Marriage and the Intergenerational Mobility of Women: Evidence from Marriage Certificates 1850-1910

Jacqueline Craig and Katherine Eriksson and Gregory T. Niemesh

Link to Supplemental Online Appendix

Abstract

The literature finds a high degree of economic mobility for men in the 19th 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 1850-1910, we link men and women to their childhood and adult census records to obtain a measure of occupational standing across two generations. Intergenerational mobility for women is higher than for men during 1850-1880. Between 1880-1910, men's mobility increases to converge with that of women. We also find evidence of assortative mating based on the correlation in occupational income score and real estate wealth between the husband's and wife's fathers.

Craig: Research Assistant, Federal Reserve Bank of Chicago. Eriksson: Assistant Professor of Economics, University of California - Davis, and NBER. Niemesh: Assistant Professor of Economics, Miami University, and NBER. We would like to thank Charles Moul, Bhashkar Mazumder, Melissa Thomasson, and participants at Northwestern University's Seminar in Economic History for helpful suggestions. The views expressed in this paper are solely those of the authors and do not reflect those of the Federal Reserve Bank of Chicago or the Federal Reserve System.

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 19th-century, less so in the 20th. For a number of reasons, 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 19th-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.

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 19th-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 1910. Linked observations across marriage certificates and censuses are essentially to estimating intergenerational mobility of women during the 19th-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 name of married women and covers a large portion of a population.[1] 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).[2] Accordingly, we estimate intergenerational mobility by occupational transitions or 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 1910. 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 two cohorts (or periods), marriages between 1850-1880, and marriages that occurred between 1880-1910.

This double-match procedure is novel and provides a way forward to analyze female mobility, but it is not without drawbacks. The large amount of information recorded on marriage certificates

[1] Birth certificates frequently list the maiden name of the mother. However, the Birth Registration Area did not begin in earnest until the early 20th-century and compliance rates were low (Eriksson et al., 2018), eliminating their usefulness for record-linkage during the 19th 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.

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

and censuses requires the use of a probabilistic method to conduct the matching, in contrast to the historical intergenerational mobility literature which has primarily used automated linking procedures that rely on the exact matching of names of a male 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 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: occupational transitions, intergenerational elasticities, and rank-rank estimates. 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 1850-1880 cohort are more mobile than men. In other words, a female 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 0.248 for 1850-1880 and 0.181 for 1880-1910. For women, the estimates are 0.192 and 0.173 for those same cohorts. Persistence was greater for men in both cohorts, but that difference had disappeared by 1910.

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 0.161 for the 1850-1880 cohort and 0.214 in the 1880-1910 cohort, a 33 percent increase. In our sample, the likelihood that women would marry out of the economic class of their fathers decreased over time. Consequently, mobility of women in the 1880-1910 cohort would have been higher than observed if they faced the lower level of assortative mating of women in the earlier 1850-1880 cohort. Further research is required to better understand the forces driving the changes in assortative mating we find over the course of the 19th-century.

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 20th-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 19th- and mid-20th centuries faced less intergenerational persistence than the modern U.S. (Ferrie, 2005; Long and Ferrie, 2013; Feigenbaum, 2017).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 19th-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).[3]

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

These studies underscore the modern literature's focus on assortative mating based on personal labor market or human capital characteristics of the spouses. Unfortunately, the economic context of married women in the 19th-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 effect separate from sorting based on education. Charles et al. (2013) show a strong correlation in parental wealth between spouses

[3] 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.

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 19th-century when females had less opportunity for formal education, and the importance of education for labor market opportunities was less influential (Goldin, 1983)

3 Creating matched father-child pairs and measuring economic status

Our dataset consists of two cohorts of father-son and father-daughter pairs who are linked between two censuses: 1850 to 1880 for cohort 1, and 1880 to 1910 for cohort 2. The novel part of our strategy is to use the listed maiden name recorded on marriage certificates to link adult married women 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. The second match is completed separately for the husband and wife. The economic status of fathers and adult children is measured with the reported occupation on the census forms and an occupational income or wealth score.

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.[4] 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 two cohorts to facilitate linking census records between the complete count censuses available at the time. Cohort 1 consists of 208,026 registered marriages between 1850-1880 where both the spouses were born between 1830 and 1850. Cohort 2 includes marriages registered from 1880-1910 in which the couple was born between 1860-1880, giving a sample of 375,195 couples. The age limits are imposed to increase the likelihood of observing the adult in their childhood home in the census taken 30 years prior.

The couples in each cohort are first matched using names and ages to a subsequent U.S. decennial census to measure the post-marriage economic status of the couple. Cohort 1 marriages are linked to the 1880 census, while cohort 2 marriages are linked to the 1910 census. Each spouse is then linked to the census taken 30 years prior using names, ages, and names of parents. Cohort 1 marriages are linked to the 1850 census, and cohort 2 is linked to the 1880 census.[5] Figure 1 is helpful to visualize the stages of our double-match record linkage process.

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

[5] Household information is drawn from the 100% 1910 census sample deposited by Ancestry.com with the NBER, and the 1850 and 1880 100% samples from IPUMS (Ruggles et al., 2017).

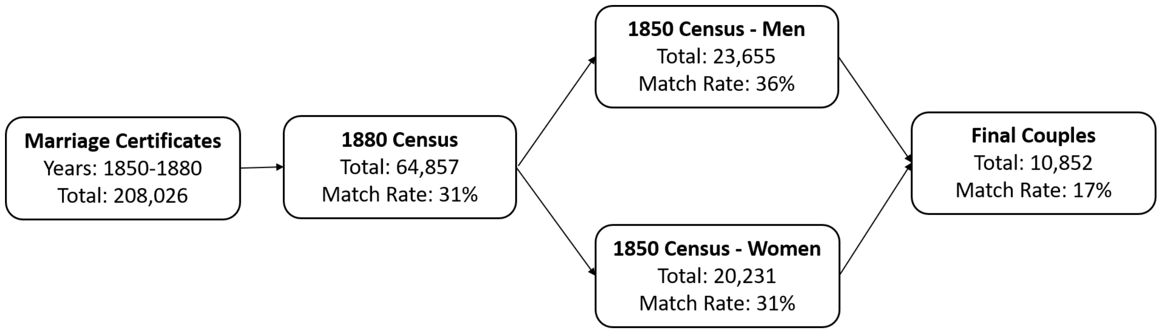
The large amount of information included on the marriage certificates requires the use of a probabilistic method to conduct a double match process. This is in contrast to the historical intergenerational mobility literature which has primarily used automated linking procedures that rely on the exact 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). In our dataset, for the first match of the married couple to a post-marriage census there are five dimensions over which to match. We have the surname, given names, and ages of the couple. The issue becomes how to balance small differences across each of these dimensions when deciding the correct the match. 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 namw and surname, father's given name, mother's given name – and the child's age and state of birth. The exact matching procedures are too conservative in our context of using a large number of strings; the methods lead 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. We briefly describe the matching procedure here with more detailed information provided in the supplemental appendix. The matching 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 probit, which is then cross-validated on the remainder of the true-links held out of the training dataset. The estimated model is then used to predict true matches for the remainder of the marriage certificate sample, for which we did not hand code "true" matches.

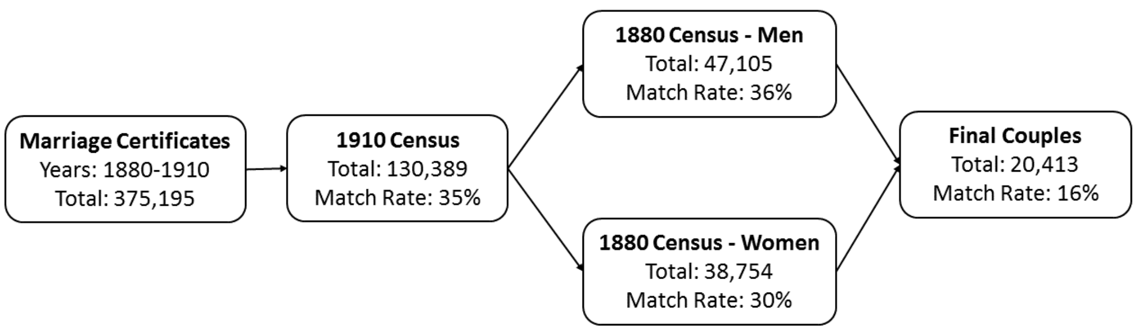
From the set of couples married in Massachusetts, we match 64,857 of the 208,026 couples in cohort 1, and 130,389 of the 375,195 couples in cohort 2 to a post-marriage census; match rates are 31% and 35%, respectively. Of the successfully matched couples to the post-marriage census, we link the husband and wife individually to their childhood household in a pre-marriage census. The final dataset includes 23,655 father-son pairs and 20,231 father-daughter pairs in cohort 1 (1850-1880), and 47,105 father-son pairs and 38,754 father-daughter pairs in cohort 2 (1880-1910). The match rates for the second match are between 30 and 36 percent. The overall match rates over both record linkages are between 9.7 and 12.6 percent. In the assortative mating analysis, we use a sample for which *both* the husband and wife from a single marriage certificate are linked successfully to their father. In this sample, we have 10,852 couples from cohort 1 and 20,413 couples from cohort 2.

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 matched certificates, and then characteristics from the adult census to the childhood census for



(a) Cohort 1: 1850-1880 Matching Results



(b) Cohort 2: 1880-1910 Matching Results

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

the second match. Results are shown in tables A2 and A3. The matched sample and full pool of marriage certificates are similar along a number of string characteristics, the age distribution, the and year of marriage distribution. Although the differences are highly statistically significant because of our large sample, none of the magnitudes of the differences appear meaningful. Repeating the process for the second match, we do find that our cohort 1 linked sample is slightly less urban and in higher ranked occupations as measured by the wealth score and occscore variables. The increased probability of successfully linking higher status observations is common in the literature (Abramitzky et al., 2012).

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. Table A4 reports the marginal effects from probit estimations for each cohort and gender. In general, an observation is more likely to be matched across all steps and enter the final data set when names are less common and less visually similar to other names. Longer husband given names lead to a *higher* match rate, whereas the length of the wife's given name and length of surnames are correlated with *lower* match rates. We also see from Table A4 that separately matching and estimating the inverse propensity weight for each group is important. Variables have different prediction power across the cohorts and genders, and in some cases the *sign* switches. For example, the length of parental given names is positively correlated with finding match for cohort 1, but has a negative association in cohort 2. In the end, the overall explanatory power of these variables to predict matches is quite small. We follow Feigenbaum (2017) in interpreting the low explanatory power as consistent with the randomness of mismatches from transcription errors by the original census enumerators and modern data indexers. In any case, 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).[6]

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 methods used in the economic history literature by 1.) using transitions between broad occupational categories, and 2.) assigning an occupational standing "score" to each occupation. This "score" can be based on either income or wealth. IPUMS census data sets assign each occupation in the historical censuses an *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 likely 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 19th century.

Instead, we follow Collins and Zimran (2018) and construct an occupational *wealth* score based on real and total property values reported in the 1870 Decennial Census 1% sample (Ruggles et al., 2017). We calculate the mean value of real property wealth and total property wealth (real plus personal) for all white male observations between the ages of 18 and 65 in each occupation and region cell. The national occupation mean is used in occupation-region cells that contain less than

[6]The propensity of being linked $P_i(L_i = 1|X_i)$ is calculated from the underlying probit coefficients of Table A4 of link status on the covariates listed X_i, with observations reweighted 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.

20 observations. 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. In response to these limitations, we also estimate mobility using broad occupational transitions which does not require the ranking of occupations (Altham and Ferrie, 2007; Ferrie, 2005; Long and Ferrie, 2007, 2013).

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 19th-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 cohort 1 are given a rank in the national distribution in 1850; adults of cohort 1 are given a rank in the national distribution in 1880, etc. 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 Intergenerational Mobility

4.1 Estimation

With our sample of matched father-son and father-daughter pairs across two cohorts of marriages (1850-79 and 1880-1910), we compare male and female intergenerational mobility within our sample, and also to other estimates in the literature from different time periods and source data. We construct a number of measures of mobility to aide in making these comparisons. In this section, we briefly discuss the estimation methods.

4.1.1 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.[7] 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

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

occupational income scores and 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:

$$\ln(Y^g_{c,i}) = \alpha + \beta^g \ln(Y^g_{f,i}) + X_i \Lambda + \epsilon_i \tag{1}$$

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 1 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 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 19th-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 1.

4.1.2 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.

4.1.3 Occupational Mobility

When using occupational scores to measure mobility, results can be highly dependent on where farmers are placed in the income or wealth distribution. In our context, the common use of the IPUMS *occscore* variable places farmers during periods where agriculture made up a large part of the workforce with large variation in farm sizes (1850, 1880, 1910) into the 1950 income distribution after much of the consolidation of agriculture had occurred.

In the absence of income or wealth data by occupation, sociologists and economic historians have used occupational transition matrices to capture the movement into or out of occupations by a son relative to a father (Ferrie, 2005; Altham and Ferrie, 2007; Long and Ferrie, 2013). Importantly, the use of occupational transition matrices do not require a ranking of occupations, and as such farmers do not need to be placed in a distribution. Altham and Ferrie (2007) present the Altham statistic as a way to easily capture the differences between two occupational transition matrices as a single number interpreted as a distance. Occupational mobility can then be measured between two time periods, places, or groups. Fathers and sons are grouped into one of four broad occupational categories: White-collar workers, Farmers, Semi-skilled workers, and Unskilled laborers. The Altham statistic measures the strength of association between the rows and columns between two transition matrices. Letting P and Q be two 4x4 transition matrices – son's occupation in rows and father's occupation in columns – where p_{ij} and q_{ij} are elements. Then the Altham statistic captures the degree of association between P and Q as:

$$d(P,Q) = \left\{ \sum_{i=1}^{4} \sum_{j=1}^{4} \sum_{l=1}^{4} \sum_{m=1}^{4} |log[(p_{ij}p_{lm}q_{im}q_{lj})/(p_{im}p_{lj}q_{ij}q_{lm})]| \right\}^{\frac{1}{2}} \tag{2}$$

The statistic ranges from zero when P and Q are identical in association between rows and columns and infinity; increases in the statistic imply more persistence in status and less mobility, decreases in the statistic imply less persistence and more mobility. We present both the occupational transition matrices and calculated Altham statistics for both men and women in the 1850-1880 and 1880-1910 cohorts, which we can than compare to historical intergenerational mobility estimates in the literature for other samples and time periods.

4.2 Mobility Results

Base results are presented in Table 1 for rank-rank regressions, "IGE" regressions, and occupational transitions using our new data 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.

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 1850-1880 between 0.225 and 0.274 across the three measures of economic status. In our sample, mobility increased from earlier to later cohort, with the rank-rank parameters declining to between 0.172 to 0.193, a statistically significant decline of about 24 to 30 percent relative to 1850-1880. We find a different pattern in mobility over time for our sample of linked women, depending on the measure of economic status used. The rank-rank parameters vary between 0.169 and 0.216 in the earlier period. Using wealth scores as outcomes, females experience a small statistically insignificant decrease of 3 to 10 percent of the earlier rank-rank parameter. Using the occscore measure, the parameter decreases by 29 percent.

Our linked sample of married women had a different mobility experience than men over the 19th-century. Women were much more mobile than men in the first 1850-1880 cohort. The difference was large and meaningful. Women's rank-rank parameter was 70 percent of that of men's in the 1850-1880 cohort – 0.173 vs. 0.248, a difference almost equivalent to the entire increase in mobility experienced by men between the two time periods. The rank-rank parameters converged between the sexes by the 1880-1910 cohort, given that mobility for men increased, whereas for women it did not. We find no statistically significantly or economically meaningful differences between intergenerational mobility for men and women in the later period when using the wealth scores as the measure of economic status. However, convergence is not yet complete in the later cohort when using occupational income scores. Women's rank-rank parameter remains 80 percent of that of men's.

Intergenerational elasticity estimates – panel B of table 1 – support the same interpretation as the rank-rank regressions. Men's economic status was more persistent than women's in the 1850-1880 sample, but had mostly converged by the 1880-1910 sample. Similar results between the two specifications is not surprising given that the "IGE" estimates are from a log-log regression on occupational income scores, which are occupation cell income medians. Applying the same score to each observation within an occupation apparently removes much of the non-linearity between father and son in the tails fo the distribution, reducing the need for the rank-rank specification (Corak and Heisz, 1999; Chetty et al., 2014).

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.[8] 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 19th-century, with full convergence between the sexes in a linked 1900-1920 sample. They do not use a sample with 1910 as an endpoint, making direct comparisons somewhat

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

Table 1: Intergenerational Mobility: Base results

	Men			Women			Male-Female Diff	
	Cohort 1 1850-1880	Cohort 2 1880-1910	Diff Cohort 1-2	Cohort 1 1850-1880	Cohort 2 1880-1910	Diff Cohort 1-2	Cohort 1 1850-1880	Cohort2 1880-1910
Panel A: Rank-Rank estimates								
Total property wealth score	0.248	0.181	-0.066	0.192	0.173	-0.020	-0.051	-0.010
	(0.008)	(0.008)	(0.010)	(0.012)	(0.012)	(0.015)	(0.015)	(0.010)
Real property wealth score	0.225	0.172	-0.053	0.169	0.165	-0.005	-0.053	-0.009
	(0.008)	(0.008)	(0.010)	(0.012)	(0.012)	(0.015)	(0.014)	(0.010)
Occupational income score	0.274	0.193	-0.082	0.216	0.154	-0.062	-0.056	-0.038
	(0.009)	(0.009)	(0.011)	(0.013)	(0.013)	(0.016)	(0.016)	(0.011)
Panel B: Intergenerational elasticity estimates "IGE"								
Total property wealth score	0.242	0.202	-0.040	0.187	0.193	0.005	-0.051	-0.012
	(0.007)	(0.007)	(0.009)	(0.009)	(0.009)	(0.011)	(0.011)	(0.008)
Real property wealth score	0.229	0.186	-0.043	0.173	0.175	0.002	-0.053	-0.014
	(0.007)	(0.007)	(0.009)	(0.009)	(0.009)	(0.011)	(0.011)	(0.009)
Occupational income score	0.276	0.208	-0.069	0.222	0.172	-0.050	-0.054	-0.036
	(0.008)	(0.008)	(0.010)	(0.012)	(0.012)	(0.015)	(0.015)	(0.011)
Panel C: Occupational mobility estimates								
Altham statistic	16.18	13.54	5.36	12.94	11.06	4.47	3.87	3.25

Notes: Heteroskedasticity robust standard errors reported in parentheses. Each entry is an estimate of intergenerational mobility from separate regressions using different measures of economic status and for different samples and time periods. 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. Panel C calculates the Altham statistic for occupational transitions between father and child (husband in the case of daughters) (Altham and Ferrie, 2007). The Altham statistic increases as mobility decreases in the economy. The estimates give the distance to independence for a single sample and cohort $d(P, I)$, and the distance between cohorts or genders when calculated for the difference between groups $d(P, Q)$. *Sources:* 1870 1% sample and complete count 1850, 1880, and 1910 Federal Census data from Ruggles et al. (2017). Marriage certificates from *FamilySearch.org*.

difficult.

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. 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 the appendix, we discuss at length results from applying the pseudo-linking procedure of Olivetti and Paserman (2015) to our direct-linked sample with a full set of estimates in Appendix Table A1. 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.

Mobility across broad occupational categories tells a similar story to the regression estimates. We focus on interpreting the Altham statistics, however a full set of occupational transition tables can be found in Appendix Table A8. Panel (c) of Table 1 presents Altham statistics for each sex and cohort relative to independence, between genders, and across cohorts. All estimates are statistically different from "no association" at the 1 percent significance level using the test outlined in Altham and Ferrie (2007). Occupational transitions support the same general pattern as the other measures of mobility. In both cohorts, women were able to marry into an occupational class different from their father's more than men were able to work in an occupational class different from their father's. Mobility for both men and women increased between the 1850-80 and 1880-1900 cohorts, more so for men than for women. In this case, we do not see the full convergence between men and women in the 1880-1910 cohort as we saw using the rank-rank regressions.

To put our estimates into context with the literature, Figure 2 plots the Altham statistics from our samples against those found in the literature. Overall, the historical and modern literature find increasing persistence and decreasing mobility over time. Here, we've plotted estimates from Long and Ferrie (2007) for the United States as a whole, and from Feigenbaum (2017) for mid-century Iowa to fill in the gap. In the national samples, each successive cohort faces an economy that delivers decreasing opportunity for sons to transition out of the broad occupational category of their father. In contrast, our estimates for men from the Massachusetts sample suggest the opposite trend, *increasing* mobility over the 19th-century. The difference in results is likely driven by the differences in the underlying sample of father-son pairs each study is meant to follow.

To summarize, all three 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 1850-1880 sample. Both sexes experienced increased in mobility, with males seeing a larger increase. Whether the sex differences fully converged by 1910 varies across specifications and measures of status. Our preferred estimates are from the rank-rank regressions using the total property

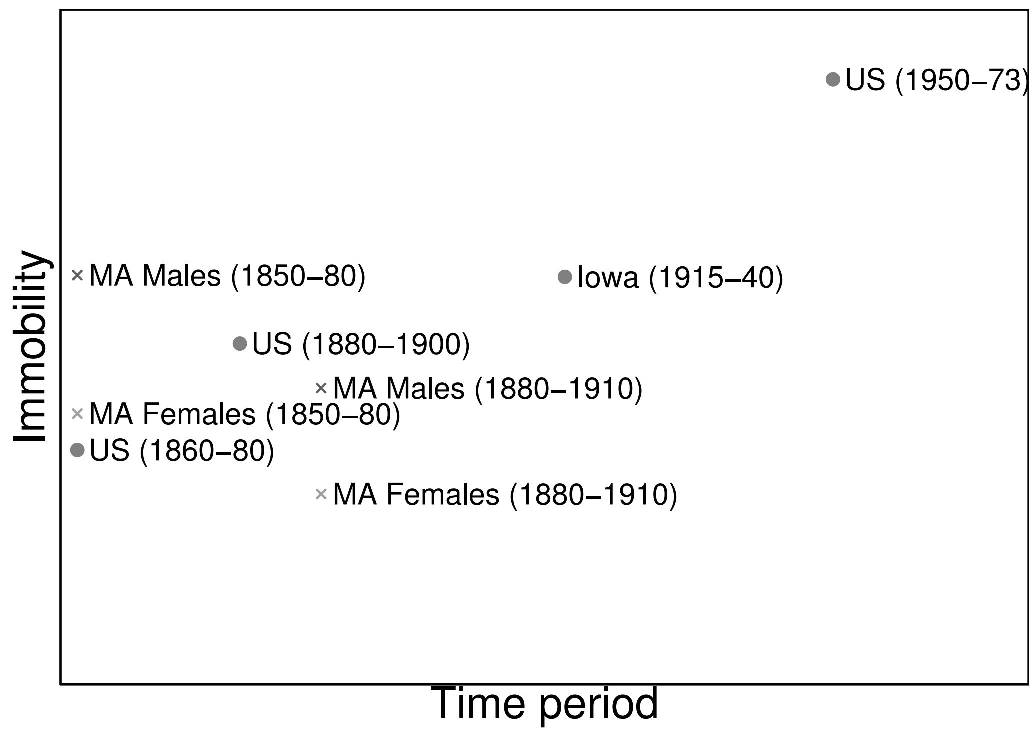


Figure 2: Occupational Mobility in the United States Over Time

Notes: Each point represents an Altham statistic of association between a father-child pair's occupations and independence. Points are plotted in the year when the child's economic status is observed as an adult. Estimates for Massachusetts (MA) come directly from Table 1. Estimates for the US come from Long and Ferrie (2007), and those for Iowa are from Feigenbaum (2017).

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 19th-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.

4.2.1 Subgroup Analysis

To attempt to describe the underlying factors behind this high level of mobility experienced, 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 panel (A) of Table

2. We construct an indicator equal to one if the observation resided in a rural area in the childhood census. Urban is defined as towns and incorporated places with at least 2,500 in population, and rural is defined as the complement.[9] We interact the rural indicator with the father's percentile rank in the total wealth distribution of his cohort, and its main effect.

Coefficients in the first line provide the rank-rank parameter for urban children. The main effect of rural suggests that men and women with rural childhoods end up with a lower occupational standing as an adult, although these results are not strong across cohorts and sex. Within the rural children however, there is weakly suggestive evidence of more persistence across generations than for urban children. These results are in contrast to modern findings. 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. Additionally, Page and Solon (2003) estimates the correlation of earnings among siblings after 1968 and finds that the majority of the correlation can be attributed to family effects rather than neighborhood effects during childhood. Our results support findings that growing up in an urban vs. rural area does not necessarily impact mobility levels in the 19th-century U.S.

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 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 effect 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.

[9] 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."*

Table 2: Intergenerational Mobility: Subgroup Analysis (Rank-Rank regressions)

	Men		Women	
	Cohort 1 1850-1880	Cohort 2 1880-1910	Cohort 1 1850-1880	Cohort 2 1880-1910
Panel A: Urban vs. Rural Childhood				
Father's total property wealth score	0.257***	0.169***	0.172***	0.163***
	(0.013)	(0.007)	(0.019)	(0.011)
Rural childhood	0.321	-1.365**	-0.816	-0.304
	(0.892)	(0.568)	(1.314)	(0.886)
Fthr's wealth * Rural	-0.017	0.031**	0.034	0.026
	(0.017)	(0.011)	(0.025)	(0.016)
Panel B: Immigrant vs. Native-born Parents				
Father's total property wealth score	0.238***	0.156***	0.188***	0.149***
	(0.009)	(0.007)	(0.012)	(0.010)
Immigrant parent	-4.501***	-4.682***	-3.291	-4.955***
	(1.051)	(0.572)	(2.112)	(0.903)
Fthr's wealth * Immigrant parent	-0.069**	0.003	-0.034	-0.003
	(0.027)	(0.011)	(0.047)	(0.019)
Panel C: Internal Migration				
Father's total property wealth score	0.274***	0.205***	0.200***	0.182***
	(0.010)	(0.006)	(0.014)	(0.009)
Internal migrant	3.402***	2.514***	3.043**	3.401**
	(0.984)	(0.661)	(1.505)	(1.077)
Fthr's wealth * Internal migrant	-0.092***	-0.082***	-0.039	-0.041**
	(0.018)	(0.012)	(0.027)	(0.019)
Panel D: First-born child				
Father's total property wealth score	0.246***	0.179***	0.195***	0.178***
	(0.010)	(0.007)	(0.014)	(0.010)
First-born child	0.976	1.374**	1.239	2.330**
	(0.930)	(0.610)	(1.462)	(0.966)
Fthr's wealth * first-born	0.003	0.004	-0.008	-0.019
	(0.018)	(0.011)	(0.028)	(0.018)
Panel E: Family size				
Father's total property wealth score	0.204***	0.123***	0.160***	0.129***
	(0.029)	(0.016)	(0.042)	(0.023)
No. of siblings	-2.576	-3.302**	1.040	-0.466
	(1.821)	(1.103)	(2.487)	(1.614)
Fthr's wealth * Siblings	0.080**	0.055**	0.005	0.018
	(0.035)	(0.019)	(0.050)	(0.029)

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. 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 (1880 and 1910). Internal migrant is defined as observing an observation in two different states in the childhood census and adult census. First-born is an indicator equal to one if the observation is the oldest child listed in the household. Number of siblings is the total number of children of the head of household listed in the census minus one for the observation at hand.

Sources: 1870 1% sample and complete count 1850, 1880, and 1910 Federal Census data from Ruggles et al. (2017). Marriage certificates from *FamilySearch.org*.

Again, we interact an indicator with father's rank in the total wealth distribution, this time equal to one if either parent was reported as foreign-born in the adult census. Both the 1880 and 1910 censuses asked about parent's place of birth, and we use this information here. The main effect of foreign parentage captures the level difference in adult incomes. Clearly, the sons and daughters of immigrants were of lower economic status as adults compared to children of native-born parents as can be seen in panel (B) of 2. The interaction captures the difference in mobility between children of foreign-born and native-born parents. Surprisingly, we find very weak evidence to suggest any mobility differences between these groups. Second generation immigrant men were more mobile than their native-born counterparts in the 1850-1880 cohort, but by the 1880-1910 cohort, the difference had disappeared. The convergence between the two groups was driven almost entirely by reductions in persistence for the sons of native-born parents. The mobility levels for second generation immigrant men remained largely the same between the cohorts.

Internal Migration

The 19th century was both an economically and geographically mobile period for the United States (Hall and Ruggles, 2004; Rosenbloom and Sundstrom, 2003; 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. We interact this indicator with father's rank in the total wealth distribution and include its main effect. 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 an important source of mobility for our sample. The interaction coefficient is statistically significant in three out of four specifications. Men who internally migrated faced persistence levels that were 33 to 40 percent lower than men that did not migrate across state lines. The effect is notably smaller for women. Importantly, we also see from the main effect of internal migration that men and women who chose to move had higher economic standing as adults, on average.

5 The Marriage Market and Assortative Mating

Married women in the 19th-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 3 captures the notion of assortative mating by regressing the husband's father's occupational standing on the wife's father's occupational standing.

$$Y_{f,i}^{husband} = \alpha + \beta Y_{f,i}^{wife} + X_i \Lambda + \epsilon_i, \tag{3}$$

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.

Table 3 presents our main results on the Massachusetts marriage market in the late 19th-century. 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 mixed, however, about the trend in within group homogamy. In our preferred rank-rank specification using the total wealth score measure, the rank-rank parameter was 0.161 for the 1850-1880 cohort and 0.214 in the 1880-1910 cohort. An increase in assortative mating is also found when using the real wealth score measure. The 1950 occupational income score, however, gives estimates of relatively high assortative mating in the early cohort, and levels of sorting in the later cohort. Finally, the Altham statistics that capture sorting across broad occupational categories suggest a significant amount of assortative mating, with a statistically significant change between the two cohorts. However, the Altham statistic makes it somewhat difficult to determine whether sorting increased or decreased. To our knowledge, the literature does not contain assortative mating estimates of this type in which to compare our own.

The online appendix discusses results from a subgroup analysis as well as presents figures from upward and downward rank mobility calculations.

Further work is necessary to understand the causes of the changing pattern of assortative mating over the 19th-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 in the 19th-century looks at re-marriage choices of Union Army widows based on economic standing and pension income (Salisbury, 2017).

6 Conclusion

This paper presents new evidence of the intergenerational mobility for men and women married in Massachusetts between 1850 and 1910. Consistent with the historical literature we find high levels of mobility in the 19th- and early 20th-century relative to the second half of the 20th-century

Table 3: Assortative Mating: Main Results

	Cohort 1 1850-1880	Cohort 2 1880-1910	Diff Cohort 2-1
Panel A: Rank-Rank estimates			
Total wealth score	0.161	0.214	0.054
	(0.012)	(0.009)	(0.014)
Real wealth score	0.148	0.191	0.043
	(0.012)	(0.009)	(0.014)
Occupation score	0.187	0.167	-0.022
	(0.011)	(0.008)	(0.014)
Panel B: Elasticity estimates (Log-Log)			
Total wealth score	0.169	0.211	0.044
	(0.012)	(0.009)	(0.015)
Real wealth score	0.165	0.191	0.027
	(0.012)	(0.009)	(0.014)
Occupation score	0.213	0.193	-0.023
	(0.012)	(0.009)	(0.015)
Panel C: Occupational mobility estimates			
Altham statistic	12.99	12.21	3.35

Notes: Heteroskedasticity robust standard errors reported in parentheses. 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. Panel C calculates the Altham statistic for occupational transitions between wife's father and husband's father (Altham and Ferrie, 2007). The estimates give the distance to independence for a single sample and cohort $d(P, I)$, and the distance between cohorts or genders when calculated for the difference between groups $d(P, Q)$.
Sources: 1870 1% sample and complete count 1850, 1880, and 1910 Federal Census data from Ruggles et al. (2017). Marriage certificates from *FamilySearch.org.*

(Long and Ferrie, 2013; Feigenbaum, 2017; Chetty et al., 2014; Solon, 1992). The mobility estimates based on income scores, wealth scores, and broad occupational group transitions 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 19th-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 19th 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 19th-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 19th-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 19th-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 1880-1910 cohort might have been higher than observed if they faced the lower level of assortative mating of women in the earlier 1850-1880 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 1850-1880 cohort. However, in the 1880-1910 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 19th-century.

7 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.

Altham, P. and J. Ferrie (2007). Comparing contingency tables: Tools for analyzing data from two groups cross-classified by two characteristics. *Historical Methods: A Journal of Quantitative and Interdisciplinary History 40*, 3–17.

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.

Bhattacharya, D. and B. Mazumder (2011, November). A nonparametric analysis of black–white differences in intergenerational income mobility in the United States. *Quantitative Economics 2*(3), 335–379.

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.

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.

Corak, M., M. J. Lindquist, and B. Mazumder (2014). A comparison of upward and downward intergenerational mobility in Canada, Sweden and the United States. *Labour Economics 30*(C), 185–200.

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 Forthcoming*.

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

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.

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* (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.

Page, M. E. and G. Solon (2003). Correlations between brothers and neighboring boys in their adult earnings: The importance of being urban. *Journal of Labor Economics 21*(4), 831–855.

Rosenbloom, J. L. and W. A. Sundstrom (2003). The decline and rise of interstate migration in the united states: Evidence from the ipums, 1850-1990. NBER Working Papers 9857, National Bureau of Economic Research, Inc.

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.