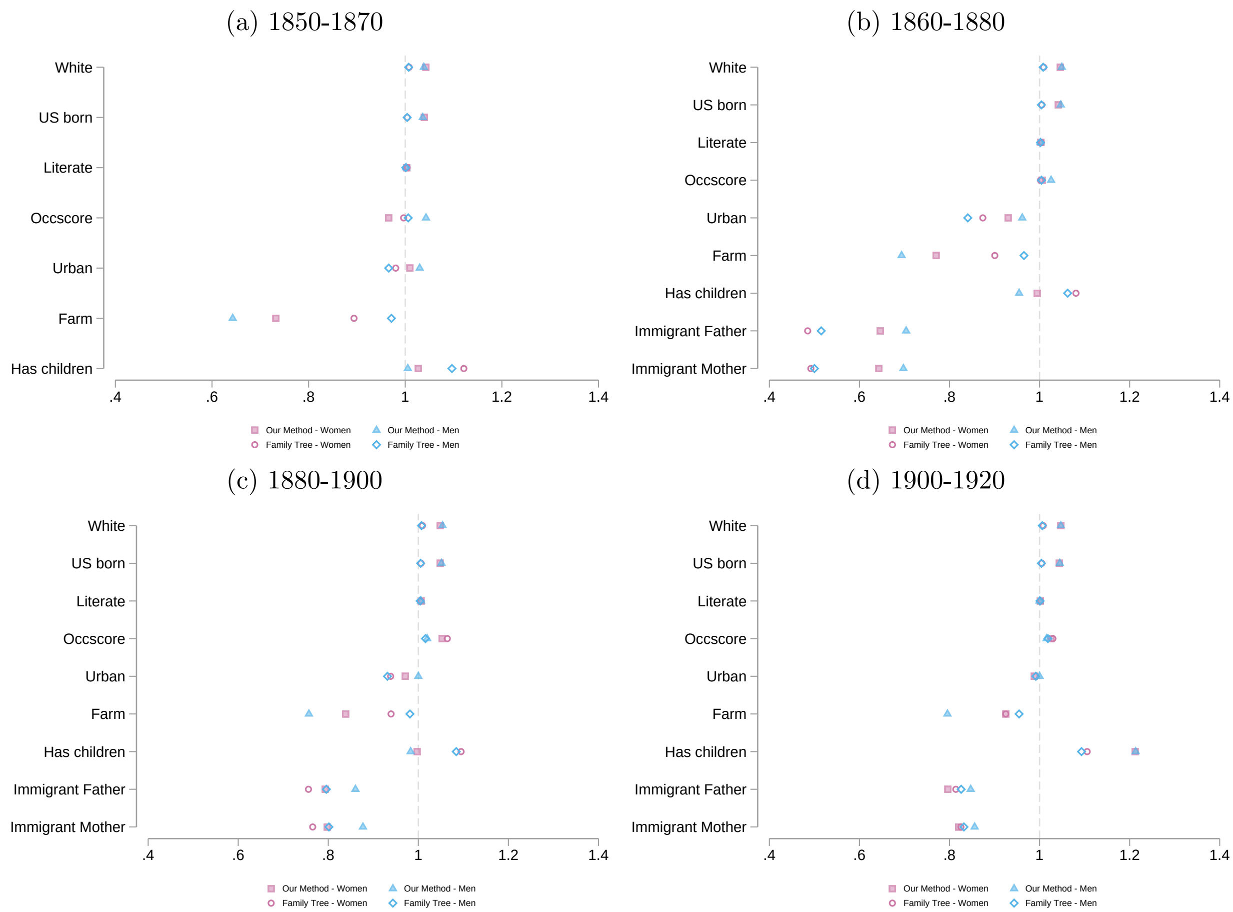
Ward, Zachary. 2023. "Intergenerational Mobility in American History: Accounting for Race and Measurement Error." *American Economic Review*, 113(12): 3213–48.

Xie, Yu, and Alexandra Killewald. 2013. "Intergenerational Occupational Mobility in Great Britain and the United States since 1850: Comment." *American Economic Review*, 103(5): 2003–2020.

Zimran, Ariell. 2019. "Sample-selection bias and height trends in the nineteenth-century United States." *The Journal of Economic History*, 79(1): 99–138.

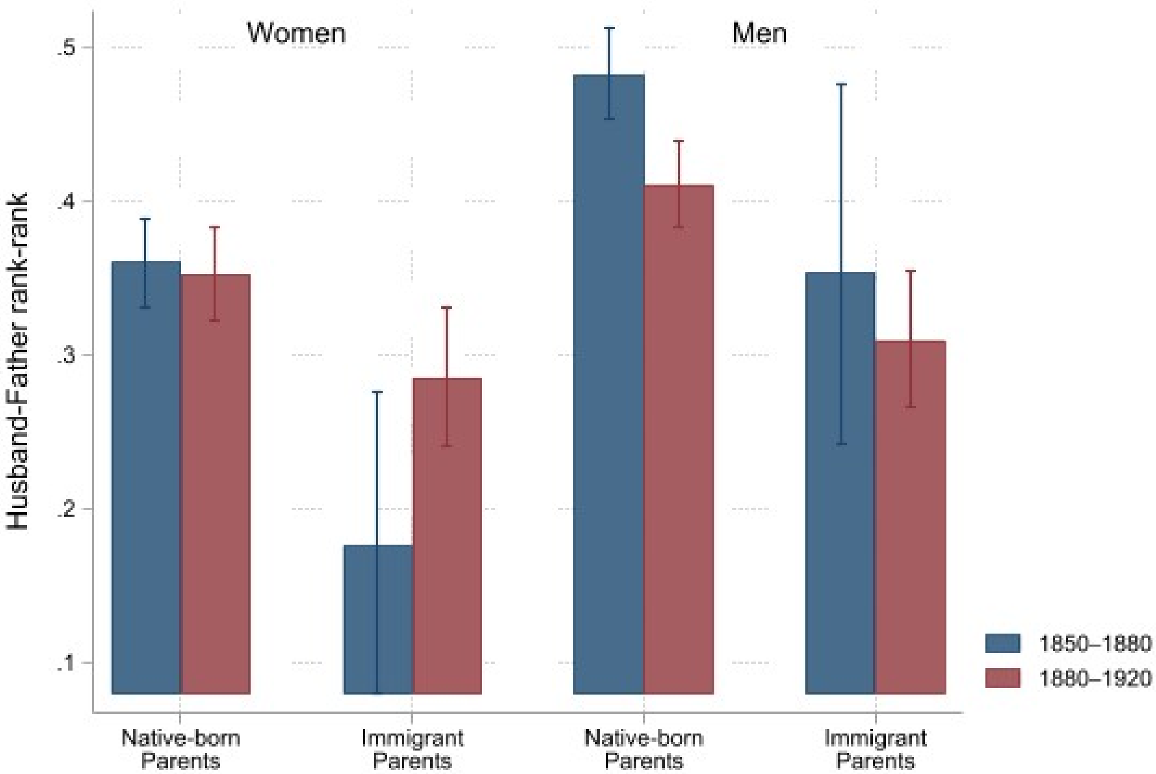
A Additional Tables and Figures

Figure A1: Representativeness of Linked Data Relative to Adult Census Population



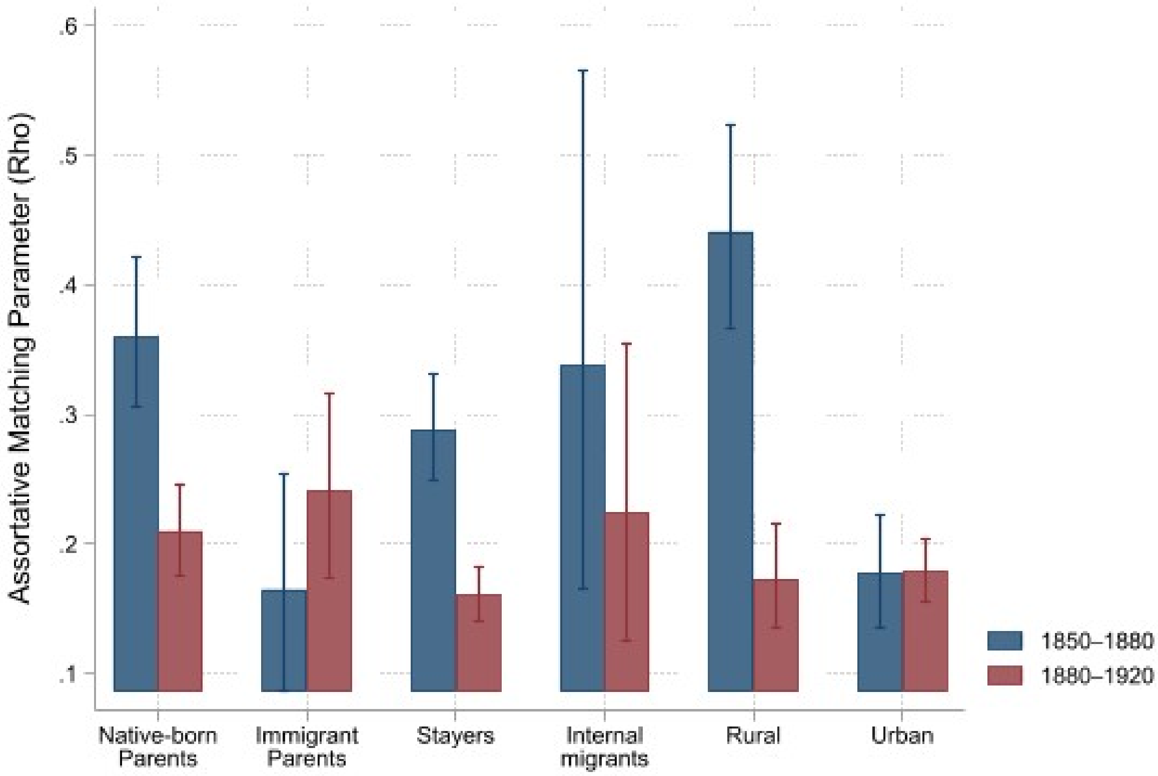
Notes: This figure illustrates the representativeness of links created by our method and genealogical links from the FamilySearch Family Tree relative to the adult census population. In each case, links are compared to the population of married individuals who are born and living in MA. Links for women (men) are compared to the population of women (men) in MA in the census. For each attribute, the x-axis represents a ratio indicating the mean of that attribute in the matched sample relative to the population mean. Dotted line for a ratio of 1 indicates perfect representativeness and values below (above) indicate under- (over-) representation of individuals with that attribute in the linked sample. Variables for mother and father's birthplace, (*mbpl* and *fbpl*), needed to classify parents as immigrants are not recorded in the 1870 census.

Figure A2: Mobility coefficient by parent nativity



Notes: This figure presents group specific mobility estimates for women and men by their own parents' nativity. The sample corresponds to that used for Figure 5. Men's mobility is the rank-rank association between a husband's status and that of his father. Women's mobility is the rank-rank association between her husband's status and that of her own father. A second observation of status is used as an instrument for father's or father-in-law's status.

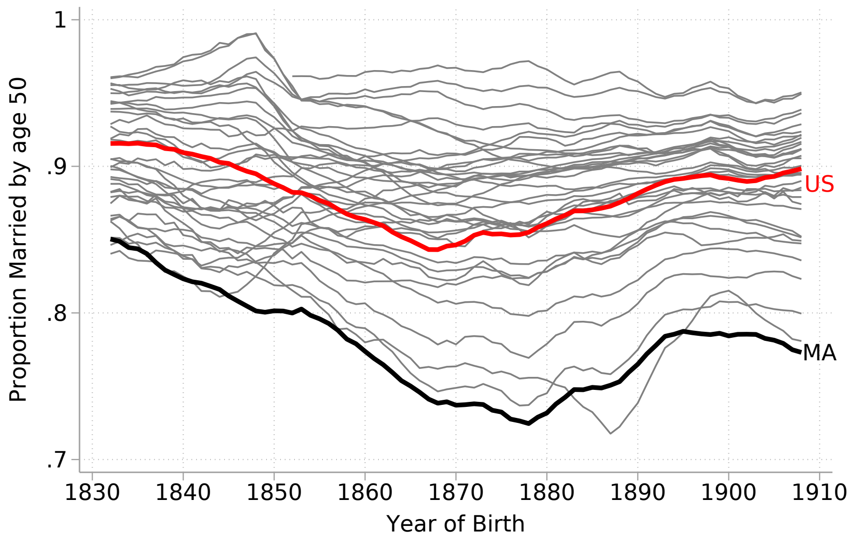
Figure A3: Estimated assortative mating by demographic group from men's vantage point



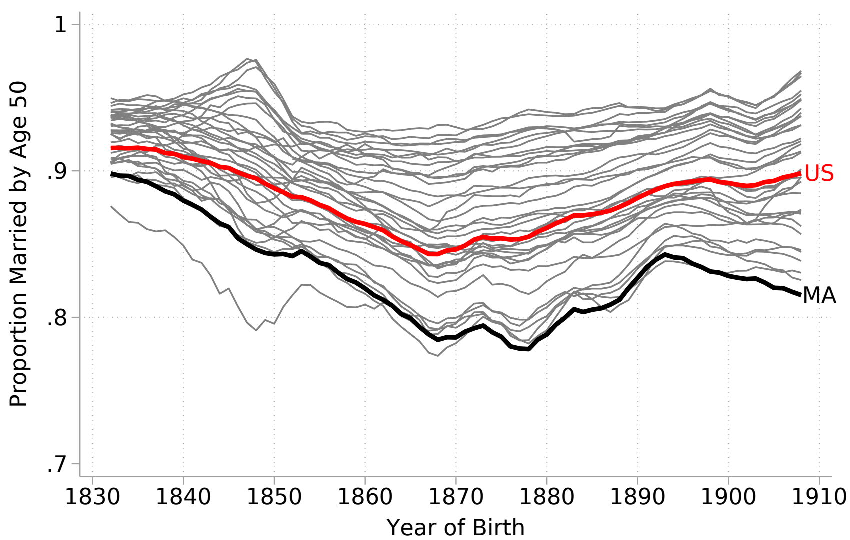
Notes: This figure plots the estimated assortative matching parameter (ρ) by demographic subgroup, calculated from the husband's perspective. Each bar represents the cohort-specific ρ for marriages formed by men in the indicated group: sons of native-born or immigrant parents, sons who remained in Massachusetts (stayers) or moved elsewhere within U.S. (internal migrants), and sons raised in rural or urban towns. Blue bars correspond to the 1850–1880 cohorts and red bars to the 1880–1920 cohorts; vertical lines denote 95% confidence intervals. Estimates are obtained from the structural model described in Section 2, using the same specification as the aggregate results in Figure 4a.

Figure A4: Prevalence of Marriage by Age 50

(a) Women



(b) Men



Notes: Figures show the proportion married of native-born white men and women by birth cohort at age 50. We use the census complete count data for 1880-1940 from Ruggles et al. (2024) and limit the sample to ages 30-50. To hold the age distribution constant, an indicator for ever married is regressed on a set of indicators for age, on separate samples for men and women. The regression-adjustment becomes: cohort rate = $\alpha + \beta_{50}$ + mean residual. We plot the five-year moving average to emphasize the trend by state of birth and for the nation as a whole.
Sources: 1850-1940 Decennial Census data from Ruggles et al. (2024).

Table A1: Accordance Rates with Family Tree Links

	1850-1870		1860-1880		1880-1900		1900-1920	
	N	Acc. rate	N	Acc. rate	N	Acc. rate	N	Acc. rate
Panel A: Married Women								
10% False Positive Rate	4,109	97.83	6,715	98.26	10,860	97.31	13,974	98.40
5% False Positive Rate	3,530	98.13	5,436	98.49	9,178	97.85	12,276	98.85
2% False Positive Rate	3,010	98.67	4,776	98.89	8,669	98.59	9,599	99.45
1% False Positive Rate	2,742	98.76	4,139	99.18	7,722	98.86	7,767	99.51
Panel B: Married Men								
10% False Positive Rate	3,782	98.36	6,341	98.33	11,084	98.40	11,824	98.11
5% False Positive Rate	3,284	98.51	5,212	98.98	9,185	98.87	10,506	98.72
2% False Positive Rate	2,659	98.68	4,739	99.26	7,817	99.31	8,889	99.25
1% False Positive Rate	2,275	98.81	4,549	99.36	7,099	99.21	8,047	99.45

Notes: This table reports the share of same matches made by our method and genealogical links from the FamilySearch Family Tree, conditional on trying to link the same person from the base year for married women (Panel A) and men (Panel B). For each linked sample, we report the number of individuals in the base year, denoted by "N", that both methods try to link and the accordance rate, "Acc. rate", which represents the share of those individuals in "N" which are matched to the same person by both methods. We report these numbers for various samples from our method created using different false positive rates. In all cases, the linked samples from both methods are restricted to men and women who are born in MA and are reported as residing there in the adult census.

Table A2: Men's and Women's Intergenerational Mobility Estimates

	1850-1870	1860-1880	1880-1900	1900-1920
	Panel A: Full Sample			
Father Rank (β_1)	0.446	0.421	0.336	0.344
	(0.018)	(0.013)	(0.010)	(0.010)
Father Rank x Woman (β_2)	-0.049	-0.040	0.019	-0.018
	(0.025)	(0.019)	(0.015)	(0.014)
Intercept (α)	41.348	40.386	47.883	45.054
	(0.920)	(0.741)	(0.587)	(0.598)
Woman (β_0)	2.443	2.223	-2.010	1.011
	(1.261)	(1.067)	(0.864)	(0.842)
No. of Observations	21,158	30,957	39,273	38,862
	Panel B: Massachusetts Born			
Father Rank (β_1)	0.382	0.376	0.265	0.208
	(0.034)	(0.026)	(0.018)	(0.018)
Father Rank x Woman (β_2)	-0.015	-0.006	0.060	0.092
	(0.042)	(0.034)	(0.024)	(0.025)
Intercept (α)	45.255	43.207	52.735	55.118
	(1.717)	(1.521)	(1.014)	(1.065)
Woman (β_0)	1.793	1.572	-3.261	-6.375
	(2.183)	(1.933)	(1.381)	(1.435)
No. of Observations	15,841	22,530	27,321	26,879

Notes: This table reports 2SLS estimates from equation 10, estimated separately for each cohort, with heteroskedasticity-robust standard errors shown in parentheses. The father's wealth score rank is instrumented using a second observation on the father. All regressions include control variables: a quartic in father's age and a quartic in the adult child's age, both measured at the time economic status is observed. Panel (a) presents results for the full sample, while Panel (b) restricts the analysis to individuals born in Massachusetts. For the Massachusetts-born sample, observations are weighted to account for differential selection into the linked sample, whereas the full-sample results in Panel (a) are unweighted. The intercept (α) captures the absolute rank mobility of individuals born to fathers at the very bottom of the wealth distribution. The coefficient (β_0) on the Woman indicator represents the difference in absolute mobility for women relative to men. The two main coefficients of interest are (β_1), which measures relative mobility for men, and (β_2), which captures the differential mobility experienced by women. Relative mobility for women is ($\beta_1 + \beta_2$). Age polynomials are centered around age 35, which only affects the interpretation of the intercept terms, not the slopes. *Sources:* 1850-1920 Decennial Census data from Ruggles et al. (2024). Marriage registrations from *FamilySearch.org*.

Table A3: Mobility Results with Different Measures of Economic Status

	1850-1870	1860-1880	1880-1900	1900-1920
	Panel A: 1870 Wealth Scores			
Father Rank (β_1)	0.429	0.481	0.370	0.310
	(0.035)	(0.032)	(0.025)	(0.024)
Father Rank x Woman (β_2)	-0.060	-0.080	0.010	0.036
	(0.045)	(0.041)	(0.032)	(0.031)
N	15,841	22,530	27,321	26,879
	Panel B: 1901 Cost of Living Survey			
Father Rank (β_1)	0.351	0.447	0.333	0.330
	(0.041)	(0.034)	(0.027)	(0.029)
Father Rank x Woman (β_2)	-0.012	-0.078	0.016	0.022
	(0.051)	(0.043)	(0.036)	(0.038)
N	15,860	22,535	27,333	26,890
	Panel C: 1950 Occscore			
Father Rank (β_1)	0.270	0.315	0.321	0.361
	(0.037)	(0.024)	(0.024)	(0.030)
Father Rank x Woman (β_2)	-0.040	-0.031	-0.033	-0.019
	(0.045)	(0.032)	(0.033)	(0.039)
N	15,841	22,530	27,321	26,879
	Panel D: Immigrant specific Song Scores			
Father Rank (β_1)	0.415	0.480	0.319	0.273
	(0.051)	(0.041)	(0.020)	(0.021)
Father Rank x Woman (β_2)	-0.037	-0.051	-0.001	0.035
	(0.063)	(0.051)	(0.027)	(0.027)
N	15,773	22,516	27,332	26,904

Notes: This table presents 2SLS estimates of the rank-rank parameter from equation 10, estimated separately for each cohort, with heteroskedasticity-robust standard errors reported in parentheses. The father's occupational score rank is instrumented using a second observation of his economic status. All regressions include control variables: a quartic in the father's age and a quartic in the adult child's age, both measured at the time their economic status is observed. Observations are weighted to adjust for differential selection into the linked sample. Each panel applies a different measure of occupational status.
Sources: 1850-1920 Decennial Census data from Ruggles et al. (2024). Marriage registrations from *FamilySearch.org*. Occupation scores based on the 1901 Cost of Living Survey are taken from Abramitzky et al. (2021*a*).

Table A4: Mobility Robustness Checks

	1850-1870	1860-1880	1880-1900	1900-1920
	Panel A: Drop Farmers			
Father Rank (β_1)	0.425	0.366	0.300	0.231
	(0.035)	(0.029)	(0.018)	(0.020)
Father Rank x Woman (β_2)	-0.048	0.041	0.060	0.090
	(0.046)	(0.037)	(0.025)	(0.026)
N	9,918	15,633	21,522	23,441
	Panel B: Limit Age to 30-50			
Father Rank (β_1)	0.410	0.362	0.264	0.269
	(0.053)	(0.049)	(0.027)	(0.025)
Father Rank x Woman (β_2)	-0.066	0.023	0.062	0.058
	(0.063)	(0.057)	(0.036)	(0.032)
N	7,895	10,516	12,385	16,154
	Panel C: IGE			
Father Status (β_1)	0.425	0.380	0.273	0.219
	(0.040)	(0.027)	(0.019)	(0.018)
Father Status x Woman (β_2)	0.004	0.004	0.072	0.083
	(0.050)	(0.035)	(0.025)	(0.024)
N	15,841	22,530	27,268	26,820

Notes: This table presents 2SLS estimates of mobility parameters from equation 10, estimated separately for each cohort, with heteroskedasticity-robust standard errors reported in parentheses. The father's occupational score is instrumented using a second observation of his economic status. All regressions include control variables: a quartic in the father's age and a quartic in the adult child's age, both measured at the time their economic status is observed. Observations are weighted to adjust for differential selection into the linked sample. Panel (A) reports the rank-rank parameters after dropping observations when either the father or adult reports their occupation as farmer. Panel (B) limits the sample to observations where both the child and father are observed at ages 30-50. Panel (C) reports log-log IGE regressions.
Sources: 1850-1920 Decennial Census data from Ruggles et al. (2024). Marriage registrations from *FamilySearch.org*.

Table A5: Main Results at Different False Positive Cutoffs Used in Linkage Algorithm

	1850-1870	1860-1880	1880-1900	1900-1920
	Panel A: 9% False Positive Cutoff			
Father Rank	0.384	0.378	0.265	0.213
	(0.034)	(0.027)	(0.018)	(0.019)
Father Rank x Woman	-0.020	-0.006	0.062	0.088
	(0.042)	(0.035)	(0.024)	(0.025)
No. of Observations	15,551	22,099	26,833	26,475
	Panel B: 8% False Positive Cutoff			
Father Rank	0.382	0.380	0.267	0.218
	(0.034)	(0.028)	(0.018)	(0.019)
Father Rank x Woman	-0.026	-0.008	0.059	0.079
	(0.043)	(0.035)	(0.025)	(0.025)
No. of Observations	15,088	21,410	26,266	25,527
	Panel C: 7% False Positive Cutoff			
Father Rank	0.375	0.380	0.274	0.210
	(0.035)	(0.029)	(0.019)	(0.020)
Father Rank x Woman	-0.023	-0.008	0.053	0.090
	(0.044)	(0.036)	(0.026)	(0.026)
No. of Observations	14,479	20,715	25,002	24,991
	Panel D: 6% False Positive Cutoff			
Father Rank	0.395	0.381	0.265	0.205
	(0.032)	(0.029)	(0.019)	(0.020)
Father Rank x Woman	-0.032	-0.001	0.059	0.097
	(0.042)	(0.038)	(0.026)	(0.026)
No. of Observations	14,042	19,766	24,071	23,986
	Panel F: 5% False Positive Cutoff			
Father Rank	0.423	0.415	0.269	0.215
	(0.033)	(0.031)	(0.020)	(0.020)
Father Rank x Woman	-0.063	-0.031	0.062	0.085
	(0.044)	(0.040)	(0.027)	(0.027)
No. of Observations	13,194	17,841	22,271	23,211

Notes: This table reports 2SLS estimates from equation 10, with heteroskedasticity-robust standard errors in parentheses. Each panel reports results for a different choice of false positive cutoff when selecting the hyperparameters (Score and Ratio) in the record linkage algorithm. The base results use a 10 percent false positive cutoff in the training data. The father's wealth score rank is instrumented using a second observation on the father. All regressions include control variables: a quartic in father's age and a quartic in the adult child's age, both measured at the time economic status is observed. Observations are weighted to account for differential selection into the linked sample
Sources: 1850-1920 Decennial Census data from Ruggles et al. (2024). Marriage registrations from *FamilySearch.org*.

Table A6: Robustness of Sorting Regressions

	1850-1870	1860-1880	1880-1900	1900-1920
	Panel A: Switching Y and X variables			
Base: Wife's Father's Rank	0.318	0.336	0.441	0.417
	(0.041)	(0.036)	(0.024)	(0.026)
Husband's Father's Rank	0.357	0.378	0.460	0.392
	(0.041)	(0.038)	(0.026)	(0.023)
	Panel B: Different measures of economic status			
1901 Cost of Living Survey	0.333	0.292	0.254	0.294
	(0.036)	(0.039)	(0.031)	(0.032)
1950 Occscore	0.304	0.276	0.354	0.237
	(0.052)	(0.042)	(0.030)	(0.030)
Immigrant specific Song Scores	0.229	0.247	0.364	0.301
	(0.044)	(0.039)	(0.026)	(0.027)

Notes: This table presents 2SLS estimates of the rank-rank parameter from equation 11, estimated separately for each cohort, with heteroskedasticity-robust standard errors reported in parentheses. The independent variable father's occupational score rank is instrumented using a second observation of his economic status. All regressions include control variables: a quartic in the father's age and a quartic in the adult child's age, both measured at the time their economic status is observed. Observations are weighted to adjust for differential selection into the linked sample. Each panel applies a different measure of occupational status.
Sources: 1850-1920 Decennial Census data from Ruggles et al. (2024). Marriage registrations from *FamilySearch.org*. Occupation scores based on the 1901 Cost of Living Survey are taken from Abramitzky et al. (2021*a*).

Table A7: Estimates Used to Calculate Parameters and Conduct Counterfactual

Cohort	b^h_f	b^h_{fl}	b^f_{fl}	ρ	β_m	β_f	Obs. Sorting	Obs. Mobility
1850-1870	0.46	0.34	0.35	0.34	0.77	0.20	3,602	7,204
1860-1880	0.45	0.33	0.33	0.33	0.83	0.17	5,099	10,198
1880-1900	0.33	0.26	0.42	0.16	0.34	0.27	6,554	13,108
1900-1920	0.26	0.23	0.41	0.13	0.36	0.21	5,296	10,592

Notes: The first two columns present 2SLS estimates of the rank-rank mobility parameters from equation 10, calculated separately for each cohort: b^h_f for men's mobility and b^h_{fl} for women's mobility. The third column, b^f_{fl}, reports the 2SLS rank-rank slope between the father and father-in-law. The next three columns display the structural parameters calculated from equations 12: assortative matching (ρ), inheritance from the mother (β_m), and inheritance from the father (β_f). The final two columns show the number of couple observations used to estimate b^f_{fl} and the number of individual observations used to estimate the mobility terms. The sample for this table includes only observations where both spouses are linked to their fathers, with at least one spouse born in Massachusetts.
Sources: 1850-1920 Decennial Census data from Ruggles et al. (2024). Marriage registrations from *FamilySearch.org*.

Table A8: Papers Studying Women's Historical Intergenerational Mobility in the US

Paper	Years / Birth-Cohort	Estimate	Mobility Metric	Sample	Data / Linking Method
Olivetti & Paserman (2015)	1850-1870	0.34	IGE	US	Pseudo-Linking
Olivetti & Paserman (2015)	1860-1880	0.40	IGE	US	Pseudo-Linking
Olivetti & Paserman (2015)	1880-1900	0.40	IGE	US	Pseudo-Linking
Olivetti & Paserman (2015)	1900-1920	0.49	IGE	US	Pseudo-Linking
Olivetti & Paserman (2015)	1910-1930	0.41	IGE	US	Pseudo-Linking
Olivetti & Paserman (2015)	1920-1940	0.37	IGE	US	Pseudo-Linking
Buckles et al. (2023)	1840	0.76	IV	US, White	Family Tree
Buckles et al. (2023)	1850	0.78	IV	US, White	Family Tree
Buckles et al. (2023)	1860	0.70	IV	US, White	Family Tree
Buckles et al. (2023)	1870	0.62	IV	US, White	Family Tree
Buckles et al. (2023)	1880	0.57	IV	US, White	Family Tree
Buckles et al. (2023)	1890	0.55	IV	US, White	Family Tree
Buckles et al. (2023)	1900	0.58	IV	US, White	Family Tree
Buckles et al. (2023)	1910	0.59	IV	US, White	Family Tree
Bailey & Lin (2024)	1865	0.29	Rank-Rank	OH	Life-M
Bailey & Lin (2024)	1870	0.36	Rank-Rank	OH	Life-M
Bailey & Lin (2024)	1875	0.31	Rank-Rank	OH	Life-M
Bailey & Lin (2024)	1880	0.31	Rank-Rank	OH	Life-M
Bailey & Lin (2024)	1885	0.35	Rank-Rank	OH	Life-M
Bailey & Lin (2024)	1890	0.30	Rank-Rank	OH	Life-M
Bailey & Lin (2024)	1895	0.29	Rank-Rank	OH	Life-M
Bailey & Lin (2024)	1900	0.26	Rank-Rank	OH	Life-M
Bailey & Lin (2024)	1905	0.19	Rank-Rank	OH	Life-M
Bailey & Lin (2024)	1910	0.19	Rank-Rank	OH	Life-M
Bailey & Lin (2024)	1915	0.15	Rank-Rank	OH	Life-M
Bailey & Lin (2024)	1920	0.15	Rank-Rank	OH	Life-M
Jácome et al. (2025)	1910s	0.69	Rank-Rank	US, White	Survey Data
Jácome et al. (2025)	1920s	0.48	Rank-Rank	US, White	Survey Data
Jácome et al. (2025)	1930s	0.40	Rank-Rank	US, White	Survey Data
Jácome et al. (2025)	1940s	0.35	Rank-Rank	US, White	Survey Data
Jácome et al. (2025)	1950s	0.46	Rank-Rank	US, White	Survey Data
Jácome et al. (2025)	1960s	0.71	Rank-Rank	US, White	Survey Data
Jácome et al. (2025)	1970s	0.61	Rank-Rank	US, White	Survey Data

Notes: This table reports historical intergenerational mobility estimates of married women found in the literature. Estimates of Olivetti and Paserman (2015) are obtained from Table 3 in their paper which presented OLS coefficients from a regression of son-in-law's income on imputed father-in-law's log occupational income. Estimates of Buckles et al. (2023) are obtained from Figure 6, Panel A which reports coefficients from an IV regression of the son-in-law's adjusted Song score on the father's, using a second father's observation as an instrument for the first. Rank-rank coefficients from Bailey & Lin (2024) are estimated from Figure 6, Panel A (precise numbers not presented). Estimates of White women from Jácome et al. (2025) are obtained from Table A.6 in their paper. Jácome et al. (2025) do not explicitly focus on married women and do not use a regression of son-in-law economic status on father-in-law's. Instead they rely on adult *family* income which in the historical period for White women is closely tied to their husbands' incomes.

B Record Linkage

Our record linkage algorithm proceeds in two steps: marriage register to post-marriage census, and post-marriage census to pre-marriage census. Both match steps follow a similar outline. The first match step begins by constructing all potential matches in the adult post-marriage census for each couple i in the marriage register. To be considered a potential match, households in the census must have both spouses present, be within 2 years difference in year of birth, and have the surname and both given names be similar in terms of string distance to couple i in the marriage register. String distance is measured using the Jaro-Winkler method and must be within 0.2 for all names. Marriages are matched to the first or second census occurring after the marriage (1850-1869 marriage to the 1870 census, 1860-1879 to the 1880 census, 1880-1889 to the 1900 census, and 1900-1919 to the 1920 census). We limit the sample to white respondents.

Next, a random sample of 20,000 married couples i (5,000 per twenty year cohort) is selected to match the proportion of couples with a single potential match and the proportion with multiple potential matches within each twenty year marriage cohort. Researchers then use the information on name similarity and years of birth to determine which, if any, potential match is the true link for couple i. Because it is conducted by hand, the choice of true match inherently relies on researcher judgment. Only potential matches that are a clear and certain match are marked as a true link. Those without a clear match, either with no good option or multiple potential matches that appear equally likely, are marked as unmatched. The set of 20,000 hand-linked marriages is then split in half into training data used to fit a prediction model and select hyper-parameters and a dataset which is used to cross-validate the linkage algorithm.

The training data is used to estimate a logistic regression prediction model to predict a true link based on functions of observable characteristics on the marriage certificates and census records. The most important characteristics are the string distances and functions of surnames, given names, and ages of the couple. Predictors in the logitistic regression include: Jaro-Winkler distance for both given names and surname, absolute value of birth year difference for both, a polynomial in number of potential matches, and indicators for if the match is exact (Jaro-Winkler distances equal 1 for all 3 names), if the match is exact and birth difference equals 0 for both, if first letter of each name matches, and if last letter of each name matches. The model is estimated separately for each 20-year marriage cohort and for marriages with single and multiple potential matches in the census.

We then select two hyper-parameters used to code links from the set of potential matches. For each couple i and potential match in the census j, the log odds ratio of being a true match is $ln(\frac{\hat{\rho_{ij}}}{1-\hat{\rho_{ij}}}) = \hat{\beta}X_{ij}$. A potential match ij will be coded as a link if it satisfies a set of requirements. For marriages with multiple potential matches, three requirements must be satisfied. First, the potential match has the highest predicted likelihood of being true match of all potential matches, $\hat{\rho_{ij}} \geq \hat{\rho_{ik}}$ for all $k \neq j$. Second, the predicted likelihood of being a true match is sufficiently high (i.e. greater than some cutoff). This cutoff is the first hyper-parameter – "Score". The requirement becomes $\hat{\rho_{ij}} >$ Score. Finally, the potential match with the second highest predicted value must be sufficiently different from the first. Here, we capture distance by the ratio of the highest to second highest predicted values. The second hyper-parameter is called the "Ratio". The final requirement for potential match ij to be coded as a link is $\frac{\hat{\rho_{ij}}}{\hat{\rho_{ik}}} > Ratio$ for all $k \neq j$. Marriage observations with single potential matches must satisfy only the first requirement: $\hat{\rho_{ij}} >$ Score.

The two hyper-parameters are chosen to maximize the True Positive Rate (TPR) while maintaining a false positive rate of no more than 10 percent in the training data. The TPR measures the proportion of the true links that our algorithm codes as links. The hyper-parameters are selected separately for the pools of single and multiple potential matches within each 20-year marriage cohort. Appendix Table B1 reports the hyper-parameters used to create linked data in our main sample.

The entire procedure is repeated for the second match step, which links the marriage observations i successfully matched to a post-marriage adult census in step 1 to a pre-marriage childhood census 20 years prior to observe the father's economic status. The procedure is identical to that used in step 1, except that individual steps are done separately for men and women. The pool of potential matches is constructed. A random sample of 5,000 marriage observations is selected for each sex and 20-year marriage cohort and hand-trained by researchers. Logit prediction models are fit on the training data sample. Predictors include Jaro-Winkler distances for individual's given name and surname, Jaro-Winkler distances for father and mother's given name, birth year difference, and indicators for if the match is exact (all four Jaro-Winkler distances equal one), if the match is exact and birth difference equals 0, if first letter of each name matches, and if last letter of each name matches. Score and Ratio hyper-parameters are selected in the same fashion and reported in Appendix Table B1.

We assess the quality of our algorithm by applying it to the hold-out sample of hand-trained links, showing that our algorithm effectively scales up work of researchers hand-training data. The cross-validation results reported in Appendix Table B2 suggests our procedure gives false positive rates comparable to other methodologies used in the literature.[18] False positives make up 11 percent of links of the full sample in the first match step, varying from 9 to 13 percent across marriage cohorts. We also capture a large portion of the true matches in the hand-trained data: 77 percent for the full sample and between 75 and 79 percent over marriage cohorts. The quality is even higher in the second match step. The false positive rate is 8 percent for both men and women in the full sample, and varies between 6 and 11 percent across cohorts. The improvement over the first match step likely derives from the fact that we have significantly more family information to use in the second step, and compared to other census to census matching algorithms used in the literature. In the census to census second link, matches are created with the surname, given name, state of birth, and birth year of the observation, as well as the given names of the mother and father. Whereas, we can only use the surname, given names, and birth years for both spouses in the marriage register to census match.

Finally, the trained prediction models are applied to the full data set, in a sense scaling up the

[18] See Abramitzky et al. (2021*b*) and Bailey et al. (2020).

judgement calls of an experienced researcher to hundreds of thousands of decisions. With a trained logit model, we estimate predicted values for all potential matches to predict the probability that a potential match is a correct match. Appendix Table B3 provides a complete breakdown of the causes for match failure.

Table B1: Selected Hyper-parameters

	Step 1: Marriage Register to Census		
	Single matches	Multiple matches	
	Score	Score	Ratio
1850-1870	0.41	0.65	76.60
1860-1880	0.50	0.83	91.40
1880-1900	0.44	0.70	36.60
1900-1920	0.41	0.73	3.60

	Step 2: Census to Census					
	Men			Women		
	Single matches	Multiple matches		Single matches	Multiple matches	
	Score	Score	Ratio	Score	Score	Ratio
1850-1870	0.30	0.50	1.00	0.30	0.50	1.00
1860-1880	0.30	0.50	10.40	0.30	0.50	1.00
1880-1900	0.30	0.58	3.80	0.30	0.61	2.60
1900-1920	0.30	0.62	1.80	0.38	0.50	2.20

Notes: This table reports the hyper-parameters used to determine whether an observation is coded as a link. A logit regression is first fit on the training sample, which then gives the predicted score for a potential match as $ln\left(\frac{\hat{\rho}}{1-\hat{\rho}}\right) = \hat{\beta}X$. The Score hyper-parameter is the cutoff such that $\hat{\rho}$ must be greater than to be coded as a link. The Ratio hyper-parameter is the cutoff for the distance between potential matches with the highest and second highest $\hat{\rho}$. The parameters were selected to give the maximum PPR while maintaining a false positive rate at or below 10 percent in the training data.

Table B2: Algorithm Quality: Cross-validation of Prediction Model and Parameter Selection

	Step 1: Marriage Register to Census			
	Couples			
	TPR	False Positives		
Full sample	0.77	0.11		
By cohort:				
1850-1870	0.76	0.13		
1860-1880	0.75	0.14		
1880-1900	0.76	0.11		
1900-1920	0.79	0.09		
	Step 2: Census to Census			
	Men		Women	
	TPR	False Positives	TPR	False Positives
Full sample	0.93	0.08	0.93	0.08
By cohort:				
1850-1870	0.96	0.06	0.96	0.06
1860-1880	0.94	0.07	0.94	0.06
1880-1900	0.92	0.10	0.93	0.08
1900-1920	0.93	0.09	0.90	0.11

Notes: This table reports cross-validation results for model parameters reported in Appendix Table B1 chosen to give a 10 percent false positive rate in the training data. TPR reports the proportion of true matches the fitted model codes as a match. False positives gives the proportion of coded matches that are not true matches.

Table B3: Summary of Matching Results and Reasons for Non-match

Panel A: Marriage Certificate to Census Link		
Category		
Total couples	**1,129,582**	**100%**
No potential matches found	**450,090**	**40%**
Causes of Match Failure	**385,387**	**34%**
Top potential match score too low (Hits = 1)	82,580	7%
Top potential match score too low (Hits >1)	5,260	0%
Top potential match ratio too low	118,949	11%
Top potential match score and ratio too low	170,975	15%
Matches	**294,105**	**26%**
Unique matches (Hits = 1)	243,221	22%
Non-unique matches (Hits >1)	50,884	5%

Panel B: Census to Census Link		
Category		
Total	**588,210**	**100%**
No potential matches found	**315,276**	**54%**
Causes of Match Failure	**86,032**	**15%**
Top potential match score too low (Hits = 1)	24,089	4%
Top potential match score too low (Hits >1)	43,360	7%
Top potential match ratio too low	16,415	3%
Top potential match score and ratio too low	2,168	0%
Matches	**186,902**	**32%**
Unique matches (Hits = 1)	127,093	22%
Non-unique matches (Hits >1)	59,809	10%

C Local measures of mobility and transition probabilities

Rank-rank slopes offer a convenient global summary of intergenerational mobility, but they can mask important variations in mobility at different points along the distribution (Deutscher and Mazumder, 2023). The binscatters in Figure C1 allow us to observe the local patterns of mobility by plotting the Conditional Expected Rank (CER)—the expected rank of children whose fathers are in a specific wealth score rank bin. This approach highlights how mobility differs across the distribution, providing insights that are obscured by the average rank-rank slope. We focus on the 1850–1870 and 1900–1920 cohorts, as they show the largest shifts in mobility over time and the most pronounced differences between genders.

For men, as shown in Panel (a), the average expected rank shows notable changes across cohorts. Sons of fathers at the 25th percentile in 1900–1920 achieved an expected rank 5.6 percentiles higher than sons of fathers at the same rank in 1850–1870. Conversely, sons of fathers at the 75th percentile experienced a decline in expected rank of 2.9 percentiles over the same period. These changes resulted in a flatter regression slope, capturing the overall increase in mobility for men.

Daughters of fathers at the 75th percentile in 1900–1920 had an expected rank 3.6 percentiles lower than daughters of fathers at the same rank in 1850–1870. In contrast, there was little to no change in the expected rank for daughters of fathers at the 25th percentile between the two cohorts. Sons from lower-ranked families saw significant average expected rank gains relative to those from higher-ranked families, while daughters experienced no comparable improvements. This stagnation for daughters at the lower end of the distribution contributes to the relatively small decline in persistence for women, and the divergence in rank-rank parameter estimates between genders over the sample period.

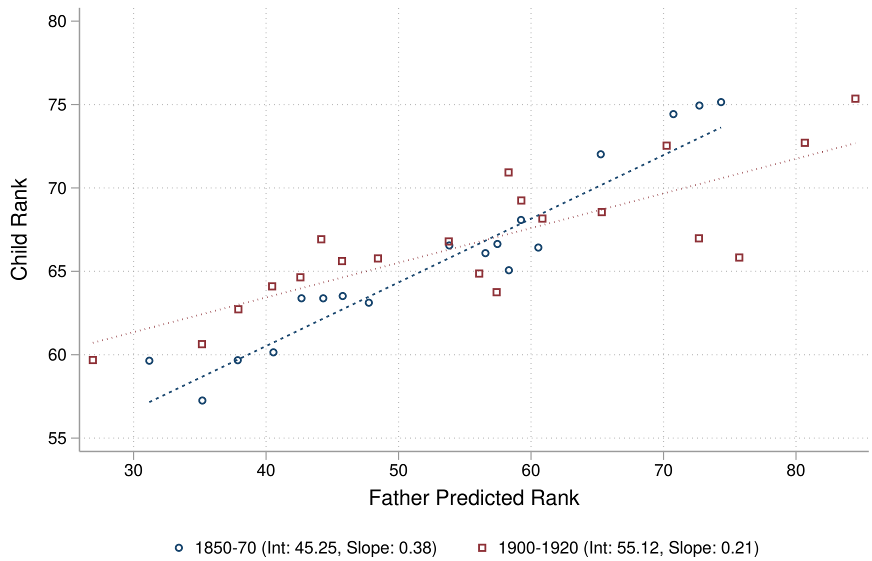
Understanding the direction of rank mobility is essential for analyzing changes in mobility across birth cohorts. For instance, how likely are individuals born to parents in the lower wealth score quintiles to rise to the top of the distribution? Table C1 presents transition probabilities by quintiles, broken down by birth cohort and sex. The top two rows of each panel highlight key mobility dynamics: cycles of poverty (1st to 1st quintile transitions) and the "rags to riches" narrative (1st to 5th quintile transitions) (Corak, 2020). We highlight key patterns in mobility, including limited cycles of poverty, relatively high upward mobility from the bottom quintile, and notable changes in persistence at the middle and top of the wealth score distribution across cohorts.

We find little evidence of persistent cycles of poverty. Across all cohorts, both men and women born in Massachusetts to parents in the 1st quintile have a low probability of remaining in the bottom quintile as adults. However, starting with the 1860–1880 cohort, the likelihood that a woman born in the 1st quintile remains there more than doubles (0.02 vs. 0.05), whereas there is no change for men. This disparity can be attributed to differences in opportunities: women's transitions are determined by the economic status of their husbands, while men's transitions reflect their own labor market outcomes. Women born in the bottom quintile are over twice as likely to marry a man in the bottom quintile as Massachusetts-born men are to remain there themselves.

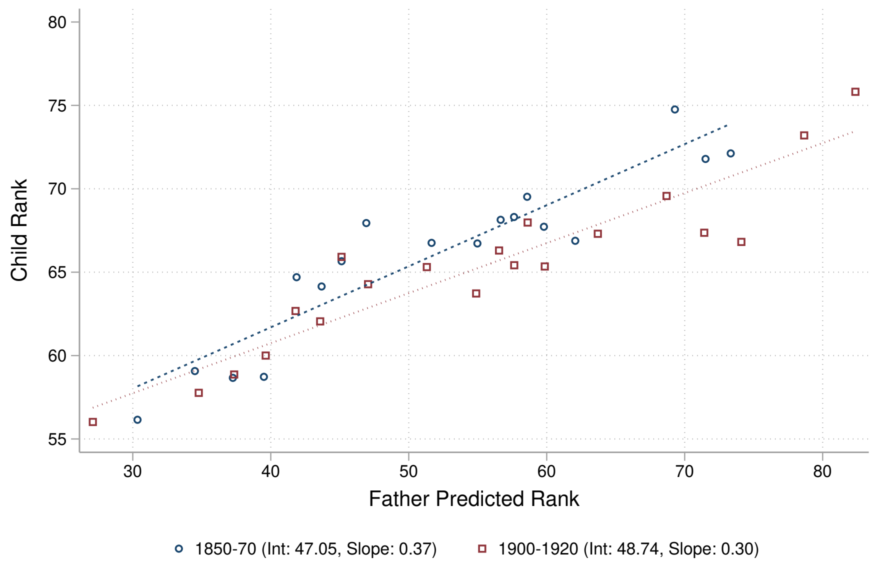
In contrast, the probability of moving from "rags to riches"—from the 1st to the 5th quintile—is

Figure C1: Binscatters for 1850-70 vs. 1900-1920 Cohorts

(a) Men



(b) Women



Notes: The figures display the average wealth score rank of Massachusetts-born children as a function of their father's predicted wealth score rank, based on data from the 1850-1870 and 1900-1920 cohorts. The father's rank represents the predicted rank derived from a first-stage regression, where the father's status is instrumented using the second observation. Accordingly, the solid lines correspond to the regression lines from the 2SLS estimates of equation 10, as reported in Appendix Table A2. The observations are weighted to account for differential selection into the linked sample. The bin widths were selected to create twenty bins, determined by the unconditional quantiles of fathers' ranks across the combined cohorts shown in the figure.
Sources: 1850-1920 Decennial Census data from Ruggles et al. (2024). Marriage certificates from *FamilySearch.org.*

relatively high. For women, this probability ranges from 0.16 to 0.20, compared to 0.08 to 0.17 for men. These figures are higher than comparable probabilities for other contexts, such as 0.114 for the Canadian 1963–1970 birth cohorts (Corak, 2020) and 0.123 for the Australian 1978–1982 cohorts (Deutscher and Mazumder, 2023). Importantly, we find no significant changes in either cycles of poverty or extreme upward mobility over time that can account for the large shifts observed in the global rank-rank measure of mobility.

At the top of the distribution, persistence is slightly higher for women than for men. There is evidence of changes in "cycles of privilege"—the 5th to 5th quintile transitions ($TP_{(5,5)}$). For women, this probability increases by 20 percent from the 1850–1870 to the 1860–1880 cohort, remains elevated for the 1880–1900 cohort, and then declines by 16 percent to near the level of the first cohort. For men, the increase is slightly smaller at 16 percent, followed by an 8 percent decline in the final cohort. These trends suggest some fluctuations in persistence at the top, though not as pronounced as might be expected given the overall changes in mobility.

While mobility at the extremes of the distribution—both the top and bottom—remained relatively stable over time, significant changes occurred in the middle. The probability of remaining in the middle quintile (3rd to 3rd transition) declines sharply between cohorts. For women, this probability was 0.43 in the first two cohorts, falling by 37 percent to 0.27 in the 1900–1920 cohort. For men, the probability declines even more, from 0.52–0.53 in the earlier cohorts to 0.32 in the final cohort, a 40 percent reduction.

Examining the full transition matrices, which report sample percentages in Appendix Tables C2 and C3, shows that the decline in middle-quintile persistence is accompanied by substantial upward mobility into the 4th quintile, with the likelihood of transitioning there more than doubling from 0.13 to 0.25 for women and from 0.13 to 0.28 for men. Smaller increases are observed for transitions to the 5th quintile, while downward mobility to the 2nd quintile is relatively limited.

Table C1: Selected Transition Probabilities

	1850-1870	1860-1880	1880-1900	1900-1920
	Panel A: Women			
TP(1st-1st) - Poverty cycle	0.02	0.05	0.05	0.05
TP(1st-5th) - Rags to riches	0.16	0.16	0.20	0.20
TP(5th-5th) - Cycles of privilege	0.46	0.55	0.57	0.48
TP(3rd-3rd) - Middle to middle	0.43	0.43	0.29	0.27
	Panel B: Men			
TP(1st-1st) - Poverty cycle	0.03	0.02	0.02	0.02
TP(1st-5th) - Rags to riches	0.13	0.08	0.17	0.15
TP(5th-5th) - Cycles of privilege	0.44	0.51	0.49	0.45
TP(3rd-3rd) - Middle to middle	0.52	0.53	0.34	0.32

Notes: This table presents conditional transition probabilities between father's rank quintile and child's rank quintile for Massachusetts-born men and women. Ranks are calculated separately for the father's and child's years using the national occupational wealth score distribution. A father's rank is determined as the average of observations of his economic status across two censuses. The observations are weighted to account for differential selection into the linked sample.
Sources: 1850-1920 Decennial Census data from Ruggles et al. (2024). Marriage registrations from *FamilySearch.org*.

Table C2: Transition Matrices - Married Massachusetts-born Women

Father rank quintile	Child rank quintile 1	2	3	4	5
1	0.10	1.18	1.94	0.28	0.68
2	0.14	3.79	9.96	1.98	4.46
3	0.17	3.11	11.53	4.25	7.98
4	0.13	3.89	12.46	5.38	16.30
5	0.13	0.58	3.07	1.77	4.71

(a) 1850-1870

Father rank quintile	Child rank quintile 1	2	3	4	5
1	0.27	1.45	1.94	0.52	0.79
2	0.18	3.66	9.80	2.39	4.41
3	0.12	3.46	12.24	3.75	9.15
4	0.17	3.24	12.14	4.42	14.70
5	0.02	0.66	2.79	1.55	6.16

(b) 1860-1880

Father rank quintile	Child rank quintile 1	2	3	4	5
1	0.33	1.40	2.09	1.05	1.24
2	0.72	4.07	6.34	4.11	5.14
3	0.45	4.57	10.15	7.99	12.22
4	0.24	1.59	3.55	4.20	8.21
5	0.29	1.46	3.21	3.69	11.69

(c) 1880-1900

Father rank quintile	Child rank quintile 1	2	3	4	5
1	0.22	1.38	1.31	0.92	0.96
2	0.76	5.35	6.64	5.92	5.20
3	0.72	5.01	7.28	6.75	7.55
4	0.42	2.97	4.97	5.97	7.56
5	0.24	2.47	3.54	5.24	10.67

(d) 1900-1920

Notes: This table presents transition matrices showing the relationship between fathers' rank quintiles and their daughters' rank quintiles for Massachusetts-born women. Each entry is the percentage of the sample in that cell. Ranks are calculated separately for the father's and daughter's years using the national occupational wealth score distribution. A father's rank is determined as the average of his economic status across two census observations. A daughter's rank is calculated from her husband's occupational wealth score. All observations are weighted to adjust for differential selection into the linked sample.
Sources: 1850-1920 Decennial Census data from Ruggles et al. (2024). Marriage registrations from *FamilySearch.org*.

Table C3: Transition Matrices - Married Massachusetts-born Men

Father rank quintile	Child rank quintile 1	2	3	4	5
1	0.09	0.67	1.74	0.27	0.42
2	0.10	2.94	11.13	2.46	2.26
3	0.11	3.23	14.18	4.26	5.75
4	0.12	4.29	14.82	6.25	14.16
5	0.00	0.68	3.59	1.80	4.70

(a) 1850-1870

Father rank quintile	Child rank quintile 1	2	3	4	5
1	0.08	0.97	2.50	0.50	0.35
2	0.04	3.42	10.57	2.72	3.46
3	0.15	3.36	14.86	3.81	6.09
4	0.13	3.37	13.03	5.71	14.24
5	0.01	0.35	3.19	1.68	5.43

(b) 1860-1880

Father rank quintile	Child rank quintile 1	2	3	4	5
1	0.09	1.11	2.03	1.39	0.97
2	0.54	2.72	7.80	4.78	4.02
3	0.30	3.24	11.86	9.41	10.01
4	0.10	1.47	4.29	5.26	7.15
5	0.16	1.84	3.78	5.10	10.59

(c) 1880-1900

Father rank quintile	Child rank quintile 1	2	3	4	5
1	0.08	1.08	1.63	1.11	0.69
2	0.34	3.33	8.52	7.45	4.43
3	0.54	3.86	8.39	7.41	5.90
4	0.35	1.97	5.45	7.12	6.38
5	0.30	2.82	3.89	6.23	10.71

(d) 1900-1920

Notes: This table presents transition matrices showing the relationship between fathers' rank quintiles and their sons' rank quintiles for Massachusetts-born men. Ranks are calculated separately for the father's and son's years using the national occupational wealth score distribution. A father's rank is determined as the average of his economic status across two census observations. All observations are weighted to adjust for differential selection into the linked sample.
Sources: 1850-1920 Decennial Census data from Ruggles et al. (2024). Marriage registrations from *FamilySearch.org*.