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

Katherine Eriksson Gregory T. Niemesh Myera Rashid

Jacqueline Craig

[Preliminary Draft – Do Not Circulate]

Abstract

The literature finds a high degree of economic mobility for men in the nineteenth century in comparison to today. However, due to data limitations, changes in female economic mobility over time are not well understood. Using a set of marriage certificates from Massachusetts over the period of 1850-1920, we link men and women 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 cohort, 1850-70. Men's mobility increases by the 1880-1900 cohort, whereas women's does not, leading to a convergence. We also find evidence of increased assortative mating over the same cohorts, as measured by the slope coefficient from a regression of the occupational standing of the husband's father on the occupational standing of the wife's father. 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 (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 Northwestern University, Yale University, University of California - Berkeley, University of Michigan, 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 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.[1]

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 in-

[1] Jácome et al. (2021) use survey data to obtain mobility estimates for women in the 20^{th} century US.

tergenerational 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[2]. The second limitation to estimating father-daughter economic mobility is the lack of historical information on women's *potential* labor market outcomes. Because 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)[3]. Accordingly, we measure women's intergenerational mobility by estimating correlations in economic status between the husband and the wife's father.

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, but it is not without drawbacks. The large amount of information recorded on marriage certificates and censuses requires the use of a probabilistic method to create record linkages, 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, 2017). 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. Using an exact match method with unique names is too stringent in this case (Ferrie, 1996). The issue becomes how to balance small differences in each of these dimensions when

[2]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 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.

[3]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).

deciding the correct the match. The problem is even more severe for the second match – post-marriage census to the pre-marriage census. Here, four strings are used in the match – child's given name and surname, father's given name, and mother's given name – and 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 use multiple measures to capture intergenerational mobility: intergenerational elasticities, rank-rank estimates and occupational transitions. Men's mobility is estimated as the association between the occupations of the husband and the husband's father, whereas we use the association between the occupations of the husband and the wife's father for female mobility. We must estimate mobility through the male the wife is connected with in both generations because a majority of females do not participate in market work. 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). The rank-rank parameter between fathers is 0.185 for the 1850-1870 cohort and 0.287 in the 1880-1900 cohort, a 55 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 to understand the relationship between marital sorting and intergenerational mobility. Using counterfactual mobility estimates obtained as a result of the model, we establish that the mobility of women in the 1880-1900 cohort would have been higher than observed if they faced the lower level of assortative mating of 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 what has been established so far in the intergenerational mobility literature. Section 3 describes our data sources and linking method. Section 4 develops a theoretical framework to capture the relation between assortative mating and intergenerational mobility. Section 5 presents the main results on intergenerational mobility. Section 6 presents the results on assortative mating. Section 7 uses the theoretical framework to determine counterfactual mobility in the absence of marital sorting. Section 8 concludes.

2 Background

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 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 flips 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) create a pseudo-link 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. While this method does provide an avenue for uncovering female mobility, it may face attenuation bias from averaged income values and the premium or penalty associated with first names on the labor market. They have since extended this method to include three generations and found that grandparents do matter in mobility level of their children and grandchildren (Olivetti et al., 2016). 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 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

[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 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. In this paper, assortative mating refers to the correlation in economic status of wife's parents and the husband's parents. Even in the modern U.S. context, sorting on parental background has been shown to have an independent association separate from sorting based on education. Charles et al. (2013) show a strong correlation in parental wealth between spouses in the PSID sample, which the strong degree of sorting based on education does not diminish. Accordingly, we might expect familial background to be even more important for marriage sorting during the nineteenth century when females had less opportunity for formal education, and the importance of education for labor market opportunities was less influential (Goldin, 1983).

3 Data

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[5]. The novel part of our strategy is to use the listed maiden name recorded on marriage certificates to link adult married women

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

to a childhood census. As such, our record linkage consists of two match procedures. Couples from the marriage index are first matched to a post-marriage decennial census to observe the economic status of the adult children. 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 on the census forms and an associated occupational income or wealth score, which we describe in more detail below.

3.1 Record linkage

We begin with a digitized set of marriage indices for the Commonwealth of Massachusetts that cover the universe of registered marriages for the period 1841-1915[6]. For both the wife and husband, the marriage index lists the full name, year of birth, and full names of both sets of parents. From the population of marriages, we construct four cohorts of marriages to link to census records: 1850-70, 1860-80, 1880-1900, and 1900-1920[7]. 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[8]. 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 and to take advantage of the wealth of information we have on individuals, their spouses, and their parents[9]. 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 birth years. The issue becomes

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

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

[8]Couples marrying between 1860 and 1869 are matched to two post-marriage censuses.

[9]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 census waves (Long and Ferrie, 2013; Abramitzky et al., 2012; Collins and Wanamaker, 2014, 2015, 2017).

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. 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.

Instead, we implement Feigenbaum (2016)'s supervised machine learning algorithm and train parameters to balance transcription errors across the collection of name strings. 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, 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 93,358 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 used in the literature focusing on father-son mobility only. Note, however, that our sample of double matched couples essentially limits the analysis to *native-born* couples. Observations with an immigrant spouse are unlikely to enter our sample because we are not able to find the immigrant father's economic standing in the childhood home unless the immigrant spouse migrated with their father at 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 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

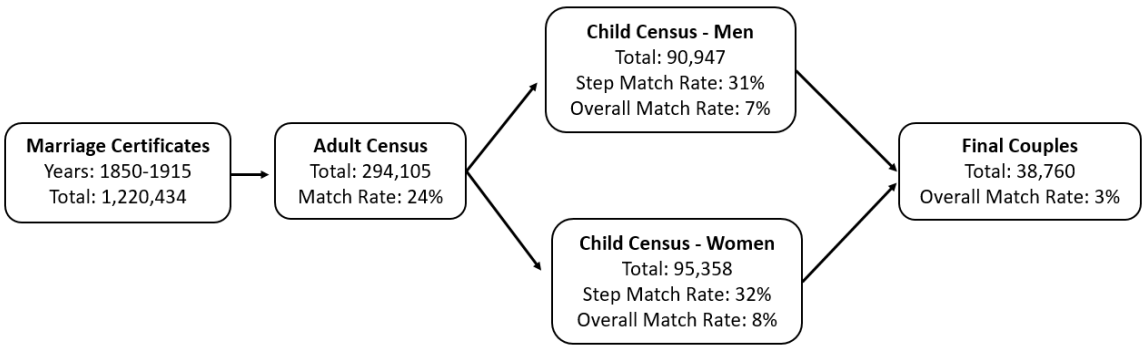


Table 1: Summary of sample sizes and match rates across the two linkage steps

	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	

3.2 Summary Statistics and Selectivity of the Sample

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 construct inverse probability weights (IPW) to adjust are linked sample to mimic the observable characteristics of the underlying population. In the first case, we want to re-weight the linked sample to represent the population of marriages registered in Massachusetts for a given cohort. In the second weighting, we want to make the linked sample mimic the population of Massachusetts-born men and women in a given childhood cohort. We will restrict our sample to observations born in Massachusetts in a number of robustness checks.

Results assessing the representativeness of the linked sample to the population of registered marriages are shown in Table A1. 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 A2 through A5 repeat the exercise separately by marriage cohorts.

In a second method of assessing the representativeness of our sample, we use the same characteristics from the marriage certificates, and functions of those characteristics, to predict the likelihood of a finding a match. 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). The propensity of being linked $P_i(L_i = 1|X_i)$ comes from estimated probit models of link status on the covariates listed in Table A1, 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.

3.3 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 - 1850, 1860, and 1870. Subsequently, all mobility estimates in the literature use reported occupations to infer economic status (Ferrie, 2005; Long and Ferrie, 2007, 2013; Olivetti and Paserman, 2015; Feigenbaum, 2017; Collins and Wanamaker, 2017). We follow the method used in the economic history literature by assigning an occupational standing "score" to each occupation. This "score" can be based on either income or wealth. One common standing measure is the *occscore*, 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. One partial solution is to use occupational income distribution information from a data source closer in time to the period under study (Abramitzky et al., 2012; Collins and Wanamaker, 2014). Another is to supplement the occupational rankings using using additional datasets (e.g. finding farmer income in the Agricultural Censuses) (Abramitzky et al., 2012; Collins and Wanamaker, 2017). None of these options are applicable to our particular case in the nineteenth century.

Instead, we follow Collins and Zimran (2018) 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. The mean provides more meaningfully variation in economic status by occupation compared to the median. A number of occupations at the lower end of the wealth distribution have median wealth values of zero. 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 – and occscore – is unable to capture the within-cell variation in wealth.

Much of the analysis will emphasize 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. As such, we want to capture the changing occupational structure of the American economy over the nineteenth century. To do this, we take each observation's wealth or income score and place it in the national distribution of scores to get the percentile rank for that cohort. For example, fathers of 1850-70 links are given a rank in the national distribution in 1850, whereas the sons of the 1850-70 links 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.

4 Theoretical Framework

This section develops a simple theoretical framework of the interaction between marriage and intergenerational mobility by adapting the model in Curtis (2021). The model helps us understand the relationship between estimated correlations in the data and underlying parameters of inheritance and marital sorting. Suppose that we are ultimately interested in a measure of economic standing, or human capital, X for individual i. The inheritance of X 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 levels of human capital, 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:

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

The ultimate measure of economic standing or human capital X 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 coefficients, and $b_f = \beta_f + \gamma\beta_m$ and $b_{lf} = \beta_{fl} + \gamma\beta_{ml}$. Already, the model shows that the estimated IGE coefficients are a combination of the inheritance terms and the degree of marital sorting on X.[10]

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 $b_{fl}^* = \gamma(\beta_{fl} + \gamma\beta_{ml}) = \gamma b_{fl}$. Fist, the IGE estimate from the husband-father-in-law regression commonly used in the literature is combination of the degree of marital sorting and the women-specific inheritance parameters. We choose to interpret this as the measure of women's intergenerational mobility because during this time period married women's economic standing was largely determined by the economic standing of the men in their lives. For our purposes, we are not so much interested in the actual strength of inheritance of human capital between fathers and daughters, as we are the persistence in economic standing between their childhood and adult years.

Second, estimates from equation 6 can be used to recover the underlying 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. The ratio of the husband-father-in-law and husband-father coefficients gives:

$$\frac{b_{fl}^*}{b_f} = \frac{\gamma(\beta_{fl} + \gamma\beta_{ml})}{(\beta_f + \gamma\beta_m)} \tag{7}$$

In the case that fathers and mothers transmit X to children at the same rate regardless of the child's gender, then the terms in parentheses cancel out of equation 7, which simplifies to the marital sorting parameter γ. Both Curtis (2021) and Clark and Cummins (2022) follow this track to recover γ and show some suggestive evidence that it was the case. On the

[10] Curtis (2021) notes that because Y_i is correlated with the u_i that the IGE estimates in equation 8 are biased towards zero as in classical measurement error case coming from: $Y_i = [\beta_f + \gamma\beta_m]X_i^f - [\beta_f + \gamma\beta_m]u_i^f + \beta_m v_i^f + e_i^h - u_i$. Curtis (2021) and Clark and Cummins (2022) remove the attenuation bias by taking ratios of correlation estimates to cancel out the bias.

other hand, Espin-Sanchez et al. (2022) find evidence of gender-specific and a child-parent gender interaction for the strength of inheritance of a noble honorific in 18th century Spain. We do not need the strength of inheritance to be the same across genders. Because we are interested in changes over time, as long as any gender-specific differences in inheritance remain stable, then a simple difference of the ratio over time will give an estimate of the change in γ. However, the data do not allow us to directly test for this. In practice, we show counterfactual mobility estimates assuming showing a range of levels and changes in gender-specific differences in inheritance.

5 Intergenerational Mobility

In this section, we estimate gender differences and long-run trends in intergenerational mobility for our sample. 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, assess robustness of estimates, and discuss differences in IGE by group characteristics.

5.0.1 Intergenerational Elasticity of Economic Status (IGE)

The most common method of estimation used in the modern U.S. literature is the intergenerational elasticity of income, which is estimated by regressing the log of a child's income measured as an adult on the log of their father's income.[11] Historical censuses prior to 1940 did not asked respondents to report their income. Instead of estimating the IGE with income, we use the log of occupational wealth scores to measure economic status. Let $Y^g_{f,i}$ be the economic status of the father and $Y^g_{c,i}$ be the economic status of the child for father-child pair i and gender $g \in \{h, w\}$ of the child, where h refers to the husband and w refers to the wife. The IGE is estimated from the following regression:

$$\ln(Y^g_{c,i}) = \alpha + \beta^g \ln(Y^g_{f,i}) + X_i\Lambda + \epsilon_i \tag{8}$$

Persistence in economic status across generations increases with the IGE (β^g), whereas mobility decreases. The level of mobility in an economy is obtained from $(1 - \beta^g)$. We refer to the estimates from equation 8 as the IGE, understanding that it is not the true IGE from a direct measure of income. For women, our IGE estimates are different from the modern interpretation. Ideally, we could capture the intergenerational elasticity of income for females by regressing the wife's income on her father's income to gain a direct estimate

[11] Reviews of this literature include Solon (1999) and Black and Devereux (2011).

of the association between a female's income level and her father's. However, the majority of married females did not participate in market work, and we must estimate female mobility through the male connected with her in each generation. Modern estimates of the IGE 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.

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. In addition to life-cycle bias, our estimates of intergenerational mobility will be increasingly attenuated in the presence of significant variance in transitory shocks to occupation (Solon, 1989, 1992). Mazumder and Acosta (2015) present evidence that persistence is largest when the father's occupation is measured late in life, and persistence is smallest when the father's occupation is measured at mid-career. They also suggest that the ideal period to measure the adult-child's occupation is at mid-career. 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). All estimates include a quartic in father's age (relative to age 40) and a quartic in the child's age at the time that economic status is measured (relative to age 40), and represented as X_i in equation 8.

5.0.2 Rank-Rank Regressions

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 (2017) and estimate a rank-rank regression in which the imposed linear relationship fits the modern data. The percentile rank of the child's economic status within the distribution of their own cohort is regressed on the percentile rank of the father's economic status within his cohort's distribution. We use rank-rank regressions as our base specification.

5.1 Mobility Results

Base results are presented in Figure 2 for rank-rank regressions and IGE regressions using our new sample linking both husbands and wives back to their fathers. Results are reported

by sex and cohort, with differences calculated for each sex between cohorts and within a cohort between sexes. Panel A reports rank-rank slope parameters and panel B reports IGE slope coefficients.

We begin by discussing comparisons of mobility over time to better place into context the differences by sex. Results for our sample of linked men suggest a rank-rank parameter for 1850-1870 of 0.231. A slight increase in persistence to 0.247 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-1900 cohorts. The decline of .038 is statistically significant and was roughly 16 percent of the 1850-70 level.

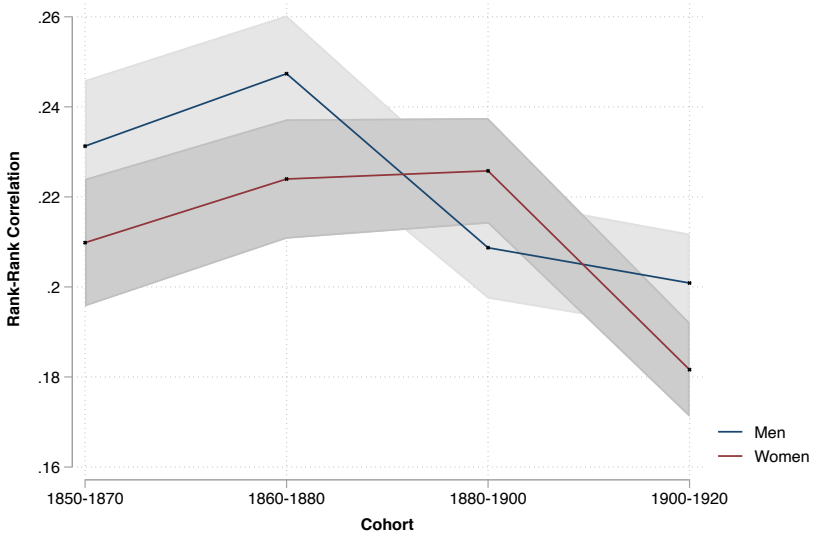
Our linked sample of married women had a different mobility experience than men over the nineteenth century. The rank-rank parameters increase from 1850-70 to 1860-80 just like for men, from 0.210 to 0.224. However, while the men rank-rank parameter begins its decline with the 1880-00 cohort, women's persistence continues to increase, or at least stays constant at around 0.226. Women's mobility did not begin to increase until at least the 1900-20 cohort, when the rank-rank parameter is estimated to 0.182 - a statistically significant difference from the 1850-70 cohort estimate.

Women were much more mobile than men in the first 1850-1870 cohort. 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 between the 1850-70 and 1900-20 cohorts. The rank-rank parameters converged between the sexes by the 1880-00 cohort, given that mobility for men increased, whereas for women it remained constant or even decreased. The comparison had flipped by this cohort so that women were less mobile than men by a statistically significant 0.017 difference in the rank-rank parameter. Over the course of the following 20 years, the rank-rank parameter for women decreased faster than for men, once again flipping the comparison. For the 1900-20 cohort, women were once again more mobile than men (0.182 vs 0.201).

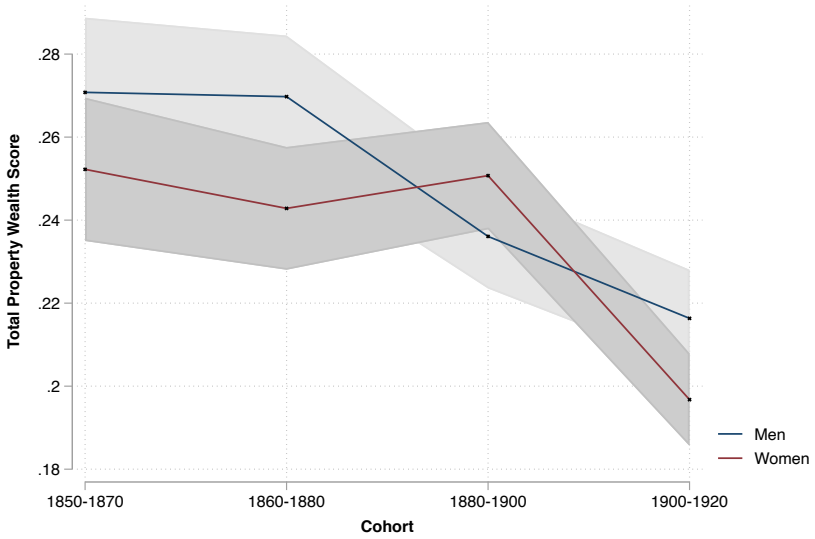
Intergenerational elasticity estimates – panel B of Figure 2 – support the same interpretation as the rank-rank regressions. Men's economic status was more persistent than women's in the 1850-1870 cohort, but had mostly converged by the post-1900 cohorts. Similar results between the two specifications is not surprising given that the "IGE" estimates are from a log-log regression on wealth scores, which are occupation cell means in wealth. Applying the same score to each observation within an occupation apparently removes much of the non-linearity between father and son in the tails of the distribution, reducing the need for the rank-rank specification (Corak and Heisz, 1999; Chetty et al., 2014).

Figure 2: Intergenerational Mobility Estimates

(a) Rank-Rank Correlations



(b) Intergenerational Elasticities



Notes: Each entry is an estimate of intergenerational mobility from separate regressions using different measures of economic status and for different samples and time periods. Shaded areas plot 95 percent confidence intervals. Panel A regresses the percentile rank of the husband's score within his cohort's distribution against the percentile rank of the father's score within his cohort's distribution. Panel B regresses the husband's score on the father's score. 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. *Sources:* 1850-1920 Decennial Census data from Ruggles et al. (2017). Marriage certificates from *FamilySearch.org*.

The magnitude and direction of our IGE estimates are at odds with the headline pseudo-linking results from Olivetti and Paserman (2015). We find IGE estimates double the magnitude found with pseudo-linking, however the difference in magnitudes cannot be directly compared. The pseudo-linking procedure introduces bias in the level of the IGE parameter, whereas our direct-linking procedure provides the true IGE parameter for our sample.[12] What we are concerned with here is a comparison of the *trend* and *sex differences* within a linking method. 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 the course of the nineteenth century, with full convergence between the sexes in a linked 1900-1920 sample.

The discrepancies between our work and that of Olivetti and Paserman (2015) can be potentially attributed to differences in the geographic restrictions on the sample (U.S. vs. Massachusetts), differences in time periods (20-year windows vs. 30-year windows), and differences in the linking methodology. The geographic differences do not seem to fully explain the discrepancies. Olivetti and Paserman (2015) construct separate estimates for each region of birth sample. Which sex is more mobile in the Northeast region varies from period to period in a way that is inconsistent with our direct estimates. To hone in on the specific cause of the discrepancies we need to isolate on differences caused by the linking method. In brief, we find that the broader the pool of father's used to construct a pseudo-linked occupational income score, the more likely estimates diverge from those in the direct-linking procedure. Importantly, using different pools of fathers can switch the sign on the sex difference and the trend over time.

To summarize, both methods and all measures of economic status provide results that are in agreement. The intergenerational persistence of economic status was higher for men than women in the 1850-1870 sample. 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 married women rarely participated in market work, or at least did not report a market work occupation on census enumerations. Intergenerational persistence in status can only be measured using the husband's economic status in place of the female's, and mobility is highly depended on the availability of potential husbands. The choice of a husband is not random, but determined in a marriage market equilibrium. After estimating mobility by subgroups, we turn to understanding changes in assortative mating, and its role in the change in mobility.

[12] Noting that our measure may suffer from other biases.

5.1.1 Subgroup Analysis

To attempt to describe the underlying factors behind this high level of mobility, we consider how different childhood conditions or marriage market outcomes may have contributed to the low level of persistence and its evolution.

Urban vs. Rural

First, we examine the impact of growing up in an urban vs. a rural area in Figure 3. Urban is defined as towns and incorporated places with at least 2,500 in population, and rural is defined as the complement.[13] We estimate rank-rank regressions separately for the urban and rural samples, and plot the rank-rank parameters for women in Panel A and men in Panel B.

Persistence is higher for men and women who grew up in urban areas, with the gap at its largest in the earliest cohort. For men in the 1850-70 cohort, the urban rank-rank parameter was 0.341 for urban and 0.209 for rural, a difference equal to 63 percent of the rural rank-rank parameter. The urban-rural difference was similar for women in the earliest cohort. The gap essentially disappeared by the 1860-80 cohort as persistence increased for rural men and decreased for urban men. The general trend over the next 40 years was for increasing mobility, however, the decline occurred first for rural men in the 1880-00 cohort, followed by urban men during the 1900-20 cohort. These results are similar to modern findings, in that rural raised children are more mobile than urban raised children. Chetty et al. (2014) suggests important geographic differences in mobility; children from rural areas were more likely to transition upward than children born in urban areas.

The meanings of urban and rural might be different in Massachusetts where the movement out of agriculture occurred much earlier than the rest of the nation, and because a significant amount of manufacturing occurred in "rural" towns during the nineteenth century. To explore these differences in more detail, we next estimate persistence for men and women by whether they resided on a farm in the childhood census.[14] Results are reported in Figure 4. We do not find any difference between mobility by childhood farm status in the 1850-70 cohort. However, for husbands who grew up on a farm and married in the 1860-80 cohort, persistence had increased dramatically. By the 1880-00 cohort and after, persistence

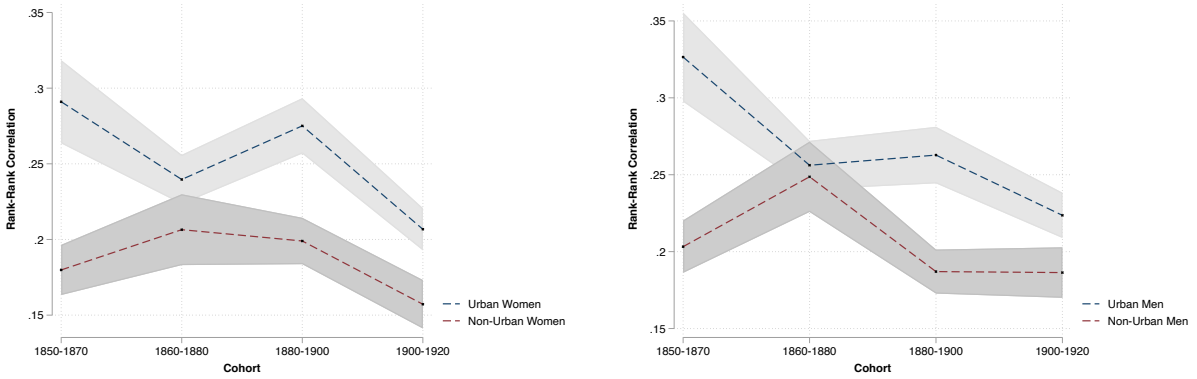
[13] We make use of the IPUMS *urban* variable. The political boundaries in Massachusetts make constructing the *urban* variable more complicated. IPUMS includes the following in the variable description *"Also includes households in Massachusetts towns (townships) containing a village or thickly settled 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."*

[14] Here we use the IPUMS *farm* variable.

continued to fall until there were no differences by childhood farm status in the 1900-20 cohort. The trends for women followed a similar to pattern as men.

Figure 3: Mobility: Urban and Rural Subgroup Estimates

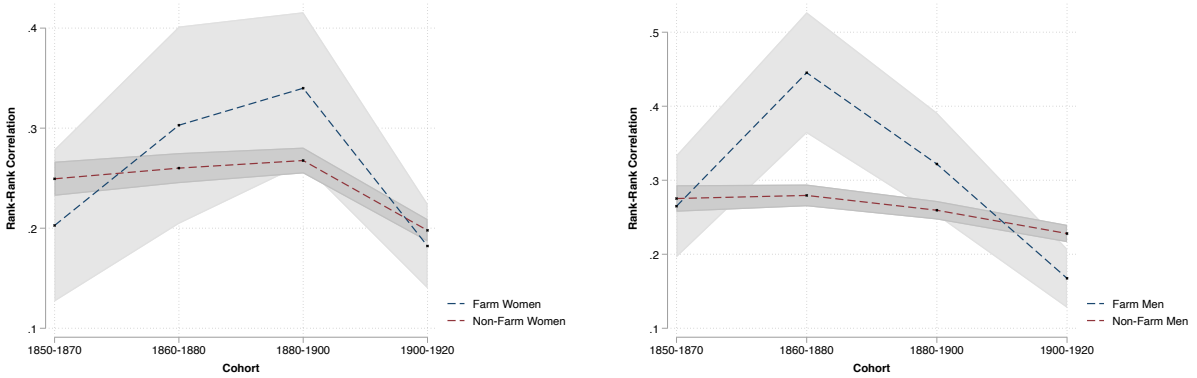
(a) Women (b) Men



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

Figure 4: Mobility: Farm and Non-farm Subgroup Estimates

(a) Women (b) Men



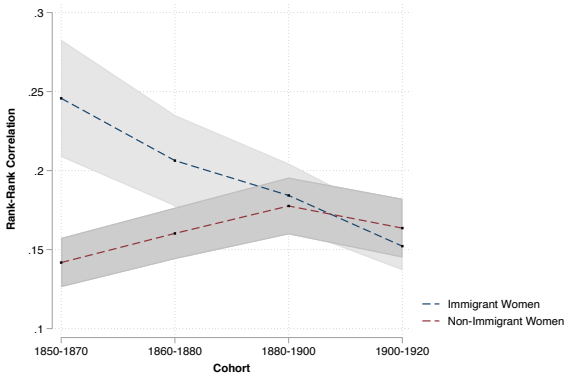
Notes: Farm is an indicator equal to one if the observation recorded residing on a farm in the childhood census.

Immigration

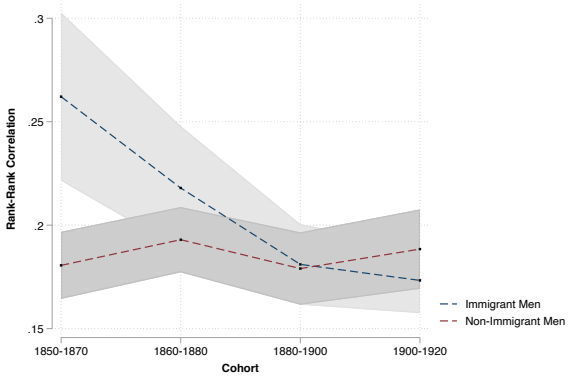
Next, we examine how mobility may have differed for children of immigrants in comparison to children of native-born parents. In our samples, about 10 percent of cohort 1

Figure 5: Mobility: Immigrant Subgroup Estimates

(a) Women



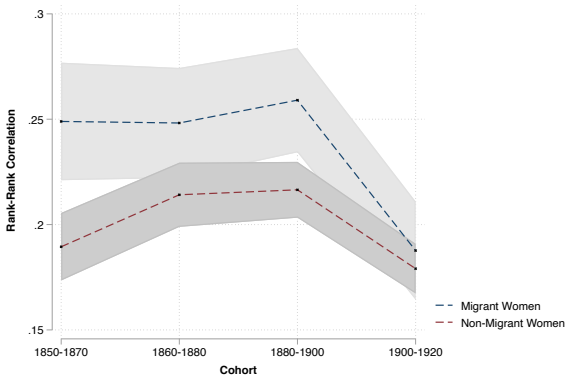
(b) Men



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 6: Mobility: 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.

individuals and 40 percent cohort 2 individuals 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. We would expect the association to be significant for the 1880-1910 period, as such an influx of immigrants should certainly have an impact on the occupational distribution, and the marriage market for the second generation children.

The parents of these children immigrated to the United States anticipating upward mobility in the "land of opportunity". Abramitzky et al. (2012) find that the economic return to migration was about 70 percent for Norwegians migrating to the United States during this period, demonstrating that, while immigration was costly, a large return resulted from this decision. Abramitzky et al. (2014) finds that immigrants did assimilate culturally over the first two generations in the United States during the "Age of Mass Migration". Children of parents that decided to give less foreign sounding names ended up with greater income levels than those that kept foreign sounding names. This assimilation led to increased incomes as adults. Over the course of several generations we might expect more intergenerational mobility for earlier-generation immigrants, and eventually more persistence as assimilation occurs.

Again, we estimate rank-rank regressions separately for the immigrant and native-born parents samples, and plot the rank-rank parameters for women in Panel A and men in Panel B. Surprisingly, we find no evidence that mobility was higher for second-generation immigrants than children born to native-born parents. In fact, the children of the earliest immigrants were much less mobile than others in the 1850-70 cohort. The difference was 0.248 vs 0.147 for women and 0.261 vs 0.179 for men. Over the subsequent cohorts, persistence fell rapidly for both men and women with immigrant parents. The gap had vanished by at least the 1880-00 cohort. The convergence between the two groups was driven almost entirely by reductions in persistence for the children of immigrant parents. The mobility levels for men born to native-born parents remained largely the same between across cohorts.

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. We find strong evidence that internal migration was not an important source of mobility for our sample. In fact, mobility was lower for internal migrants in our sample than for the non-migrants.

6 The Marriage Market and Assortative Mating

Married women in the nineteenth century United States rarely participated in market work and 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. 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 9 captures the notion of assortative mating by regressing the husband's father's occupational standing on the wife's father's occupational standing.

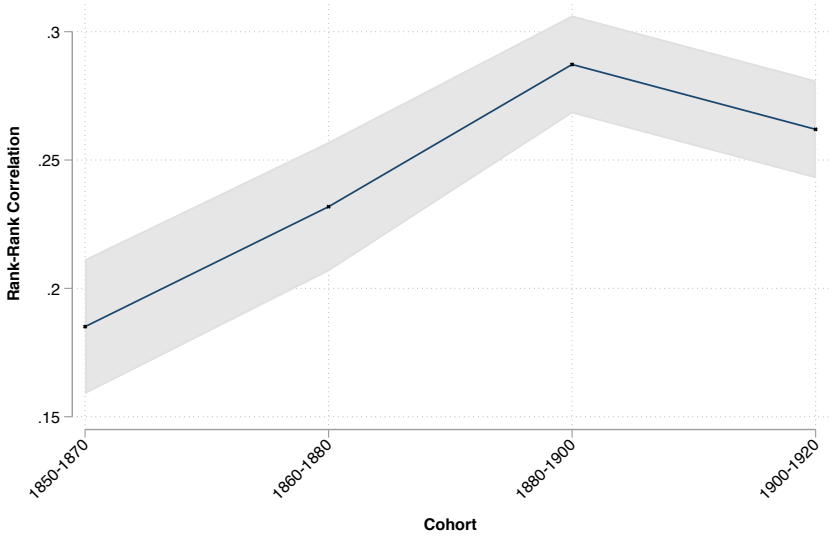
$$Y_{f,i}^{husband} = \alpha + \beta Y_{f,i}^{wife} + X_i\Lambda + \epsilon_i, \tag{9}$$

where $Y_{f,i}^{husband}$ and $Y_{f,i}^{wife}$ are the occupational standing measures for the husband's and wife's father, respectively. β provides the measure of assortative mating as a rank-rank parameter when standing is measured as the percentile rank, and provides the elasticity estimate of assortative mating when measured as the natural log of either a wealth or income score. Each regression includes a quartic in the husband's father's age and a quartic in the wife's father's age, both measured at the time of the childhood census.

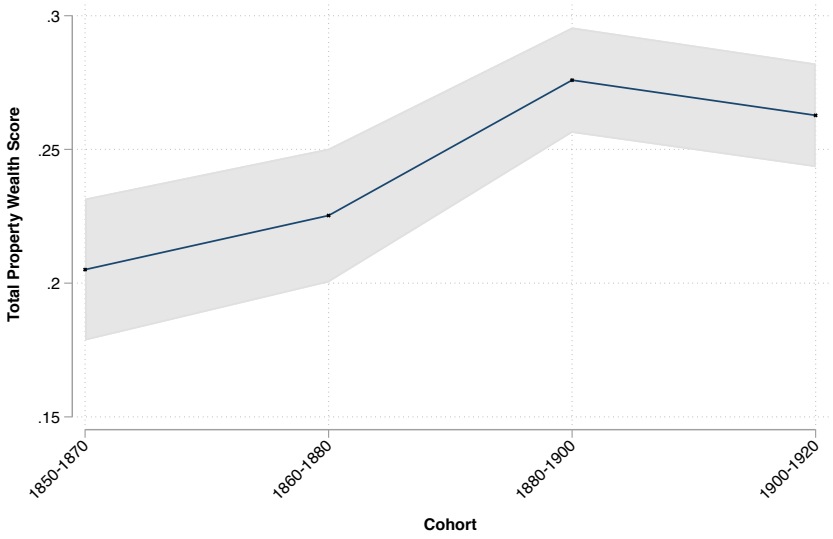
Figure 7 presents our main results on the Massachusetts marriage market in the nineteenth and early twentieth centuries. Unsurprisingly, we find evidence of assortative mating in all periods and all specifications, and can definitively say that couples did not randomly choose mates across parental socioeconomic status. The evidence is also clear about the trend in within group homogamy. In our preferred rank-rank specification using the total wealth score measure in panel A, the rank-rank parameter increased from 0.185 in the 1850-70 cohort to a peak of 0.287 in the 1880-00 cohort, and then decreased slightly to 0.262 in the 1900-20 cohort. An increase in assortative mating is also found when using the log total wealth score as seen in panel B.

Figure 7: Marital Sorting Estimates

(a) Rank-Rank Slope Parameters



(b) Log Score Slope Coefficients



Notes: Each entry is an estimate of assortative mating from regressing the husband's father's economic status on the wife's father's economic status from separate regressions using different measures of economic status and for different samples and time periods. Panel A uses the percentile rank of the father's score within his cohort's distribution. Panel B regresses log score on log score. All regressions include as controls a quartic in each father's age at the time economic status is measured.
Sources: 1850-1920 Decennial Census data from Ruggles et al. (2017). Marriage certificates from *FamilySearch.org*.

6.1 Assortative Mating: Subgroup Analysis

Figures 8 through 11 report estimates of the strength of marital sorting by subgroups. Figure 8 splits groups by rural or urban childhood. We find evidence that marital sorting may have been stronger in urban areas for both men and women. The gap was widest during the earliest cohorts and converged as sorting for both groups increased. However, the majority of the convergence was driven by increased strength of sorting in the rural group.

Figure 9 repeats the exercise by childhood farm status. The results are striking. Sorting for non-farm men and women have the same levels and follow the same trend, which mirrors the total trend in sorting found in Figure 7. Large differences are found by gender. The strength of sorting for men who resided on a farm during childhood is much lower than non-farm men throughout the entire period. Given that their fathers were farmers, this must mean that they were highly unlikely to marry daughters of farmers. However, the daughters of farmers faced sorting levels similar to non-farm women.

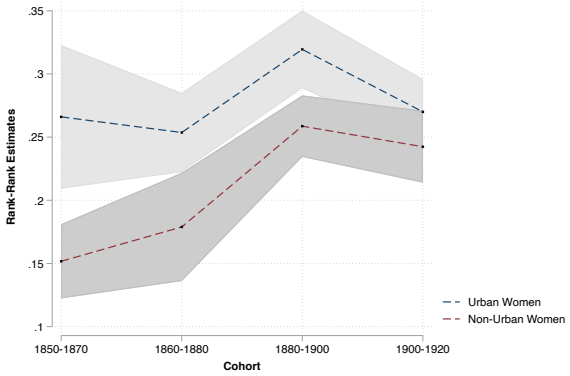
Figure 10 reports sorting estimates for children of immigrant vs native-born parents. Estimates suggest that marital sorting for the children of immigrants was stronger for both men and women in the earliest cohort. However, by the 1900-20 cohort of marriages sorting had converged between the groups. The patterns of convergence differ across genders. In the case of men born to native-born parents, sorting stays relatively constant across the cohorts. Convergence for men is driven almost entirely by the declines in marital sorting for the children of immigrant parents. In contrast, the strength of sorting declined for women of immigrant parents and increased for women born to native-born parents, roughly sharing the causation of convergence.

Note that the our sample of children of immigrants is quite restricted by the construction of the marital sorting sample. We require a match for both the husband and wife to their pre-marriage childhood census to observe the father's occupation. This restriction effectively eliminates from our sample any second-generation immigrant who marries a first-generation immigrant. Moreover, the origin countries of immigrants changed dramatically between our cohorts. The first cohort requires we observe the father in the 1850 census prior to much of the Irish, German, and southern European immigration, whereas those immigrant groups will get included in the sample for the second cohort.

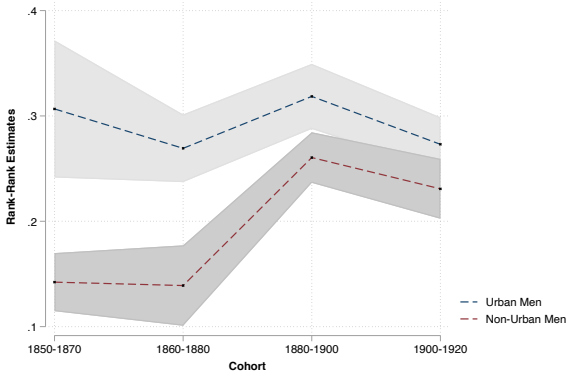
Further work is necessary to understand the causes of the changing pattern of assortative mating over the nineteenth century. Much of the work in sociology and demography has focused on the idea of a "marriage squeeze", where uneven age-specific sex-ratios leads to more, maybe better, marriage opportunities for women when the squeeze is in their favor. This literature has estimated the effects of the "marriage squeeze" on age at first marriage and the probability of marriage (Schoen, 1983; Hacker et al., 2010). Other work on marriage

Figure 8: Sorting: Urban and Rural Subgroup Estimates

(a) Women



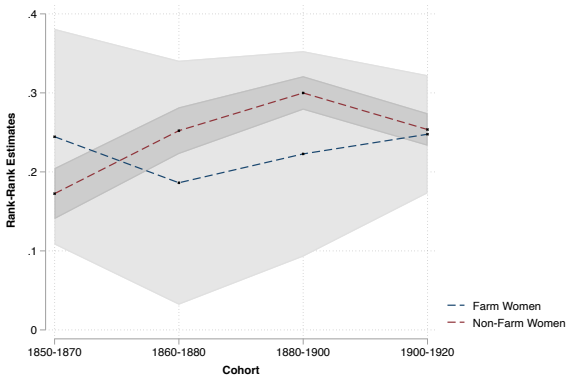
(b) Men



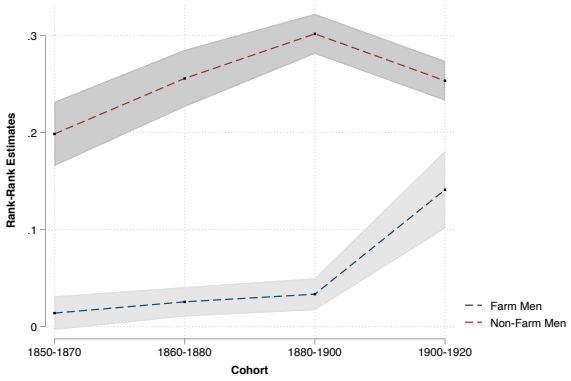
Notes: Urban is an indicator equal to one if the observation resided in an urban area as a childhood. Urban is defined following the IPUMS definition of towns and incorporated places of at least 2,500 in population.

Figure 9: Sorting: Farm and Non-farm Subgroup Estimates

(a) Women



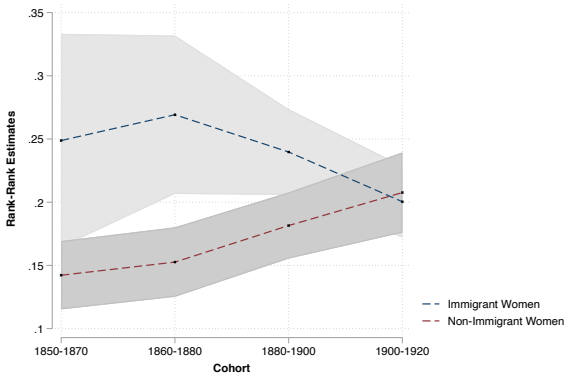
(b) Men



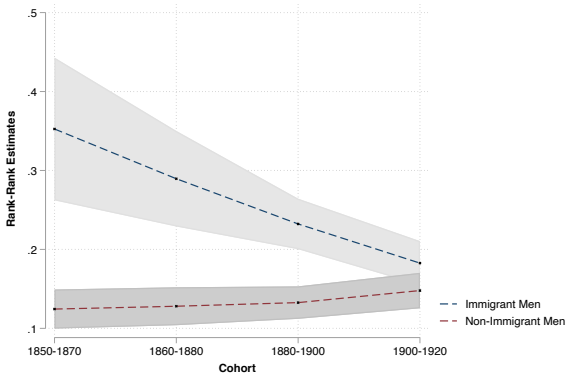
Notes: Farm is an indicator equal to one if the observation recorded residing on a farm in the childhood census.

Figure 10: Sorting: Immigrant Subgroup Estimates

(a) Women



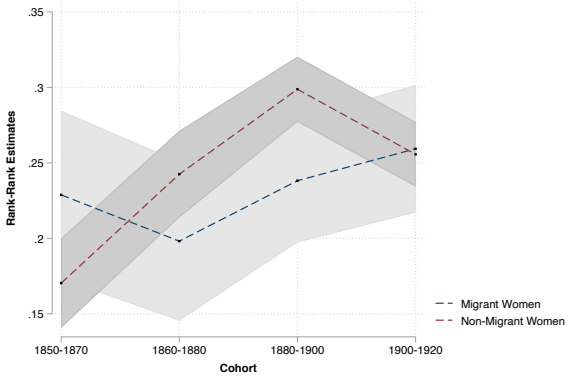
(b) Men



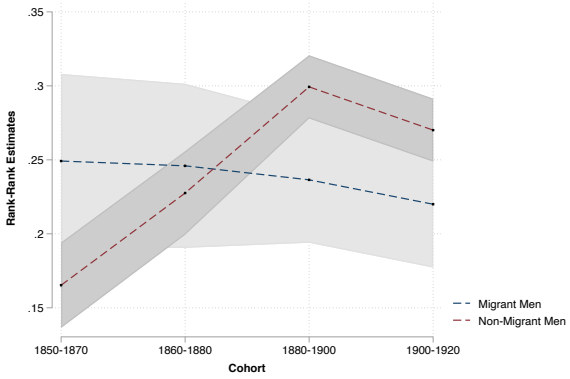
Notes: Immigrant is an indicator equal to one if at least one of the parents was born outside the United States according to FBPL and MBPL IPUMS variables in the adult census.

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.

in the nineteenth century looks at re-marriage choices of Union Army widows based on economic standing and pension income (Salisbury, 2017).

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

Finally, we rely on the theoretical framework developed in section 4 to estimate counterfactual mobility rates under different levels of marital sorting. Using the ratio of husband-father-in-law to husband-father coefficients from the rank-rank regressions in Figure 2 we get estimates of the marital sorting parameter γ.

The first set of estimates assume no differences in inheritance by gender and is simple ratio of b_w^* to b_f. According to this measure, marital sorting increased from .909 in the 1850-70 period to 1.08 in the 1880-1900 period. The second estimate of the strength of marital sorting allows for the existence of gender differences in the strength of inheritance. We take the father and mother specific inheritance estimates for sons and daughters from Espin-Sanchez et al. (2022) Table 2 and substitute them into equation 7 and then solve for γ. Allowing for gender differences meaningfully affects the estimates - now 0.915 for 1850-70 and 1.047 in 1880-00. While seemingly small, these differences are quantitatively important when used to calculate the counterfactual rank-rank parameters. In both cases the estimates of γ suggest an increase in marital sorting over time just as in our earlier analysis of the association of occupational standing between fathers and fathers-in-law. Note that these estimates do not allow for any change in gender-specific inheritance over time.

The last part of Table 2 provides counterfactual estimates of women's intergenerational mobility if marital sorting for the 1880-1900 cohort had stayed at its 1850-1870 level and not increased. These estimates are calculated as the Husband/Father rank-rank parameter from 1880-1900 times the γ estimate for 1850-1870. The method of calculating γ is important. The counterfactual Husband/Father-in-law rank-rank parameter is 0.19 assuming no gender differences and 0.22 allowing for gender differences in inheritance, compared to the actual rank-rank parameter for women of 0.226 in the 1880-1900 period. The counterfactual change in persistence, -0.036 is 60 percent larger than the actual decline for men, and flips the sign for the change in women's rank-rank parameter, which increased by 0.016. In the absence of the increase in marital sorting over the two periods women would have continued to enjoy their higher intergenerational mobility of the earlier period, and men's mobility would not have converged with women's. The increase in marital sorting between 1850 and 1900 almost entirely counteracted the potential mobility improvements women would have experienced

if looking solely at their husbands' changes in economic standing.

Table 2: Marital Sorting and Counterfactual Women's Mobility

	1850-1870	1880-1900
Rank-rank parameter		
Husband/Father	0.231	0.209
Husband/Father-in-law	0.210	0.226
Strength of marital sorting (γ)		
No gender diffs	0.909	1.08
Espin-Sanchez et al. gender diffs	0.915	1.047
Counterfactual women's rank-rank		
No gender diffs		0.19
Espin-Sanchez et al. gender diffs		0.22

Notes: Rank-rank parameters are drawn from row (1) in Figure 2. γ measures the strength of marital sorting. For the no gender differences in inheritance row, γ is estimated as the ratio of the Husband/Father-in-law rank-rank parameter to the Husband/Father rank-rank parameter. We allow for gender differences in the strength of inheritance by using the father, mother, father-in-law, and mother-in-law specific estimates of inheritance from Espin-Sanchez et al. (2022) Table 2 and substituting in equation 7 and solving for γ.

8 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 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 other 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 starting from the mid-1840s. 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 mobility of females. 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 1880-1900 cohort might have been higher than observed if they faced the lower level of assortative mating of women in the earlier 1850-1870 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.

9 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. 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.

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 (2017, May). Up from Slavery? African American Intergenerational Economic Mobility Since 1880. NBER Working Papers 23395, National Bureau of Economic Research, Inc.

Collins, W. J. and A. Zimran (2018). The economic assimilation of irish famine migrants to the us. Unpublished manuscript.

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.

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. (1996). A new sample of males linked from the public use microdata sample of the 1850 u.s. federal census of population to the 1860 u.s. federal census manuscript schedules. *Historical Methods: A*

Journal of Quantitative and Interdisciplinary History 29(4), 141–156.

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.

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.

Hacker, D. J., L. R. Hilde, and J. H. Jones (2010). The effect of the civil war on southern marriage patterns. *Journal of Southern History 76*(1), 39–70.

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.

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.

Mazumder, B. and M. Acosta (2015). Using occupation to measure intergenerational mobility. *The ANNALS of the American Academy of Political and Social Science 657*(1), 174–193.

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.

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.

Salisbury, L. (2017). Women's income and marriage markets in the united states: Evidence from the civil war pension. *The Journal of Economic History 77*(1), 1–38.

Schoen, R. (1983). Measuring the tightness of a marriage squeeze. *Demography 20*(1), 61–78.

Solon, G. (1989, February). Biases in the Estimation of Intergenerational Earnings Correlations. *The Review of Economics and Statistics 71*(1), 172–174.

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.

Appendices

Table A1: 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)

Notes:
Sources:.

Table A2: 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)

Notes:
Sources:.

Table A3: 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)

Notes:
Sources:.

Table A4: Representativeness of linked sample to population of registered marriages: 1900

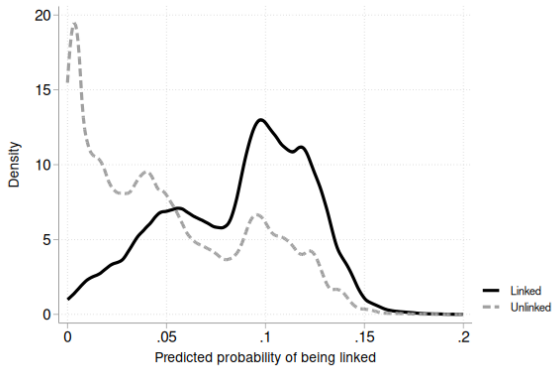
	Women			Men		
	Population	Linked unweighted	Linked weighted	Population	Linked unweighted	Linked weighted
Age	24.96	22.66	24.87	28.57	24.87	26.47
	(0.01)	(0.02)	(0.93)	(0.01)	(0.02)	(0.39)
Given name length	7.73	8.32	7.33	8.12	9.01	7.91
	(0.00)	(0.01)	(0.19)	(0.00)	(0.01)	(0.08)
Surname length	6.74	6.41	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)

Notes:
Sources:.

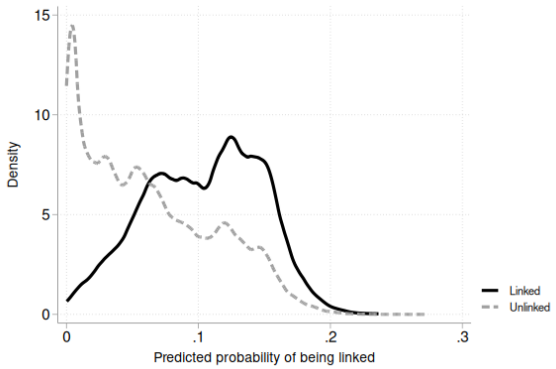
Table A5: Representativeness of linked sample to population of registered marriages: 1920

	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)

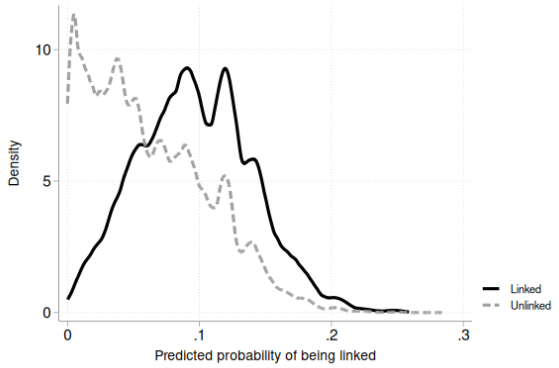
Notes:
Sources:.



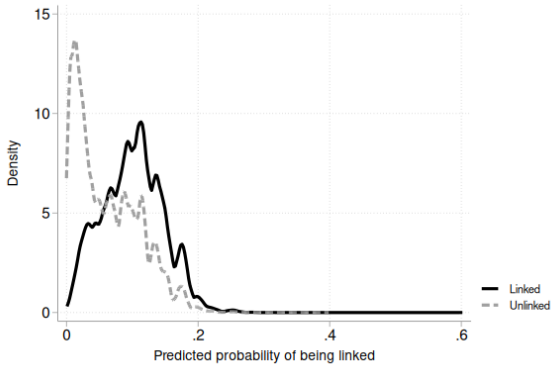
(a) 1850-1870 Marriages



(b) 1860-1880 Marriages

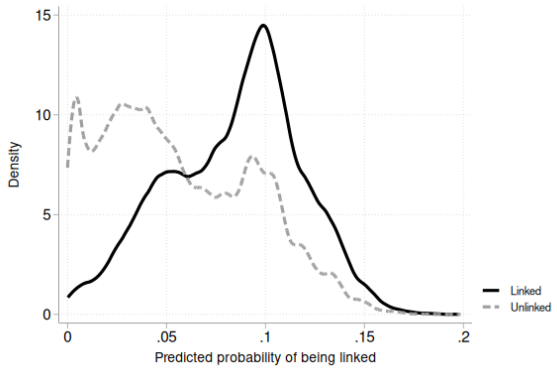


(c) 1880-1900 Marriages

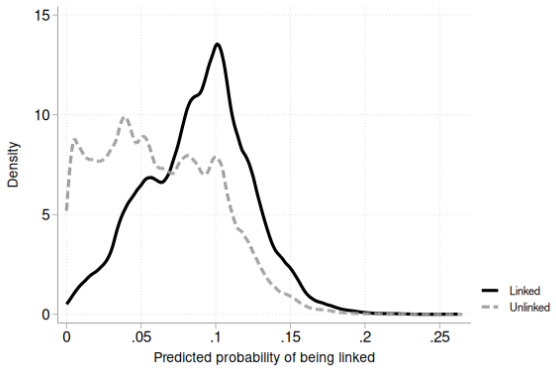


(d) 1900-1920 Marriages

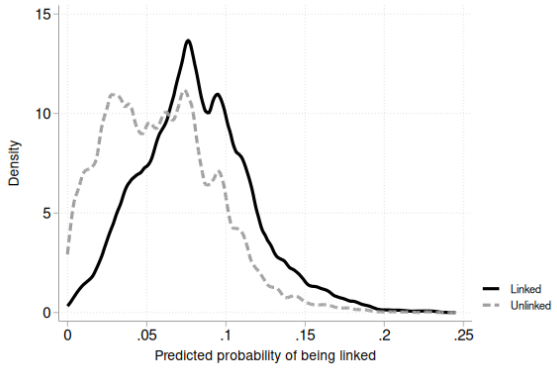
Figure A1: Men: Overlap of predicted probability of being in the linked sample by cohort



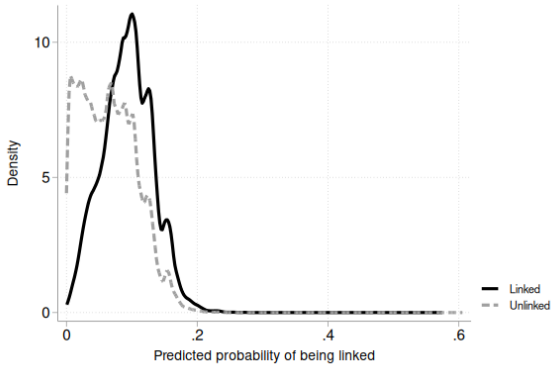
(a) 1850-1870 Marriages



(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 A6: 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:
Sources:.

Table A7: 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:
Sources:.

Table A8: 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:
Sources:.

Table A9: Representativeness of children in linked sample: 1880

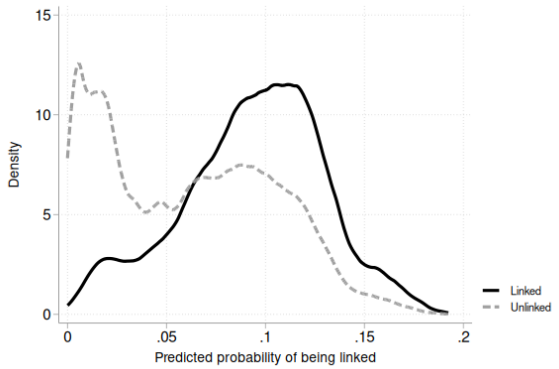
	Women			Men		
	Population	Linked unweighted	Linked weighted	Population	Linked unweighted	Linked weighted
Child age	8.58	10.17	8.36	8.63	11.53	8.10
	(0.01)	(0.04)	(0.07)	(0.01)	(0.04)	(0.14)
Urban	0.50	0.41	0.51	0.49	0.38	0.51
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)
Reside in MA	0.91	0.97	0.90	0.91	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:
Sources:.

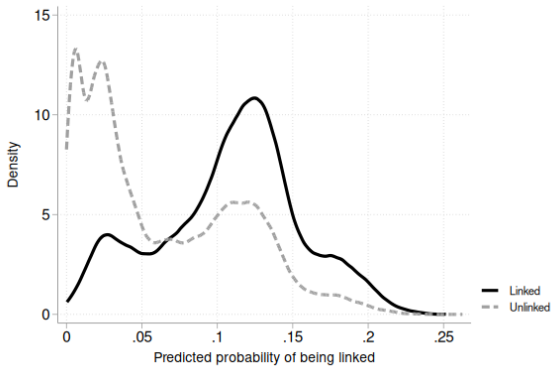
Table A10: Representativeness of children in linked sample: 1900

	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)

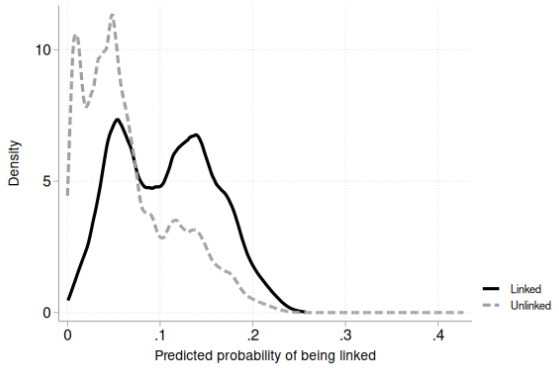
Notes:
Sources:.



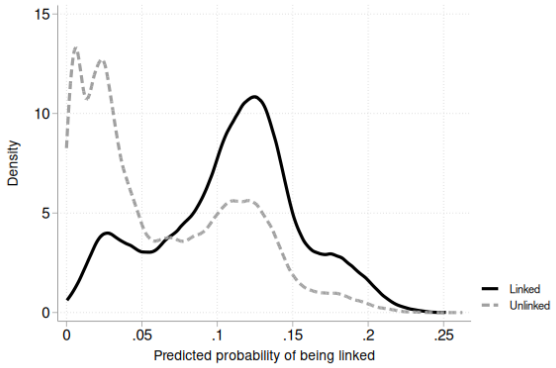
(a) 1850 Census



(b) 1860 Census



(c) 1880 Census



(d) 1900 Census

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