Marriage and the Intergenerational Mobility of Women: Evidence from Marriage Certificates 1850-1920*

Katherine Eriksson
UC Davis & NBER

Gregory Niemesh
Miami University & NBER

Myera Rashid
Northwestern Univeristy

Jacqueline Craig
McKinsey & Company

Abstract

Due to data limitations, long-run changes in women's economic mobility are not well understood. Using a set of marriage certificates from Massachusetts over the period of 1850-1920, we link women and men to their childhood and adult census records to obtain a measure of occupational standing across two generations. Intergenerational mobility was higher for women than for men in the earliest 1850-70 cohort. Men's mobility increases by the 1880-1900 cohort, whereas women's does not, leading to a convergence. During a period with low married women's labor force participation, the choice of a partner was crucial for women's economic status. We find evidence of strong and increasing assortative matching prior to 1880, followed by declines to the 1900-20 cohort. Absent the increase in marital sorting, married women would have experienced the same increases in intergenerational mobility as did men in the sample. Finally, both men and women in the youngest cohort experience an increase in mobility and decreases in marital sorting, consistent with the widespread expansion of educational attainment during the "High School Movement."

*Corresponding Author: Niemesh, Miami University, Department of Economics, 800 East High Street, Oxford, OH 45056, and NBER (email: niemesgt@miamioh.edu); Eriksson: University of California-Davis, Department of Economics, Davis, CA 95616, NBER, and Stellenbosch University (email: kaeriksson@ucdavis.edu); Rashid: Northwestern University, Department of Economics, 2211 Campus Dr., Evanston, IL 60208 (email: mrashid@u.northwestern.edu); Craig: McKinsey & Company (email: craigjk2@miamioh.edu). We acknowledge financial support from the Northwestern Center for Economic History. We have benefited from helpful comments from Mathew Curtis, Charles Moul, Bhashkar Mazumder, José-Antonio Espín-Sánchez, 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 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.

1 Introduction

*"The American Dream is that dream of a land in which life should be better and richer and fuller for everyone, with opportunity for each according to ability or achievement... It is a dream of social order in which each man and each **woman** shall be able to attain to the fullest stature of which they are innately capable, and be recognized by others for what they are, regardless of the fortuitous circumstances of birth or position."* -James Truslow Adams, *The Epic of America*, 1931

The "American Dream" captures the idea that equality of opportunity is a foundational tenet of the United States' ethos; those who work hard should be able to rise, socially and economically, regardless of their familial beginnings. Much is known about the persistence of earnings, income, wealth, and occupations of sons and fathers, in both modern and historical periods for the United States. The "American Dream" seemed plausible for men in the nineteenth century, less so in the twentieth. For a number of reasons, including a lack of suitable data, we know much less about the social mobility of women. This rather large knowledge gap remains despite the – seemingly needless to say – fact that women make up roughly half the population, half the genetic endowment transferred across generations, and play an important role in the provision and guidance of human capital investment in children, extending their reach across generations.

How economically mobile were women in the nineteenth century United States relative to their husbands? Did men and women face the same trend in mobility? Because women changed surnames upon marriage during this period, observing father-daughter adult pairs across time is extremely difficult. Thus, little is known about women's intergenerational mobility. What we do know about the nineteenth century is based on work by Olivetti and Paserman (2015), which develops a pseudo-linking strategy that relies on the economic content of given names. For each adult-daughter-father pair, the father's economic status is measured as the average economic status of men with daughters of the same given name as the adult daughter observation. Thus, a pseudo-link. Moving later in time, Jácome et al. (2021) use survey data to obtain mobility estimates for women in the mid-twentieth century US.

In contrast, ours is the first paper to use a direct-linking method to quantify the intergenerational persistence of economic status faced by women over the course of the late nineteenth century United States. We overcome the data limitations by constructing a new longitudinal dataset of father-son and father-daughter pairs linked across censuses for several cohorts of women that registered their marriage in Massachusetts between 1850 and 1920. Linked observations across marriage certificates and censuses are essential to estimating intergenerational mobility of women during the nineteenth century. First, no longitudinal datasets exist in this period that contain parent and child economic outcomes, as in the PSID. Second, marriage certificates are one of the few sources of data that list the maiden names of married women and cover a large portion of a population.[1] The second

[1]Birth certificates frequently list the maiden name of the mother. However, the Birth Registration Area did not begin in earnest until the early twentieth century and compliance rates were low (Eriksson et al., 2018), eliminating their usefulness for record-linkage during the nineteenth century. Miles (1999) also used marriage certificates to

limitation to estimating father-daughter economic mobility is the lack of historical information on women's *potential* labor market outcomes. Because outside-the-home market work was rare for married women during this period, a married woman's economic status was often determined primarily by the income, wealth, and status of her husband (Goldin, 1983).[2] Accordingly, we measure women's intergenerational mobility by estimating correlations in economic status between the husband and the wife's father, following the earlier literature (Olivetti and Paserman, 2015; Olivetti et al., 2018).

We construct a set of father-daughter and father-son pairs using two linkages across three datasets. We begin with a set of over 1 million recorded marriages between 1850 and 1920. Using given names, surname, and ages, couples from the marriage certificates are first matched to a post-marriage census to observe the husband's occupation. Of the marriages that are successfully linked in the first step, the husband and wife are individually matched to a pre-marriage childhood census in which the father's occupation can be observed for each spouse. For much of the analysis, we split the data into four cohorts (or periods) of marriages: 1850-70, 1860-80, 1880-1900, and 1900-20.

The double-match procedure is novel and provides a way forward to estimate women's mobility. We are able to take advantage of the large amount of family information recorded on marriage certificates and censuses by using a probabilistic method to create record linkages. This is in contrast to the historical intergenerational mobility literature, which has primarily used automated linking procedures that rely on the exact matching of men's names across two census waves (Long and Ferrie, 2013; Abramitzky et al., 2012; Collins and Wanamaker, 2014, 2015, 2022; Ward, 2022). In our dataset, for the first match of the married couple to a post-marriage census there are five dimensions over which to match: surname, two given names, and two ages. The issue becomes how to balance small differences in each of these dimensions when deciding the correct match. In the second match – post-marriage census to the pre-marriage census – we have even more information to find the correct record link. Here, four strings are used in the match – child's given name and surname, father's given name, and mother's given name – as well as the child's age and state of birth. The exact matching procedures are too conservative leading to many false negatives and a small sample. Instead, we implement Feigenbaum (2016)'s supervised machine learning algorithm and train parameters to balance transcription errors across the collection of name strings.

With our doubly matched sample completed, we estimate men's relative intergenerational mobility as the rank-rank association in economic status between the husband and husband's father, whereas we use the rank-rank association between the husband and the wife's father for female mobility. We must estimate mobility through the male the wife is connected with in both genera-

estimate occupational transitions, but for men, because the certificates list the occupation of both the husband and his father at the time of marriage. While this provides an avenue for studying mobility, it is not especially accurate, as the occupations are recorded at different points in life for the father and son. For this reason we are forced to conduct the double-match linking procedure.

[2] Labor force participation rates of white married women aged 18-50 were 1.8 percent in 1880, 2.2 percent in 1900, and 5.5 percent in 1910 (Authors calculations from 1% IPUMS samples). Under-reporting of occupations by married women does not substantially bias the estimates. Census estimates of married women's labor force participation are similar to those of working class families around 1900 (Goldin, 1983).

tions because a majority of females do not participate in outside-the-home market work.[3] We can also interpret the estimate as a woman's ability to marry upwards or downwards, as opposed to her ability to acquire a higher or lower income level on her own compared to her father.

Across all specifications and measures of economic status, women in the 1850-1870 cohort are more mobile than men. In other words, a woman had a greater ability to marry upwards or downwards in comparison to a male's ability to make a higher or lower level of income level than his father. Using our preferred specification, a rank-rank regression, with an occupational wealth score based off of total wealth in the 1870 census, we find a father-son rank-rank parameter of 0.231 for the 1850-1870 cohort and 0.201 for the 1900-20 cohort. For women, the estimates are 0.210 and 0.182 for those same cohorts. Persistence was greater for men in the early cohort, but that difference had largely disappeared by 1920.

Using a sample of our linked father-daughter and father-son pairs for which both the husband and wife were successfully matched, we estimate the degree to which the economic status of fathers are correlated. A high degree of marriage within social background as proxied by occupational status is found, although less than in modern estimates (Charles et al., 2013), but more than using the pseudo-linking method for the same time period (Olivetti et al., 2020). The rank-rank parameter between fathers is 0.196 for the 1850-1870 cohort and 0.296 in the 1880-1900 cohort, a 51 percent increase. In our sample, the likelihood that women would marry out of the economic class of their fathers decreased over time.

We develop a simple theoretical framework following recent work by Espín-Sánchez et al. (2023) to understand the relationship between marital sorting and intergenerational mobility. Using counterfactual assortative matching estimates obtained from the model, we establish that the mobility of women in the 1860-80 and 1880-1900 cohorts would have been significantly higher than observed if they faced the lower level of assortative matching experienced by women in the earlier 1850-1870 cohort. Further research is required to better understand the forces driving the changes in assortative mating we find over the course of the nineteenth century.

The rest of this paper is structured as follows. Section 2 provides background on marriage trends and mobility in our cohorts. Section 3 develops a theoretical framework to capture the relation between assortative matching and intergenerational mobility. Section 4 describes our data sources and linking method. Section 5 presents the main results on intergenerational mobility, the evolution of sorting in the marriage market, and uses the theoretical framework to determine counterfactual intergenerational mobility at different levels of assortative matching. We explore heterogeneity in mobility and marital sorting across groups in Section 6. Section 7 concludes.

2 Literature

The literature on modern intergenerational mobility for sons is extensive. Solon (1999) and Black and Devereux (2011) provide thorough reviews, which suggest that the United States in the

[3]More specifically, the decennial census does not record an occupational related to market work for the vast majority of married women, who are most often recorded as some form of "housekeeper".

late twentieth century had relatively low mobility for sons compared to other developed European nations. Until recently the mobility of daughters has received less attention (Chadwick and Solon, 2002; Jäntti et al., 2006), and when estimated, differences in mobility between the sexes depends on dataset, method, and time period. Mazumder (2005) uses IRS records and finds that estimates of the intergenerational elasticity (IGE) for women in the United States are similar to those of men. Chadwick and Solon (2002) using the PSID and Jäntti et al. (2006) using the NLSY find that daughters are more mobile than sons, although a significant amount of persistence still exists.

In contrast, historical estimates for the United States suggest that cohorts in the nineteenth and mid-twentieth centuries faced less intergenerational persistence than the modern U.S. (Ferrie, 2005; Long and Ferrie, 2013; Feigenbaum, 2017; Song et al., 2020). However, recent work by Ward (2022) finds that the trend reverses when measurement error in father's status is accounted for; the level of intergenerational persistence actually declined between the nineteenth and twentieth centuries.

This long decline in mobility comes during a period of significant structural change in the U.S. economy. Railroads lowered the transport costs of goods and decreased the cost of migration. Millions of Europeans immigrated to the United States as a land of perceived opportunity. Industrialization lured individuals to urban areas, prompting the rise of city life. This period, commonly known as the Gilded Age, can easily be considered as a period of impressive growth but also may have developed the roots of a pivotal divide between individuals in the United States who rose to the top and fell to the bottom. The literature lists a number of potential causes of the decline in economic mobility: changes in the selection of movement out of farming, propensity of internal migration, regional differences in wages, assimilation patterns of immigrants across generations, returns to human capital, and trends in inequality (Long and Ferrie, 2013; Abramitzky et al., 2014; Olivetti and Paserman, 2015; Salisbury, 2014).

Even more than that for the modern period, the historical literature has focused on father-son transitions without much attention to father-daughter transitions. Linkage of adults to their childhood household in a prior census requires a search for similar given names and surnames. In the case of women in the nineteenth century U.S., of which the majority are married, surnames change between census at the time of marriage. A direct census to census link of an adult married woman to her father in a previous pre-marriage census is impossible without additional information.

Instead of directly linking fathers and adult daughters across censuses, Olivetti and Paserman (2015) make a major advance by creating pseudo-links based on the fact that given names convey socioeconomic status. They identify the occupational income score (*occscore*) of an individual in a specific census and calculate the average occupational income score for all fathers in the previous census who have a child with that individual's name. For example, for a daughter named "Katherine", the income level of her father is calculated as the average income of all fathers in the previous census with a daughter named "Katherine". They use these pseudo-links to calculate estimates for intergenerational elasticity of income for both men and women from 1850-1940. Their method does not capture a "true" estimate for elasticity of income because of the absence of direct links between generations. Rather, it calculates a measure that can be compared over time, assuming equal bias over time and across genders. They have since extended the pseudo-linking method to include three generations and found that grandparents do matter in mobility level of their children and grandchildren (Olivetti et al., 2016), and to examining the probability of marriage and marital sorting based on the father's economic status (Olivetti et al., 2020). Due to the methodological differences, our estimate of intergenerational mobility will not be comparable in magnitude, but will be useful to compare the mobility of women relative to men and the trends over time. Our direct linking method, and the results we obtain, are complementary to those found by Olivetti and Paserman (2015)[4].

We find that women are more mobile than men in our sample, with a gap that disappears over time. What causes this difference between sexes in economic persistence? The extent of marital sorting by socioeconomic background has important implications for the persistence of economic status of women across generations. How much of role does marriage have to play as a source of social mobility? Building off of Becker and Tomes (1979), theoretical frameworks have been developed and applied to this question by Oddbjørn et al. (2008) and Ermisch et al. (2006), both finding a large role for assortative mating. For Germany and the U.K., Ermisch et al. (2006) finds that assortative mating on earnings potential accounts for 40 to 50 percent of the correlation in parental and total family income. Using own and total family earnings, Oddbjørn et al. (2008) finds a high correlation between parental earnings and total family earnings for women, but in the U.S., a small correlation with the woman's own earnings. They argue that while sorting in marriage explains the high persistence in total family earnings, the labor supply decisions of married women (i.e. the income effect from high earning husbands) reduces the correlation with the wife's own earnings.

These studies underscore the modern literature's focus on assortative mating based on personal labor market or human capital characteristics of the spouses. Unfortunately, the economic context of married women in the nineteenth century and the lack of information about education in historical censuses excludes this mode of analysis followed in the modern literature. Instead, we note that marriage was the prime vehicle for economic mobility for women during this period, and that sorting based on social background – originating from preferences or availability of potential mates through social interaction – is important to understanding how much of parents' socioeconomic conditions are transferred to their children.

3 Theoretical Framework

We begin by developing a simple theoretical framework of the interaction between marriage and intergenerational mobility by adapting the model in Espín-Sánchez et al. (2023). The model helps us tie together the empirical relationships estimated in the data and underlying parameters

[4]In the supplemental appendix, we show how estimates differ by applying the methodology of Olivetti and Paserman (2015) to our sample of direct-linked data. Under certain sample restrictions of the pool of fathers, the two methods provide consistent stories for the evolution of differences between the sexes in intergenerational mobility.

of inheritance and marital sorting. Moreover, we use the model to conduct counterfactual analysis of what women's mobility would have looked liked under different levels of assortative mating. The inheritance of income from one generation to the next is captured by the gender specific inheritance equations:

$$
\begin{aligned}
X_i^h &= \beta_f X_i^f + \beta_m X_i^m + e_i^h \\
X_i^w &= \beta_{fl} X_i^{fl} + \beta_{ml} X_i^{ml} + e_i^w
\end{aligned} \tag{1}
$$

, where h and w denote husband and wife, f and m denote the mother and father of the husband, and fl and ml indicate father- and mother-in-law of the husband (father and mother of the wife). The equations in 1 make explicit that in its most flexible form, inheritance can differ by both gender and the child-parent gender interaction (Espin-Sanchez et al., 2022).

Men and women sort into marriages based on incomes, which can be summarized in a linear relationship as:

$$
X_i^w = \gamma X_i^h + v_i^h \tag{2}
$$

, where v_i^h is an uncorrelated error term. The sorting equation 2 also holds in the parent generation where m and f replace w and h. This sorting equation can be substituted into the inheritance equations of 1 to get:

$$
\begin{aligned}
X_i^h &= [\beta_f + \gamma\beta_m] X_i^f + \beta_m v_i^f + e_i^h \\
X_i^w &= [\beta_{fl} + \gamma\beta_{ml}] X_i^{fl} + \beta_{ml} v_i^{fl} + e_i^w
\end{aligned} \tag{3}
$$

Income is observed with noise or through some observable proxy Y for individual i:

$$
Y_i = X_i + u_i \tag{4}
$$

, where u_i is an error term uncorrelated with X. With observable measures of standing for both men and women, Y_i^h and Y_i^w, we can estimate father-child intergenerational mobility regressions of the form:

$$
\begin{aligned}
Y_i^h &= \alpha + b_f X_i^f + e_i^h \\
Y_i^w &= \alpha + b_{fl} X_i^{fl} + e_i^w
\end{aligned} \tag{5}
$$

, where b_f and b_{fl} are the estimated gender-specific IGE or rank-rank coefficients. These are the empirical correlations estimated from modern data and which the literature refers to as intergenerational mobility (Jácome et al., 2021; Ward, 2022). Note that the model shows they are different from the underlying structural inheritance terms. Estimated IGEs are a combination of the inheritance terms and the degree of marital sorting on X: $b_f = \beta_f + \gamma\beta_m$ and $b_{fl} = \beta_{fl} + \gamma\beta_{ml}$.

In historical data, we do not observe married women's own economic standing as measured by their own income or occupation, Y_i^w, meaning that the second equation of 5 cannot be estimated.

However, we can estimate the husband-father-in-law IGE with:

$$Y_i^h = \alpha + b_{fl}^* X_i^{fl} + e_i^{w*} \tag{6}$$

where $b_{fl}^* = \beta_f \mathbb{E}[X_f X_{fl}] + \beta_m \mathbb{E}[X_m X_{fl}])$. First, the IGE estimate from the husband-father-in-law regression is commonly interpreted as "women's mobility" in the literature (Olivetti and Paserman, 2015; Jácome et al., 2021). It is the combination of inheritance parameters and the correlation in income between the father and father-in-law. The empirical association in income between the fathers of the spouses can be thought of as one measure of marital sorting based on family status (Olivetti et al., 2020), which can be estimated with the following equation:

$$Y_i^f = \alpha + \lambda Y_i^{fl} + v_i \tag{7}$$

where λ is our estimate of $\mathbb{E}[X_f X_{fl}]$. Second, estimates from equation 6 can be used to recover the structural marital sorting parameter γ which we can then use to estimate counterfactual women's mobility if the strength of marital sorting remained constant over time.

With two generations of data with outcomes measured only for men, Espín-Sánchez et al. (2023) show that a multigenerational structural model of marriage and social mobility can be captured by a system of three equations:

$$b_f = \beta_f + \gamma \beta m \tag{8}$$

$$b_{fl}^* = \lambda(\beta_f + \beta m) \tag{9}$$

$$\gamma = \lambda(\beta_f^2 + \beta_m^2 + \lambda \beta_f \beta_m) \tag{10}$$

b_f, b_{fl}^*, and λ can be estimated with the data, and we are left with three equations and three unknown structural parameters $(\beta_f, \beta_m, \gamma)$. Note that to point identify these parameters we need to assume that assortative matching in the parent's generation is the same as in the children's generation, and that the empirical correlation in status between all four parents across marriages are equal.[5]

Our end goal with the model is twofold. First, it allows us to use the estimated empirical associations (b_f, b_{fl}, b_{fl}^*) to find the structural assortative matching parameter (γ) and plot how it changes over time. Second, we can use estimates of the three structural parameters to conduct counterfactual analysis on women's intergenerational mobility. Specifically, after having solved for the inheritance and sorting parameters for two separate cohorts, we can plug the estimate of λ from cohort 1 into equation 9 to get a sense of what women's mobility would have been in cohort 2 if marital sorting had remained at the level of cohort 1.

[5] $\mathbb{E}[X_f X_{fl}] = \mathbb{E}[X_f X_{ml}] = \mathbb{E}[X_{fl} X_m] = \mathbb{E}[X_m X_{ml}]$

4 Data

We measure the evolution of intergenerational mobility across the nineteenth and early twentieth centuries for both men *and* women using a new dataset of linked census observations. The novel part of our strategy is to use the listed maiden name recorded in marriage registers to link adult married women to their fathers in a childhood census. As such, our record linkage consists of two match procedures. Couples from the marriage index are first linked to a post-marriage decennial census to observe their economic status as an adult. The sample of successfully matched married couples is then matched to the childhood household in a pre-marriage census to observe the father's economic status. This second match is completed separately for the husband and wife. The economic status of fathers and adult children is measured using the reported occupation from the census and an associated occupational income or wealth score, which we describe in more detail below. Our dataset consists of four cohorts of father-son and father-daughter pairs linked between two censuses: 1850-70, 1860-80, 1880-1900, and 1900-20.[6]

4.1 Linking Sons and Daughters to Fathers

We begin with a digitized set of marriage registers for the Commonwealth of Massachusetts that cover the universe of registered marriages for the period 1841-1915.[7] For both the wife and husband, the marriage index lists the full name, year of birth, and full names of both sets of parents. We construct four cohorts of marriages to link to census records from the population of marriages: 1850-70, 1860-80, 1880-1900, and 1900-1920.[8] We restrict the sample such that at least one of the spouses would have been 20 years or younger in the year of the childhood census, so that it is more likely to observe them in their childhood home.

The couples in each cohort are first matched using names and years of birth to a subsequent U.S. decennial census to measure the post-marriage economic status of the couple. Marriages that occurred between 1850 and 1869 are linked to the 1870 census, those that occurred between 1860 and 1879 are linked to the 1880 census, and so on.[9] Of the successfully matched couples to a post-marriage census, each spouse is then separately linked to the childhood census taken 20 years prior using names, year and state of birth, and names of parents. Figure 1 is helpful to visualize the stages of our double-match record linkage process.

We use a probabilistic match method to link records by taking advantage of the wealth of information we have on individuals, their spouses, and their parents.[10] In our dataset, for the first

[6] The 1890 Decennial Census population schedules burned in a fire in 1921. Thus, we cannot create 1870-90 or 1890-1910 links.

[7] FamilySearch.org digitized the microfilmed copies of the marriage indices and kindly provided us access (FamilySearch, 2016).

[8] Household information is drawn from the complete count decennial censuses from 1850-1920 deposited by Ancestry.com and IPUMS with the NBER

[9] 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.

[10] This is in contrast to the historical intergenerational mobility literature which has primarily used automated linking procedures that rely on the exact matching or unique matching of names and functions of names across two

match of the married couple to a post-marriage census there are five dimensions over which to match: surname, two given names, and two birth years. The issue becomes how to balance small differences across each of these dimensions when choosing a link. The problem is even more severe for the second match from the post-marriage census to the pre-marriage census. In this case, we match on four strings – child's given name and surname, father's given name, mother's given name – and the child's year and state of birth.

We implement Feigenbaum (2016)'s supervised machine learning record linkage algorithm to take advantage of the rich information on family structure provided by the marriage registers. We train a model and choose validated parameters to balance transcription errors across the numerous name strings.[11] We briefly describe the matching procedure here with more detailed information provided in the supplemental appendix. The method developed by Feigenbaum (2016) uses machine learning to train an algorithm to replicate and scale up the careful manual-linking that would be done by a researcher. The process starts by selecting a random sample of couples from the marriage certificates and manually constructing the true links for this sample. A portion of the true links are used to train a model, in this case a logit regression, which is then cross-validated on the remainder of the true links held out of the training sample. The estimated model is then used to predict matches for the remainder of the marriage certificate sample for which we did not hand code "true" matches.

As seen in Figure 1, we match 294,105 of the 1,220,434 couples in the marriage registrations that meet our sample restrictions, for a 24 percent match rate in the first step. Of the successfully matched couples to the post-marriage adult census, we link the husband and wife individually to their childhood household in a pre-marriage childhood census. The final dataset includes 90,947 father-son pairs and 95,358 father-daughter pairs. The match rates for the second match are between 30 and 31 percent, respectively. The total match rates over both steps are 7 and 8 percent. In the marital sorting analysis, we use a sample for which *both* the husband and wife from a single marriage registration are linked successfully to their fathers. In this sample, we have 38,760 couples (3 percent match rate).

We would like to highlight that linking both spouses to their fathers is an innovation in our work. For this sample, we get a lower match rate than typical in the literature on father-son mobility which requires a link between two data sources, instead of three in our case. In our analaysis of the association of status between fathers (i.e. father/father-in-law), both spouses must be linked to their fathers independently. Note, that this requirement limits, in practice, the analysis to *native-born* couples. Observations with an immigrant spouse are unlikely to enter our sample because we are unlikely to find the immigrant father's economic standing in the childhood home unless the immigrant spouse migrated with their father as a young child. Thus, the true match rate for the double matched couples is actually higher than reported, because the denominator used is larger

census waves (Long and Ferrie, 2013; Abramitzky et al., 2012; Collins and Wanamaker, 2014, 2015, 2022; Abramitzky et al., 2020; Ward, 2022).

[11] Linkage procedures requiring exact matches on all strings are too conservative in our context with more name strings than is typical, leading to more false negatives and a smaller sample size than is necessary.

than in reality; it should be limited to marriage registrations where both spouses are native-born. However, we cannot make this limitation in our marriage index data.

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

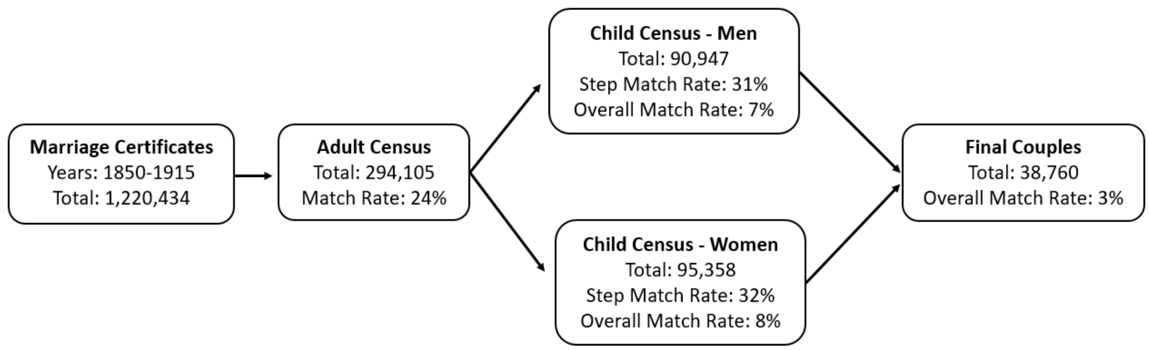


Table 1: Summary of sample sizes and match rates across the two linkage steps

	Pool of marriages	Matches: Marriage to adult census		Matches: Adult census to child census			Both matched to child census
	Obs	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	0.06	6,330
				W: 14,979	0.38	0.07	
1860-80	250,598	56,745	0.23	M: 19,214	0.34	0.08	7,887
				W: 19,359	0.34	0.08	
1880-00	372,213	91,235	0.25	M: 27,743	0.30	0.07	11,034
				W: 27,134	0.30	0.07	
1900-20	379,765	106,572	0.28	M: 30,847	0.29	0.08	13,509
				W: 33,866	0.32	0.09	
Total	1,220,434	294,105	0.24	M: 90,947	0.31	0.07	38,760
				W: 95,358	0.32	0.08	

4.2 Summary Statistics and Selectivity of the Sample

Nonrandom selection of records into the linked sample is common (Bailey et al., 2018; Abramitzky et al., 2021). We assess the representativeness of our matched sample in two ways. For the first match, we compare sample means of characteristics from the full pool of marriage certificates to the set of linked certificates, and then characteristics for all Massachusetts born respondents aged

0-20 in the childhood census to the same characteristics for the linked sample. In the case where characteristics differ between the population and the linked sample, we follow the advice of Bailey et al. (2018) and reweight the linked sample to resemble the observable characteristics of the full sample using the inverse propensity-score weight (IPW). For the full sample, we use characteristics contained in the marriage registers, and functions of those characteristics, to predict the likelihood of a finding a match. The weights thus make the linked sample mimic the population of marriages registered in Massachusetts. In some specifications we restrict the sample to observations born in Massachusetts. In a second scheme, we reweight the linked sample to mimic the population of Massachusetts-born men and women from a childhood cohort. See Appendix B.1 for more details on the construction of weights.

Results assessing the representativeness of the linked sample to the population of registered marriages are shown in Table A5. The matched sample and full pool of marriage certificates are similar along a number of string characteristics and the age at marriage. Although the differences between the population and the unweighted linked sample are highly statistically significant because of our large sample, none of the magnitudes of the differences appear meaningful. Tables A6 through A9 repeat the exercise separately by marriage cohorts.

4.3 Measuring economic status

A major distinction between the modern and historical mobility literature is the lack of income information in the historical censuses. The U.S. Decennial Census did not record income prior to 1940, and only occasionally records household wealth (1860 and 1870). Subsequently, the literature estimates historical mobility using proxies for individual income inferred from the reported occupation of an individual (Ferrie, 2005; Long and Ferrie, 2007, 2013; Olivetti and Paserman, 2015; Feigenbaum, 2017; Collins and Wanamaker, 2014, 2022; Ward, 2022). Our preferred approach is to follow Ferrie (1999) and Collins and Zimran (2023) and construct an occupational *wealth* score using the value of total property owned reported by the head of household in the 1870 Decennial Census complete count microdata (Ruggles et al., 2017). We calculate the mean value of total property wealth (real plus personal) for all white male observations between the ages of 18 and 65 in each occupation, region, and immigrant/non-immigrant cell. When less than 20 observations are present in the occupation-region-immigrant cell, we first use the mean wealth in the occupation by region cell, and then the national occupation mean if needed. Three limitations arise from using wealth scores. First, respondents were asked to report gross values, not net of any debt. Thus, the wealth scores will be biased to the extent that property was *differentially* debt-financed across occupations. Second, the wealth score measure does not capture any changes in the wealth distribution between occupation-region cells between 1850 and 1910. Finally, the wealth score is unable to capture the within-cell variation in wealth.

Much of the analysis emphasizes the rank-rank mobility of father-son and father-daughter pairs, capturing the idea of "upward" or "downward" mobility in the wealth distribution relative to one's peers. To capture the changing occupational structure of the American economy over the nineteenth and early twentieth centuries, we take each observation's income score and place it in the cohort specific national distribution, and calculate its percentile rank. For example, fathers of the 1850-70 cohort are given a rank in the national distribution in 1850, whereas the sons in the 1850-70 cohort are placed in the 1870 national distribution. Recalculating occupational ranks in each census implies that a son that stays in the same occupation as their father does not necessarily keep the same rank. That son could be upwardly- or downwardly-mobile. Similarly, a son that transitions to an occupation with a higher wealth score than his father does not necessarily improve the rank. The entire distribution of occupations could have been upgraded so that the son's rank stayed the same or even decreased relative to the father's.

We further show that our results are similar when a number of additional proxies for income or economic status are used. Appendix C shows figures for mobility and marital sorting using the IPUMS *occscore* variable derived from the 1950 income distribution, income scores from a 1901 Cost of Living report and the 1900 Census of Agriculture for farmers, the literacy based occupational score (song score) (Song et al., 2020), a race and region adjusted song score (Ward, 2022), and an immigrant and region specific song score.

5 Intergenerational Mobility and the Marriage Decision

In this section, we estimate gender differences in and the long-run trend of relative mobility, and link those changes to sorting in the marriage market. We compare other estimates in the literature from different time periods and source data to our own. To begin, we briefly discuss estimation methods, then report results, and finally assess the robustness of the mobility estimates. Next, we estimate the strength of assortative matching in the marriage market, and assess its importance for trends in women's intergenerational mobility.

5.1 Intergenerational Mobility

We measure intergenerational relative mobility with rank-rank regressions of an adult's economic status rank on their father's rank, allowing the intercept and slope to differ by gender:[12]

$$Adult\ Rank = \alpha + \beta_0 Woman + \beta_1 Rank\ Father + \beta_2 Woman \times Rank\ Father + \epsilon \tag{11}$$

The intercept term α measures the absolute rank mobility of men born to the fathers at the very bottom of the wealth score distribution. The differential absolute rank mobility for women is measured by β_0. Our main coefficients of interest are β_1 which provides the relative rank mobility for men, and β_2 which gives any estimated difference in mobility for women. The regression is estimated separately on each of the four cohorts. To address life-cycle bias, all estimates include

[12] We provide IGE estimates in the appendix. The log-log estimate of the IGE imposes a linear relationship between father's log economic status and the child's log economic status. However, recent work shows this linear relationship fails to hold in the upper and lower tails of the income distribution (Corak and Heisz, 1999; Chetty et al., 2014). We follow Dahl and DeLeire (2008); Chetty et al. (2014); Mazumder (2014); Collins and Wanamaker (2022) and estimate a rank-rank regression in which the imposed linear relationship fits the modern data.

a quartic in father's age (relative to age 40) and a quartic in the adult child's age at the time that economic status is measured (relative to age 40).[13] The main text reports mobility estimates unweighted for the likelihood of linking a record. The appendix reports results with a number of weighting schemes, and our results are not sensitive to weighting.[14]

For women, we follow the historical literature and report the association between her husband's rank and her father's rank, and call it women's mobility (Olivetti and Paserman, 2015). This is different from how modern estimates are interpreted. Ideally, we could capture a direct estimate of the association of a woman's status with her father's by regressing the wife's own personal status on her father's. However, the majority of married white women did not participate in market work "full-time" like men, and more importantly did not report an occupation in the census. Thus, we must estimate female mobility through the male connected with her in each generation (i.e. regressing her husband's rank on her own father's rank). Modern estimates of the rank-rank association for women suffer from a different set of issues, mainly that the labor supply and educational choices of married women are jointly determined with the income and wealth prospects of potential husbands. This issue is less relevant in the nineteenth century. Note that what we call "mobility" is a mixture of the assortative marriage patterns and the transferability of traits from parents to child, as discussed earlier in the theoretical framework.

We start by discussing the evolution of mobility across time, and then put any differences by gender in context. Figure 2 plots the mobility estimates for men (β_1) and for women ($\beta_1 + \beta_2$). Results for our sample of linked men suggest a rank-rank parameter for 1850-1870 of 0.231. An increase in persistence to 0.249 in the 1860-70 cohort was followed by a large decrease to 0.209 in the 1880-1900 cohort. Mobility was essentially unchanged for the 1900-20 cohort. In sum, we see a brief decrease in mobility for men followed by a large and sustained increase in mobility for the post-1880 cohorts. The decline of .034 between the first and last cohorts is statistically significant and was roughly 15 percent of the 1850-70 level. The even bigger decline of .052 between the 1860-1880 and 1900-1920 cohorts was 21 percent of the 1860-1880 level.

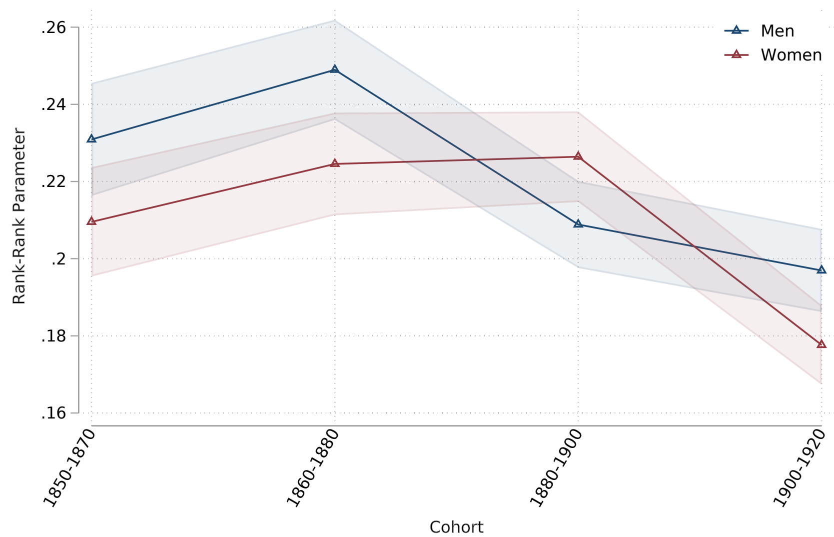
The linked sample of married women had a different mobility experience than men over the nineteenth century. The rank-rank parameters increase from 1850-70 to 1860-80 just like for men, from 0.210 to 0.225. However, while the men's rank-rank parameter begins its decline with the 1880-1900 cohort, women's persistence continues to increase, or at least stays constant at around 0.227. Women's mobility did not begin its increase until at least the 1900-1920 cohort, when we

[13] Life-cycle bias may enter our estimates from the fact that not all fathers (or adult children) are observed at the same age (Grawe, 2006). Haider and Solon (2006) find that life-cycle bias is minimized when income is measured at age 40. In our case, we observe occupation at a single point in an individual's life. To address the concerns of life-cycle bias, we first include a set of controls to capture the life-cycle pattern of economic status (Aaronson and Mazumder, 2008; Lee and Solon, 2009).

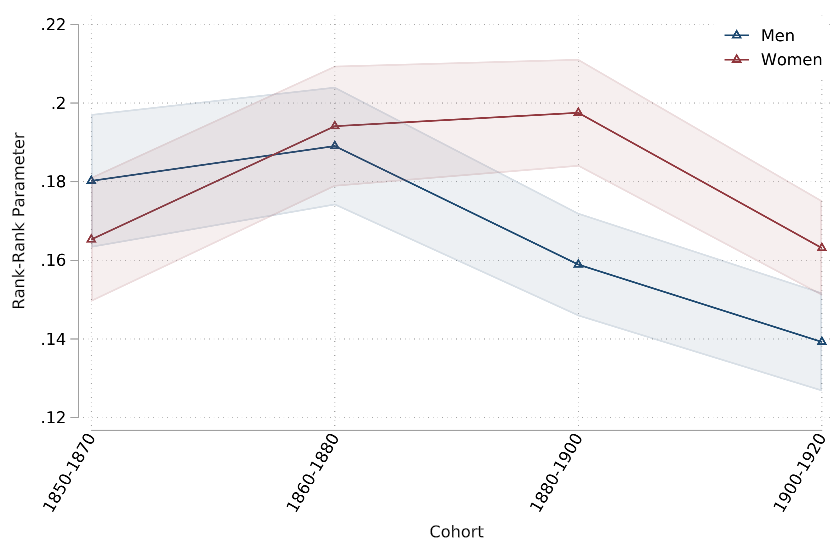
[14] Bailey et al. (2018) and Abramitzky et al. (2021) suggest methods to assess the representativeness of a linked sample. When unrepresentativeness poses a problem, reweighting a linked sample to mimic the original base population of records may be a solution. In our case, the choice of a population and its characteristics to reweight the linked sample is not straightforward. We begin with the full set of registrations of marriages occurring in the Commonwealth of Massachusetts. The first weighting scheme we apply is to make our linked sample mimic the observed characteristics based on names and functions of names and ages in the population of marriage records. However, this does not necessarily capture the underlying representatives of economic status that the researcher cares about.

Figure 2: Intergenerational Mobility Estimates: Rank-Rank Correlations

(a) Full Sample



(b) Massachusetts Born



Notes: Each entry is an estimate of the rank-rank parameter from separate regressions of own wealth score rank on father's wealth score rank for men and women and by cohort. The mobility estimate for men is β_1 and the estimate for women is $(\beta_1 + \beta_2)$ from equation 11. Panel A plots results for the full sample, whereas Panel B restricts the sample to observations born in Massachusetts. Shaded areas plot 95 percent confidence intervals. All regressions include as controls a quartic in father's age and a quartic in husband's age, both at the time economic status is measured. Regressions are unweighted.
Sources: 1850-1920 Decennial Census data from Ruggles et al. (2017). Marriage certificates from *FamilySearch.org*.

estimate the rank-rank parameter as 0.178 - a decline of 0.032 (15 percent) from the 1850-70 cohort of women.

We directly compare economic mobility between men and women in Table 2, which reports the estimated coefficients from equation 11 for each cohort. The intercept (α) suggests that the mean rank of men born to fathers at the very bottom of the wealth score distribution had a mean expected rank ranging from the 54th to the 58th percentile across all four cohorts. Moreover, the positive values of β_0 suggest women experienced small, but statistically significant, higher levels of absolute mobility in all cohorts except for 1880-1900. Moving to the top two columns, β_2 shows that women experienced lower persistence (higher mobility) than men in, again, all but the 1880-1900 birth cohort. These differences are highly statistically significant. In the 1850-1870 cohort, a woman's association in rank with her father was 2.1 percentiles less than a man's with his father. The difference was large and meaningful. Women's rank-rank parameter was 91 percent of that of men's in the 1850-1870 cohort – 0.210 vs. 0.231, a difference equivalent to 70 percent of the entire increase in mobility experienced by men between the 1850-70 and 1900-20 cohorts. The gender gap widened in the 1860-80 cohort with women becoming even more mobile than men. The clear outlier is the 1880-1900 birth cohort, where women abruptly lost their mobility advantage relative to men. The gap flipped signs and turned positive. Women did not share in the mobility gains experienced by men between the 1860-80 and 1880-1900 cohorts. However, women's mobility returned to being higher than men's by the 1900-1920 cohort, with the gap similar to that in the initial 1850-70 cohort.

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. We focus on the Massachusetts-born sample for much of the remainder of the text, as is done in Panel B of Table 2. The level of the rank-rank parameter (β_1) is roughly 22 to 29 percent lower for men in the Massachusetts-born sample relative to the full sample; the association of economic status between father and son was lower for those born in the state compared to in-migrants. The change in mobility for men was similar for both samples. Mobility declined slightly moving from the 1850-70 to 1860-1880 cohort, but then made large increases across the next two cohorts. Additionally, absolute rank mobility was also higher in the Massachusetts born sample - (α) was 8 to 10 percent higher. Massachusetts-born women, however, began with slightly more mobility than men in the 1850-1870 cohort, but had fully converged with Massachusetts-born men by the 1860-1880 cohort, twenty years earlier than in the full sample. By the 1880-1900 cohort and later, Massachusetts-born women experienced a significantly higher rank-rank parameter than Massachusetts-born men.

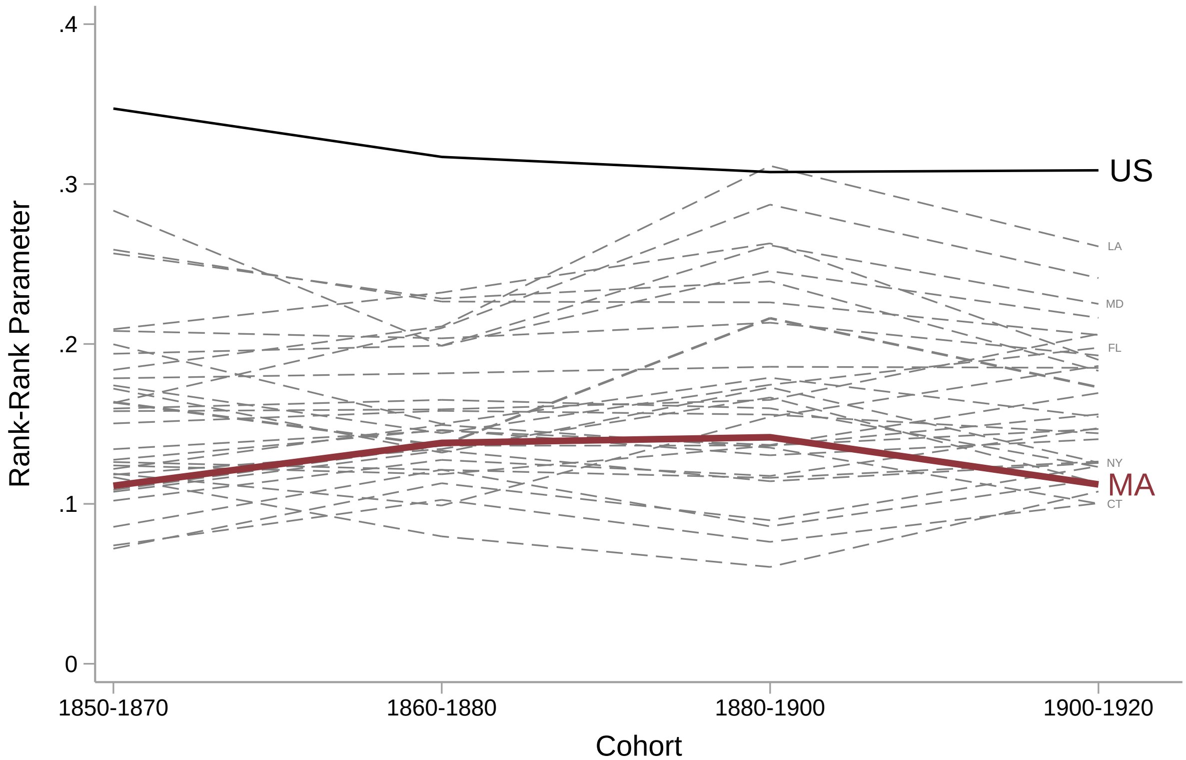
Massachusetts in the nineteenth and early twentieth centuries was a relatively mobile place. Hence, our estimates of mobility for men are not directly comparable to those in the recent literature using nationally representative samples of men covering a similar time period (Long and Ferrie, 2013; Song et al., 2020; Ward, 2022). To show that the Massachusetts-born population of men experience a different level and change in mobility than the rest of the nation, we use a standard record-linkage methodology and our constructed wealth scores. Figure 3 estimates rank-rank parameters for each state of birth and for the nation using a set of linked white men from the Census Linking Project (Abramitzky et al., 2020). Note that this figure includes married and single

Table 2: Men's and Women's Intergenerational Mobility Estimates

	1850-1870	1860-1880	1880-1900	1900-1920
	Panel A: Full Sample			
Father Rank (β_1)	0.231	0.249	0.209	0.197
	(0.007)	(0.007)	(0.006)	(0.005)
Father Rank x Woman (β_2)	-0.021	-0.024	0.018	-0.019
	(0.010)	(0.009)	(0.008)	(0.008)
Intercept (α)	55.910	53.959	58.102	56.535
	(0.479)	(0.436)	(0.388)	(0.372)
Woman (β_0)	1.418	1.580	-1.962	1.350
	(0.669)	(0.621)	(0.558)	(0.512)
No. of Observations	27,355	34,451	46,226	53,103
	Panel B: Massachusetts Born			
Father Rank (β_1)	0.180	0.189	0.159	0.139
	(0.009)	(0.008)	(0.007)	(0.006)
Father Rank x Woman (β_2)	-0.015	0.005	0.039	0.024
	(0.012)	(0.011)	(0.010)	(0.009)
Intercept (α)	60.333	59.438	63.409	62.101
	(0.546)	(0.505)	(0.453)	(0.435)
Woman (β_0)	1.027	-0.891	-3.853	-2.895
	(0.756)	(0.724)	(0.656)	(0.605)
No. of Observations	20,320	25,001	31,982	36,811

Notes: This table reports coefficients from equation 11 estimated separately on each cohort with heteroskedasticity-robust standard errors in parentheses. Panel A reports results from regressions on the full sample of linked data, whereas Panel B limits the sample to observations born in Massachusetts. Both regressions include quartics in the age of the father and the age of the husband at the time economic status is observed. Reported coefficients are evaluated at age 35 for both the father and the adult. The Intercept (α) measures the absolute rank mobility of children born to fathers at the very bottom of the wealth score distribution. The (β_0) on a Woman indicator measures the difference in absolute mobility for women relative to men. The two main coefficients of interest (β_1) on Father's Rank and (β_2) on Father's rank x Woman measure relative mobility for men and the differential mobility experienced by women.
Sources: 1850-1920 Decennial Census data from Ruggles et al. (2017). Marriage registrations from *FamilySearch.org*.

Figure 3: Intergenerational Mobility of Massachusetts-born Men Relative to the Nation



Notes: This figure plots rank-rank parameters 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 parameter for men born in other states (limited to states in the 1850 census). Record links were created using the Census Linking Project and the conservative ABE-method with exact name matching. Regressions are unweighted. Economic status for father and son is measured using the wealth score ranks discussed in the Data section. *Sources:* Abramitzky et al. (2020) and Ruggles et al. (2017).

men. Persistence estimates for Massachusetts-born men are in the bottom third of state rankings in each cohort. Moreover, individual states as subgroups do not follow the overall national trend in mobility.[15]

Two recent papers are more appropriate with which to compare our results. The mobility estimates for our direct-link sample match the patterns found in the pseudo-linking results in Olivetti and Paserman (2015), with some caveats.[16] While their national and regional estimates

[15]The slope parameter for the entire nation can be above that of all the individual subgroup slope parameters. 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 et al., 2021). In our case, there are large regional differences in wealth scores, and to the extent that fathers and sons tend to live in the same region, the full population slope parameter will be higher than parameters for individual states.

[16]Summarizing their main findings, Olivetti and Paserman (2015) find that women were *less* mobile than men in a national sample measured across census records from 1850-1870. Both sexes experience a decrease in mobility over

are at odds with our own, the differences vanish when we apply the pseudo-linking procedure to our our linked sample and estimation methods. In brief, we find that the broader the pool of fathers used to construct a pseudo-linked occupational income score, the more likely estimates diverge from those in the direct-linking procedure.[17] Again, we are concerned with a comparison of the *trend* and *sex differences* within a linking method. A second recent paper, Althoff et al. (2023) uses the Social Security Numident file and its information on maiden names to link women to a childhood home in the census. Again, our level estimates are lower than their nationally representative sample, but are in agreement that women were more mobile than men.

To be clear, we do not claim that our mobility results for men and women are representative of the trends for the entire United States. Our purpose is to link changes in assortative matching in the marriage market to mobility outcomes. Marriages in Massachusetts prove to be a fruitful context to explore this link. We compare our results to a standard linking procedure, the standard ABE-method with NYSIIS adjustment of names Abramitzky et al. (2021) with links taken from the Census Linking Project Abramitzky et al. (2020). Table 3 reports results for the standard census-to-census linkage method for men born in Massachusetts in Panel A. Panel B and Panel C split the sample by single men and married men, respectively. Panel D reports are main results for Massahcusetts-born men from Table 2 for reference. We see that the levels and trends are similar for the married men using the Census Linking Project links as in our sample of links using the marriage certificates. The lone difference is in the timing of the decline in persistence. In our sample, the decline begins between the 1860-1880 and 1880-1900 cohorts, whereas in the CLP links it occurs between the final two cohorts.

The general patterns of mobility across cohorts are not sensitive to the choice of alternative measures of occupational status. Appendix Table A15 reports the mobility coefficient estimates for men and women using four different constructions of occupational scores: occupational earnings from the 1901 Cost of Living Survey, income scores based on the earnings distribution of the 1940 decennial census following Abramitzky et al. (2021), 1950 decennial census income scores, and a human capital based score – "Song score" – from Song et al. (2020). There is no single ideal proxy for economic status. Trade-offs exists for each measure of status. Some are further away in time from the 70 year period under consideration, such as the 1940 and 1950 based income scores. Others do not allow for differences in status by age, region, or immigration status – the 1901 Cost of Living Survey, 1950 occupational income score, and Song score. Mobility estimates for all choices of economic status measures follow the inverted-U shape for intergenerational persistence for men that we found in our main results in Table 2. The timing of the peak, however, is shifted from the 1860-80 cohort to the 1880-1900 cohort. The results suggest women are more mobile than men in all cohorts for three of the measures. In the case of the 1901 Cost of Living Survey based scores,

the course of the nineteenth century, with full convergence between the sexes in a linked 1900-1920 sample.

[17]Reported results in Olivetti and Paserman (2015) are for the entire northeast, use IGEs instead of rank-rank specifications, use a national pool of fathers to estimate name based income differences, and use the 1950 based IPUMS *occscore* variable, all of which are reasonable choices to make. In Appendix D we show that the gaps in mobility for men and women as well as the trends match our those on our direct-linked sample when using a pool of fathers residing in the Northeast to measure name based occupational income scores in the pseudo-linking method.

Table 3: Mobility for Massachusetts-born Men by Marital Status

	1850-1870	1860-1880	1880-1900	1900-1920
	Panel A: CLP Links - All Men			
Father Rank (β_1)	0.115	0.139	0.141	0.111
	(0.004)	(0.004)	(0.004)	(0.004)
No. of Observations	36,504	39,558	47,518	55,686
	Panel B: CLP Links - Single Men			
Father Rank (β_1)	0.091	0.119	0.112	0.097
	(0.006)	(0.006)	(0.006)	(0.005)
No. of Observations	17,444	18,816	25,165	27,277
	Panel C: CLP Links - Married Men			
Father Rank (β_1)	0.142	0.157	0.168	0.121
	(0.006)	(0.006)	(0.006)	(0.005)
No. of Observations	19,060	20,742	22,353	28,409
	Panel D: Main Results - Married Men			
Father Rank (β_1)	0.180	0.189	0.159	0.139
	(0.009)	(0.008)	(0.007)	(0.006)

Notes: This table reports coefficients from equation 11 estimated separately on each cohort with heteroskedasticity-robust standard errors in parentheses. Record links for Panels A through C are from the Census Linking Project (Abramitzky et al., 2020), and use the standard ABE-method with NYSIIS name transformations. Regressions are unweighted. The sample consists of Massachusetts-born men. Panel A reports results for married and single men combined. Panel B limits the sample to single men, whereas Panel C limits the sample to married men. Panel D reprints results from Table 2 Panel B for the links created using the marriage registers. All regressions include as controls quartics in the age of the father and the age of the son at the time economic status is measured.
Sources: 1850-1920 Decennial Census data from Ruggles et al. (2017). Record linkages from the Census Linking Project (Abramitzky et al., 2020). Marriage registrations from *FamilySearch.org*.

women's mobility converges with men's by the 1900-1920 cohort.

Our main results are not sensitive to choices on sample limitations and do not suffer from bias arising from the linkage procedure providing an unrepresentative sample. Bailey et al. (2018) and Abramitzky et al. (2021) argue that nonrandom selection of records into the linked sample is common and can bias results. They suggest reweighting the linked sample to better reflect observable characteristics in some base population. In our case, the choice of a base population with which to compare is not straightforward. We begin with a set of marriages occurring in Massachusetts, a population that cannot directly be found in decennial census microdata. We choose to reweight the full linked sample to mimic the observable characteristics (functions of names, ages, and marriage dates) of the marriage registrations. Results reported in Panel A of Appendix Table A1 follow the patterns for mobility of the unweighted regressions. Panel B limits the sample to the Massachusetts born and reweights observations to reflect the observable characteristics of the Massachusetts-born population, married and single, in the appropriate adult decennial census. Again, mobility patterns follow a similar path as in the unweighted regressions. Results are similar when estimating a log-log IGE instead of the rank-rank base specification, as can be seen in Panel C. During times of structural transformation embodied in the movement away from agriculture and to industry, mobility estimates may be sensitive to how farmers and farm income is handled (Xie and Killewald, 2013). Panel D of Appendix Table A1 drops observations where the father reports their occupation as a farmer. Finally, life-cycle bias may still enter our estimates even though we control for quartics in the ages of both the father and husband. Panel E further limits the sample so that wealth score measurement occurs at ages 30-50 when measurement error is less likely to occur (Mazumder, 2005; Haider and Solon, 2006).

To summarize, the intergenerational persistence of economic status was higher for men than women in the 1850-1870 sample across all measures of economic status. Both sexes experienced increases in mobility, with earlier improvements for men. Our preferred estimates are from the rank-rank regressions using the total property wealth score. Using these estimates the question becomes "what caused the initial gap between the sexes and the difference in changes over time?". Note that in the nineteenth century white married women often did not participate in market work, or at least did not report an occupation tied to outside the home market work on census enumerations. Intergenerational persistence can then only be measured using the husband's economic status, and thus mobility for women is highly dependent on the availability of potential husbands. Moreover, the choice of a spouse is not random, but determined in a marriage market equilibrium. We now turn to estimating the importance of marital sorting and its role in mobility changes.

5.2 The Marriage Market and Marital Sorting

Married white women in the nineteenth century United States rarely participated in out-of-home market work and often lacked sources of income independent of their husbands (Goldin, 1983). Accordingly, the marriage market played an important role in determining women's economic status and intergenerational mobility. In this section, we estimate the strength of sorting into marriages based on parental and family background, and use the model of Espín-Sánchez et al. (2023) to uncover the underlying structural assortative matching parameter.

We begin by exploring the association in economic status between the fathers of the spouses. The data used in this exercise is the portion of the linked sample in which we were able to match both husband and wife to their childhood household in a pre-marriage census. We begin with regressions similar to those used previously. Equation 12 captures the notion of assortative mating by regressing the husband's father's occupational wealth score rank on that of the wife's father and is the empirical equivalent to equation 7 in the theoretical framework.

$$Y_i^f = \alpha + \lambda Y_i^{fl} + v_i \tag{12}$$

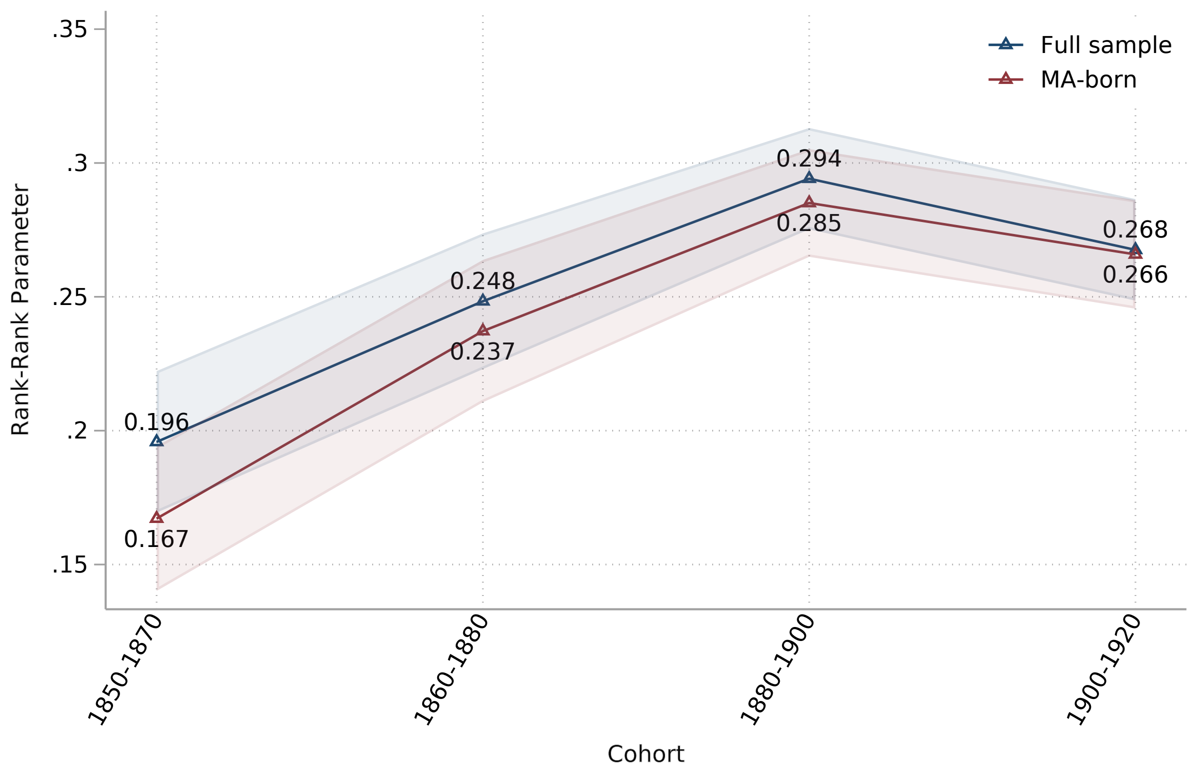
where Y_i^f and Y_i^{fl} are the occupational standing measures for the husband's and wife's father, respectively. λ provides a measure of sorting in the marriage market. Each regression includes a quartic in the husband's father's age and a quartic in the wife's father's age, both taken at the time when economic status is measured.

Figure 4 presents our first results on the Massachusetts marriage market in the nineteenth and early twentieth centuries. Unsurprisingly, we find evidence of marital sorting based on the status of fathers in all periods and all specifications, and can definitively say that couples did not randomly choose mates across parental socioeconomic status. The evidence is also clear about the trend in within group homogamy. The association in fathers' ranks in the full sample of couples (in blue) increased from 0.196 in the 1850-70 cohort to a peak of 0.294 in the 1880-1900 cohort, and then decreased slightly to 0.268 in the 1900-20 cohort. Trends and levels are similar when we restrict the sample of couples to having at least one spouse be born in Massachusetts.

These estimates are somewhat smaller than those for more recent periods in the United States. Charles et al. (2013) find a correlation in parental wealth of 0.4, while Gihleb and Lang (2016) report that correlation in spouses education is around 0.6. The most comparable study to ours is that of Olivetti et al. (2020), which applies the pseudo-linking methodology of Olivetti and Paserman (2015) to marital sorting over a period roughly similar to ours. For the Northeast, they find a correlation in the occupational status of the spouses' fathers between .008 and .03, levels that are 10 to 30 magnitudes less than our estimates. Note that the authors make clear that the pseudo-linking methodology likely biases magnitudes downward because of the use of an imperfect proxy for parental status for both values in the correlation. When adjusting for the bias, the correlation in spouses' parental status is between 0.1 and 0.3 for the entire nation, values much more in line with ours. Interestingly, we find roughly the same pattern over time for marital sorting as Olivetti et al. (2020). Direct comparisons are difficult because we build on our data on a marriage cohort model, whereas they use a birth cohort framework. Nonetheless, they find increases in marital sorting over the the pre-1880 birth cohorts, followed by declines for the 1880-1900 birth cohorts, roughly in line with the patterns we see in our estimates for couples born in Massachusetts.

Next, we show the evolution of the underlying assortative matching parameter. The correlation in spouses' parental status, while an interesting empirical relationship in its own right, is not

Figure 4: Marital Sorting - Father on Father-in-law Estimates



Notes: Each entry is an estimate of marital sorting from regressing the husband's father's wealth score rank on the wife's father's wealth score rank. Results in blue are for the full sample and those in red limit the sample to at least one spouse born in Massachusetts. All regressions include as controls a quartic in each father's age at the time economic status is measured. Shaded areas represent 95 percent confidence intervals for the point estimates.
Sources: 1850-1920 Decennial Census data from Ruggles et al. (2017). Marriage certificates from *FamilySearch.org*.

equivalent to what economists define as assortative matching between the spouses.[18] We use the structural model of Espín-Sánchez et al. (2023) detailed in our theoretical framework to uncover γ, the correlation in economic status between husband and wife. Note that we must use the structural model to do this because the economic status of women is not directly observed in the data. We first show results in a model using strong assumptions similar to those in Curtis (2021) and Clark and Cummins (2022), followed by those when we relax the assumptions.

The assortative matching parameter γ can be recovered from the data using only the correlation in economic status between the father and husband and the correlation of status between father-in-law and husband. Espin-Sanchez et al. (2022) show that this can be done as

$$\begin{aligned}\gamma &= \mathbb{E}[X_{husband}X_{father}] \times \mathbb{E}[X_{husband}X_{father-in-law}] \\ &= b_f \times b^*_{fl}\end{aligned} \tag{13}$$

[18]See the Theoretical Framework section and Espín-Sánchez et al. (2023) for more details.

when we assume that i.) the mother does not transfer status to the child ($B_m = 0$) and ii.) the husband's parents' status is uncorrelated with the wife's parent's status: $\mathbb{E}[X_f X_{fl}] = \mathbb{E}[X_f X_{ml}] = \mathbb{E}[X_{fl} X_m] = \mathbb{E}[X_m X_{ml}] = 0$. These assumptions are quite strong, and in fact we've already shown that ii.) is not true in that the status of the spouses' fathers are correlated. However, this exercise allows us to use the much larger set of linked data where we only need to match one of the spouses back to their father. We relax these assumptions when using the sample where we are able to link both spouses back to fathers, and show the results are similar to the more restrictive model with the full sample.

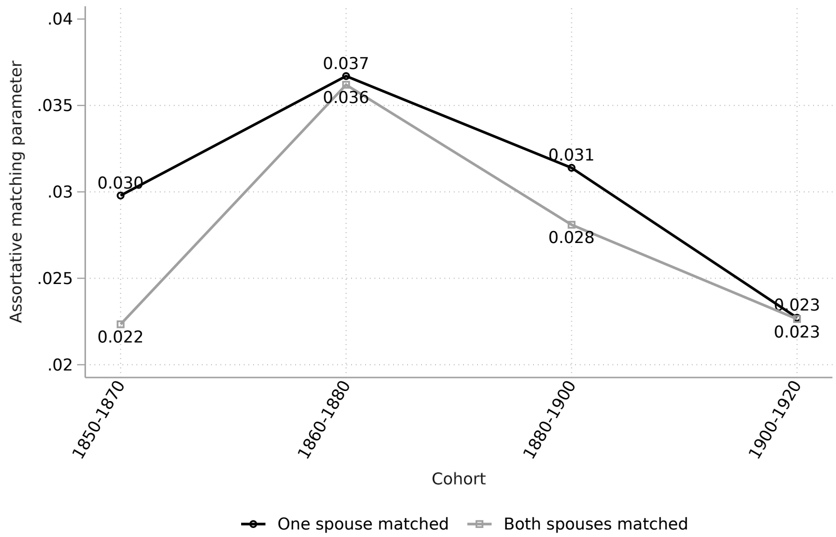
Results reported in Figure 5a suggest that assortative matching is important in mid-19th century marriage markets and followed an inverted-U shape pattern. The strength of matching increased from 0.03 in the 1850-1870 cohort to 0.037 in the 1860-1880 cohort, after which it declined consistently to a low of 0.023 by the 1900-1920 cohort. For the sample with both spouses matched to their fathers, we find lower initial levels of assortative matching, but a similar inverted-U shaped pattern. Note that the magnitudes for the assortative matching parameter are difficult to interpret and to compare to empirical relationships in studies on assortative matching in more recent periods. What we will show shortly through a counterfactual analysis is that the observed changes in assortative matching found here are economically important for observed mobility estimates for women.

Assortative matching follows a similar trajectory when we relax the assumptions used in the restrictive model above. Here, we i.) allow the inheritance term for mothers (B_m) to be non-zero and different from the inheritance term for men (B_f), and ii.) all correlations between the parents' status are equal, but not uncorrelated - $\mathbb{E}[X_f X_{fl}] = \mathbb{E}[X_f X_{ml}] = \mathbb{E}[X_{fl} X_m] = \mathbb{E}[X_m X_{ml}]$. Espín-Sánchez et al. (2023) show that with these assumptions, and now adding in the estimated correlation in status between father and father-in-law, the structural inheritance parameters (B_m) and (B_f) can be recovered along with the assortative matching parameter using the system of equations 8 through 10. Note that this now requires us to move to the sample in which both spouses are linked to fathers.

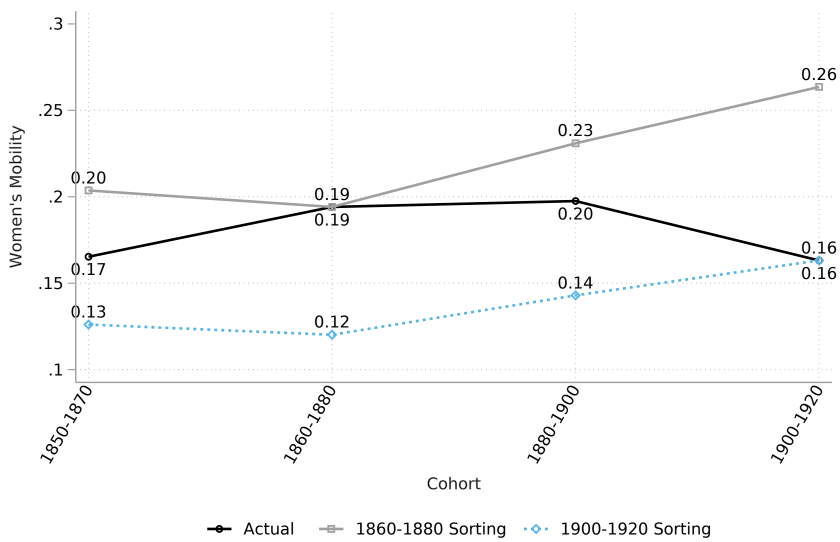
Table 4 lists the solutions for these parameters by cohort. Assortative matching began at 0.084 in the 1850-70 cohort, and increased by 27 percent moving to the 1860-1880 cohort. Over the next two cohorts, γ declined dramatically by 42 percent to a low of 0.062 in the 1900-1920 cohort. These values follow the same path as in the more restrictive model in Figure 5a, but are generally 2 to 4 times higher. We find higher levels of assortative matching here because we allow mothers to matter and for the inheritance parameters to change differentially over time. Interestingly, we find that mothers matter more for the economic status of children than do fathers. The inheritance parameter from mothers is 0.66 and 0.14 from fathers in the initial 1850-1870 cohort. Mothers remain more important over the entire period, however the gap declines. By the 1900-1920 cohort the inheritance from mothers declined by a third to 0.44, whereas inheritance from fathers increased slightly to 0.17 in 1880-1900 and 0.16 in 1900-1920.

Figure 5: Restrictive Model - Parameter Estimates and Counterfactual

(a) Structural Assortative Matching Parameter Estimates (γ)



(b) Counterfactual Women's Mobility by Strength of Assortative Matching



Notes: Each entry of Panel A is an estimate of the structural assortative matching parameter γ, by multiplying the father/husband correlation in status (b_f) times the father-in-law/husband correlation in status (b^*_{fl}). The sample is limited to observations born in Massachusetts. The black line reports results for the sample where at least one spouse is linked back to their father. The gray line reports results for the couple sample where both spouses are required to be linked back to fathers. All regressions include as controls quartics in father's age and husband's age at the time economic status is measured. Panel B shows results of the counterfactual analysis for the restrictive model. The black line plots actual women's mobility from the data ($\beta_1 + \beta_2$) from equation 11. The gray line plots the counterfactual women's mobility if assortative matching was at the 1860-1880 cohort level. The blue line plots counterfactual women's mobility using the 1900-1920 cohort assortative matching level.
Sources: 1850-1920 Decennial Census data from Ruggles et al. (2017). Marriage certificates from *FamilySearch.org*.

Table 4: Structural Parameter Estimates from Less Restrictive Model

Cohort	γ	B_m	B_f
1850-1870	0.084	0.657	0.135
1860-1880	0.107	0.652	0.146
1880-1900	0.063	0.408	0.171
1900-1920	0.062	0.437	0.158

Notes: This table reports solutions for structural inheritance parameters from the mother (B_m) and father (B_f) and assortative matching parameter (γ) from the system of equations 8 through 10. The data used to estimate the empirical relationships is Massachusetts-born sample in which both spouses are linked to the father.
Sources: 1850-1920 Decennial Census data from Ruggles et al. (2017). Record linkages from the Census Linking Project (Abramitzky et al., 2020). Marriage registrations from *FamilySearch.org*.

5.3 The Contribution of Marital Sorting to Women's Intergenerational Mobility

Finally, we rely on the theoretical framework developed by Espín-Sánchez et al. (2023) and outlined in section 3 to estimate counterfactual mobility rates under different levels of marital sorting. The model shows that assortative matching is a function of estimated correlations in status from the data. In the restrictive model, it is equal to the correlation in status between father and husband times the correlation between father-in-law and husband - $\gamma = b_f \times b^*_{fl}$. To create the counterfactual in this case, we substitute γ from a different cohort into the equation with the actual b_f and then solve for b^*_{fl}, a counterfactual women's mobility holding men's mobility constant.

Figure 5b compares women's mobility estimates from the data to what they would have been if assortative matching was at its highest level as in the 1860-1880 cohort (in gray) or its lowest level as in the 1900-1920 cohort (in blue). We can see from these graphs that stronger assortative matching leads to higher intergenerational persistence for women. With assortative matching at its 1860-1880 level, women's rank-rank parameter would have been 18 percent higher at 0.20 in the 1850-1870 cohort, and by 62.5 percent to 0.26 in the 1900-1920 cohort. This high level of assortative matching would have reversed women's gain in mobility over the cohorts, and would have made women's rank-rank parameter double that of men's in the 1900-1920 cohort.

Conversely, the importance of woman's father's economic status would have been much less important for their own adult economic status if the strength of assortative matching would have remained at the low level of the 1900-1920 cohort. Women's rank-rank parameter would have been lower by 24 percent (0.13 vs. 0.17) in 1850-1870, 37 percent (0.12 vs. 0.19) in 1860-1880, and 30 percent (0.14 vs. 0.20) in 1880-1900.

6 Heterogeneity

Changes in the strength of assortative matching were clearly important for the evolution of women's intergenerational mobility over the 19th and early 20th centuries. Everyone might have experienced these changes in the marriage market similarly, or certain subgroups might have been affected more than others by changes in the economy. The changes in assortative matching and mobility could have been driven by changes in i.) level differences in economic status between groups, ii.) the share of the population made up by different subgroups, and 3.) differences in mobility by subgroups. In this section, we briefly explore heterogeneity in mobility and marital sorting for individuals who grew up in urban vs rural places, those who migrate within the US vs those who stay in MA, and children of immigrants vs US-borns.

6.1 Mobility

To describe the underlying factors behind our estimates of mobility, we consider how different childhood conditions or later life migration choices may have contributed to the mobility estimates and their evolution.

Urban vs Rural Childhood

First, we examine the impact of growing up in an urban vs. a rural area. Figure 6 plots a binned scatterplot showing the mean income rank of children growing up in rural and urban areas by father's income ventile rank separately for women (Panel A) and men (Panel B) for the 1850-1870 cohort. Urban is defined as towns and incorporated places with at least 2,500 in population, and rural is defined as the complement.[19] Panels A and B suggest that both men and women who grew up in rural areas are more mobile than those growing up in urban areas and this holds true especially for children with fathers in the upper tail of the income distribution. However, much of the mobility difference between those from rural areas and urban areas, as captured by the slope coefficients, is driven by children born to fathers at the 80th percentile or above. Much of the higher mobility of those from rural areas is in fact downward mobility where children of fathers in the upper tail of the income distribution are expected to do worse than their fathers. For instance, a woman born to a father at the 80th percentile of the income distribution is expected to marry a man at the 70th percentile of his income distribution. In the Appendix, we show that these patterns also hold for later cohorts of our analysis.

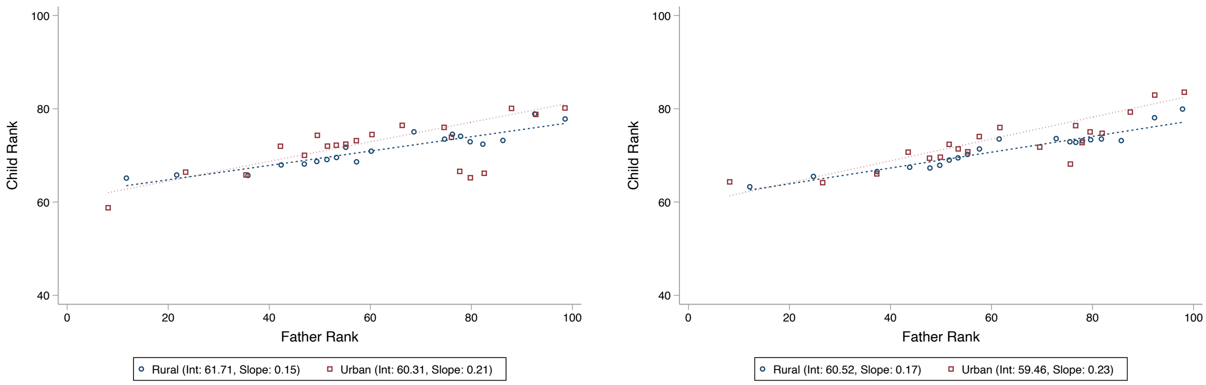
Children of Immigrants vs US-born

Next, we examine how mobility may have differed for children of immigrants in comparison to children of native-born parents. In our samples, about 10 percent of the 1850-1870 cohort, 20 percent of the 1860-1880 cohort, 38 percent of the 1880-1900 and 50 percent of the 1900-1920 cohort were children of immigrants, defined as at least one of the parents reported as foreign born. The relative magnitudes of the size of second-generation immigrant children fits with the wave of immigration that began in the mid-1840s, which peaked in the early 1900s.

[19] We make use of the IPUMS *urban* variable. The political boundaries in Massachusetts make constructing the *urban* variable more complicated. IPUMS includes the following in the variable description *"Also includes households in Massachusetts towns (townships) containing a village or thickly settled area of 2,500 or more inhabitants and comprising, either by itself or when combined with other villages within the same town, more than 50 percent of the total population of the town. Also includes townships and other political subdivisions (not incorporated municipalities) with a total population of 10,000 or more and a population density of 1,000 or more per square mile."*

Figure 6: Intergenerational Mobility of Children from Rural and Urban Areas 1850-1870

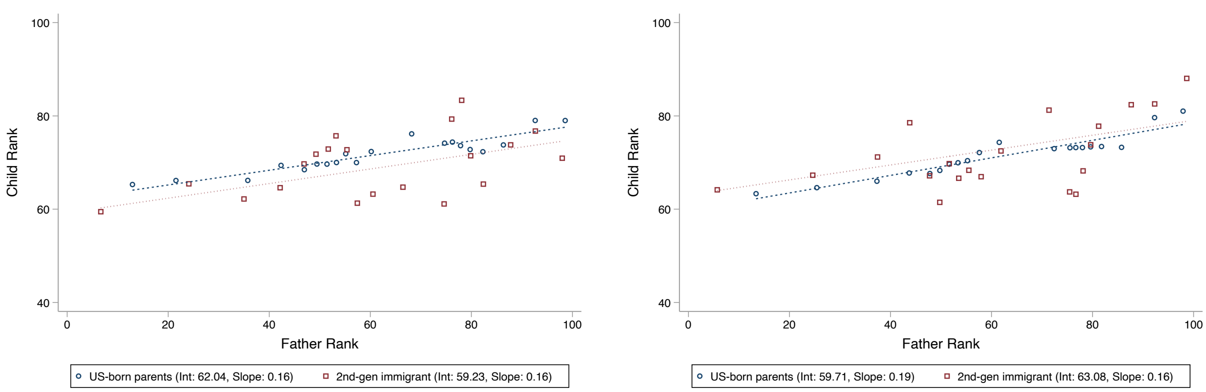
(a) Women (b) Men



Notes: Urban is an indicator equal to one if the observation resided in an urban area as a childhood. Urban is defined following the IPUMS definition of towns and incorporated places of at least 2,500 in population.

Figure 7: Intergenerational Mobility of Children of Immigrants and US-born 1850-1870

(a) Women (b) Men



Notes: Immigrant is an indicator equal to one if at least one of the parents was born outside the United States according to FBPL and MBPL IPUMS variables in the adult census.

Figure 7, Panels A (women) and B (men) show mean income rank of children of immigrants and US-born by father's income ventile rank for the 1850-1870 cohort. For nearly every ventile rank of father's income, sons of immigrants are more upwardly mobile than sons of the US-born. This fact is supported by Abramitzky et al. (2021) who find that sons of immigrants in the US from nearly every sending country are more mobile than sons of US born in the 19th century. However, daughters of immigrants do not enjoy this mobility advantage. While the daughters of immigrants have the same rank-rank estimate as the daughters of US-born (given by the slopes), for every ventile of father's rank, daughters of immigrants are marrying men of slightly lower rank in their income distribution.

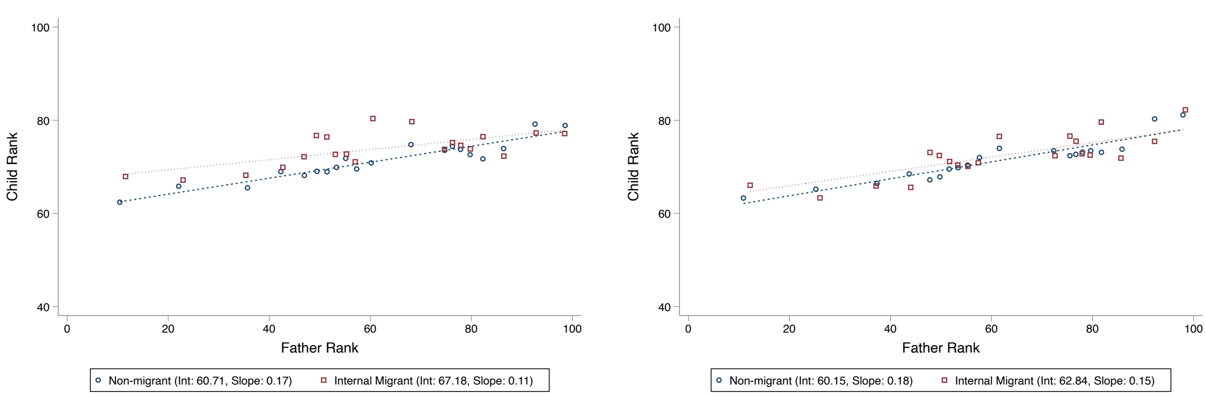
Internal Migration

The nineteenth century was both an economically and geographically mobile period for the United States (Hall and Ruggles, 2004; Rosenbloom and Sundstrom, 2004; Ferrie, 2005; Long and Ferrie, 2013; Salisbury, 2014). Low rates of economic persistence across generations may have been partially driven by the ability to move to the frontier, a new region, or to a city, where economic opportunities were more favorable. Did internal migration have an impact on mobility in our sample? We define an internal migrant as an adult that reports residing in a state different from that in the childhood census. About half the sample of husbands and wives in both periods are coded as internal migrants. In contrast to immigration, where the decision to move was made by the parents' generation, this decision was made by the child's generation.

Figure 8 Panel A shows that for nearly every ventile rank of father's income in the 1850-1870 cohort, women who migrated within the US were more upwardly mobile than women who did not migrate. This upward mobility advantage is also enjoyed by men who migrate, though the upward mobility gap between migrants and non-migrants is much smaller for men.

Figure 8: Intergenerational Mobility of Internal Migrants and Stayers 1850-1870

(a) Women (b) Men



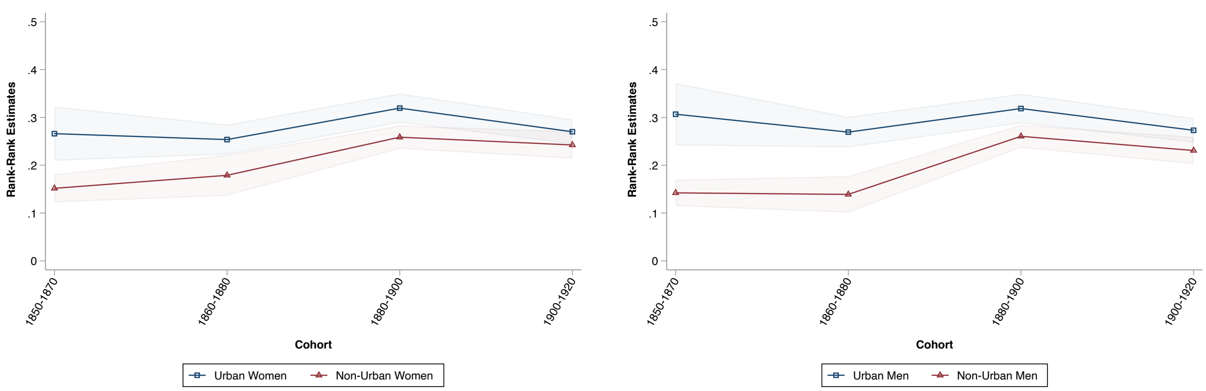
Notes: Internal migrant is defined as observing an observation in two different states in the childhood census and adult census.

6.2 Marital Sorting

To better understand whether the overall trend in assortative matching in our sample is driven by any particular subgroup, we plot the strength of marital sorting from Equation 12 over time by subgroups in Figures 9-11. We find evidence that marital sorting may have been stronger in urban areas for both men and women (Figure 9 Panels A and B). The gap in marital sorting between urban and rural areas is widest during the earliest cohorts and converges as sorting for both groups increased. However, the majority of the convergence and overall observed increase in assortative mating in the full sample is driven by increased strength of sorting in the rural group. We do not find evidence that the increase in assortative mating across our four cohorts as observed in Figure 4 is driven by assortative mating patterns of immigrants or internal migrants (Figures 10 and 11).

Figure 9: Marital Sorting: Urban and Rural Subgroup Estimates

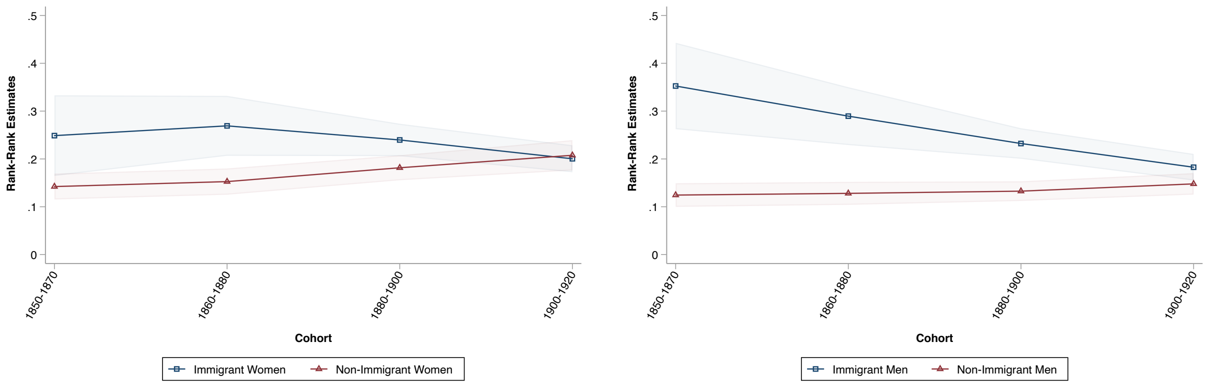
(a) Women (b) Men



Notes: Urban is an indicator equal to one if the observation resided in an urban area as a child. Urban is defined following the IPUMS definition of towns and incorporated places of at least 2,500 in population.

Figure 10: Sorting: Immigrant Subgroup Estimates

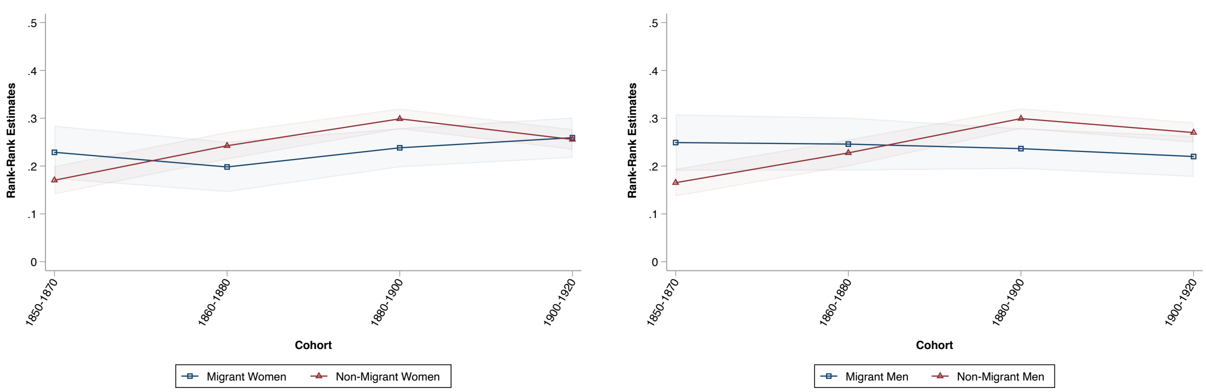
(a) Women (b) Men



Notes: Immigrant is an indicator equal to one if at least one of the parents was born outside the United States according to FBPL and MBPL IPUMS variables in the adult census.

Figure 11: Sorting: Internal Migrant Subgroup Estimates

(a) Women (b) Men



Notes: Internal migrant is defined as observing an observation in two different states in the childhood census and adult census.

7 Conclusion

This paper presents new evidence of the intergenerational mobility for men and women married in Massachusetts between 1850 and 1920. Consistent with the historical literature we find high levels of mobility in the nineteenth and early twentieth century relative to the second half of the twentieth century (Long and Ferrie, 2013; Feigenbaum, 2017; Chetty et al., 2014; Solon, 1992). The mobility estimates based on various measures all suggest that there was more persistence in economic status for sons than for daughters during the middle of the 1800s. By the end of the nineteenth century this difference in mobility between the sexes had largely vanished. In our sample, we find mobility levels to have increased over the course of the second-half of the nineteenth century, whereas others have found decreases for national samples (Long and Ferrie, 2013; Olivetti and Paserman, 2015).

An open question remains as to why the sons and daughters of Massachusetts faced higher rates of social mobility than the U.S. as a whole. The American economy faced massive structural transformation over the nineteenth century: westward expansion and the eventual closing of the frontier, the growth of factories and industry, and dramatic inflows of immigrants. Massachusetts' experience was a harbinger for the rest of the country. The Commonwealth early on transitioned its labor force out of agriculture and into manufacturing and white collar occupations: it ranked second only to Rhode Island throughout the latter half of the nineteenth century as having the lowest proportion of workers in agriculture. Massachusetts received large flows of immigrants from the mid-1840s onward. The majority of the rail network was built out as early as 1850, and with the opening of the Erie Canal brought competition from New York and Midwestern crops. All of this is to say that the labor market opportunities available to men in Massachusetts did not mimic those of the rest of the country, especially in agriculture. Our results do suggest that internal migration remained an important avenue for social mobility throughout the nineteenth century for men who had at least some attachment to Massachusetts.

Marriage was the prime vehicle for economic mobility for women during this period. Accordingly, an understanding of the marriage market is crucial to understand female mobility. We provide evidence that sorting based on social background – originating from preferences or availability of potential mates through social interaction – is important to understanding how much of parents' socioeconomic conditions are transferred to their children. Using a sample of our linked father-daughter and father-son pairs for which both the husband and wife were successfully matched, we find a high degree of sorting in marriage within social background as proxied by occupational status. However, the likelihood that women would marry out of the economic class of their fathers decreased over time. Consequently, mobility of women in the 1860-1880 cohort might have been higher than observed if they faced the lower level of assortative matching of women in the later 1900-1920 cohort. Another way to interpret the results is that for women's own economic standing, the importance to marry upward with respect to fathers' standing decreases over our time period. For example, a woman marrying into a family with a higher economic standing than her own is 50% more likely to move up relative to her own father during the 1850-1870 cohort. However, in the 1880-1900 cohort, marrying into a higher ranked family is only associated with a 30% higher likelihood of moving up relative to her own father. Combining the mobility and marriage sorting results, marriage into a specific family background becomes less important for women's mobility because *men* are becoming more mobile. The economic standing of sons depends less on their father's standing. Further research is required to better understand the forces driving the changes in assortative mating we find over the course of the nineteenth century.

8 Bibliography

Aaronson, D. and B. Mazumder (2008). Intergenerational economic mobility in the u.s., 1940-2000. *Journal of Human Resources 43*(1), 139–172.

Abramitzky, R., L. Boustan, K. Eriksson, J. Feigenbaum, and S. Pérez (2021, September). Automated Linking of Historical Data. *Journal of Economic Literature 59*(3), 865–918.

Abramitzky, R., L. Boustan, K. Eriksson, S. Pérez, and M. Rashid (2020). Census linking project: Version 2.0 [dataset]. https://censuslinkingproject.org.

Abramitzky, R., L. Boustan, E. Jacome, and S. Perez (2021, February). Intergenerational Mobility of Immigrants in the United States over Two Centuries. *American Economic Review 111*(2), 580–608.

Abramitzky, R., L. P. Boustan, and K. Eriksson (2012, May). 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, R., L. P. Boustan, and K. Eriksson (2014). A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration. *Journal of Political Economy 122*(3), 467–506.

Althoff, L., H. B. Gray, and H. Reichardt (2023). The missing link: Women and intergenerational mobility.

Bailey, M., C. Cole, M. Henderson, and C. Massey (2018). How well do automated linking methods perform in historical samples? evidence from new ground truth. Working Paper.

Becker, G. and N. Tomes (1979). An equilibrium theory of the distribution of income and intergenerational mobility. *Journal of Political Economy 87*(6), 1153–89.

Black, S. and P. Devereux (2011). Recent developments in intergenerational mobility. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics* (1 ed.), Volume 4B, Chapter 16, pp. 1487–1541. Elsevier.

Chadwick, L. and G. Solon (2002). Intergenerational income mobility among daughters. *The American Economic Review 92*(1), 335–344.

Charles, K., E. Hurst, and A. Killewald (2013, February). Marital Sorting and Parental Wealth. *Demography 50*(1), 51–70.

Chetty, R., N. Hendren, P. Kline, and E. Saez (2014). Where is the land of opportunity? the geography of intergenerational mobility in the united states. *Quarterly Journal of Economics 129*(4), 1553–623.

Chetty, R., N. Hendren, P. Kline, E. Saez, and N. Turner (2014, May). Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility. *American Economic Review 104*(5), 141–147.

Clark, G. and N. Cummins (2022). ClarkCummins2022.pdf.

Collins, W. J. and M. H. Wanamaker (2014, January). 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, W. J. and M. H. Wanamaker (2015, December). 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, W. J. and M. H. Wanamaker (2022, July). African American Intergenerational Economic Mobility since 1880. *American Economic Journal: Applied Economics 14*(3), 84–117.

Collins, W. J. and A. Zimran (2023, July). 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.

Corak, M. and A. 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.

Curtis, M. (2021, December). jmp-matt-curtis.pdf.

Dahl, M. and T. DeLeire (2008). The association between children's earnings and fathers' lifetime earnings: estimates using administrative data. Discussion Paper No. 1342-08. http://irp.wisc.edu/publications/dps/pdfs/dp134208.pdf.

Eriksson, K., G. T. Niemesh, and M. Thomasson (2018). Revising infant mortality rates for the early 20th century united states. *Demography 55*, 2001–2024.

Ermisch, J., M. Francesconi, and T. Siedler (2006, July). Intergenerational Mobility and Marital Sorting. *Economic Journal 116*(513), 659–679.

Espin-Sanchez, J.-A., S. Gil-Guirado, and C. Vickers (2022, March). 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.

Espín-Sánchez, J.-A., J. P. Ferrie, and C. Vickers (2023, August). Women and the Econometrics of Family Trees.

FamilySearch (2016, June). "massachusetts marriages, 1695-1910". "Database: http://FamilySearch.org". Index based upon data collected by the Genealogical Society of Utah, Salt Lake City.

Feigenbaum, J. (2016). A machine learning approach to census record linking. Working Paper.

Feigenbaum, J. (2017). Multiple measures of historical intergenerational mobility: Iowa 1915 to 1940. Forthcoming in Economic Journal.

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

Gihleb, R. and K. Lang (2016, December). Educational Homogamy and Assortative Mating Have Not Increased.

Goldin, C. (1983). The changing economic role of women: A quantitative approach. *The Journal of Interdisciplinary History 13*(4), 707–733.

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

Haider, S. and G. Solon (2006). Life-cycle variation in the association between current and lifetime earnings. *The American Economic Review 96*(4), 1308–1320.

Hall, P. K. and S. Ruggles (2004, December). "restless in the midst of their prosperity": New evidence on the internal migration of americans, 1850–2000. *Journal of American History 91*(3), 829–846.

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

Jácome, E., I. Kuziemko, and S. Naidu (2021). Mobility for all: Representative intergenerational mobility estimates over the 20th century. *National Bureau of Economic Research.*

Jäntti, M., B. Bratsberg, K. Røed, O. Raaum, R. Naylor, E. Österbacka, A. Björklund, and T. Eriksson (2006, January). American Exceptionalism in a New Light: A Comparison of Intergenerational Earnings Mobility in the Nordic Countries, the United Kingdom and the United States. IZA Discussion Papers 1938, Institute for the Study of Labor (IZA).

Lee, C.-I. and G. Solon (2009). Trends in intergenerational income mobility. *The Review of Economics and Statistics 91*(4), 766–772.

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

Long, J. and J. Ferrie (2013, June). Intergenerational occupational mobility in great britain and the united states since 1850. *American Economic Review 103*(4), 1109–37.

Mazumder, B. (2005, May). 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, B. (2014). Black–White Differences in Intergenerational Economic Mobility in the U.S. *Economic Perspectives 38*(Q I), 1–18.

Miles, A. (1999). *Social Mobility in Nineteenth- and Early Twentieth-Century England.* New York: MacMillan.

Oddbjørn, R., B. Bernt, R. Knut, Ö. Eva, E. Tor, J. Markus, and N. R. A (2008, January). Marital Sorting, Household Labor Supply, and Intergenerational Earnings Mobility across Countries. *The B.E. Journal of Economic Analysis & Policy 7*(2), 1–49.

Olivetti, C. and M. D. Paserman (2015, August). 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, C., M. D. Paserman, and L. Salisbury (2016). Three-generation mobility in the united states, 1850-1940: The role of maternal and paternal grandparents. NBER Working Paper No. 22094.

Olivetti, C., M. D. Paserman, and L. Salisbury (2018, October). Three-generation mobility in the United States, 1850–1940: The role of maternal and paternal grandparents. *Explorations in Economic History 70*, 73–90.

Olivetti, C., M. D. Paserman, L. Salisbury, and E. A. Weber (2020, October). Who Married, (to) Whom, and Where? Trends in Marriage in the United States, 1850-1940. Technical Report w28033, National Bureau of Economic Research, Cambridge, MA.

Preston, S. H. and M. R. Haines (1991). *Fatal Years: Child Mortality in Late Nineteenth-Century America.* Princeton University Press.

Rosenbloom, J. L. and W. A. Sundstrom (2004). The decline and rise of interstate migration in the united states: Evidence from the ipums, 1850-1990. In A. J. Field (Ed.), *Research in Economic History*, Chapter 22, pp. 289–235. Elsevier.

Ruggles, S., K. Genadek, R. Goeken, J. Grover, and M. Sobek (2017). Integrated public use microdata series. Version 7.0 [Machine-readable database].

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

Solon, G. (1992). Intergenerational income mobility in the united states. *American Economic Review 82*(3), 393–408.

Solon, G. (1999). Intergenerational mobility in the labor market. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3 of *Handbook of Labor Economics*, Chapter 29, pp. 1761–1800. Elsevier.

Song, X., C. G. Massey, K. A. Rolf, J. P. Ferrie, J. L. Rothbaum, and Y. Xie (2020, January). 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, Z. (2022, July). Intergenerational Mobility in American History: Accounting for Race and Measurement Error.

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

Appendices

A Additional Tables and Figures

Table A1: Mobility Robustness Checks

	1850-1870	1860-1880	1880-1900	1900-1920
	Panel A: Full Sample - Weighted			
Father Rank (β_1)	0.281	0.296	0.241	0.216
	(0.011)	(0.009)	(0.008)	(0.008)
Father Rank x Woman (β_2)	-0.003	-0.019	0.025	-0.025
	(0.015)	(0.013)	(0.011)	(0.010)
N	27,133	34,196	45,680	52,577
	Panel B: Massachusetts Born - Weighted			
Father Rank (β_1)	0.177	0.171	0.146	0.121
	(0.010)	(0.010)	(0.008)	(0.008)
Father Rank x Woman (β_2)	-0.002	0.037	0.051	0.025
	(0.015)	(0.014)	(0.012)	(0.011)
N	20,320	25,001	31,982	36,750
	Panel C: Massachusetts Born - IGE			
Father Rank (β_1)	0.296	0.267	0.270	0.213
	(0.013)	(0.010)	(0.009)	(0.008)
Father Rank x Woman (β_2)	-0.038	-0.024	-0.027	-0.012
	(0.017)	(0.014)	(0.012)	(0.011)
N	20,318	24,997	31,981	36,822
	Panel D: Drop Children of Farmers			
Father Rank (β_1)	0.214	0.212	0.200	0.168
	(0.010)	(0.009)	(0.007)	(0.007)
Father Rank x Woman (β_2)	-0.022	0.009	0.031	0.009
	(0.014)	(0.012)	(0.010)	(0.009)
N	13,595	19,028	26,956	32,584
	Panel E: Limit to Prime Age Fathers			
Father Rank (β_1)	0.180	0.185	0.156	0.133
	(0.010)	(0.009)	(0.008)	(0.008)
Father Rank x Woman (β_2)	-0.017	0.013	0.044	0.023
	(0.014)	(0.013)	(0.012)	(0.011)
N	14,982	18,333	22,919	27,513

Notes: This table reports coefficients from equation 11 estimated separately on each cohort with heteroskedasticity-robust standard errors in parentheses. All regressions include quartics in the age of the father and the age of the husband at the time economic status is observed. Reported coefficients are evaluated at age 35 for both the father and the adult. Panel A weights the linked sample to mimic observed characteristics in the set of marriage registrations. Panel B limits the sample to observations born in Massachusetts and weights the sample to mimic the observable characteristics of the Massachusetts born population in the appropriate adult decennial census. See Appendix Section B.1 for details on how weights are constructed. All other regressions are unweighted. Panel C reports log-log IGE regressions on the Massachusetts born sample. Panel D drops any observation where the father reported being a farmer. Panel E limits the sample to prime aged fathers (30-50).
Sources: 1850-1920 Decennial Census data from Ruggles et al. (2017). Marriage registrations from *FamilySearch.org*.

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 logit 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 logit 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.

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 predicted probability of being a true match is $ln\left(\frac{\hat{\beta}X_{ij}}{1-\hat{\beta}Xij}\right) = \hat{\rho_{ij}}$. 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 $k \neq i$. 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 $k \neq i$. 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 A2 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.

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 A3 suggests our procedure gives false positive rates comparable to other methodologies used in the literature.[20] 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 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 A4 provides a complete breakdown of the

[20]See Abramitzky et al. (2021) and Bailey et al. (2018).

causes for match failure.

Table A2: 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 $ln(\frac{\hat{\beta}X}{1-\hat{\beta}X}) = \hat{\rho}$. 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 A3: 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 A2 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 A4: 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%

B.1 Weighting

We construct inverse probability weights (IPW) following the advice of Bailey et al. (2018) to make the linked sample mimic the observable characerістics of some base population. We construct two weighting schemes. The first weights the full sample to represent the characteristics of the population of registered marriages in Massachusetts. Here we estimate the probability of being linked within a cohort of marriages; a marriage in 1879 is compared to the population of marriages in the 1860-1880 cohort of marriages. The second scheme weights the linked sample of marriages with a Massachusetts born member to represent the characteristics of the population of Massachusetts born population of children in that observations childhood census year. For example, a marriage with a Massachusetts born husband observed as a child in the 1850 census is compared to the population of Massachusetts born children in the 1850 census.

The propensity of being linked $P_i(L_i = 1|X_i)$ comes from estimated probit models of link status on a rich set of observable characteristics, X_i, with observations re-weighted by $(1 - P_i(L_i = 1|X_i))/P_i(L_i = 1|X_i) * q/(1 - q)$, where q is the proportion of records linked. For the full sample, the prediction model includes variables for age, name length for the surname and given names of couple, and name commonness indices for the same set of names, and a set of indicators for the year of marriage. The variables used in the prediction for the sample limited to the Massachusetts born include: an urban residence dummy, age of the child, an indicator for residence in Massachusetts, a set of indicators for county of residence within Massachusetts, father's occscore, mother and father immigrant status, age of father and mother, indicators for father and mother literacy, and an indicator for whether mother is present in the household. All mother specific variables are interacted with the mother present indicator. Note that having the father present in the household is a requirement to enter the linked sample. All continuous variables are discretized in both prediction models. Table A5 compares the observable characteristics of the full linked sample to the population of marriage certificates separately for men and women. Tables A6 through A9 repeats the exercise for individual years. Tables A10 through A14 compare the observable characteristics of the linked Massachusetts born sample to the entire Massachusetts born child population in the appropriate childhood census year. Finally, Figures A1 through A3 plot the overlap between the linked sample and unlinked population over the predicted probability entering the linked sample.

Table A5: Representativeness of linked sample to population of registered marriages

	Unweighted			Weighted		
	Population	Linked sample	Diff (Linked - Pop)	Population	Linked sample	Diff (Linked - Pop)
	Panel A: Women					
Age	24.78	22.30	-2.64	23.90	23.60	-0.33
	(0.01)	(0.01)	(0.01)	(0.02)	(0.33)	(0.33)
Given name length	7.85	8.56	0.75	7.78	7.46	-0.34
	(0.00)	(0.01)	(0.01)	(0.01)	(0.07)	(0.07)
Surname length	6.72	6.38	-0.36	6.67	6.77	0.10
	(0.00)	(0.01)	(0.01)	(0.01)	(0.14)	(0.14)
Father name length	6.32	6.98	0.70	6.33	6.44	0.11
	(0.00)	(0.01)	(0.01)	(0.00)	(0.05)	(0.05)
Mother name length	5.98	6.53	0.58	5.98	5.93	-0.05
	(0.00)	(0.01)	(0.01)	(0.00)	(0.05)	(0.05)
Surname commenness	0.05	0.05	0.00	0.05	0.05	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Given name commenness	0.76	0.23	-0.56	0.84	1.51	0.72
	(0.00)	(0.00)	(0.00)	(0.00)	(0.07)	(0.07)
Father name commonness	1.42	0.93	-0.53	1.48	1.88	0.43
	(0.00)	(0.01)	(0.01)	(0.01)	(0.11)	(0.11)
Mother name commonness	2.33	1.46	-0.92	2.38	2.85	0.50
	(0.00)	(0.01)	(0.01)	(0.01)	(0.13)	(0.13)
	Panel B: Men					
Age	28.47	24.48	-4.23	26.79	25.60	-1.26
	(0.01)	(0.01)	(0.01)	(0.01)	(0.18)	(0.18)
Given name length	8.07	8.97	0.96	8.02	7.80	-0.24
	(0.00)	(0.01)	(0.01)	(0.00)	(0.06)	(0.06)
Surname length	6.46	6.36	-0.10	6.48	6.46	-0.02
	(0.00)	(0.01)	(0.01)	(0.00)	(0.03)	(0.03)
Father name length	6.54	7.09	0.59	6.56	6.57	0.02
	(0.00)	(0.01)	(0.01)	(0.00)	(0.03)	(0.03)
Mother name length	6.21	6.76	0.59	6.21	6.15	-0.06
	(0.00)	(0.01)	(0.01)	(0.00)	(0.07)	(0.07)
Surname commenness	0.05	0.05	-0.01	0.05	0.05	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Given name commenness	0.71	0.22	-0.52	0.79	1.54	0.80
	(0.00)	(0.00)	(0.00)	(0.01)	(0.10)	(0.10)
Father name commonness	1.27	0.81	-0.49	1.32	1.70	0.41
	(0.00)	(0.01)	(0.01)	(0.01)	(0.07)	(0.07)
Mother name commonness	2.06	1.23	-0.88	2.10	2.49	0.41
	(0.00)	(0.01)	(0.01)	(0.01)	(0.16)	(0.16)

Table A6: Representativeness of linked sample to population of registered marriages: 1870

	Women			Men		
	Population	Linked unweighted	Linked weighted	Population	Linked unweighted	Linked weighted
Age	24.10	21.87	23.15	28.16	24.59	25.81
	(0.01)	(0.03)	(0.20)	(0.02)	(0.03)	(0.52)
Given name length	7.53	8.41	6.81	7.48	8.42	6.90
	(0.00)	(0.02)	(0.11)	(0.00)	(0.02)	(0.24)
Surname length	6.35	6.25	6.66	6.30	6.33	6.40
	(0.00)	(0.01)	(0.15)	(0.00)	(0.01)	(0.10)
Father name length	5.66	6.49	5.93	6.01	6.52	5.88
	(0.00)	(0.02)	(0.06)	(0.00)	(0.02)	(0.10)
Mother name length	4.49	5.54	4.48	4.87	6.03	4.78
	(0.01)	(0.02)	(0.09)	(0.01)	(0.02)	(0.26)
Surname commenness	0.06	0.06	0.06	0.06	0.06	0.06
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Given name commenness	1.08	0.29	2.88	1.19	0.32	3.10
	(0.00)	(0.01)	(0.22)	(0.00)	(0.01)	(0.36)
Father name commonness	2.01	1.16	2.77	1.80	1.05	2.85
	(0.01)	(0.02)	(0.18)	(0.01)	(0.02)	(0.28)
Mother name commonness	5.94	3.42	5.99	5.01	2.40	5.43
	(0.01)	(0.05)	(0.21)	(0.01)	(0.04)	(0.60)

Table A7: Representativeness of linked sample to population of registered marriages: 1880

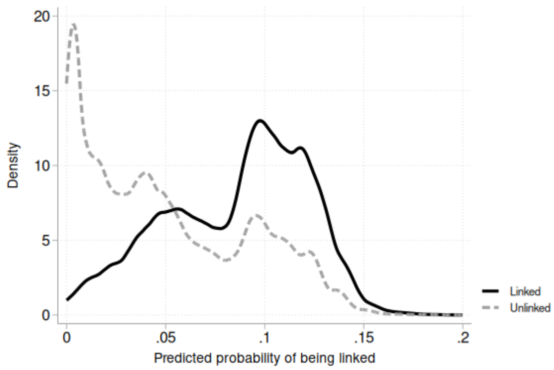
	Women			Men		
	Population	Linked unweighted	Linked weighted	Population	Linked unweighted	Linked weighted
Age	24.46	22.07	23.35	28.49	24.55	25.64
	(0.01)	(0.03)	(0.56)	(0.02)	(0.03)	(0.35)
Given name length	7.61	8.31	7.41	7.79	8.71	7.53
	(0.00)	(0.02)	(0.05)	(0.00)	(0.01)	(0.07)
Surname length	6.46	6.27	6.30	6.32	6.27	6.35
	(0.00)	(0.01)	(0.08)	(0.00)	(0.01)	(0.06)
Father name length	6.11	6.83	6.35	6.37	6.96	6.47
	(0.00)	(0.01)	(0.03)	(0.00)	(0.01)	(0.07)
Mother name length	5.91	6.64	6.16	6.32	6.96	6.21
	(0.00)	(0.02)	(0.10)	(0.00)	(0.02)	(0.08)
Surname commenness	0.05	0.05	0.05	0.05	0.05	0.05
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Given name commenness	1.06	0.32	1.76	1.00	0.30	1.80
	(0.00)	(0.01)	(0.12)	(0.00)	(0.01)	(0.12)
Father name commonness	1.73	1.10	1.96	1.58	0.98	1.90
	(0.00)	(0.02)	(0.05)	(0.00)	(0.01)	(0.10)
Mother name commonness	2.13	1.37	2.61	1.90	1.21	2.32
	(0.01)	(0.02)	(0.09)	(0.01)	(0.02)	(0.12)

Table A8: Representativeness of linked sample to population of registered marriages: 1900

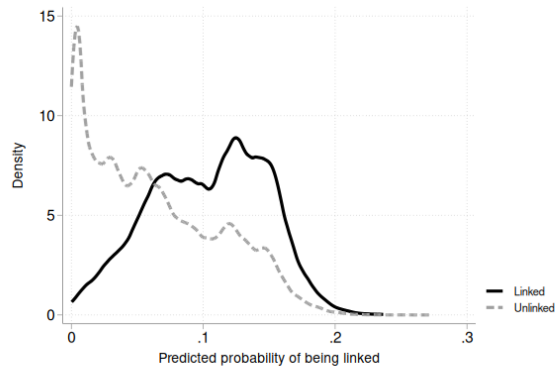
	Women			Men		
	Population	Linked unweighted	Linked weighted	Population	Linked unweighted	Linked weighted
Age	24.96	22.66	24.87	28.57	24.87	26.47
	(0.01)	(0.02)	(0.93)	(0.01)	(0.02)	(0.39)
Given name length	7.73	8.32	7.33	8.12	9.01	7.91
	(0.00)	(0.01)	(0.19)	(0.00)	(0.01)	(0.08)
Surname length	6.74	6.41	7.10	6.46	6.35	6.47
	(0.00)	(0.01)	(0.43)	(0.00)	(0.01)	(0.04)
Father name length	6.50	7.18	6.45	6.71	7.34	6.79
	(0.00)	(0.01)	(0.15)	(0.00)	(0.01)	(0.06)
Mother name length	6.38	6.85	6.22	6.60	7.08	6.68
	(0.00)	(0.01)	(0.14)	(0.00)	(0.01)	(0.05)
Surname commenness	0.06	0.05	0.04	0.05	0.04	0.05
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Given name commenness	0.75	0.26	1.19	0.65	0.21	1.27
	(0.00)	(0.01)	(0.11)	(0.00)	(0.00)	(0.11)
Father name commonness	1.47	1.02	2.14	1.29	0.84	1.54
	(0.00)	(0.01)	(0.32)	(0.00)	(0.01)	(0.05)
Mother name commonness	1.73	1.22	2.58	1.57	1.12	1.88
	(0.00)	(0.01)	(0.40)	(0.00)	(0.01)	(0.11)

Table A9: Representativeness of linked sample to population of registered marriages: 1920

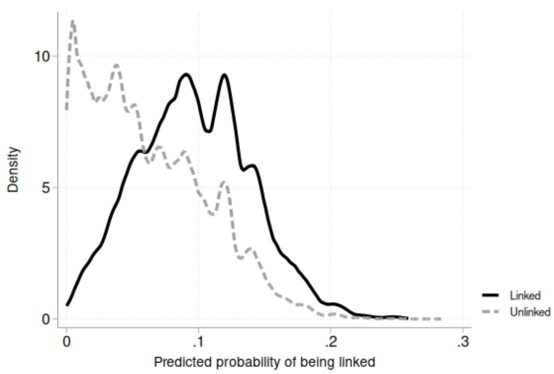
	Women			Men		
	Population	Linked unweighted	Linked weighted	Population	Linked unweighted	Linked weighted
Age	25.16	22.34	22.75	28.53	24.05	24.47
	(0.01)	(0.02)	(0.05)	(0.01)	(0.02)	(0.04)
Given name length	8.28	8.96	8.06	8.50	9.34	8.53
	(0.00)	(0.02)	(0.04)	(0.00)	(0.02)	(0.03)
Surname length	7.06	6.47	6.80	6.62	6.44	6.55
	(0.00)	(0.01)	(0.03)	(0.00)	(0.01)	(0.02)
Father name length	6.64	7.13	6.81	6.77	7.20	6.92
	(0.00)	(0.01)	(0.02)	(0.00)	(0.01)	(0.02)
Mother name length	6.44	6.64	6.42	6.48	6.67	6.55
	(0.00)	(0.01)	(0.02)	(0.00)	(0.01)	(0.02)
Surname commenness	0.04	0.06	0.04	0.05	0.04	0.04
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Given name commenness	0.41	0.14	0.77	0.35	0.14	0.46
	(0.00)	(0.00)	(0.05)	(0.00)	(0.00)	(0.03)
Father name commonness	0.89	0.64	0.99	0.79	0.56	0.86
	(0.00)	(0.01)	(0.03)	(0.00)	(0.01)	(0.02)
Mother name commonness	1.11	0.85	1.24	1.05	0.84	1.09
	(0.00)	(0.01)	(0.03)	(0.00)	(0.01)	(0.02)



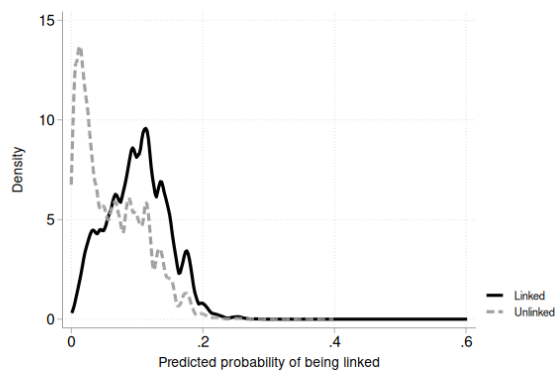
(a) 1850-1870 Marriages



(b) 1860-1880 Marriages

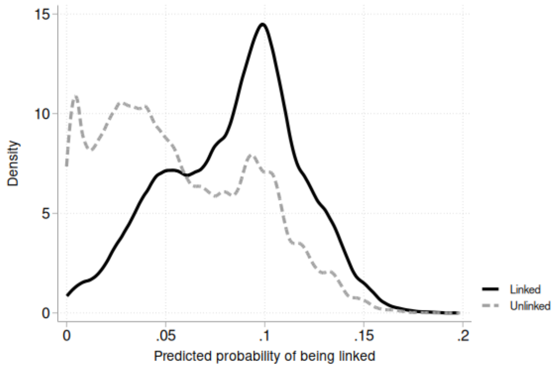


(c) 1880-1900 Marriages

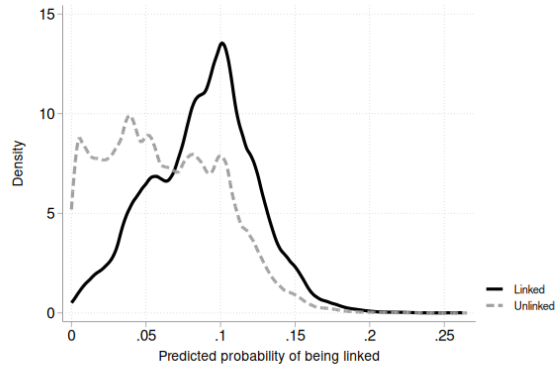


(d) 1900-1920 Marriages

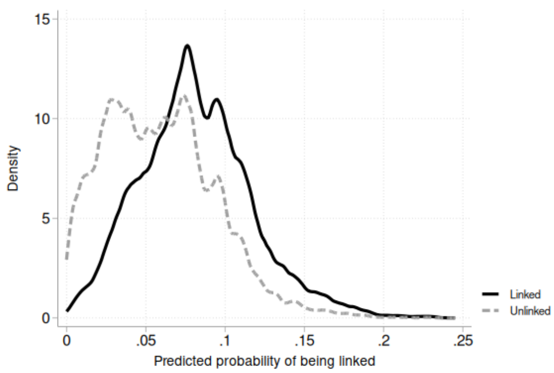
Figure A1: Men: Overlap of predicted probability of being in the linked sample by cohort



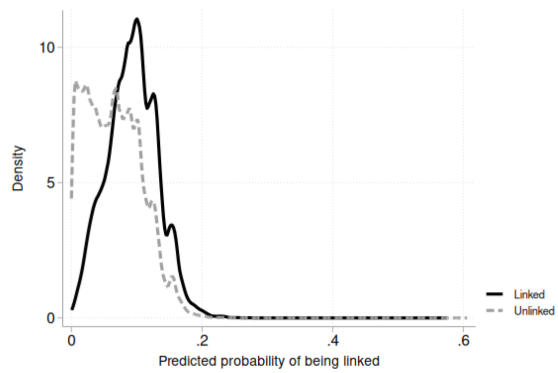
(a) 1850-1870 Marriages



(b) 1860-1880 Marriages



(c) 1880-1900 Marriages



(d) 1900-1920 Marriages

Figure A2: Women: Overlap of predicted probability of being in the linked sample by cohort

Table A10: Representativeness of children in linked sample

	Unweighted			Weighted		
	Population	Linked sample	Diff (Linked - Pop)	Population	Linked sample	Diff (Linked - Pop
	Panel A: Women					
Child age	8.32	10.88	2.75	8.12	8.01	-0.12
	(0.01)	(0.02)	(0.02)	(0.01)	(0.08)	(0.08)
Urban	0.58	0.49	-0.09	0.58	0.58	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
Reside in MA	0.92	0.98	0.07	0.91	0.89	-0.02
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Immigrant father	0.42	0.25	-0.18	0.43	0.43	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
Literate father	0.95	0.97	0.02	0.95	0.95	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Dad occscore	22.97	23.04	0.08	22.96	22.92	-0.04
	(0.01)	(0.04)	(0.05)	(0.01)	(0.12)	(0.12)
Father's age	41.00	43.59	2.78	40.81	40.71	-0.10
	(0.01)	(0.04)	(0.04)	(0.01)	(0.11)	(0.12)
Mother present	0.97	0.97	0.01	0.97	0.97	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Immigrant mother	0.41	0.24	-0.18	0.42	0.43	0.01
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
Literate mother	0.94	0.97	0.04	0.93	0.93	-0.01
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	Panel B: Men					
Child age	8.33	12.07	3.97	8.08	7.75	-0.35
	(0.01)	(0.02)	(0.02)	(0.02)	(0.21)	(0.21)
Urban	0.57	0.48	-0.10	0.58	0.59	0.01
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
Reside in MA	0.92	0.98	0.06	0.91	0.86	-0.05
	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.02)
Immigrant father	0.42	0.23	-0.19	0.43	0.46	0.03
	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.02)
Literate father	0.95	0.98	0.03	0.95	0.95	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Dad occscore	22.90	22.94	0.04	22.89	22.70	-0.21
	(0.01)	(0.05)	(0.05)	(0.02)	(0.22)	(0.22)
Father's age	40.96	44.79	4.07	40.71	40.52	-0.20
	(0.01)	(0.04)	(0.04)	(0.01)	(0.18)	(0.18)
Mother present	0.97	0.97	0.01	0.97	0.97	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Immigrant mother	0.41	0.23	-0.18	0.42	0.44	0.02
	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.02)
Literate mother	0.94	0.97	0.04	0.93	0.93	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)

Notes: Representativeness of linked sample with respect to the son or daughter observation in the childhood census.

Table A11: Representativeness of children in linked sample: 1850

	Women			Men		
	Population	Linked unweighted	Linked weighted	Population	Linked unweighted	Linked weighted
Child age	8.46	10.14	8.03	8.44	11.63	6.78
	(0.01)	(0.05)	(0.10)	(0.01)	(0.05)	(0.63)
Urban	0.30	0.20	0.32	0.29	0.18	0.31
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.04)
Reside in MA	0.90	0.98	0.88	0.90	0.98	0.76
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.07)
Immigrant father	0.15	0.05	0.16	0.15	0.03	0.24
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.06)
Literate father	0.97	0.99	0.96	0.97	0.99	0.97
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)
Dad occscore	23.49	22.70	23.69	23.27	22.43	22.26
	(0.03)	(0.09)	(0.16)	(0.03)	(0.10)	(0.91)
Father's age	41.24	43.00	40.70	41.22	44.68	39.93
	(0.02)	(0.09)	(0.18)	(0.02)	(0.09)	(0.63)
Mother present	0.97	0.98	0.97	0.97	0.98	0.97
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Immigrant mother	0.13	0.04	0.15	0.13	0.03	0.15
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.03)
Literate mother	0.95	0.99	0.94	0.95	0.99	0.96
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)
Mother's age	37.25	39.11	36.88	37.22	40.81	35.57
	(0.02)	(0.08)	(0.15)	(0.02)	(0.09)	(0.79)

Notes: Representativeness of linked sample with respect to the son or daughter observation in the childhood census.

Table A12: Representativeness of children in linked sample: 1860

	Women			Men		
	Population	Linked unweighted	Linked weighted	Population	Linked unweighted	Linked weighted
Child age	8.16	10.28	7.82	8.18	11.44	7.31
	(0.01)	(0.04)	(0.09)	(0.01)	(0.04)	(0.23)
Urban	0.72	0.68	0.73	0.72	0.67	0.74
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)
Reside in MA	0.88	0.98	0.85	0.88	0.97	0.84
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)
Immigrant father	0.33	0.12	0.36	0.33	0.10	0.41
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.02)
Literate father	0.94	0.98	0.93	0.94	0.98	0.93
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)
Dad occscore	21.22	21.48	21.26	21.07	21.40	21.36
	(0.03)	(0.10)	(0.18)	(0.03)	(0.11)	(0.32)
Father's age	40.67	43.13	40.29	40.65	44.32	39.56
	(0.02)	(0.08)	(0.15)	(0.02)	(0.08)	(0.35)
Mother present	0.97	0.98	0.97	0.97	0.98	0.97
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Immigrant mother	0.32	0.12	0.35	0.32	0.09	0.40
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.02)
Literate mother	0.91	0.97	0.89	0.91	0.97	0.87
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.02)
Mother's age	36.49	38.87	36.10	36.45	40.20	35.11
	(0.02)	(0.07)	(0.13)	(0.02)	(0.08)	(0.33)

Notes: Representativeness of linked sample with respect to the son or daughter observation in the childhood census.

Table A13: Representativeness of children in linked sample: 1880

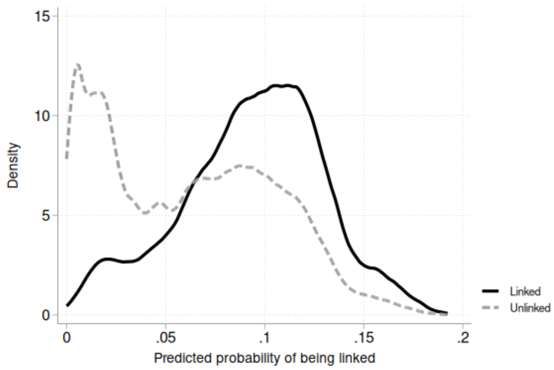
	Women			Men		
	Population	Linked unweighted	Linked weighted	Population	Linked unweighted	Linked weighted
Child age	8.58	10.17	8.36	8.63	11.53	8.10
	(0.01)	(0.04)	(0.07)	(0.01)	(0.04)	(0.14)
Urban	0.50	0.41	0.51	0.49	0.38	0.51
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)
Reside in MA	0.91	0.97	0.90	0.91	0.98	0.89
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)
Immigrant father	0.50	0.30	0.52	0.50	0.28	0.54
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)
Literate father	0.99	1.00	0.99	0.99	1.00	0.99
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Dad occscore	24.31	24.53	24.44	24.21	24.27	24.05
	(0.02)	(0.08)	(0.09)	(0.02)	(0.08)	(0.17)
Father's age	41.73	43.29	41.50	41.71	44.64	40.77
	(0.02)	(0.07)	(0.10)	(0.02)	(0.07)	(0.23)
Mother present	0.96	0.97	0.97	0.96	0.97	0.97
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Immigrant mother	0.48	0.30	0.50	0.48	0.29	0.52
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)
Literate mother	0.99	0.99	0.99	0.99	0.99	0.99
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Mother's age	37.38	38.78	37.15	37.37	40.25	36.59
	(0.02)	(0.06)	(0.09)	(0.02)	(0.06)	(0.20)

Notes: Representativeness of linked sample with respect to the son or daughter observation in the childhood census.

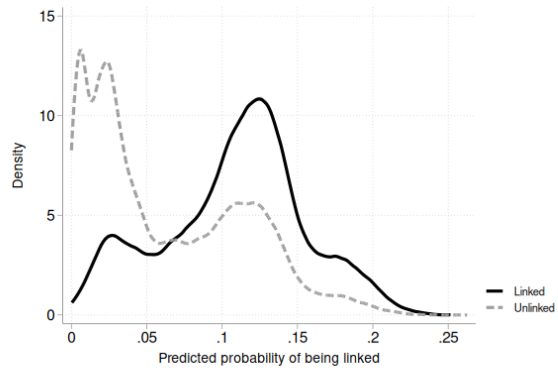
Table A14: Representativeness of children in linked sample: 1900

	Women			Men		
	Population	Linked unweighted	Linked weighted	Population	Linked unweighted	Linked weighted
Child age	8.15	12.14	7.87	8.15	13.19	8.26
	(0.01)	(0.03)	(0.20)	(0.01)	(0.03)	(0.48)
Urban	0.66	0.58	0.67	0.66	0.57	0.70
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.02)
Reside in MA	0.94	0.98	0.93	0.94	0.98	0.91
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.02)
Immigrant father	0.52	0.38	0.51	0.52	0.37	0.55
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.03)
Literate father	0.93	0.95	0.92	0.93	0.95	0.93
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)
Dad occscore	22.76	22.97	22.46	22.82	22.97	22.72
	(0.02)	(0.08)	(0.29)	(0.02)	(0.09)	(0.33)
Father's age	40.58	44.38	40.40	40.51	45.30	41.22
	(0.02)	(0.06)	(0.27)	(0.02)	(0.06)	(0.26)
Mother present	0.96	0.97	0.96	0.96	0.97	0.97
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Immigrant mother	0.52	0.38	0.53	0.52	0.38	0.56
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.03)
Literate mother	0.91	0.94	0.90	0.91	0.94	0.91
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)
Mother's age	36.84	40.46	36.67	36.77	41.44	37.60
	(0.01)	(0.05)	(0.33)	(0.01)	(0.06)	(0.45)

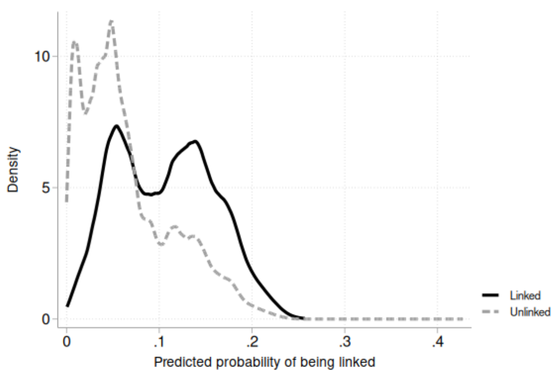
Notes: Representativeness of linked sample with respect to the son or daughter observation in the childhood census.



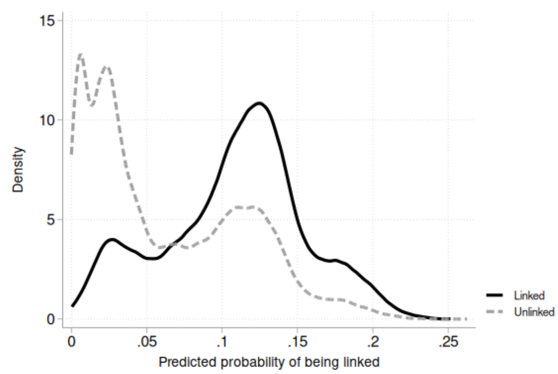
(a) 1850 Census



(b) 1860 Census



(c) 1880 Census



(d) 1900 Census

Figure A3: Overlap of predicted probability of being in the linked sample by child year

C Measuring Status

Income Scores

We create income based scores following Abramitzky et al. (2021) by using a statistical model to predict income for white men aged 18-65 in the 1940 census. We then use the results of this model to predict income for men in earlier census years in which information on income does not exist. To create these "income scores", we regress log 1940 wage and salary income on indicators for the 3-digit occupation (IPUMS *occ1950* variable), state of residence, age and age-squared, country of origin, 1-digit occupation, and interactions of census division, country of origin, and occupation. The 1940 Census excludes self-employment from the income variable, and thus the majority of farm income is not included. We follow Collins and Wanamaker (2022) in taking the ratio of income of farmers relative to farm laborers in the 1960 census and applying to the wages of farm laborers in 1940 to estimate the income of farmers. Farm income is specific to region and immigration status, but not country of origin. In-kind payment for services made up a significant portion of compensation for farm laborers and farm managers. We thus scale up the income for these occupations by an additional 26 percent using the calculations from Collins and Wanamaker (2022).

We also follow the method used in the economic history literature by measuring status with the IPUMS *occscore* variable, which is calculated from an occupation's median income in the 1950 census, placing each occupation in the 1950 occupational income distribution. To make comparable estimates of mobility for women to those in Olivetti and Paserman (2015), we report results using the *occscore* variable. However, the income distribution from 1950 is radically different than that for the nineteenth century, and placing each historical occupation in the list of occupations available from the 1950 census is not straightforward.

The 1901 Cost of Living Survey also provides income information by occupation during the middle of our study period. To begin, we estimate income by using the 1900 occupational earnings distribution and tabulated by occupation by Preston and Haines (1991). The survey was aimed at examining the living expenses of families residing in industrial areas within the United States. However, the 1901 survey gathered data for the standard urban family. This results in a compressed income distribution relative to a nationally representative sample. Moreover, we find significant regional differences in wealth in 1870, as well as regional differences in income in 1940. The Preston and Haines estimates of occupational income do not allow for any regional differentiation in occupational income scores.

Human Capital Scores

Song et al. (2020) measure the status of an occupation based on human capital. They calculate the average human capital level in an occupation x 10-year birth cohort cell, relative to the human capital distribution of that birth cohort. Between 1850 and 1930, literacy is used the measure of human capital, and from 1940 onward education is the measure of human capital. The measure is built up from the decennial census microdata samples. The main benefit of this score is that it allows for relative changes in the status of occupations over time as occupational literacy gaps evolve. Using their measure, Song et al. (2020) show that a number of initially high-status occupations in 1850 lost status by the 20th century. The drawbacks from this measure are twofold. First, Song et al. (2020) did not calculate status by region, not allowing for regional differences in human capital. Secondly, there is no literacy gap between immigrants and the native-born population throughout most of the period under study. To the extent that wealth and income gaps exist and are important for the definition of "status" desired by the researcher, the Song score may bias levels and trends in mobility.

C.1 Sensitivity of Results to Measures of Economic Status

Table A15: Mobility Results with Different Measures of Economic Status

	1850-1870	1860-1880	1880-1900	1900-1920
	Panel A: 1901 Cost of Living Survey			
Father Rank (β_1)	0.203	0.213	0.208	0.189
	(0.010)	(0.009)	(0.008)	(0.008)
Father Rank x Woman (β_2)	-0.033	-0.025	-0.017	-0.009
	(0.013)	(0.012)	(0.011)	(0.010)
N	20,320	25,001	31,985	36,826
	Panel B: 1940 Income Scores			
Father Rank (β_1)	0.298	0.249	0.266	0.225
	(0.015)	(0.012)	(0.011)	(0.010)
Father Rank x Woman (β_2)	-0.076	-0.050	-0.036	-0.020
	(0.019)	(0.015)	(0.014)	(0.013)
N	20,088	24,707	31,461	36,255
	Panel C: 1950 Occscore			
Father Rank (β_1)	0.225	0.242	0.267	0.216
	(0.011)	(0.009)	(0.009)	(0.008)
Father Rank x Woman (β_2)	-0.067	-0.052	-0.052	-0.050
	(0.015)	(0.012)	(0.012)	(0.011)
N	20,320	25,001	31,982	36,811
	Panel D: Song Scores			
Father Rank (β_1)	0.220	0.222	0.269	0.224
	(0.011)	(0.010)	(0.007)	(0.008)
Father Rank x Woman (β_2)	-0.047	-0.053	-0.053	-0.022
	(0.015)	(0.014)	(0.010)	(0.011)
N	19,975	24,828	31,633	35,712

Notes: This table reports coefficients from equation 11 estimated separately on each cohort with heteroskedasticity-robust standard errors in parentheses. Each panel uses a different measure of occupational status. See Appendix Section C for details on the construction of each occupational score. All regressions include quartics in the age of the father and the age of the husband at the time economic status is observed. Reported coefficients are evaluated at age 35 for both the father and the adult. Regressions are unweighted.
Sources: 1850-1920 Decennial Census data from Ruggles et al. (2017). Marriage registrations from *FamilySearch.org*. Song Scores from Song et al. (2020). Occupation scores based on the 1901 Cost of Living Survey are taken from Abramitzky et al. (2021).

D Comparison to Pseudo-Linking Methodology

Olivetti and Paserman (2015) make the first attempt at estimating long-run female mobility in the United States separately from males. Without the ability to link a female to her childhood economic status due to name change at the time of marriage, they create a pseudo-link using the assumption that names convey socioeconomic status. They identify the occupational income score (*occscore*) of an individual in a specific census and calculate the average occupational income score for all fathers in the previous census who have a child with that individual's name. For example, for an individual named "John", the income level of his father is calculated as the average income of all fathers in the previous census with a son named "John". They use these pseudo-links to calculate estimates for intergenerational elasticity of income for both men and women from 1850-1940. As a majority of women did not work during this time period, intergenerational elasticity of income for women was calculated as the elasticity between a female's husband and a female's father. Their method does not capture a "true" estimate for elasticity of income because of the absence of direct links between generations. Rather, it calculates a measure that can be compared over time, assuming equal bias over time and across genders. Due to the methodological differences, our estimate of intergenerational mobility will not be comparable in magnitude to Olivetti and Paserman (2015), but will be useful to compare the mobility of women relative to men and the trends over time.

Our results consistently demonstrate that female mobility is higher than male mobility, when applying a direct-linking methodology between fathers and their children. To identify how our estimates compare with Olivetti and Paserman (2015) pseudo-linked estimates, we apply their pseudo-linking methodology to the sample of direct links from Massachusetts marriages. This section uses 30-year cohort links created in a previous version of this paper. However, the comparison of the methods is still valid because we apply pseudo-link occupational status scores to our direct links. We compare our direct and pseudo-linked estimates to Olivetti and Paserman (2015)'s northeastern region elasticities, as our dataset consists mainly of individuals from Massachusetts. To do so, we take all male and female first names from the full sample of direct matches and impute the pseudo-linked father's occscore based on all fathers in the United States.[21] On average, both the pseudo-linked and direct-linked father's occscore in 1880 are less than the husband's occscore in 1910. Significantly less variation in occscore occurs for pseudo-linked fathers, which is not surprising given that this variable is already an average of occscores for fathers with similarly named children.

We perform identical regressions as before, except replacing the direct-linked father's occcore with the imputed pseudo-linked father's occscore. Results show an elasticity of income for females ranging from 0.217 to 0.303 and for males ranging from 0.178-0.374. We compare these estimates to the Northeast region results reported by Olivetti and Paserman (2015). They find an elasticity

[21] About 100 first names did not exist as a child in the 1880 census, therefore preventing the imputation of the father's occscore. These names can be assumed to have some sort of small error that prevented exact matching to a child in the 1880 census, but allowed for the probabilistic matching process to identify a direct match.

of income from 1880-1900 for females of 0.3111 and for males of 0.1677, an increase in mobility for men and a decrease in mobility for women. Pseudo-linked results on the Massachusetts sample also find an increase in mobility for men, but of a smaller magnitude, but, instead, a decrease in mobility for women as well. We find some striking differences between the methods. The direct-linking procedure on the sample of couples finds the IGE for men to be greater than women in both cohorts, with decreases from the 1850-1880 to the 1880-1910 cohort. Depending on whether we use *all fathers* in the census, *fathers residing in New England*, or *fathers born in New England*, we find that the relative difference in the IGE between men and women changes. In the all fathers case, we get a reversal in the trend, an increase in persistence, whereas direct-linking finds a decrease in persistence. Limiting the sample pool of fathers to more closely align with the underlying population for which estimates are being made tends to lead estimates more in line with the direct-linking procedure. We suggest that researchers leverage the complete count census microdata recently made available for the historical U.S. censuses and construct a pool of fathers that mimics the underlying population of daughters in the sample when using the pseudo-linking procedure.[22]

[22] At the time, the complete count censuses were not available to (Olivetti and Paserman, 2015).

Table A16: Direct-linking comparison to pseudo-linking

	Direct-linking		Pseudo-linking					
			All fathers		Fathers reside in NE		Fathers born in NE	
	Men	Women	Men	Women	Men	Women	Men	Women
Panel A: Massachusetts Data								
Occscore IGE (1850-1880)	0.278	0.222	0.374	0.217	0.334	0.303	0.372	0.278
	(0.008)	(0.012)	(0.056)	(0.047)	(0.055)	(0.054)	(0.055)	(0.051)
Obs	9,568	9,568	9,568	9,568	9,568	9,568	9,568	9,568
Occscore IGE (1880-1910)	0.209	0.172	0.178	0.249	0.258	0.250	0.268	0.249
	(0.006)	(0.009)	(0.047)	(0.044)	(0.049)	(0.044)	(0.049)	(0.044)
Obs	18,021	18,021	18,021	18,021	18,021	18,021	18,021	18,021
Panel B: Olivetti & Paserman (2015) Estimates								
Occscore IGE (1850-1870)			0.35	0.34	0.295	0.201		
			(0.02)	(0.02)	(0.04)	(0.04)		
Occscore IGE (1880-1900)			0.344	0.399	0.168	0.311		
			(0.02)	(0.02)	(0.03)	(0.04)		

Heteroskedasticity robust standard errors reported in parentheses. All estimates are statistically significant at the 1 percent level. Direct-linking reprints our main estimates of mobility from Table 1. Pseudo-linking refers to creating father-child pairs by imputing father's occupational income score using the mean occscore of father's in a given pool that have a child listed in the household with the same given name as the child observed as an adult in a later census. The pool of fathers over which mean occscore is computed are: all fathers in childhood census, fathers residing in New England at time of childhood census, and fathers born in New England at time of childhood census. Panel B reprints IGE estimates for the Northeast region from table 8 of Olivetti and Paserman (2015).